# NELSON VICmaths

VCE UNITS 3 + 4

Dirk Strasser
Greg Neal
Contributing author:
Neale Woods

# general mathematics 12

Nelson

Nelson VICmaths General Mathematics 12
1st Edition
Dirk Strasser
Greg Neal
ISBN 9780170448406

Publishers: Dirk Strasser, Robert Yen
Project editor: Tanya Smith
Editors: Karen Chin, Liz Waud
Series text design: Alba Design (Rina Gargano)
Series cover design: Watershed Art & Design (Leigh Ashforth)
Series designer: Nikita Bansal
Permissions researcher: Liz McShane
Production controller: Karen Young
Typeset by: Nikki M Group Pty Ltd

Any URLs contained in this publication were checked for currency during the production process. Note, however, that the publisher cannot vouch for the ongoing currency of URLs.

Acknowledgements
Selected VCE Examination questions and extracts from the VCE Study Designs are copyright Victorian Curriculum and Assessment Authority (VCAA), reproduced by permission. VCE® is a registered trademark of the VCAA. The VCAA does not endorse this product and makes no warranties regarding the correctness or accuracy of this study resource. To the extent permitted by law, the VCAA excludes all liability for any loss or damage suffered or incurred as a result of accessing, using or relying on the content Current VCE Study Designs, past VCE exams and related content can be accessed directly at www.vcaa.vicedu.au.

TI-Nspire: Images used with permission by Texas Instruments, Inc
Casio ClassPad: Images used with permission by Shriro Australia Pty. Ltd.



National Library of Australia Cataloguing-in-Publication Data
A catalogue record for this book is available from the National Library of Australia.

Cengage Learning Australia
Level 7, 80 Dorcas Street
South Melbourne, Victoria Australia 3205

Cengage Learning New Zealand
Unit4B Rosedale Office Park
331 Rosedale Road, Albany, North Shore 0632, NZ

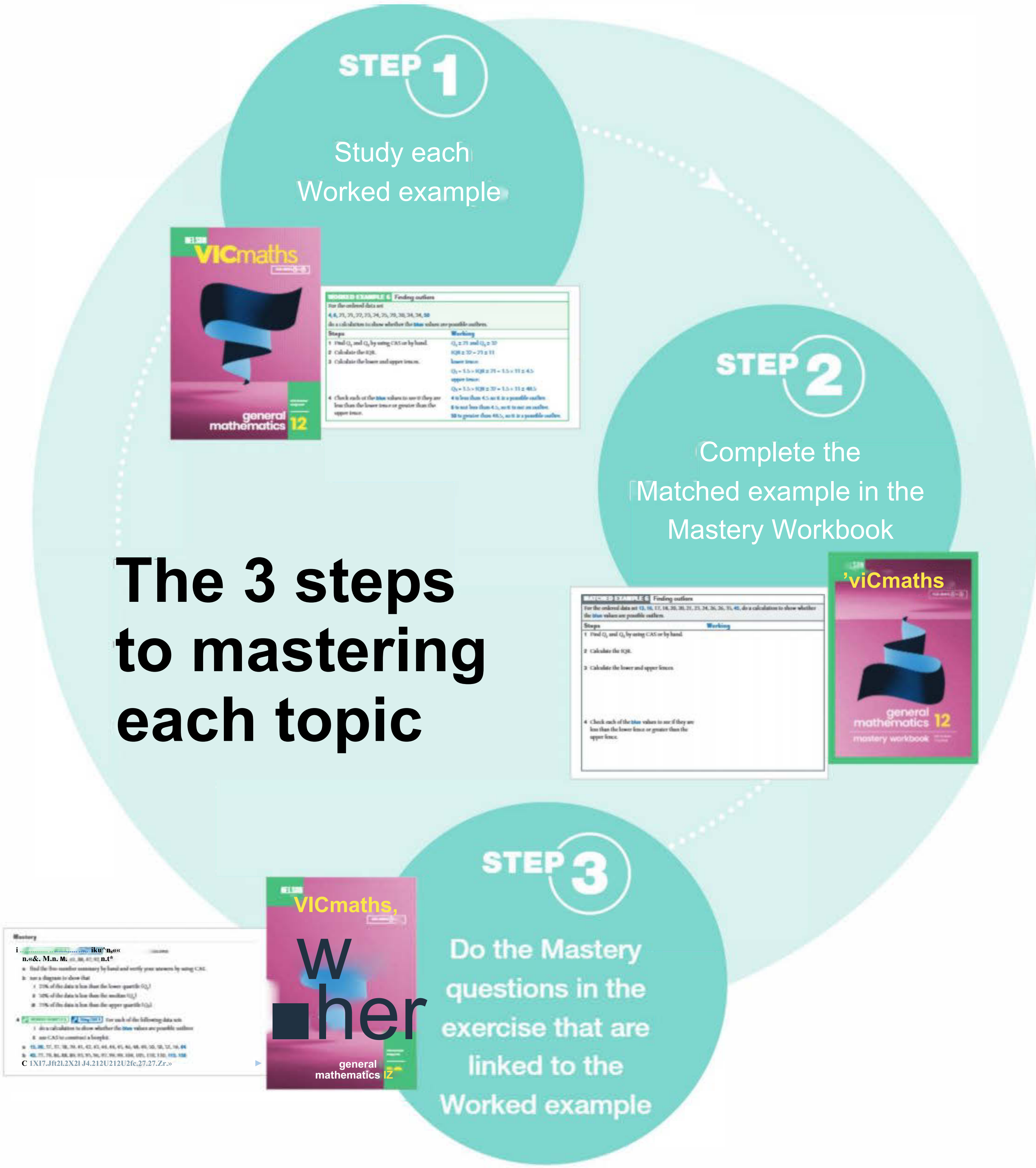
For learning solutions, visit cengage.com.au

Printed in China by 1010 Printing International Limited.
1 2 3 4 5 6 7 25 24 23 22 21

# Contents

# To the student

*Nelson VICmaths* is your best friend when it comes to studying General Mathematics in Year 12. It has been written to help you maximise your learning and success this year. Every explanation, every exam hack and every worked example has been written with the exams in mind.



# To the teacher

Now there's a better way to VCE maths mastery.

*Nelson VICmaths 11—12* is a new VCE mathematics series that is backed by research into the science of learning. The design and structure of the series has been informed by teacher advice and evidence-based pedagogy, with the focus on preparing VCE students for their exams and maximising their learning achievement.

- Using backwards learning design, this series has been built by analysing past VCE exam questions and ensuring that all theory and examples are precisely mapped to the VCE Study Design.
- To reduce the cognitive load for learners, explanations are clear and concise, using the technique of chunking text with accompanying diagrams and infographics.
- The student book and workbook combination has been designed for mastery of the learning content.
- The exercise structure of Recap, Mastery and Exam practice leads students from procedural fluency to higher-order thinking using the learning technique of interleaving.
- Exam practice includes exam-style questions and graded past VCE exam questions with success percentages based on VCAA performance data.
- The cumulative structure of Exercise Recaps and chapter-based Cumulative examinations is built on the learning and memory techniques of spacing and retrieval.

# About the authors

Dirk Strasser is an experienced teacher, a former Head of Mathematics and a lead author and senior publisher of mathematics series for over 30 years. He has published and co-written eight best-selling mathematics series for Heinemann and Pearson, and won several Australian Educational Publishing Awards. He is the Manager of Secondary Mathematics at Nelson Cengage.

Greg Neal has been teaching for over 40 years and was Head of Mathematics at Ballarat High School. He has been a VCAA examination assessor for Mathematical Methods for over 20 years and has expertise with CAS technology. Greg has written a range of Nelson resources for Further Mathematics, Mathematical Methods and Specialist Mathematics. Greg presents at MAV conferences and has worked recently as a senior mathematics tutor at Phoenix College, Ballarat.

## Contributing authors

Neale Woods coordinated the *Using CAS* content. He is a retired mathematics teacher with over 45 years of teaching experience. Neale is a CAS specialist who has presented at numerous conferences over the years and continues to conduct technology workshops for teachers and students.

Darren Smyth (maffsguru.com) created the VCE question analysis videos.

# Study Design grid

| Area of study | *Nelson VICmaths General Mathematics 12* chapter | |
|---|---|---|
| 1 Data analysis, probability and statistics | | |
| Investigating data distributions | 1 | Data distributions |
| Investigating association between two variables | 2 | Associations between two variables |
| Investigating and modelling linear associations | 3 | Linear associations |
| Investigating and modelling time series data | 4 | Time series |
| 2 Discrete mathematics | | |
| Depreciation of assets | 5 | Interest and depreciation |
| Compound interest investments and loans | 5 | Interest and depreciation |
| Reducing balance loans | 6 | Loans, investments and finance solvers |
| Annuities and perpetuities | 6 | Loans, investments and finance solvers |
| Compound interest investment with periodic and equal additions to the principal | 6 | Loans, investments and finance solvers |
| Matrices and their applications | 7 | Matrices and their applications |
| Transition matrices | | Transition matrices |
| Graphs and networks | 9 | Undirected graphs |
| Exploring and travelling problems | 9 | Undirected graphs |
| Trees and minimum connector problems | 9 | Undirected graphs |
| Shortest path problems | 9 | Undirected graphs |
| Graphs and networks (digraphs) | 10 | Directed graphs |
| Flow problems | 10 | Directed graphs |
| Shortest path problems (digraphs) | 10 | Directed graphs |
| Matching problems | 10 | Directed graphs |
| Scheduling problems and critical path analysis | 10 | Directed graphs |

# About this book

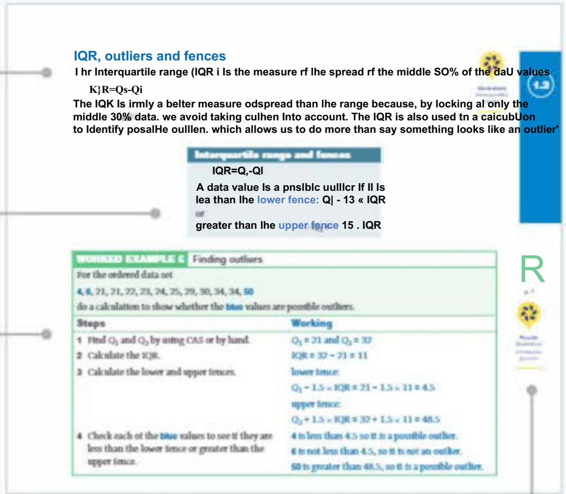
## In each chapter

Study Design coverage and extracts are shown at the start of the chapter, along with a listing of Nelson MindTap chapter resources.

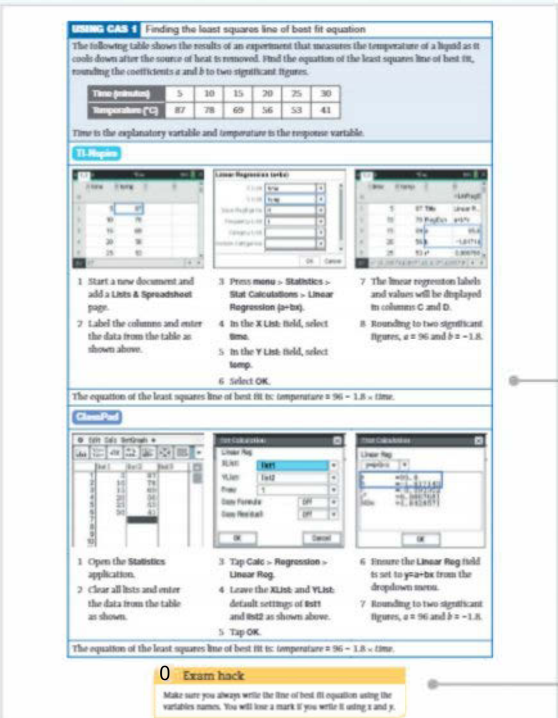

Important words and phrases are printed in blue and listed in the Glossary and index at the back of the book.

Important facts and formulas are highlighted in a shaded box.

Worked examples are explained clearly step-by-step, with the mathematical working shown on the right-hand side.



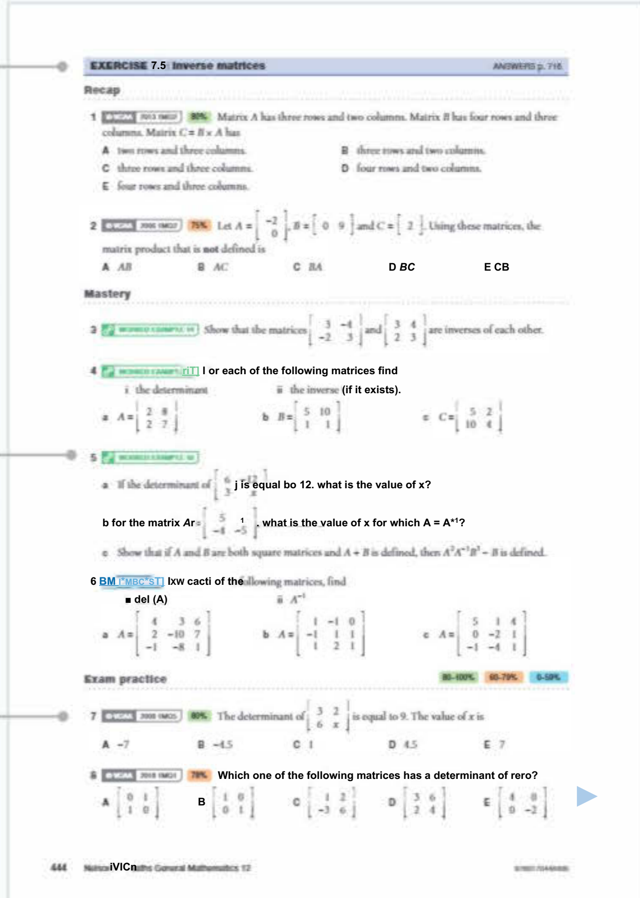

Links to scaffolded Matched examples in the Mastery Workbook (WB).




Using CAS provides clear instructions for TI-Nspire and Casio ClassPad calculators.

Exam hacks highlight valuable exam hints and common student errors.

Graded exercises, including Recap, Mastery and Exam practice questions, are linked to worked examples and Using CAS, and include past VCE exam and exam-style questions.




Recap questions revise skills from the previous exercise, Mastery questions provide skill practice, while Exam practice applies learned skills to VCE exam and exam-style problems.

©VCAA 2008 1MQ5 80%

Past VCE exam questions are clearly tagged and graded by colour-coded success percentages based on VCAA data, presented in the order: Exam 1, Exam 2.

KEY

| | |
|---|---|
| 2021 | 2021 exam year |
| <year>N | Northern hemisphere exam |
| <year>S | VCAA sample exam |
| 1 | Exam 1 |
| 2 | Exam 2 |
| C | Core |
| M | Matrices |
| N | Networks and decision mathematics |
| BRM | Business-related mathematics |
| NP | Number patterns |

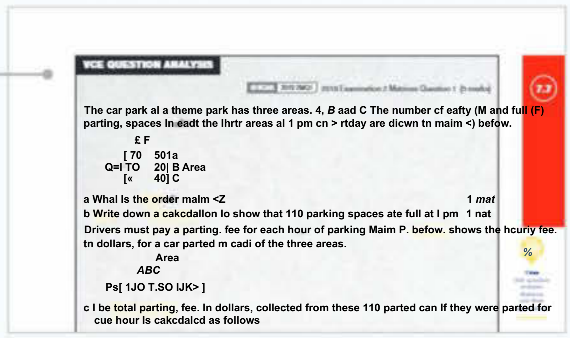
For multiple-choice questions and other 1 -mark questions, the success percentage is the percentage of students who answered correctly. For questions worth more than 1 mark, the success percentage is the mean percentage result scored by students.

| Success percentage range | Example |
|---|---|
| 80-100% (straightforward) | 83% |
| 60-79% (standard) | 62% |
| 0-59% (complex) | 45% |

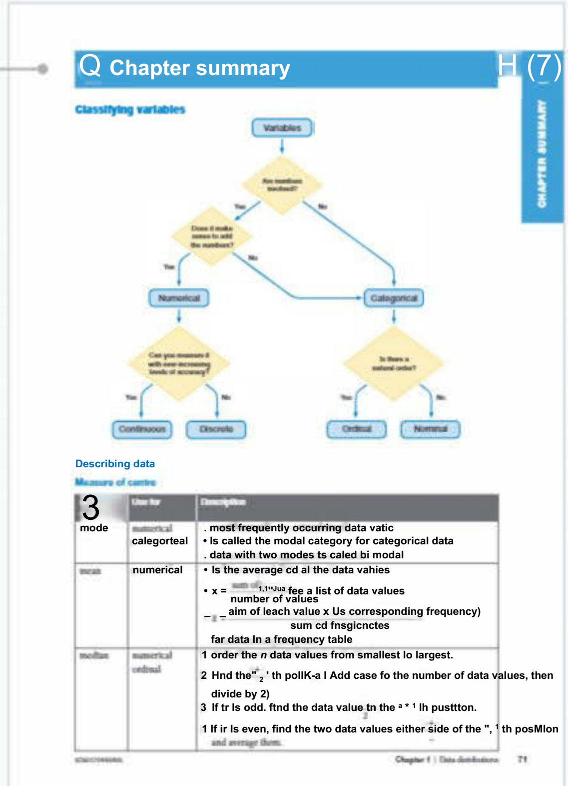
Note: Questions from Northern hemisphere and VCAA sample exams do not have success percentages.

# At the end of each chapter

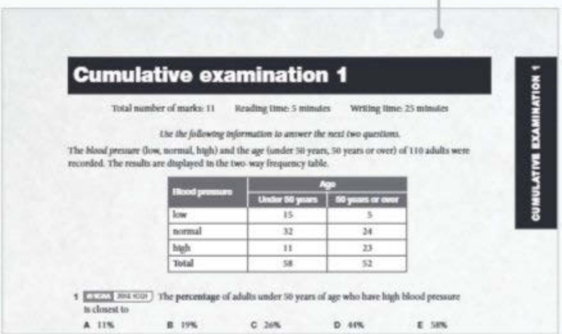
VCE question analysis leads students through a past VCE exam question that exemplifies the chapter, discussing how to approach the question, providing advice on interpreting the question, common student errors and statistics on student performance.



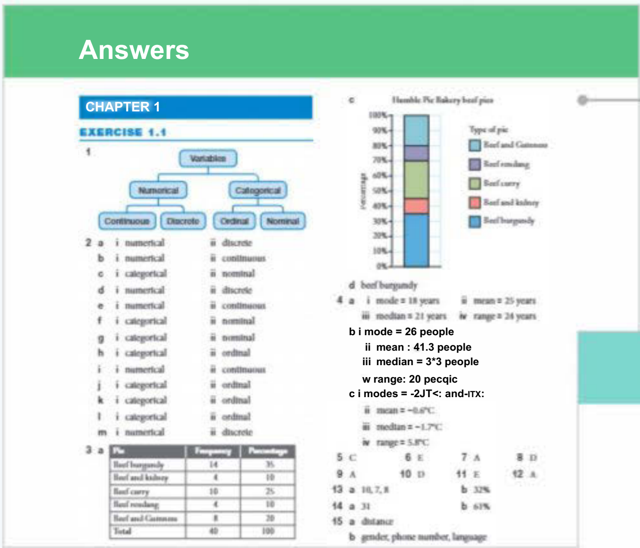
Chapter summary for easy reference




Cumulative examinations 1 and 2 are mini-exams based on the format of the VCE examinations 1 and 2, with around 50% of questions focusing on the chapter in which they appear.




# At the end of the book




Answers (with worked solutions provided on Nelson MindTap for teachers to allocate to students).


A combined Glossary and index.

# Nelson MindTap

Nelson MindTap is an online learning space that provides students with tailored learning experiences. Access tools and content that make learning simpler yet smarter to help you achieve VCE maths mastery.

Nelson MindTap includes an eText with integrated interactives and online assessment.

Margin links in the student book signpost multimedia student resources found on MindTap.

Video playlists

Worksheets

Skillsheets

Nelson MindTap for students:

- Watch video tutorials featuring expert teacher advice to unpack new concepts and develop your understanding.
- Revise using quizzes, worksheets and skillsheets to practise your skills and build your confidence.
- Navigate your own path, accessing the content, analytics and support as you need it.

Nelson MindTap for teachers*:

- Tailor content to different learning needs - assign directly to the student, or the whole class.
- Monitor progress using assessment tools like Gradebook and Reports.
- Integrate content and assessment directly within your schools LMS for ease of access.
- Access topic tests, teaching plans and worked solutions to each exercise set.

*Complimentary access to these resources is only available to teachers who use this book as part of a class set, book hire or booklist. Contact your Cengage Education Consultant for information about access and conditions.

# Nelson VICmaths 11-12 series


# Companion resources

WB

## Mastery Workbook

- Step-by-step scaffolding to guide students towards mastery of the course content.
- Write-in support that encourages students to show working.
- Matched examples for *every* worked example in the student book.
- Full integration between student book and workbook.
- Answers (worked solutions for teachers on Nelson MindTap)

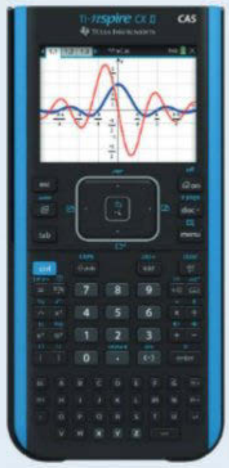
## ExampZus

- Create and simulate exam-like conditions in minutes.
- Save time with an extensive bank of filterable and difficulty-graded questions for Year 12.
- Over 1000 past VCAA and unseen exam-style questions and solutions all in the one place.
- Extensively researched and user-tested.

# TI-Nspire CAS introduction

The latest TI-Nspire model is TI-Nspire CXII CAS. When purchasing a new handheld, you also gain access to the student software. If you purchase a used handheld, then you can pay an additional fee for the student software. Alternatively, you can connect your handheld to your computer using the TI-Nspire™ CX II Connect web-based app, which enables you to perform a variety of functions such as screen captures, file transfers and operating system updates.

Note that TI-Nspire non-CAS technology is also available. It is vital that you use the CAS technology.



TI-Nspire CX II CAS

## Student book instructions

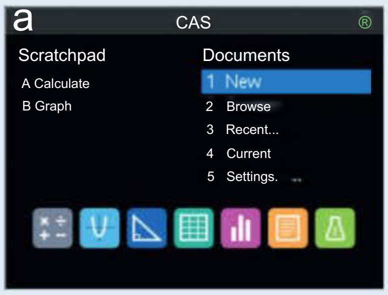
The instructions in this student book use words instead of symbols. Most keys on the keypad are clearly labelled with a word or an abbreviation. Four words that are used to represent less obvious keys are:

| Word | Key |
|---|---|
| home | \|tf}on\| |
| catalog | 0 |
| template | 0 |
| Scratchpad | rs~i |

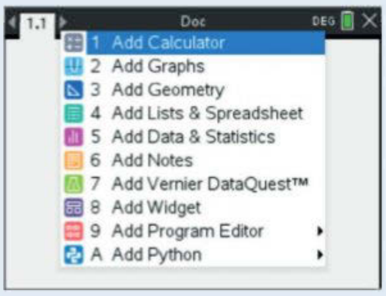
Several keys have a second function highlighted in blue above the key. For example, press Ctrl + $x^2$ to access the square root function *J"*.

## Applications

The applications available are outlined below.




Press home to view the home page. The Scratchpad options on the left are available for quick calculations and graphing. The Document options on the right are used for navigation. The seven icons on the bottom are the main applications.

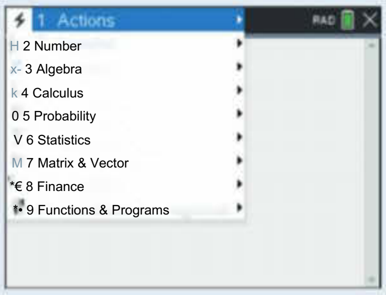



When you select Documents > New from the home page, a list of the seven applications plus three additional menu options will be displayed. From any application, press Ctrl + I (for insert) to display this list and add a new page to the document.

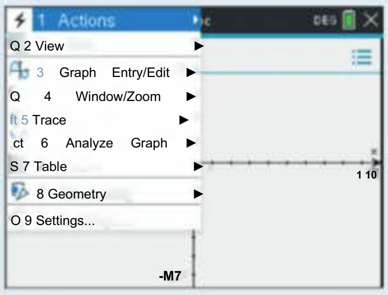
# Menus

The instructions in this student book primarily use the Calculator, Graphs, Lists & Spreadsheet and Data & Statistics applications (see Hints on page xvii). The following figures show the initial menu options for these four applications. These menu options link to submenu options, which are not shown below. On the handheld and software, the applications are referred to as Documents. In the student book instructions, the applications are referred to as pages of the document (e.g. Add a Graphs page).

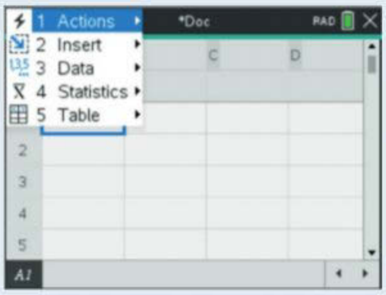
Calculator



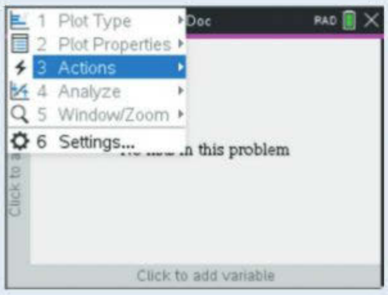

Graphs




Lists & Spreadsheet




Data & Statistics




If your document contains more than one page, move among them by clicking on the numbered tabs at the top of the screen. Alternatively, press Ctrl + left arrow or Ctrl + right arrow. To view all the pages of a document, press Ctrl + up arrow.

All menu and submenu options include numbers. The student book instructions do not include numbers. For example, the instruction in a Calculator page for clearing all calculations is ‘press menu > Actions > Clear History’. The shortcut is ‘press menu > 1 > 5’. For efficiency, you are encouraged to learn the sequence of numbers for frequently used commands.

# Document Settings

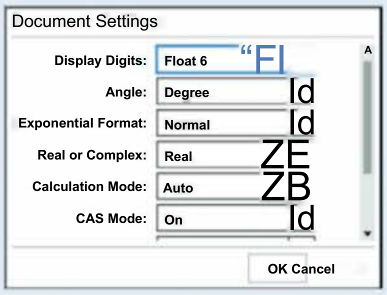
The document settings can be accessed in the following ways:

1 From the home page, press Settings > Document Settings.

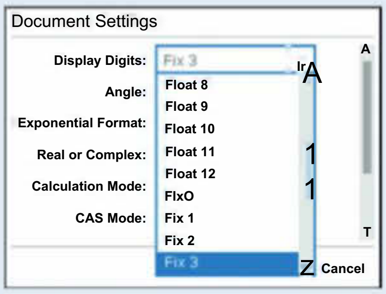
2 From a document page, press the doc key or click on Doc at the top of the page, then select Settings & Status > Document Settings.

3 Click on the battery icon in the top right-hand corner of the page, then select Document Settings.

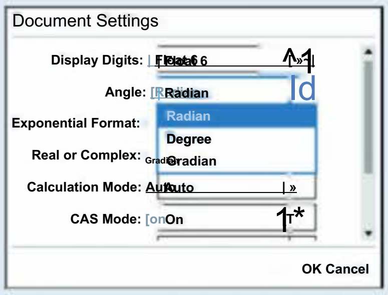
The document settings shown below are the primary ones you will be using.



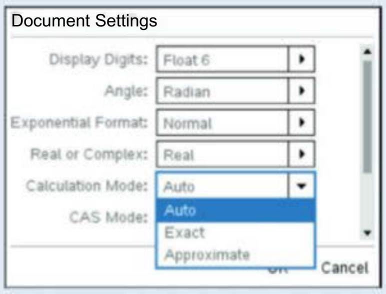

The screen above shows the default document settings. The Display Digits field is set to Float 6, which means up to six significant figures will be displayed.




Click in any field to display the options. The screen above shows the Display Digits options. Scroll down to select a specific number of decimal places.



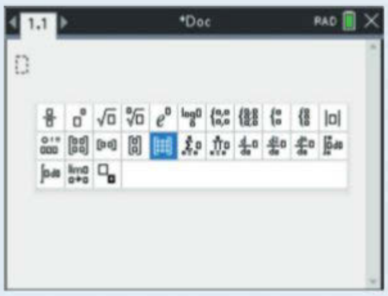

There are three Angle options shown above. Select either Radian or Degree. These two angle settings can be toggled at any time by clicking on DEG or RAD in the top right-hand corner of the screen.



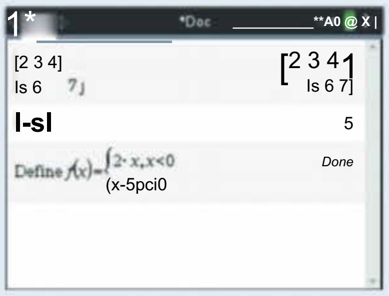

There are three Calculation Mode options shown above. It is recommended that you keep the default setting of Auto. If you require an approximate or decimal answer, press Ctrl + enter or include a decimal point in your calculation.

## Templates

The template key is located to the right of the 9 key.



Press template to view the template options. Most templates will be directly inserted but for some, you will be prompted with a dialogue box. For example, after selecting the 3x3 matrix template, you will be prompted for the number of rows and columns of the matrix.

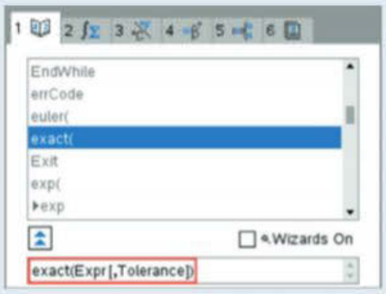



The screen above shows three examples using templates: a 2x3 matrix, an absolute value and a piecewise function. Many of the templates can be accessed using keys on the keyboard, for example, fraction, square root etc.

## Catalog

The catalog key is located to the right of the template key. In addition to using the menus and submenus, you can access all the commands using the catalog. The advantage of using the catalog is that it shows the parameters required for each command at the bottom of the screen. Optional parameters appear in square brackets.

For example:





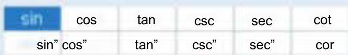
Press catalog and ensure tab 1 is selected. Press E to jump to the commands starting with E. Scroll down to exact(. The parameters for the exact command appear at the bottom of the screen (see red rectangle). Expr is a required parameter. The square brackets around Tolerance means it is an optional parameter.

When you use the exact command to convert 0.666 to a fraction, the answer is $\frac{333}{500}$.

If you include the optional tolerance of 3 then 0.666 converts to $\frac{2}{3}$.

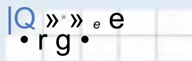
## Symbols

All symbols are listed at the end of the catalog. To access the symbol palette, press Ctrl + catalog. Frequently used symbols are available in mini-palettes by pressing the keys shown below.




Press trig to access the trigonometry functions.


Press Ctrl + = to access the inequalities and the constraint symbol.

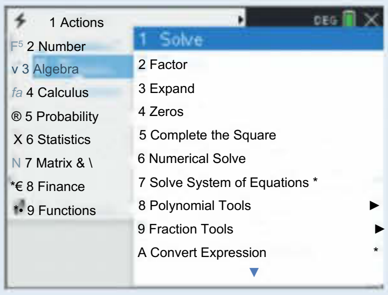


Press TT to access commonly used symbols.

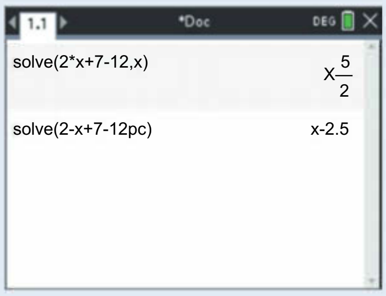
# Hints

The instructions below provide a few hints to assist with using the four main applications.

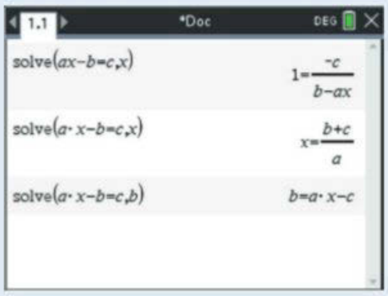
## Calculator



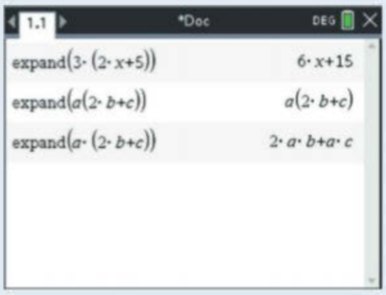

Press menu > Algebra to access the algebra functions. Select Solve. Other useful menu options include Factor and Expand.



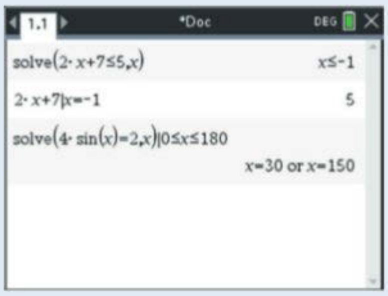

Enter the equation followed by ,x. Press enter for the exact answer. Press Ctrl + enter for the decimal answer.



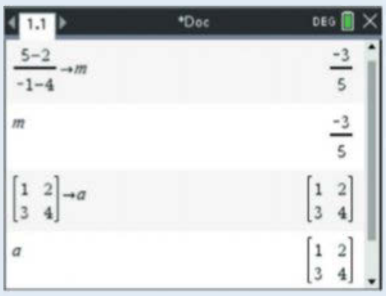

When multiplying variables, you *must* include a multiplication sign. The first answer above is incorrect as there is no multiplication sign between the a and x. CAS assumes this is the variable ax, not a x x. The second answer is correct. The third answer shows a solve example using the variable b instead of x.




When you enter a left or right bracket, both brackets appear. When you enter a number in front of a bracket, CAS automatically inserts a multiplication sign as in the first example above. However, if you enter a variable in front of a bracket, you *must* include the multiplication sign or you will get an error, or an incorrect answer. The second example above is incorrect, whereas the third example is correct.

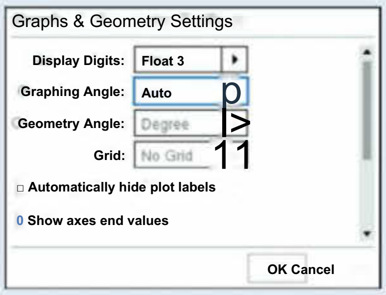



Press Ctrl + = to access the mini-palette for the inequality symbols. This mini-palette also has the constraint symbol (|), which can be used for substitution and to restrict domains.

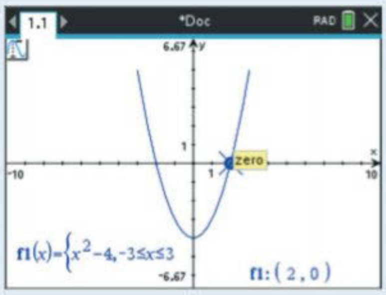



Press Ctrl + var for the store symbol, which can be used to store values and matrices. Note that TI-Nspire converts all letters to lower case.

## Graphs & Geometry

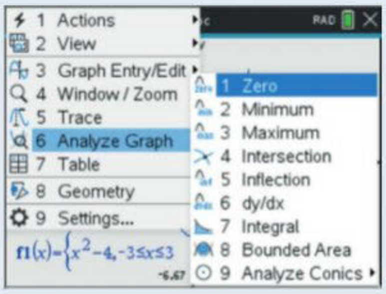



Press menu > Settings to view and/or change the settings for the Graphs and Geometry applications. The default setting for Display Digits is Float 3 so change this if you need greater accuracy. The Graphing Angle is set to Auto but can be changed to Degree or Radian.

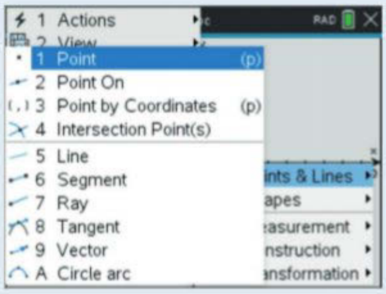



The constraint command can also be used in a Graphs page to specify a domain.

Press menu > Trace > Graph Trace to identify points on a graph.

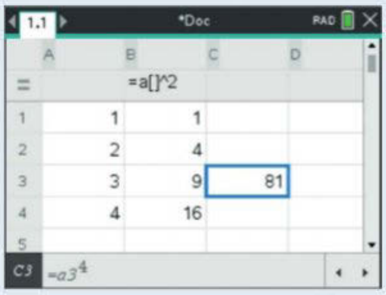



A second option for identifying key points on a graph is to press menu > Analyze Graph. These options prompt you for a lower bound and upper bound to locate the point.

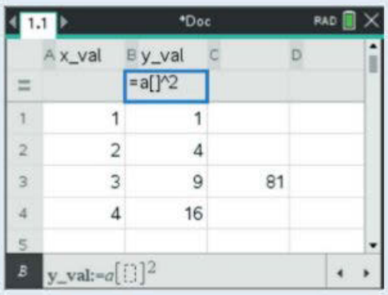



A third option is to press menu > Geometry > Points & Lines. Locate points of intersection or place points on the graph and move them along the graph. Click on a coordinate to manually change a value.

## Lists & Spreadsheet




The columns in the Lists & Spreadsheet application can be used for lists. Above, the values in the list in column B are the squares of the values of the list in column A. It can also be used as a spreadsheet. The value in cell C3 is the value in cell A3 raised to the power of 4.

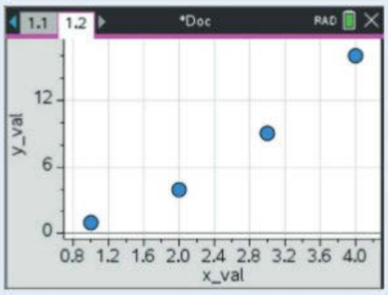


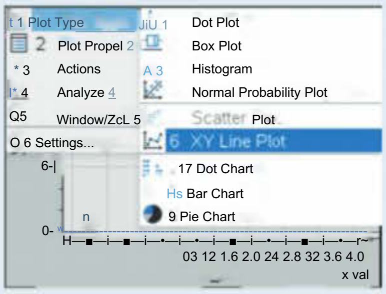

Calculations using lists can be completed using the column headings A, B, C etc. Lists need to be labelled to be used in other applications. The list A above has been labelled x_val and list B labelled y_val. List names cannot have spaces so spaces are replaced by asterisks. Alternatively, press Ctrl > space bar to insert the underscore character instead of a space.



9780170448406

## Data & Statistics




The Data & Statistics application is reliant on lists generated in other applications. It is designed for ungrouped data. The plot above displays the x_val and y_val lists from the Lists & Spreadsheet application.

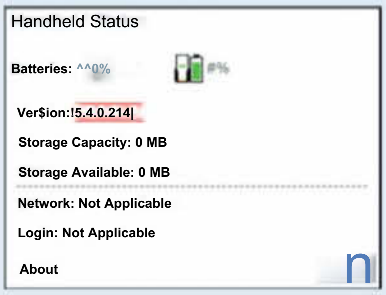
Press menu > Plot Type to view the various graphing options. Options 1 to 4 are for univariate data (one list). Options 5 and 6 are for bivariate data (two lists). Options 7 to 9 are for categorical data (one list).

# Operating systems

Ensure the latest operating system is installed on your handheld and software.

Installing the latest operating system is relatively straightforward. Using the USB cable provided, connect the handheld to a computer with the student or computer link software installed. Select Help > Check for OS Updates. If you see a message that a new OS is available, follow the links to install it. Alternatively, go to the TI website at [https://education.ti.com/](https://education.ti.com/) to download the latest operating system. Select Tools > Install OS then select the downloaded file.

To determine the version of your operating system, press home > Settings > Status. At the time of publication, the operating system for the CXII is version 5.4.0.214.




9780170448406

# Casio ClassPad introduction

The latest model of the Casio ClassPad is the fx-CP400. The connectivity software Screen Receiver, Share Assistant and Program Link Software can be downloaded for free. The ClassPad Manager software emulator is a separate program available at an additional cost.


Casio ClassPad

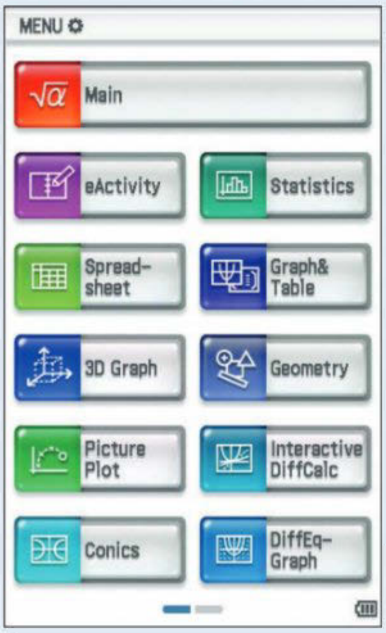
## Student book instructions

The instructions in this student book use words instead of symbols. ClassPad tools are located at the top of the screen. These tools vary with each application. Initially, these instructions will show a tool enclosed in a red rectangle with the corresponding word highlighted in red. Examples are shown here.

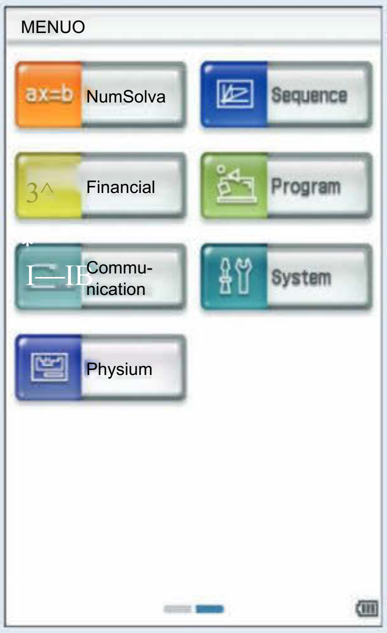
| Word | Tool |
|---|---|
| Graph |  |
| View Window | B |
| Table | H |
| Table Input |  |

## Applications

The applications available are outlined below.



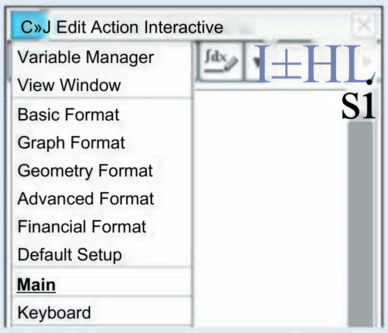

Tap Menu to view the applications.



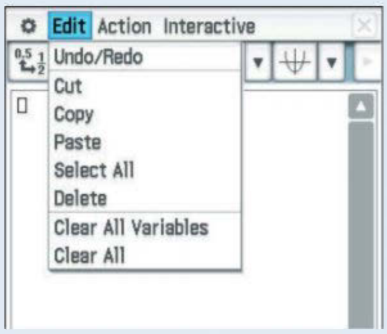

Slide the scroll bar at the bottom of the screen to access the full list.

# Menus

The instructions in this student book will primarily use the Main, Statistics, Spreadsheet, Graph&Table, Sequence and Financial applications (see Applications on page xxiii). All these applications have the G (systems) and Edit menus available at the top left of the screen.

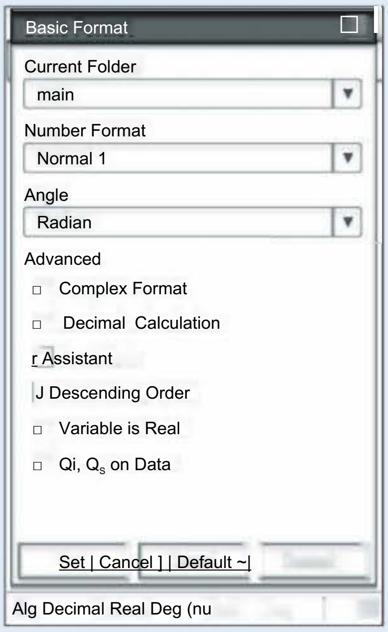



Tap O to view the system menu. The menu options allow you to manage variables and format applications.

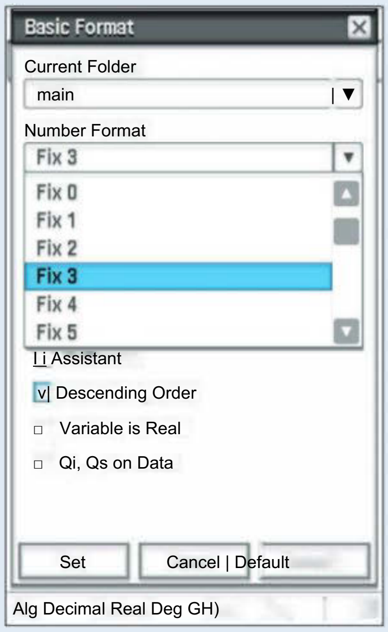



Tap Edit to view the edit menu. The menu options allow you to cut, copy, paste and delete screen content and clear variables. The Edit menu varies with each application.

# Document Settings




Tap O > Basic Format. The screen above shows the default document settings. The Number Format field is set to Normal 1.

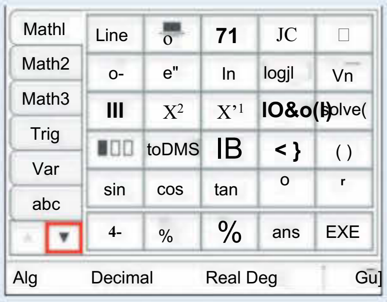



Tap the Number Format field to display the options. The screen above shows the Display Digits options. Scroll down to fix a specific number of decimal places.

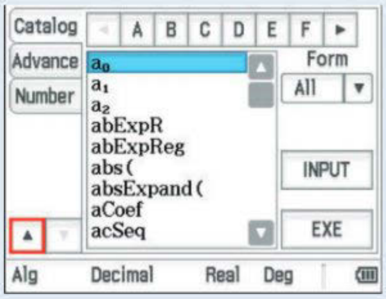
The settings Standard/Decimal, Real/Cplx and Rad/Deg/Gra can be toggled at the bottom of the screen. The recommended settings are Decimal, Real and Deg.

# Keyboard

There are nine soft keyboards available.




Press Keyboard to view the soft keyboards. The Math1 soft keyboard is shown above. Tap the left tabs to access the other keyboards. Press the down arrow to view the second screen.




All functions can be accessed from Catalog. Tap on the letters at the top to jump through the list. Press the up arrow to return to the first screen.


From the first screen, tap Var to access the variables. Use variables, not letters, in your algebraic calculations.

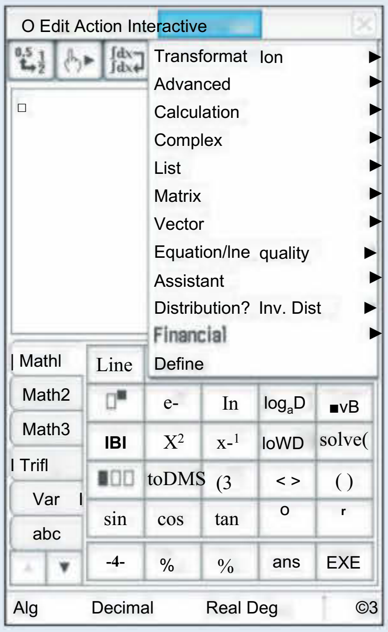

Tap abc to access letters and symbols. Use letters to name functions, matrices etc. Tap the tabs at the top of the screen to access the range of symbols. Press back to return to the main screen.

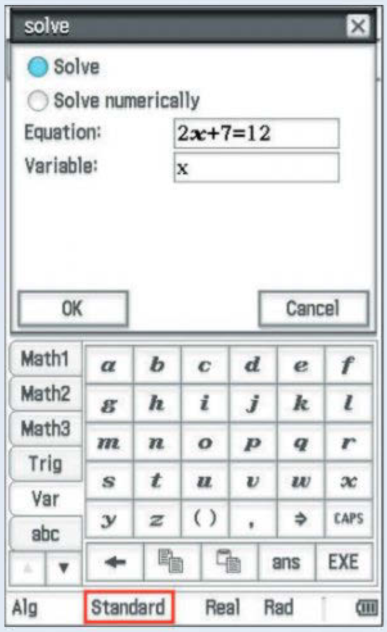
 9780170448406

# Applications

The instructions below provide a few hints to assist with using the main applications.

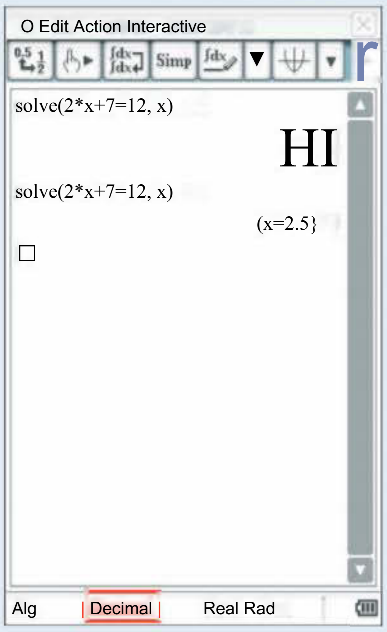
## Main



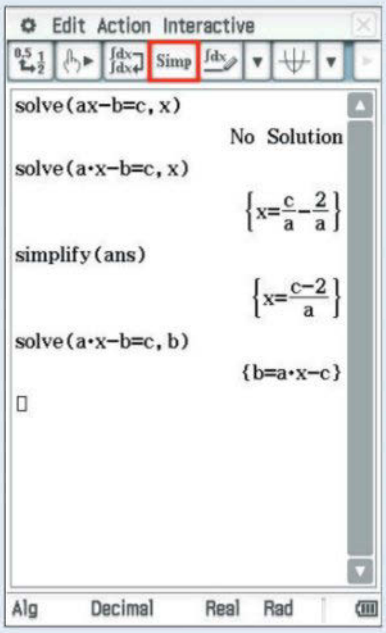




In Main, there are two menus available to enter functions: Active and Interactive. With the Interactive menu, enter the expression first then highlight and operate from there. Alternatively, tap Interactive then enter the expression in the dialogue box. With few exceptions, the instructions in this student book are written using the Interactive menu.

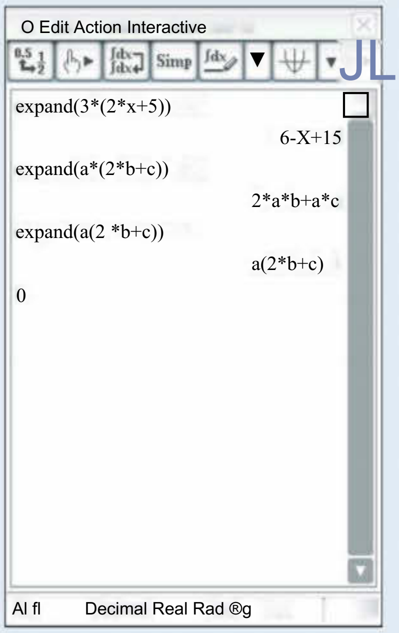
Tap Interactive > Advanced > solve to open the solve dialogue box. In the Equation field, enter the equation. In the Variable field, the default variable is set to x. Press OK to solve the equation. Note that the mode is set to Standard so the answer will be exact.



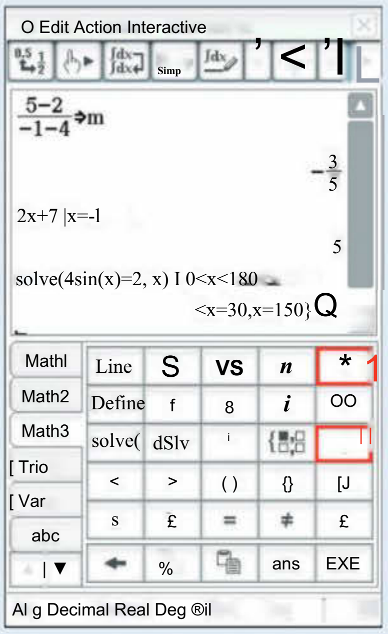

The first answer above is exact as the mode was set to Standard. To find the decimal answer, tap to change the mode to Decimal. Press EXE to display the second answer above, which is the decimal value. Alternatively, you can select the solve numerically option in the solve dialogue box (see previous screen).




When multiplying variables, be sure to use the Var menu, not the a be menu. The first answer has no solution as letters were used, and there was no multiplication sign inserted between the a and x. CAS assumes this is the variable ax, not axx. The second answer used variables, not letters so it is correct. Tap Simp to simplify to get the third answer. The fourth answer shows a solve example using the variable b instead of x.




Tap Interactive > Transformation > expand. When you enter a number or variable in front of a bracket, CAS automatically includes the multiplication sign, as in the first and second examples above. The third example is incorrect as a letter was used instead of a variable, and the multiplication sign between the letter and the left bracket was not inserted.




The store (|*|) arrow is available from the Math and Trig soft keyboards. Store can be used to store values and matrices. Use the abc soft keyboard to label them.

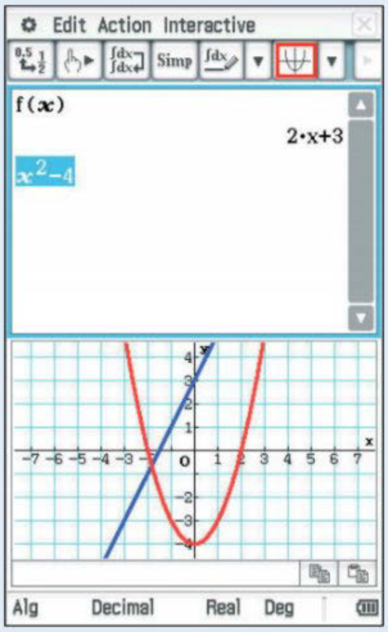
The constraint symbol | is available in the Math3 soft keyboard. Use constraint for substitution and to restrict domains.



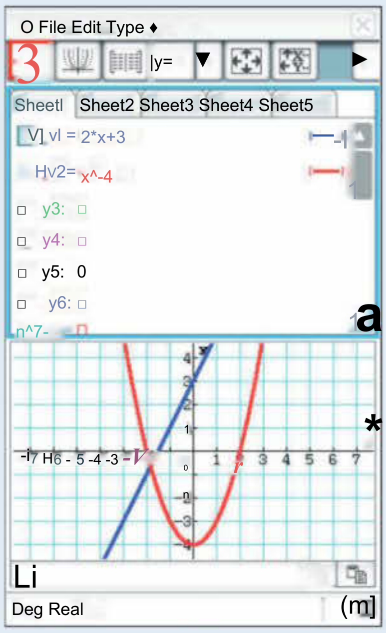
9780170448406

## Graphing

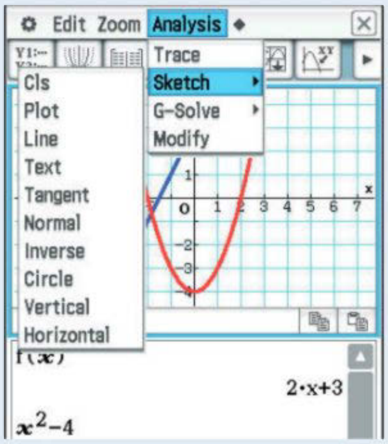
There are two main options for graphing functions and relations.



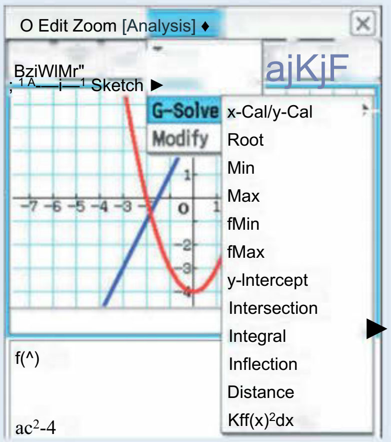

In Main, tap Graph to open the graph window. Enter and highlight the function or expression and drag it into the Graph window.




Tap Menu > Graph&Table. Enter the function then tap Graph. The instructions in this student book use the Main option because the y= is not required, but ensure you are familiar with both graphing methods.

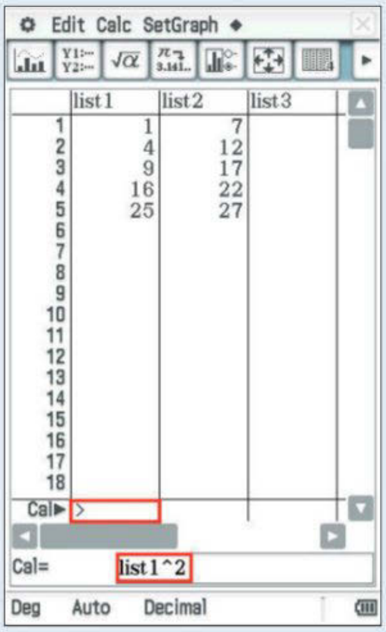



There are several menu options available to analyse graphs. Tap Analysis > Trace and press the arrow keys to move along the graph. Tap Analysis > Sketch for the options shown above.

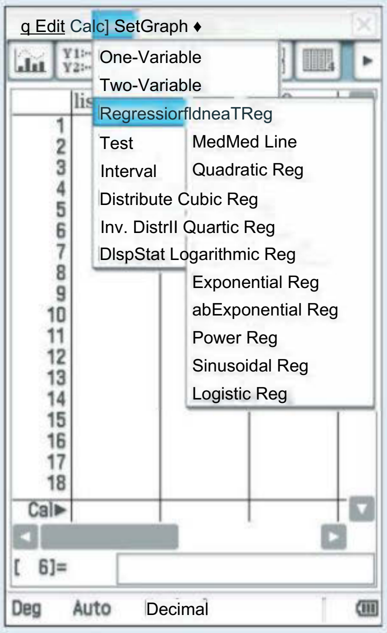



Tap Analysis > G-Solve to identify key features of a graph. For example, select Root to find the x-intercepts. Select Intersection to locate points of intersection of two graphs.

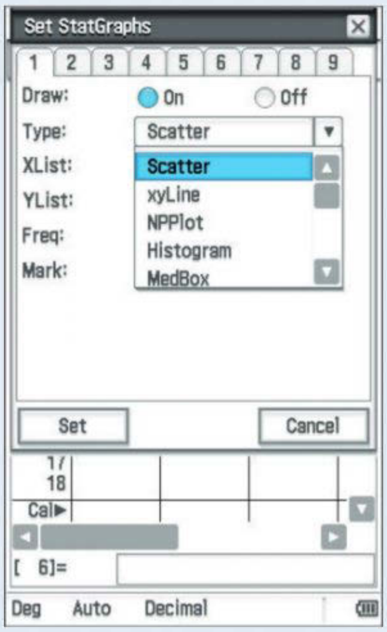
## Statistics



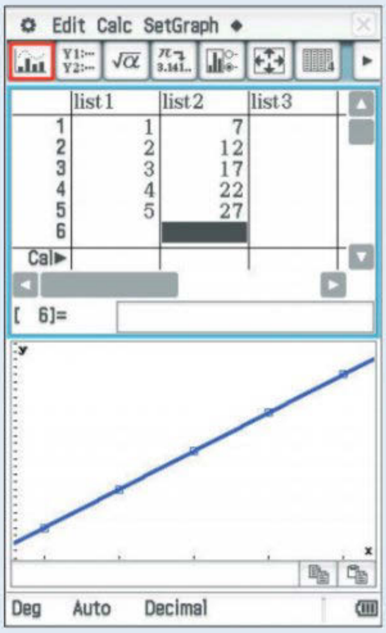

Tap Menu > Statistics to open the Statistics application. The default settings show listl, Iist2, list3 etc. If required, tap on the list name to enter a new heading. If you need to perform a calculation, tap in the Cal row at the bottom of the list to enter the formula.




Tap Calc to view the Calculation menu options. Tap One-Variable for the dialogue box used to calculate statistical analysis of a list. Tap Regression to access the Regression submenu options shown above.

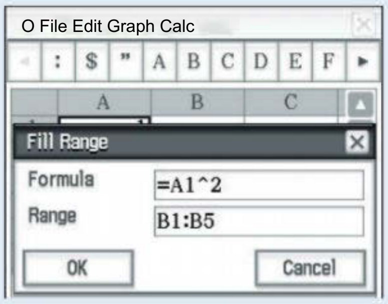



Tap SetGraph to set up the statistics graphs. Tap Type: to view the graphing options for statistical graphs. When finished, tap Set.

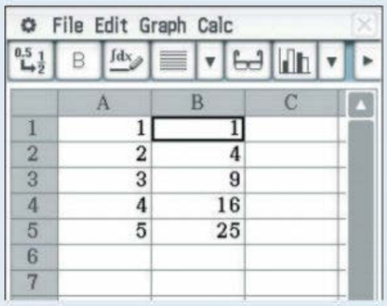



Tap Graph to display the statistical graph. The graph above displays a scatterplot of the listl and Iist2 values with the linear regression line joining the points.

## Spreadsheet




Tap Menu > Spreadsheet to open the Spreadsheet application. Tap Edit > Fill > Fill Range. In the dialogue box, enter the formula and range.




In the screen above, the values 1 to 5 were entered into column A. The formula and range entered squared these values and placed them in column B.

# Operating systems

Ensure you have the latest operating system installed on your handheld.

Tap Menu then tap Settings (located in the bottom left comer of the screen). Select Version. At the time of pubheation, the latest version is 02.01.7001.

To download the latest operating system, go to Australian Shriro website at http://www.casio.edu.shriro.com.au/classpad.php. Using the USB cable provided, connect the handheld to a computer. Start the installation program and follow the prompts. Some of these prompts will be on the computer and others will be on the handheld.

CHAPTER

# 1 DATA DISTRIBUTIONS

STUDY DESIGN COVERAGE

## Study Design coverage

### AREA OF STUDY 1: DATA ANALYSIS, PROBABILITY AND STATISTICS

#### Investigating data distributions

- types of data
- representation, display and description of the distributions of categorical variables: data tables, two-way frequency tables and their associated segmented bar charts
- representation, display and description of the distributions of numerical variables: dot plots, stem plots, histograms; the use of a logarithmic (base 10) scale to display data ranging over several orders of magnitude and their interpretation in terms of powers of ten
- use of the distribution(s) of one or more categorical or numerical variables to answer statistical questions
- summary of the distributions of numerical variables; the five-number summary and boxplots (including the use of the lower fence ($Q_1$ - 1.5 x IQR) and upper fence ($Q_3$ + 1.5 x IQR) to identify and display possible outliers); the sample mean and standard deviation and their use in comparing data distributions in terms of centre and spread
- the normal model for bell-shaped distributions and the use of the 68-95-99.7% rule to estimate percentages and to give meaning to the standard deviation; standardised values (z-scores) and their use in comparing data values across distributions.

Note: Two-way frequency tables and distributions of more than one variable are covered in Chapter 2: Associations between two variables.



### Video playlists (9):

1.1 Introduction to data distributions
1.2 Histograms
1.3 Boxplots
1.4 Log scales
1.5 Dot plots and stem plots
1.6 The mean and standard deviation
1.7 Bell-shaped distributions
1.8 Standardised values
**VCE question analysis** Data distributions

### Skillsheets (1):

1.1 Statistical measures

### Worksheets (23):

1.1 Statistical data match-up • Frequency distribution tables • Frequency tables • Mode, median and mean • Measures of central tendency • Mean, median, mode and range
1.2 Histograms • Shapes of distributions
1.3 Five-number summaries • Interquartile range • Boxplots • Boxplots 1 • Boxplots 2
1.4 Significant figures
1.5 Stem-and-leaf plots
1.6 Standard deviation • Statistical calculations • Statistics review • Calculating and interpreting summary statistics • Data and statistics crossword • Statistics crossword
1.7 The normal curve
1.8 z-scores

### Puzzles (2):

1.2 Statistical match-up
1.3 Statistical measures puzzle

Nelson MindTap

To access resources above, visit cengage.com.au/nelsonmindtap

Video playlist
Introduction to data distributions

# (^) Introduction to data distributions

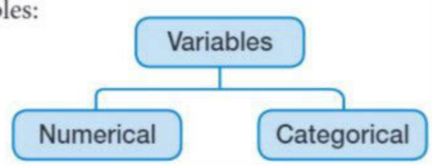
## Variables and data

In statistics, a variable is something that can be counted, measured or categorised. Data are the specific values of a variable.

The table shows examples of variables and data measured in a survey of ten students.

| Variables | Data |
|---|---|
| *Number* of mobile phones in the home | 4, 3, 5, 2,1,4, 6, 3,4,4 |
| *Age* (in years) | 17,17,18,18,17,17,18,17,16,17 |
| *Time* taken to run 50 metres | 7 sec, 5 sec, 7 sec, 6 sec, 8 sec, 5 sec, 8 sec, 6 sec, 6 sec, 6 sec |
| *Colour* of mothers' car | grey, white, grey, white, red, blue, grey, red, red, silver |
| *Rating* given to the latest superhero movie where 1 = great, 2 = okay, 3 = awful, 4 = didn't see it | 1,1,1,2, 3, 3, 4,2,1,2 |
| *Postcode* of home address | 3149, 3166, 3148, 3149, 3149, 3149, 3150, 3147, 3149, 3148 |

There are two main types of variables:




Numerical variables involve numbers that can be counted or measured. One way to test this is to ask, 'Does it make sense to add the numbers together?* For example:

- *number* of mobile phones in the home
  When counting phones, adding the numbers makes sense: 4 phones + 5 phones = 9 phones
- *time* taken to run 50 metres
  When measuring time, adding the numbers makes sense: 8sec + 6sec = Msec

Categorical variables involve either numbers where adding makes no sense or categories that don't involve any numbers. For example:

- postcode of home address
  Adding these numbers makes no sense:
  3149 (Mt Waverley, Vic.) + 3166 (Oakleigh, Vic.)
  *6315 (Dongolocking, WA)
- *colour* of mothers' car
  Grey, white, red etc. are not numbers.

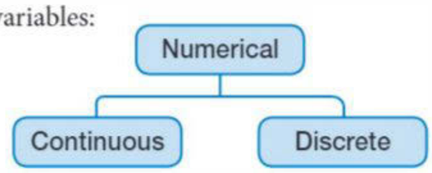
**Exam hack**

Exam questions about variables involving numbers with no mathematical meaning are common. Remember, just because a variable involves numbers doesn't mean it's a numerical variable.

Worksheet
Statistical data match-up

## Classifying variables

There are two types of numerical variables:




Continuous variables are numerical variables that can be measured to ever-increasing levels of accuracy. For example:

- *time* taken to run 50 metres
  This can be measured to ever-increasing levels of accuracy:
  7 sec, 5 sec, 7 sec ... or 7.4 sec, 5.1 sec, 6.9 sec ... or 7.42 sec, 5.13 sec, 6.94 sec ... etc.


9780170448406

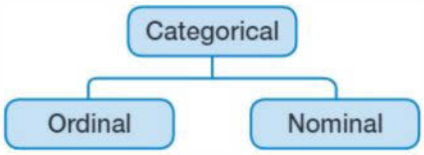
1.1

Discrete variables are numerical variables that can only take specific values and *can't* be measured to ever-increasing levels of accuracy. For example:

- *number* of mobile phones in the home
  The variable can only take whole number values: 0,1, 2, 3 ... It is impossible, for example, to have 1.5 or 2.8 of a mobile phone.
- *age* (in years)
  It specifically states age in years, so the variable can only take whole numbers. An age of 17.3 years, for example, is not allowed.

> **i©l Exam hack**
>
> Think of continuous variables as a footpath and discrete variables as stepping stones.

There are two types of categorical variables:




Ordinal variables are categorical variables that have a natural order. For example:

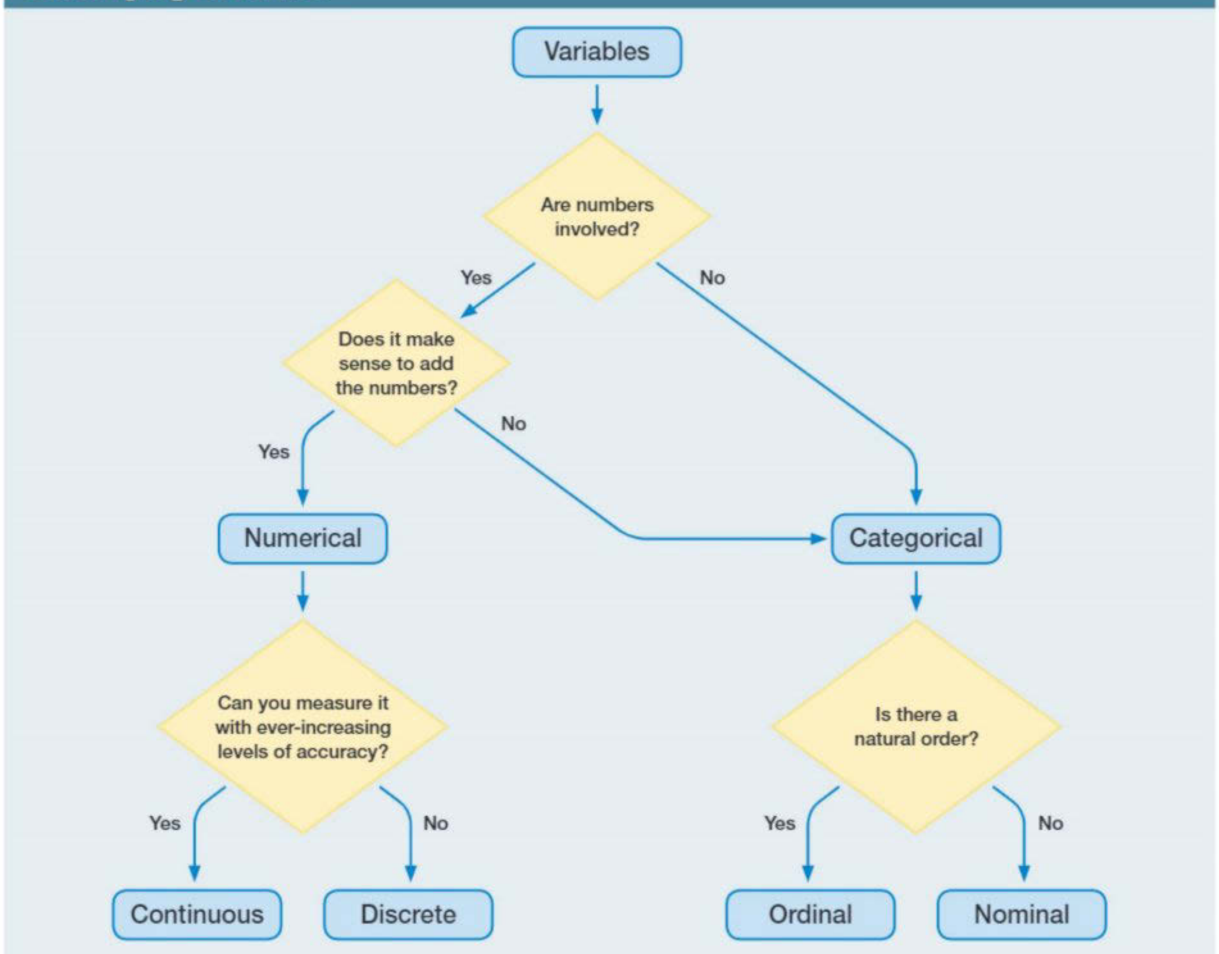
- *rating* given to the latest superhero movie where 1 = great, 2 = okay, 3 = awful, 4 = didn't see it.

Nominal variables are categorical variables that have *no* natural order. For example:

- *colour* of mother's car
- *postcode* of home address

> **Exam hack**
>
> If a question asks whether a variable is discrete, continuous, ordinal or nominal, always decide ***first*** if the variable is numerical or categorical.

**Classifying variables**



WB

p-i

| WORKED EXAMPLE 1 | Classifying variables |
|---|---|
| State whether the following variables shown in *italics* are<br>i numerical or categorical<br>ii discrete, continuous, ordinal or nominal. | |
| **Steps** | **Working** |
| a *Country* of birth | |
| i Are numbers involved? | No, so it is categorical. |
| ii Is there a natural order? | No, so it is nominal. |
| b *Weight* at birth | |
| i 1 Are numbers involved? | yes |
| 2 Does it make sense to add the numbers? | Yes, so it is numerical. |
| ii Can it be measured with increasing levels of accuracy? | Yes, so it is continuous. |
| c *Speed* of a car (under 40 km/h, 40-60 km/h, over 60 km/h) | |
| i 1 Are numbers involved? | yes |
| 2 Does it make sense to add the numbers? | No, so it is categorical. |
| ii Is there a natural order? | Yes, so it is ordinal. |
| d *Attendance numbers* at a series of concerts | |
| i 1 Are numbers involved? | yes |
| 2 Does it make sense to add the numbers? | Yes, so it is numerical. |
| ii Can it be measured with increasing levels of accuracy? | No, so it is discrete. |
| e *Price* of most recent book bought | |
| i 1 Are numbers involved? | yes |
| 2 Does it make sense to add the numbers? | Yes, so it is numerical. |
| ii Can it be measured with increasing levels of accuracy? | No, so it is discrete. |

%

Worksheets
Frequency distribution tables

Frequency tables

## Displaying data

Raw data is unorganised data. It is easier to analyse data if it is organised and displayed clearly. There are a number of different ways to display data.

A frequency table involves counting how often each data value occurs, with frequencies often shown as percentages. The following frequency table has information about a categorical variable where 57 students were asked how often they left home without their mobile phone (always, sometimes or never).

Frequency table

| Leave home without mobile phone | Frequency | Percentage |
|---|---|---|
| Always | 12 | 21.1 |
| Sometimes | 13 | 22.8 |
| Never | 32 | 56.1 |
| Total | 57 | 100.0 |

$\text{percentage} = \frac{\text{frequency}}{\text{total}} \times 100\%$

$= \frac{12}{57} \times 100\%$

$= 0.211 \times 100\%$

$= 21.1\%$

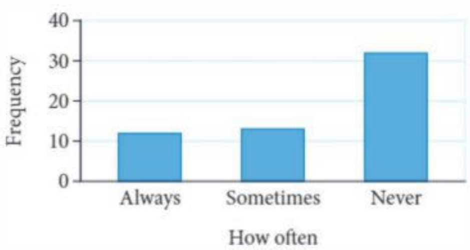
If the percentages have been rounded, this sometimes totals 99.9% or 100.1%.



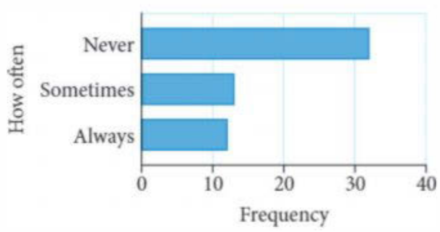
9780170448406

Bar charts organise and display data in vertical or horizontal bars and are particularly useful when dealing with categorical variables.

Bar chart with vertical bars



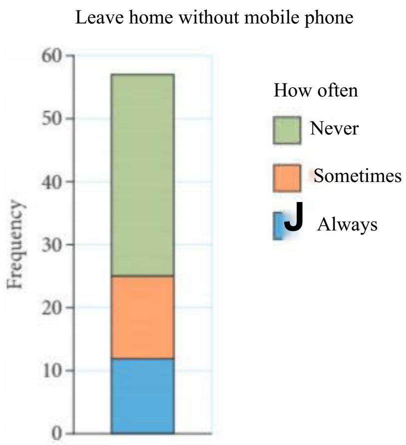

Bar chart with horizontal bars



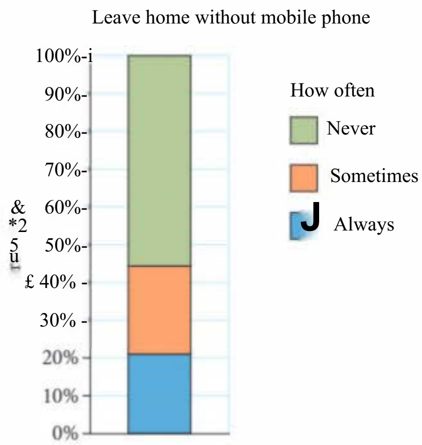

A segmented bar chart involves one bar with several segments that show the frequency of each category.

A percentage segmented bar chart has percentages rather than frequencies, and the length of the bar is 100%. Segmented bar charts require a key and should not have too many segments.

Segmented bar chart




Percentage segmented bar chart




## Displaying data

Frequency tables

- display both categorical and numerical variables
- can include frequencies or percentages

. percentage-^“^x 100%
* total

Bar charts

- display categorical variables
- can have vertical or horizontal bars.

Segmented bar charts

- can be shown as a frequency or a percentage
- need a key to explain what each segment represents
- become difficult to read if they have too many segments.

WB

p.2

## WORKED EXAMPLE 2 Displaying categorical variables

The following is raw data of the types of hot drinks the first 40 people ordered one day at the Little Bean Coffee Shop, where C = Cappuccino, F = Flat white, L = Latte, T = Turmeric latte, M = Matcha latte, H = Hot chocolate.

C, C, C, L, M, L, C, C, T, F, L, L, F, L, L, T, H, M, C, L,
T, H, L, L, L, C, T, C, C, F, H, C, L, C, L, T, F, C, C, C

**Steps** **Working**

**a** Set up a frequency table that includes both the frequency and percentage of each type of hot drink ordered.

**1** Set up a table with three columns and list the categories in the first column. Count the number in each category and record the frequency.

Check that the total frequency equals the total number of data values given in the question.

| Type of hot drink | Frequency | Percentage |
|---|---|---|
| Cappuccino | 14 | |
| Flat white | 4 | |
| Latte | 12 | |
| Turmeric latte | 5 | |
| Matcha latte | 2 | |
| Hot chocolate | 3 | |
| **Total** | 40 | |

**2** Calculate the percentage for each category using

$$\text{percentage} = \frac{\text{frequency}}{\text{total}} \times 100\%$$

Check that the total percentage is equal to 100% (or 99.9% or 100.1% if the percentages have been rounded).

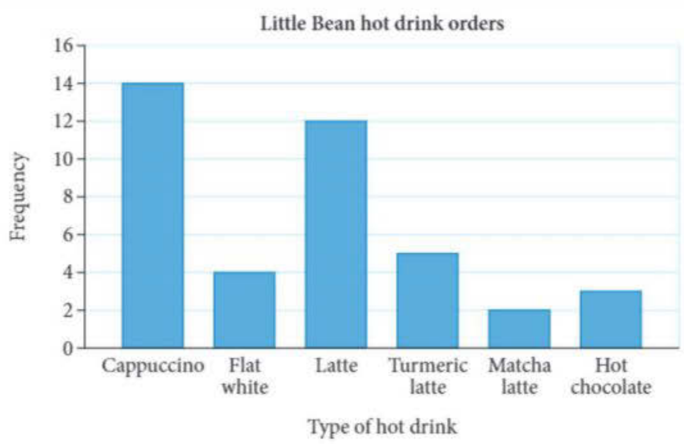
| Type of hot drink | Frequency | Percentage |
|---|---|---|
| Cappuccino | 14 | 35 |
| Flat white | 4 | 10 |
| Latte | 12 | 30 |
| Turmeric latte | 5 | 12.5 |
| Matcha latte | 2 | 5 |
| Hot chocolate | 3 | 7.5 |
| **Total** | 40 | 100.0 |

**b** Draw a bar chart of this frequency table showing the number of each type of hot drink ordered, with the categories on the horizontal axis.

Draw a bar chart with the categories on the horizontal axis.

Add a title, label the horizontal axis with the variable name, and label the vertical axis 'Frequency'.

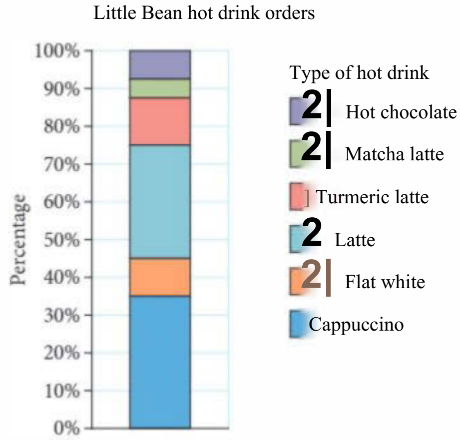
Note: As the variables are categorical nominal and therefore don't have a natural order, the bars can appear in any order.



c Draw a percentage segmented bar chart of this frequency table.

Draw a percentage segmented bar chart with the vertical axis labelled 'Percentage'.

Add a title and a key matching the order in which the categories appear in the bar chart.




d State the most frequently occurring category.

Find the category that occurs most often.

The most frequently occurring category is 'Cappuccino',

## Measures of centre and spread

Two key features to look for when analysing data are centre and spread. The centre of a distribution is a single value that best describes the distribution. We will be looking at three ways to measure the centre of a distribution:

- The mode is the most frequently occurring data value and is often called the modal category for categorical data.
  - There can be more than one mode.
  - Data with two modes is called bi-modal.
  - If *every* data value appears exactly once, there is no mode.
- The mean is the average of all the data values and is calculated by dividing the sum of all the data values by the total number of values.
- The median is the middle value. To find the median:
  1. Order the $n$ data values from smallest to largest, where $n$ is the number of data values.
  2. Find the $\frac{n+1}{2}$th position. (Add one to the number of data values, then divide by 2.)
  3. If $n$ is odd, find the data value in the $\frac{n+1}{2}$th position.
  4. If $n$ is even, find the two data values either side of the $\frac{n+1}{2}$th position and average them.

Skillsheet
Statistical measures

Worksheets
Mode, median and mean

Measures of central tendency

Mean, median, mode and range

Odd number of ordered data values ($n = 11$)

Median position is $\frac{11+1}{2} = \frac{12}{2} = 6$

Median is the 6th data value:

3, 6, 7, 7, 7, 8, 13, 19, 20, 20, 22

5 data values | 5 data values

Median

Even number of data values ($n = 12$)

Median position is $\frac{12+1}{2} = \frac{13}{2} = 6.5$

Median is between the 6th and 7th data value:

3, 6, 7, 7, 7, 8, 13, 17, 19, 20, 22, 26

6 data values | 6 data values

Median $= \frac{8+13}{2} = 10.5$

When analysing data, we also look for the spread of a distribution. One measure of the spread of a distribution is the range.

range = largest value - smallest value

The more spread out the data, the greater the range.

**Exam hack**

Not every measure of centre and spread can be used with every type of variable.

**Centre, spread and data types**

| Categorical: Nominal data | Categorical: Ordinal data | Numerical data |
|---|---|---|
| Measures of centre | | |
| mode | mode<br>median | mode<br>median<br>mean |
| Measures of spread | | |
| – | range | range |

WB

p. 4

**WORKED EXAMPLE 3** Finding measures of centre and spread

For each of the following data sets, find the

i mode ii mean iii median iv range.

| Steps | Working |
|---|---|
| **a** Ages of the oldest living pet (in years) in ten households: 9, 3, 6, 1, 0, 1, 10, 7, 8, 0 | |
| i Find the most frequently occurring data value(s). | modes = 0 years and 1 year |
| ii Divide the sum of all the data values by the total number of values. | $\text{mean} = \frac{9+3+6+1+0+1+10+7+8+0}{10}$<br>$= \frac{45}{10}$<br>$= 4.5$ years |
| iii 1 Order the $n$ data values from smallest to largest. | 0, 0, 1, 1, 3, 6, 7, 8, 9, 10 |
| 2 Find $n$ and note if the number is odd or even. | $n = 10$; even |
| 3 Find the position of the median. | $\frac{n+1}{2} = \frac{10+1}{2} = \frac{11}{2} = 5.5$<br>The median is between the 5th and 6th ordered data values. |
| 4 If $n$ is odd, find the middle value.<br>If $n$ is even, average the two middle values. | 0, 0, 1, 1, 3, 6, 7, 8, 9, 10<br>$\text{median} = \frac{3+6}{2} = \frac{9}{2} = 4.5$ years |
| iv range = largest value – smallest value | range = 10 – 0 = 10 years |

b Maximum daily temperatures (°C to one decimal place) at the Victorian Alpine National Park during a week in August:

2.7,0.0,-1.3, -3.1, 2.9, 0.8,-1.3

| | |
|---|---|
| i Find the most frequently occurring data value(s). | mode = -1.3°C |
| ii Dividing the sum of all the data values by the total number of values. | mean = $\frac{2.7 + 0.0 - 1.3 - 3.1 + 2.9 + 0.8 - 1.3}{7}$ <br> $= \frac{07}{7}$ <br> = 0.1°C |
| iii 1 Order the n data values from smallest to largest. | -3.1, -1.3, -1.3, 0.0, 0.8, 2.7,2.9 |
| 2 Find $n$ and note if the number is odd or even. | n = 7; odd |
| 3 Find the position of the median. | $\frac{n+1}{2} = \frac{7+1}{2} = \frac{8}{2} = 4$ <br> The median is the 4th ordered data value. |
| 4 If $n$ is odd, find the middle value. | -3.1, -1.3, -1.3, 0.0,0.8, 2.7,2.9 |
| If $n$ is even, average the two middle values. | median = 0.0°C |
| iv range = largest value - smallest value | range = 2.9 - (-3.1) = 6.0°C |

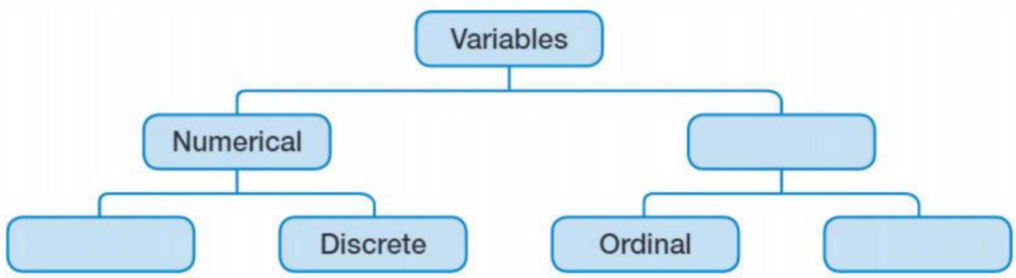
1.1

## EXERCISE 1.1 Introduction to data distributions

ANSWERS p. 692

### Mastery

1 Copy and complete the following diagram:




2 E3 WORKED EXAMPLE 1 I State whether the following variables shown in *italics* are

i numerical or categorical

ii discrete, continuous, ordinal or nominal,

a *Number* of television sets in the house

b *Time* spent doing homework last night

c *Name* of the suburb in home address

d *Number* of students sitting the General Mathematics examination

e *Speed* of cars at a certain point on a freeway

f The *candidate* people voted for in an election

g *Numbers* on soccer players uniforms

h Peoples' *opinion* of football on a scale of 1 to 5, where 1 is dislike and 5 is love

i *Fuel consumption* of a car

j Mothers *salary* classified as high, medium or low

k *Number* of students in schools in a region (less than 500, between 500 and 1000, more than 1000)

I *How often* students walk to school (never, sometimes, often, always)

m *Price* of mobile phones

3E WORKED EXAMPLE 2 The following is raw data of the types of beef pies the first 40 people ordered one day at the Humble Pie Bakery, where B = Beef burgundy, K = Beef and kidney, C = Beef curry, R = Beef rendang and G = Beef and Guinness.

C, C, B, K, G, G, C, B, K, B, C, B, B, C, R, G, K, C, B, C,
R, B, B, B, G, C, G, G, B, R, G, B, G, B, C, B, R, B, K, C

a Set up a frequency table that includes both the frequency and percentage of each type of pie ordered,

b Draw a bar chart of this frequency table showing the number of each type of pie ordered, with the categories on the horizontal axis.

c Draw a percentage segmented bar chart of this frequency table.

d State the most frequently occurring category.

4 H WORKED EXAMPLE 3 For each of the following data sets, find the

i mode ii mean iii median iv range.

a Ages of people (in years) working in a restaurant:

42,21, 18, 30, 19, 18,27

b Number of people buying coffees at a cafe on six consecutive mornings:

39, 36, 36, 38, 56, 44

c Average maximum monthly temperatures (°C to one decimal place) at a ski resort from April to November:

3.0, -2.0, -1.7, -2.0, -1.7, -2.8, 0.3, 2.1

## Exam practice

80-100% 60-79% 0-59%

5 ©VCAA 20111CQ4 91% The variables *region* (city, urban, rural) and *population density* (number of people per square kilometre)

A are both categorical.

B are both numerical.

C are categorical and numerical respectively.

D are numerical and categorical respectively.

E are neither categorical nor numerical.

6 ^^^J2O19N1CQ8 ~ The variables *recovery time after exercise* (in minutes) *and fitness level* (below average, average, above average) are

A both numerical.

B both categorical.

C an ordinal variable and a nominal variable respectively.

D a numerical variable and a nominal variable respectively.

E a numerical variable and an ordinal variable respectively.

7 ©VCAA 2018N 1CQ5 The variables *height* (less than 1.83 m, 1.83 m and over) and *enthusiasm for playing basketball* (low, medium, high) are

A both ordinal variables.

B both nominal variables.

C a nominal and an ordinal variable respectively.

D an ordinal and a nominal variable respectively.

E a numerical and an ordinal variable respectively.

8 ©VCAA 2019NICQ3 The total birth weight of a sample of 12 babies is 39.0 kg. The mean birth weight of these babies, in kilograms, is

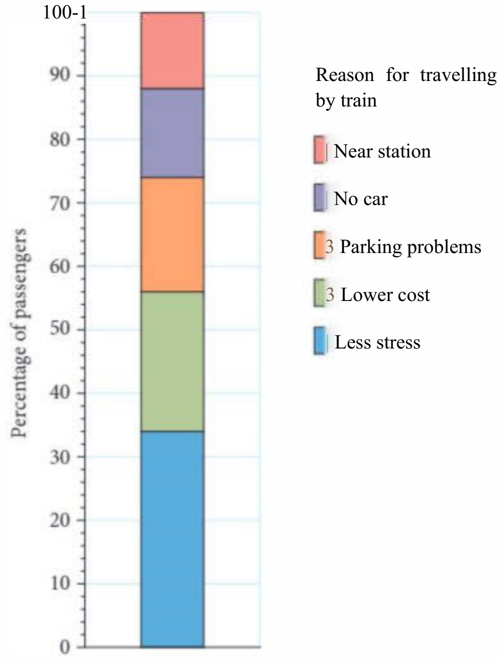
A 2.50 B 2.75 C 3.00 D 3.25 E 3.50

9 ©VCAA 2010 1CQ4 86% The passengers on a train were asked why they travelled by train. Each reason, along with the percentage of passengers who gave that reason, is displayed in the segmented bar chart shown. The percentage of passengers who gave the reason 'no car' is closest to

A 14% B 18%

C 26% D 74%

E 88%




1.1

*Use the following information to answer the next two questions.*

The percentage investment returns of seven superannuation funds for the year 2002 are

-4.6%, -4.7%, 2.9%, 0.3%, -5.5%, -4.4%, -1.1%

10 ©VCAA 2003 1CQ1 83% The median investment return is

A −4.7% B -4.6% C -4.5% D -4.4% E 0.3%

11 ©VCAA 2003 1CQ2 73% The range of investment returns is

A 2.6% B 3.5% C 4.0% D 5.5% E 8.4%

12 ©VCAA 2020 1CQ7 69% Data relating to the following five variables was collected from insects that were caught overnight in a trap:

- *colour*
- *name of species*
- *number of wings*
- *body length* (in millimetres)
- *body weight* (in milligrams)

The number of these variables that are discrete numerical variables is

A 1 B 2 C 3 D 4 E 5

13 ©VCAA 2008 2CQ1 95% (2 marks) In a small survey, twenty-five Year 8 girls were asked what they did (walked, sat, stood, ran) for most of the time during a typical school lunch time. Their responses are recorded below.

| | | | | |
|---|---|---|---|---|
| sat | stood | sat | ran | sat |
| walked | walked | sat | walked | ran |
| sat | walked | walked | walked | ran |
| walked | ran | walked | ran | walked |
| ran | sat | ran | ran | walked |

Use the data to

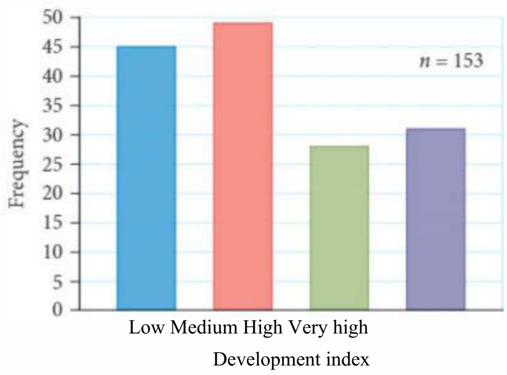
a copy and complete the frequency table 1 mark

| Activity | Frequency |
|---|---|
| walked | |
| sat or stood | |
| ran | |
| Total | 25 |

b determine the percentage of Year 8 girls who ran for most of the time during a typical school lunch time. 1 mark

14 ©VCAA 2013 2CQ1 J 83% (2 marks)

A development index is used as a measure of the standard of living in a country. The bar chart displays the development index for 153 countries in four categories: low, medium, high and very high.




a How many of these countries have a very high development index? 1 mark

b What percentage of the 153 countries has either a low or medium development index? Write your answer correct to the nearest percentage. 1 mark

15 (7 marks) The following table shows the data collected from a random sample of eight Year 12 students in a school. The variables in the table are:

- *gender* - the gender identity of the student (F = female, M = male, O = other)
- *phone number* - the phone number of the student
- *number of family members* - the number of family members living at home
- *language* - the language spoken at home (1 = English, 2 = Mandarin, 3 = Cantonese, 4 = Sinhala, 5 = other)
- *distance* - the distance that each student travels to school from their home

| *Gender* | *Phone number* | *Number of family members* | *Language* | *Distance* (km) |
|---|---|---|---|---|
| M | 0491 570 006 | 6 | 1 | 1.4 |
| F | 0491 578 957 | 3 | 2 | 0.4 |
| M | 0491 574 632 | 4 | 1 | 3.2 |
| F | 0491 571 266 | 4 | 3 | 1.1 |
| M | 0491 575 254 | 4 | 4 | 2.1 |
| O | 0491 579 455 | 5 | 5 | 1.6 |
| M | 0491 571 804 | 3 | 1 | 0.9 |
| F | 0491 572 549 | 4 | 1 | 1.3 |

a List the variables in this data set that are continuous. 1 mark

b List the variables in this data set that are nominal. 1 mark

c What percentage of those who identify as male speak English at home? 1 mark

d What is the modal category of language spoken at home? 1 mark

e Find the median distance that students travel to school from their home. 1 mark

f Find the range of the number of family members living at home. 1 mark

g Explain why it does not make sense to calculate the median of the phone numbers. 1 mark

16 (5 marks) The following bar chart shows the number of confirmed cases of COVID-19 in each Australian state and territory at 5.30 pm on 17 April 2020.


Source: State Health Departments, Monash University (quoted in *The Age* 18 April 2020 p. 7)

a What were the total number of COVID-19 cases in Australia at 5.30 pm on 17 April 2020? 1 mark

b What percentage of the total cases did each of New South Wales and Victoria have, rounded to the nearest percentage? 2 marks

The following table gives the percentages of the total Australian population in each state and territory.

| | Population % |
|---|---|
| New South Wales | 32 |
| Victoria | 26 |
| Queensland | 20 |
| South Australia | 7 |
| Western Australia | 10 |
| Tasmania | 2 |
| Northern Territory | 1 |
| Australian Capital Territory | 2 |
| Australia | 100 |

Source: ABS (Australian Bureau of Statistics) ABS website, accessed 18 April 2020

c Which state, Victoria or New South Wales, was more effective in restricting the spread of COVID-19 up until 17 April 2020? Provide evidence for your answer. 2 marks

# @ Histograms

Video playlist
Histograms

Worksheet
Histograms

Puzzle
Statistical match-up

## Grouped frequency tables and histograms

A grouped frequency table is a frequency table where data has been grouped into regular intervals to make it easier to deal with large amounts of data. A histogram is a graphical display of data from a grouped frequency table with data grouped into 5 to 15 intervals. Grouped frequency tables and histograms are used for numerical data only.

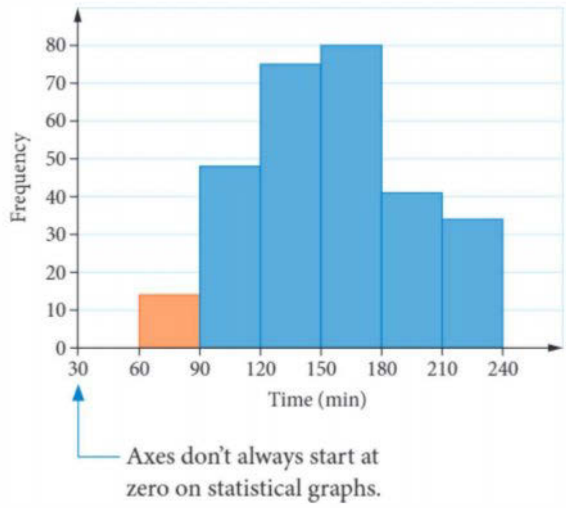
The following grouped frequency table and histogram both show the times (in minutes) taken by the 292 participants to complete a half marathon.

Grouped frequency table

| Time (min) | Frequency |
|---|---|
| 60–<90 | 14 |
| 90–<120 | 48 |
| 120–<150 | 75 |
| 150–<180 | 80 |
| 180–<210 | 41 |
| 210–<240 | 34 |
| **Total** | 292 |

60-<90 means "60 minutes or more but less than 90 minutes'. Someone with a time of exactly 90 minutes is recorded in the second interval.

Histogram




### Exam hack

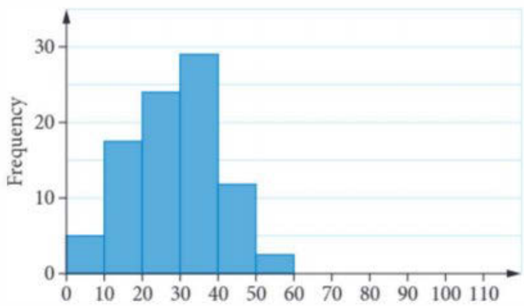
Although at first glance a histogram looks similar to a bar chart, there are a number of differences:

- Histograms display numerical data, whereas bar charts are best used to display categorical data.
- Histograms don't have any spaces between the columns.
- Histograms are *always* vertical.

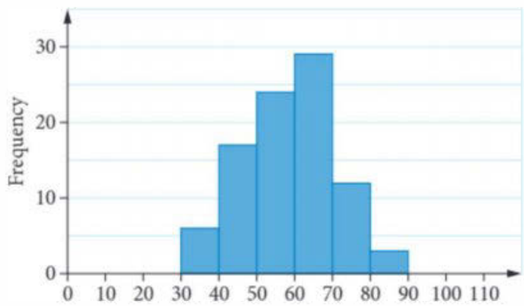
## Centre and spread of histograms

Histograms make it easier to see if the centre or spread of one distribution is different to another. We refer to the modal interval, or most frequently occurring interval, rather than the mode when looking at centres of histograms.

Here are two distributions with the same spread but different centres:




The modal interval is 30-<40.

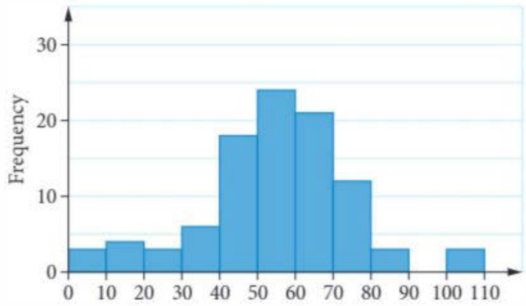


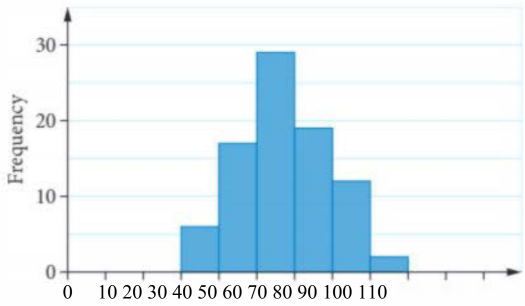

The modal interval is 60-<70.


9780170448406

1.2

Here are two distributions with approximately the same centres but different spreads:




The modal interval for both of the above histograms is 50-<60.

## Shape of histograms

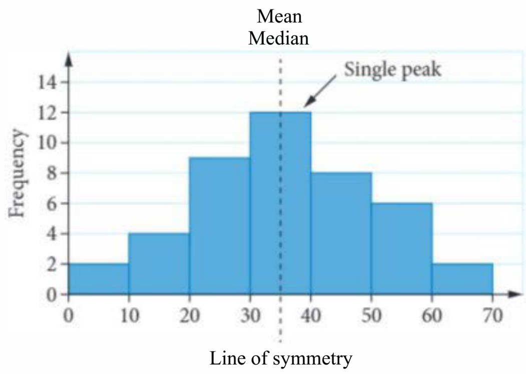
A histogram can also be used to describe the shape of a distribution. The shape of a distribution can be

- approximately symmetric
- positively skewed
- negatively skewed.

Worksheet
Shapes of distributions

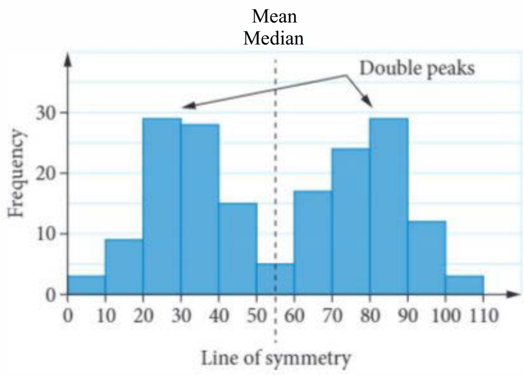
A histogram shows an approximately symmetric distribution when the left side closely mirrors the right side. For approximately symmetric distributions, the mean and median are both always near the line of symmetry, but the modal interval can be somewhere else.

Single-peaked
approximately symmetric distribution




The modal interval is where a single peak occurs. Here it is 30-<40.

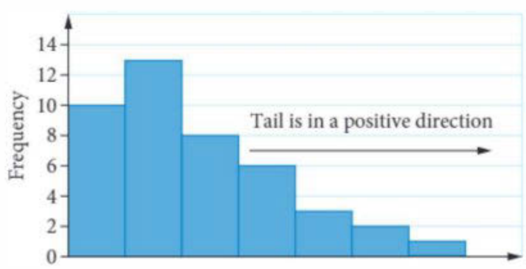
Double-peaked
approximately symmetric distribution




When a histogram has two equal peaks, it is a bi-modal distribution and has two modal intervals. Here they are 20-<30 and 80-<90.

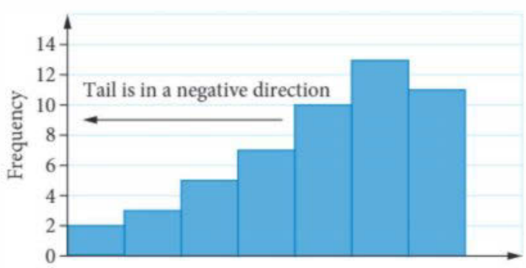
The following two histograms show skewed distributions.

Positively skewed




The mean is greater than the median.

Negatively skewed



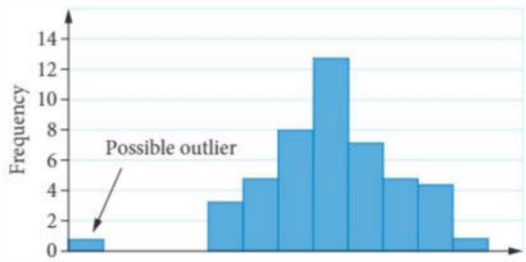

The mean is less than the median.

## Exam hack

To help you remember which skew is positive and which is negative, identify where the 'tail' of the histogram is. Think of the positive and negative directions of a number line. If the tail is in the positive direction, the distribution is positively skewed. If the tail is in the negative direction, the distribution is negatively skewed.

## Outliers

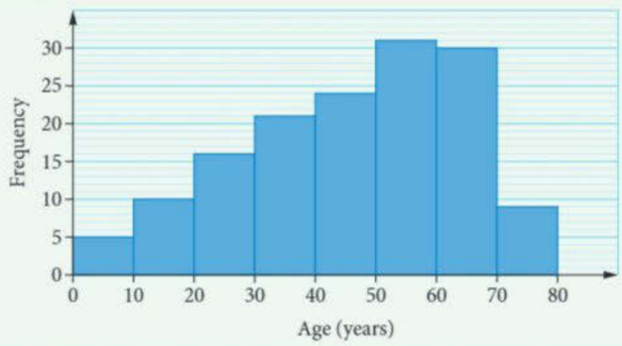
We can also comment on a distribution by referring to its outliers. An outlier is an extreme high or low value in the data. Outliers can indicate an error made in dealing with the data and can sometimes contaminate calculations and conclusions drawn from data sets. However, they can also occur without an error being involved. Histograms often make it easier to identify possible outliers.




WB
p.5

**WORKED EXAMPLE 4** Working with histograms

The following histogram shows the ages of people (in years) in a small town.




| Steps | Working |
|---|---|
| **a** How many people are in their 50s or 60s? | |
| Find the frequencies from the histogram. | 31 + 30 = 61 people are in their 50s or 60s. |
| **b** Is the histogram approximately symmetric, positively skewed or negatively skewed? Does it have any possible outliers? | |
| The histogram has a negative tail. | The histogram is negatively skewed with no possible outliers. |
| **c** What is the modal interval? | |
| Which age group has the highest frequency? | The modal interval is 50–<60 years. |
| **d** In which interval is the median? | |
| 1 Create a grouped frequency table for the histogram. | (see table below) |

| Ages | Frequency |
|---|---|
| 0–<10 | 5 |
| 10–<20 | 10 |
| 20–<30 | 16 |
| 30–<40 | 21 |
| 40–<50 | 24 |
| 50–<60 | 31 |
| 60–<70 | 30 |
| 70–<80 | 9 |
| **Total** | 146 |

2 Count the number of data values $n$. Note whether it is odd or even, and find the position of the median.

$n = 146$; even

$\frac{n+1}{2} = \frac{146+1}{2} = \frac{147}{2} = 73.5$

The median is between the 73rd and 74th ordered data values.

3 Add the frequencies in order from the grouped frequency table until the interval with the median position is reached.

Add the frequencies of the first four intervals:

5 + 10+ 16 + 21 = 52

This means the 52nd data value is in 30–<40.

Add the frequencies of the first five intervals:

5+10+16 + 21+24 = 76

This means the 76th data value is in 40–<50.

So, the 73rd and 74th data value are both in 40–<50.

The median is in the interval 40–<50 years.

**Exam hack**

A ruler can often help to read values from histograms and other statistical charts.

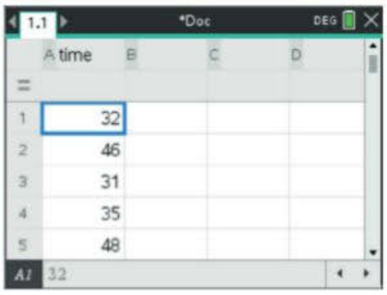
## USING CAS 1 Constructing a histogram from raw data

The following data shows the times (in seconds) for swimmers at a swim school to complete a 50-metre freestyle time trial.

32, 46, 31, 35, 48, 40, 40, 55, 33, 47, 53, 36, 58, 45, 41,
40, 55, 31, 56, 48, 39, 42, 43, 46, 39, 34, 51, 40, 44

Construct a histogram for the data using intervals of 30–<34, 34–<38 and so on.

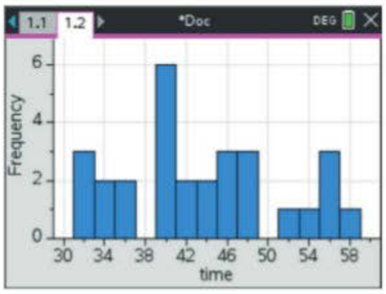
**TI-Nspire**




1 Start a new document and add a **Lists & Spreadsheet** page.

2 Label column **A** as **time**.

3 Enter the ungrouped scores as shown above.




4 Insert a **Data & Statistics** page.

5 For the horizontal axis, select **time**.

6 Press **menu > Plot Type > Histogram**.

7 A histogram of the data will be displayed.

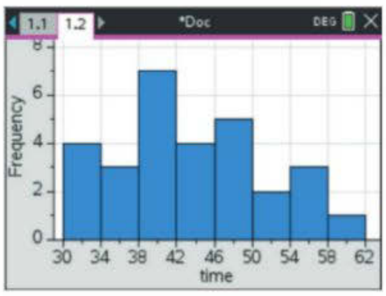

8 Press **menu > Plot Properties > Histogram Properties > Bin Settings > Equal Bin Width**.

9 Enter the following settings:

**Width:** 4

**Alignment:** 30

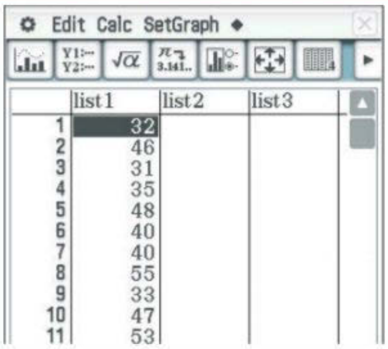
10 Select **OK**.

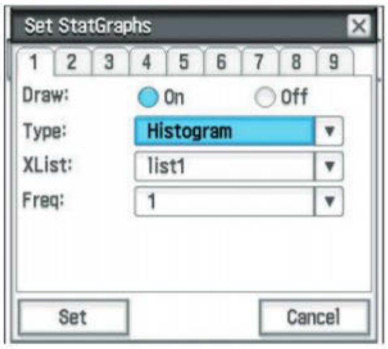



11 A histogram of the grouped data in intervals of 4 will be displayed.

12 Adjust the **Window/Zoom** settings or grab the horizontal and vertical axes and drag them to view the full histogram.

## ClassPad




1 Tap Menu and open the Statistics application.

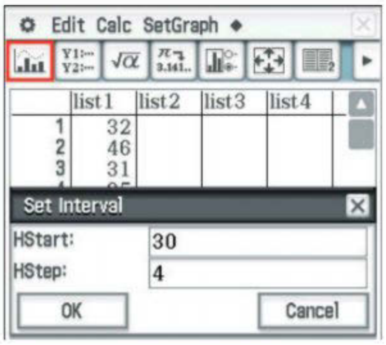
2 Clear all lists.

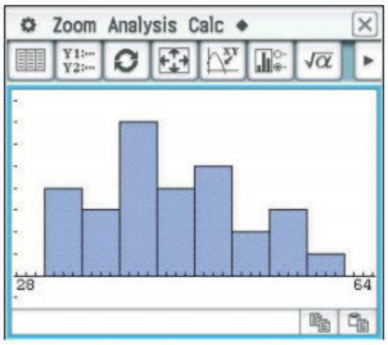
3 Enter the ungrouped scores into listl as shown above.

4 Tap SetGraph > Setting.

5 In the dialogue box, select the following:
Type: Histogram
XList: listl
Freq: 1

6 Tap Set.




7 Tap Graph.

8 In the dialogue box, set the following:
HStart: 30
HStep: 4

9 Tap OK.

10 A histogram of the data in intervals of 4 will appear in the lower window.

## EXERCISE 1.2 Histograms

ANSWERS p. 693

### Recap

1 ©VCAA 2Q17N1CQ4 The association between the *amount of solar energy* captured by a solar panel, in megajoules, and the *capital city* (Melbourne, Adelaide etc.) in which it is captured will be investigated. The variables *amount of solar energy* and *capital city* are

A both numerical variables.

B both categorical variables.

C a nominal variable and a numerical variable respectively.

D a numerical variable and a nominal variable respectively.

E a numerical variable and an ordinal variable respectively.

▶ 2 ©VCAA | 20171CQ7 I 46% A study was conducted to investigate the association between the *number of moths* caught in a moth trap (less than 250, 250-500, more than 500) and the *trap type* (sugar, scent, light). The variables *number of moths* and *trap type* are

A both nominal variables.

B both ordinal variables.

C a numerical variable and a categorical variable respectively.

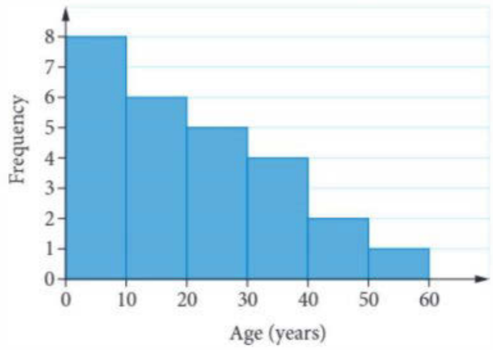
D a nominal variable and an ordinal variable respectively.

E an ordinal variable and a nominal variable respectively.

1.2

## Mastery

30 WORKED EXAMPLE 4 The following histogram shows the ages of people (in years) at an indoor playground.



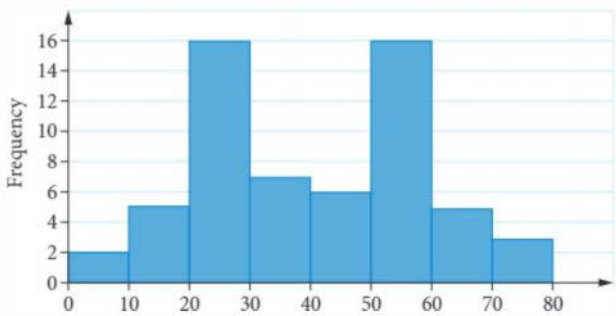

a How many people are in their 20s or 30s?

b Is the histogram approximately symmetric, positively skewed or negatively skewed? Does it have any possible outliers?

c What is the modal interval?

d In which interval is the median?

4 For the histogram shown, state whether each of the statements is true or false.




a The histogram is positively skewed,

b The histogram is double-peaked,

c The histogram has a possible outlier,

d The histogram is bi-modal.

e The histogram is approximately symmetric,

f The modal interval is 10-<20.

g The mean and median are both around 40.

5 a using CAS 1 The data below shows the time taken (in seconds) for athletes at an athletics club to complete a 400-metre sprint time trial.

72, 66, 80, 65, 78, 60, 81, 75, 64, 67, 74, 66, 58, 65, 61, 82, 75, 61, 76, 78, 69, 62, 64, 76, 49, 64, 51, 72, 64

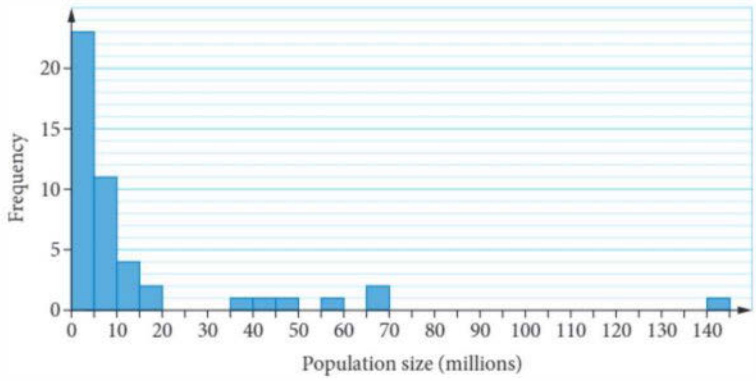
Construct a histogram for the above data using intervals of 48-<52, 52-<56 and so on.

## Exam practice

80-100% 60-79% 0-59%

*Use the following information to answer the next two questions.*

The histogram below shows the distribution of the *population size* of 48 countries in 2018.




Datasource: Worldometers, www.worldometers.info

6 ©VCAA 2019 1CQ1 92% The number of these countries with a *population size* between 5 million and 20 million people is

A 11 B 17 C 23 D 34 E 35

7 ©VCAA 20191CQ2 86% The shape of this histogram is best described as

A positively skewed with no outliers.

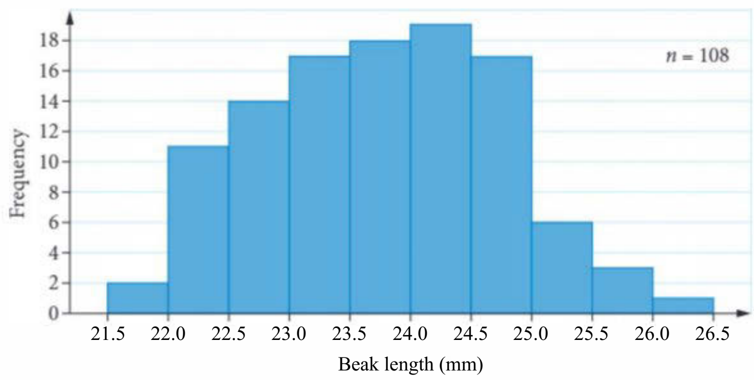
B positively skewed with outliers.

C approximately symmetric.

D negatively skewed with no outliers.

E negatively skewed with outliers.

8 ©VCAA 2018N 1CQ3 The histogram shows the distribution of *beak length,* in millimetres, of a sample of 108 birds of the same species. Both male and female birds are included in this sample.




The *beak length* for this sample of 108 birds is most frequently

A greater than or equal to 22.5 mm and less than 23.0 mm.

B greater than or equal to 23.0 mm and less than 23.5 mm.

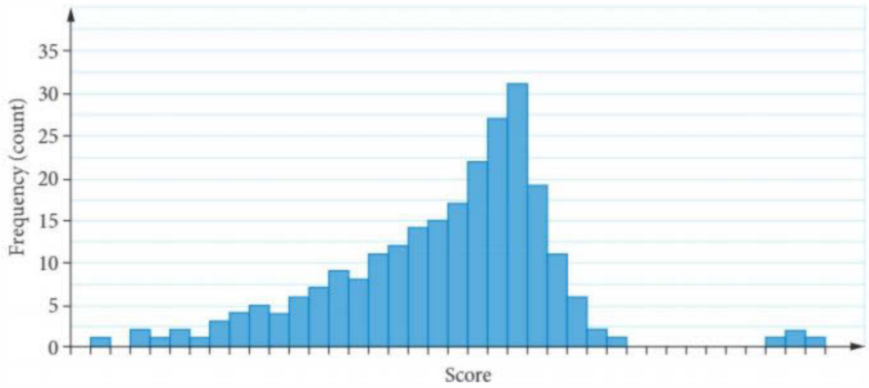
C greater than or equal to 23.5 mm and less than 24.0 mm.

D greater than or equal to 24.0 mm and less than 24.5 mm.

E greater than or equal to 24.5 mm and less than 25.0 mm.

1.2

**9** ©VCAA 20051CQ3 | 74% The histogram below is best described as




A negatively skewed.

B positively skewed.

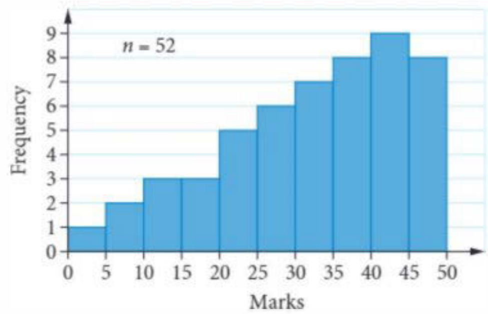
C symmetric.

D negatively skewed with outliers.

E positively skewed with outliers.

*Use the following information to answer the next three questions.*

The histogram below shows the distribution of marks obtained by 52 students on a test.




10 **ES** 2017N1CQ1 J The shape of the distribution of marks is

A symmetric.

B bell-shaped.

C positively skewed.

D negatively skewed.

E approximately normal.

11 © VCA 20[illegible]2 The pass mark for the test was set at 25. The percentage of students who passed this test was closest to

A 14%

B 28%

C 32%

D 52%

E 73%

12 ©VCAA 2017N 1CQ3 The median mark for the test was

A greater than 20 but less than 25.

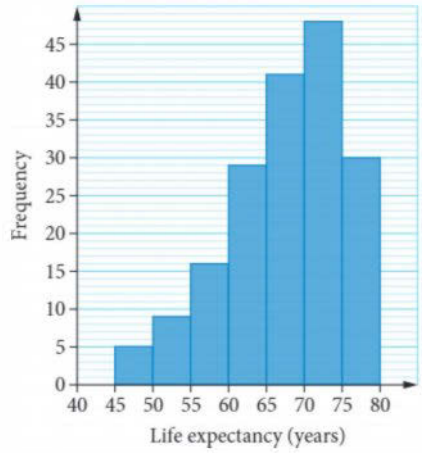
B greater than 25 but less than 30.

C greater than 30 but less than 35.

D greater than 35 but less than 40.

E greater than 40 but less than 45.

13 ©VCAA | 20i5 2CQibcn (2 marks) The histogram shows the distribution of life expectancy of people for 183 countries.

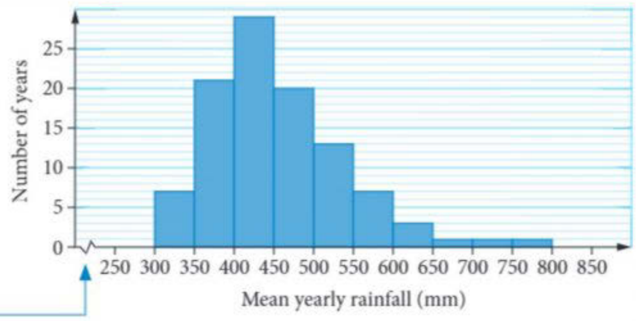



a 77% In how many of these countries is life expectancy less than 55 years? 1 mark

b 68% In what percentage of these 183 countries is life expectancy between 75 and 80 years? Write your answer correct to one decimal place. 1 mark

14 © VCAA 2007 2CQ1 I 69% (3 marks) The histogram shows the distribution of mean yearly rainfall (in mm) for Australia over 103 years.

Sometimes a break like this is used to indicate that not all of the scale marks from zero have been included.




Data source: Australian Bureau of Statistics (ABS) 2007

a Describe the shape of the histogram. 1 mark

b Use the histogram to determine

i the number of years in which the mean yearly rainfall was 500 mm or more 1 mark

ii the percentage of years in which the mean yearly rainfall was between 500 mm and 600 mm. Write your answer correct to one decimal place. 1 mark

# 1.3 Boxplots

Video playlist Box plots

Worksheet Five-number summaries

## The five-number summary

The median is the value that divides an ordered data set in half. The quartiles are three values that divide an ordered data set in quarters.

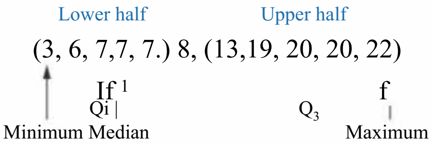
- Qj or the lower quartile is the median of the lower half of the data.
- $Q_2$ is the median of the whole data set.
- $Q_3$ or the upper quartile is the median of the upper half of the data.

The five-number summary provides a good overview of a distribution. It is created using the maximum and minimum data values and the three quartiles.

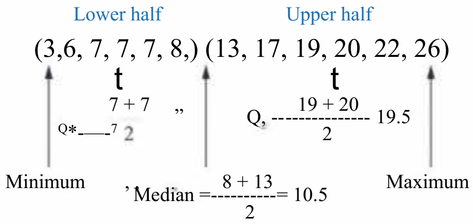


9780170448406

Example with an odd number of data values:




Example with an even number of data values:

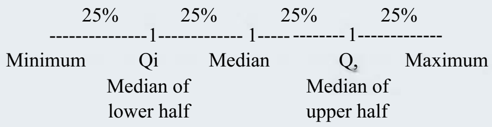





## The five-number summary

The five-number summary divides the data into segments of 25%:

- 25% of the data is less than $Q_1$
- 50% of the data is less than the median ($Q_2$).
- 75% of the data is less than $Q_3$.




These percentages are not exact for an odd-numbered data set because the median is one of the data values and we can't split a data value in half.

**WORKED EXAMPLE 5** Finding the five-number summary by hand

WB p.6

For the following data

34, 25, 23, 22, 24, 34, 21, 48, 6, 30, 21, 29

**a** find the five-number summary by hand

**b** use a diagram to show that

  **i** 25% of the data is less than the lower quartile ($Q_1$)

  **ii** 50% of the data is less than the median ($Q_2$)

  **iii** 75% of the data is less than the upper quartile ($Q_3$).

| Steps | Working |
|---|---|
| **a 1** Order the data from smallest to largest. | 6, 21, 21, 22, 23, 24, 25, 29, 30, 34, 34, 48 |
| **2** Find the minimum and maximum value. | min = 6, max = 48 |
| **3** Find the median. There is an even number of data values, so find the average of the two middle points. | $Q_2$ = median = $\frac{24 + 25}{2}$ = 24.5 |
| **4** Find $Q_1$, the median of the lower half of the data. | The lower half of the data is 6, 21, 21, 22, 23, 24. $Q_1$ = lower quartile = $\frac{21 + 22}{2}$ = 21.5 |
| **5** Find $Q_3$, the median of the upper half of the data. | The upper half of the data is 25, 29, 30, 34, 34, 48. $Q_3$ = upper quartile = $\frac{30 + 34}{2}$ = 32 |
| **6** List the five-number summary. | min = 6, $Q_1$ = 21.5, median = 24.5, $Q_3$ = 32, max = 48 |
| **b 1** Draw a diagram showing the three quartiles. |  |

2 Use the diagram to calculate the percentage of data less than each quartile.

i 3 out of a total of 12 data values are less than the lower quartile. $\frac{3}{12} = 25\%$

ii 6 out of a total of 12 data values are less than the median. $\frac{6}{12} = 50\%$

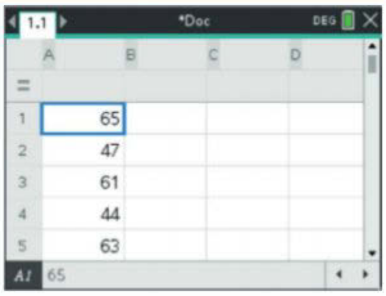
iii 9 out of a total of 12 data values are less than the upper quartile. $\frac{9}{12} = 75\%$

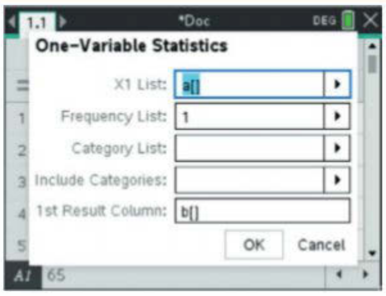
## USING CAS 2 Finding the five-number summary

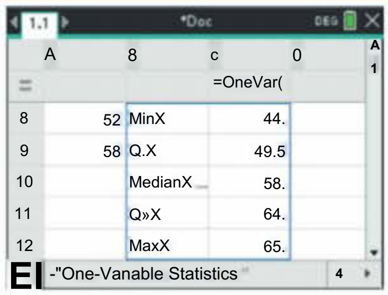
Use CAS to calculate the five-number summary for the following data:

65, 47, 61, 44, 63, 56, 65, 52, 58

### TI-Nspire




1 Start a new document and add a Lists & Spreadsheet page.

2 Enter the data into column A as shown.

3 Press menu > Statistics > Stat Calculations > One-Variable Statistics.

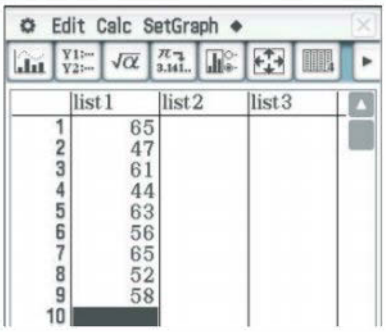
4 On the next screen, keep the number of lists default setting of 1 and select OK.

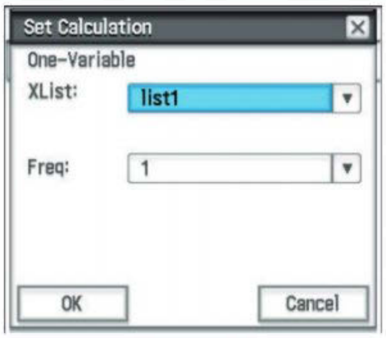
5 Leave the X1 List default setting of a[ ] then select OK.

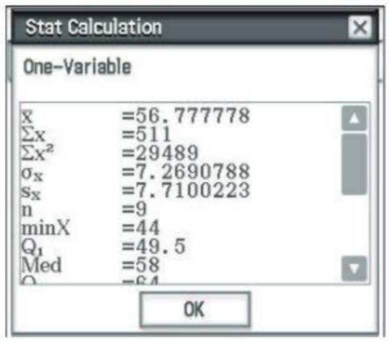
6 The labels will appear in column B and the corresponding one-variable statistics will appear in column C.

7 Scroll down to view the five-number summary values.

### ClassPad




1 Tap Menu and open the Statistics application.

2 Clear all lists and enter the data as shown.

3 Tap Calc > One-Variable.

4 Leave the XList default setting of listl.

5 Tap OK.

6 The one-variable statistics will be displayed.

7 Scroll down to view the five-number summary values.

## Exam hack

If a data set is already ordered, it can sometimes be quicker to find the five-number summary by hand than by using CAS.

## IQR, outliers and fences

The interquartile range (IQR) is the measure of the spread of the middle 50% of the data values.

$IQR=Q_3-Q,$

The IQR is usually a better measure of spread than the range because, by looking at only the middle 50% of data, we avoid taking outliers into account. The IQR is also used in a calculation to identify possible outliers, which allows us to do more than say something 'looks like an outlier'.

Worksheet
Interquartile range

1.3

**Interquartile range and fences**

IQR-Q3-Q,

A data value is a possible outlier if it is
less than the lower fence: Qj - 1.5 x IQR
or
greater than the upper fence: $Q_3$ + 1.5 x IQR.

**WORKED EXAMPLE 6** Finding outliers

For the ordered data set

4, 6, 21, 21, 22, 23, 24, 25, 29, 30, 34, 34, 50

do a calculation to show whether the blue values are possible outliers.

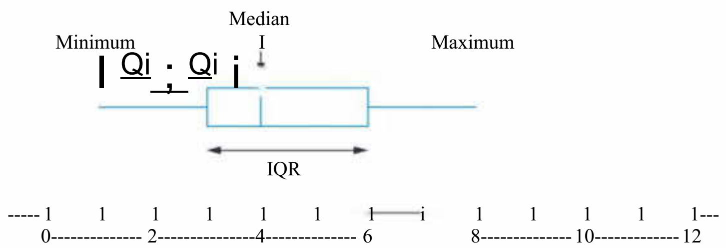
| Steps | Working |
|---|---|
| 1 Find Qj and $Q_3$ by using CAS or by hand. | Qj - 21 and $Q_3$ - 32 |
| 2 Calculate the IQR. | IQR =32-21-11 |
| 3 Calculate the lower and upper fences. | lower fence:<br>Qi - 1.5 x IQR= 21 - 1.5 x 11 = 4.5<br>upper fence:<br>$Q_3$ + 1.5 x IQR= 32 + 1.5 x 11 = 48.5 |
| 4 Check each of the blue values to see if they are less than the lower fence or greater than the upper fence. | 4 is less than 4.5 so it *is* a possible outlier.<br>6 is not less than 4.5, so it is *not* an outlier.<br>50 is greater than 48.5, so it *is* a possible outlier. |

WB
p.7

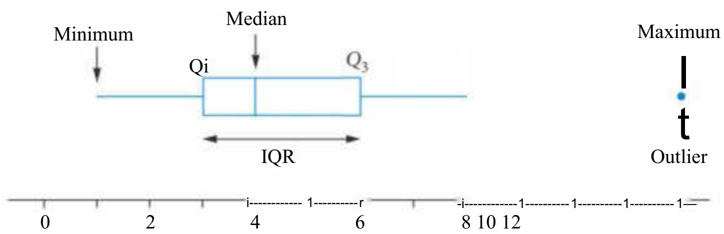
Puzzle
Statistical measures puzzle

## Boxplots

Boxplots, also known as box-and-whisker plots, display numerical data based on the five-number summary, IQR and outliers. If there are no outliers, the whiskers show the minimum and maximum values.




Outliers are shown as dots. If there are outliers, the lowest outlier is the minimum value and the highest outlier is the maximum value.



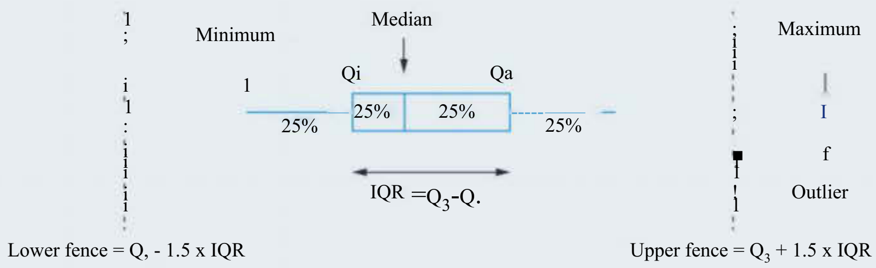

Worksheets
Boxplots
Boxplots 1
Box plots 2

**Exam hack**

You need to include outliers when finding the minimum and maximum values.

## Boxplots

Boxplots provide the following information:




Boxplots can also be displayed vertically.

### @ Exam hack

Fences are not part of a boxplot. You don't need to draw them.

## USING CAS 3 Constructing boxplots

The weekly wages ($) of twenty university students who work part-time is given below.

304, 285, 331, 270, 306, 294, 299, 302, 295, 299, 294, 315, 291, 299, 314, 270, 299, 284, 270, 302

Construct a boxplot of the above data.

**TI-Nspire**


1 Start a new document and add a **Lists & Spreadsheet** page.

2 Label column **A 'wages'** and enter the weekly wages as shown.

3 Insert a **Data & Statistics** page.

4 Click on the horizontal axis to add the variable '**wages**'.

5 The data will be displayed as a dot plot.

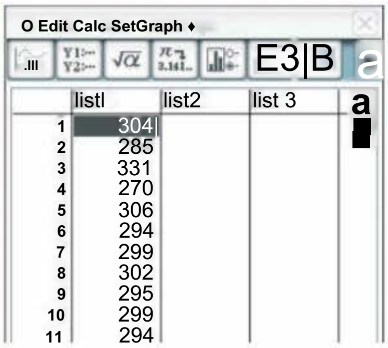
6 Press **menu > Plot Type > Box Plot.**

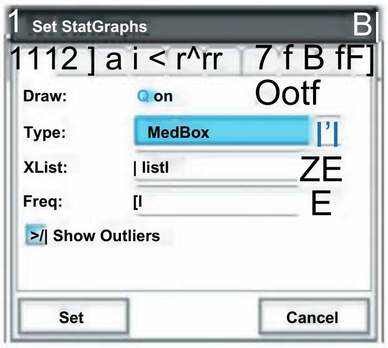
7 The data will be displayed as a boxplot.

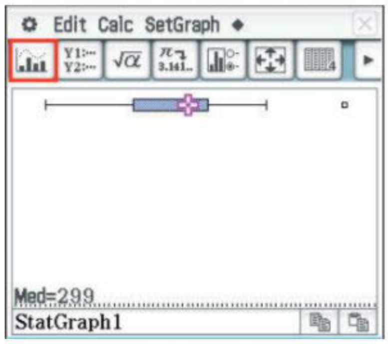
8 As you move the cursor over the boxplot, the quartile and outlier values will appear.

1.3

## ClassPad




1 Tap Menu and open the Statistics application.

2 Clear all lists and enter the weekly wages, as shown.

3 Tap SetGraph > Setting.

4 Tap on the Type: field. From the dropdown menu, select MedBox.

5 Tap on Show Outliers (optional).

6 Tap Set.

7 Tap on the Graph tool to display the data as a boxplot in the lower window.

8 While in the graph window, tap Analysis > Trace.

9 Press the left and right arrows to display the quartile and outlier values.

**WORKED EXAMPLE 7** Reading boxplots

WB

p. 8

The boxplot shows the distribution of 64 student test scores marked out of 20.


Find the

**a** five-number summary

**b** percentage of students who scored more than 17

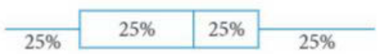
**c** percentage of students who scored less than 15

**d** percentage of students who scored between 10 and 17

**e** number of students who scored less than 11

**f** scores at the lower end that would be considered outliers

**g** scores at the upper end that would be considered outliers.

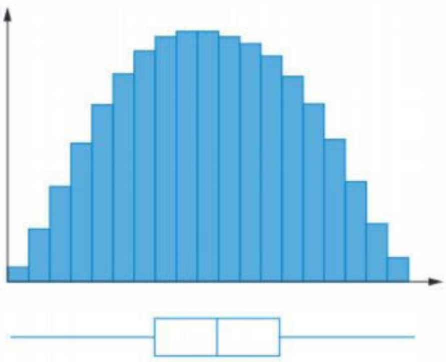
| Steps | Working |
|---|---|
| **a** Read directly from the boxplot. | min = 10, $Q_1$ = 11, median = 15, $Q_3$ = 17, max = 19 |
| **b–d** Use the fact that quartiles divide data into four equal groups, so 25% of the data is in each group.  | $Q_3$ = 17 so 25% of students scored more than 17. median = 15, so 50% of students scored less than 15. $Q_3$ = 17, so 75% of students scored between 10 and 17. |
| **e** Find the percentage first and then multiply by the total number. | $Q_1$ = 11, so 25% of students scored less than 11. Total number of students = 64 Number of students who scored less than 11 = 64 × 25% = 16 |

| | |
|---|---|
| f Use the IQR to calculate the lower fence,<br>lower fence - Qj - 1.5 x IQR | IQR = $Q_3$-QJ = 17-11 = 6<br>lower fence =11-1.5x6=11-9 = 2<br>Scores less than 2 would be considered outliers. |
| g Use the IQR to calculate the upper fence,<br>upper fence = $Q_3$ + 1.5 x IQR | upper fence = 17 + 1.5 x 6 = 17 + 9 = 26<br>Scores greater than 26 would be considered outliers.<br>The maximum score possible is 20, so no scores at the upper end would be considered outliers. |

## Comparing boxplots and histograms

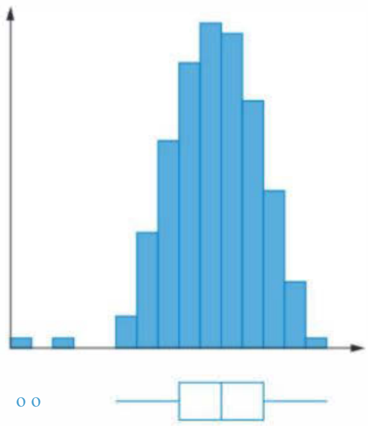
The histogram of a distribution can often provide an idea of what the boxplot of the data looks like.

Approximately symmetric distributions



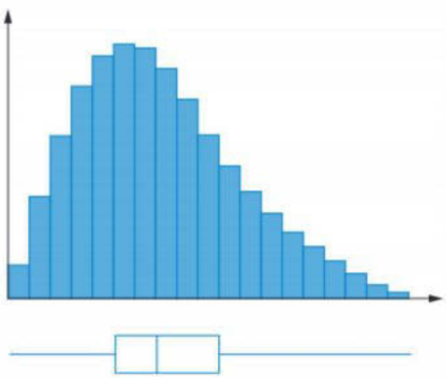
The median is approximately in the middle of the box and the whiskers are about the same length.

Distributions with outliers



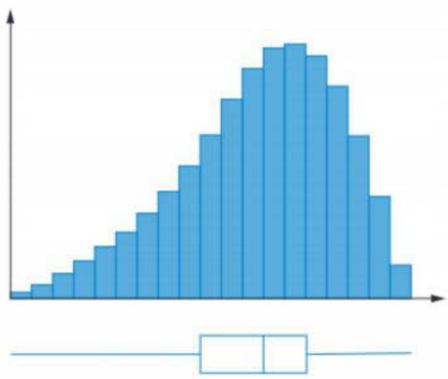
The boxplot matches the histogram with the outliers shown by dots.

Positively skewed distributions



The box and whisker in the positive direction are longer than the box and whisker in the negative direction.

Negatively skewed distributions



The box and whisker in the negative direction are longer than the box and whisker in the positive direction.

## EXERCISE 1.3 Boxplots

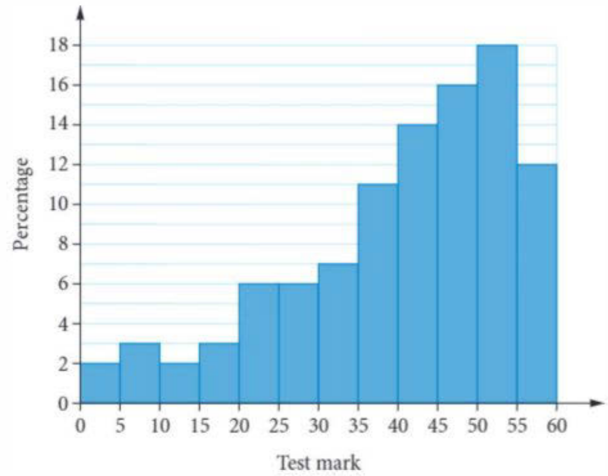
ANSWERS p. 693

1.3

### Recap

*Use the following information to answer the next two questions.*

The distribution of test marks obtained by a large group of students is displayed in the percentage frequency histogram.




1 Which of the following best describes the histogram?

A It has four outliers.
B It is negatively skewed.
C It is symmetric.
D It is positively skewed.
E It is bi-modal.

2 © VCAA 2006 1CQ5 85% The pass mark on the test was 30 marks. The percentage of students who passed the test is

A 7%
B 22%
C 50%
D 78%
E 87%

### Mastery

3G WORKED EXAMPLE 5 J a using CAS 2 J For the following test scores

73, 65, 54, 90, 74, 51, 61, 88, 47, 92, 71, 66

a find the five-number summary by hand and verify your answers by using CAS.

b use a diagram to show that

i 25% of the data is less than the lower quartile (Qj)

ii 50% of the data is less than the median ($Q_2$)

iii 75% of the data is less than the upper quartile ($Q_3$).

4 gj WORKED EXAMPLE 6 J a using CAS 3^] For each of the following data sets

i do a calculation to show whether the blue values are possible outliers

ii use CAS to construct a boxplot.

a 15, 26, 37, 37, 38, 39, 41,42, 43, 44,44, 45,46, 48, 49, 50, 50, 52, 54, 64

b 40, 77, 79, 86, 88, 89, 93,95, 96, 97, 99, 99, 104,105, 110, 110,115,156

c 15, 17, 20,21, 22,23, 24,25,25, 25, 25, 26, 27,27, 27,29

▶ **5S** WORKED EXAMPLE 7 The boxplots below show the distribution of 84 student test scores marked out of 20. For each of the boxplots, find the

i five-number summary

ii percentage of students who scored more than 8

iii percentage of students who scored less than 11

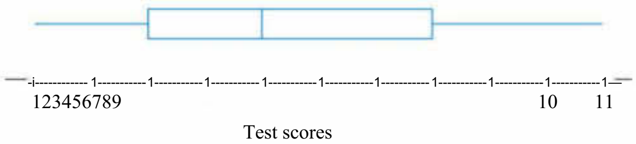
iv percentage of students who scored between 3 and 11

v number of students who scored less than 3

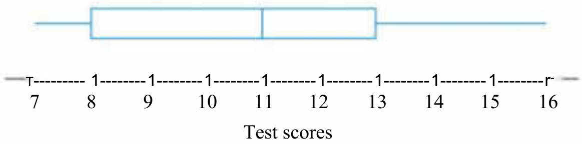
vi scores at the lower end that would be considered outliers

vii scores at the upper end that would be considered outliers.

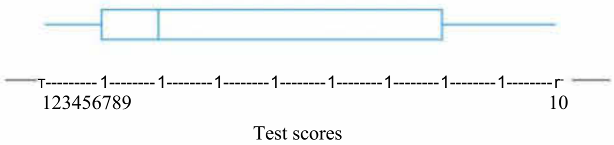
**a**




**b**

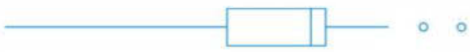



**c**




6 For each of the following boxplots, state whether the distribution is approximately symmetric, positively skewed or negatively skewed and whether it has outliers, giving a reason for your answer.

**a**



**b**

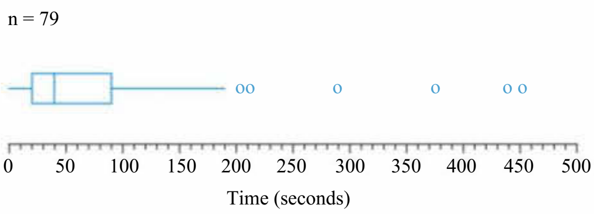
**c**

## Exam practice

80–100% 60–79% 0–59%

*Use the following information to answer the next three questions.*

The boxplot shows the distribution of the time, in seconds, that 79 customers spent moving along a particular aisle in a large supermarket.




Data source: www.stars.ac.uk

7 ©VCAA | 20081CQ2 J 83% The shape of the distribution is best described as

A symmetric.

B negatively skewed.

C negatively skewed with outliers.

D positively skewed.

E positively skewed with outliers.

8 © VCAA 2008 1CQ1 82% The longest time, in seconds, spent moving along this aisle is closest to

A 40 B 60 C 190 D 450 E 500

9 ©VCAA 2008 1CQ3 J 42% The number of customers who spent more than 90 seconds moving along this aisle is closest to

A 7 B 20 C 26 D 75 E 79

*Use the following information to answer the next two questions.*

The times between successive nerve impulses *(time)*, in milliseconds, were recorded. The table shows the five-number summary calculated using 800 recorded data values.

| | Time (milliseconds) |
|---|---|
| Mean | 220 |
| Minimum value | 10 |
| First quartile (Qj) | 70 |
| Median | 150 |
| Third quartile ($Q_3$) | 300 |
| Maximum value | 1380 |

Data: adapted from P Fatt and B Katz, 'Spontaneous subthreshold activity at motor nerve endings', *The Journal of Physiology*, 117,1952, pp. 109-128

### Exam hack

With questions involving boxplots, always check whether the value you are being asked to find is one of the quartiles.

10 ©VCAA I20201CQ2J 73% Of these 800 times, the number of times that are longer than 300 milliseconds is closest to

A 20 B 25 C 75 D 200 E 400

11 ©VCAA I20201CQ3 59% The shape of the distribution of these 800 times is best described as

A approximately symmetric.

B positively skewed.

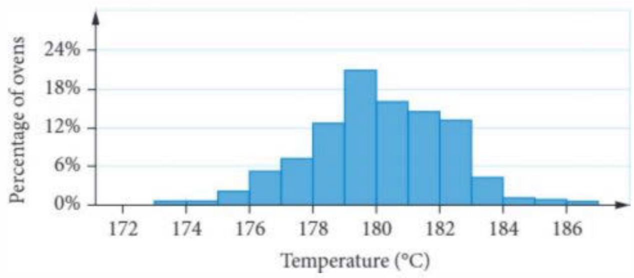
C positively skewed with one or more outliers.

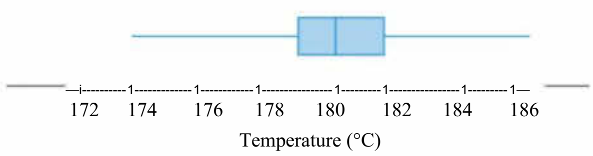
D negatively skewed.

E negatively skewed with one or more outliers.

*Use the following information to answer the next two questions.*

To test the temperature control on an oven, the control is set to 180°C and the oven is heated for 15 minutes. The temperature of the oven is then measured. Three hundred ovens were tested in this way. Their temperatures were recorded and are displayed using both a histogram and a boxplot.




12 ©VCAA 2010 1CQ2 70% The interquartile range for temperature is closest to

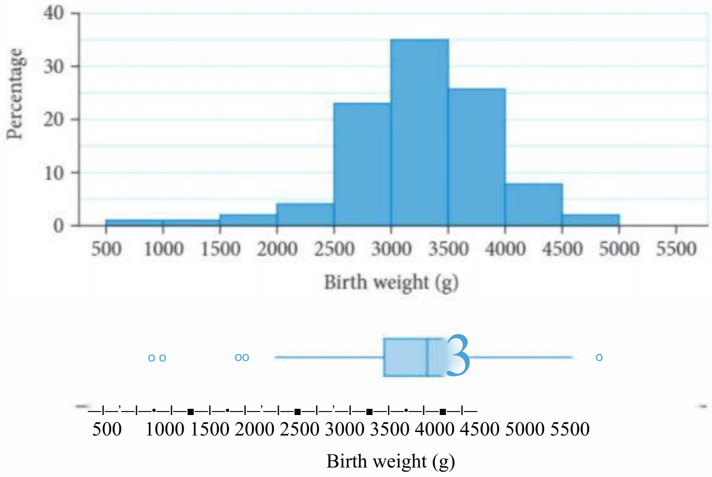
A 1.3°C B 1.5°C C 2.0°C D 2.7°C E 4.0°C

13 ©VCAA 2010 1CQ1 55% A total of 300 ovens were tested and their temperatures were recorded.

The number of these temperatures that lie between 179°C and 181 °C is closest to

A 40 B 50 C 70 D 110 E 150

▶ 14 ^^^3~2oi~9N ICQI-2 MODIFIED J (4 marks) The histogram and boxplot shown below both display the distribution of the *birth weight,* in grams, of 200 babies.




a Describe the shape of the distribution of the babies' *birth weight.* 2 marks

b How many babies are there with a *birth weight* between 3000 g and 3500 g? 1 mark

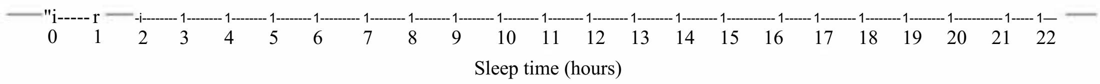
c What is the range of the distribution? 1 mark

15 ©VCAA 2019N 2CQ2 (3 marks) The five-number summary was determined from the *sleep time,* in hours, of a sample of 59 types of mammals.

| Statistic | Sleep time (hours) |
|---|---|
| Minimum | 2.5 |
| First quartile | 8.0 |
| Median | 10.5 |
| Third quartile | 13.5 |
| Maximum | 20.0 |

a Show, with calculations, that a boxplot constructed from this five-number summary will not include outliers. 2 marks

b Copy the following number line and use it to construct the boxplot. 1 mark




# 1 @ Log scales

Video playlist Log scales

Worksheet Significant figures

## Rounding to significant figures

Rounding is an important skill when dealing with real-life data. In practical situations, it often makes sense to round an answer to a number of significant figures rather than to a number of decimal places, particularly when a combination of large whole numbers and decimals are involved. We need to be able to do both methods of rounding.

1.4

In the following examples, the significant figures are in red.

| Example | No. of significant figures | What type of digits are significant? | What type of digits are not significant? |
|---|---|---|---|
| 367.268 | 6 | non zero digits | |
| 190.0043 | 7 | zeros between non-zero digits | |
| 3.540 | 4 | trailing zeros in decimals | |
| 0.0043 | 2 | | leading zeros in decimals |
| 1390 | 3 | | trailing zeros in whole number |

When rounding to significant figures, use the same rounding rules as for rounding to a number of decimal places:

- *0-4 round down and '5-9 round up'
  e.g. 27 501 rounded to two significant figures is 28000
- round the 9 to 0 and carry the rounding over to the next digit on the left
  e.g. 3.9722 rounded to two significant figures is 4.0
  497 rounded to two significant figures is 500

### Significant figures

Significant figures:

- any non-zero digit
- zeros between non-zero digits
- trailing zeros in decimals.

Not significant figures:

- leading zeros in decimals
- trailing zeros in whole numbers.

When rounding to significant figures, use usual rounding rules:

- '0-4 round down and '5-9 round up'
- round the 9 to 0 and carry the rounding over to the next digit on the left
- include trailing zeros in decimals, if necessary.

### Exam hack

Note that the two methods of rounding can give very different results.

20718.039 rounded to two decimal places is 20718.04.

20718.039 rounded to two significant figures is 21000.

### WORKED EXAMPLE 8 [-Rounding to decimal places versus significant figures

WB p. 9

Round each number to

i two decimal places ii two significant figures.

a 1.333 b 6.268 c 0.5563 d 14700

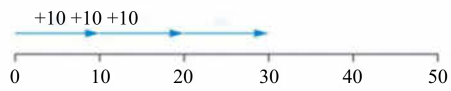
e 10.882 f 76008.037 g 12.8989

| Steps | Working |
|---|---|
| a i Focus on the first two decimal places. | 1.333 rounded to two decimal places is 1.33. |
| ii Focus on the first two significant figures. | 1.333 rounded to two significant figures is 1.3. |
| b i Focus on the first two decimal places. | 6.268 rounded to two decimal places is 6.27. |
| ii Focus on the first two significant figures. | 6.268 rounded to two significant figures is 6.3. |
| c i Focus on the first two decimal places. | 0.5563 rounded to two decimal places is 0.56. |
| ii Focus on the first two significant figures. | 0.5563 rounded to two significant figures is 0.56. |
| d i Focus on the first two decimal places. | 14700 rounded to two decimal places is 14700.00. |
| ii Focus on the first two significant figures. | 14700 rounded to two significant figures is 15000. |
| e i Focus on the first two decimal places. | 10.882 rounded to two decimal places is 10.88. |
| ii Focus on the first two significant figures. | 10.882 rounded to two significant figures is 11. |

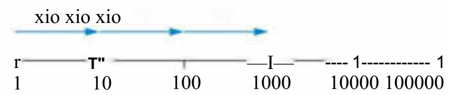
| | |
|---|---|
| f i Focus on the first two decimal places. | 76 008.037 rounded to two decimal places is 76008.04. |
| ii Focus on the first two significant figures. | 76 008.037 rounded to two significant figures is 76000. |
| g i Focus on the first two decimal places. | 12.8989 rounded to two decimal places is 12.90. |
| ii Focus on the first two significant figures. | 12.8989 rounded to two significant figures is 13. |

## Linear and log scales

All the plots and graphs weve used so far have had linear scales. On a linear scale, we *add* the same number to move from one scale mark to the next. In the example on the right, the linear scale involves adding 10 each time.




There are some situations where it is better to use a log scale (or logarithmic scale). On a log scale, we *multiply* the same number to move from one scale mark to the next. In the example on the right, the log scale involves multiplying by 10 each time.

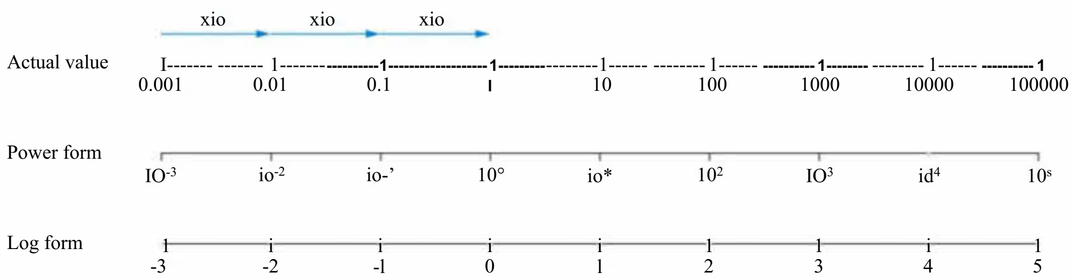



This is called a 'log base 10' or '$\log_{10}$' scale. Just as a linear scale can increase by numbers other than +10, a log scale can increase by numbers other than xlO. We will only be considering $\log_{10}$ scales, so when we refer to a log scale, we assume it's $\log_{10}$.

A log scale allows us to plot values such as 2,17, 2567 and 98654 on the *same* graph, whereas a linear base 10 scale would need to be very long to plot all of these values! Log scales are used in predicting the spread of corona viruses and trends in social media, and when measuring the brightness of stars and the magnitude of earthquakes.

## Reading log scales

Log scales let us plot very small values and very large values together. The following shows three different ways that the $\log_{10}$ scale from 0.001 to 100000 can be written:




We will be using the log form. The log form is a rearrangement of the power form.:

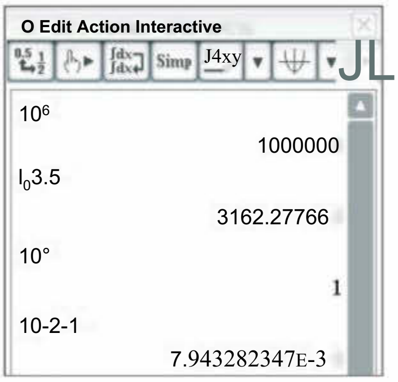
For example $10^4$ = 10000 is rearranged to $\log_{10}10000 = 4$.

To work out the actual values involved in a log form scale, we will need to find what the numbers appearing on the scale are as powers of 10. Unless its a simple whole number, we will need to use CAS or a scientific calculator. For example:

| Number appearing on log form scale | Actual value |
|---|---|
| 6 | $10^6$= 1000 000 |
| 3.5 | $10^{3.5}$ = 3162.28 (rounded to two decimal places) |
| 0 | $10^0$= 1 |
| -2.1 | $10^{-2.1}$ = 0.008 (rounded to three decimal places) |

**TI-Nspire**

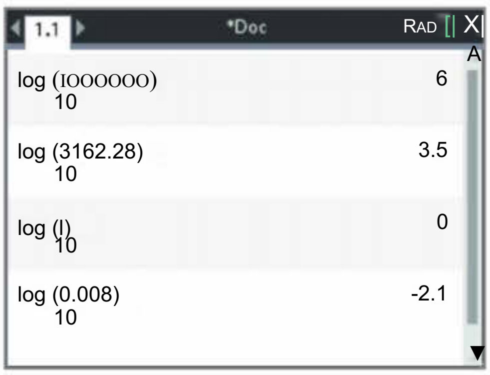

**ClassPad**



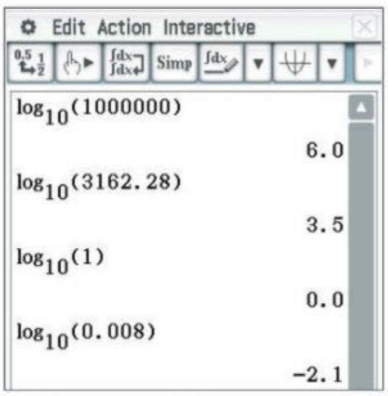

If the actual value is known and we want to find the number appearing on the log form scale, we use the $\log_{10}$ function on CAS or a scientific calculator.

| Actual value | Number appearing on log form scale |
|---|---|
| 1000000 | $\log_{10}$1000 000 - 6 |
| 3162.28 | $\log_{10}$ 3162.28 = 3.5 (rounded to two significant figures) |
| 1 | $\log_{10}$1 - 0 |
| 0.008 | $\log_{10}$0.008 = -2.1 (rounded to two significant figures) |

**TI-Nspire**




**ClassPad**



WB

p. 10

## WORKED EXAMPLE 9 Reading histograms with log scales

The histogram shows the *number of hits* on 78 websites surveyed by a market research company during one week.

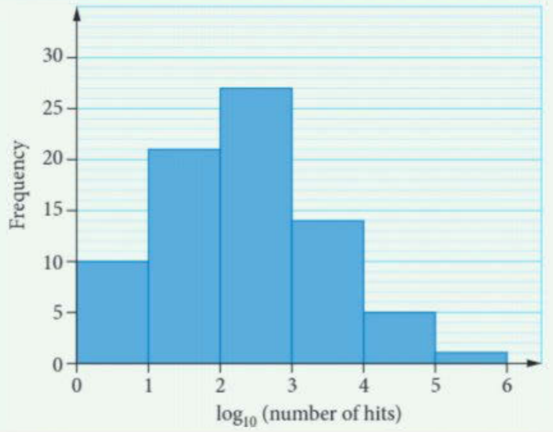
a How many websites had

  i over 100 000 hits during the week?

  ii over 10 000 hits during the week?

  iii under 100 hits during the week?

b What percentage of websites had between 1000 and 10000 hits during the week? Round to three significant figures.




| **Steps** | **Working** |
|---|---|
| a 1 Rewrite the log scale to show the actual values. | 1 10 100 1000 10000 100000 1000000<br>Number of hits |
| 2 Read the result using these actual values. | i 1 site had over 100000 hits.<br>ii 1 + 5 = 6 sites had over 10000 hits.<br>iii 10 + 21 = 31 sites had under 100 hits. |
| b Read the result using these actual values and convert it to a percentage, using<br>$\text{percentage} = \frac{\text{frequency}}{\text{total}} \times 100\%$<br>and rounding to three significant figures. | 14 sites<br>$\frac{14}{78} \times 100\% = 0.17948... \times 100\%$<br>$= 17.948...\%$<br>$= 17.9\%$ |

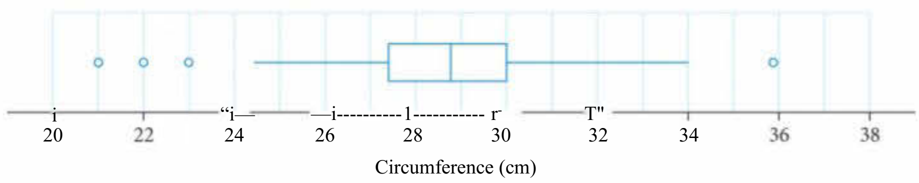
## EXERCISE 1.4 Log scales

ANSWERS p. 694

### Recap

*Use the following information to answer the next two questions.*

The boxplot shows the distribution of the forearm *circumference,* in centimetres, of 252 people.




1 ©VCAA 2017 1CQ1 J 93% The percentage of these 252 people with a forearm *circumference* of less than 30 cm is closest to

A 15% B 25% C 50% D 75% E 100%

2 ©VCAA 20171CQ2J 58% The five-number summary for the forearm *circumference* of these 252 people is closest to

A 21,27.4,28.7,30,34 B 21,27.4,28.7,30,35.9 C 24.5,27.4,28.7,30,34

D 24.5, 27.4, 28.7, 30, 35.9 E 24.5,27.4,28.7,30,36

## Mastery

3 R3 WORKED EXAMPLE 8 I Round each number to

i two decimal places ii two significant figures.

a 7.421 b 12.919 c 0.363 d 1800

e 20.666 f 72.0037 g 9.7962

4H WORKED EXAMPLE 9 The histogram shows the *number of subscribers* that 81 YouTubers have,

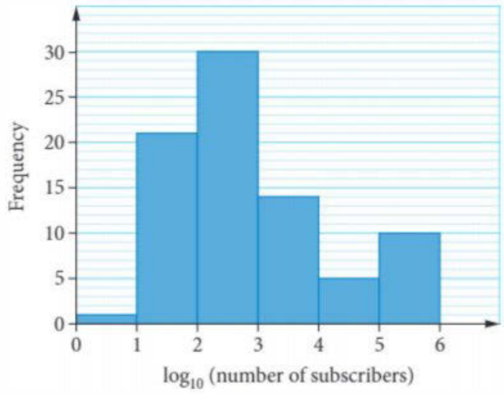
a How many YouTubers had

i under 10 subscribers?

ii under 100 subscribers?

iii over 1000 subscribers?

b What percentage of YouTubers had between 100000 and 1000000 subscribers? Round to four significant figures.




5 a For each of the following numbers appearing on a log scale, find their actual value, giving your answers rounded to three decimal places if necessary.

i 2 ii 1.3 iii 1 iv 0.7

v 0 vi -1 vii -1.2

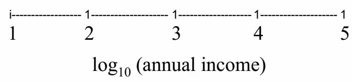
b For each of the following actual values, find the number that would appear on a log scale, giving your answers rounded to three significant figures.

i 100 ii 100000000 iii 65 iv 896

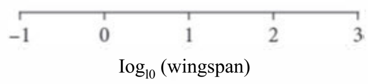
v 0.5 vi 20000 vii 8

6 Rewrite each of these $\log_{10}$ scales to show the actual values involved.

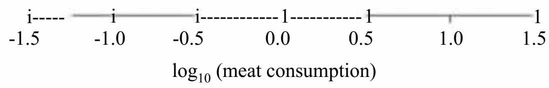
a Annual income of the population of an island in dollars.




b Wingspan of flying animals in millimetres.




c Annual per capita *meat consumption* of countries, in kilograms, rounded to one decimal place.




## Exam practice

80-100% 60-79% 0-59%

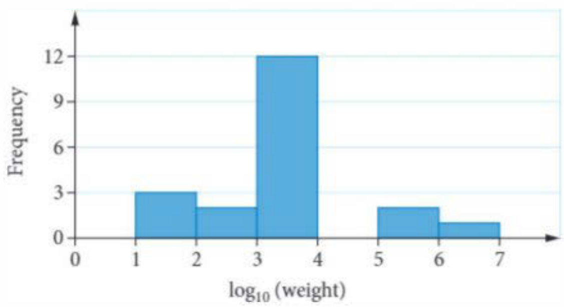
7 ©VCAA | 20201CQ5 78% The histogram shows the distribution of *weight,* in grams, for a sample of 20 animal species. The histogram has been plotted on a $\log_{10}$ scale.

The percentage of these animal species with a *weight* of less than 10000g is

A 17% B 70%

C 75% D 80%

E 85%



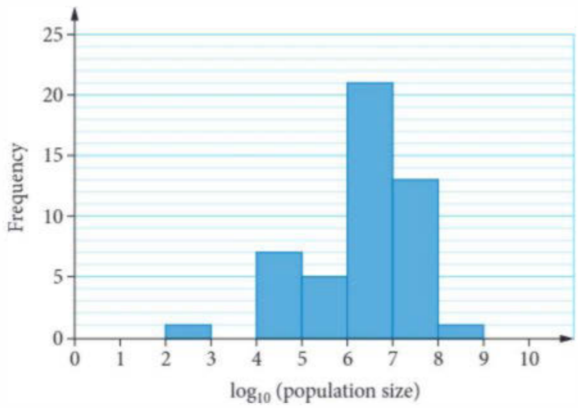
**8** ©VCAA | 20191CQ3 | 71% The histogram shows the *population size* for 48 countries plotted on a $\log_{10}$ scale.

Based on this histogram, the number of countries with a *population size* that is less than 100000 people is

A 1 B 5

C 7 D 8

E 48




Data: Worldometers, www.worldometers.info

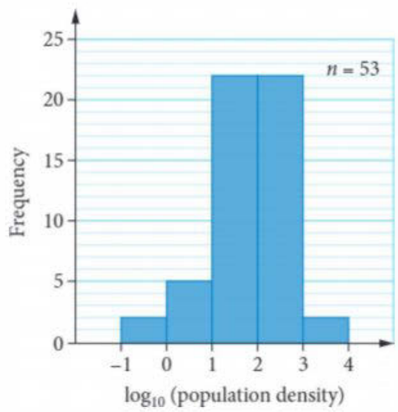
9 ©VCAA 2017N1CQ5 The histogram displays the distribution of *population density,* in people per square kilometre, for 53 countries. The horizontal scale of the histogram is log *(population density).*

Based on the histogram, how many countries have a population density that is less than 10 people per square kilometre?

A 2 B 5

C 7 D 29

E 46




Data: United Nations, Department of Economic and Social Affairs, Population Division (2015), 'World Population Prospects: The 2015 Revision' (DVD edition)

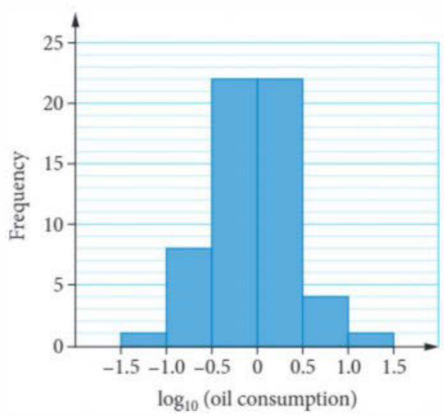
10 ©VCAA | 2016N1CQ7 The histogram displays the distribution of the annual per capita *oil consumption* (tonnes) for 58 countries plotted on a log scale.

The percentage of countries with an annual per capita *oil consumption* of more than 10 tonnes is closest to

A 1% B 2%

C 27% D 57%

E 98%



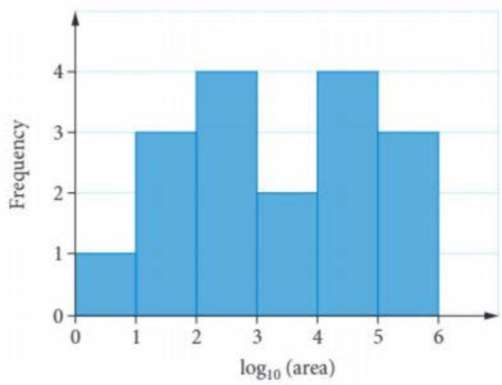

**11** ©VCAA | 20171CQ4 | 62% The histogram shows the distribution of the $\log_{10}$ (area), with *area* in square kilometres, of 17 islands.

The median area of these islands, in square kilometres, is between

A 2 and 3 B 3 and 4

C 10 and 100 D 1000 and 10000

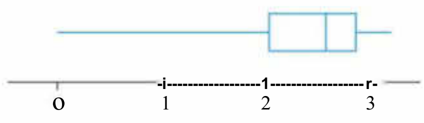
E 10000 and 100000



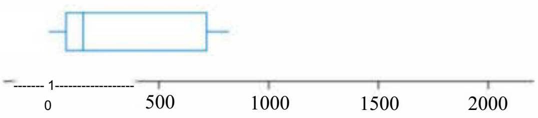



9780170448406

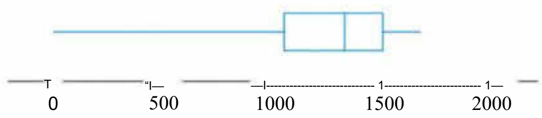
▶ 12 The boxplot to the right has a log base 10 scale. Which one of the following boxplots would best match its linear scale equivalent?



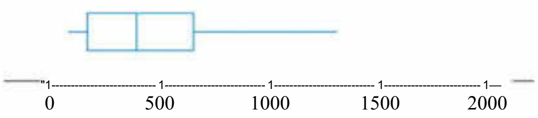

A



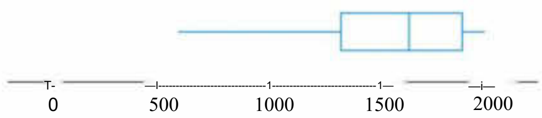

B



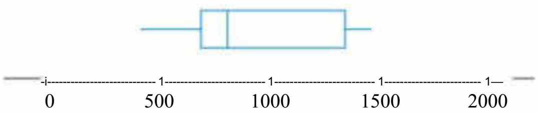

C




D

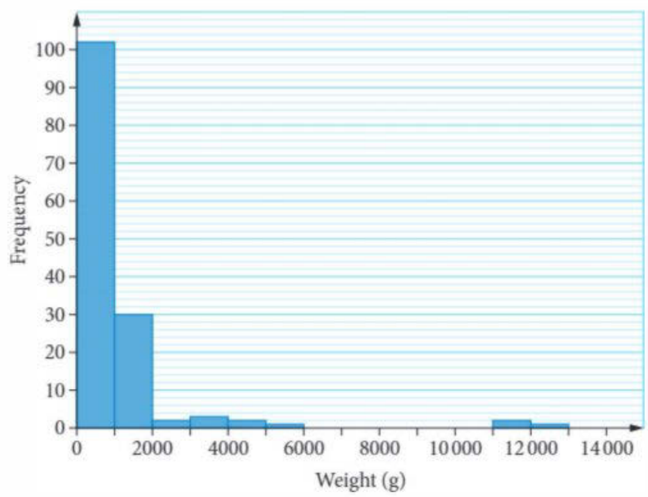



E




13 © VCAA 2018N 2CQ3 (3 marks) Histogram 1 displays the *weight* distribution of 143 birds of different species living in a small zoo.

Histogram 1

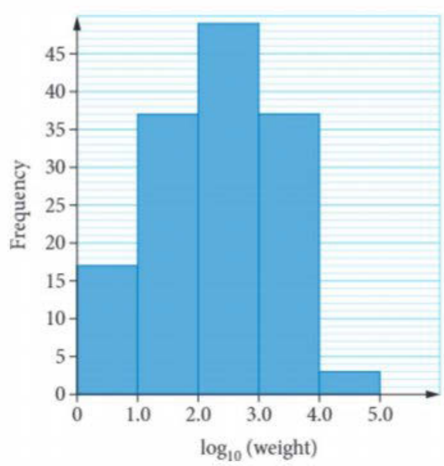



a Describe the shape of the distribution displayed in Histogram 1. Note the number of possible outliers, if any. 1 mark

b What percentage of these birds weigh less than 1000 g? Round your answer to one decimal place. 1 mark

c Histogram 2 displays the *weight* distribution of the same 143 birds plotted on a $\log_{10}$ scale.

Histogram 2




How many of these birds weigh between 10 g and 100 g? 1 mark

# 1 @ Dot plots and stem plots

Video playlist
Dot plots and stem plots

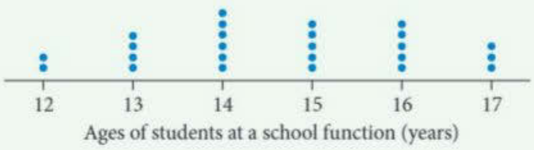
## Dot plots

Dot plots can be used for both categorical and numerical data. They are useful when there are not too many data values involved and the values are not too spread out.

WB
p. 11

**WORKED EXAMPLE 10** Using dot plots

The dot plot shows the ages of students at a school function.

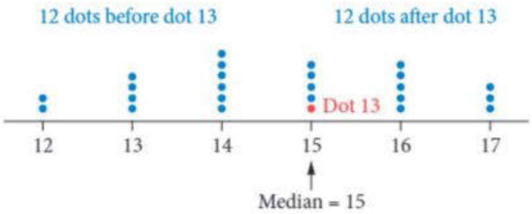
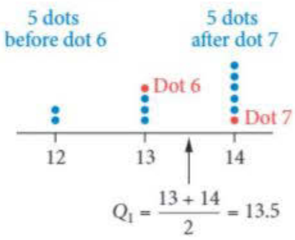



**a** Find the

**i** mode **ii** range **iii** median

**iv** lower quartile ($Q_1$) **v** upper quartile ($Q_3$) **vi** interquartile range (IQR).

**b** What could best describe the shape of the distribution: approximately symmetric, positively skewed, or negatively skewed?

| Steps | Working |
|---|---|
| **a** **i** Find the most common value. | mode = 14 years |
| **ii** range = largest value − smallest value | range = 17 − 12 = 5 years |
| **iii** **1** Count the number of dots $n$, note whether $n$ is odd or even, and find the position of the median. | $n = 25$; odd<br>$\frac{n+1}{2} = \frac{25+1}{2} = \frac{26}{2} = 13$<br>The median is the 13th ordered data value. |
| **2** If $n$ is odd, find the data value of the middle dot.<br>If $n$ is even, find the average of the data values for the two middle dots.<br>Count each column of dots from the bottom up to reach the median. |  |
| **3** Write the answer. | median = 15 years |
| **iv** **1** To find the lower quartile $Q_1$, find the median of the lower half.<br>Count the number of dots $n$, note whether $n$ is odd or even, and find the position of $Q_1$. | $n = 12$; even<br>$\frac{n+1}{2} = \frac{12+1}{2} = \frac{13}{2} = 6.5$<br>$Q_1$ is between the 6th and 7th ordered data values in the lower half. |
| **2** If $n$ is odd, find the data value of the middle dot.<br>If $n$ is even, find the average of the data values for the two middle dots. Count each column of dots from the bottom up to reach the lower quartile. |  |
| **3** Write the answer. | $Q_1 = 13.5$ years |


9780170448406

1.5

v 1 To find the upper quartile $Q_3$, find the median of the upper half.

Count the number of dots n, note whether $n$ is odd or even, and find the position of $Q_3$.

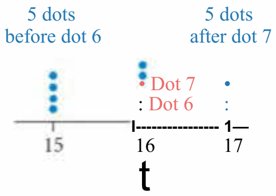
$n = 12$; even

$$\frac{n+1}{2} = \frac{12+1}{2} = \frac{13}{2} = 6.5$$

$Q_3$ is between the 6th and 7th ordered data values in the upper half.

2 If $n$ is odd, find the data value of the middle dot.

If $n$ is even, find the average of the data values for the two middle dots. Count each column of dots from the bottom up to reach the upper quartile.




$Q_3 = \frac{1\wedge}{2} = 16$

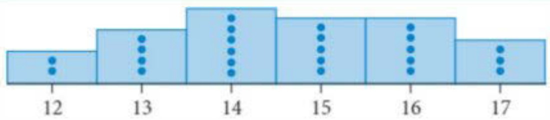
3 Write the answer.

$Q_3 = 16$ years

vi Use $Q_3$ and Qj to calculate the interquartile range.

IQR = $Q_3$-QJ
= 16-13.5
= 2.5 years

b Picture the dot plot as a histogram.




The distribution is approximately symmetric.

## Stem plots

Worksheet Stem-and-leaf plots

Stem plots, also known as stem-and-leaf plots, are an alternative to histograms where actual data values appear. The data values are ordered from smallest to largest where the stem is made up of the leading digits and the leaf is the last digit. Stem plots are best used up to a maximum of 50 data values. When there are a small number of stems, we split the stem to see the distribution more clearly.

See the following stem plots for the ordered data values:

80, 81, 84, 85, 89, 89, 91, 92, 92, 96,105,107,108, 109, 109,112,114,118

Stem plot

| Stem | Leaf | |
|---|---|---|
| 8 | 0 1 4 5 9 9 | |
| 9 | 12 2 6 | |
| 10 | 5 7 8 9 9 | Data values from 100 to 109 |
| 11 | 7 4 | |

Key:9|1=91

Split stem plot

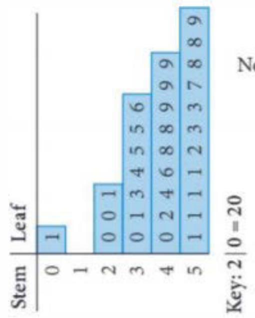
| Stem | Leaf | |
|---|---|---|
| 8 | 0 1 4 | |
| 8 | 5 9 9 | |
| 9 | 1 2 2 | |
| 9 | 6 | |
| 10 | | Datavalues from 100 to 104 |
| 10 | 5 7 8 9 9 | Data values from 105 to109 |
| 11 | 2 4 | |
| 11 | 8 | |

Key: 911 - 91

To decide whether a stem plot is approximately symmetric, positively skewed or negatively skewed, rotate it 90° anticlockwise so that the stem forms the horizontal axis and picture it as a histogram. For example, we can see after rotating the following stem plot that the distribution is negatively skewed:

| Stem | Leaf |
|---|---|
| 0 | 1 |
| 1 | |
| 2 | 0 0 1 |
| 3 | 0 1 3 4 5 5 6 |
| 4 | 024688999 |
| 5 | 1 1 1 1 2 3 3 7 8 8 9 |

Key: 2|0 = 20




Negatively skewed

### Stem plots

Stem plots
- can be used with both continuous and discrete data
- are best used with up to a maximum of 50 data values
- always require a key.

WB p. 13

## WORKED EXAMPLE 11 Using stem plots

The stem plot shows the test marks out of 60 for a class of 33 students.

a Find the
- i mode
- ii range
- iii median
- iv lower quartile (Qj)
- v upper quartile ($Q_3$)
- vi interquartile range (IQR).

b Is there an outlier? Justify your answer.

| stem | Leaf |
|---|---|
| 0 | 1 |
| 1 | |
| 2 | 00111233788 |
| 3 | 0 1 3 4 5 5 6 |
| 4 | 023334578 |
| 5 | 115 6 7 |

Key: 3 | 1 = 31 marks

| Steps | Working |
|---|---|
| a i Find the most common value. | Marks of 21 and 43 each appear three times so this data set is bi-modal. Modes are 21 marks and 43 marks. |
| ii range = largest value - smallest value | range = 57 - 1 = 56 marks |
| iii 1 Count the number of data values H, note whether $n$ is odd or even, and find the position of the median. | $n$ = 33; odd<br>$\frac{n+1}{2} = \frac{33+1}{2} = \frac{34}{2}$<br>The median is the 17th ordered data value. |
| 2 If $n$ is odd, find the middle data value.<br>If $n$ is even, find the average of the two middle data values. | (stem plot below) |
| 3 Write the answer. | median = 35 marks |
| iv 1 To find the lower quartile Qp find the median of the lower half.<br>Count the number of data values n, note whether $n$ is odd or even, and find the position of Qj. | $n$ = 16; even<br>$\frac{n+1}{2} = \frac{16+1}{2} = \frac{17}{2} = 8.5$<br>$Q_1$ is between the 8th and 9th ordered data values in the lower half. |
| 2 If $n$ is odd, find the middle data value.<br>If $n$ is even, find the average of the two middle data values. | (stem plot below) |
| 3 Write the answer. | Qi = 23 marks |

Stem plot for iii 2:

| Stem | Leaf |
|---|---|
| 0 | 1 |
| 1 | 16 values before the median |
| 2 | 001 1 1233788 |
| 3 | 0 1 3 4 5 5 6 — Median = 35 |
| 4 | 023334578 |
| 5 | 1 1 5 6 7 — 16 values after the median |

Stem plot for iv 2:

| Stem | Leaf |
|---|---|
| 0 | 1 |
| 1 | 7 values before the 8th value |
| 2 | 00111233788 — $Q_1 = \frac{23+23}{2} = 23$ |
| 3 | 0 1 3 4 — 7 values after the 9th value |


9780170448406

v 1 To find the upper quartile $Q_3$, find the median of the upper half.

Count the number of data values n, note whether $n$ is odd or even, and find the position of $Q_3$.

$n = 16$; even

$\frac{n+1}{2} - \frac{16+1}{2} - \frac{17}{2} = 8.5$

$Q_3$ is between the 8th and 9th ordered data values in the upper half.

2 If $n$ is odd, find the middle data value.

If $n$ is even, find the average of the two middle data values.

| Stem | Leaf |
|---|---|
| | 7 values before the 8th value |
| 3 | 5 6 |
| 4 | 0 2 3 3 3 4 5 7 8 $Q_3 = \frac{44 + 45}{2} = 44.5$ |
| 5 | 115 6 7 |
| | 7 values after the 9th value |

3 Write the answer.

$Q_3$- 44.5 marks

vi Use $Q_3$ and $Q_1$ to calculate the interquartile range.

IQR = Q3 - Qi
= 44.5 - 23
= 21.5 marks

b Check any value that appears to be an outlier against the upper or lower fence.

The 1 mark result may be an outlier. Check using the lower fence.

Qj - 1.5 x IQR = 23-1.5x21.5
= -9.25

Since 1 isn't less than -9.25, it is *not* an outlier.

## Exam hack

Always look to see if the total number of data values is given in the question. That way you don't waste time counting them.

## Displays and data types

| Categorical: Nominal data | Categorical: Ordinal data | Numerical data |
|---|---|---|
| Displays | | |
| dot plot<br>bar chart | dot plot<br>bar chart | dot plot<br>histogram<br>boxplot<br>stem plot |

1.5

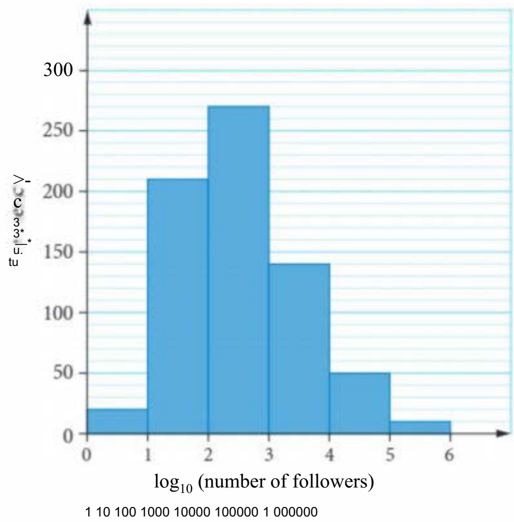
## EXERCISE 1.5 Dot plots and stem plots

ANSWERS p. 694

### Recap

*Use the following information to answer the next two questions.*

A blogger has recorded the number of Twitter followers for each of her subscribers and has displayed the data in the histogram.




1 The number of the bloggers' subscribers who have between 1000 and 10000 followers is closest to

A 140 B 210 C 270 D 480 E 1000

2 Which one of these statements is *not* true?

A The number of the bloggers' subscribers who have more than 10000 but fewer than 100000 followers is 50.

B The number of the bloggers' subscribers who have fewer than 100 followers is 210.

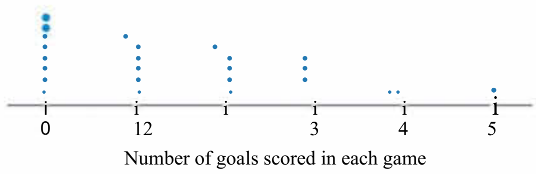
C The number of the bloggers' subscribers who have more than 10000 followers is 60.

D None of the blogger's subscribers has more than 1000 000 followers.

E The number of the bloggers' subscribers with fewer than 10 followers is twice the number of those with more than 100000 followers.

### Mastery

3 E3 WORKED EXAMPLE io The dot plot shows the number of goals scored by a soccer team in each game during a season.




a Find the

i mode ii range iii median

iv lower quartile (Qj) v upper quartile *(Qf)* vi interquartile range (IQR).

b What could best describe the shape of the distribution: approximately symmetric, positively skewed or negatively skewed?

4& WORKED EXAMPLE 11 I The stem plot shows the percentage of students with a Further Mathematics study score of 40 or more in the top-performing 35 schools in 2019.

| Stem | Leaf |
|---|---|
| 1 | 666667889 |
| 2 | 0011112235566889 |
| 3 | 2 3 3 4 5 8 |
| 4 | 3 4 |
| 5 | 1 |
| 6 | 8 |

Key:2|6 = 26%

Source: Better Education https://bettereducation.com.au/Resu lts/vceTopl n Su b j ects. as px

a Find the

i mode ii range iii median

iv lower quartile (Qj) v upper quartile ($Q_3$) vi interquartile range (IQR)

b Is there a possible outlier? Justify your answer.

5 This unordered stem-and-leaf plot represents the number of points scored per match by the Blues in a disappointing football season,

| Stem | Leaf |
|---|---|
| 4 | 5 3 9 |
| 5 | 7 2 0 8 |
| 6 | 4 7 8 5 1 2 |
| 7 | 2 9 3 0 |
| 8 | 9 4 2 |

Key:6|7 = 67

a Display the data using an ordered stem-and-leaf plot.

b How many matches were played in the season?

c What was the Blues' highest score for a match?

d For what percentage of matches did the Blues score below 56 points?

e What could best describe the shape of the distribution: approximately symmetric, positively skewed, or negatively skewed?

## Exam practice

80–100% 60–79% 0–59%

*Use the following information to answer the next two questions.*

The stem plot shows the distribution of mathematics *test scores* for a class of 23 students.

| Stem | Leaf |
|---|---|
| 4 | 0 1 4 4 |
| 5 | 2 7 9 9 9 |
| 6 | 5 6 8 8 9 9 |
| 7 | 0 0 5 6 7 8 |
| 8 | 5 9 |

Key:4|2 = 42 n = 23

6 ©VCAA 2019 1CQ4 J 96% For this class, the range of *test scores* is

A 22 B 40 C 45 D 49 E 89

7 ©VCAA 2019 1CQ5 J 87% For this class, the interquartile range (IQR) of *test scores* is

A 14.5 B 17.5 C 18 D 24 E 49

8 ©VCAA 2016 1CQ3 87% The stem plot displays 30 temperatures recorded at a weather station.

The modal temperature is

A 2.8°C B 2.9°C C 3.7°C

D 8.0°C E 9.0°C

Temperature

| Stem | Leaf |
|---|---|
| 2 | 2 2 4 4 |
| 2 | 578888889999 |
| 3 | 1 2 3 3 4 4 4 |
| 3 | 5 6 7 7 7 7 |
| 4 | 1 |

Key: 2|2 = 2.2°C

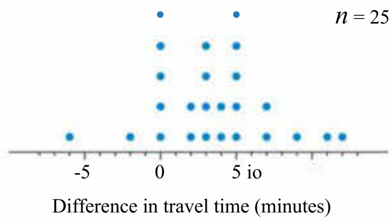
### © Exam hack

Always check the stem plot key in exam questions to see whether your data has decimal places or not.

9780170448406

*Use the following information to answer the next two questions.*

The dot plot displays the *difference in travel time* between the morning peak and the evening peak travel times for the same journey on 25 days.




9 © VCAA j 2O181CQ1~I 85% The percentage of days when there was five minutes *difference in travel time* between the morning peak and the evening peak travel times is

A 0% B 5% C 20% D 25% E 28%

10 ©VCAA | 20181CQ2~1 85% The median *difference in travel time* is

A 3.0 minutes. B 3.5 minutes. C 4.0 minutes.

D 4.5 minutes. E 5.0 minutes.

11 ©VCAA 2015 1CQ1 70% The stem plot displays the average number of decayed teeth in 12-year-old children from 31 countries.

| Stem | Leaf |
|---|---|
| 0 | 2 |
| 0 | 5 6 7 7 8 9 |
| 1 | 0 0 0 0 1 4 4 4 |
| 1 | 5 6 7 |
| 2 | 3 3 4 |
| 2 | 7 7 8 9 |
| 3 | 0 4 |
| 3 | 5 6 |
| 4 | 1 |
| 4 | 7 |

Key: 0|2 = 0.2

Data: Gapminder

Based on this stem plot, the distribution of the average number of decayed teeth for these countries is best described as

A negatively skewed with a median of 15 decayed teeth and a range of 45

B positively skewed with a median of 15 decayed teeth and a range of 45

C approximately symmetric with a median of 1.5 decayed teeth and a range of 4.5

D negatively skewed with a median of 1.5 decayed teeth and a range of 4.5

E positively skewed with a median of 1.5 decayed teeth and a range of 4.5

12 © VCAA 2018N ICQI Y The stem plot displays the distribution of *beak length,* in millimetres, of a sample of 33 female birds of the same species.

| Stem | Leaf |
|---|---|
| 22 | |
| 22 | 8 9 |
| 23 | 0 1 2 3 4 4 |
| 23 | 5 6 6 7 8 9 9 |
| 24 | 0 3 4 4 |
| 24 | 6 7 8 9 |
| 25 | 0 1 2 |
| 25 | 5 7 |
| 26 | 3 4 |
| 26 | 6 |
| 27 | 3 4 |
| 27 | |

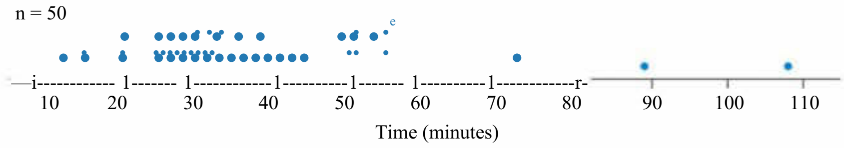
Key: 23|9 = 23.9 $n$ = 33

For these 33 female birds, the median *beak length* is

A 24.0 mm B 24.1mm C 24.3 mm

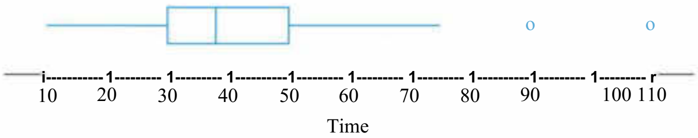
D 24.6 mm E 25.0 mm

▶ 13 ©VCAA 2016SICQ8 The dot plot shows the distribution of the time, in minutes, that 50 people spent waiting to get help from a call centre.

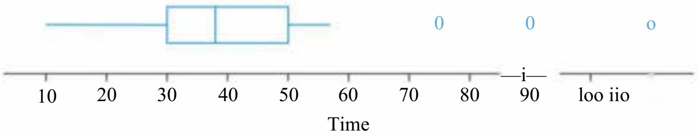



Which one of the following boxplots best represents the data?

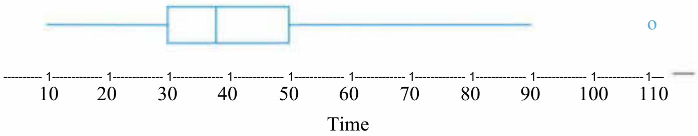
A



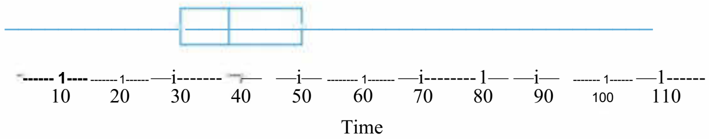

B



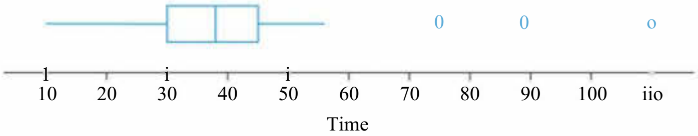

C




D




E




14 ©VCAA 2019 2CQ1 (4 marks) The table shows the *day number* and the *minimum temperature*, in degrees Celsius, for 15 consecutive days in May 2017.

| Day number | Minimum temperature (°C) |
|---|---|
| 1 | 12.7 |
| 2 | 11.8 |
| 3 | 10.7 |
| 4 | 9.0 |
| 5 | 6.0 |
| 6 | 7.0 |
| 7 | 4.1 |
| 8 | 4.8 |
| 9 | 9.2 |
| 10 | 6.7 |
| 11 | 7.5 |
| 12 | 8.0 |
| 13 | 8.6 |
| 14 | 9.8 |
| 15 | 7.7 |

Data: Australian Government, Bureau of Meteorology, www.bom.gov.au

a 83% Which of the two variables in this data set is an ordinal variable? 1 mark ▶

The incomplete ordered stem plot has been constructed using the data values for days 1 to 10.

Minimum temperature (°C)

| Stem | Leaf |
|---|---|
| 4 | 1 8 |
| 5 | |
| 6 | 0 7 |
| 7 | 0 |
| 8 | |
| 9 | 0 2 |
| 10 | 7 |
| 11 | 8 |
| 12 | 7 |

Key: 4| 1 = 4.1 n=15

b 83% Copy and then complete the stem plot by adding the data values for days 11 to 15. 1 mark

c The ordered stem plot shows the *maximum temperature,* in degrees Celsius, for the same 15 days.

Maximum temperature (°C)

| Stem | Leaf |
|---|---|
| 9 | 2 |
| 10 | |
| 11 | 5 6 |
| 12 | 2 5 |
| 13 | 5 5 7 |
| 14 | 9 9 |
| 15 | 0 2 5 6 |
| 16 | 0 |

Key:9|2 = 9.2 n=15

Data: Australian Government, Bureau of Meteorology, www.bom.gov.au

Use this stem plot to determine

i 80% the value of the first quartile (Qj) 1 mark

ii 85% the percentage of days with a maximum temperature higher than 15.3°C. 1 mark

15 ©VCAA | 2020 2CQ1 J (3 marks) *Body mass index* (BMI), in kilograms per square metre, was recorded for a sample of 32 men and displayed in the ordered stem plot.

| Stem | Leaf |
|---|---|
| 21 | 6 9 9 |
| 22 | 12 5 6 |
| 23 | 0 1 4 6 6 7 8 |
| 24 | 4 5 6 7 7 9 |
| 25 | 6 8 |
| 26 | 1 7 9 |
| 27 | 3 7 |
| 28 | 2 |
| 29 | 1 8 |
| 30 | 4 |
| 31 | 1 |

Key: 2116 = 21.6 n = 32

a 86% Describe the shape of the distribution, 1 mark

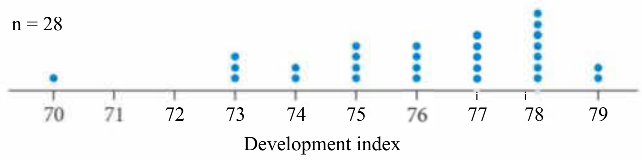
b 70% Determine the median BMI for this group of men. 1 mark

c 84% People with a BMI of 25 or over are considered to be overweight. What percentage of these men would be considered to be overweight? 1 mark



9780170448406

▶ 16 © VCAA J 2016S2CQ2 (3 marks) The development index for a country is a whole number between 0 and 100. The dot plot shown displays the values of the development indices for 28 countries.




a Using the information in the dot plot, determine the mode and the range. 1 mark

b Write down an appropriate calculation and use it to explain why the country with a development index of 70 is an outlier for this group of countries. 2 marks

1.6

# @ The mean and standard deviation

Video playlist The mean and standard deviation

Worksheets Standard deviation

Statistical calculations

Statistics review

Calculating and interpreting summary statistics

Data and statistics crossword

Statistics crossword

## The mean

We will now have a closer look at the mean, often called the average, the most commonly used measure of the centre of a distribution. The symbol for the mean of a data set is $\bar{x}$ (called x bar'). There are shortcuts for calculating the mean depending on how the data is displayed.

### The mean

The mean for a list of data values:

$$\bar{x} = \frac{\text{sum of all values}}{\text{number of values}}$$

The mean for data in a frequency table:

$$\bar{x} = \frac{\text{sum of (each value x its corresponding frequency)}}{\text{sum of frequencies}}$$

## Comparing the mean and median

We often need to choose between the mean and the median for the best measure of the centre.

- For symmetric distributions, the mean = the median.
- For distributions that are approximately symmetric, the mean and the median will be very close in value.
- The mean is greater than the median for positively skewed distributions.
- The mean is less than the median for negatively skewed distributions.
- Outliers usually don't affect the median, but they can often significantly affect the mean.

### Mean versus median

Choosing between the mean and median as the measure of the centre of a distribution:

| Shape of distribution | Choose |
|---|---|
| Approximately symmetric distribution with no outliers | mean or median |
| Approximately symmetric distribution with outliers | median |
| Skewed distribution | median |

## The standard deviation

The standard deviation, like the range and the interquartile range, is a measure of the spread of the distribution. While the interquartile range measures the spread of data around the median, the standard deviation measures the spread of data around the mean. We will always calculate the standard deviation using CAS. The symbol for the standard deviation is $s_x$.

**Standard deviation versus IQR**

Choosing between the standard deviation and IQR as the measure of the spread of a distribution:

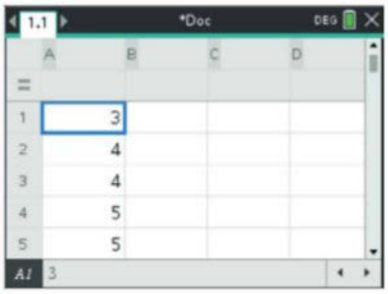
| Shape of distribution | Choose |
|---|---|
| Approximately symmetric distribution with no outliers | standard deviation or IQR |
| Approximately symmetric distribution with outliers | IQR |
| Skewed distribution | IQR |

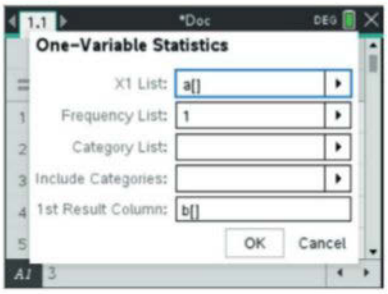
**USING CAS 4** Finding the mean and standard deviation for ungrouped data

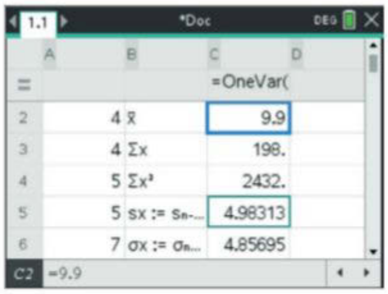
Find the mean $\bar{x}$ and the standard deviation $s$, rounded to two decimal places, for the ungrouped data shown.

3, 4, 4, 5, 5, 7, 7, 7, 8, 9, 9, 9, 12, 12, 13, 15, 15, 16, 18, 20

**TI-Nspire**



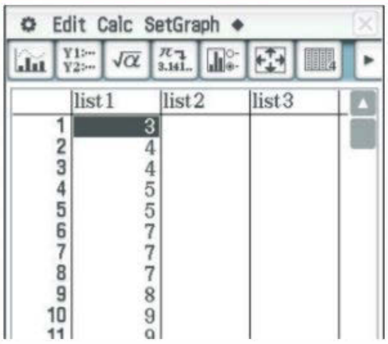

1. Start a new document and add a **Lists & Spreadsheet** page.
2. Enter values into column **A**.
3. Press **menu > Statistics > Stat Calculations > One-Variable Statistics**.
4. On the next screen, keep the number of lists default setting of **1** and select **OK**.
5. In the **X1 List:** field, keep the default setting of **a[ ]***.
6. Select **OK**.

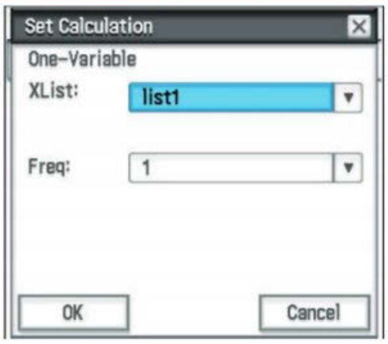
*Alternatively, label the column and use the variable name.

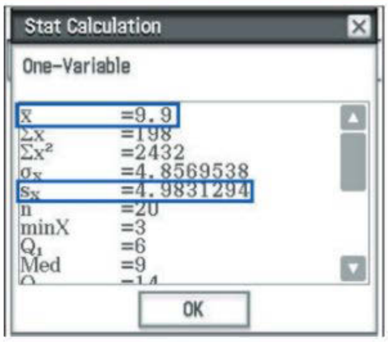
7. The one-variable labels and values will be displayed in columns **B** and **C**.
8. Scroll down to view the mean and standard deviation values.

$\bar{x} = 9.90$, $s = 4.98$

**ClassPad**




1 Tap Menu and open the Statistics application.

2 Clear all lists and enter the data as shown.

3 Tap Calc > One-Variable.

4 Leave the default settings of XList as listl and Freq: as 1.

5 Tap OK.

6 The mean and standard deviation values will be displayed.

$\bar{x} = 9.90$, s = 4.98

## Exam hack

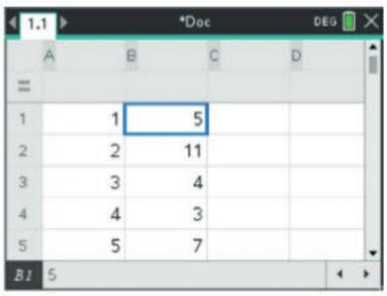
The steps for finding the mean and standard deviation for a data set using CAS are the same as for finding the five-number summary.

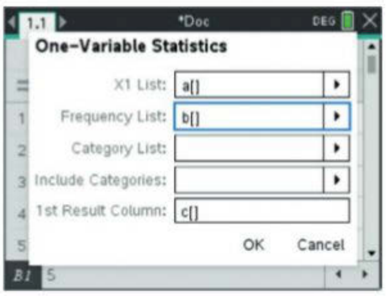
## USING CAS 5 Finding the mean and standard deviation for grouped data

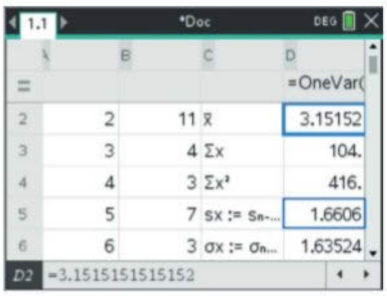
Find the mean $\bar{x}$ and the standard deviation s, rounded to two decimal places, for the grouped data shown.

| Score | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Frequency | 5 | 11 | 4 | 3 | 7 | 3 |

**TI-Nspire**




1 Start a new document and add a Lists & Spreadsheet page.

2 Enter the Score values into column A and the Frequency values into column B.

3 Press menu > Statistics > Stat Calculations > One-Variable Statistics.

4 On the next screen, keep the number of lists default setting of 1 and select OK.

5 In the X1 List: field, enter a[ ]*.

6 In the Frequency List: field, enter b[ ]*.

7 Select OK.

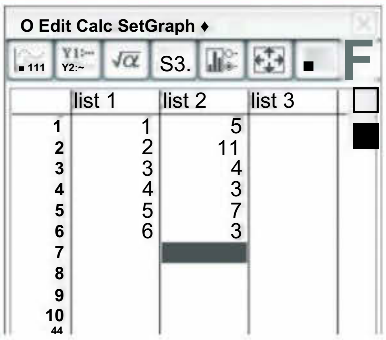
*Alternatively, label the columns and use the variable names.

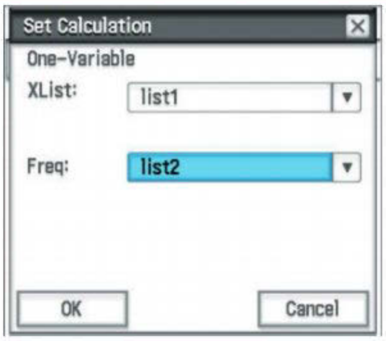
8 The one-variable labels and values will be displayed in columns C and D.

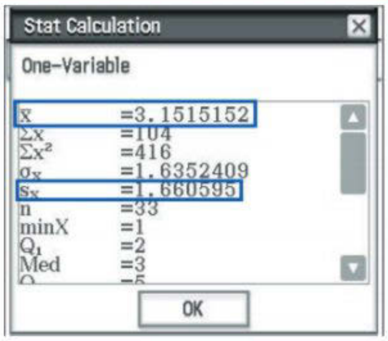
9 Scroll down to view the mean and standard deviation values.

$\bar{x} = 3.15$, $s = 1.66$

**ClassPad**




1 Tap Menu and open the Statistics application.

2 Clear all lists and enter the data as shown.

3 Tap Calc > One-Variable.

4 Leave the XList default setting of list 1. Change the Freq: setting to list2.

5 Tap OK.

6 The mean and standard deviation values will be displayed.

$\bar{x}$ =3.15, $s$ = 1.66

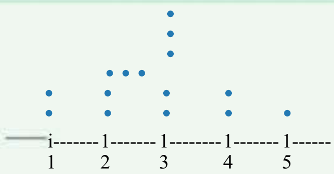
WB p. 15

## WORKED EXAMPLE 12 Working with the mean and standard deviation from a display

For each the following displays

- **i** find the mean
- ii find the median
- iii state whether the mean or median or both are appropriate measures of the centre and why
- iv find the standard deviation.
- v find the IQR
- vi state whether the standard deviation or IQR or both are appropriate measures of the spread and why.

Round your answers to one decimal place where necessary.

| **Steps** | **Working** |
|---|---|
| a Number of items bought in a store by 15 customers |   |
| **i** Use CAS by entering the values 1,1,2,2, 2,3... and selecting the mean. | mean = 2.9 items |
| ii Use CAS and select the median. | median = 3 items |
| iii Is the distribution approximately symmetric or skewed? Does the distribution have outliers? | The mean and median are both appropriate measures of centre because the distribution is approximately symmetric with no outliers. |
| iv Use CAS and select the standard deviation. | standard deviation =1.1 items |
| v Use CAS to find the IQR. | IQR = $Q_3$ - Qj = 4-2 = 2 items |
| vi Is the distribution approximately symmetric or skewed? Does the distribution have outliers? | The standard deviation and IQR are both appropriate measures of spread because the distribution is approximately symmetric with no outliers. |



9780170448406

b Number of people in a store in each hour during a 24-hour sale.

| Stem | Leaf |
|---|---|
| 1 | 2235678899 |
| 2 | 0 0 1 1 2 3 5 |
| 3 | 2 3 3 4 |
| 4 | 0 4 |
| 5 | 1 |

Key: 1 | 3 = 13

| | |
|---|---|
| i Use CAS by entering the values 12,12,13,15 ... and selecting the mean. | mean = 24.1 people |
| ii Use CAS and select the median. | median = 20.5 people |
| iii Is the distribution approximately symmetric or skewed? Does the distribution have outliers? | The median is an appropriate measure of centre because the distribution is skewed. |
| iv Use CAS and select the standard deviation. | standard deviation = 10.4 people |
| v Use CAS to find the IQR. | IQR = $Q_3 - Q_1$ = 32.5 - 17.5 = 15 people |
| vi Is the distribution approximately symmetric or skewed? Does the distribution have outliers? | The IQR is an appropriate measure of spread because the distribution is skewed. |

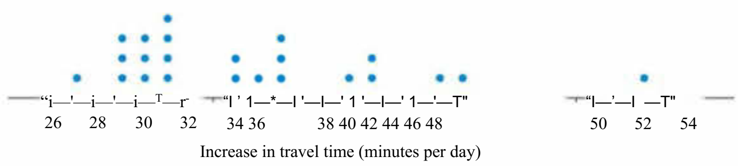
1.6

## EXERCISE 1.6 The mean and standard deviation

ANSWERS p. 695

### Recap

1 ©VCAA | 20182CQigM0DiFiEDJ Traffic congestion can lead to an increase in travel times in cities. The following dot plot shows the increase in travel time due to traffic congestion, in minutes per day, for 23 UK cities.




The upper quartile ($Q_3$) is

A 11 B 30 C 34 D 39 E 40

2 ©VCAA 2018N1CQ2 The stem plot displays the distribution of *beak length,* in millimetres, of a sample of 33 female birds of the same species.

The percentage of these 33 female birds with a beak length of less than 25.4 mm is closest to

A 21.2% B 25.4% C 27.0%

D 75.8% E 78.8%

| Stem | Leaf |
|---|---|
| 22 | |
| 22 | 8 9 |
| 23 | 0 1 2 3 4 4 |
| 23 | 5 6 6 7 8 9 9 |
| 24 | 0 3 4 4 |
| 24 | 6 7 8 9 |
| 25 | 0 1 2 |
| 25 | 5 7 |
| 26 | 3 4 |
| 26 | 6 |
| 27 | 3 4 |
| 27 | |

Key: 23|9 = 23.9 n = 33 ►

## ▶ Mastery

30 using CAS 4] Find the mean $\bar{x}$ and the standard deviation s, rounded to two decimal places, for the ungrouped data shown:

26, 27, 23, 24, 27,25, 23, 19, 25, 21, 22, 20, 27, 21

4 Using CAS 5 Find the mean $\bar{x}$ and the standard deviation $s$, rounded to two decimal places, for each of the following grouped data.

a

| Score | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Frequency | 3 | 7 | 4 | 11 | 13 | 10 | 6 | 5 | 3 | 2 | 2 |

b

| Score | Frequency |
|---|---|
| 3 | 2 |
| 4 | 3 |
| 5 | 5 |
| 6 | 6 |
| 7 | 7 |
| 8 | 6 |
| 9 | 3 |
| 10 | 3 |
| 11 | 3 |
| 12 | 2 |

5 WORKED EXAMPLE 12 Using CAS 4 For each the following displays

i find the mean

ii find the median

iii state whether the mean or median or both are appropriate measures of the centre and why

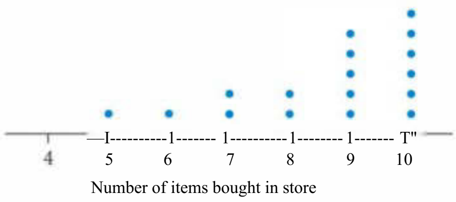
iv find the standard deviation.

v find the IQR

vi state whether the standard deviation or IQR or both are appropriate measures of the spread and why.

Round your answers to one decimal place where necessary.

a Number of items bought in a store by 17 customers




b Number of people in a store in each hour during a 21-hour sale.

| Stem | Leaf |
|---|---|
| 2 | 3 6 8 |
| 3 | 4 4 5 |
| 4 | 02235678 |
| 5 | 12 3 5 |
| 6 | 8 9 9 |

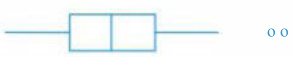
Key:3|4 = 34

6 For each of these displays, state whether

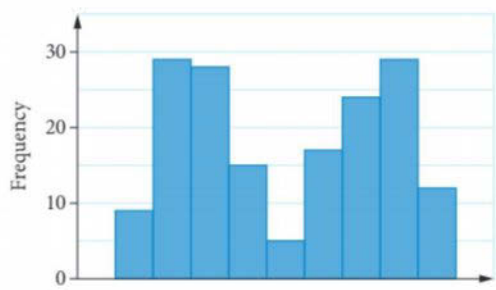
i the mean or median or both are appropriate measures of the centre

ii the standard deviation or IQR or both are appropriate measures of the spread.

a



b



1.6

## Exam practice

80-100% 60-79% 0-59%

7 ©VCAA 2017 1CQ3 J 88% The table shows the forearm *circumference,* in centimetres, of a sample of 10 people selected from a group of 252 people.

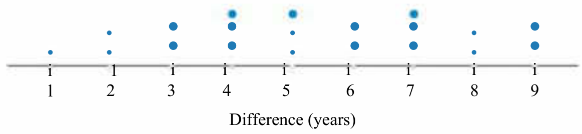
| Circumference | 26.0 | 27.8 | 28.4 | 25.9 | 28.3 | 31.5 | 28.2 | 25.9 | 27.9 | 27.8 |
|---|---|---|---|---|---|---|---|---|---|---|

The mean, $\bar{x}$, and the standard deviation, $s_x$, of the forearm *circumference* for this sample of people are closest to

A $\bar{x} = 1.58$ *x = 27.8 B $\bar{x} = 1.66$ ^ = 27.8 C x = 27.8 $s_x$=1.58

D $\bar{x} = 27.8$ $s_x$=1.66 E $\bar{x} = 27.8$ $s_x = 2.30$

8 ©VCAA 2015 1CQ3 J 66% The dot plot displays the difference between female and male life expectancy, in years, for a sample of 20 countries.




The mean ($\bar{x}$) and standard deviation (s) for this data are

A mean = 2.32 standard deviation = 5.25

B mean = 2.38 standard deviation = 5.25

C mean = 5.0 standard deviation = 2.0

D mean = 5.25 standard deviation = 2.32

E mean = 5.25 standard deviation = 2.38

*Use the following information to answer the next two questions.*

The number of DVD players in each of 20 households is recorded in the frequency table below.

| Number of DVD players | 0 | 1 | 2 | 3 | 4 | 5 | Total |
|---|---|---|---|---|---|---|---|
| Frequency | 6 | 9 | 3 | 1 | 0 | 1 | 20 |

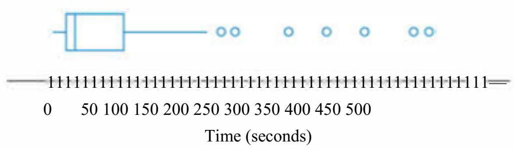
9 ©VCAA 2004 1CQ4 J 62% For this sample of households, the percentage of households with at least one DVD player is

A 30% B 45% C 50% D 70% E 90%

10 ©VCAA 2004 1CQ5 54% For this sample of households, the mean number of DVD players in these 20 households is

A 0.75 B 1.00 C 1.15 D 1.64 E 2.00

11 ©VCAA 2008 1CQ4 53% The boxplot shows the distribution of the time, in seconds, that 79 customers spent moving along a particular aisle in a large supermarket.




Data source: www.stars.ac.uk

From the boxplot, it can be concluded that the median time spent moving along the supermarket aisle is

A less than the mean time.

B equal to the mean time.

C greater than the mean time.

D half of the interquartile range.

E one quarter of the range.

### Exam hack

In median versus mean questions, work out whether the distribution is symmetric, positively skewed or negatively skewed.

12 ©VCAA 2010 2CQ1J (5 marks) This table shows the percentage of women ministers in the parliaments of 22 countries in 2008.

| Country | Percentage of women ministers | Country | Percentage of women ministers |
|---|---|---|---|
| Norway | 56 | Australia | 24 |
| Sweden | 48 | Italy | 24 |
| France | 47 | United States | 24 |
| Spain | 44 | Belgium | 23 |
| Switzerland | 43 | United Kingdom | 23 |
| Austria | 38 | Ireland | 21 |
| Denmark | 37 | Liechtenstein | 20 |
| Iceland | 36 | Canada | 16 |
| Germany | 33 | Luxembourg | 14 |
| Netherlands | 33 | Japan | 12 |
| New Zealand | 32 | Singapore | 0 |

a 79% What proportion of these 22 countries had a higher percentage of women ministers in their parliament than Australia? 1 mark

b 79% Determine the median, range and interquartile range of this data. 2 marks

The ordered stem plot displays the distribution of the percentage of women ministers in parliament for 21 of these countries. The value for Canada is missing.

| Stem (10s) | Leaf (units) |
|---|---|
| 0 | 0 |
| 1 | 2 4 |
| 2 | 0 1 3 3 4 4 4 |
| 3 | 2  3 3 6 7 8 |
| 4 | 3  4 7 8 |
| 5 | 6 |

c 69% Copy and complete the stem plot by adding the value for Canada, 1 mark

d 69% Both the median and the mean are appropriate measures of centre for this distribution. Explain why. 1 mark

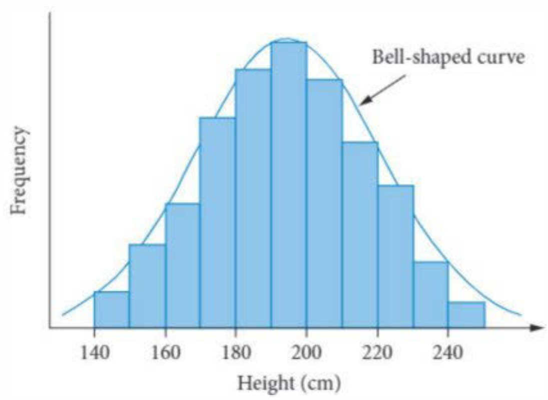
# @ Bell-shaped distributions

Video playlist Bell-shaped distributions

Worksheet The normal curve

## The normal or bell-shaped distribution

Many variables in real life have what is known as a normal distribution or bell-shaped distribution. This means they have an approximate bell shape as in this example.

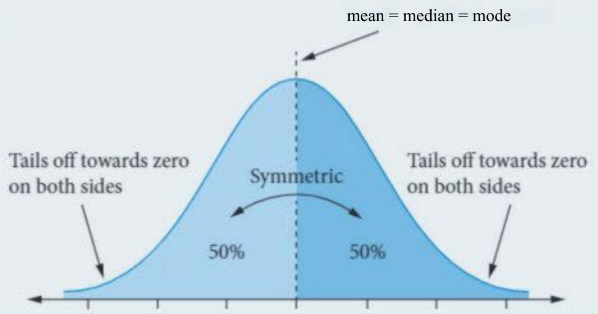



 9780170448406

## Bell-shaped distributions

Bell-shaped distributions

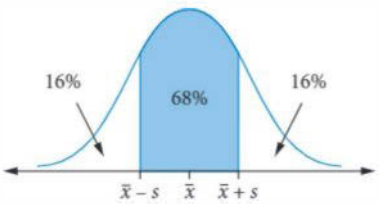
- are symmetric about the mean
- have 50% of the data either side of the mean
- have mode = mean = median
- have a peak in the centre and tail off towards zero on both sides.



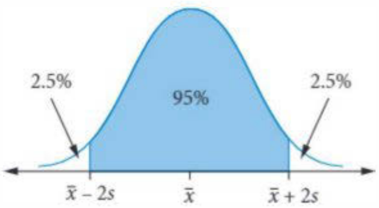

## The 68-95-99.7% rule

For a normal distribution, we can use the 68-95-99.7% rule:

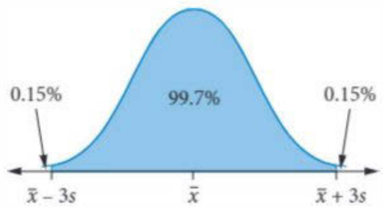
Around 68% of the data values lie within *one* standard deviation of the mean.




Around 95% of the data values lie within *two* standard deviations of the mean.

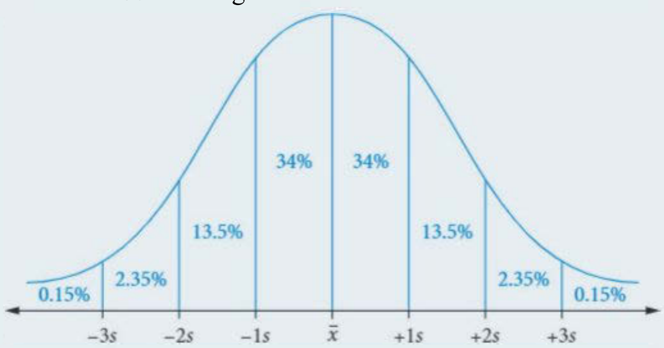



Around 99.7% of the data values lie within *three* standard deviations of the mean.

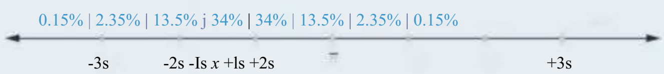



## Representing the 68-95-99.7% rule

68-95-99.7% rule diagram




68-95-99.7% rule scale

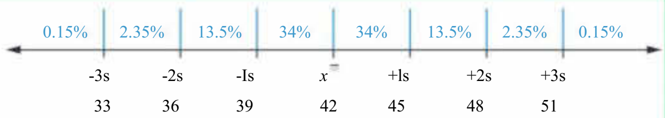


WB

p. 16

## WORKED EXAMPLE 13 Working with the 68-95-99.7% rule

After a lengthy study, it was found that the number of chockbits in packets approximated a bell-shaped distribution with a mean of 42 and a standard deviation of 3.

| Steps | Working |
|---|---|
| **a** Write the 68-95-99.7% rule scale for this. | |
| Write a 68-95-99.7% rule scale that includes the mean and standard deviations given. |  |
| **b** Find the percentage of packets that have between 36 and 48 chockbits. | |
| 1 Add the required percentages from the 68-95-99.7% rule scale. | Adding the percentages between 36 and 48:<br>13.5% + 34% + 34% + 13 5% — 95% |
| 2 Write the answer. | 95% of packets have between 36 and 48 chockbits, |
| **c** Find the percentage of packets with more than 45 chockbits. | |
| 1 Add the required percentages from the 68-95-99.7% rule scale. | Adding the percentages greater than 45:<br>13.5% + 2.35% + 0.15% = 16% |
| 2 Write the answer. | 16% of packets have more than 45 chockbits. |
| **d** Find the percentage of packets with more than 39 chockbits. | |
| 1 Add the required percentages from the 68-95-99.7% rule scale. | Adding the percentages greater than 39:<br>34% + 34% + 13.5% + 2.35% + 0.15% = 84% |
| 2 Write the answer. | 84% of packets have more than 39 chockbits. |
| **e** A supermarket has bought 4000 chockbits packets.<br>i How many of these would they expect to have fewer than 33 chockbits? | |
| 1 Add the required percentages from the 68-95-99.7% rule scale. | Adding the percentages less than 33:<br>0.15% |
| 2 Find this percentage of the total given and write the answer. | 0.15% of4000 = 6<br>The supermarket would expect 6 packets to have fewer than 33 chockbits. |
| ii How many of these would they expect to have between 33 and 51 chockbits? | |
| 1 Add the required percentages from the 68-95-99.7% rule scale. | Adding the percentages between 33 and 51:<br>2.35% + 13.5% + 34% + 34% + 13.5% + 2.35% = 99.7% |
| 2 Find this percentage of the total given and write the answer. | 99.7% of 4000 = 3988<br>The supermarket would expect 3988 packets to have between 33 and 51 chockbits. |



9780170448406

## EXERCISE 1.7 Bell-shaped distributions

ANSWERS p. 695

### Recap

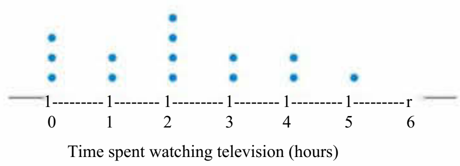
1 ©VCAA 20111CQ6 MODIFIED ) 67% The table displays the systolic blood pressure readings, in mmHg, that result from fifteen successive measurements of the same persons blood pressure.

| Reading number | Systolic blood pressure | Reading number | Systolic blood pressure |
|---|---|---|---|
| 1 | 121 | 9 | 125 |
| 2 | 126 | 10 | 121 |
| 3 | 141 | 11 | 118 |
| 4 | 125 | 12 | 134 |
| 5 | 122 | 13 | 125 |
| 6 | 126 | 14 | 127 |
| 7 | 129 | 15 | 119 |
| 8 | 130 | | |

Correct to one decimal place, the mean and standard deviation of this persons systolic blood pressure measurements are respectively

A 124.9 and 4.4 B 125.0 and 5.8 C 125.0 and 6.0 D 125.9 and 5.8 E 125.9 and 6.0

2 ©VCAA | 20081CQ5 60% A sample of 14 people were asked to indicate the time (in hours) they had spent watching television on the previous night. The results are displayed in the dot plot.



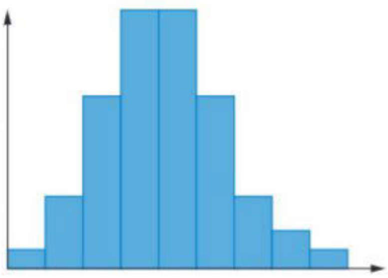

Correct to one decimal place, the mean and standard deviation of these times are respectively

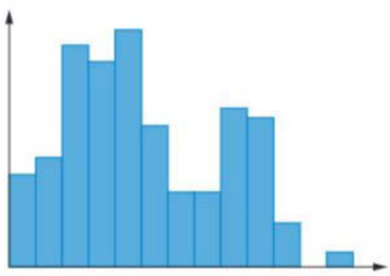
A $\bar{x}$ = 2.0 5 = 1.5 B $\bar{x}$=2.1 5=1.5 C $\bar{x}$=2.1 5=1.6

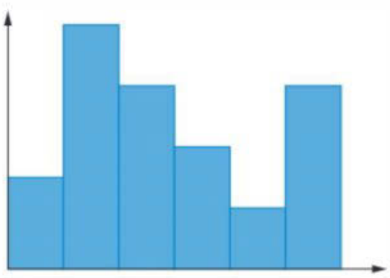
D $\bar{x}$ = 2.6 5=1.2 E $\bar{x}$ = 2.6 5= 1.3

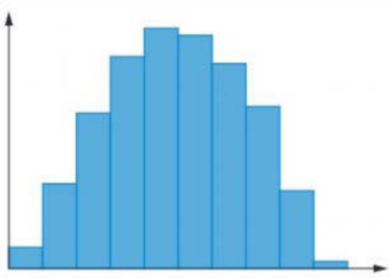
### Mastery

3 State whether each of these histograms represent an approximate normal distribution.

a


b


c


d


4E WORKED EXAMPLE 13 A restaurant chain records the total time customers spend in their restaurants. It was found that the times approximated a bell-shaped distribution with the mean amount of time spent in the restaurant being 73 minutes and the standard deviation 8 minutes.

a Write the 68-95-99.7% rule scale for this.

b Find the percentage of times spent in the restaurant that are less than 73 minutes.

c Find the percentage of times spent in the restaurant that are between 49 and 97 minutes,

d Find the percentage of times spent in the restaurant that are less than 57 minutes,

e Find the percentage of times spent in the restaurant that are more than 49 minutes,

f One of the restaurants in the chain had 426 customers in one day. How many of these customers would they expect to spend

i more than 81 minutes in the restaurant?

ii between 65 and 81 minutes in the restaurant?

5 In a particular school, the number of days that students are late to class in a year has an approximate normal distribution with a mean of 21 and a standard deviation of 3.

a Write the 68-95-99.7% rule scale for this.

Use the scale to state whether the following are reasonable estimates.

b 50% of students are late to class on more than 21 days in a year.

c 25% of students are late to class on fewer than 21 days in a year.

d 68% of students are late to class between 18 and 24 days in a year.

e 99.7% of students are late to class between 12 and 30 days in a year.

f 16% of students are late to class on fewer than 18 days in a year.

g 2.5% of students are late to class on more than 27 days in a year.

h 0.15% of students are late to class on more than 12 days in a year.

i 84% of students are late to class on more than 18 days in a year.

j 97.5% of students are late to class on fewer than 27 days in a year.

k 99.85% of students are late to class on fewer than 30 days in a year.

## Exam practice

80-100% 60-79% 0-59%

6 ©VCAA 20191CQ6J 79% The time taken to travel between two regional cities is approximately normally distributed with a mean of 70 minutes and a standard deviation of 2 minutes. The percentage of travel times that are between 66 minutes and 72 minutes is closest to

A 2.5% B 34% C 68% D 81.5% E 95%

7 ©VCAA 2020 ICQS 78% The *wing length* of a species of bird is approximately normally distributed with a mean of 61 mm and a standard deviation of 2 mm. Using the 68-95-99.7% rule, for a random sample of 10000 of these birds, the number of these birds with a *wing length* of less than 57 mm is closest to

A 50 B 160 C 230 D 250 E 500

1.7

8 ©VCAA J 20201CQ9 77% The lifetime of a certain brand of light globe, in hours, is approximately normally distributed. It is known that 16% of the light globes have a lifetime of less than 655 hours and 50% of the light globes have a lifetime that is greater than 670 hours. The mean and the standard deviation of this normal distribution are closest to

A mean = 655 hours standard deviation =10 hours

B mean = 655 hours standard deviation =15 hours

C mean = 670 hours standard deviation =10 hours

D mean = 670 hours standard deviation =15 hours

E mean = 670 hours standard deviation = 20 hours

*Use the following information to answer the next two questions.*

The time, in hours, that each student spent sleeping on a school night was recorded for 1550 secondary school students. The distribution of these times was found to be approximately normal with a mean of 7.4 hours and a standard deviation of 0.7 hours.

9 ©VCAA 2013 1CQ5 77% The time that 95% of these students spent sleeping on a school night could be

A less than 6.0 hours.

B between 6.0 and 8.8 hours.

C between 6.7 and 8.8 hours.

D less than 6.0 hours or more than 8.8 hours.

E less than 6.7 hours or more than 9.5 hours.

10 ©VCAA 1 20131CQ6 76% The number of these students who spent more than 8.1 hours sleeping on a school night was closest to

A 16 B 248 C 1302 D 1510 E 1545

11 ©VCAA 20191CQ7 72% The volume of a cup of soup served by a machine is normally distributed with a mean of 240 mL and a standard deviation of 5 mL. A fast-food store used this machine to serve 160 cups of soup. The number of these cups of soup that are expected to contain less than 230 mL of soup is closest to

A 3 B 4 C 8 D 26 E 156

12 ©VCAA | 20091CQ7 69% The level of oil use in certain countries is approximately normally distributed with a mean of 42.2 units and a standard deviation of 10.2 units. The percentage of these countries in which the level of oil use is greater than 32 units is closest to

A 5% B 16% C 34% D 84% E 97.5%

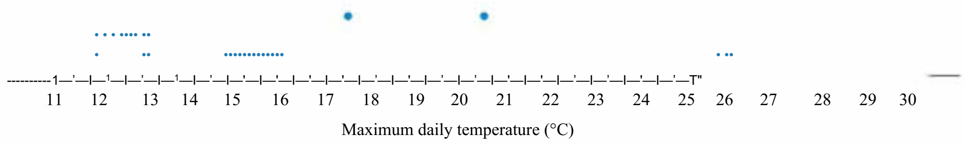
13 ©VCAA | 20071CQ4 61% The length of 3-month-old baby boys is approximately normally distributed with a mean of 61.1 cm and a standard deviation of 1.6 cm. The percentage of 3-month-old baby boys with length greater than 59.5 cm is closest to

A 5% B 16% C 68% D 84% E 95%

14 ©VCAA 2016S1CQ2 The time spent by shoppers at a hardware store on a Saturday is approximately normally distributed with a mean of 31 minutes and a standard deviation of 6 minutes. If 2850 shoppers are expected to visit the store on a Saturday, the number of shoppers who are expected to spend between 25 and 37 minutes in the store is closest to

A 16 B 68 C 460 D 1900 E 2400

**15** ©VCAA 2012 2CQ1 J 70% (3 marks) The dot plot below displays the maximum daily temperature (in °C) recorded at a weather station on each of the 30 days in November 2011.




a From this dot plot, determine

i the median maximum daily temperature, correct to the nearest degree — 1 mark

ii the percentage of days on which the maximum temperature was less than 16°C. Write your answer, correct to one decimal place. — 1 mark

Records show that the minimum daily temperature for November at this weather station is approximately normally distributed with a mean of 9.5°C and a standard deviation of 2.25°C.

b Determine the percentage of days in November that are expected to have a minimum daily temperature less than 14°C at this weather station. Write your answer, correct to one decimal place. — 1 mark

Video playlist Standardised values

Worksheet z-scores

# @ Standardised values

## ^-scores

Standardised values, also known as z-scores, allow us to compare values from different normal distributions. The standardised values

- always have mean = 0
- always have standard deviation = 1
- tell us the number of standard deviations each actual value lies from the mean.

### Standardised values

To calculate standardised values from actual values, use the formula:

$$\text{standardised value} = \frac{\text{actual value} - \text{mean}}{\text{standard deviation}} \quad \text{or} \quad z = \frac{x - \bar{x}}{s_x}$$

To calculate actual values from standardised values, use the formula:

$$\text{actual value} = (\text{standardised value} \times \text{standard deviation}) + \text{mean} \quad \text{or} \quad x = z \times s_x + \bar{x}$$

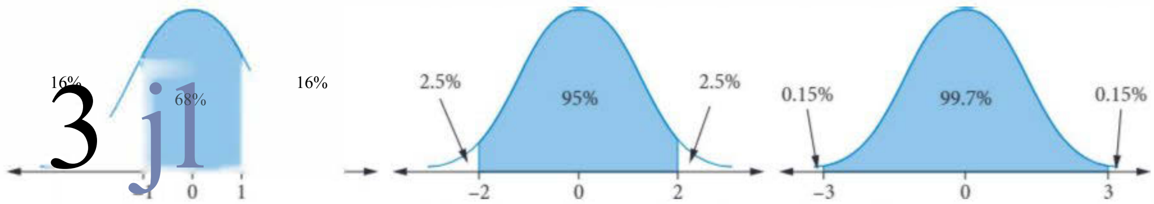
We can say the following about the actual value if we know its z-score:

| z-score | What does it mean? |
|---|---|
| z-score is positive | The actual value is above the mean. |
| z-score is negative | The actual value is below the mean. |
| z-score = 0 | The actual value equals the mean. |
| z-score = 1 | The actual value is 1 standard deviation above the mean. |
| z-score —2 | The actual value is 2 standard deviations below the mean. |

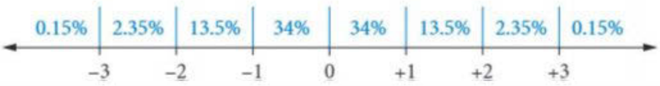

9780170448406

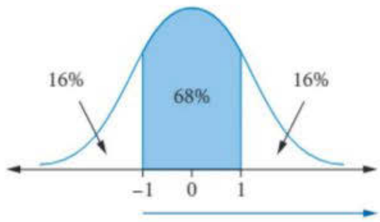
## Exam hack

When dealing with z-scores rather than the original data, the three standardised 68-95-99.7% rule diagrams simplify to:




and the 68–95–99.7% rule scale simplifies to:




| **WORKED EXAMPLE 14** | Working with $z$-scores |
|---|---|
| The lengths of the handspans of adults are known to be normally distributed with a mean of 21.5 cm and a standard deviation of 1.5 cm. A person has a handspan of 20 cm. | |
| **Steps** | **Working** |
| **a** Calculate the standardised value for this actual value. | |
| **1** Write the $z$-score formula. | $z = \dfrac{x - \bar{x}}{s_x}$ |
| **2** Substitute in the actual value, mean and standard deviation. | $z = \dfrac{20 - 21.5}{1.5} = \dfrac{-1.5}{1.5} = -1$ |
| **b** Show that this standardised value means that 84% of people have handspans wider than this person. | |
| **1** State how the $z$-score relates to the number of standard deviations. | $z = -1$ means the person's handspan is one standard deviation below the mean. |
| **2** Sketch the relevant standardised 68–95–99.7% rule diagram and show the required region or use the 68-95-99.7% rule scale. |   |
| **3** Read the percentage from the diagram. | From the diagram, 68% + 16% = 84%. So 84% of people have a handspan greater than this person. |
| **c** Another person has a standardised handspan of $z = 2.5$. What is this person's actual handspan? | |
| **1** Write the formula for finding the actual value from the standardised value, and the values for $z$, $s_x$ and $\bar{x}$. | $x = z \times s_x + \bar{x}$<br>$z = 2.5,\ s_x = 1.5,\ \bar{x} = 21.5$ |
| **2** Substitute the values into the formula and solve for $x$. | $x = z \times s_x + \bar{x}$<br>$x = 2.5 \times 1.5 + 21.5$<br>$x = 25.25$ |
| **3** Write the answer. | The person's actual handspan is 25.25 cm. |

1.8

WB
p. 17

## Using ^-scores to compare

Suppose you achieved a mark in the General Mathematics exam of 91% and a mark in the Specialist Mathematics exam of 49%. You couldn't necessarily say you performed better in General Mathematics because Specialist Mathematics is a more difficult subject. The way to compare the two results, assuming both distributions are bell-shaped, is to standardise them.

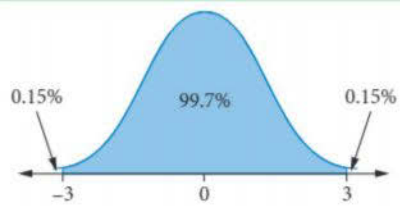
WB
p. 18

**WORKED EXAMPLE 15** Using $z$-scores to compare

The table below shows the marks out of 100 that a student has achieved on her exams in three mathematics subjects, plus the means and standard deviations for each of the subjects.

| | Mark | Mean | Standard deviation |
|---|---|---|---|
| General Mathematics | 91 | 73 | 6 |
| Mathematical Methods | 67 | 51 | 4 |
| Specialist Mathematics | 49 | 31 | 5 |

Assume the results for each of the three subjects approximates a bell-shaped distribution.

| Steps | Working |
|---|---|
| **a** In which mathematics subject did the student perform the best? | |
| **1** Write the $z$-score formula. | $z = \frac{x - \bar{x}}{s_x}$ |
| **2** Substitute the values for each subject. | General Mathematics<br>$z = \frac{91 - 73}{6} = \frac{18}{6} = 3$<br>Mathematical Methods<br>$z = \frac{67 - 51}{4} = \frac{16}{4} = 4$<br>Specialist Mathematics<br>$z = \frac{49 - 31}{5} = \frac{18}{5} = 3.6$ |
| **3** State which $z$-score is the largest. | The student performed best in Mathematical Methods. |
| **b** In which of these subjects was she in the top 0.15% of students? Draw a diagram that shows how you obtained your answer. | |
| **1** Sketch the relevant standardised 68–95–99.7% rule diagram or use the 68–95–99.7% rule scale. |  |
| **2** Read the answer from the diagram. | From the diagram, the top 0.15% of students had a $z$-score of 3 or more. The student had $z$-scores of 3, 4 and 3.6, so she was in the top 0.15% of students in all three mathematics subject. |

## VCE QUESTION ANALYSIS

©VCAA 2017 2CQ1 2017 Examination 2 Core Question 1 (5 marks)

1 The number of eggs counted in a sample of 12 clusters of moth eggs is recorded in the table.

| Number of eggs | 172 | 192 | 159 | 125 | 197 | 135 | 140 | 140 | 138 | 166 | 136 | 131 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|

a From the information given, determine

i the range — 1 mark

ii the percentage of clusters in this sample that contain more than 170 eggs. — 1 mark

In a large population of moths, the number of eggs per cluster is approximately normally distributed with a mean of 165 eggs and a standard deviation of 25 eggs.

b Using the 68-95-99.7% rule, determine

i the percentage of clusters expected to contain more than 140 eggs — 1 mark

ii the number of clusters expected to have less than 215 eggs in a sample of 1000 clusters. — 1 mark

c The standardised number of eggs in one cluster is given by z —2.4.
Determine the actual number of eggs in this cluster. — 1 mark

1.8

Video
VCE question analysis: Data distributions

### Reading the question

- The standardised score formula is on the formula sheet, but you will need to know the range formula.
- Make sure you are clear on what the 68-95-99.7% rule is.
- Identify in each question part whether you are being asked for a percentage or a number.

### Thinking about the question

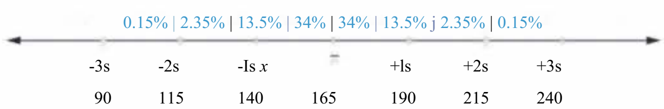
- You should expect the first parts of question 1 in Examination 2 to be relatively straightforward, but expect the later parts to be more difficult.
- Note that questions involving the 68-95-99.7% rule are best solved by drawing a scale or diagram.

### Worked solution (/ = 1 mark)

a i range = largest value - smallest value = 197 - 125 = 72 /

ii There are 3 clusters with more than 170 eggs and 12 clusters in total, so $\frac{3}{12} \times 100\% = 25\%$ / of clusters contain more than 170 eggs.

b i Write a 68-95-99.7% rule scale that includes the mean and standard deviations given.




From the scale, the percentage of clusters expected to contain more than 140 eggs is

34% + 34% + 13.5% + 2.35% + 0.15% = 84% /

ii From the scale, the percentage of clusters expected to have less than 215 eggs is

0.15% + 2.35% + 13.5% + 34% + 34% + 13.5% = 97.5%

Alternatively: 100% - (2.35% + 0.15%)
= 100% - 2.5% = 97.5%

The number of clusters expected to have less than 215 eggs in a sample of 1000 clusters is
1000 x 97.5% = 975 /

**UF Exam hack**

Always note carefully if a question is asking for a number or a percentage. Its a common mistake in exams to calculate the wrong one.

c Use the formula for converting standardised values into actual values.

$x = zs_x + \bar{x}$

$z = -2.4$, $s_x = 25$, $\bar{x} = 165$

$x = -2.4 \times 25 + 165$

$x = 105$

The actual number in this cluster is 105 / eggs.

**Student performance** 80-100% 60-79% 0-59%

a i 87% This was generally answered well, but a small number of students wrote the answer as an interval (e.g. '125 to 197'), which was not accepted.

ii 87%

b i 69%

ii 47% This was generally very poorly done. 97.5% was a very common incorrect answer. These students did not read the question properly and gave the answer as a percentage rather than a number.

c 67%

## EXERCISE 1.8 Standardised values

ANSWERS p. 695

### Recap

*Use the following information to answer the next two questions.*

The beak length of small birds in a large population is approximately normally distributed with a mean of 9.5 mm and a standard deviation of 0.50 mm.

1 ©VCAA Which one of the following statements relating to this population of birds is not true?

A No bird will have a beak length that is less than 8.0 mm.

B More than 99% of the birds will have a beak length that is less than 11 mm.

C Approximately half of the birds will have a beak length that is less than 9.5 mm.

D Approximately 2.5% of the birds will have a beak length that is greater than 10.5 mm.

E Approximately 34% of the birds will have a beak length that is between 9.5 mm and 10.0 mm.

2 ©VCAA 2017N 1CQ9 A random sample of 250 of these birds is captured and the beak length of each bird is measured. The expected number of these captured birds with beak lengths that are greater than 9 mm is closest to

A 6 B 13 C 170 D 210 E 244

### Mastery

3 E3 WORKED EXAMPLE 14 The heights of Year 12 teachers in Australia are known to be normally distributed with a mean of 175.6 cm and a standard deviation of 6.5 cm. A particular Year 12 teacher has a height of 162.6 cm.

a Calculate the standardised value for this actual value.

b Show that this standardised value means that 97.5% of Year 12 teachers are taller than this particular teacher.

c Another Year 12 teacher has a standardised height of $z = -2.3$. What is this teachers actual height?



9780170448406

4 H WORKED EXAMPLE 15 J The table shows the marks out of 100 that a student has achieved on the final exams, plus the means and standard deviations for each of the subjects.

| | Marks | Mean | Standard deviation |
|---|---|---|---|
| General Mathematics | 78 | 72 | 5 |
| German | 47 | 48 | 2 |
| Hospitality | 77 | 71 | 2 |
| Psychology | 50 | 54 | 4 |
| Systems Engineering | 62 | 68 | 5 |

Assuming the results for each of the subjects approximate a bell-shaped distribution:

a In which of the subjects did the student perform the best?

b In which of these subjects was the student in the bottom 16%? Draw a diagram which shows how you obtained your answer.

## Exam practice a0-10007' 6o-79% <«9%

*Use the following information to answer the next two questions.*

The lengths of the left feet of a large sample of Year 12 students were measured and recorded. These foot lengths are approximately normally distributed with a mean of 24.2 cm and a standard deviation of 4.2 cm.

5 ©VCAA 1 2010 1CQ5 J 80% A Year 12 student has a foot length of 23cm. The students standardised foot length (standard z-score) is closest to

A -1.2 B –0.9 C -0.3 D 0.3 E 1.2

6 ©VCAA | 20101CQ6 J 70% The percentage of students with foot lengths between 20.0 and 24.2 cm is closest to

A 16% B 32% C 34% D 52% E 68%

7 ©VCAA 72% A student obtains a mark of 56 on a test for which the mean mark is 67 and the standard deviation is 10.2. The students standardised mark (standard z-score) is closest to

A -1.08 B -1.01 C 1.01 D 1.08 E 49.4

*Use the following information to answer the next three questions.*

The pulse rates of a population of Year 12 students are approximately normally distributed with a mean of 69 beats per minute and a standard deviation of 4 beats per minute.

8 ©VCAA 20181CQ3 J 86% A student selected at random from this population has a standardised pulse rate of z = -2.5. This students actual pulse rate is

A 59 beats per minute. B 63 beats per minute. C 65 beats per minute.
D 73 beats per minute. E 79 beats per minute.

9 ©VCAA | 2018 1CQ4 , 68% A student selected at random from this population has a standardised pulse rate of z = -1. The percentage of students in this population with a pulse rate greater than this student is closest to

A 2.5% B 5% C 16% D 68% E 84%

▶ 10 ©VCAA 20181CQ5 69% A sample of 200 students was selected at random from this population. The number of these students with a pulse rate of less than 61 beats per minute or greater than 73 beats per minute is closest to

A 19 B 37 C 64 D 95 E 190

11 ©VCAA 2012JCQ4 68% A class of students sat for a Biology test and a Legal Studies test. Each test had a possible maximum score of 100 marks. The table shows the mean and standard deviation of the marks obtained in these tests. The class marks in each subject are approximately normally distributed. Sashi obtained a mark of 81 in the Biology test.

| | Subject | |
|---|---|---|
| | Biology | Legal Studies |
| Class mean | 54 | 78 |
| Class standard deviation | 15 | 5 |

The mark that Sashi would need to obtain on the Legal Studies test to achieve the same standard score for both Legal Studies and Biology is

A 81 B 82 C 83 D 87 E 95

12 (3 marks) The *weight* of a species of bird is approximately normally distributed with a mean of 71.5 g and a standard deviation of 4.5 g.

a What is the standardised weight (z-score) of a bird weighing 67.9 g? 1 mark

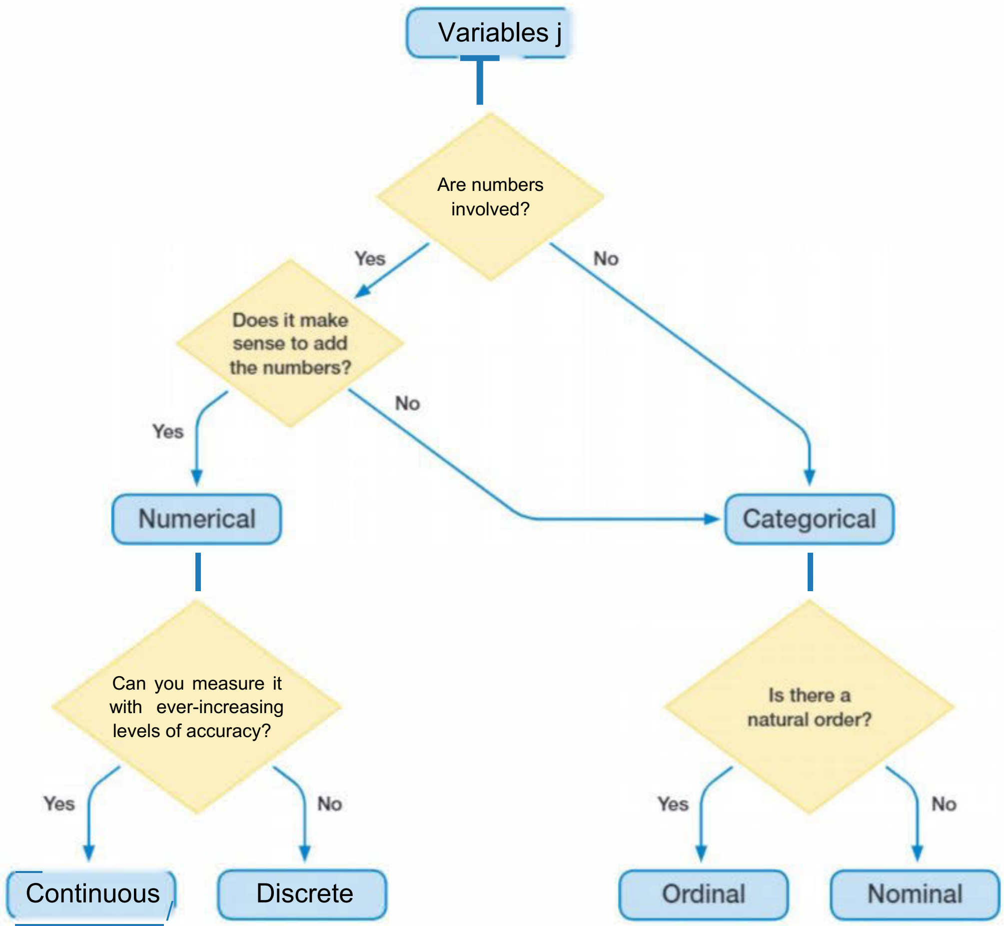
b Use the 68-95-99.7% rule to estimate

i the expected percentage of these birds that weigh less than 67 g 1 mark

ii the expected number of birds that weigh between 62.5 g and 76.0 g in a flock of 200 of these birds. 1 mark

# Chapter summary

## Classifying variables




## Describing data

### Measure of centre

| Measure of centre | Use for | Description |
|---|---|---|
| mode | numerical<br>categorical | • most frequently occurring data value<br>• is called the modal category for categorical data<br>• data with two modes is called bi-modal |
| mean | numerical | • is the average of all the data values<br>• $\bar{x} = \frac{\text{sum of all values}}{\text{number of values}}$ for a list of data values<br>• $\bar{x} = \frac{\text{sum of (each value} \times \text{its corresponding frequency)}}{\text{sum of frequencies}}$<br>for data in a frequency table |
| median | numerical<br>ordinal | 1 Order the *n* data values from smallest to largest.<br>2 Find the $\frac{H+1}{2}$th position. (Add one to the number of data values, then divide by 2.)<br>3 If n is odd, find the data value in the $\frac{n+1}{2}$th position.<br>4 If *n* is even, find the two data values either side of the $\frac{n}{2}$th position and average them. |

## The five-number summary

```
          25%           25%           25%           25%
-----------------1--------------- 1--------------- -----------------
Minimum Q, Median Q3 Maximum
              Median of                     Median of
              lower half                    upper half
```

## Measure of spread

| Measure of spread | Use for | Description |
|---|---|---|
| range | numerical | • measures the spread of the entire data set<br>• range = largest value - smallest value |
| interquartile range | numerical | • measures the spread of the middle 50% of the data values<br>. IQR=$Q_3$-$Q_1$ |
| standard deviation | numerical | • measures the spread around the mean<br>• use CAS to calculate $s_x$ |

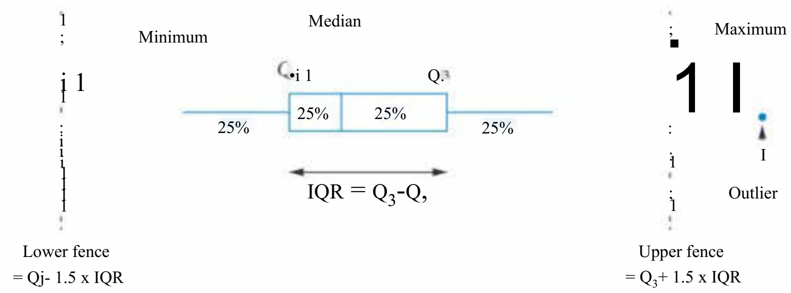
## Outliers

- An outlier is an extreme high or low data value.
- A data value is a possible outlier if it is less than the lower fence - 1.5 x IQR or greater than the upper fence $Q_3$ + 1.5 x IQR.

# Centre, spread, display and data type summary

| Categorical: Nominal data | Categorical: Ordinal data | Numerical data |
|---|---|---|
| | Measures of centre | |
| mode | mode<br>median | mode<br>median<br>mean |
| | Measures of spread | |
| – | range<br>IQR | range<br>IQR<br>standard deviation |
| | Displays | |
| dot plot<br>bar chart | dot plot<br>bar chart | dot plot<br>histogram<br>boxplot<br>stem plot |

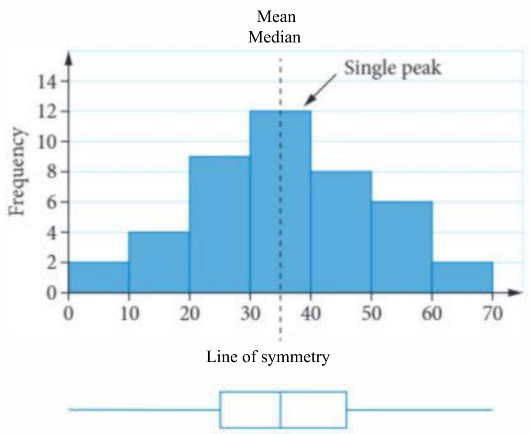
## Data displays

| Display | Description |
|---|---|
| Frequency table | • involves counting the number of times each data value occurs<br>• frequencies often shown as percentages<br>• percentage = $\frac{\text{frequency}}{\text{total}} \times 100\%$ |
| Bar chart | • bars can be horizontal or vertical |
| Segmented bar chart | • involves one bar with several segments and requires a key<br>• should not have too many segments |
| Percentage segmented bar chart | • segments represent percentages<br>• bar has height of 100 |
| Grouped frequency table | • involves numerical data that has been grouped into regular intervals<br>• makes it easier to deal with large amounts of data |
| Histogram | • graphical display of data from a grouped frequency table<br>• best if data is grouped into 5 to 15 intervals |
| Dot plot | • used for both categorical and numerical data<br>• should not involve too many data values or a large data spread |
| Stem plot | • involves actual data values and requires a key<br>• best used with a maximum of 50 data values and to see all the data values |
| Boxplot | • best if we want to read the five-number summary easily<br>• helps to find the IQR and outliers<br><br> |

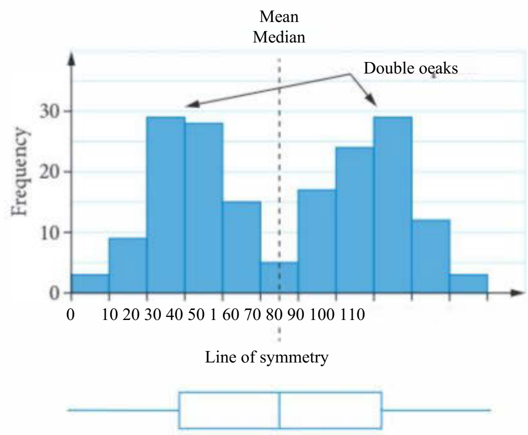
## Describing distributions

### Symmetric distributions

Single-peaked




Double-peaked



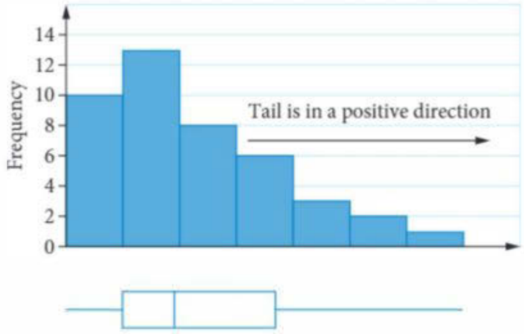

Modal interval: 30-<40

Bi-modal: 20-<30 and 80-<90

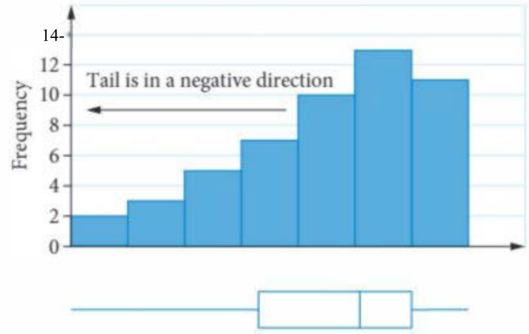
The mean and median are both close to the line of symmetry.

### Skewed distributions

Positively skewed



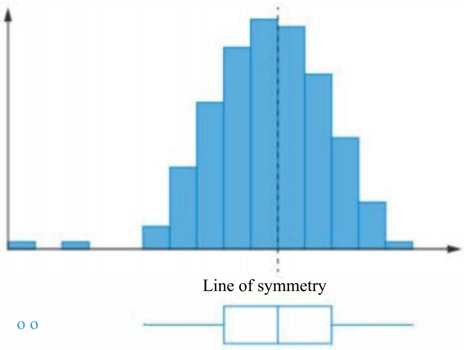

Negatively skewed




The mean is greater than the median.

The mean is less than the median.

### Distributions with outliers




In the above distribution, the mean and median are both close to the line of symmetry.

## Mean versus median / Standard deviation versus IQR

| Shape of distribution | Choose: | |
|---|---|---|
| Approximately symmetric distribution with no outliers | mean or median | standard deviation or IQR |
| Approximately symmetric distribution with outliers | median | IQR |
| Skewed distribution | median | IQR |

## Significant figures

Significant figures are

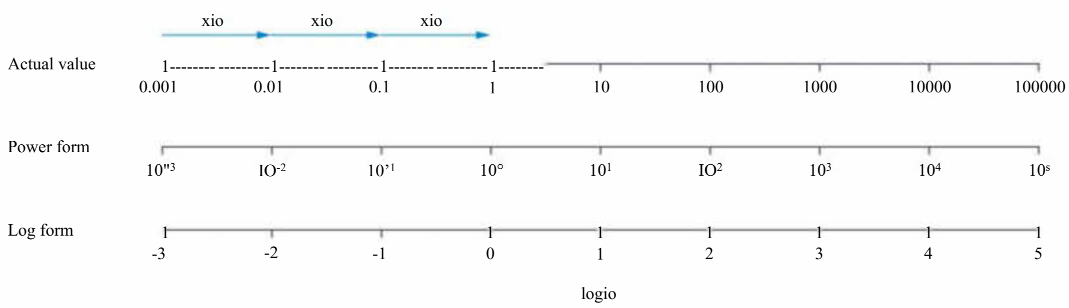
- any non-zero digit
- zeros between non-zero digits
- trailing zeros in decimals.

When rounding to significant figures, use the usual rounding rules: '0-4 round down and '5-9 round up'.

## Scales

- On a linear scale, we *add* the same number to move from one scale mark to the next.
- On a log scale, we *multiply* the same number to move from one scale mark to the next.

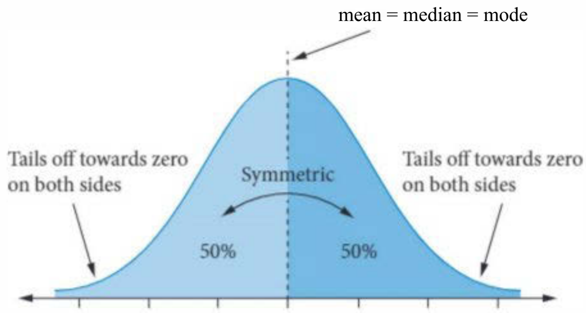
  Log scales can be written in three different ways:





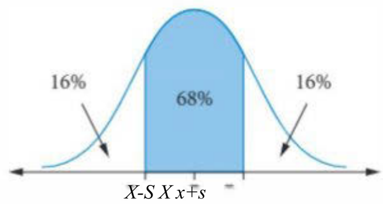

9780170448406

## Normal or bell-shaped distributions

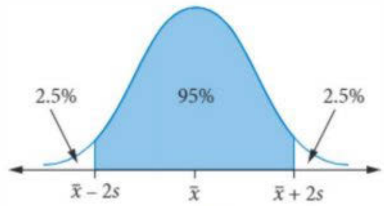



. The 68-95-99 .7% rule for approximate bell-shaped distributions says:

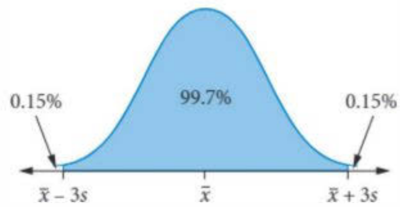
Around 68% of the data values lie within one standard deviation of the mean.



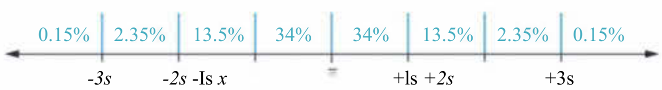

Around 95% of the data values lie within two standard deviations of the mean.




Around 99.7% of the data values lie within three standard deviations of the mean.



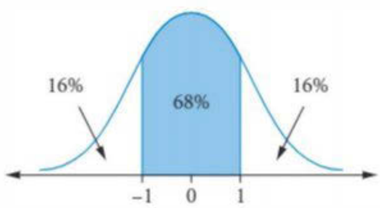

## 68-95-99.7% rule scale

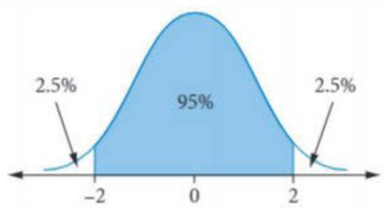


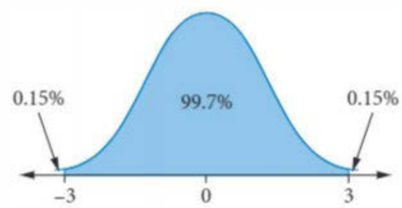

68-95-99.7% rule

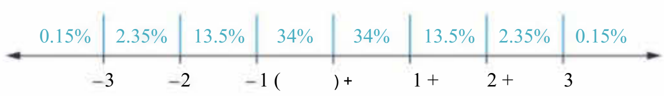
## Standardised values (^-scores)

- standardised value $= \dfrac{\text{data value - mean}}{\text{standard deviation}}$ or $z = \dfrac{x - \bar{x}}{s_x}$
- actual value = (standardised value x standard deviation) + mean or $x = z \times s_x + \bar{x}$
- 68-95-99.7% rule for z-scores:



# Cumulative examination 1

Total number of marks: 14 Reading time: 6 minutes Writing time: 32 minutes

*Use the following information to answer the next three questions.*

The following table shows the data collected from a random sample of seven drivers drawn from the population of all drivers who used a supermarket car park on one day. The variables in the table are:

- *distance* - the distance that each driver travelled to the supermarket from their home
- *sex* - the sex of the driver (female, male)
- *number of children* - the number of children in the car
- *type of car* - the type of car (sedan, wagon, other)
- *postcode* - the postcode of the drivers home.

| Distance (km) | Sex (F = female, M = male) | Number of children | Type of car (1 = sedan, 2 = wagon, 3 = other) | Postcode |
|---|---|---|---|---|
| 4.2 | F | 2 | 1 | 8148 |
| 0.8 | M | 3 | 2 | 8147 |
| 3.9 | F | 3 | 2 | 8146 |
| 5.6 | F | 1 | 3 | 8245 |
| 0.9 | M | 1 | 3 | 8148 |
| 1.7 | F | 2 | 2 | 8147 |
| 2.5 | M | 2 | 2 | 8145 |

1 ©VCAA 2016S 1CQ4 The number of discrete numerical variables in this data set is

A 0 B 1 C 2 D 3 E 4

2 ©VCAA 2016S 1CQ5 The number of ordinal variables in this data set is

A 0 B 1 C 2 D 3 E 4

3 ©VCAA 2016S 1CQ6 The number of female drivers with three children in the car is

A 0 B 1 C 2 D 3 E 4

4 ©VCAA The following ordered stem plot shows the areas, in square kilometres, of 27 suburbs of a large city.

| Stem | Leaf |
|---|---|
| 1 | 5 6 7 8 |
| 2 | 1 2 4 5 6 8 9 9 |
| 3 | 0 1 1 2 2 8 9 |
| 4 | 0 4 7 |
| 5 | 0 1 |
| 6 | 1 9 |
| 7 | |
| 8 | 4 |

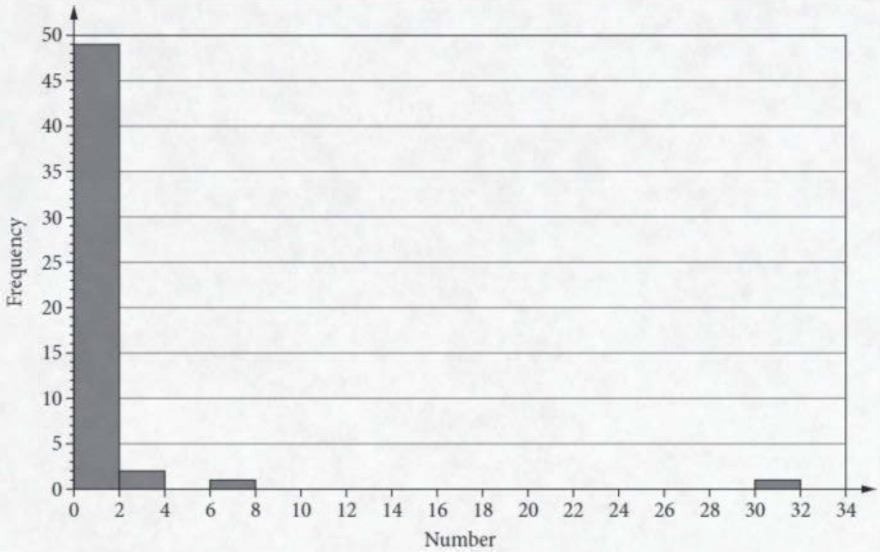
Key: 1|6 = 1.6 km$^2$

The median area of these suburbs, in square kilometres, is

A 3.0 B 3.1 C 3.5 D 30.0 E 30.5

*Use the following information to answer the next two questions.*

The histogram shows the distribution of the *number* of billionaires per million people for 53 countries.

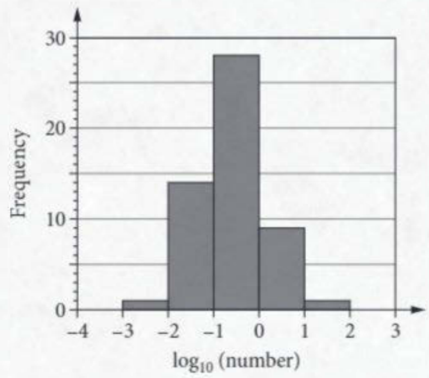



Data: Gapminder

5 ©VCAA 2016 1CQ6 Using this histogram, the percentage of these 53 countries with less than two billionaires per million people is closest to

A 49% B 53% C 89% D 92% E 98%

6 ©VCAA 2016 1CQ7 The histogram below shows the distribution of the *number* of billionaires per million people for the same 53 countries as in the previous question, but this time plotted on a $\log_{10}$ scale.




Data: Gapminder

Based on this histogram, the number of countries with one or more billionaires per million people is

A 1 B 3 C 8 D 9 E 10

7 ©VCAA 2021 1CQ5 The stem plot below shows the *height,* in centimetres, of 20 players in a junior football team.

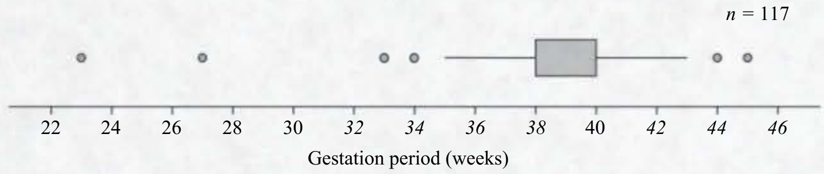
| Stem | Leaf |
|---|---|
| 14 | 2 2 4 7 8 8 9 |
| 15 | 0 0 1 2 5 5 6 8 |
| 16 | 0 1 1 2 |
| 17 | 9 |

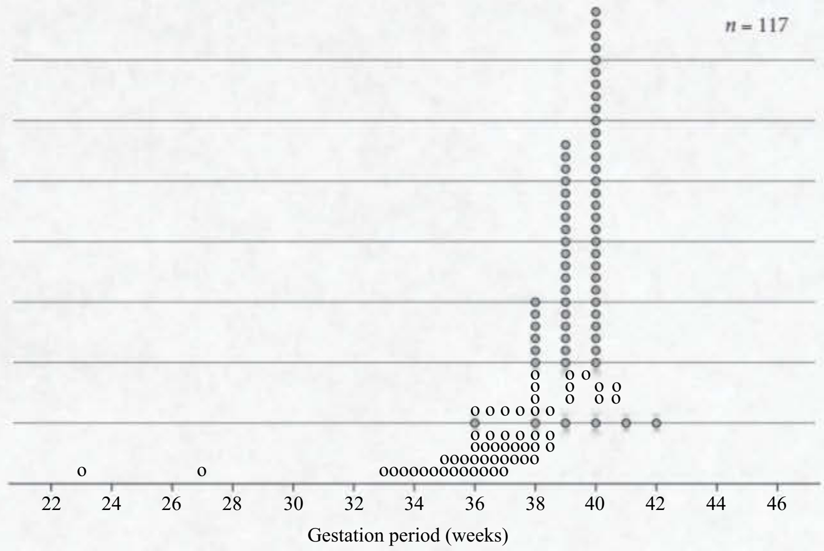
Key: 14|2 = 142 cm $n = 20$

A player with a height of 179 cm is considered an outlier because 179 cm is greater than

A 162 cm B 169 cm C 172.5 cm D 173 cm E 175.5 cm

8 ©VCAA 2019N1CQ4 The boxplot and dot plot shown below both display the distribution of the *gestation period,* in weeks, for 117 baby girls.




The median *gestation period,* in weeks, of these baby girls is

A 38 B 38.5 C 39 D 39.5 E 40

*Use the following information to answer the next two questions.*

The weights of male players in a basketball competition are approximately normally distributed with a mean of 78.6 kg and a standard deviation of 9.3 kg.

9 ©VCAA 2016 1CQ4 There are 456 male players in the competition. The expected number of male players in the competition with weights above 60 kg is closest to

A 3 B 11 C 23 D 433 E 445

**Exam hack**

Always aim to have 60% of the exam done by halfway through the exam.

10 ©VCAA 2016ICQ5J Brett and Sanjeeva both play in the basketball competition. When the weights of all players in the competition are considered, Brett has a standardised weight of $z = -0.96$ and Sanjeeva has a standardised weight of $z = -0.26$. Which one of the following statements is not true?

A Brett and Sanjeeva are both below the mean weight for players in the basketball competition.

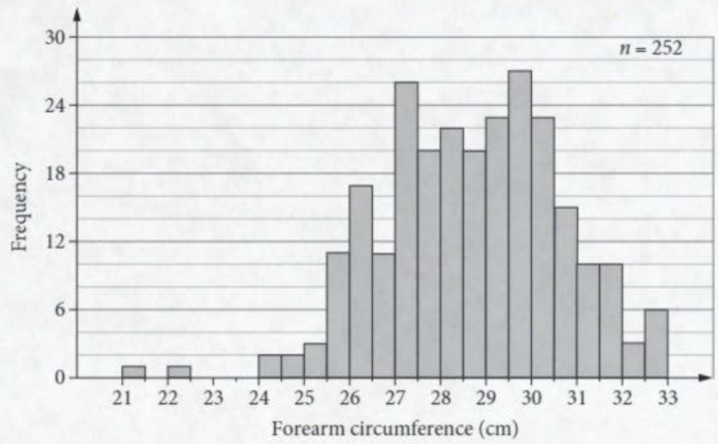
B Sanjeeva weighs more than Brett.

C If Sanjeeva increases his weight by 2 kg, he would be above the mean weight for players in the basketball competition.

D Brett weighs more than 68 kg.

E More than 50% of the players in the basketball competition weigh more than Sanjeeva.

11 ©VCAA | 20201CQ4 The histogram below shows the distribution of the *forearm circumference,* in centimetres, of 252 men. Assume that the *forearm circumference* values were all rounded to one decimal place.




The third quartile ($Q_3$) for this distribution could be

A 29.3 B 29.8 C 30.3 D 30.8 E 31.3

*Use the following information to answer the next three questions.*

The birth weights of a large population of babies are approximately normally distributed with a mean of 3300 g and a standard deviation of 550 g.

12 ©VCAA 2019N 1CQ5 A baby selected at random from this population has a standardised weight of $z = -0.75$. Which one of the following calculations will result in the *actual birth weight* of this baby?

A *actual birth weight* = 550 – 0.75 × 3300

B *actual birth weight* = 550 + 0.75 x 3300

C *actual birth weight* = 3300 - 0.75 x 550

D *actual birth weight* = $3300 + \frac{0.75}{550}$

E *actual birth weight* = $3300 - \frac{0.75}{550}$

13 EnZ3P°1 9N 1CQ6 J Using the 68-95-99.7% rule, the percentage of babies with a birth weight of less than 1650g is closest to

A 0.14% B 0.15% C 0.17% D 0.3% E 2.5%

14 ©VCAA 2Q19N1CQ7 J A sample of 600 babies was drawn at random from this population. Using the 68-95-99.7% rule, the number of these babies with a birth weight between 2200 g and 3850 g is closest to

A 111 B 113 C 185 D 408 E 489

# Cumulative examination 2

Total number of marks: 26 Reading time: 7 minutes Writing time: 39 minutes

1 © VCAA 2019N2CQ1J (4 marks) The table displays the *average sleep time,* in hours, for a sample of 19 types of mammals.

| Type of mammal | Average sleep time (hours) |
|---|---|
| cat | 14.5 |
| squirrel | 13.8 |
| mouse | 13.2 |
| rat | 13.2 |
| grey wolf | 13.0 |
| arctic fox | 12.5 |
| raccoon | 12.5 |
| gorilla | 12.0 |
| jaguar | 10.8 |
| baboon | 9.8 |
| red fox | 9.8 |
| rabbit | 8.4 |
| guinea pig | 8.2 |
| grey seal | 6.2 |
| cow | 3.9 |
| sheep | 3.8 |
| donkey | 3.1 |
| horse | 2.9 |
| roedeer | 2.6 |

Data: T Allison and DV Cicchetti, 'Sleep in Mammals: Ecological and Constitutional Correlates', In Science, American Association for the Advancement of Science, vol. 194, no. 4266, pp. 732-734, 12 November 1976; accessed from OzDASL, StatSci.org, www.statsci.org/data/general/sleep.html

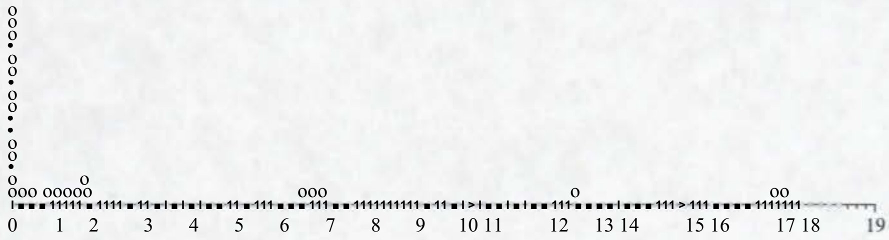
a Which of the two variables, *type of mammal* or *average sleep time,* is a nominal variable? 1 mark

b Determine the mean and standard deviation of the variable *average sleep time* for this sample of mammals. Round your answers to one decimal place. 1 mark

c The average sleep time for a human is eight hours. What percentage of this sample of mammals has an *average sleep time* that is less than the average sleep time for a human? Round your answer to one decimal place. 1 mark

d The sample is increased in size by adding in the average sleep time of the little brown bat. Its average sleep time is 19.9 hours. By how many hours will the range for *average sleep time* increase when the average sleep time for the little brown bat is added to the sample? 1 mark

2 ©VCAA 2016 2CQ1 (7 marks) The dot plot shows the distribution of daily rainfall, in millimetres, at a weather station for 30 days in September.




a Write down the

i range — 1 mark

ii median. — 1 mark

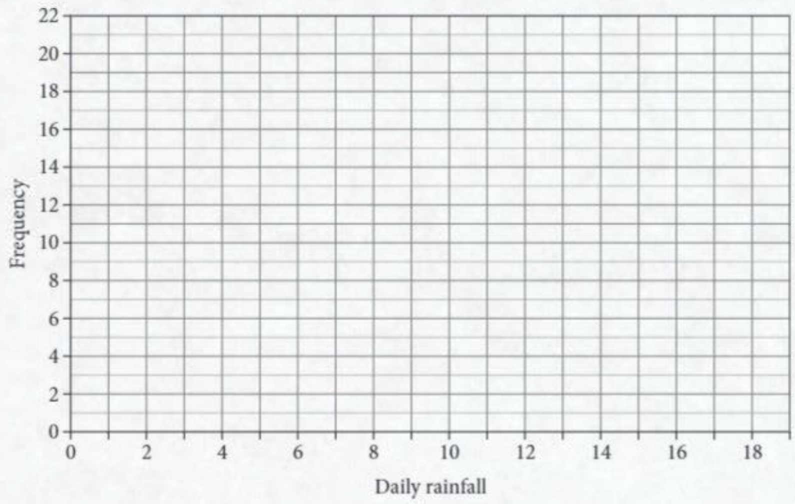
b Which data point on the dot plot corresponds to the third quartile ($Q_3$)? — 1 mark

c Write down

i the number of days on which no rainfall was recorded — 1 mark

ii the percentage of days on which the daily rainfall exceeded 12 mm. — 1 mark

d Construct a histogram on a grid like the one shown that displays the distribution of daily rainfall for the month of September. Use interval widths of two with the first interval starting at 0. — 2 marks




3 tawj 2011 2CQ1 I (5 marks) The stem plot in Figure 1 shows the distribution of the average age, in years, at which women first marry in 17 countries.

Figure 1: Average age, in years, of women at first marriage

| Stem | Leaf |
|---|---|
| 24 | |
| 25 | 0 |
| 26 | 6 |
| 27 | 113 4 7 |
| 28 | 2 2 2 3 3 6 |
| 29 | 1 1 |
| 30 | 1 4 |
| 31 | |

Key: 2713 = 27.3 years

a For these countries, determine

i the lowest average age of women at first marriage — 1 mark

ii the median average age of women at first marriage. — 1 mark

The stem plot in Figure 2 shows the distribution of the average age, in years, at which men first marry in 17 countries.

Figure 2: Average age, in years, of men at first marriage

| Stem | Leaf |
|---|---|
| 25 | |
| 26 | 0 |
| n | |
| 28 | 9 |
| 29 | 0 9 9 |
| 30 | 0 0 3 5 6 7 9 |
| 31 | 0 0 2 |
| 32 | 5 9 |
| 33 | |

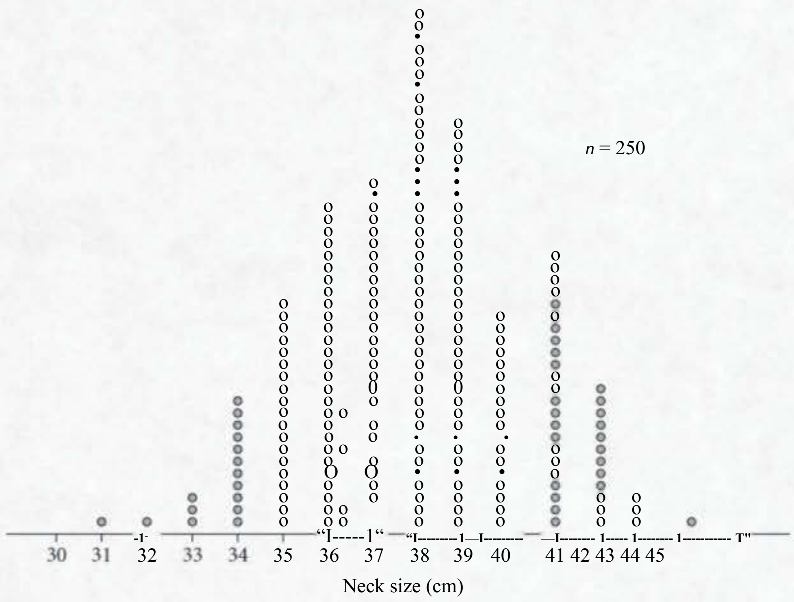
Key: 3215 = 32.5 years

b For these countries, determine the interquartile range (IQR) for the average age of men at first marriage. 1 mark

c If the data values displayed in Figure 2 were used to construct a boxplot with outliers, then the country for which the average age of men at first marriage is 26.0 years would be shown as an outlier.

Explain why this is so. Show an appropriate calculation to support your explanation. 2 marks

4 OVCAA 2020 2CQ2 (5 marks) The *neck size,* in centimetres, of 250 men was recorded and displayed in the dot plot below.




Data: RW Johnson, 'Fitting percentage of body fat to simple body measurements', *Journal of Statistics Education*, 4:1,1996, https://doi.Org/10.1080/10691898.1996.11910505

a Write down the modal *neck size,* in centimetres, for these 250 men. 1 mark

b Assume that this sample of 250 men has been drawn at random from a population of men whose *neck size* is normally distributed with a mean of 38 cm and a standard deviation of 2.3 cm.

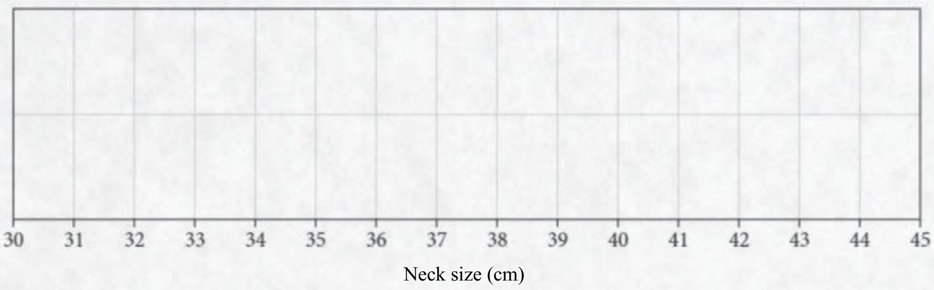
i How many of these 250 men are expected to have a *neck size* that is more than three standard deviations above or below the mean? Round your answer to the nearest whole number. 1 mark

ii How many of these 250 men actually have a *neck size* that is more than three standard deviations above or below the mean? 1 mark

c The five-number summary for this sample of neck sizes, in centimetres, is given below.

| Minimum | First quartile (Qd | Median | Third quartile (Qj) | Maximum |
|---|---|---|---|---|
| 31 | 36 | 38 | 39 | 44 |

Copy the grid below and use the five-number summary to construct a boxplot, showing any outliers if appropriate. 2 marks



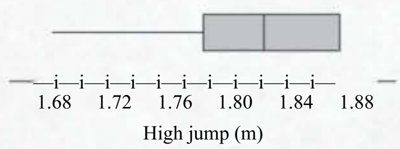

5 ©VCAA 2021 3CO1adef (5 marks) In the sport of heptathlon, athletes compete in seven events. These events are the 100 m hurdles, high jump, shot-put, javelin, 200 m run, 800 m run and long jump. Fifteen female athletes competed to qualify for the heptathlon at the Olympic Games. Their results for three of the heptathlon events - high jump, shot-put and javelin - are shown in the table.

| Athlete number | High jump (metres) | Shot-put (metres) | Javelin (metres) |
|---|---|---|---|
| 1 | 1.76 | 15.34 | 41.22 |
| 2 | 1.79 | 16.96 | 42.41 |
| 3 | 1.83 | 13.87 | 46.53 |
| 4 | 1.82 | 14.23 | 40.62 |
| 5 | 1.87 | 13.78 | 45.64 |
| 6 | 1.73 | 14.50 | 42.33 |
| 7 | 1.68 | 15.08 | 40.88 |
| 8 | 1.82 | 13.13 | 39.22 |
| 9 | 1.83 | 14.22 | 42.51 |
| 10 | 1.87 | 13.62 | 42.75 |
| 11 | 1.87 | 12.01 | 38.12 |
| 12 | 1.80 | 12.88 | 42.65 |
| 13 | 1.83 | 12.68 | 45.68 |
| 14 | 1.87 | 12.45 | 41.32 |
| 15 | 1.78 | 11.31 | 42.88 |

a Write down the number of numerical variables in the table. 1 mark

b In the qualifying competition, the heights jumped in the high jump are expected to be approximately normally distributed. Charas jump in this competition would give her a standardised score of $z = -1.0$. Use the 68-95-99.7% rule to calculate the percentage of athletes who would be expected to jump higher than Chara in the qualifying competition. 1 mark

c The boxplot shown was constructed to show the distribution of high jump heights for all 15 athletes in the qualifying competition.




Explain why the boxplot has no whisker at its upper end. 1 mark

d For the javelin qualifying competition, another boxplot is used to display the distribution of athletes' results. An athlete whose result is displayed as an outlier at the upper end of the plot is considered to be a potential medal winner in the event. What is the minimum distance that an athlete needs to throw the javelin to be considered a potential medal winner? 2 marks

**Exam hack**

A minimum or maximum is always a single number, not a range of numbers.

CHAPTER

2

# ASSOCIATIONS BETWEEN TWO VARIABLES

## Study Design coverage

### AREA OF STUDY 1: DATA ANALYSIS, PROBABILITY AND STATISTICS

#### Investigating association between two variables

- response and explanatory variables and their role in investigating associations between variables
- contingency (two-way) frequency tables, their associated bar charts (including percentage segmented bar charts) and their use in identifying and describing associations between two categorical variables
- back-to-back stem plots, parallel dot plots and boxplots and their use in identifying and describing associations between a numerical variable and a categorical variable
- scatterplots and their use in identifying and qualitatively describing the association between two numerical variables in terms of direction (positive/negative), form (linear/non-linear) and strength (strong/moderate/weak)
- answering statistical questions that require a knowledge of the associations between pairs of variables
- Pearson correlation coefficient, r, and its calculation and interpretation
- cause and effect; the difference between observation and experimentation when collecting data and the need for experimentation to definitively determine cause and effect.



STUDY DESIGN COVERAGE

**Video playlists (6):**

- **2.1** Explanatory and response variables
- **2.2** Associations between two categorical variables
- **2.3** Associations between numerical and categorical variables
- **2.4** Associations between two numerical variables
- **2.5** Correlation and causation
- **VCE question analysis** Associations between two variables

**Worksheets (11):**

- **2.2** Two-way tables 1 • Two-way tables 2 • Percentage tables
- **2.3** Comparing data • Comparing group measures • Calculating and interpreting statistics
- **2.4** Scatterplots and associations • A page of scatterplots • Height vs shoe size • Body measurements
- **2.5** Relationships between variables

# Nelson MindTap

To access resources above, visit cengage.com.au/nelsonmindtap

# @ Explanatory and response variables

Video playlist Explanatory and response variables

## Dealing with two variables

All the examples in Chapter 1 have involved only one variable. From now on we will be exploring associations between two variables. Data associated with two related variables is called bivariate data.

Here are some examples of questions that involve analysing the association between two variables.

- Does human activity explain global warming?
- Is there a relationship between vaping and lung cancer?
- Does the number of books in a home predict academic success?
- How likely is it that the amount of time spent on social media affects VCE results?
- Is there an association between the amount of chocolate eaten and happiness?

## Identifying explanatory and response variables

It's important to decide which one of the variables is the explanatory variable and which is the response variable. An explanatory variable is a variable that we use to predict or explain the changes observed in another variable, which is called a response variable.

**Exam hack**

When deciding whether a variable is explanatory or response, look for the words 'explain changes' or 'predict'. If they don't appear, think about which of the two variables is the most likely to affect the other variable.

WB p. 19

**WORKED EXAMPLE 1** Identifying explanatory and response variables

For each of the following

i identify the two variables

ii state whether each one is categorical or numerical

iii identify the explanatory variable, giving a reason for your answer.

| Steps | Working |
|---|---|
| **a** An investigation is done to see whether the size of the crowd at an AFL match can be predicted by the team's position on the AFL ladder. | |
| i Consider the sort of data that will be collected. | *crowd size* and *position on ladder* |
| ii Are numbers involved? Does it make sense to add the numbers? | *Crowd size* is numerical and *position on ladder* is categorical. |
| iii Which variable is 'explaining' or 'predicting' the other? If the wording isn't used, which variable is most likely to affect the other? | The aim of the investigation is to predict crowd size from the team's position on the AFL ladder, so *position on ladder* is the explanatory variable. |
| **b** Researchers wish to investigate the association between the age of a person and the amount of time they sleep. | |
| i Consider the sort of data that will be collected. | *age* and *time spent sleeping* |
| ii Are numbers involved? Does it make sense to add the numbers? | Both variables are numerical. |
| iii Which variable is 'explaining' or 'predicting' the other? If the wording isn't used, which variable is most likely to affect the other? | A person's age is likely to affect the time they spend sleeping. However, the time a person spends sleeping is not likely to affect their age. So, *age* is the explanatory variable. |


9780170448406

| | |
|---|---|
| **c** A statistical analysis is done to establish whether there is a connection between the cost of avocados and the season. | |
| i Consider the sort of data that will be collected. | *cost of avocados* and *season* |
| ii Are numbers involved? Does it make sense to add the numbers? | *Cost of avocados* is numerical and *season* is categorical. |
| iii Which variable is 'explaining' or 'predicting' the other? If the wording isn't used, which variable is most likely to affect the other? | The season is likely to affect the cost of avocados. However, the cost of avocados is not likely to affect the season. So, *season* is the explanatory variable. |
| **d** A study is undertaken to establish whether a person's alcohol consumption (measured by the average number of standard alcoholic drinks per week) can explain depression levels (mild, moderate or severe). | |
| i Consider the sort of data that will be collected. | *alcohol consumption* and *depression levels* (mild, moderate or severe) |
| ii Are numbers involved? Does it make sense to add the numbers? | *Alcohol consumption* is numerical and *depression levels* (mild, moderate or severe) is categorical. |
| iii Which variable is 'explaining' or 'predicting' the other? If the wording isn't used, which variable is most likely to affect the other? | The aim of the study is to explain depression levels from alcohol consumption, so *alcohol consumption* is the explanatory variable in this study. Note: it is also possible that depression levels affect alcohol consumption, so in another study, *depression levels* could be the explanatory variable. |

## EXERCISE 2.1 Explanatory and response variables

ANSWERS p. 696

### Mastery

1 **a** WORKED EXAMPLE 1 I For each of the following

i identify the two variables

ii state whether each variable is categorical or numerical

iii identify the explanatory variable, giving a reason for your answer.

a A researcher investigates the association between the average daily screen time of Year 12 students and their ATAR scores.

b Research is undertaken to see whether stress levels (scale of 0 to 5, where 0 is none and 5 is extremely high) can predict headache levels (1 = mild, 2 = moderate, 3 = severe).

c A study is done to establish whether there is an association between the time taken to complete a Fun Run and the age of a person.

d A statistical analysis is undertaken to investigate whether there is a relationship between a person's gender and whether they are left- or right-hand dominant.

e An experiment is conducted to see whether driving response times (in milliseconds) can be explained by levels of sleep deprivation (1 = low, 2 = medium, 3 = high).

f A researcher wants to determine whether there is a connection between Year 12 students' English study scores and the number of television sets in their homes.

## Exam practice

80-100% 60-79% 0-59%

2 The temperatures (°C) at 8 am and 3 pm were measured each day in December 2021 to see whether it was possible to predict the temperature at 3 pm from the temperature at 8 am. The explanatory variable is

A *temperature* (°C).
B *temperature at 3pm.*
C *temperature in December.*
D *temperature at 8 am.*
E *temperature in 2021.*

3 A consumer group conducts research to determine whether the price of a laptop is affected by its weight. They weigh 50 laptops and note each of their prices on a spreadsheet. The explanatory and response variables are, respectively

A *50 laptops* and *laptop prices.*
B *laptop prices* and *laptop weight.*
C *laptop weight* and *laptop prices.*
D *laptop prices* and *spreadsheet.*
E *consumer group* and *study.*

4 A study is done to understand the association between the breed of a dog and its life span (months). Which of the following statements is not true?

A One of the variables is numerical and the other is categorical.
B The explanatory variable is *breed.*
C The response variable is *life span.*
D The breed of the dog may affect its life span.
E The lifespan of the dog may affect its breed.

Video playlist Associations between two categorical variables

Worksheets Two-way tables 1

Two-way tables 2

Percentage tables

# @ Associations between two categorical variables

Two variables are said to be associated when they are linked in some way.

We say two categorical variables are

- associated, if the explanatory variable categories give considerably *different* results
- *not* associated, if the explanatory variable categories give *similar* results.

To find whether two categorical variables are associated or not, we compare percentages in

- percentage two-way frequency tables
- parallel percentage segmented bar charts.

## Two-way frequency tables

A two-way frequency table (or contingency table) is an effective way to display two categorical variables. Here is an example of a two-way frequency table. It looks at the association between the variables *gender* and *exercise habits.*

It's possible to have more than two categories for each variable.

Explanatory variable

Response variable

| | Gender | | |
|---|---|---|---|
| Exercise habits | Male | Female | Total |
| Exercise | 3028 | 1084 | 4112 |
| No exercise | 1532 | 946 | 2478 |
| Total | 4560 | 2030 | 6590 |

Row totals

Number of males who exercise

Column totals

### © Exam hack

The variable appearing in the columns of a two-way frequency table is usually, *but not always,* the explanatory variable.



9780170448406

**WORKED EXAMPLE 2** Creating two-way frequency tables

Sixty people were interviewed about their biscuit preference. Of the 35 women, 20 said they preferred chocolate biscuits, whereas 19 of the men preferred biscuits without chocolate.

a Name the explanatory and the response variables.

b Construct a two-way frequency table to show this information, using *gender* for the columns.

| Steps | Working |
|---|---|
| a Name the explanatory and the response variables. | The explanatory variable is *gender*.<br>The response variable is *biscuit preference*. |

b 1 Create a table using *gender* as the explanatory variable for the columns and *biscuit preference* as the response variable for the rows.

| | Gender | | |
|---|---|---|---|
| Biscuit preference | Female | Male | Total |
| Chocolate | | | |
| Without chocolate | | | |
| Total | | | |

2 Fill in the information from the question.

- 60 people
- 35 women
- 20 women preferred chocolate biscuits
- 19 men preferred biscuits without chocolate

| | Gender | | |
|---|---|---|---|
| Biscuit preference | Female | Male | Total |
| Chocolate | 20 | | |
| Without chocolate | | 19 | |
| Total | 35 | | 60 |

3 Complete the table using column and row totals.

20+15 = 35
35 + 25 = 60
15+19 = 34
6+19 = 25
20 + 6 = 26

| | Gender | | |
|---|---|---|---|
| Biscuit preference | Female | Male | Total |
| Chocolate | 20 | 6 | 26 |
| Without chocolate | 15 | 19 | 34 |
| Total | 35 | 25 | 60 |

## Percentage two-way frequency tables

To find whether there are differences between categories, and therefore to find if there is an association between the variables, convert the two-way frequency table to a percentage two-way frequency table. There are a number of ways to calculate the percentages in a two-way frequency table, depending on the association we are finding, but we usually calculate each data value as a percentage of the explanatory variable (the column total, not the row total).

Percentaging gender in the two-way frequency table of gender and exercise habits gives the table on the right. The percentages are rounded to one decimal place.

From this percentage two-way frequency table, we can see that there is a *difference* in exercise habits between males and females: 66.4% of males exercise

Two-way frequency table

| | Gender | | |
|---|---|---|---|
| Exercise habits | Male | Female | Total |
| Exercise | 3028 | 1084 | 4112 |
| No exercise | 1532 | 946 | 2478 |
| Total | 4560 | 2030 | 6590 |

Percentage two-way frequency table

| | Gender | |
|---|---|---|
| Exercise habits | Male | Female |
| Exercise | 66.4% | 53.4% |
| No exercise | 33.6% | 46.6% |
| Total | 100.0% | 100.0% |

$$\frac{\text{data value}}{\text{column total}} \times 100\% = \frac{3028}{4560} \times 100\%$$

2.2

WB p. 21

whereas only 53.4% of females exercise, which is a difference of over 10%. This suggests that there is an association between gender and exercise habits.

If the male and female percentages had been approximately equal, we would have said there is no association, which would mean knowing a persons gender doesn't tell us anything about their exercise habits.

WB
p. 22

## WORKED EXAMPLE 3 Working with percentage two-way frequency tables

a Convert the following two-way frequency table into a percentage two-way frequency table by percentaging the explanatory variable. Round to the nearest percentage.

b What does the table suggest about the association between a persons gender and biscuit preference? Give reasons by referring to percentages.

| | Gender | | |
|---|---|---|---|
| Biscuit preference | Female | Male | Total |
| Chocolate | 28 | 32 | 60 |
| Without chocolate | 10 | 26 | 36 |
| Total | 38 | 58 | 96 |

**Steps** | **Working**

a 1 The explanatory variable forms the columns, so redraw the two-way table using only the column totals.

| | Gender | |
|---|---|---|
| Biscuit preference | Female | Male |
| Chocolate | 28 | 32 |
| Without chocolate | 10 | 26 |
| Total | 38 | 58 |

2 Calculate the required percentages.

Females preferring chocolate: $\frac{28}{38}$ x 100 « 74%

Females preferring without chocolate: $\frac{10}{38}$ x 100 « 26%

Males preferring chocolate: $\frac{32}{58}$ x 100 « 55%

Males preferring without chocolate: $\frac{26}{58}$ x 100 « 45%

3 Write the percentages in the two-way table.

| | Gender | |
|---|---|---|
| Biscuit preference | Female | Male |
| Chocolate | 74% | 55% |
| Without chocolate | 26% | 45% |
| Total | 100% | 100% |

b An association means the explanatory variable categories give considerably different results. Refer to percentages in your answer.

The table suggests there is an association between gender and chocolate biscuit preferences. 74% of females prefer biscuits with chocolate compared to only 55% of males. The difference in biscuit preference between males and females is nearly 20%, indicating that females have a greater preference for chocolate biscuits than males.

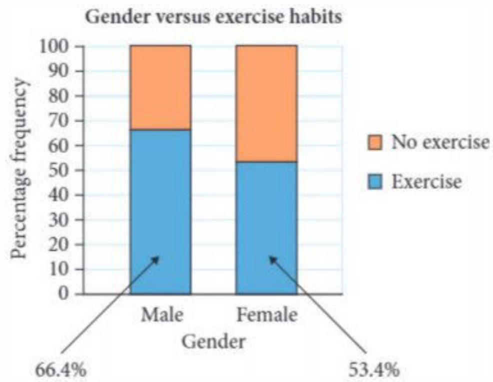
### Exam hack

Don't use the words 'evidence' or 'proof' when discussing associations. Results are never clear-cut when it comes to associations. Also, don't use the word 'significant' because this has a specific meaning in statistics. Instead, use words like 'indicate', 'suggest', 'considerably higher', 'similar', and 'noticeably different'.

## Parallel percentage segmented bar charts

It is often easier to see patterns in data when it is displayed as a chart rather than a table. Information from a percentage two-way frequency table can be displayed as a parallel percentage segmented bar chart where each cell in the table corresponds to a segment in the bar chart.

For example, the percentage two-way frequency table of gender and exercise habits can be shown as the following parallel percentage segmented bar chart:

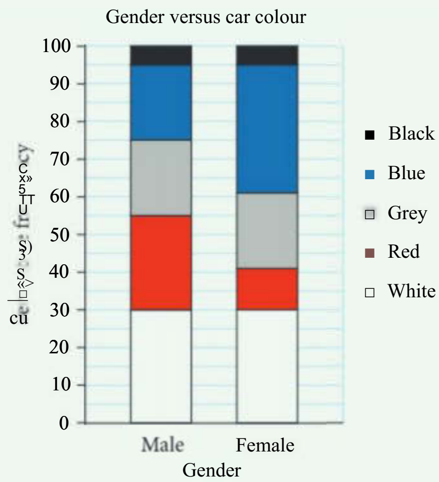



Parallel percentage segmented bar charts are particularly useful when we are dealing with data from across several populations.

WB p. 23

### WORKED EXAMPLE 4 Reading parallel percentage segmented bar charts

The parallel percentage segmented bar chart shows the results of a survey of preferred car colour of males and females. Discuss whether this suggests there is an association between car colour and gender by comparing percentages.




| Steps | Working |
| --- | --- |
| Look at how many of the segments are similar and how many are different. | There are considerable differences in the percentages of males and females who prefer red cars (male 25% and female 11%) and blue cars (male 20% and female 34%), which by itself would suggest that there may be an association between car colour and gender.<br>However, the three other colours had very similar percentages for males and females: white (30%), grey (20%) and black (5%). This suggests that there may be an association between car colour and gender but only with certain colours. |

## EXERCISE 2.2 Associations between two categorical variables

ANSWERS p. 697

### Recap

1 A study is done to investigate the relationship between the monthly electricity cost of a household and the number of people in the household.

The response variable is

A categorical.

B *number of people in household.*

C *household cost.*

D nominal.

E *monthly electricity cost.*

2 Research is undertaken to see whether stress scores can predict study scores in VCE examinations. Which of the following statements is not true?

A Both variables are ordinal.

B The response variable is numerical.

C The explanatory variable is numerical.

D The response variable is *study score.*

E The explanatory variable is *stress score.*

### Mastery

3 State whether it is possible to set up a two-way frequency table for each of these pairs of variables,

a *height* and *weight*

b *stress level* (low, medium or high) and *gender* (male, female, other)

c *hair colour* and *eye colour*

d *attitude to school subject* (scale of 1 to 5 where 1 is hate and 5 is love) *and favourite reality TV series*

e *favourite colour* and *age*

4 EJ WORKED EXAMPLE 2 Seventy people were asked about whether or not they owned more than one mobile phone. Of the 32 women in the survey, 12 owned more than one mobile phone, whereas 15 of the men in the survey owned only one.

a Name the explanatory variable and the response variable.

b Construct a two-way frequency table to show this information, using *gender* for the columns.

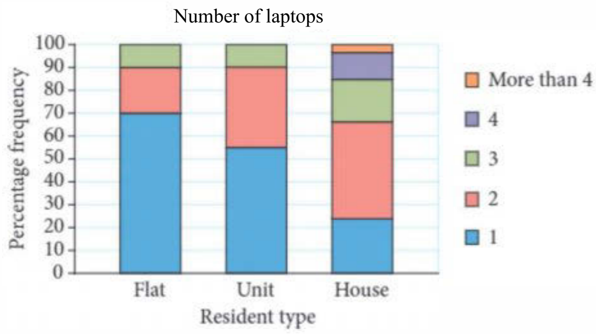
5 EJ WORKED EXAMPLE 3J

a Convert the following two-way frequency table into a percentage two-way frequency table by percentaging the explanatory variable. Round to the nearest percentage.

| | Gender | | |
|---|---|---|---|
| Sporting club membership | Mal<e | Female | Total |
| Sporting club member | 24 | 38 | 62 |
| Not a Sporting club member | 12 | 17 | 29 |
| Total | 36 | 55 | 91 |

b What does the table suggest about the association between a persons gender and sporting club membership? Give reasons by referring to percentages.

▶ **60** WORKED EXAMPLE 4 | The following parallel percentage segmented bar chart shows the results of a survey about the *type of residence* and the *number of laptops* in it. Discuss whether this suggests there is an association between the type of residence and the number of laptops by comparing percentages.




2.2

## Exam practice

80-100% 60-79% 0-59%

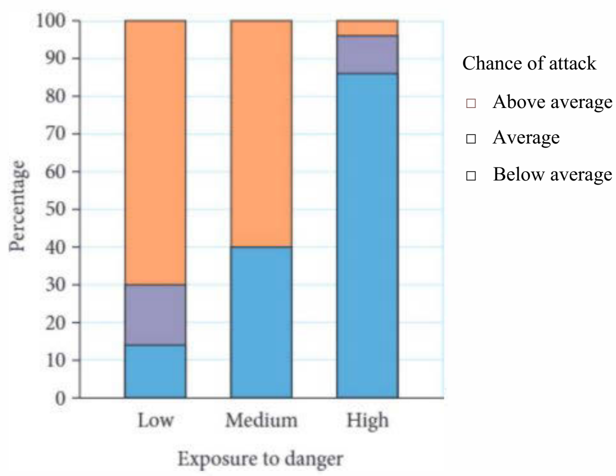
7 ©VCAA | 20091CQ8 | 88% An animal study was conducted to investigate the association between *exposure to danger* during sleep (high, medium, low) and *chance of attack* (above average, average, below average). The results are summarised in the percentage segmented bar chart.

The percentage of animals whose *exposure to danger* during sleep is high, and whose *chance of attack* is below average, is closest to

A 4% B 12%

C 28% D 72%

E 86%




8 ©VCAA 2012 1CQ6 | 84% The table below shows the percentage of households with and without a computer at home for the years 2007, 2009 and 2011.

| | Year | | |
|---|---|---|---|
| | 2007 | 2009 | 2011 |
| Households with a computer | 66.4% | 77.7% | 84.5% |
| Households without a computer | 33.6% | 22.3% | 15.5% |

In the year 2009, a total of 5170 000 households were surveyed.

The number of households without a computer at home in 2009 was closest to

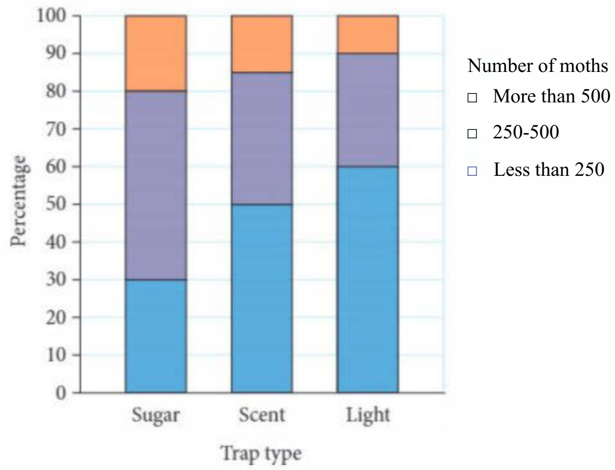
A 801000 B 1 153000 C 1737000

D 3433000 E 4017000

▶

*Use the following information to answer the next two questions.*

A study was conducted to investigate the association between the *number of moths* caught in a moth trap (less than 250, 250-500, more than 500) and the *trap type* (sugar, scent, light). The results are summarised in the percentage segmented bar chart.




9 ©VCAA 20171CQ5 76% There were 300 sugar traps. The number of sugar traps that caught less than 250 moths is closest to

A 30 B 90 C 250 D 300 E 500

10 ©VCAA 20171CQ6 71% The data displayed in the percentage segmented bar chart supports the contention that there is an association between the *number of moths* caught in a moth trap and the *trap type* because

A most of the light traps contained less than 250 moths.

B 15% of the scent traps contained 500 or more moths.

C the percentage of sugar traps containing more than 500 moths is greater than the percentage of scent traps containing less than 500 moths.

D 20% of sugar traps contained more than 500 moths while 50% of light traps contained less than 250 moths.

E 20% of sugar traps contained more than 500 moths while 10% of light traps contained more than 500 moths.

11 ©VCAA 20201CQ6 72% A percentage segmented bar chart would be an appropriate graphical tool to display the association between *month of the year* (January, February, March, etc.) and the

A *monthly average rainfall* (in millimetres).

B *monthly mean temperature* (in degrees Celsius).

C *annual median wind speed* (in kilometres per hour).

D *monthly average rainfall* (below average, average, above average).

E *annual average temperature* (in degrees Celsius).

12 ©VCAA 2012CQ7 70% The table below shows the percentage of students in two age groups (15-19 years and 20-24 years) who regularly use the internet at one or more of three locations:

- at home
- at an educational institution
- at work.

| | Age group | |
|---|---|---|
| Location of Internet use | 15-19 years | 20-24 years |
| at home | 95% | 95% |
| at an educational institution | 85% | 18% |
| at work | 38% | 74% |

For the students surveyed, which one of the following statements, by itself, supports the contention that the location of internet use is associated with the age group of the internet user?

A 85% of students aged 15-19 years used the internet at an educational institution.

B 95% of students aged 15-19 years used the internet at home, but only 38% of 15-19 year olds used it at work.

C 95% of students aged 15-19 years used the internet at home and 18% of 20-24 year olds used the internet at an educational institution.

D The percentage of students who used the internet at an educational institution decreased from 85% for those aged 15-19 years to 18% for those aged 20-24 years.

E The percentage of students who used the internet at home was 95% for those aged 15-19 years and 95% for those aged 20-24 years.

13 ©VCAA 2019CQ8 63% Percy conducted a survey of people in his workplace. He constructed a two-way frequency table involving two variables.

One of the variables was *attitude towards shorter working days* (for, against).

The other variable could have been

A *age* (in years).
B *sex* (male, female).
C *height* (to the nearest centimetre).
D *income* (to the nearest thousand dollars).
E *time* spent travelling to work (in minutes).

14 ©VCAA 2003CQ6 53% The level of water usage of 250 houses was rated in a survey as low, medium or high, and the size of the houses as small, standard or large. The results of the survey are displayed in the table below.

| | Size of house | | |
|---|---|---|---|
| Level of water usage | Small | Standard | Large |
| Low | 15 | 14 | 9 |
| Medium | 22 | 71 | 11 |
| High | 15 | 47 | 46 |

The percentage of standard-sized houses rated as having a high level of water usage is

A 18.8% B 35.6% C 43.5% D 47.0% E 53.8%

15 ©VCAA 2018 2CQ1a-c (6 marks) The data in Table 1 relates to the impact of traffic congestion in 2016 on travel times in 23 cities in the United Kingdom (UK).

The four variables in this data set are:

- *city* - name of city
- *congestion level* - traffic congestion level (high, medium, low)
- *size* - size of city (large, small)
- *increase in travel time* - increase in travel time due to traffic congestion (minutes per day).

Table 1

| City | Congestion level | Size | Increase in travel time (minutes per day) |
|---|---|---|---|
| Belfast | high | small | 52 |
| Edinburgh | high | small | 43 |
| London | high | large | 40 |
| Manchester | high | large | 44 |
| Brighton and Hove | high | small | 35 |
| Bournemouth | high | small | 36 |
| Sheffield | medium | small | 36 |
| Hull | medium | small | 40 |
| Bristol | medium | small | 39 |
| Newcastle-Sunderland | medium | large | 34 |
| Leicester | medium | small | 36 |
| Liverpool | medium | large | 29 |
| Swansea | low | small | 30 |
| Glasgow | low | large | 34 |
| Cardiff | low | small | 31 |
| Nottingham | low | small | 31 |
| Birmingham-Wolverhampton | low | large | 29 |
| Leeds-Bradford | low | large | 31 |
| Portsmouth | low | small | 27 |
| Southampton | low | small | 30 |
| Reading | low | small | 31 |
| Coventry | low | small | 30 |
| Stoke-on-Trent | low | small | 29 |



a 72% How many variables in this data set are categorical variables? 1 mark

b 71% How many variables in this data set are ordinal variables? 1 mark

c 80% Name the large UK cities with a medium level of traffic congestion. 1 mark d | 97% Use the data in Table 1 to copy and complete the following two-way frequency table, Table 2. 2 marks

Table 2

| | City size | |
|---|---|---|
| Congestion level | Small | Large |
| high | 4 | |
| medium | | |
| low | | |
| Total | 16 | |

e 81% What percentage of the small cities have a high level of traffic congestion? 1 mark

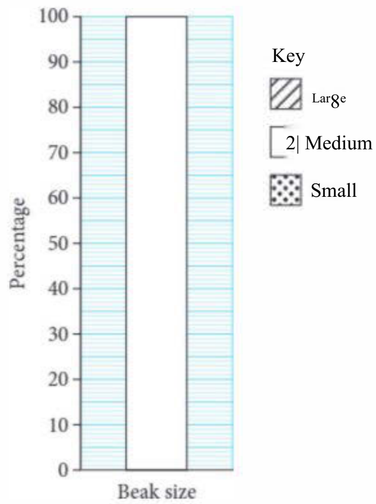
16 ©VCAA 2018N2CQ4 (4 marks) A sample of 96 birds are grouped according to their *beak size* (small, medium, large). The percentage of birds in each group is calculated. The results are displayed in Table 1.

Table 1

| Beak size | Percentage (%) |
|---|---|
| small | 25 |
| medium | 44 |
| large | 31 |
| Total | 100 |

a How many of the 96 birds have small beaks? 1 mark

b Use the percentages in Table 1 to construct a percentage segmented bar chart. Copy the template below to assist you in completing this task. Use the key to indicate the segment of your bar chart that corresponds to each beak size. 1 mark




c In order to investigate a possible association between *beak size* and *sex*, the same birds are grouped by both their *beak size* (small, medium, large) and their *sex* (male, female). The results of this grouping are shown in Table 2.

Table 2

| | Sex | |
|---|---|---|
| Beak size | Male | Female |
| small | 1 | 23 |
| medium | 26 | 16 |
| large | 27 | 3 |
| Total | 54 | 42 |

Does the information provided above support the contention that *beak size* is associated with sex? Justify your answer by quoting appropriate percentages. It is sufficient to consider one *beak size* only when justifying your answer. 2 marks

Video playlist
Associations between numerical and categorical variables

Worksheets
Comparing data

Comparing group measures

Calculating and interpreting statistics

# 2.3 Associations between numerical and categorical variables

We can also find associations where one variable is categorical and the other variable is numerical.

As with two categorical variables, we say a categorical and numerical variable are

- associated, if the explanatory variable categories give considerably *different* results
- *not* associated, if the explanatory variable categories give *similar* results.

To find whether a numerical variable and categorical variable are associated or not, we can compare

1 shape: symmetry and skewness
2 centre: median
3 spread: range and IQR

The following displays can be used:

- back-to-back stem plots
- parallel dot plots
- parallel boxplots.

## Back-to-back stem plots

Back-to-back stem plots are a good choice if we are dealing with just two sets of data values for the same variable and want to see the actual data values. A back-to-back stem plot has two sets of leaves, one on the left of the stem and one on the right. This allows us to display the data for the two groups being compared, as in the example below.

Test results for two classes

This represents the data value 52%.

| Class A | Stem | Class B |
|---|---|---|
| 4 3(T | 5j | |
| 8 5 4 3 1 | 6 | 0 0 2 4 |
| 9 6 4 4 3 1 0 | 7 | 35678889 |
| 087532® | 8 | 0 14 6 9 |
| 1 0< | X | 2 3 5 6 7 7 8 |

4|6 = 64% 6|4 = 64%

Largest value in this leaf.

Smallest value in this leaf.

### Exam hack

As with one-sided stem plots, when there are a small number of stems for back-to-back stem plots, we split the stem to see the distribution more clearly.



9780170448406

## WORKED EXAMPLE 5 Interpreting back-to-back stem plots

An investigation is undertaken to find whether there is an association between the sex of a particular species of snake and its length in metres. The results are summarised in the following back-to-back stem plot of 27 males and 27 females.

| Female | Stem | Male |
|---|---|---|
| | 1 | 0 6 8 |
| | 2 | 15 5 7 |
| 9 3 2 | 3 | 0 0 2 3 3 |
| 86655431 10 | 4 | 0 1 1 3 4 8 9 |
| 8876633200 | 5 | 1 2 3 3 6 6 7 8 |
| 8 7 1 | 6 | |
| 2 | 7 | |

2|1|3 = 3.2 metres 3|1|2 = 3.2 metres

a For both males and females, calculate the

i range ii median iii IQR.

b Describe the shape of the

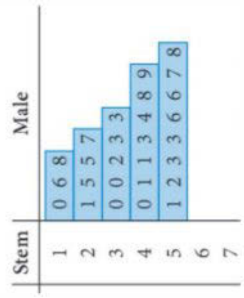
i male data ii female data.

c Does the back-to-back stem plot support the contention that the snake species' length is associated with its sex? Refer to the values of three appropriate statistics and the shape of the data in your response.

| Steps | Working |
|---|---|
| a i range = largest value - smallest value | male range = 5.8 - 1.0 = 4.8 m<br>female range = 7.2 - 3.2 = 4 m |
| ii If the number of data values is odd, find the middle data value. If the number of data values is even, find the average of the two middle data values. | male median = 4.1 m<br>female median = 5 m |
| iii IQR=$Q_3$-QJ | male IQR = 5.2 - 2.7 = 2.5 m<br>female IQR = 5.7 - 4.3 = 1.4 m |

b i To see the shape of the right leaf data, rotate the page 90° anticlockwise so that the stem forms the horizontal axis and picture it as a histogram.

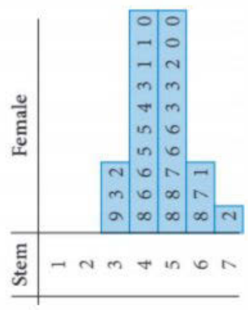
| Stem | Male |
|---|---|
| 1 | 0 6 8 |
| 2 | 1 5 5 7 |
| 3 | 0 0 2 3 3 |
| 4 | 0 1 1 3 4 8 9 |
| 5 | 1 2 3 3 6 6 7 8 |
| 6 | |
| 7 | |




The male data is negatively skewed.

ii To see the shape of the left leaf data, picture the data on the right and repeat the right leaf data steps.

| Stem | Female |
|---|---|
| 1 | |
| 2 | |
| 3 | 9 3 2 |
| 4 | 8 6 6 5 5 4 3 1 1 0 |
| 5 | 8 8 7 6 6 3 3 2 0 0 |
| 6 | 8 7 1 |
| 7 | 2 |




The female data is approximately symmetric.

c Compare the ranges, medians, IQRs, and the shapes. Do the two stem plots show differences?

Yes, the back-to-back stem plot supports the contention that the snake species' length is associated with its sex. The male range (4.8 m) is considerably more than the female range (4 m). The male median (4.1 m) is considerably less than the female median (5 m). The male IQR (2.3 m) is considerably more than the female IQR (1.3 m). The male distribution is negatively skewed whereas the female distribution is approximately symmetric.

2.3

WB
p. 24

## Parallel dot plots

Parallel dot plots can be used when we are dealing with *two or more* categories for the same numerical variable and want to be able to read the data values from the plot.

WB
p. 26

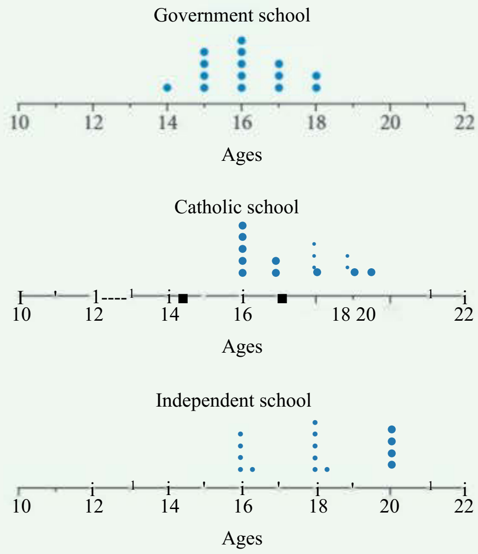
**WORKED EXAMPLE 6** Interpreting parallel dot plots

People who attended three different school systems in Victoria were asked to record the age they were when they started their first paid employment.

a Describe the shape of each distribution.

b Calculate the median, range and IQR for each school system.

c Do the dot plots support the contention that the age people are when they first start employment is associated with the school system they attended? Quote the IQR values in your response.




| Steps | Working |
| --- | --- |
| a Comment on symmetry and skewness for each distribution. | The Government school data is approximately symmetric.<br>The Catholic school data is positively skewed.<br>The Independent school data is symmetric. |
| b Calculate the median, range and IQR from each of the parallel dot plots. | Government school:<br>median = 16, range = 4, IQR =17-15 = 2<br>Catholic school:<br>median = 18, range = 4, IQR =19-16 = 3<br>Independent school:<br>median =18, range = 4, IQR = 20 - 16 = 4 |
| c Compare the IQRs to see if the dot plots show differences. | Yes, the dot plots support the contention that the age people are when they start paid employment is associated with the school system they attended. The Government school IQR (2 years) is noticeably less than Catholic school system IQR (3 years), which is noticeably less than the Independent school IQR (4 years). |

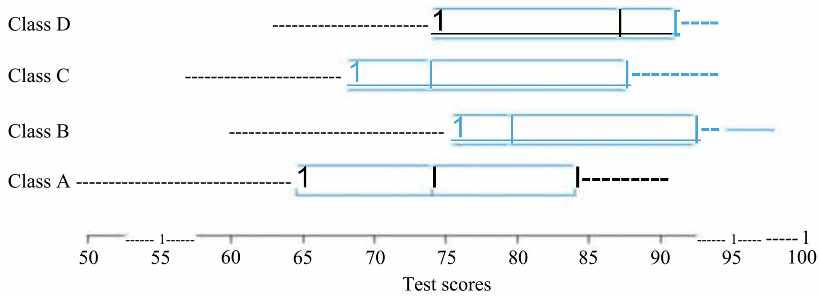

9780170448406

2.3

## Parallel boxplots

Parallel boxplots are a good choice when we are dealing with *two or more* categories for the same numerical variable, particularly when the data set is large. It is also easier to compare medians and quartiles from parallel boxplots than from back-to-back stem plots and parallel dot plots.

Here is an example based on the test results for four classes. Its relatively easy to find which class has the highest median, lowest Qj or highest maximum value etc.




**USING CAS 1** Constructing parallel boxplots

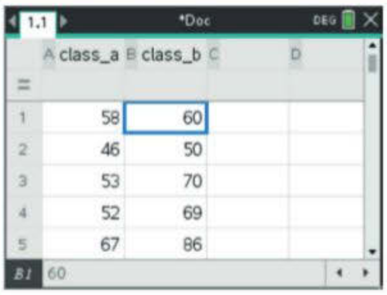
The test results of two Year 12 classes in General Mathematics are shown below.

**Class A:** 58, 46, 53, 52, 67, 36, 61, 49, 47, 59, 66, 53, 94, 69, 46, 44, 57

**Class B:** 60, 50, 70, 69, 86, 43, 60, 60, 44, 56, 49, 50, 56, 56, 42, 65, 47, 67, 25, 46

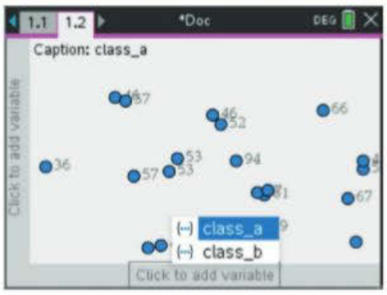
Construct parallel boxplots for the data.

**TI-Nspire**



1 Start a new document and add a **Lists & Spreadsheet** page.

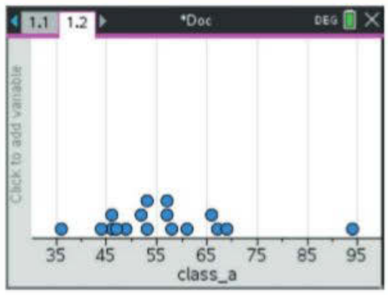
2 Label the columns and enter the data from the table as shown above.



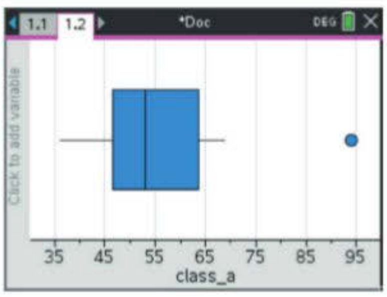
3 Insert a **Data & Statistics** page.

4 Click on the horizontal axis to add a variable.

5 Select **class_a**.

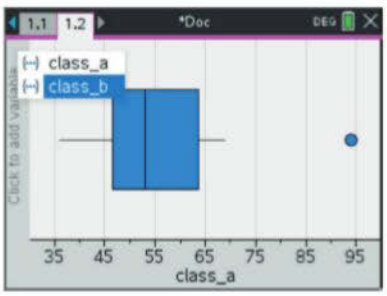


6 The **class_a** data will appear as a dot plot.



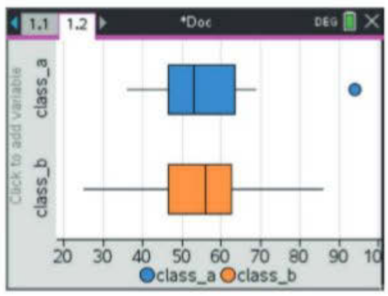
7 Press **menu > Plot Type > Box Plot.**

8 The **class_a** data will display as a boxplot.



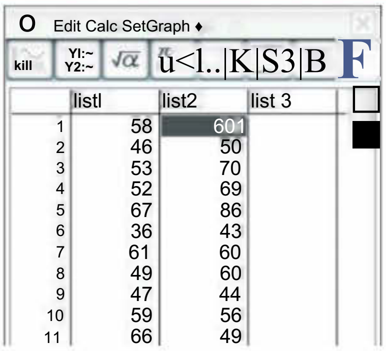
9 Press **menu > Plot Properties > Add X Variable.**

10 Select **class_b**.



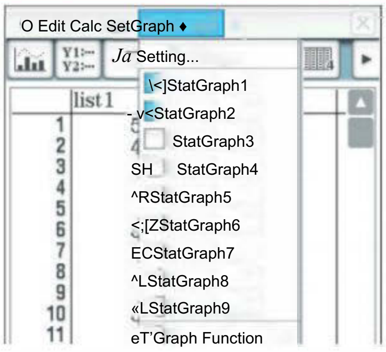
11 The **class_a** and **class_b** data will display as parallel boxplots.

## ClassPad



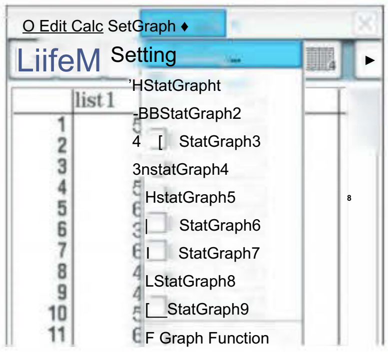

1 Tap Menu and open the Statistics application.

2 Clear all lists and enter the data from the table as shown.



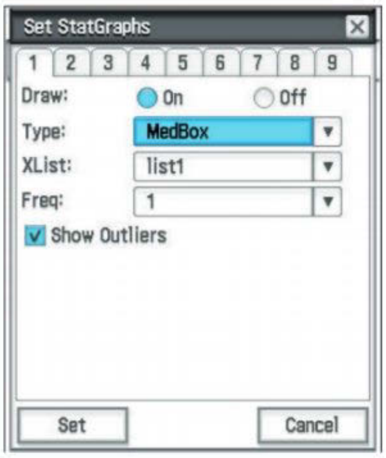

3 Tap SetGraph.

4 Select both StatGraphl and StatGraph2.




5 Tap SetGraph.

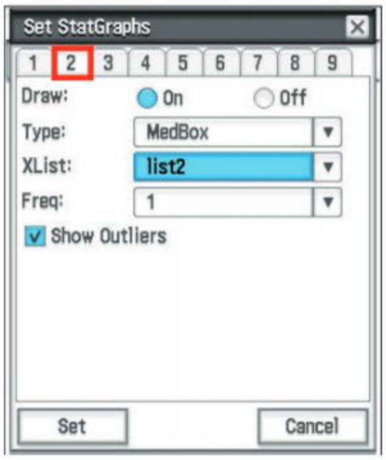
6 Tap Setting.




7 In the Type: field select MedBox.

8 Keep the XList setting as listl.

9 Select Show Outliers.




10 Tap tab 2 at the top of the screen.

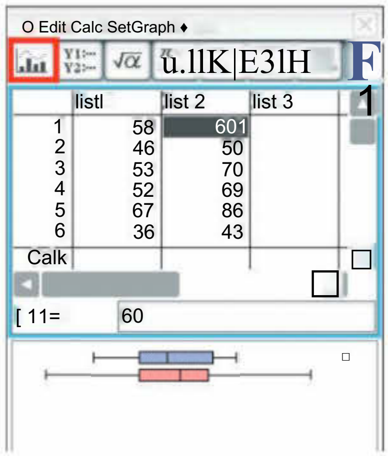
11 Ensure Draw: is set to On.

12 Set Type: to MedBox.

13 Set XList: to Iist2.

14 Select Show Outliers.

15 Tap Set.



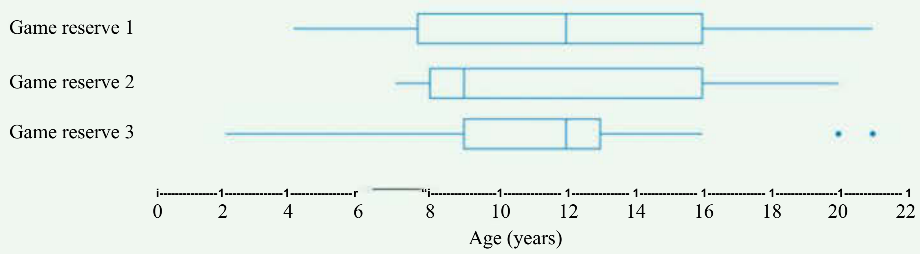

16 Tap Graph to display the parallel boxplots.



9780170448406

**WORKED EXAMPLE 7** Interpreting parallel boxplots

A study is undertaken to find whether there is an association between the lifespans of lions and the game reserve in which they live. The results are summarised in the following parallel boxplots.




a Describe the shape of each distribution.

b Calculate the median, range and IQR for each game reserve.

c Do the boxplots support the contention that there is an association between the lifespans of lions and the game reserve in which they live? Quote the shape and the range of the data in your response.

| Steps | Working |
|---|---|
| a Comment on symmetry and skewness for each distribution. | The game reserve 1 data is approximately symmetric.<br>The game reserve 2 data is positively skewed.<br>The game reserve 3 data is negatively skewed. |
| b Calculate the median, range and IQR from each of the parallel boxplots. | Game reserve 1: median =12, range =17, IQR = 16 - 8 = 8<br>Game reserve 2: median = 9, range =13, IQR = 16-8 = 8<br>Game reserve 3: median =12, range =19, IQR = 13-9 = 4 |
| c Compare the shapes and ranges to see if the boxplots show differences.<br><br>**iOl Exam hack**<br>Include the outliers in range calculations unless you are told to exclude them for some reason. | Yes, the boxplots support the contention that the lifespans of lions are associated with the game reserve in which they live. The shape of the three boxplots are different. Game reserve 1 data is approximately symmetric, game reserve 2 data is positively skewed, and game reserve 3 data is negatively skewed. The game reserve 2 range (13 years) is considerably lower than the game reserve 1 range (17 years), which is considerably lower again than the game reserve 3 range (19 years). |

2.3

WB
p. 27

## EXERCISE 2.3 Associations between numerical and categorical variables ANSWERS p. 697

### Recap

*Use the following information to answer the next two questions.*

Text messaging use (never, sometimes, every day) and the number of mobile phones in the household were recorded for a sample of 154 households. The results are shown in the table below.

| | Number of mobile phones in household | | | | |
|---|---|---|---|---|---|
| Text messaging use | 0 | 1 | 2 | 3 | Total |
| Never | 34 | 10 | 3 | 0 | 47 |
| Sometimes | 0 | 23 | 12 | 2 | 37 |
| Every day | 0 | 45 | 15 | 10 | 70 |
| Total | 34 | 78 | 30 | 12 | 154 |

▶ 1 ©VCAA 20051CQ4 80% Of the households with two mobile phones in the sample, the percentage that never used text messaging is

A 0% B 6% C 10% D 20% E 30%

2 ©VCAA | 20051CQ5 43% The mean number of mobile phones in these 154 households is closest to

A 1.13 B 1.45 C 1.50 D 1.54 E 2.00

## Mastery

3B WORKED EXAMPLE 5 An investigation is undertaken to find whether there is an association between the sex of a particular bird of prey and its lifespan. The results are summarised in the following back-to-back stem plot of 25 males and 25 females.

| Female | Stem | Male |
|---|---|---|
| | 1 | 02334569 |
| | 2 | 0 1 3 3 6 9 9 |
| 8 | 3 | 0 5 5 7 9 |
| 7 1 | 4 | 3 3 4 8 |
| 88655531 10 | 5 | |
| 8766332200 | 6 | 2 |
| 5 | 7 | |
| 1 | 8 | |

813 = 38 years 310 = 39 years

a For both males and females, calculate the

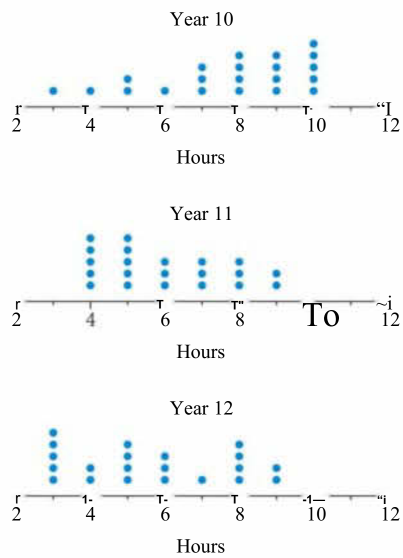
i range ii median iii IQR.

b Describe the shape of the

i male data ii female data.

c Does the back-to-back stem plot support the contention that the bird of prey lifespan is associated with the sex of the bird? Refer to the values of three appropriate statistics and the shape of the data in your response.

4 WORKED EXAMPLE 6 Students from Years 10,11 and 12 were asked to record the number of hours they spent binge-watching series during a school week. The results were presented as the following parallel dot plots,




a Describe the shape of each distribution,

b Calculate the median, range and IQR for each class,

c Do the dot plots support the contention that the number of hours spent binge-watching series during a school week is associated with the school year level? Quote the median values in your response.

5 Q using CAS 1"] The recorded car speeds (in km/h) from two speed cameras positioned on different roads are shown.

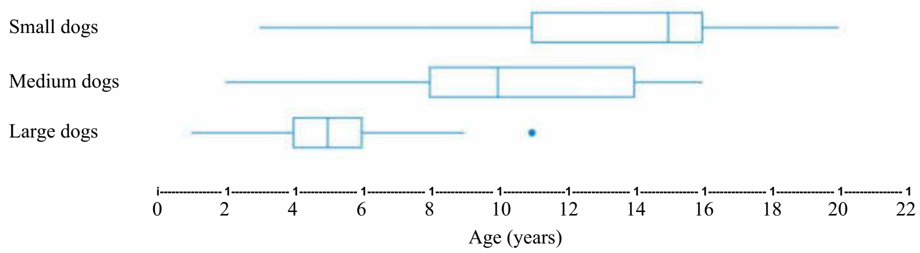
Camera 1: 78, 63, 75, 69, 71, 83, 80, 67, 74, 72, 73, 74, 90, 83, 65, 73, 69, 89, 76,102, 83, 78, 69, 71

Camera 2: 112,139,120,116, 116,136,140, 123,135,131, 120,117,138,131,127,119,125,130, 130, 134,123, 148, 169, 130

Construct parallel boxplots for the data.

▶

▶ 60 WORKED EXAMPLE 7 | A study is undertaken to find whether there is an association between the lifespans of dogs and their sizes (small, medium or large). The results are summarised in the following parallel boxplots.




a Describe the shape of each distribution.

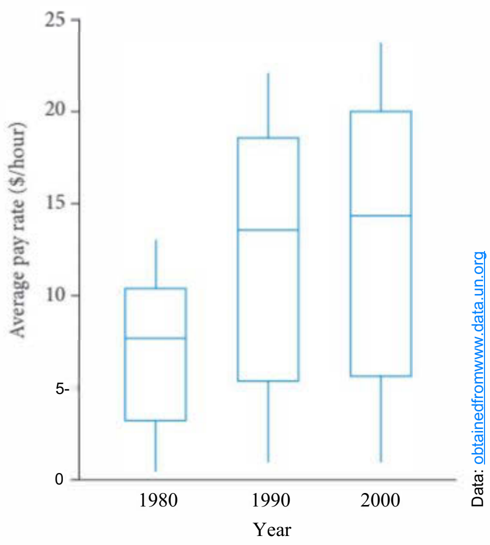
b Calculate the median, range and IQR for each size.

c Do the boxplots support the contention that there is an association between lifespans and the sizes of dogs (small, medium or large)? Quote the shape and the range of the data in your response.

## Exam practice

80-100% 60-79% 0-59%

7 ©VCAA | 2011 1CQ5 | 54% The boxplots display the distribution of average pay rates, in dollars per hour, earned by workers in 35 countries for the years 1980,1990 and 2000. Based on the information contained in the boxplots, which one of the following statements is not true?




A In 1980, over 50% of the countries had an average pay rate less than $8.00 per hour.

B In 1990, over 75% of the countries had an average pay rate greater than $5.00 per hour.

C In 1990, the average pay rate in the top 50% of the countries was higher than the average pay rate for any of the countries in 1980.

D In 1990, over 50% of the countries had an average pay rate less than the median average pay rate in 2000.

E In 2000, over 75% of the countries had an average pay rate greater than the median average pay rate in 1980.

8 ©VCAA | 2016 1CQ8 | 45% Parallel boxplots would be an appropriate graphical tool to investigate the association between the monthly median rainfall, in millimetres, and the

A monthly median wind speed, in kilometres per hour.

B monthly median temperature, in degrees Celsius.

C month of the year (January, February, March etc.).

D monthly sunshine time, in hours.

E annual rainfall, in millimetres. ▶

9 ^^^J~2oi~41CQ8 28% A single back-to-back stem plot would be an appropriate graphical tool to investigate the association between a cars speed, in kilometres per hour, and the

A drivers age, in years.

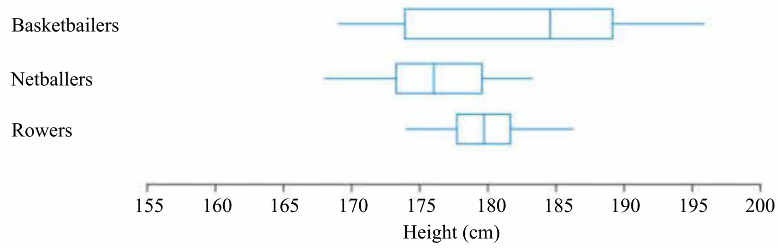
B cars colour (white, red, grey, other).

C cars fuel consumption, in kilometres per litre.

D average distance travelled, in kilometres.

E drivers sex (female, male).

10 ©VCAA 2018N 1CQ6 The parallel boxplots display the distribution of height for three groups of athletes: rowers, netballers and basketbailers.




Which one of the following statements is not true?

A The shortest athlete is a netballer.

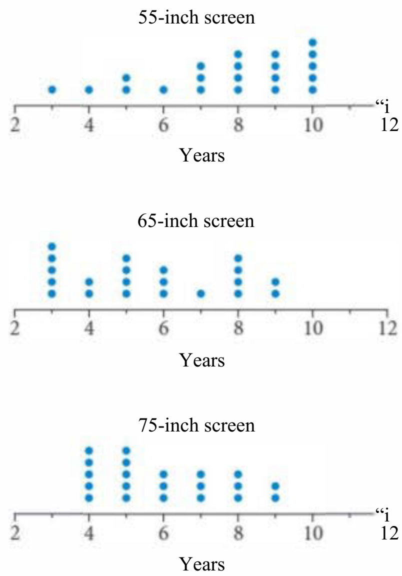
B The rowers have the least variable height.

C More than 25% of the netballers are shorter than all rowers.

D The basketballers are the tallest athletes in terms of median height.

E More than 50% of the basketballers are taller than any of the rowers or netballers.

11 (8 marks) A brand of television set comes in three screen sizes: 55-inch, 65-inch and 75-inch. A long-term study was performed to see whether there is an association between the number of years before the television needs to be replaced and its screen size. The results are shown in the parallel dot plots.




a Do the shapes of the distributions suggest there is an association?

Include a description of the shapes in your response. 4 marks

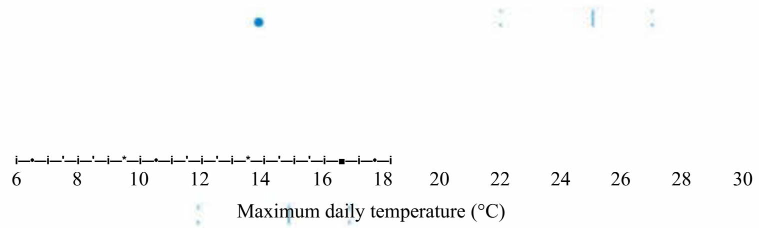
b Do the medians suggest there is an association?

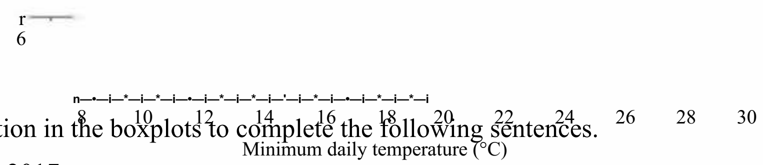
Give the median values in your response. 4 marks



9780170448406

▶ 12 ©VCAA | 2019 2CQ2 I (4 marks) The parallel boxplots below show the *maximum daily temperature* and *minimum daily temperature,* in degrees Celsius, for 30 days in November 2017.




Data: Australian Government, Bureau of Meteorology, www.bom.gov.au

a Use the information in the boxplots to complete the following sentences.

For November 2017

i 89% the interquartile range for the *minimum daily temperature* was°C — 1 mark

ii 87% the median value for *maximum daily temperature* was°C higher than the median value for *minimum daily temperature* — 1 mark

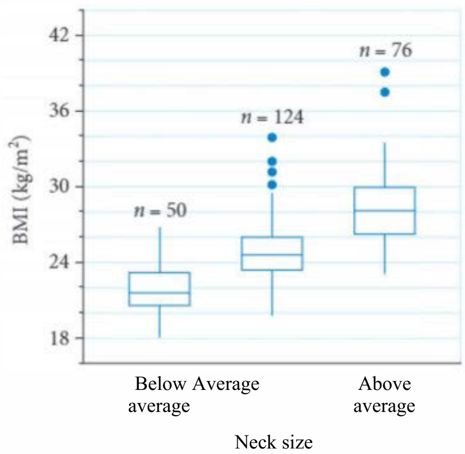
iii 79% the number of days on which the *maximum daily temperature* was less than the median value for *minimum daily temperature* was — 1 mark

b 58% The *temperature difference* between the *minimum daily temperature* and the *maximum daily temperature* in November 2017 at this location is approximately normally distributed with a mean of 9.4°C and a standard deviation of 3.2°C.

Determine the number of days in November 2017 for which this *temperature difference* is expected to be greater than 9.4°C. — 1 mark

13 ©VCAA | 2020 2CQ3 J (5 marks) In a study of the association between BMI and *neck size,* 250 men were grouped by *neck size* (below average, average and above average) and their BMI recorded. Five-number summaries describing the distribution of BMI for each group are displayed in the table along with the group size. The associated boxplots are shown below the table.

| | | BMI ($kg/m^2$) | | | | |
|---|---|---|---|---|---|---|
| Neck size | Group size | Min. | $Q_1$ | Median | $Q_3$ | Max. |
| Below average | 50 | 18.1 | 20.6 | 21.6 | 23.2 | 26.8 |
| Average | 124 | 19.8 | 23.4 | 24.6 | 26.0 | 33.9 |
| Above average | 76 | 23.1 | 26.25 | 28.1 | 29.95 | 39.1 |




Data: RW Johnson, 'Fitting percentage of body fat to simple body measurements', Journal of Statistics Education, 4:1,1996, https://doi.org/10.1080/10691898.1996.11910505

**Exam hack**

When you're asked for one statistic, don't refer to more than one. If your second statistic is incorrect, you will lose marks. Always use an exact value if it's available rather than reading the value from a graph.

a 93% What percentage of these 250 men are classified as having a below average neck size? 1 mark

b 89% What is the interquartile range (IQR) of BMI for the men with an average neck size? 1 mark

c 22% People with a BMI of 30 or more are classified as being obese. Using this criterion, how many of these 250 men would be classified as obese? Assume that the BMI values were all rounded to one decimal place. 1 mark

d 49% Do the boxplots support the contention that BMI is associated with *neck size?* Refer to the values of an appropriate statistic in your response. 2 marks

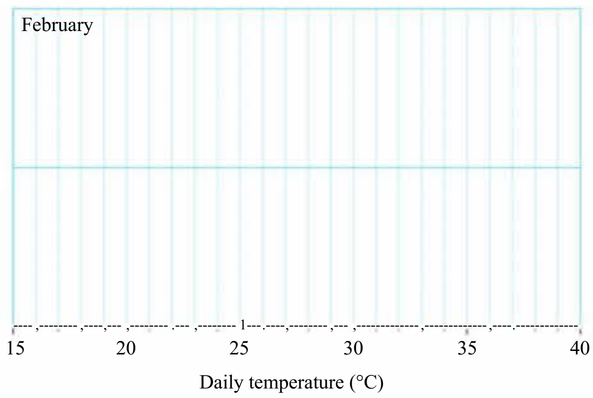
14 ©VCAA 2016 2CQ2 (5 marks) A weather station records daily maximum temperatures.

a The five-number summary for the distribution of maximum temperatures for the month of February is displayed in the following table.

There are no outliers in this distribution.

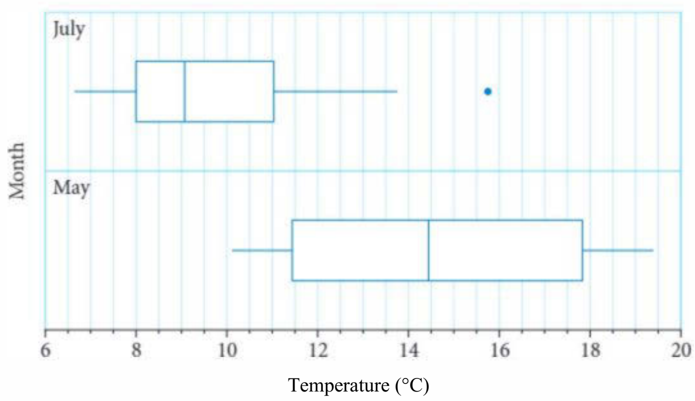
| | Temperature (°C) |
|---|---|
| Minimum | 16 |
| Qi | 21 |
| Median | 25 |
| $Q_3$ | 31 |
| Maximum | 38 |

j 96% Use the five-number summary to construct a boxplot on a grid like the one shown. 1 mark




jj 71% What percentage of days had a maximum temperature of 21 °C, or greater, in this particular February? 1 mark

b The boxplots below display the distribution of maximum daily *temperature* for the months of May and July.




**Exam hack**

You can only estimate the mean from a boxplot if the distribution is clearly symmetric.

i 56% Describe the shapes of the distributions of daily *temperature* (including outliers) for July and for May. 1 mark

jj 60% Determine the value of the upper fence for the July boxplot. 1 mark

jjj 30% Using the information from the boxplots, explain why the maximum daily temperature is associated with the *month* of the year. Quote the values of appropriate statistics in your response. 1 mark

▶ 15 © VCAA 2017 2CQ2 , (8 marks) The back-to-back stem plot displays the *wingspan,* in millimetres, of 32 moths and their *place of capture* (forest or grassland).

Wingspan (mm)

| Forest (*n* = 13) | Stem | Grassland (n = 13) |
|---|---|---|
| 6 | 1 | 8 |
| 2 1 1 0 0 0 0 | 2 | 2 2 4 4 |
| 7 | 2 | 5 5 9 |
| 4 0 | 3 | 0 0 1 2 3 4 |
| 5 | 3 | 6 8 |
| | 4 | 0 3 |
| | 4 | 5 |
| 2 | 5 | |

Key: 118 =18

a 92% Which variable, *wingspan* or *place of capture,* is a categorical variable? 1 mark

b 87% Write down the modal wingspan, in millimetres, of the moths captured in the forest. 1 mark

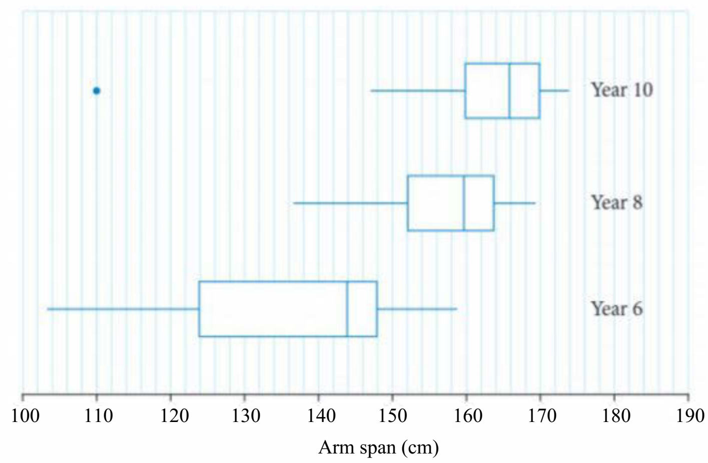
c 87% Use the information in the back-to-back stem plot to complete the following table. 2 marks

| | Wingspan (mm) | | | | |
|---|---|---|---|---|---|
| Place of capture | Minimum | $Q_1$ | Median (M) | $Q_3$ | Maximum |
| Forest | | 20 | 21 | 32 | 52 |
| Grassland | 18 | 24 | 30 | | 45 |

d 73% Show that the moth captured in the forest that had a *wingspan* of 52 mm is an outlier. 2 marks

e 51% The back-to-back stem plot suggests that *wingspan* is associated with *place of capture.* Explain why, quoting the values of an appropriate statistic. 2 marks

16 © VCAA 2008 2CQ3 51% (4 marks) The arm spans (in cm) of Years 6, 8 and 10 girls were recorded in a survey. The results are summarised in the three parallel boxplots displayed.




a Copy and complete the following sentence.

The middle 50% of Year 6 students have an arm span betweenandcm. 1 mark

b The three parallel boxplots suggest that arm span and year level are associated. Explain why. 1 mark

c The arm span of 110 cm of a Year 10 girl is shown as an outlier on the boxplot. This value is an error. Her real arm span is 140 cm. If the error is corrected, would this girl's arm span still show as an outlier on the boxplot? Give reasons for your answer, showing an appropriate calculation. 2 marks

Video playlist
Associations between two numerical variables

# l@ Associations between
# 1 two numerical variables

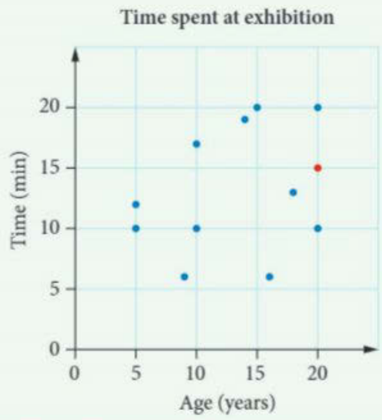
## Scatterplots

A scatterplot is used to display the results when we investigate the association between two numerical variables. A scatterplot is constructed by plotting points onto a Cartesian plane where the horizontal or x-axis is used for the explanatory variable and the vertical or 7-axis is used for the response variable.

WB
p. 28

**WORKED EXAMPLE 8** Interpreting scatterplots

A study was conducted on the length of time (in minutes) people of various ages spent at an exhibition at the Royal Melbourne Show and a scatterplot was plotted of the data.




| Steps | Working |
|---|---|
| **a** What is the explanatory variable? | |
| The explanatory variable appears on the $x$-axis. | *age* (years) |
| **b** What is the response variable? | |
| The response variable appears on the $y$-axis. | *time* (min) |
| **c** How many people were in the study? | |
| Count the number of dots. | 12 people |
| **d** What does the red dot represent? | |
| Read from both axes. | A 20-year-old who was at the exhibition for 15 minutes. |
| **e** How many teenagers (of ages 13 to 19) were in the study? | |
| Count the number of points for the ages 13 to 19. | 4 |
| **f** What was the longest time spent at the exhibition? | |
| Read from the $y$-axis. | 20 minutes |

**Exam hack**

To remember which axis is used for which variable, use the fact that ‘explanatory’ starts with ‘ex’ for x-axis.

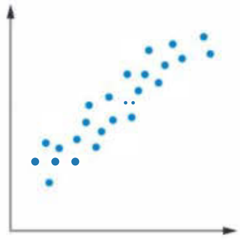

9780170448406

## Scatterplots and association

There are three ways that scatterplots can be used to describe the association between two numerical variables.

1 Direction

Positive association



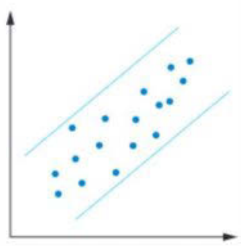
The direction of a positive association *rises* from left to right. This indicates that the *y* (response) variable tends to *increase* as the *x* (explanatory) variable increases.

Negative association

The direction of a negative association *falls* from left to right. This indicates that the *y* (response) variable tends to *decrease* as the *x* (explanatory) variable increases.

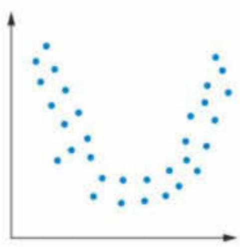
2 Form

Linear



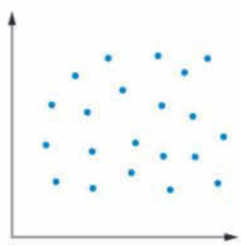
Data in general follows a straight line pattern.

Non-linear



Data does not occur in a straight line pattern but does follow a curved pattern.

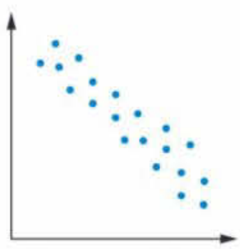
No association



Data is randomly scattered and shows no pattern at all.

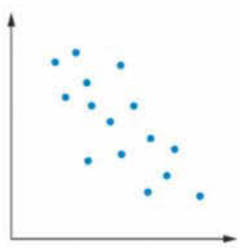
3 Strength

Strong association



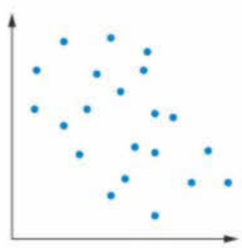
Data points are all reasonably close together.

Moderate association



Data points are more spread out than a strong association.

Weak association



Data points are widely spread out.

2.4

Worksheets
ScatterpLots and associations
A page of scatterplots
Height vs shoe size
Body measurements

WB

p. 29

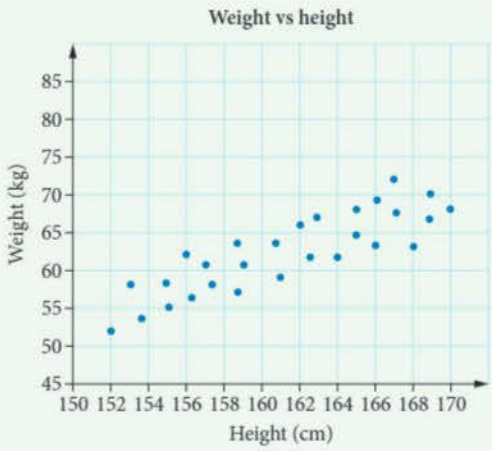
## WORKED EXAMPLE 9 Describing association using scatterplots

For each of the following scatterplots

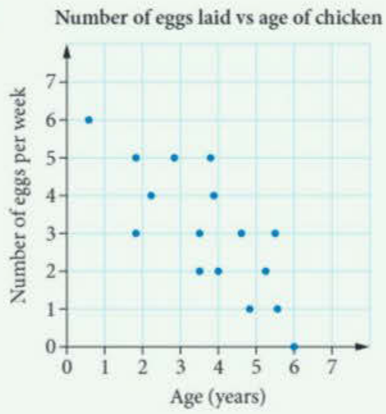
i describe the association between the two variables in terms of direction, form and strength

ii explain what this means in terms of the variables.

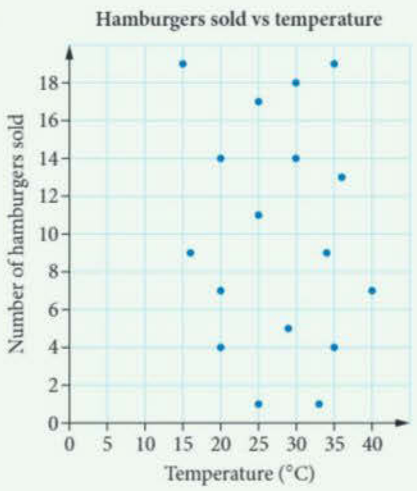
a




b




c




| **Steps** | **Working** |
|---|---|
| **a** i Is the data sloping up or down?<br>Does the data follow a linear pattern?<br>How spread out are the data points? | positive, linear and strong |
| ii Refer to the variables. | Weight increases as height increases. |
| **b** i Is the data sloping up or down?<br>Does the data follow a linear pattern?<br>How spread out are the data points? | negative, linear and moderate |
| ii Refer to the variables. | The number of eggs a chicken lays tends to decrease as the chicken gets older. |
| **c** i Is the data sloping up or down?<br>Does the data follow a linear pattern?<br>How spread out are the data points? | no association |
| ii Refer to the variables. | There appears to be no association between the temperature and the number of hamburgers sold. |



9780170448406

## USING CAS 2 Constructing scatterplots

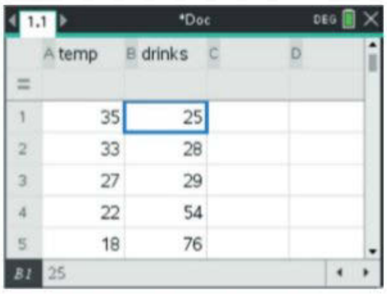
The number of hot drinks sold at a cafe each day and the day's maximum temperature were recorded over a period of two weeks.

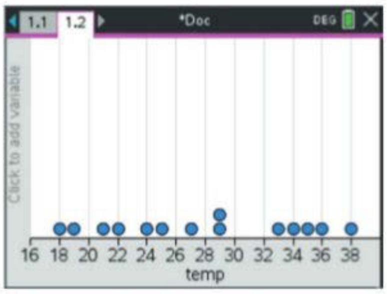
| Temperature | 35 | 33 | 27 | 22 | 18 | 29 | 38 | 36 | 24 | 25 | 29 | 34 | 21 | 19 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| No. of drinks sold | 25 | 28 | 29 | 54 | 76 | 48 | 18 | 39 | 42 | 61 | 36 | 49 | 68 | 53 |

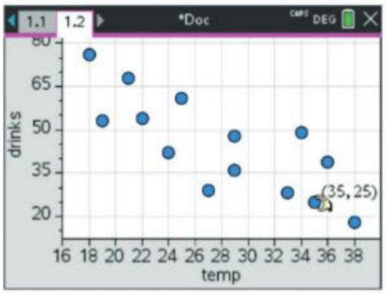
Construct a scatterplot for this information and describe the association shown in the scatterplot.

Decide first which is the explanatory variable and which is the response variable. Temperature affects the number of drinks sold, so *temperature* is the explanatory variable (x) and the *number of drinks sold* is the response variable (y).

### TI-Nspire




1 Start a new document and add a Lists & Spreadsheet page.

2 Label the columns and enter the data from the table as shown above.

3 Insert a Data & Statistics page.

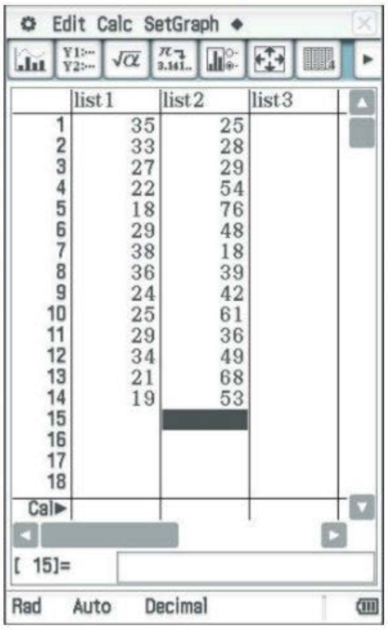
4 Click on the horizontal axis and select temp.

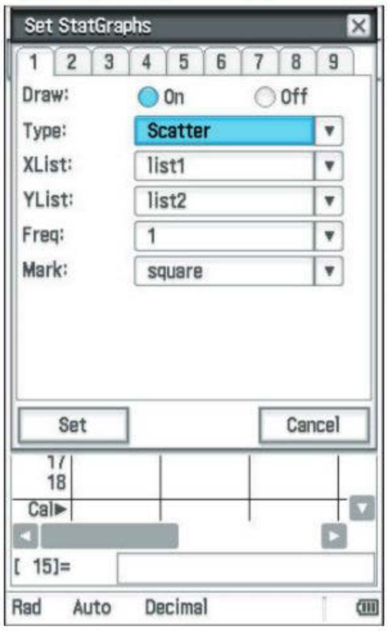
5 Click on the vertical axis and select drinks.

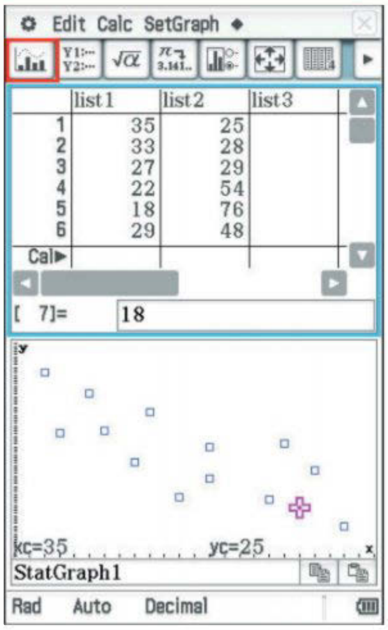
6 The scatterplot will be displayed.

7 Click on any data point to display the pair of values.

### ClassPad




1 Tap Menu and open the Statistics application.

2 Clear all lists and enter the data from the table as shown.

3 Tap SetGraph > Settings to confirm the default values shown above.

4 Tap Set.

5 Tap Graph.

6 The scatterplot will be displayed in the lower window.

7 Tap Analysis > Trace then use the arrow keys to display the data point values.

The association between the day's maximum temperature and the number of hot drinks sold at a cafe per day can be described as negative, linear and moderate.

## Graphs showing association between two variables

| Explanatory variable | Response variable | Graph | Advantage |
|---|---|---|---|
| categorical (2 or more categories) | categorical (2 or more categories) | parallel percentage segmented bar charts | |
| categorical (2 categories only) | numerical | back-to-back stem plot | Individual data values are shown. |
| categorical (2 or more categories) | numerical | parallel dot plots | Data values can be read from the graph. |
| categorical (2 or more categories) | numerical | parallel boxplots | It is easier to compare medians and quartiles. |
| numerical | numerical | scatterplot | |

WB
p. 30

**WORKED EXAMPLE 10** Choosing a graph for two variables

For the following

i determine which graph(s) can be used to display the association between the two variables

ii give a reason for your answer.

| Steps | Working |
|---|---|
| a the AFL team a person supports and their age | |
| i Determine the graph(s) to display the association | parallel dot plots and parallel boxplots |
| ii Refer to the types of variables involved. | *AFL team supported* is categorical with more than two categories and *age* is numerical. |
| b the number of times per year a person attends an AFL match and their age | |
| i Determine the graph(s) to display the association | scatterplot |
| ii Refer to the types of variables involved. | Both *number of yearly AFL match attendances* and *age* are numerical. |
| c the AFL team a person supports and the persons postcode | |
| i Determine the graph(s) to display the association | parallel percentage segmented bar charts |
| ii Refer to the types of variables involved. | Both *AFL team supported* and *postcode* are categorical. |
| d the gender and the average number of minutes a person spend online per day, where its important to see the data values. | |
| i Determine the graph(s) to display the association | back-to-back stem plot |
| ii Refer to the types of variables involved. | *Gender* (male, female) is categorical with two categories and *age* is numerical, and seeing the data values is important. |



9780170448406

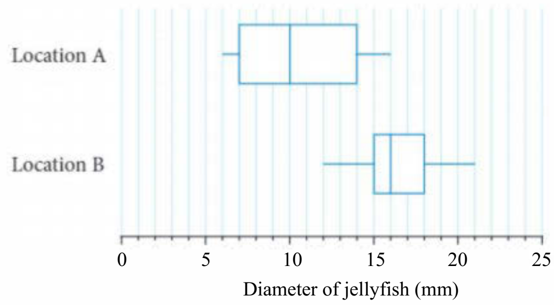
## EXERCISE 2.4 Associations between two numerical variables

ANSWERS p. 698

### Recap

*Use the following information to answer the next two questions.*

Samples of jellyfish were selected from two different locations, A and B. The diameter (in mm) of each jellyfish was recorded and the resulting data is summarised in the boxplots shown.




1 ©VCAA 2007 1CQ5 66% The percentage of jellyfish taken from location A with a diameter greater than 14 mm is closest to

A 2% B 5% C 25% D 50% E 75%

2 ©VCAA 2007 1CQ6 75% From the boxplots, it can be concluded that the diameters of the jellyfish taken from location A are generally

A similar to the diameters of the jellyfish taken from location B.

B less than the diameters of the jellyfish taken from location B and less variable.

C less than the diameters of the jellyfish taken from location B and more variable.

D greater than the diameters of the jellyfish taken from location B and less variable.

E greater than the diameters of the jellyfish taken from location B and more variable.

### Mastery

3 WORKED EXAMPLE 8 A study was made on how long people of various ages spent at a café and a scatterplot was plotted of the data.

a What is the explanatory variable?

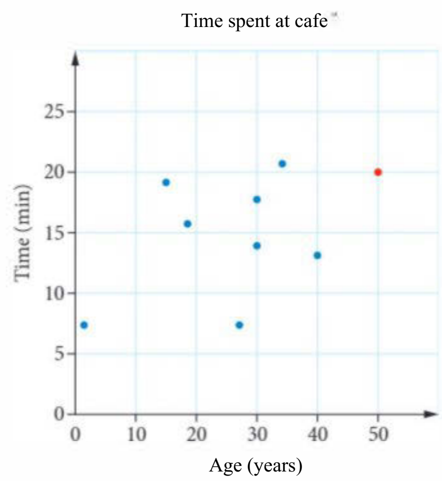
b What is the response variable?

c How many people were in the study?

d What does the red dot represent?

e How many people in their 30s were in the study?

f Why could we reasonably conclude that the two people who stayed for the shortest time were together?

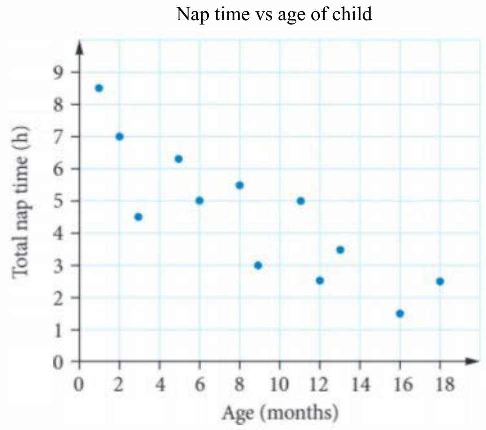


4 WORKED EXAMPLE 9 For each of the following scatterplots

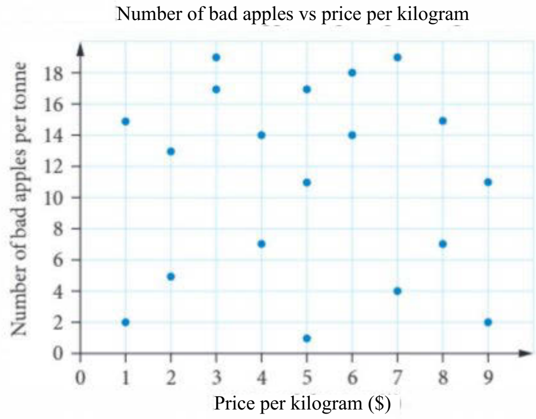
i describe the association between the two variables in terms of direction, form and strength

ii explain what this means in terms of the variables.

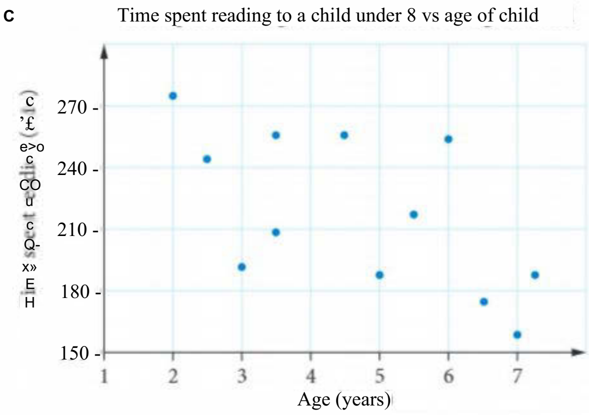
a



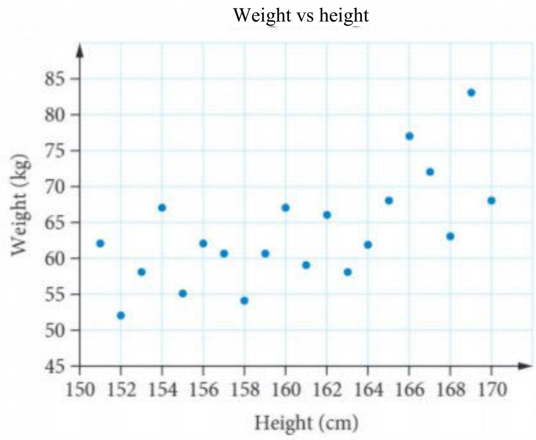

b




c




d




5 For each pair of variables, state whether they would generally have a positive association, a negative association or no association.

a *height* and *shoe size* of a person

b *salary level* and *lung capacity* of an employee

c *price* of a brand of car and the *number of cars* of that brand sold

d *amount of time spent studying* for an exam and the *exam score*

e *number of cigarettes smoked* and *incidence of lung cancer*

f *number of police* on roads and *number of speeding cars*

6 using CAS 2 J For each of the following, use CAS to construct a scatterplot for the information and describe the association shown in the scatterplot.

a The hearing sensitivities (highest audible frequency measured in kilohertz) of different people from the same extended family were tested and the results are shown below.

| Age | 5 | 12 | 15 | 21 | 46 | 50 | 62 | 70 | 75 |
|---|---|---|---|---|---|---|---|---|---|
| Frequency (kHz) | 30 | 25 | 23 | 22 | 19 | 18 | 16 | 17 | 15 |

b People at a convenience store were asked how many chocolate bars they ate on average in a month.

| Age (years) | 8 | 16 | 23 | 87 | 46 | 38 | 24 | 76 | 11 | 33 | 19 | 26 | 31 | 59 | 65 | 50 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of chocolate bars | 4 | 13 | 15 | 4 | 13 | 11 | 11 | 6 | 12 | 9 | 16 | 8 | 12 | 6 | 3 | 8 |


9780170448406

▶ 7H WORKED EXAMPLE 10 I For the following

i determine which graph(s) can be used to display the association between the two variables

ii give a reason for your answer.

a the number of people at a concert and the amount of money spent on merchandise

b the make of car a person drives and the type of pet they have

c the make of car a person drives and the amount of money they earn

d whether a child is a boy or a girl and the number of pets they have

e the types physical activities a person does and the amount of time spent watching sport on television, where a quick comparison of medians is needed.

8 State whether each the following is true or false. If your answer is false, give a reason.

a A scatterplot could be used to present results from a survey where

i the explanatory variable is numerical and the response variable is categorical.

ii the explanatory variable is the *number of times someone has been badly sunburnt* and the response variable is the *number of skin cancer occurrences.*

iii the explanatory variable is the *average hours in week a VCE student spends working at a paying job* and the response variable is the *students ATAR.*

iv the explanatory variable is a persons *annual income* and the response variable is the persons *make of car.*

b It is possible to present a data set as both

i a scatterplot and parallel percentage segmented bar charts.

ii parallel dot plots and parallel boxplots.

iii a back-to-back stem plot and three parallel boxplots.

## Exam practice

80-100% 60–79% 0–59%

9 ©VCAA 20161CQ12 J 67% There is a strong positive association between a country's Human Development Index and its carbon dioxide emissions. From this information, it can be concluded that

A increasing a country's carbon dioxide emissions will increase the Human Development Index of the country.

B decreasing a country's carbon dioxide emissions will increase the Human Development Index of the country.

C this association must be a chance occurrence and can be safely ignored.

D countries that have higher human development indices tend to have higher levels of carbon dioxide emissions.

E countries that have higher human development indices tend to have lower levels of carbon dioxide emissions.

10 ©VCAA 2007 1CQ10 37% The association between the variables *size of car* (1 = small, 2 = medium, 3 = large) and *salary level* (1 = low, 2 = medium, 3 = high) is best displayed using

A a scatterplot. B a histogram.

C parallel boxplots. D a back-to-back stem plot.

E a percentage segmented bar chart.

▶

11 ©VCAA 2016S1CQ12 A large study of secondary-school male students shows that there is a negative association between the time spent playing sport each week and the time spent playing computer games. From this information, it can be concluded that

A male students who spend a lot of time playing computer games do not play sport.

B encouraging male students to spend less time playing sport will increase the time they spend playing computer games.

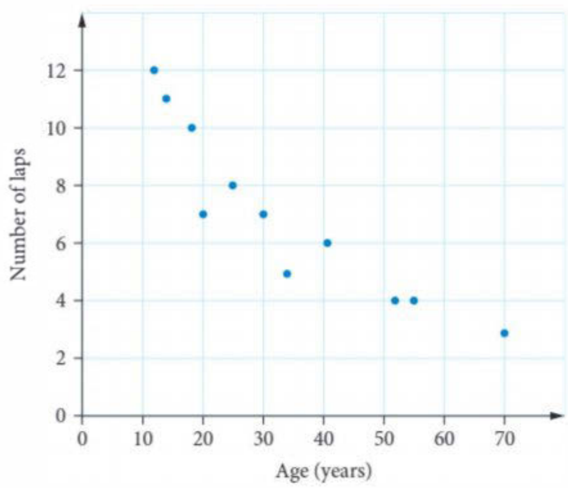
C encouraging male students to spend more time playing sport will reduce the time they spend playing computer games.

D male students who tend to spend more time playing sport tend to spend less time playing computer games.

E male students who tend to spend more time playing sport tend to spend more time playing computer games.

12 Eleven people of different ages ran around an oval for 30 minutes. The number of laps they ran were recorded and the results are displayed in the scatterplot.

Number of laps around oval vs age




Which one of the following statements is false?

A The youngest person in the group ran the most laps.

B *Age* is the explanatory variable and *number of laps* is the response variable.

C The association between *age* and *number of laps* can be described as strong.

D The association between *age* and *number of laps* can be described as positive.

E The association between *age* and *number of laps* can be described as non-linear.

13 The ages of both parents for students in a Year 12 class were recorded. Which of the following is true?

A The association between the two variables can be displayed using a scatterplot.

B The association between the two variables can be displayed using a histogram.

C The association between the two variables can be displayed using a percentage segmented barchart.

D This involves looking at the association between a numerical and a categorical variable.

E This involves looking at the association between two categorical variables.



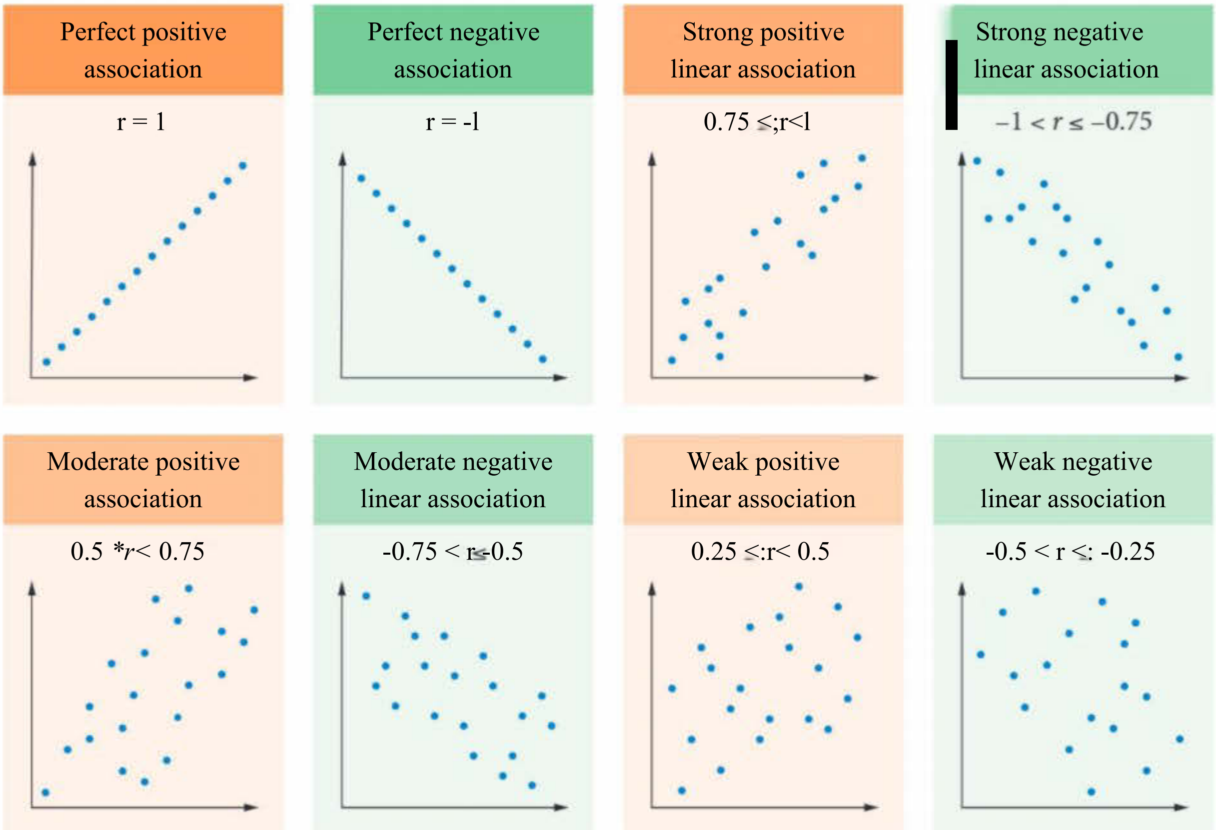
9780170448406

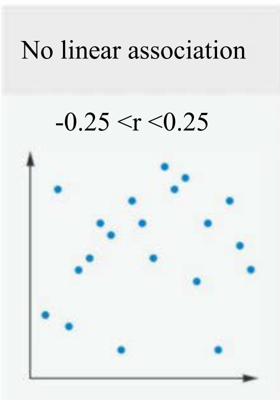
2.5

# @ Correlation and causation

## The Pearson correlation coefficient

A scatterplot can only give us an indication of the association between two variables. To get an accurate measure, we calculate the Pearson correlation coefficient, r, (also simply called the correlation coefficient), which is a number between -1 to 1 that measures the strength and direction of *linear* associations. We will be using CAS to calculate r.




**Pearson correlation coefficient assumptions**

If we are using the Pearson correlation coefficient to measure the association of two variables, we are making three assumptions:

1 The variables are both numerical.
2 The association is linear.
3 There are no outliers.

Video playlist Correlation and causation

Worksheet Relationships between variables

WB
p. 31

## WORKED EXAMPLE 11 Interpreting correlation coefficient values

Interpret the correlation coefficient values and write a sentence describing the association for each of the studies described below, beginning with 'The data suggests...'.

| Steps | Working |
|---|---|
| a A study is investigating whether there is an association between *temperature* and the *number of heaters sold.* The correlation coefficient was found to be $r$ = -0.923. | |
| What level of strength of association does $r$ indicate? Is the association positive or negative? | The data suggests there is a strong negative linear association between *temperature* and *number of heaters sold.* |
| b A study is investigating whether there is an association between the *number of steps* a person takes in a day and their *weight.* The correlation coefficient was found to be r = 0.283. | |
| What level of strength of association does $r$ indicate? Is the association positive or negative? | The data suggests there is a weak positive linear association between *number of steps* and *weight.* |
| c A study is investigating whether there is an association between a persons' *height* and *salary.* The correlation coefficient was found to be r = -0.031. | |
| What level of strength of association does r indicate? Is the association positive or negative? | The data suggests there is no association between *height* and *salary.* |

## USING CAS 3 Calculating the Pearson correlation coefficient

Calculate the correlation coefficient, correct to two decimal places, for the data in the table showing the number of umbrellas sold at a shopping centre and the amount of rain in the area recorded over 8 weeks.

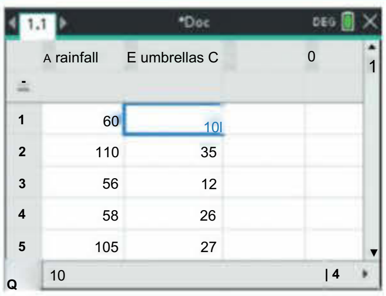
a What are the three assumptions we make when calculating the correlation coefficient?

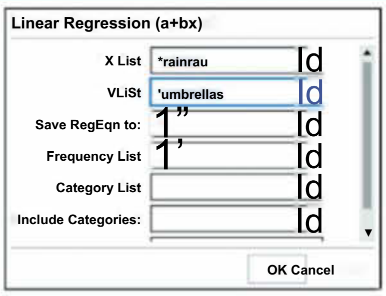
b Discuss if there is an association between the number of umbrellas sold and the amount of rainfall.

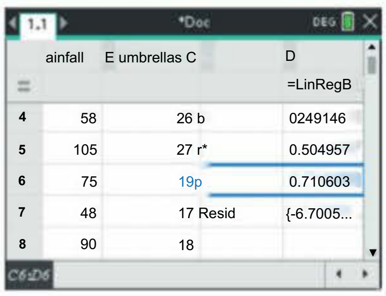
| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Umbrellas sold | 10 | 35 | 12 | 26 | 27 | 19 | 17 | 18 |
| Rainfall (mm) | 60 | 110 | 56 | 58 | 105 | 75 | 48 | 90 |

Decide first which variable is the explanatory (x) and which is the response (y). Rainfall affects the number of umbrellas sold, so *rainfall* is the explanatory variable and *umbrellas sold* is the response variable.

### TI-Nspire




1 Start a new document and add a Lists & Spreadsheet page.

2 Label the columns and enter the data from the table as shown above.

3 Press menu > Statistics > Stat Calculations > Linear Regression (a+bx).

4 In the X List: field, select rainfall.

5 In the Y List: field, select umbrellas.

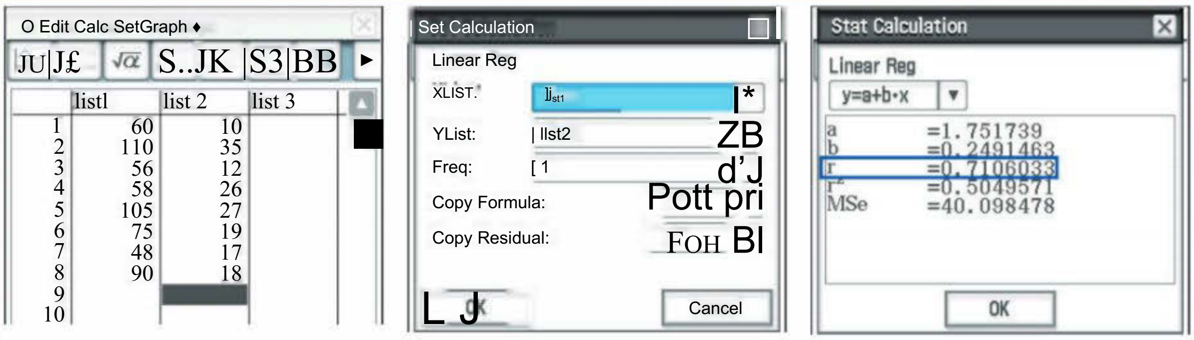
6 Select OK.

7 The linear regression labels and values will be displayed in columns C and D.

8 Scroll down to view the r value, which is 0.71.


9780170448406

**ClassPad**




1 Tap Menu and open the Statistics application.

2 Clear all lists and enter the data from the table as shown.

3 Tap Calc > Regression > Linear Reg.

4 Leave the XList: and YList: default settings of list1 and list2 as shown above.

5 Tap OK.

6 Tap on the Linear Reg dropdown menu to ensure the field is set to y=a+bx by tapping on the dropdown menu.

7 The r value is 0.71.

a The three assumptions are that both variables are numerical, the association is linear and there are no outliers.

b An $r$ value of 0.71 indicates that there is a moderate, positive linear association between rainfall and the numbers of umbrellas sold.

**Exam hack**

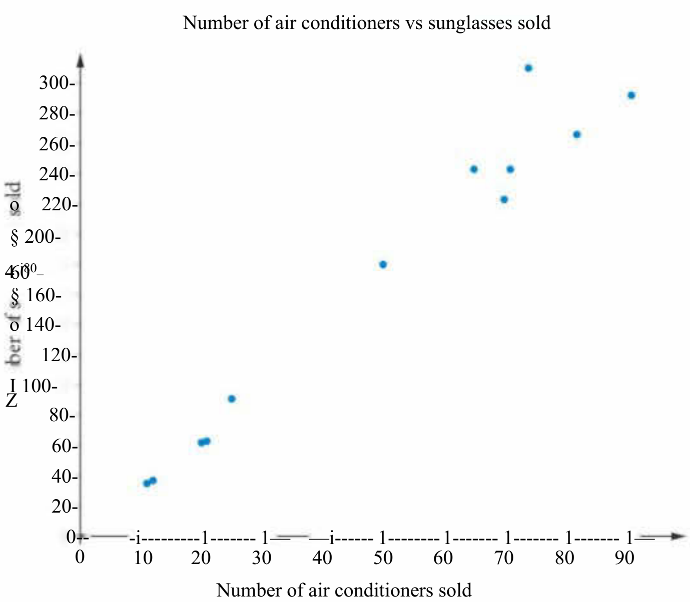
The calculation of $r$ gives the same value regardless of which variable you take to be the explanatory variable and which one to be the response variable.

## Cause and effect

If two variables have a correlation or association, it doesn't necessarily mean that one *causes* the other. For example, a store recorded the following monthly sales of air conditioners and sunglasses over a year:

| | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of air conditioners sold | 74 | 65 | 50 | 25 | 21 | 20 | 11 | 12 | 70 | 71 | 82 | 91 |
| Number of sunglasses sold | 310 | 243 | 180 | 91 | 63 | 62 | 35 | 37 | 223 | 243 | 266 | 292 |

These sales numbers give a correlation of $r = 0.98$ and the following scatterplot.



The correlation coefficient and scatterplot both indicate that there is an extremely strong positive correlation between the variables *number of air conditioners sold* and *number of sunglasses sold*. However clearly neither one of these variables is *causing* the other. What is most likely happening is that another variable *(outdoor temperature)* is contributing to the changes in both variables.

Correlation measures association, not cause. Its much harder to show that one variable *causes* another variable.

WB p. 32

**WORKED EXAMPLE 12** Exploring causation

For each of the following correlations between pairs of variables, suggest another variable that could be the underlying cause of the correlation between the two.

| Steps | Working |
|---|---|
| a A positive correlation between the *number of cars sold in Victoria* and the *number of pizza slices eaten in Victoria*. | |
| Which variable might be causing changes in both? | *Population* changes could be the underlying cause of the correlation between the two. |
| b A negative correlation between *umbrellas sold* ($) and *ice creams sold ($)*. | |
| Which variable might be causing changes in both? | *Outdoor temperature* changes could be the underlying cause of the correlation between the two. |
| c A positive correlation between the *number offirefighters called out to fight a blaze* and the *level of fire damage ($)*. | |
| Which variable might be causing changes in both? | The *size of the fire* could be the underlying cause of the correlation between the two. |
| d A positive correlation between *amount spent on private tutor hours per household* and *amount spent on streaming service subscriptions per household ($)*. | |
| Which variable might be causing changes in both? | *Household income* could be the underlying cause of the correlation between the two. |

## Experimentation and causation

An association between two variables is *never* enough to show causation, no matter how obvious the causation appears. The only way to reasonably claim causation is through an experimental study.

Data can be gathered by either observation or experimentation. In an observational study, the researcher passively observes an existing situation, whereas in an experimental study, the researcher actively manipulates a situation to eliminate other possible variables *before* observing the results.

Despite the advantages of experimental studies, we often have to use observational studies because experimental studies can be impractical, unethical or too expensive. For example, if we wanted to establish whether women who smoked through pregnancy gave birth to underweight babies, it would be unethical to set up an experiment that involved getting non-smoking pregnant women to smoke.

**Experimentation and causation**

- Only an experimental study can reasonably claim to show causation.
- Experimental studies eliminate other variables by:
  - controlling the factors that can be controlled
  - randomisation
  - repeating the experiment.



9780170448406

## VCE QUESTION ANALYSIS

©VCAA 2004 2CQ2 2004 Examination 2 Core Question 2 (11 marks)

2.5

Video VCE question analysis: Associations between two variables

*Body mass index* (BMI) is defined as

$$\text{BMI} = \frac{\text{mass}}{(\text{height})^2}$$

where *mass* is measured in kilograms and *height* in metres.

a Determine the body mass index of a person who weighs 66 kg and who is 1.69 m tall.
Write your answer correct to one decimal place. 1 mark

The BMI for each person in a sample of 17 males and 21 females is recorded in Table 1.

Table 1 BMI of males and females

| Body mass index (kg/m$^2$) | | | |
|---|---|---|---|
| Males | Females | Males | Females |
| 31.4 | 27.0 | 22.0 | 21.8 |
| 30.1 | 26.9 | 21.8 | 21.5 |
| 26.8 | 25.2 | 21.6 | 21.4 |
| 25.7 | 24.6 | 21.1 | 20.9 |
| 25.5 | 24.2 | 20.9 | 20.6 |
| 25.5 | 24.2 | 20.6 | 20.3 |
| 23.6 | 23.4 | | 20.1 |
| 23.3 | 23.4 | | 19.9 |
| 22.5 | 22.8 | | 18.8 |
| 22.4 | 22.5 | | 17.5 |
| 22.3 | 22.4 | | |

b Write the range of the BMI data for males in the sample. 1 mark

c A BMI greater than 25 is sometimes taken as an indication that a person is overweight.
Use this criterion and the data in Table 1 to copy and complete Table 2, the two-way frequency table. 2 marks

Table 2 Weight rating by gender

| | Gender | |
|---|---|---|
| Weight rating | Male | Female |
| overweight | | |
| not overweight | | |
| Total | 17 | 21 |

d Does the data support the contention that, for this sample, weight rating is associated with gender? Justify your answer by quoting appropriate percentages correct to one decimal place. 2 marks e The parallel boxplots in Figure 1 have been constructed to compare the distribution of BMI for males and females in this sample.

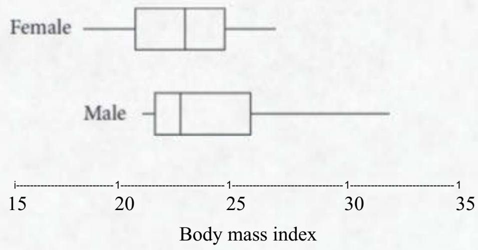



Figure 1 Parallel boxplots comparing BMI for males and females

i Use the parallel boxplots to identify and name two similar properties of the BMI distributions for males and females. 2 marks

ii Use the information in Table 1 to determine the mean BMI for the males in this sample. Write your answer correct to one decimal place. 1 mark

iii The median BMI for males is 22.5. Of the mean or median, which measure gives a better indication of the typical BMI for males? Explain your answer. 2 marks

## Reading the question

- When dealing with two data sets, always note which set each question is referring to.
- Highlight when you are asked for a percentage and the sort of rounding required.
- Note terms whose meaning you might not be 100% clear on, such as similar properties'.

## Thinking about the question

- You will need to know what information a boxplot shows.
- Think about why some questions are worth two marks.
- Do you know the reasons for choosing a median over a mean?

## Worked solution (/ = 1 mark)

a Substitute the values into the BMI formula:

$$\text{BMI} = \frac{\text{mass}}{(\text{height})^2} = \frac{66}{(1.69)^2}$$

= 23.108...

= 23.1 / $\text{kg/m}^2$, correct to one decimal place.

b range = highest male BMI value - lowest male BMI value

= 31.4-20.6

= 10.8 / $\text{kg/m}^2$

c From Table 1 we can see that 6 males have a BMI over 25 (overweight) and 11 under 25 (underweight), and for females 3 are overweight and 18 are underweight. Insert these numbers in the two-way table.

| | Gender | |
|---|---|---|
| Weight rating | Male | Female |
| overweight | 6 | 3 |
| not overweight | 11 ✓ | 18 □ |
| Total | 17 | 21 |



9780170448406

2.5

d From the two-way table:

percentage of overweight males = $\frac{6}{17} \times 100\%$

$= 35.3\%$

and

percentage of overweight females = $\frac{3}{21} \times 100\%$

$= 14.3\%$ (correct to one decimal place).

The percentage of overweight males to females is 35.3% to 14.3%, which is considerably different /, so yes, the data supports the contention / that, for this sample, weight rating is associated with gender.

**Exam hack**

If the question asks, 'Does the data support...', you need to quote some figures to justify what you're saying, not just give an opinion.'

e i Note all the properties that are shown on a boxplot: minimum, $Q_p$ median($Q_2$), IQR, $Q_3$ and maximum, and select the two that are similar in the parallel boxplots. The median / and IQR / (length of the boxes) are similar for male and female.

ii Mean BMI for males = $\frac{\text{sum of male BMIs}}{\text{no. of male BMIs}}$

$= \frac{407.1}{17}$

$= 23.947...$

$= 23.9$ / $kg/m^2$ correct to one decimal place (or use CAS),

iii The mean takes into account every data value, so it can be affected by extreme values.

The top two male BMIs, 31.4 and 30.1, could be classed as extreme values. /
The median gives the better indication of the typical BMI because it is not affected by extreme values, and the mean is affected. /

## Student performance

80-100% 60-79% 0-59%

a This was generally done well. Some students forgot to square the *height.*

b 93% This was also done well, although some incorrectly wrote '20.6 - 31.4*.

c 93% Some students entered percentages in the table. Since this was a 2-mark question, if the percentages were correct, students were awarded one mark.

d 51 % Answers had to quote percentages not raw numbers. No marks were given for 'yes' without justification. Some students ignored the data and expressed their own observations of weight distribution according to gender. Other incorrect answers used fractions, such as $\frac{6}{11}$ or $\frac{3}{18}$, or their percentage equivalent.

e i 51% Most students gave *median* as one answer. The most common incorrect term was *range.*

ii 51% The most common error involved incorrect rounding. Many students appeared to find the correct number of 23.947, which they rounded to 23.95 with two decimal places, and some rounded 23.95 up to 24.0. These students were not awarded the mark for the question.

iii 51% This was not done well. Simply giving the median as the answer without an explanation was not enough for a mark. The most common incorrect response was that the *mean* gave the better indication because 'the *mean* takes all data values into account while the *median* is only the middle number'. Another common incorrect interpretation was 'the *mean* gives the average but the *median* is only the midpoint'. Many students who correctly named the *median* then referred to it not being affected as much by *outliers.* This answer scored only one mark because although 31.4 and 30.1 can be describes as *extreme values* from Table 1, the boxplots didn't indicate any *outliers.*

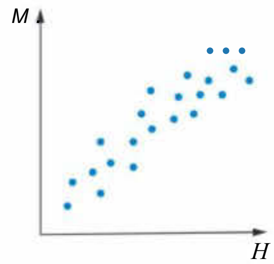
## EXERCISE 2.5 Correlation and causation

ANSWERS p. 699

### Recap

*Use the following information to answer the next two questions.*

Consider this scatterplot, which shows the relationship between variables *M* and *H.*




1 Which statement best describes the association between *M* and H?

A no association

B non-linear association

C negative, linear and weak

D positive, linear and weak

E positive, linear and strong

2 Which of the following sentences is correct?

A It can be concluded that *M* should decrease as *H* increases.

B It has been proven that *M* will always increase as *H* increases.

C It can be concluded that *M* is likely to increase as *H* increases.

D There is clear evidence to support that *H* will increase as *M* increases.

E It can be concluded that *H* is likely to increase as *M* decreases.

### Mastery

3 H WORKED EXAMPLE 11 Interpret the correlation coefficient values and write a sentence describing the association for each of the studies described below, beginning with 'The data suggests...'.

a A study is investigating whether there is an association between the *amount of exercise* a person does and their *height.* The correlation coefficient was found to be r = 0.294.

b A study is investigating whether there is an association between *weight* and the *number of hours a person spends sitting down.* The correlation coefficient was found to be r = 0.528.

c A study is investigating whether there is an association between the *percentage of good peaches* in a container and their *number of weeks in storage.* The correlation coefficient was found to be r=-0.910.

d A study is investigating whether there is an association between the number of *sales* at a local store and the *temperature* over a year. The correlation coefficient was found to be *r* = 0.724.

e A study is investigating whether there is an association between the *number of pets* owned and the average *temperature* over a year. The correlation coefficient was found to be r = -0.062.

f A study is investigating whether there is an association between the *number of home-cooked meals* per week and *income.* The correlation coefficient was found to be r = -0.463.



9780170448406

▶ 40 using CAS 3~] At a local cinema, people were asked their *yearly income* and the *number of times they had been to the movies in the pastyear.* Calculate the correlation coefficient, correct to two decimal places, for the data in the table.

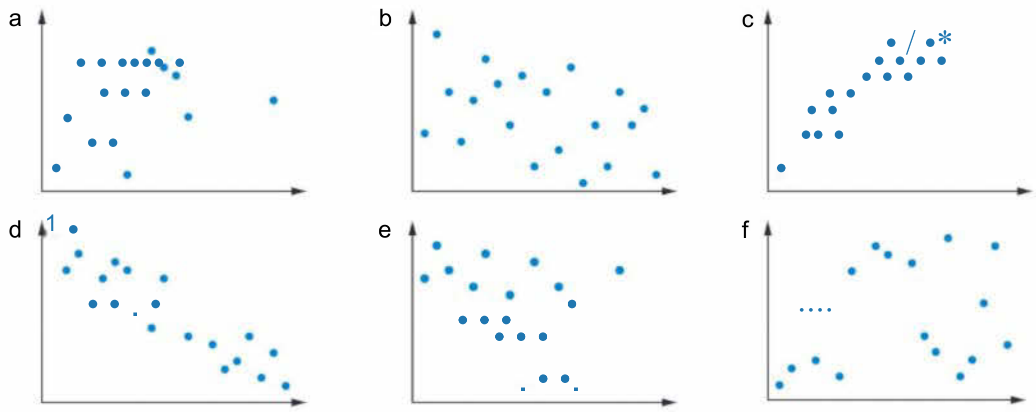
a What are the three assumptions we make when calculating the correlation coefficient?

b Discuss whether there is an association between *income* and the *number of cinema movies seen in the past year.*

| Income ($1000s per year) | 63 | 125 | 32 | 19 | 136 | 49 | 102 | 67 | 82 | 25 | 91 | 42 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of times attended a cinema in the past year | 4 | 12 | 1 | 8 | 16 | 13 | 13 | 6 | 7 | 6 | 12 | 10 |

5 Match the following scatterplots with their correlation coefficients:

0.2, -0.9, -0.4, -0.5, 0.8, 0.6




6 H WORKED EXAMPLE 12 I For each of the following correlations between pairs of variables, suggest another variable that could be the underlying cause of the correlation between the two.

a A negative correlation between the *number of people who drown* and the *number of ski jackets sold.*

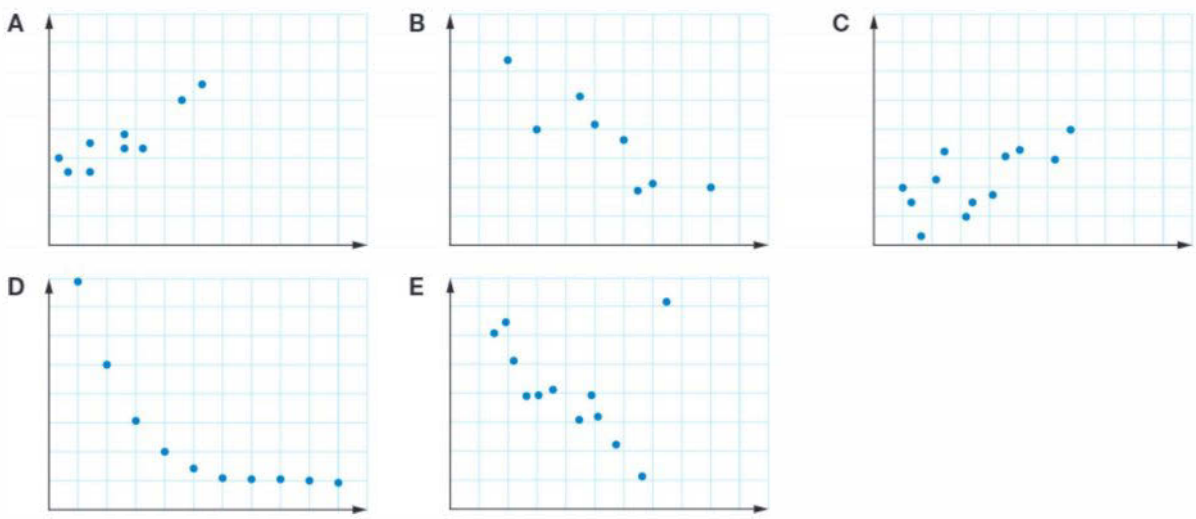
b A positive correlation between the *number of weddings in Victoria* and the *total consumption of apples in Victoria* (kg).

c A positive correlation between the *amount of money spent on eating out in a household* ($) and the *number of laptops in a household.*

d A positive correlation between *spelling ability in children and foot length.*

**Exam practice** 80-100% 60-79% 0-59%

7 Which of the following scatterplots have a correlation coefficient $r$ closest to 0.921?



▶ 8 ©VCAA 2013 1CQ8 I 79% The table shows the hourly rate of pay earned by 10 employees in a company in 1990 and in 2010.

The value of the correlation coefficient, r, for this set of data is closest to

A 0.74 B 0.86

C 0.92 D 0.93

E 0.96

| | Hourly rate of pay ($) | |
|---|---|---|
| Employee | 1990 | 2010 |
| Ben | 9.53 | 17.02 |
| Lani | 9.15 | 16.71 |
| Freya | 8.88 | 15.10 |
| Jill | 8.60 | 15.93 |
| David | 7.67 | 14.40 |
| Hong | 7.96 | 13.32 |
| Stuart | 6.42 | 15.40 |
| Mei Lien | 11.86 | 19.79 |
| Tim | 14.64 | 23.38 |
| Simon | 15.31 | 25.11 |

9 ©VCAA | 20021CQ10J 66% The following data was recorded from measurements made on 12 men.

The value of the correlation coefficient, r, for mass against waist measurement, is closest to

A 0.6061 B -0.7785

C 0.8675 D 0.9314

E 0.9651

| Age (years) | Mass (kg) | Waist (cm) |
|---|---|---|
| 26 | 84 | 84 |
| 29 | 72 | 74 |
| 32 | 67 | 89 |
| 32 | 59 | 75 |
| 34 | 97 | 106 |
| 37 | 112 | 114 |
| 39 | 67 | 80 |
| 40 | 91 | 101 |
| 41 | 98 | 101 |
| 43 | 89 | 94 |
| 45 | 117 | 126 |
| 51 | 62 | 82 |

10 ©VCAA | 2011 icon"] 62% For a group of 15-year-old students who regularly played computer games, the correlation between the time spent playing computer games and fitness level was found to be $r$ = -0.56. On the basis of this information it can be concluded that

A 56% of these students were not very fit.

B these students would become fitter if they spent less time playing computer games.

C these students would become fitter if they spent more time playing computer games.

D the students in the group who spent a short amount of time playing computer games tended to be fitter.

E the students in the group who spent a large amount of time playing computer games tended to be fitter.

11 A worldwide study has shown that there is a high positive correlation between the number of cars a person owned in their lifetime and the age they live to. Which of the following variables would best explain this association?

A *persons' wealth ($)* B *number of cars owned*

C *population* D *age when started to drive*

E *number of cars in the country*

▶



9780170448406

12 A controlled experiment involving a random selection of subjects has shown that people who use calorie-free sweeteners in place of sugar tend to gain weight. Which of the following is the most plausible explanation for this association?

A Calorie-free sweeteners are not actually calorie free.

B Calorie-free sweeteners cause people to gain weight.

C Calories don't have anything to do with weight gain.

D There is another variable causing the association.

E The people who say they use calorie-free sweeteners are lying.

13 Which of the following studies can best be described as an experimental study?

A The number of accidents at train level crossing across Victoria over a year are recorded to identify which crossings should be replaced by overpasses.

B The main meals ordered by people in a restaurant are recorded over a month to see which dishes could be cut from the menu to make room for new ones.

C The effect of a particular fertiliser on a fruit tree is tested by randomly selecting sections of orchard and using the fertiliser on those sections.

D To establish whether VCE students in Victoria would be happy for the school day to start and finish an hour later each day, all the VCE students at a school are surveyed.

E To establish whether eating breakfast affects school marks, a particular school asks its students if they eat breakfast and correlates this against their marks.

14 ^^^^20032CQI, (4 marks) The table shows the number of telephone calls made on a given day by a sample of 12 people working in a large company. Also given is the cost of each persons calls for the day.

| Person | Number of calls | Cost (dollars) |
|---|---|---|
| A | 33 | 4.54 |
| B | 15 | 1.00 |
| C | 22 | 5.96 |
| D | 27 | 4.47 |
| E | 52 | 8.87 |
| F | 34 | 8.50 |
| G | 55 | 11.09 |
| H | 47 | 8.51 |
| I | 11 | 3.98 |
| J | 18 | 2.42 |
| K | 36 | 11.30 |
| L | 27 | 7.48 |

a 60% Determine the value of the Pearson correlation coefficient, $r$. Write your answer correct to four decimal places. 1 mark

b What level of strength of association does $r$ indicate? 1 mark

c 41% Copy and complete the following:

The value of the Pearson correlation coefficient measures the strength and direction of theassociation between call cost and number of calls, 1 mark

d 41% If we wish to predict the cost of calls from the number of calls made, what is the response variable? 1 mark

15 (9 marks) Two researchers want to find out whether taking longer strides increases running speed,

a What are the two variables involved? 2 marks

The first researcher decides to look at video footages of Olympic 100 m runners, count the number of strides it takes them to complete the race, calculate the average stride length and compare this with their recorded speed.

b Does this study involve observation or experimentation? Give a reason for your answer. 2 marks

The second researcher decides to randomly select 50 people, video and time them running 100 m races on a number of occasions under different conditions over a year, then count strides to calculate their average stride length and compare this with their speed.

c Does this study involve observation or experimentation? Give a reason for your answer. 2 marks

d Both studies show a strong positive association between *length of stride* and *running speed.* Which one can we reasonably say has shown that taking longer strides causes an increase in running speed? Give reasons for your answer. 2 marks

e List an unknown factor that could have influenced the first researchers' study. 1 mark

16 (5 marks) Centuries ago, the people of the Hebrides, a chain of islands north of Scotland where head lice were common, were convinced that the head lice cured people who were sick with fever. They had noticed that while healthy people nearly always had head lice, sick people didn't have any.

a What is the causal link that the people of the Hebrides were making? 1 mark

b Is this association based on observation or experimentation? 1 mark

c What do you think was really happening to cause this association? 1 mark

d Explain the mistake the people of the Hebrides were making in terms of explanatory and response variables. 1 mark

e Briefly describe an experiment that you could set up to reasonably establish whether or not head lice cause people to be healthy. Comment on the ethics of your experiment. 1 mark


9780170448406

# (7) Chapter summary

2

CHAPTER SUMMARY

## Explanatory and response variables

- An explanatory variable is a variable that we use to predict or explain the changes observed in another variable, which is called a response variable.
- If the words 'explain changes' or 'predict' don't appear in the question, decide which of the two variables is the most likely to affect the other.

## Associations between variables

Two variables are said to be associated when they are linked in some way.

To find whether two categorical variables are associated or not, we compare percentages in

- percentage two-way frequency tables
- parallel percentage segmented bar charts.

To find whether a numerical variable and categorical variable are associated or not, we compare

1 shape: symmetry and skewness

2 centre: median

3 spread: range and IQR

The following displays can be used:

- back-to-back stem plots
- parallel dot plots
- parallel boxplots.

To find whether two numerical variables are associated or not, we use

- scatterplots
- Pearson correlation coefficient.

## Two-way tables

- Two-way frequency tables can be used to look at the association between two categorical variables.
- Percentaging two-way frequency tables gives us more information. We usually percentage the explanatory variables.
- Information from a percentage two-way frequency table can be displayed as a parallel percentage segmented bar chart where each cell in the table corresponds to a segment in the bar chart.

## Graphs showing association between two variables

| Explanatory variable | Response variable | Graph | Comments |
|---|---|---|---|
| categorical | categorical | parallel percentage segmented bar charts | |
| categorical | numerical | back-to-back stem plot | Lets us see data values, but only two categories are possible. |
| categorical | numerical | parallel dot plots | Allows us to read data values from graph, but comparisons are not as easily made as with a parallel boxplot. |
| categorical | numerical | parallel boxplots | Makes it easier to compare medians and quartiles. |
| numerical | numerical | scatterplot | |

## Scatterplots

Three ways scatterplots can be used to describe the association between two numerical variables:

1 Direction: positive or negative

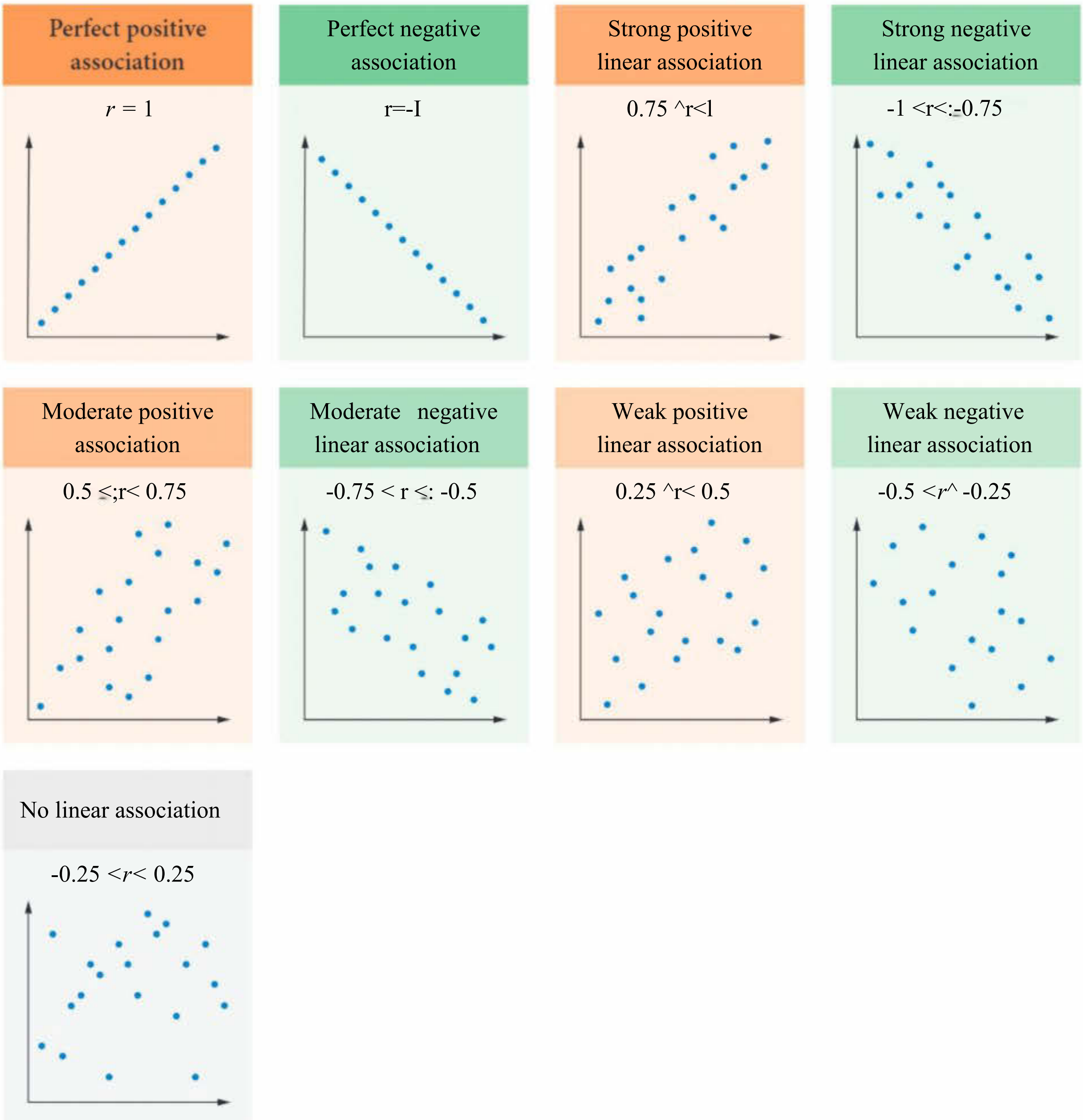
2 Form: linear, non-linear or no association

3 Strength: strong association, moderate association, weak association

## The Pearson correlation coefficient

- The Pearson correlation coefficient, r, is a number on the scale from -1 to 1 that measures the strength and direction of *linear* associations.
- The calculation gives the same value regardless of which variable we take to be the explanatory variable and which one to be the response variable.

It's possible to estimate the value of r from the shape of the scatterplot using these guidelines:




- If we are using the Pearson correlation coefficient to measure the association of two variables, we are generally making three assumptions:

  1 The variables are both numerical.

  2 The association is linear.

  3 There are no outliers.



9780170448406

## Causation

- Just because two variables have a high correlation or association, doesn't necessarily mean that one *causes* the other.
- Data can be gathered by either observation or experimentation. In an observational study, the researcher passively observes an existing situation, whereas in an experimental study, the researcher actively manipulates a situation to eliminate other variables before observing it.
- Only an experimental study can reasonably be said to show causation, but sometimes an experimental study isn't possible for practical, expense or ethical reasons.

# Cumulative examination 1

Total number of marks: 11 Reading time: 5 minutes Writing time: 25 minutes

1 ©VCAA 2017N1CQ6 Which of the following could be used to identify and describe the association between the variables *height* (short, medium, tall) and *hat size* (small, medium, large)?

A a histogram
B a scatterplot
C parallel boxplots
D a segmented bar chart
E a back-to-back stem plot

*Use the following information to answer the next two questions.*

The table lists the speed (in km/h) of ten cars recorded in a 60 km/h zone. Also recorded are the ages (in years) of the drivers.

| Speed | Age |
|---|---|
| 71.8 | 27 |
| 68.3 | 38 |
| 65.1 | 22 |
| 63.2 | 64 |
| 62.8 | 57 |
| 62.6 | 37 |
| 62.5 | 21 |
| 61.3 | 19 |
| 60.1 | 57 |
| 59.8 | 61 |

2 ©VCAA 2005 1CQ1 The median speed (in km/h) of the ten cars is

A 62.6
B 62.7
C 62.8
D 63.0
E 63.5

3 ©VCAA 2005 1CQ2 The percentage of the drivers over the age of 25 years is

A 30%
B 40%
C 50%
D 60%
E 70%

*Use the following information to answer the next two questions.*

The foot lengths of a sample of 2400 women were approximately normally distributed with a mean of 23.8 cm and a standard deviation of 1.2 cm.

4 ©VCAA 2015 1CQ4 The expected number of these women with foot lengths less than 21.4 cm is closest to

A 60
B 120
C 384
D 2280
E 2340

5 ©VCAA 2015 1CQ5 The standardised foot length of one of these women is $z = -1.3$. Her actual foot length, in centimetres, is closest to

A 22.2
B 22.7
C 25.3
D 25.6
E 31.2

*Use the following information to answer the next three questions.*

The data in the table was collected in a study of the association between the variables *frequency of nightmares* (low, high) and *snores* (no, yes).

| Frequency of nightmares | Snores | | Total |
|---|---|---|---|
| | No | Yes | |
| low | 80 | 58 | 138 |
| high | 11 | 12 | 23 |
| Total | 91 | 70 | 161 |

Data: adapted from RA Hicks and J Bautista/Snoring and nightmares', *Perceptual and Motor Skills*, 1 October 1993, https://doi.org/10.2466/pms.1993.77.2.433

6 ^^^j20201 CQI o J The variables in this study, *frequency of nightmares* (low, high) and *snores* (no, yes), are

A ordinal and nominal respectively.

B nominal and ordinal respectively.

C both numerical.

D both ordinal.

E both nominal.

**7** ©VCAA 20201CQ1 IThe percentage of participants in the study who did not snore is closest to

A 42.0% B 43.5% C 49.7% D 52.2% E 56.5%

8 ©VCAA 20201CQ12 J Of those people in the study who did snore, the percentage who have a high frequency of nightmares is closest to

A 7.5% B 17.1% C 47.8% D 52.2% E 58.0%

*Use the following information to answer the next three questions.*

The back-to-back ordered stem plot shows the distribution of maximum temperatures (in °Celsius) of two towns, Beachside and Flattown, over 21 days in January.

| Beachside | Stem | Flattown |
|---|---|---|
| 9 8 7 5 | 1 | 8 9 |
| 4 3 2 2 1 1 0 0 | 2 | |
| 9 9 8 7 6 5 | 2 | 8 9 |
| 3 2 | 3 | 3 3 4 |
| 8 | 3 | 55677788 |
| | 4 | 0 0 12 |
| | 4 | 5 6 |

9 ©VCAA 2006 1CQ1 The variables temperature (°Celsius) and town (Beachside or Flattown) are

A both categorical variables.

B both numerical variables.

C categorical and numerical variables respectively.

D numerical and categorical variables respectively.

E neither categorical nor numerical variables.

10 ©VCAA J 20061CQ2 For Beachside, the range of maximum temperatures is

A 3°C B 23°C C 32°C D 33°C E 38°C

11 ©VCAA 2006 1CQ3 The distribution of maximum temperatures for **Flattown** is best described as

A negatively skewed.

B positively skewed.

C positively skewed with outliers.

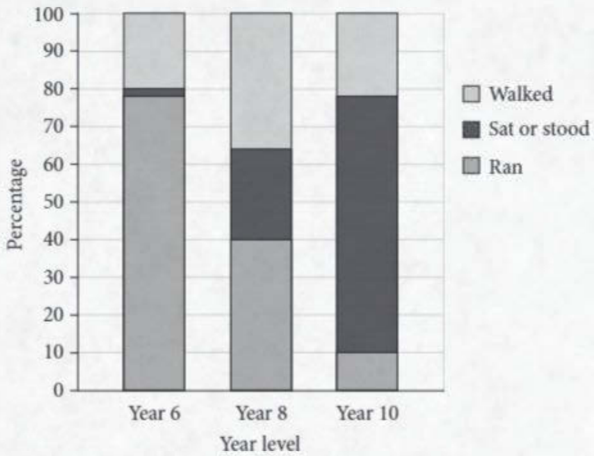
D approximately symmetric.

E approximately symmetric with outliers.

# Cumulative examination 2

Total number of marks: 18 Reading time: 5 minutes Writing time: 27 minutes

1 ©VCAA 2008 2CQ2 J (2 marks) In a large survey, Years 6, 8 and 10 girls were asked what they did (walked, sat, stood, ran) for most of the time during a typical school lunch time. The results are displayed in the percentage segmented bar chart.




Does the percentage segmented bar chart support the opinion that, for these girls, the lunch time activity (walked, sat or stood, ran) undertaken is associated with Year level? Justify your answer by quoting appropriate percentages.

2 ©VCAA I 2021 2CQ1bc] (2 marks) In the sport of heptathlon, athletes compete in seven events. These events are the 100 m hurdles, high jump, shot-put, javelin, 200 m run, 800 m run and long jump. Fifteen female athletes competed to qualify for the heptathlon at the Olympic Games. Their results for three of the heptathlon events - high jump, shot-put and javelin - are shown in Table 1 below.

Table 1

| Athlete number | High jump (metres) | Shot-put (metres) | Javelin (metres) |
|---|---|---|---|
| 1 | 1.76 | 15.34 | 41.22 |
| 2 | 1.79 | 16.96 | 42.41 |
| 3 | 1.83 | 13.87 | 46.53 |
| 4 | 1.82 | 14.23 | 40.62 |
| 5 | 1.87 | 13.78 | 45.64 |
| 6 | 1.73 | 14.50 | 42.33 |
| 7 | 1.68 | 15.08 | 40.88 |
| 8 | 1.82 | 13.13 | 39.22 |
| 9 | 1.83 | 14.22 | 42.51 |
| 10 | 1.87 | 13.62 | 42.75 |
| 11 | 1.87 | 12.01 | 38.12 |
| 12 | 1.80 | 12.88 | 42.65 |
| 13 | 1.83 | 12.68 | 45.68 |
| 14 | 1.87 | 12.45 | 41.32 |
| 15 | 1.78 | 11.31 | 42.88 |

a Copy and complete Table 2 by calculating the mean height jumped for the high jump, in metres, by the 15 athletes. 1 mark

Table 2

| Statistic | High jump (metres) | Shot-put (metres) |
|---|---|---|
| Mean | | 13.74 |
| Standard deviation | 0.06 | 1.43 |

b In shot-put, athletes throw a heavy spherical ball (a shot) as far as they can.

Athlete number six, Jamilia, threw the shot 14.50 m. Calculate Jamilias standardised score (z). Round your answer to one decimal place. 1 mark

3 ©VCAA 2006 2CQ1 (7 marks) Table 1 below shows the heights (in cm) of three groups of randomly chosen boys aged 18 months, 27 months and 36 months respectively.

Table 1

| Height (cm) | | |
|---|---|---|
| 18 months | 27 months | 36 months |
| 76.0 | 82.0 | 88.0 |
| 78.5 | 83.1 | 88.8 |
| 78.6 | 84.0 | 90.0 |
| 80.0 | 86.8 | 92.3 |
| 80.5 | 87.2 | 93.0 |
| 81.2 | 87.6 | 94.1 |
| 82.8 | 88.3 | 94.2 |
| 83.2 | 90.7 | 95.8 |
| 83.4 | 91.0 | 96.9 |
| 83.7 | 92.3 | 97.1 |
| 85.8 | 92.5 | 97.8 |
| 86.6 | 93.1 | 99.2 |
| 87.3 | 94.8 | 100.6 |
| 89.8 | 97.2 | 103.8 |

a Complete Table 2 below by calculating the standard deviation of the heights of the 18-month-old boys. Write your answer correct to one decimal place. 1 mark

Table 2

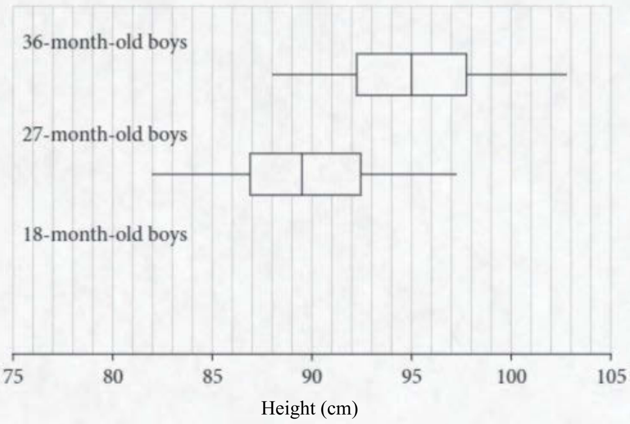
| Age | 18 months | 27 months | 36 months |
|---|---|---|---|
| Mean | 82.7 | 89.3 | 95.1 |
| Standard deviation | | 4.5 | 4.5 |

b A 27-month-old boy has a height of 83.1 cm.

Calculate his standardised height (z-score) relative to this sample of 27-month-old boys. Write your answer correct to one decimal place. 1 mark c The heights of the 36-month-old boys are normally distributed. A 36-month-old boy has a standardised height of 2. Approximately what percentage of 36-month-old boys will be shorter than this child? 1 mark

d Using the data from Table 1, boxplots have been constructed to display the distributions of heights of 36-month-old and 27-month-old boys as shown.

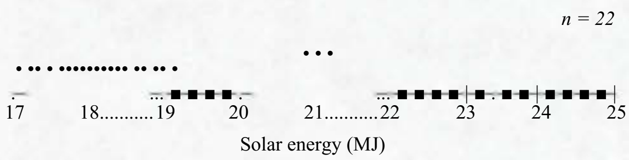
Copy and complete the display by constructing and drawing a boxplot that shows the distribution of heights for the 18-month-old boys. 2 marks




e Use the appropriate boxplot to determine the median height (in centimetres) of the 27-month-old boys. 1 mark

f The three parallel boxplots suggest that *height* and *age* (18 months, 27 months, 36 months) are positively related. Explain why, giving reference to an appropriate statistic. 1 mark

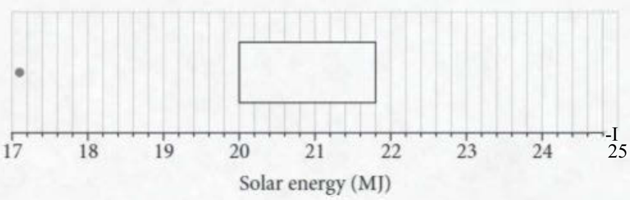
4 ©VCAA | 20i7N2CQi3 (7 marks) A 1 $m^2$ solar array is located at a weather station. The total amount of energy generated by the solar array, in megajoules, is recorded each month. The data for the month of February for the last 22 years is displayed in the dot plot.




a Determine the number of years in which the energy generated during February was greater than 23 MJ. 1 mark

b For the data in the dot plot above, the first quartile Qj = 20 and the third quartile $Q_3$ = 21.8. Show that the data value 17.1 is an outlier. 2 marks

c Using the data in the dot plot, copy and complete the boxplot below. 2 marks



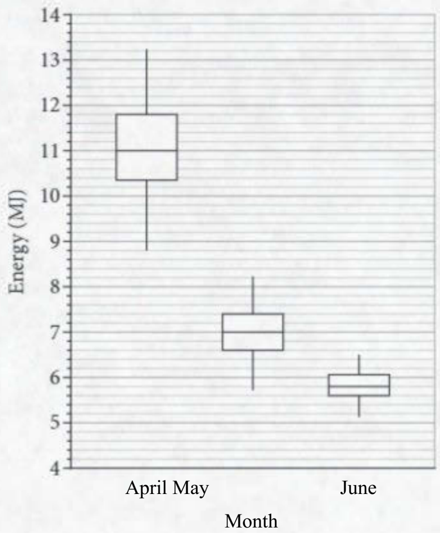


9780170448406

d The distribution of the amount of energy generated by the solar array for the months of April, May and June for the last 22 years is displayed in the following parallel boxplots.

The parallel boxplots suggest that the amount of energy generated is associated with the month of the year. Explain why, quoting the values of an appropriate statistic. 2 marks





CHAPTER

3

# LINEAR ASSOCIATIONS

## Study Design coverage

### AREA OF STUDY 1: DATA ANALYSIS, PROBABILITY AND STATISTICS

#### Investigating and modelling linear associations

- least squares line of best fit $y = a + bx$, where $x$ represents the explanatory variable, and $y$ represents the response variable; the determination of the coefficients $a$ and $b$ using technology, and the formulas $b = r\frac{s_y}{s_x}$ and $a = \bar{y} - b\bar{x}$
- modelling linear association between two numerical variables, including the:
  - identification of the explanatory and response variables
  - use of the least squares method to fit a linear model to the data
- interpretation of the slope and intercepts of the least squares line in the context of the situation being modelled, including:
  - use of the rule of the fitted line to make predictions being aware of the limitations of extrapolation
  - use of the coefficient of determination, $r^2$, to assess the strength of the association in terms of explained variation
  - use of residual analysis to check quality of fit
- data transformation and its use in transforming some forms of non-linear data to linearity using a square, logarithmic (base 10) or reciprocal transformation (applied to one axis only)
- interpretation and use of the equation of the least squares line fitted to the transformed data to make predictions.



STUDY DESIGN COVERAGE

**Video playlists (6):**

**3.1** The least squares Line of best fit
**3.2** The coefficient of determination
**3.3** Making predictions
**3.4** Residual analysis
**3.5** Data transformations
**VCE question analysis** Linear associations

**Worksheets (6):**

**3.1** Line of fit • Least squares regression line
**3.2** Coefficient of determination
**3.3** Interpolation and extrapolation
**3.4** Predictions and residuals • Residual plots

Nelson MindTap

To access resources above, visit cengage.com.au/nelsonmindtap

Video playlist The Least squares Line of best fit

Worksheets Line of fit

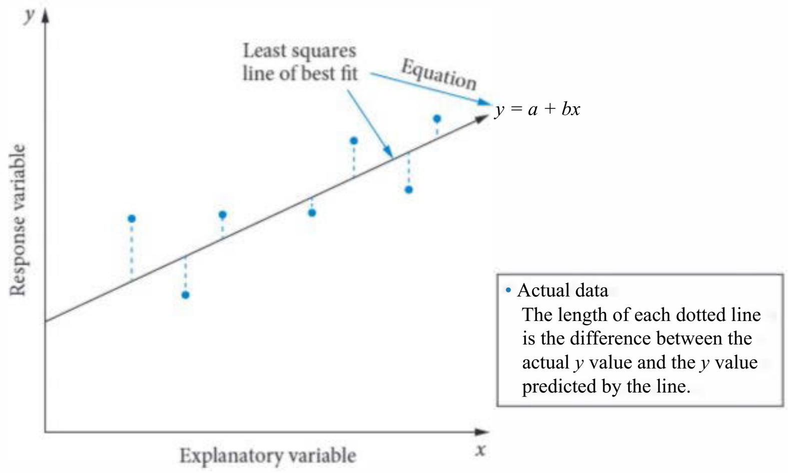
Least squares regression Line

# (sj) The least squares line of best fit

## Least squares line of best fit

A line of best fit is a straight line that best shows the association between two numerical variables on a scatterplot and helps us to make predictions.

The least squares line of best fit (also known as a least squares regression line or just a least squares line) can be determined by finding the line that minimises the sum of the squares of the vertical distances (the dotted lines in the diagram) between the line and each data point in a scatterplot. These vertical distances are the differences between the predictions made by the line and the actual data points, so it makes sense that we would want to minimise these prediction errors. The vertical distances above the line are positive and the vertical distances below the line are negative.




The x-axis variable is the explanatory variable and they-axis variable is the response variable. When a least squares line of best fit equation is calculated, the equation should be written in terms of the variable names rather than using the letters x and *y*.

### Least squares line of best fit equation

The general form of the equation for the least squares line of best fit is

$$y = a + bx$$

where the slope or gradient of the line is

$$b = r\frac{s_y}{s_x}$$

the y-intercept of the line is

$$a = \bar{y} - b\bar{x}$$

and

- r is the Pearson correlation coefficient
- $s_x$ and $s_y$ are the sample standard deviations of x and y respectively
- $\bar{x}$ and $\bar{y}$ are the sample means of x and y respectively.

We always calculate *b* first because *b* is needed in the calculation for *a*.

### Exam hack

The values *a* and *b* in the least squares line of best fit equation are called the 'coefficients' of the equation. This is a general term for these values in an equation. Don't confuse this with the use of the same word in the term 'correlation coefficient'.

This least squares line of best fit equation appears on the examination formula sheet, so you don't need to memorise it, but you do need to know what values r, x, $\bar{y}$, $s_x$ and $s_y$ represent, and how to calculate them.



9780170448406

3.1

**WORKED EXAMPLE 1** Finding the least squares line of best fit equation using the formula

The total runs scored and the total balls faced by a batsman over a cricket season were recorded and the values of the following statistics were determined.

The sample mean of the runs scored was 14.91.

The sample standard deviation of the runs scored was 11.74.

The sample mean of the balls faced was 22.45.

The sample standard deviation of the balls faced was 15.81.

The Pearson correlation coefficient was 0.89.

a State the explanatory and response variables.

b Write the values for $\bar{x}$, $s_x$, $\bar{y}$, $s_y$ and r.

c Calculate the equation for the least squares line of best fit that models this data, rounding the coefficients *a* and *b* to two significant figures.

| **Steps** | **Working** |
| --- | --- |
| a Which variable affects the other? | explanatory variable: *balls faced*<br>response variable: *runs scored* |
| b Which is the *x* variable and which is the *y* variable? | $\bar{x}$ =22.45, $s_x$= 15.81<br>$\bar{y}$ = 14.91, $s_y$= 11.74, r= 0.89 |
| c 1 Calculate the slope *(b)* using the formula and round to two significant figures. | $b = r \times \frac{s_y}{s_x}$<br>- 0.89 x $\frac{1174}{15.81}$<br>= 0.660885 5...<br>» 0.66 |
| 2 Calculate the *y*-intercept (a) using the formula and round to two significant figures. Use the *unrounded* answer for *b* when calculating the value of *a*. | $a = \bar{y} - b\bar{x}$<br>= 14.91 -0.6608855... x 22.45<br>= 0.07312...<br>« 0.073 |
| 3 Write the equation for the least squares line of best fit. | *a* - 0.073 and *b* - 0.66<br>*y* = 0.073 + 0.66x |
| 4 Replace *y* and *x* in the equation with the correct variable names. | *runs scored* = 0.073 + 0.66 x *balls faced* |

WB
p. 33

**Exam hack**

Never use rounded values in calculations. Only round for final answers.

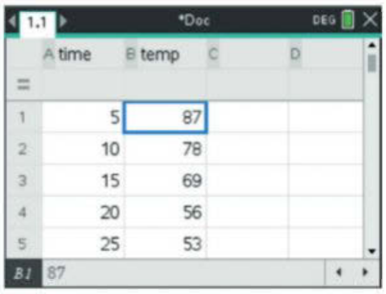
## USING CAS 1 Finding the least squares line of best fit equation

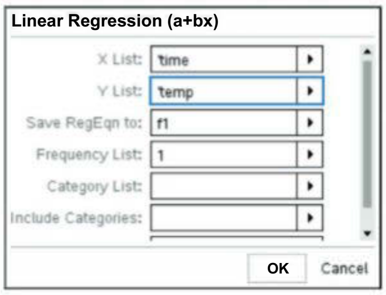
The following table shows the results of an experiment that measures the temperature of a liquid as it cools down after the source of heat is removed. Find the equation of the least squares line of best fit, rounding the coefficients $a$ and $b$ to two significant figures.

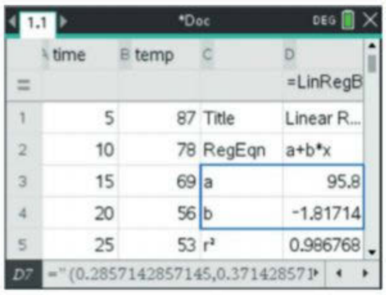
| Time (minutes) | 5 | 10 | 15 | 20 | 25 | 30 |
|---|---|---|---|---|---|---|
| Temperature (°C) | 87 | 78 | 69 | 56 | 53 | 41 |

*Time* is the explanatory variable and *temperature* is the response variable.

**TI-Nspire**




1 Start a new document and add a Lists & Spreadsheet page.

2 Label the columns and enter the data from the table as shown above.

3 Press menu > Statistics > Stat Calculations > Linear Regression (a+bx).

4 In the X List: field, select time.

5 In the Y List: field, select temp.

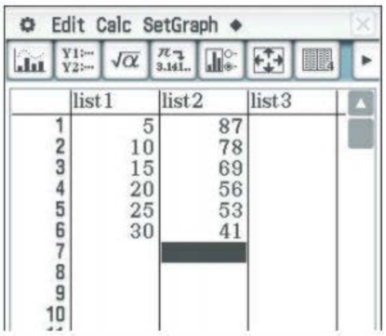
6 Select OK.

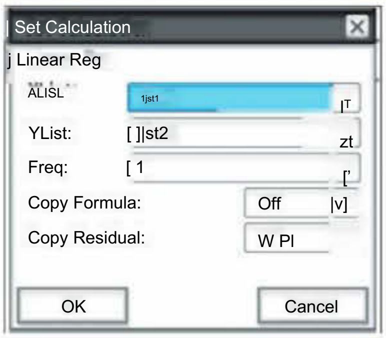
7 The linear regression labels and values will be displayed in columns C and D.

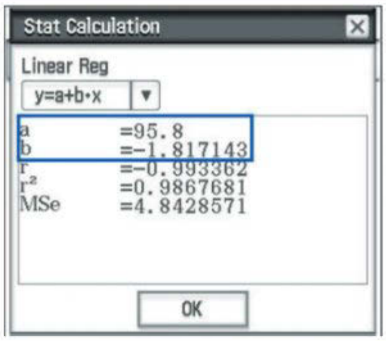
8 Rounding to two significant figures, $a = 96$ and $b$ —1.8.

The equation of the least squares line of best fit is: *temperature* = 96 - 1.8 x *time.*

**ClassPad**




1 Open the Statistics application.

2 Clear all lists and enter the data from the table as shown.

3 Tap Calc > Regression > Linear Reg.

4 Leave the XList: and YList: default settings of listl and Iist2 as shown above.

5 Tap OK.

6 Ensure the Linear Reg field is set to y=a+bx from the dropdown menu.

7 Rounding to two significant figures, $a = 96$ and $b = -1.8$.

The equation of the least squares line of best fit is: *temperature* = 96 - 1.8 x *time.*

### Exam hack

Make sure you always write the line of best fit equation using the variables names. You will lose a mark if you write it using $x$ and $y$.



9780170448406

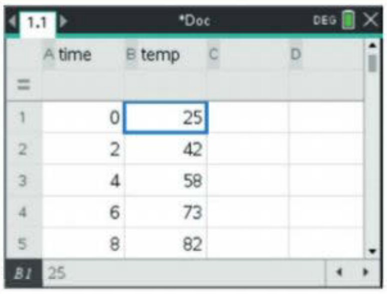
3.1

## USING CAS 2 Graphing the least squares line of best fit

Graph the least squares line of best fit for the data below, which measures the temperature of a liquid on a heat source as it heats up over time.

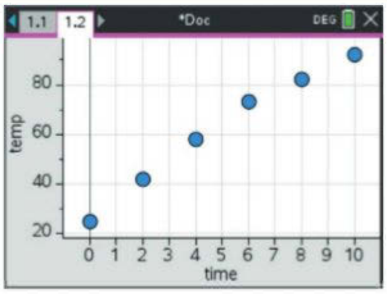
| Time (minutes) | 0 | 2 | 4 | 6 | 8 | 10 |
|---|---|---|---|---|---|---|
| Temperature (°C) | 25 | 42 | 58 | 73 | 82 | 92 |

### TI-Nspire




1 Start a new document and add a Lists & Spreadsheet page.

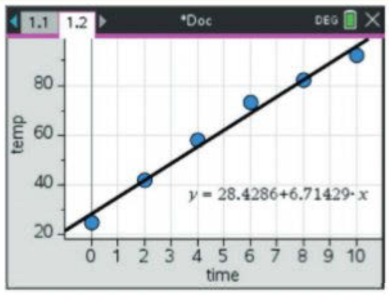
2 Label the columns and enter the data from the table as shown above.




3 Insert a Data & Statistics page.

4 For the horizonal axis, select time.

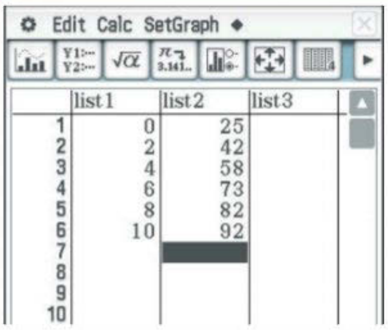
5 For the vertical axis, select temp.




6 Press menu > Analyze > Regression > Show Linear (a+bx).

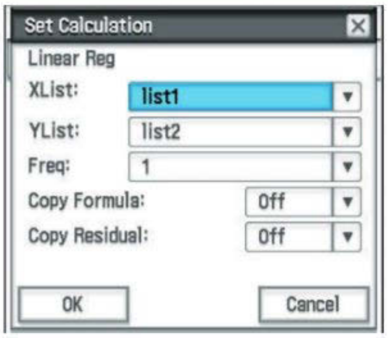
7 The least squares line of best fit will be displayed.

### ClassPad




1 Open the Statistics application.

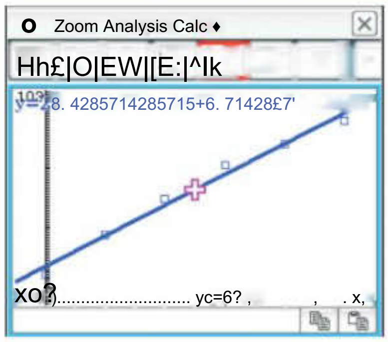
2 Clear all lists and enter the data from the table as shown.




3 Tap Calc > Regression > Linear Reg.

4 Tap OK.

5 On the next screen displaying the values, tap OK.




6 The data points and the line of best fit will be displayed (in the lower window).

7 Tap the Equation tool to display the equation of the line of best fit.

## Interpreting the least squares line of best fit

Let's look again at the equation of the least squares line of best fit line:

*temperature* = 96 - 1.8 x *time*

where the *temperature* (°C) of a liquid was measured as it cools down over *time* (minutes) after the source of heat was removed.

The intercept is 96. This tells us what the *temperature* was when we started measuring (i.e.» when *time* = 0). So we know the initial temperature was 96°C when the source of heat was removed.

The slope is -1.8, which tells us the temperature on average decreases by 1.8°C for every one-minute increase in time.

**Interpreting the least squares line of best fit equation**

We interpret the equation of a least squares line of best fit, jy = $a + bx$, by saying:

- the $y$-intercept is $a$. This means the $y$ variable is $a$ units when the $x$ variable is zero units.
- the slope is $b$. This means the $y$ variable on average increases/decreases by $b$ units for every 1-unit increase in the $x$ variable.

Use the word 'increases' when $b$ is positive and 'decreases' when $b$ is negative.

Replace the words in bold with the appropriate variable names or units of measure for each variable.

$x$ is the explanatory variable and $y$ is the response variable.

WB

p. 34

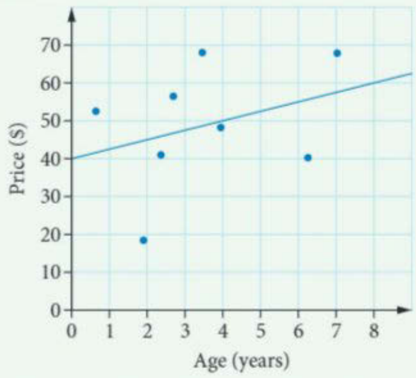
**WORKED EXAMPLE 2** Finding the least squares line of best fit from the graph

A study of the association between the age and price of a particular wine has resulted in the following least squares line of best fit fitted to a scatterplot. Find

**a** the $y$-intercept and interpret what it means

**b** the slope rounded to three significant figures and interpret what it means

**c** the least squares line of best fit equation.




| Steps | Working |
|---|---|
| **a** **1** Where does the line of best fit cross the $y$-axis? | The $y$-intercept is 40. |
| **2** Interpret what the $y$-intercept means in terms of the variables and units. | The price of the wine was $40 when it was released. |
| **b** **1** Choose two easy-to-identify points on the line. | (0, 40) and (8, 60) |
| **2** Calculate the slope from these two points and give the answer to the required rounding. | slope = $\frac{\text{rise}}{\text{run}} = \frac{\text{vertical distance from 40 to 60}}{\text{horizontal distance from 0 to 8}} = \frac{20}{8}$ = 2.5 = 2.50 rounded to 3 significant figures |
| **3** Interpret what the slope means in terms of the variables and units. | The price of the wine on average increases by $2.50 for every one-year increase in age. |
| **c** The least squares line of best fit equation is $y = a + bx$ where $a$ is $y$-intercept and $b$ is the slope. Use the variables given instead of $x$ and $y$. | *price* = 40 + 2.50 × *age* |

3.1

### WORKED EXAMPLE 3 Interpreting the least squares line of best fit equation

For the following least squares line of best fit equation

$$hand\ span = -11 + 0.19 \times height$$

where hand span and height are measured in centimetres,

a identify and interpret the slope

b identify and interpret the y-intercept and comment on your result.

| Steps | Working |
|---|---|
| a The slope is the value that height is multiplied by in the equation.<br>Interpret the slope in terms of the variables and units. | The slope is +0.19.<br>This means on average hand span increases by 0.19 cm for every 1-cm increase in height. |
| b Identify and interpret the y-intercept and comment. | The y-intercept is -11.<br>This means that hand span is -11 cm when height is 0 cm. A person of zero height can't have a -11 cm hand span. This least squares line of best fit only applies from a certain minimum height. |

WB p. 35

## EXERCISE 3.1 The least squares line of best fit

ANSWERS p. 700

### Mastery

1 ▶ WORKED EXAMPLE 1 For each of the following

i state the explanatory and response variables

ii write the values for $\bar{x}$, $s_x$, $\bar{y}$, $s_y$ and $r$

iii calculate the equation for the least squares line of best fit that models this data, rounding the coefficients $a$ and $b$ to two significant figures.

a Temperatures at sunset during summer (°C) and the volume of cicadas in decibels (dB) in a field were recorded and the values of the following statistics were determined.

The sample mean of the temperatures was 25.3°C.
The sample standard deviation of the temperatures was 5.6 °C.
The sample mean of the volume of cicadas was 83 dB.
The sample standard deviation of the volume of cicadas was 7 dB.
The Pearson correlation coefficient was 0.79.

b The annual number of bushfires and annual rainfall (cm) were recorded for a region and the values of the following statistics were determined.

The sample mean of the annual number of bushfires was 18.38.
The sample standard deviation of the annual number of bushfires was 8.16.
The sample mean of the annual rainfall was 19.88 cm.
The sample standard deviation of the annual rainfall was 6.31 cm.
The Pearson correlation coefficient was -0.88.

#### Exam hack

When finding the equation of the least squares line of best fit, decide which is the explanatory (x) variable and which is the response (y) variable. Don't assume that the first variable listed is always the explanatory variable.

▶ 2H Using CAS 1 JS using CAS 2~] The following table shows the results of an investigation by a home loan company into the association between the interest rate and the number of loan applications they had in eight consecutive years.

| Interest rate (p.a.) | 8.3 | 9.7 | 10.4 | 9.5 | 8.1 | 9.1 | 10.8 | 10.0 |
|---|---|---|---|---|---|---|---|---|
| Number of applications | 55 | 46 | 29 | 36 | 47 | 45 | 26 | 32 |

a Find the equation of the least squares line of best fit, rounding the coefficients *a* and *b* to two significant figures.

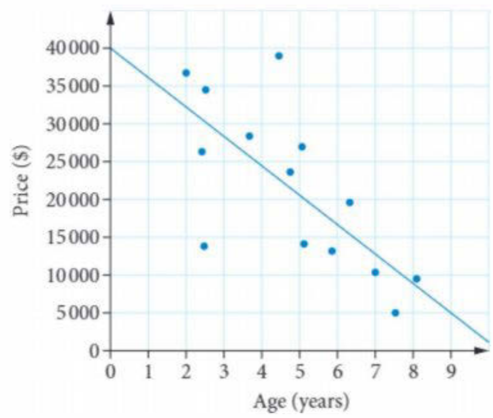
b Graph the least squares line of best fit for the data using CAS.

3B WORKED EXAMPLE 2 A study of the association between the age and price of a particular car model has resulted in a least squares line of best fit fitted to a scatterplot. Find

a the *y*-intercept and interpret what it means

b the slope rounded to three significant figures and interpret what it means

c the least squares line of best fit equation.




4 Ej WORKED EXAMPLE V For the following least squares line of best fit equation

*weight* = -44 + 0.7 x *height*

where height is measured in centimetres and weight is measured in kilograms,

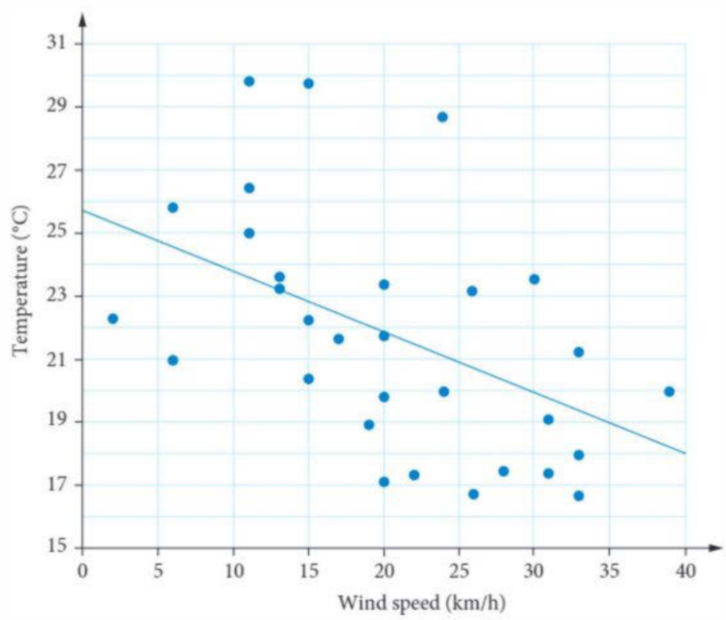
a identify and interpret the slope

b identify and interpret the *y*- intercept and comment on your result.

**Exam practice**

80-100% 60-79% 0-59%

**5** ©VCAA | 2012 1CQ8 73% The maximum wind speed and maximum temperature were recorded each day for a month. The data is displayed in the scatterplot and a least squares line of best fit has been fitted. The response variable is *temperature.* The explanatory variable is *wind speed.* The equation of the least squares line of best fit is closest to




A *temperature* = 25.7 - 0.191 x *wind speed*

B *wind speed* = 25.7 - 0.191 x *temperature*

C *temperature* = 0.191 + 25.7 x *wind speed*

D *wind speed* = 25.7 + 0.191 x *temperature*

E *temperature* = 25.7 + 0.191 x *wind speed*

3.1

▶ 6 ©VCAA J 20151CQ10 J 69% For a set of bivariate data that involves the variables $x$ and 7:

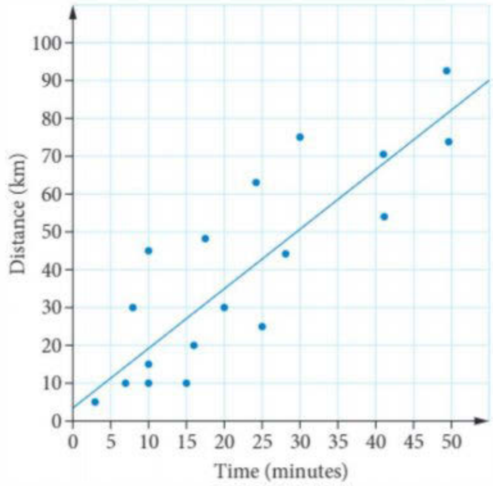
$r$ = -0.47, * = 1.8, $s_x$ = 1.2, 7 = 7.2, $s_y$ = 0.85

Given this information, the least squares line of best fit predicting $y$ from $x$ is closest to

A 7 = 8.4 - 0.66* B 7 = 8.4 + 0.66* C 7=7.8-0.33*

D 7=7.8 + 0.33* E 7= 1.8+ 5.4*

7 ©VCAA | 20151CQ9 I 62% A least squares line of best fit has been fitted to the scatterplot to enable *distance,* in kilometres, to be predicted from *time,* in minutes.




The equation of this line is closest to

A *distance* = 3.5 + 1.6 x *time*

B *time* = 3.5 + 1.6 x *distance*

C *distance* = 1.6 + 3.5 x *time*

D *time* = 1.8 + 3.5 x *distance*

E *distance* = 3.5 + 1.8 x *time*

8 ©VCAA J 20141CQ9I 58% The equation of a least squares line of best fit is used to predict the fuel consumption, in kilometres per litre of fuel, from a cars weight, in kilograms. The equation predicts that a car weighing 900 kg will travel 10.7 km per litre of fuel, while a car weighing 1700 kg will travel 6.7 km per litre of fuel. The slope of this least squares line of best fit is closest to

A -250 B -0.005 C -0.004 D 0.005 E 200

9 ©VCAA 20181CQ10 51% In a study of the association between a persons *height,* in centimetres, and *body surface area,* in square metres, the following least squares line was obtained.

*body surface area* = -1.1 +0.019x *height*

Which one of the following is a conclusion that can be made from this least squares line?

A An increase of 1 $m^2$ in *body surface area* is associated with an increase of 0.019 cm in *height.*

B An increase of 1 cm in *height* is associated with an increase of 0.019 $m^2$ in *body surface area.*

C The correlation coefficient is 0.019.

D A persons *body surface area,* in square metres, can be determined by adding 1.1 cm to their *height.*

E A persons *height,* in centimetres, can be determined by subtracting 1.1 from their *body surface area,* in square metres.

10 ©VCAA 20181CQ13J 49% The statistical analysis of a set of bivariate data involving variables * and 7 resulted in the information displayed in the table.

| | | |
|---|---|---|
| Mean | $\bar{x} = 27.8$ | $\bar{y} = 33.4$ |
| Standard deviation | $s_x = 2.33$ | |
| Equation of the least squares line | $y = -2.84 + 1.31x$ | |

Using this information, the value of the correlation coefficient $r$ for this set of bivariate data is closest to

A 0.88 B 0.89 C 0.92 D 0.94 E 0.97

▶ 11 ©VCAA | 2Q2O1CQ13^] 34% A least squares line of the form $y = a + bx$ is fitted to a scatterplot. Which one of the following is always true?

A As many of the data points in the scatterplot as possible will lie on the line.

B The data points in the scatterplot will be divided so that there are as many data points above the line as there are below the line.

C The sum of the squares of the shortest distances from the line to each data point will be a minimum.

D The sum of the squares of the horizontal distances from the line to each data point will be a minimum.

E The sum of the squares of the vertical distances from the line to each data point will be a minimum.

12 ©VCAA 2019 2CQ4 J (3 marks) The relative humidity (%) at 9 am and 3 pm on 14 days in November 2017 is shown in the table.

A least squares line is to be fitted to the data with the aim of predicting the relative humidity at 3 pm *(humidity 3pm)* from the relative humidity at 9 am *(humidity 9am)*.

| Relative humidity (%) | |
|---|---|
| 9 am | 3 pm |
| 100 | 87 |
| 99 | 75 |
| 95 | 67 |
| 63 | 57 |
| 81 | 57 |
| 94 | 74 |
| 96 | 71 |
| 81 | 62 |
| 73 | 53 |
| 53 | 54 |
| 57 | 36 |
| 77 | 39 |
| 51 | 30 |
| 41 | 32 |

Data: Australian Government, Bureau of Meteorology, www.bom.gov.au

a | 84% Name the explanatory variable. 1 mark

b | 59% Determine the values of the intercept and the slope of this least squares line. Round both values to three significant figures. Copy the equation below and write your answers in the appropriate boxes. 1 mark

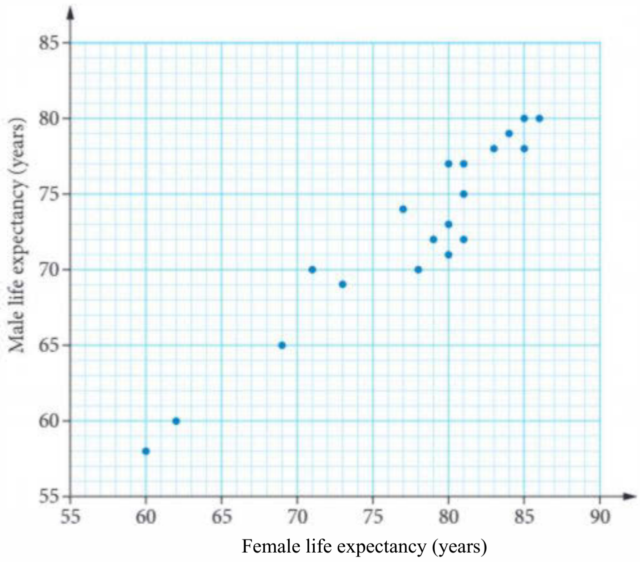
*humidity 3pm* = [ ] + [ ] x *humidity 9am*

c 76% Determine the value of the correlation coefficient for this data set.

Round your answer to three decimal places. 1 mark ▶


9780170448406

▶ 13 ©VCAA 2015 2CQ4 J (2 marks) The table shows male life expectancy *(male)* and female life expectancy *(female)* for a number of countries in 2013. The scatterplot has been constructed from this data.




| Life expectancy (in years) in 2013 | |
|---|---|
| Male | Female |
| 80 | 85 |
| 60 | 62 |
| 73 | 80 |
| 70 | 71 |
| 70 | 78 |
| 78 | 83 |
| 77 | 80 |
| 65 | 69 |
| 74 | 77 |
| 70 | 78 |
| 75 | 81 |
| 58 | 60 |
| 80 | 86 |
| 69 | 73 |
| 79 | 84 |
| 72 | 81 |
| 78 | 85 |
| 72 | 79 |
| 77 | 81 |
| 71 | 80 |

3.1

a 63% Use the scatterplot to describe the association between *male* life expectancy *and female* life expectancy in terms of strength, direction and form. 1 mark

b 49% Determine the equation of a least squares line of best fit that can be used to predict *male* life expectancy *from female* life expectancy for the year 2013. Copy and complete the equation for the least squares line of best fit by writing the intercept and slope in the boxes. Write these values correct to two decimal places.

*male* = ☐ + ☐ x *female* 1 mark

14 ©VCAA 2008 2CQ4 I (4 marks) The arm spans (in cm) and heights (in cm) for a group of 13 boys have been measured. The results are displayed in the table.

The aim is to find an equation that allows arm span to be predicted from height.

| Arm span (cm) | Height (cm) |
|---|---|
| 152 | 152 |
| 153 | 155 |
| 174 | 168 |
| 141 | 149 |
| 170 | 172 |
| 165 | 168 |
| 163 | 163 |
| 155 | 157 |
| 165 | 165 |
| 152 | 150 |
| 143 | 146 |
| 156 | 153 |
| 174 | 174 |

a 51% What will be the explanatory variable in the equation? 1 mark

b 51% Assuming a linear association, determine the equation of the least squares line of best fit that enables *arm span* to be predicted from *height*. Write this equation in terms of the variables *arm span* and *height*. Give the coefficients rounded to two significant figures. 2 marks

c 51% Using the equation that you have determined in part b, interpret the slope of the least squares line of best fit in terms of the variables *height* and *arm span*. 1 mark

# 3.2 The coefficient of determination

Video playlist The coefficient of determination

Worksheet Coefficient of determination

## Calculating the coefficient of determination

Now that we have used the least squares line of best fit to model sets of data, we should ask ourselves, 'How well does our line of best fit actually represent our set of data?' To answer this question, we use the coefficient of determination, or $r^2$.

**The coefficient of determination**

The coefficient of determination ($r^2$)

- can be calculated by squaring the Pearson correlation coefficient (r)
- is a value between 0 and 1
- is a measure of how useful a line of best fit is as a linear model for a particular set of data (0 means it is a totally useless measure and 1 means it is a perfect measure).

The higher the coefficient of determination

- the stronger the association between the variables
- the better the line of best fit is as a model for the data.


9780170448406

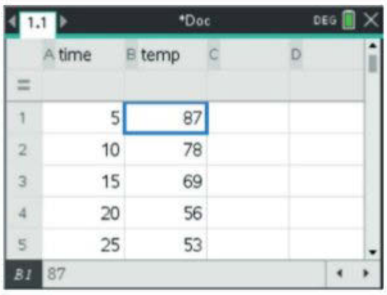
## USING CAS 3 Finding the coefficient of determination

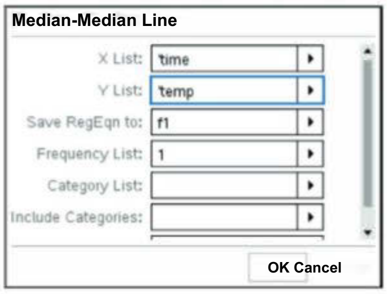
To calculate the coefficient of determination from a set of data, calculate the linear regression data. The r[1] ² value is one of the values that appears on the screen along with the values *a* and *b* for the least squares line of best fit equation.

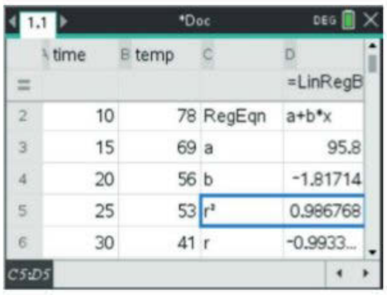
| Time (minutes) | 5 | 10 | 15 | 20 | 25 | 30 |
|---|---|---|---|---|---|---|
| Temperature (°C) | 87 | 78 | 69 | 56 | 53 | 41 |

*Time* is the explanatory variable and *temperature* is the response variable.

### TI-Nspire




1 Start a new document and add a Lists & Spreadsheet page.

2 Label the columns and enter the data from the table as shown above.

3 Press menu > Statistics > Stat Calculations > Linear Regression (a+bx).

4 In the X List: field, select time.

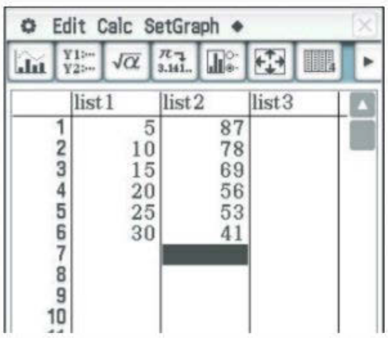
5 In the Y List: field, select temp.

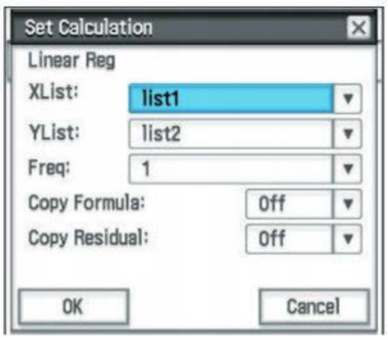
6 Select OK.

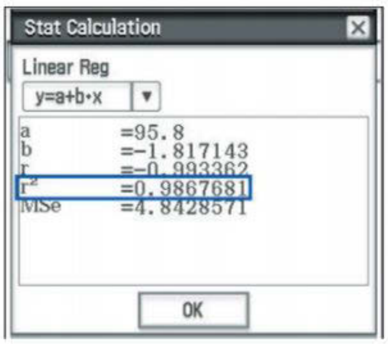
7 The linear regression labels and values will be displayed in columns C and D.

8 Rounding to two decimal places, the coefficient of determination $r^2$ is equal to approximately 0.99.

### ClassPad




1 Open the Statistics application.

2 Clear all lists and enter the data from the table as shown.

3 Tap Calc > Regression > Linear Reg.

4 Leave the XList: and YList: default settings of listl and Iist2 as shown above.

5 Tap OK.

6 Ensure the Linear Reg field is set to y=a+bx from the dropdown menu.

7 Rounding to two decimal places, the coefficient of determination $r^2$ is equal to approximately 0.99.

## Interpreting the coefficient of determination

The coefficient of determination is usually given as a decimal between 0 and 1, but we can convert it to a percentage between 0 and 100%. This value tells us the percentage of the variation in the response variable that is explained by the explanatory variable.

For example, if $r^2 = 0.85$ for a data set comparing *height* (the explanatory variable) and *hand span* (the response variable), then we can say that 85% of the variation in hand span can be explained by the variation in height. Alternatively, we can say that 15% of the variation in the hand span is *not* explained by the variation in height.

### Interpreting the coefficient of determination

When interpreting the coefficient of determination, the following general sentence can be used:

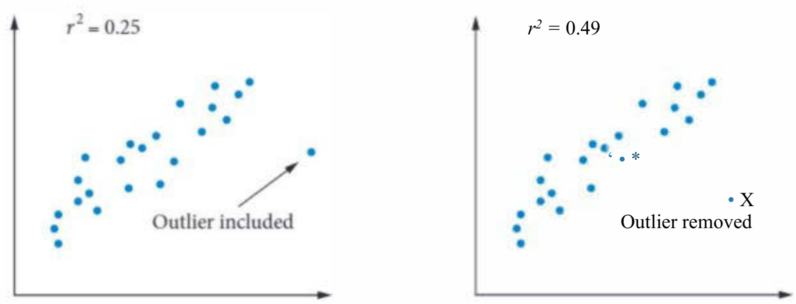
> $r^2$ x 100% of the variation in the response variable can be explained by the variation in the explanatory variable.

where the bold text is replaced with the appropriate percentage and variable names of the set of data being investigated.

A coefficient of determination of 70% and above is considered high. A high coefficient of determination indicates that the line of best fit is an appropriate model for the data.

### © Exam hack

The coefficient of determination measures the predictive power of the association and is affected by outliers. Removing an outlier increases the coefficient of determination. See the example below.




**WORKED EXAMPLE 4** Interpreting the coefficient of determination

WB
p. 36

Data was collected to investigate the association between the minimum daily temperature and the maximum daily temperature and is displayed in the table below.

| Minimum temperature (°C) | 12 | 15 | 13.5 | 14.1 | 10.2 | 11.7 | 13.2 | 11.3 |
|---|---|---|---|---|---|---|---|---|
| Maximum temperature (°C) | 23.4 | 25.6 | 23.1 | 25.3 | 20 | 21.1 | 22.6 | 21 |

a Assuming that *minimum temperature* is the explanatory variable, calculate the coefficient of determination, correct to three decimal places.

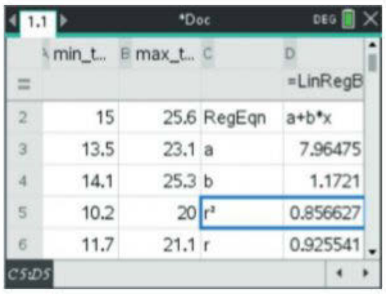
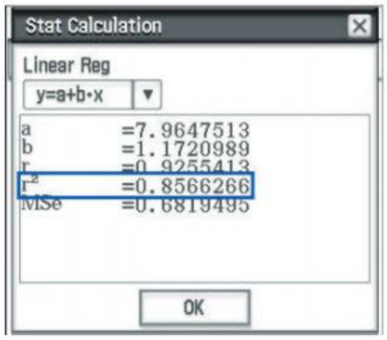
b Interpret the coefficient of determination.

c What percentage of variation in the *maximum temperature* is not explained by the variation in the *minimum temperature?* Round your answer to the nearest whole number.

d What is the least squares line of best fit equation that models this data? Round the coefficients *a* and *b* to two significant figures.

e Do you think that the model is appropriate? Justify your answer.


9780170448406

| Steps | Working |
|---|---|
| a We are told the *minimum temperature* is the explanatory variable. If we weren't told, we would have to determine this. Use a CAS to find $r^2$. The *a* and *b* values for the least squares line of best fit equation will also be calculated. | **TI-Nspire**   **ClassPad**   $r^2 = 0.857$ |
| b Calculate $r^2$ x 100%, then interpret the result using the general sentence. | 85.7% of the variation in the *maximum temperature* can be explained by the variation in the *minimum temperature.* |
| c Subtract the coefficient of determination percentage from 100% and round to the nearest whole number. | 100 - 85.7 = 14.3<br>14% of variation in the *maximum temperature* is not explained by the variation in the *minimum temperature.* |
| d 1 Read the *a* and *b* values for the line of best fit equation from the screen and write the equation using these values. | $y$ = 7.964 75 + 1.1721* |
| 2 Round *a* and *b* to two significant figures. | $y$ = 8.0 + 1.2x |
| 3 Replace *x* and *y* in the equation with the correct variable names. | *maximum temperature* = 8.0 + 1.2 x *minimum temperature* |
| e Determine the appropriateness of the model, using $r^2$ to support your decision. | Yes, this is an appropriate model because of the high $r^2$ value of 85.7%. |

## Positive or negative association

The coefficient of determination is a squared number, so it will always be positive. It doesn't tell us about the direction (positive or negative) of the association. For example, if

$r$ - 0.54, then $r^2$ - $(0.54)^2$ = 0.2916

and also if

r = -0.54, then $r^2 = (-0.54)^2$ = 0.2916

So if we know, for example, that $r^2 = 0.2916$ then the Pearson correlation coefficient could be either

r - $\sqrt{0.2916}$ = 0.54

or

$r$ - -$\sqrt{0.2916}$ - -0.54

### © Exam hack

Your calculator will most likely only give the positive square root, so remember when finding $r$ from $r^2$, you need to use the extra information in the question to work out if r is positive or negative.

### Finding the correlation coefficient from the coefficient of determination

If we know the coefficient of determination $r^{1\ 2\ 3}$ and want to find the correlation coefficient r, we calculate the square root *and* decide whether it is positive or negative, by looking at either:

- the line of best fit equation to see if the slope $b$ is positive or negative, or
- the scatterplot to see if the association is positive or negative.

WB

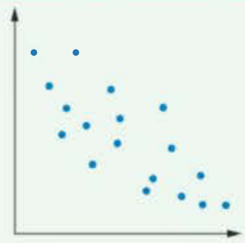
p. 37

**WORKED EXAMPLE 5** Finding the correlation coefficient from the coefficient of determination

Find the value of the Pearson correlation coefficient, correct to two decimal places, for each of the following.

**Steps** **Working**

a The coefficient of determination for the data displayed in the scatterplot is 0.78.



1 Calculate the square root of the coefficient of determination rounded to the given decimal places. Remember that this value could be positive or negative.

r=±V078

« ±0.88

2 Is the association shown on the scatterplot positive or negative?

The association is negative.

3 The association allows us to determine if $r$ is positive or negative.

r = -0.88

b The least squares line of best fit that enables the percentage mark on a test to be determined from the number of hours spent studying is

*% mark* = 15 + 9.5 x *study hours*

The coefficient of determination for this data is 0.92.

1 Calculate the square root of the coefficient of determination rounded to the given decimal places. Remember that this value could be positive or negative.

$r = \pm\sqrt{0.92}$

$\approx \pm 0.96$

2 Find the slope of the least squares line of best fit and check if it is positive or negative.

The slope of the least squares line of best fit is positive 9.5.

3 The sign of the slope allows us to determine if $r$ is positive or negative.

r=0.96



9780170448406

## EXERCISE 3.2 The coefficient of determination

ANSWERS p. 701

3.2

### Recap

*Use the following information to answer the next two questions.*

The table lists the average life span (in years) and average sleeping time (in hours/day) of 12 animal species.

| species | Life span (years) | Sleeping time (hours/day) |
|---|---|---|
| baboon | 27 | 10 |
| cow | 30 | 4 |
| goat | 20 | 4 |
| guinea pig | 8 | 8 |
| horse | 46 | 3 |
| mouse | 3 | 13 |
| Pig | 27 | 8 |
| rabbit | 18 | 8 |
| rat | 5 | 13 |
| red fox | 10 | 10 |
| rhesus monkey | 29 | 10 |
| sheep | 20 | 4 |

1 © VCAA 20091 co9 J 49% Using *sleeping time* as the explanatory variable, a least squares line of best fit is fitted to the data. The equation of the least squares line of best fit is closest to

A *life span* = 38.9 - 2.36 x *sleeping time.*

B *life span* = 11.7 - 0.185 x *sleeping time.*

C *life span* = -0.185 - 11.7 x *sleeping time.*

D *sleeping time* = 11.7 - 0.185 x *life span.*

E *sleeping time* = 38.9 - 2.36 x *life span.*

2 © VCAA 2009 1CQ10 70% The value of the Pearson correlation coefficient for *life span* and *sleeping time* is closest to

A -0.6603 B -0.4360 C -0.1901 D 0.4360 E 0.6603

### Mastery

3 EJ WORKED EXAMPLE 4 Using CAS 3 J The shoe sizes and heights of some students were measured and the results are shown below.

| Shoe size | 8 | 7.5 | 11 | 6 | 7 | 7.5 | 7 | 7.5 | 10 | 8 | 8 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Height (cm) | 169 | 167 | 182 | 162 | 163 | 157 | 168 | 171 | 180 | 171 | 168 |

a Assuming that *height* is the explanatory variable, calculate the coefficient of determination, correct to three decimal places.

b Interpret the coefficient of determination.

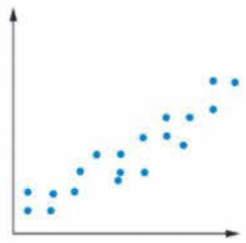
c What percentage of variation in the *height* is not explained by the variation in the *shoe size?* Round your answer to the nearest whole number.

d What is the least squares line of best fit equation that models this data? Round the coefficients $a$ and $b$ to two significant figures.

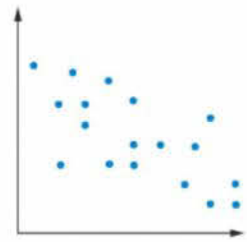
e Do you think the model is appropriate? Justify your answer.

4H WORKED EXAMPLE 5 Find the value of the Pearson correlation coefficient, correct to two decimal places, for each of the following:

a The coefficient of determination for the data displayed in the scatterplot is 0.91.



b The coefficient of determination for the data displayed in the scatterplot is 0.62.



c The least squares line of best fit that enables the distance travelled to be determined from the time spent travelling is

*distance travelled* =31 + 1.5 x *time*

The coefficient of determination for this data is 0.88.

d The least squares line of best fit that enables a persons score on a fitness test to be determined by their weight is

*fitness test score* = 142 - 2.3 x *weight*

The coefficient of determination for this data is 0.31.

**Exam practice** 80–100% 60–79% 0–59%

*Use the following information to answer the next two questions.*

A least squares line is used to model the relationship between the monthly *average temperature* and *latitude* recorded at seven different weather stations. The equation of the least squares line is found to be

*average temperature* = 42.9842 - 0.877 447 x *latitude*

5 © VCAA 2019 1Q9 67% When the numbers in this equation are correctly rounded to three significant figures, the equation will be

A *average temperature* = 42.984 - 0.877 x *latitude*

B *average temperature* = 42.984 - 0.878 x *latitude*

C *average temperature* = 43.0 - 0.878 x *latitude*

D *average temperature* = 42.9 - 0.878 x *latitude*

E *average temperature* = 43.0 - 0.877 x *latitude*

6 ©VCAA 20191 QI Oj 41% The coefficient of determination was calculated to be 0.893 743.

The value of the correlation coefficient, rounded to three decimal places, is

A -0.945 B -0.898 C 0.806 D 0.898 E 0.945

 9780170448406

7 ©VCAA 2011 1CQ8 58% When blood pressure is measured, both the systolic (or maximum) pressure and the diastolic (or minimum) pressure are recorded. The table displays the blood pressure readings, in mmHg, that result from fifteen successive measurements of the same persons blood pressure.

| | Blood pressure | |
|---|---|---|
| Reading number | systolic | diastolic |
| 1 | 121 | 73 |
| 2 | 126 | 75 |
| 3 | 141 | 73 |
| 4 | 125 | 73 |
| 5 | 122 | 67 |
| 6 | 126 | 74 |
| 7 | 129 | 70 |
| 8 | 130 | 72 |
| 9 | 125 | 69 |
| 10 | 121 | 65 |
| 11 | 118 | 66 |
| 12 | 134 | 77 |
| 13 | 125 | 70 |
| 14 | 127 | 64 |
| 15 | 119 | 69 |

From the fifteen blood pressure measurements for this person, it can be concluded that the percentage of the variation in systolic blood pressure that is explained by the variation in diastolic blood pressure is closest to

A 25.8% B 50.8%

C 55.4% D 71.9%

E 79.0%

3.2

8 ©VCAA 20071CQ7 I 55% The lengths and diameters (in mm) of a sample of jellyfish were recorded and displayed in the scatterplot. The least squares line of best fit for this data is shown.

The equation of the least squares line of best fit is

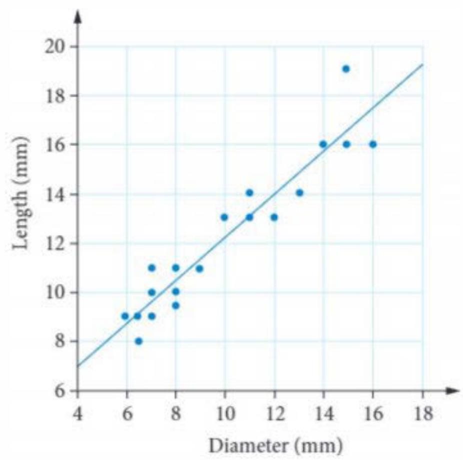
*length* = 3.5 + 0.87 x *diameter*

The correlation coefficient is r = 0.9034.

Written as a percentage, the coefficient of determination is closest to

A 0.816% B 0.903% C 81.6%

D 90.3% E 95.0%



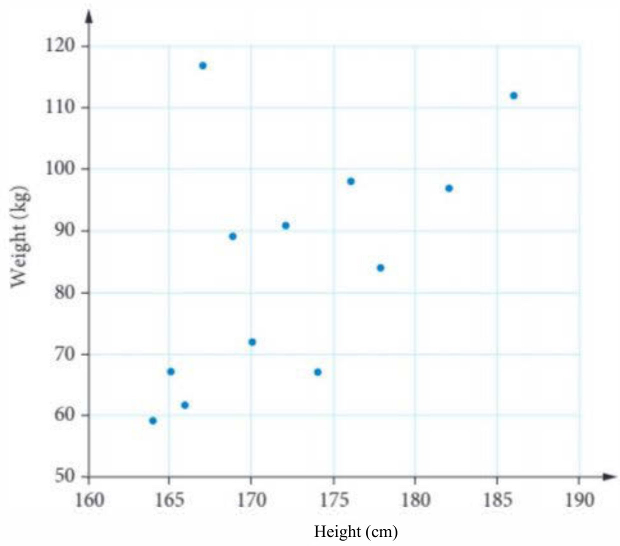

9 ©VCAA | 20021CQ12J 52% A persons weight is known to be positively associated with their height. To investigate this association for 12 men, a scatterplot is constructed as shown.

While there is a moderately strong positive linear association between weight and height, there is a clear outlier. When a least squares line of best fit is used to model this data, the coefficient of determination is found to be 0.3146. If the outlier is removed from the data, and a least squares line of best fit refitted to the data of the remaining 11 men, the value of the coefficient of determination will

A remain the same. B increase.

C decrease. D be halved.

E not be able to be determined.

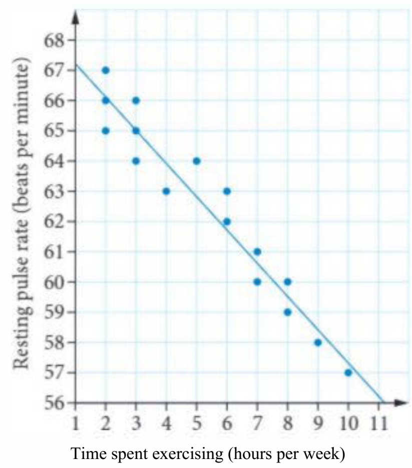


**10** ©VCAA 2018 1CQ9 45% The scatterplot displays the *resting pulse rate,* in beats per minute, and the *time spent exercising,* in hours per week, of 16 students. A least squares line has been fitted to the data.

The coefficient of determination is 0.8339. The correlation coefficient r is closest to

A -0.913 B -0.834 C -0.695

D 0.834 E 0.913




**11** ©VCAA 20031CQ9 33% Eighteen students sat for a 15-question multiple-choice test. In the scatterplot, the number of errors made by each student on the test is plotted against the time they reported studying for the test. A least squares line of best fit has been determined for this data and is also displayed on the scatterplot.

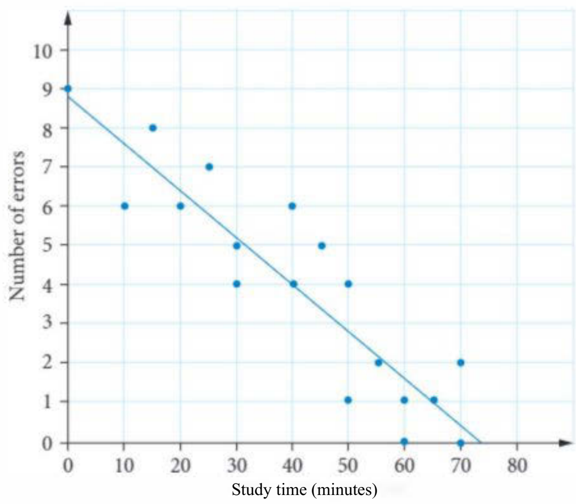
The equation for the least squares line of best fit is *number of errors* = 8.8 - 0.120 x *study time* and the coefficient of determination is 0.8198.

The value of the Pearson correlation coefficient, *r,* for this data, correct to two decimal places, is

A -0.91 B -0.82

C 0.67 D 0.82

E 0.91




12 ©VCAA 2020 2CQ4 (5 marks) The *age,* in years, *body density,* in kilograms per litre, and *weight,* in kilograms, of a sample of 12 men aged 23 to 25 years are shown in the table.

| Age (years) | Body density (kg/litre) | Weight (kg) |
|---|---|---|
| 23 | 1.07 | 70.1 |
| 23 | 1.07 | 90.4 |
| 23 | 1.08 | 73.2 |
| 23 | 1.08 | 85.0 |
| 24 | 1.03 | 84.3 |
| 24 | 1.05 | 95.6 |
| 24 | 1.07 | 71.7 |
| 24 | 1.06 | 95.0 |
| 25 | 1.07 | 80.2 |
| 25 | 1.09 | 87.4 |
| 25 | 1.02 | 94.9 |
| 25 | 1.09 | 65.3 |

### iOl Exam hack

Don't round if you're not asked to.

a For these 12 men, determine

j (97%) their median *age,* in years 1 mark

ii (77%) the mean of their *body density,* in kilograms per litre. 1 mark



9780170448406

b A least squares line is to be fitted to the data with the aim of predicting *body density* from *weight.*

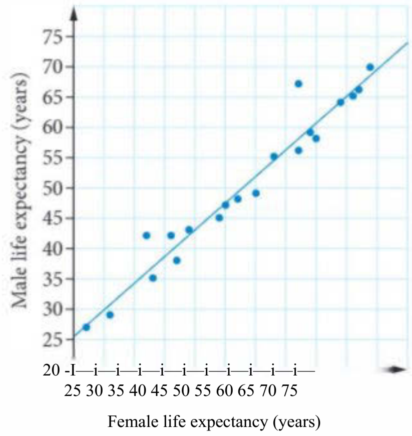
i 78% Name the explanatory variable for this least squares line. 1 mark

ii 29% Determine the slope of this least squares line. Round your answer to three significant figures. 1 mark

c 54% What percentage of the variation in *body density* can be explained by the variation in *weight?* Round your answer to the nearest percentage. 1 mark

13 ©VCAA 2015 2CQ3 (3 marks) The scatterplot plots male life expectancy *(male)* against female life expectancy *(female)* in 1950 for a number of countries. A least squares line of best fit has been fitted to the scatterplot as shown.

The slope of this least squares line of best fit is 0.88.




a 40% Interpret the slope in terms of the variables *male* life expectancy and *female* life expectancy. 1 mark

b 78% The equation of this least squares line of best fit is

*male* = 3.6 + 0.88 x *female*

In a particular country in 1950.*female* life expectancy was 35 years. Use the equation to predict *male* life expectancy for that country. 1 mark

c 52% The coefficient of determination is 0.95. Interpret the coefficient of determination in terms of male life expectancy and female life expectancy. 1 mark

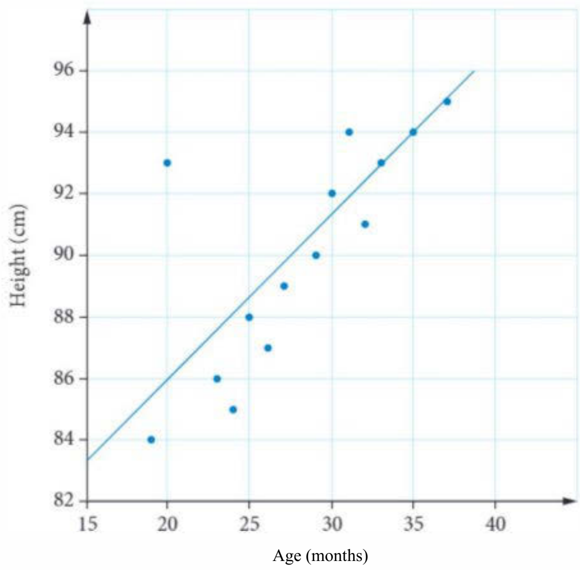
14 ©VCAA I 2006 2CQ2 I 44% (3 marks)

The heights (in cm) and ages (in months) of a random sample of 15 boys have been plotted in the scatterplot. The least squares line of best fit has been fitted to the data.

The equation of the least squares line of best fit is

*height* = 75.4 + 0.53 x *age*

The correlation coefficient is r = 0.7541.




## Exam hack

When you are asked to round your answer to a certain number of decimal places, don't round anything until the very last step.

a Copy and complete the following sentence.

On average, the height of a boy increases by ____ cm for each one-month increase in age. 1 mark

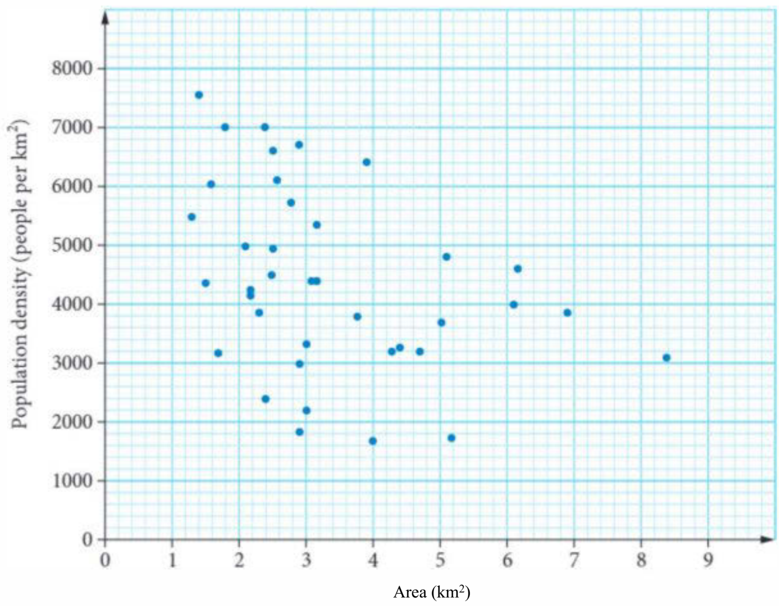
b i Evaluate the coefficient of determination.

Write your answer, as a percentage, correct to one decimal place. 1 mark

ii Interpret the coefficient of determination in terms of the variables *height* and *age.* 1 mark ►

15 ©VCAA J 2014 2CQ4 J (4 marks) The scatterplot shows the *population density,* in people per square kilometre, and the *area,* in square kilometres, of 38 inner suburbs of a city.

For this scatterplot, $r^{i\,2} = 0.141$.




a 30% Describe the association between the variables *population density* and *area* for these suburbs in terms of strength, direction and form. 1 mark

b 42% The mean and standard deviation of the variables *population density* and *area* for these 38 inner suburbs are shown in the table.

| | Population density (people per km$^2$) | Area (km$^2$) |
|---|---|---|
| Mean | 4370 | 3.4 |
| Standard deviation | 1560 | 1.6 |

i One of these suburbs has a population density of 3082 people per square kilometre.

Determine the standard z-score of this suburbs population density.

Write your answer correct to one decimal place. 1 mark

Assume the areas of these inner suburbs are approximately normally distributed.

ii How many of these 38 suburbs are expected to have an area that is two standard deviations or more above the mean?

Write your answer correct to the nearest whole number. 1 mark

iii How many of these 38 inner suburbs actually have an area that is two standard deviations or more above the mean? 1 mark

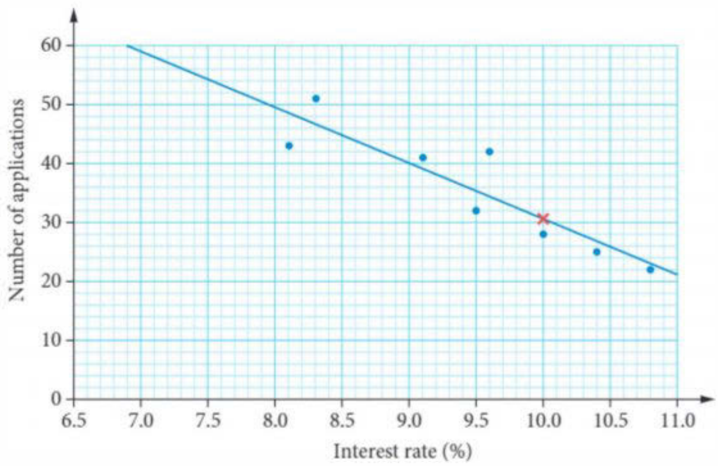

9780170448406

## 3.3 Making predictions

### Interpolation and extrapolation

It is possible to make predictions directly from a least squares line of best fit graph. For example, it is possible to predict from the graph below that an interest rate of 10% will have between 30 and 31 applicants. See the red cross on the graph. However, this method is not accurate. Its difficult to determine by eye whether the prediction is closer to 30 or 31.

A better way to predict values is to use the equation of the least squares line of best fit by substituting the given values into the equation and then solving. The equation can be used to predict *within* the original data range, which is called interpolation, as well as *outside* the original data range, which is called extrapolation.




Predictions based on extrapolation are not as reliable as those based on interpolation because we can't be certain that the equation applies to values outside the range of data values we have.

**WORKED EXAMPLE 6** Making predictions from the line of best fit

Data was collected from people aged between 7 and 19 years of age, and a least squares line of best fit was found to have the equation

*height* (cm) = 90 + 5 x *age* (years)

a Predict the height of a 15-year-old. Does this involve interpolation or extrapolation?

b Predict the height of a 24-year-old. Does this involve interpolation or extrapolation?

c Which of the predictions in parts a and b is more reliable? Justify your answer,

d Predict the age of a person of height 145 cm.

| Steps | Working |
|---|---|
| a Substitute the value into the equation in place of *age* and solve for *height*.<br>Was the value used within or outside the original data range? | *height* = 90 + 5 x *age*<br>= 90 + 5 x 15<br>= 165 cm<br>15 is within the original data range of 7 to 19 years, so this involves interpolation. |
| b Substitute the value into the equation in place of *age* and solve for *height*.<br>Was the value used within or outside the original data range? | *height* = 90 + 5 x *age*<br>= 90 + 5 x 24<br>= 210cm<br>24 is outside of the original data range of 7 to 19 years, so this involves extrapolation. |


3.3

Video playlist Making predictions

Worksheet Interpolation and extrapolation

WB p. 38

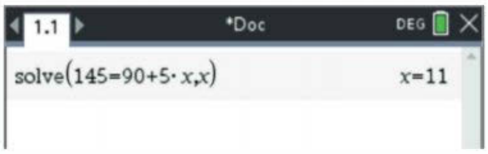
c Decide which of the two is the more reliable prediction and justify your decision.

The prediction for the 15-year-old involves interpolation, so it is more reliable than the prediction for a 24-year-old, which involves extrapolation.

d Substitute the value into the equation in place of *height* and solve for *age*, using CAS if necessary.

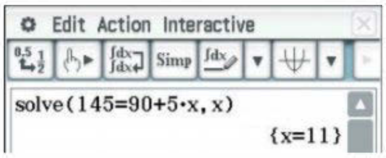
145 = 90 + 5 x *age*
5 x *age* = 145-90 (subtracting 90 from both sides)
5 x *age* = 55
*age* =55-5-5 (dividing both sides by 5)
*age* =11 years

**TI-Nspire**




1 Press menu > Algebra > Solve.
2 Enter the equation followed by ,x, where *x* represents the variable *age*.
3 Press enter.

**ClassPad**




1 Tap Interactive > Advanced > solve.
2 In the dialogue box, enter the equation in the Equation: field.
3 In the Variable field, keep the default variable x.
4 Tap OK.

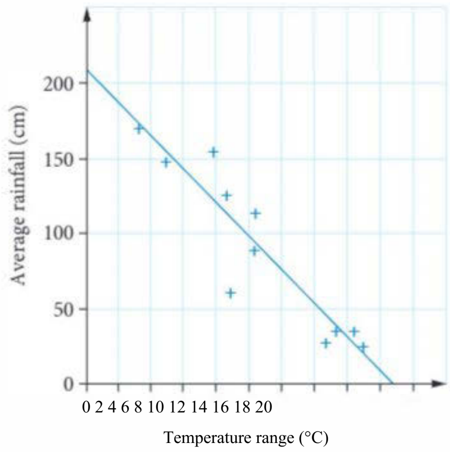
## EXERCISE 3.3 Making predictions

ANSWERS p. 701

### Recap

*Use the following information to answer the next two questions.*

The average rainfall and temperature range at several different locations in the South Pacific region are displayed in the scatterplot.




1 © VCAA 2004 1CQ8 72% A least squares line of best fit has been fitted to the data as shown.
The equation of this line is closest to

A *average rainfall* = 210-llx *temperature range*
B *average rainfall* = 210 + 11 x *temperature range*
C *average rainfall* = 18 - 0.08 x *temperature range*
D *average rainfall* = 18 + 0.08 x *temperature range*
E *average rainfall* = 250 - 13 x *temperature range*

2 © VCAA 2004 1CQ9 38% The value of the Pearson correlation coefficient, $r$, for the data, is $r = -0.9260$.
The value of the coefficient of determination is

A -0.9260 B -0.8575 C 0.8575 D 0.9260 E 0.9623


9780170448406

## Mastery

3 The least squares line of best fit with equation

*number of laps* = 10.87 - 0.16 x *age*

models the association between the number of laps run around an oval in 30 minutes and the age of the runner (in years). Use the model to predict, correct to two decimal places, the number of laps run in 30 minutes by a runner aged

a 12 b 34 c 55

**Exam hack**

When stating the predicted value, make sure that you include the units of measure as well.

3.3

4 The least squares line of best fit with the equation given below models data relating the outside temperature (°C) to the amount of gas consumed (kWh) by a household over a three-month period.

*gas consumed* = 2212 - 125 x *outside temperature*

Use the model to predict the outside temperature, to the nearest degree, when the gas consumed is

a 2000 kWh b 1200 kWh c OkWh

5 H WORKED EXAMPLE 6 I The weight (in grams) of boxes of chocolates containing between 15 and 50 chocolates were recorded. The least squares line of best fit for the data was found to have the equation

*weight* = 20 + 5 x *number of chocolates*

a Predict the weight of a box containing 25 chocolates. Does this involve interpolation or extrapolation?

b Predict the weight of a box containing 5 chocolates. Does this involve interpolation or extrapolation?

c Predict the weight of a box containing 70 chocolates. Does this involve interpolation or extrapolation?

d Which of these predictions is the most reliable? Justify your answer.

e Predict the number of chocolates that would be in a box weighing 125 grams.

## Exam practice

80-100% 60-79% 0-59%

*Use the following information to answer the next three questions.*

A least squares line of best fit modelling the height (in cm) of 20 children aged between 5 and 15 years was found to have the equation

*height* = 99.97 + 2.59 x *age*

6 The predicted height for a 14-year-old is

A 136 cm B 136.23 cm C 137.5 cm D 146 cm E 146.34 cm

7 The predicted age of a child who is 120.69 cm is

A 5 B 6 C 7 D 8 E 9

8 Which one of the following statements is *false?*

A Predicting the height of a child aged 9 years is an example of interpolation.

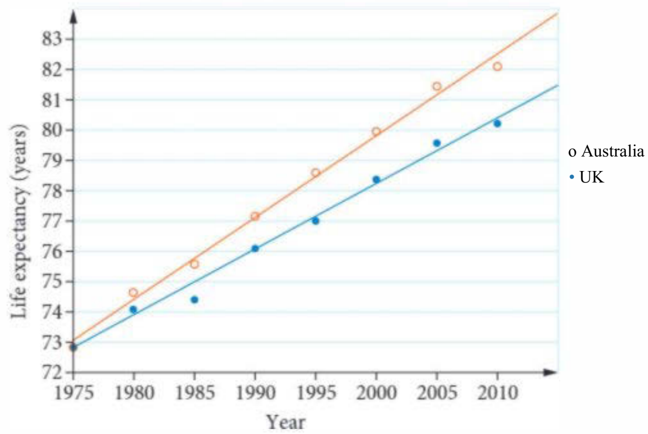
B For every one year of increase in age, the model predicts a child's height increases by 2.59 cm.

C Predicting the height of a 6-year-old will be more reliable than predicting the height of a 12-year-old.

D A child's height will increase as their age increases.

E Predicting the height of a 10-year-old is more likely to be reliable than predicting the height of an 18-year-old.

▶ 9 ©VCAA 2015 2CQ5b] (3 marks) In 1975, the life expectancies in Australia and the UK were very similar. From 1975, the gap between the life expectancies in the two countries increased, with people in Australia having a longer life expectancy than people in the UK. To investigate the difference in life expectancies, least squares lines of best fit were fitted to the data for both Australia and the UK for the period 1975 to 2010. The results are shown.




The equations of the least squares lines of best fit are as follows.

Australia: *life expectancy* = -451.7 + 0.2657 x *year*

UK: *life expectancy* = -350.4 + 0.2143 x *year*

a 72% Use these equations to predict the difference between the life expectancies of Australia and the UK in 2030. Give your answer correct to the nearest year. 2 marks

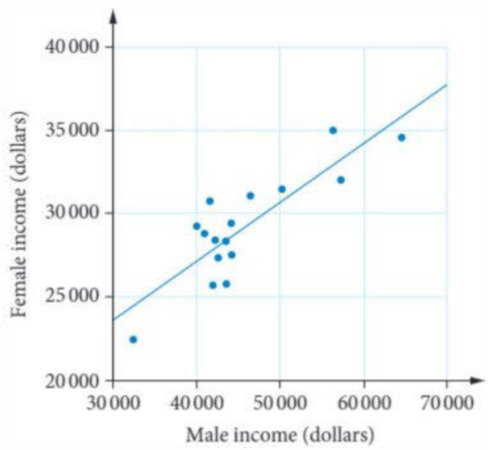
b 45% Explain why this prediction may be of limited reliability. 1 mark

10 ©VCAA 2010 2CQ2 57% (4 marks)

In the scatterplot, average annual *female income,* in dollars, is plotted against average annual *male income,* in dollars, for 16 countries. A least squares line of best fit is fitted to the data.

The equation of the least squares line of best fit for predicting female income from male income is

*female income* = 13000 + 0.35 x *male income*




a What is the explanatory variable? 1 mark

b Copy and then complete the following statement by filling in the missing information.

From the least squares line of best fit equation it can be concluded that, for these countries, on average, female income increases by $for each $1000 increase in male income. 1 mark

c i Use the least squares line of best fit equation to predict the average annual female income (in dollars) in a country where the average annual male income is $15000. 1 mark

ii The prediction made in part c i is not likely to be reliable. Explain why. 1 mark

**Exam hack**

For 'explain' questions, refer only to the information in the question. Don't offer your own personal views.

# @ Residual analysis

3.4

## Residual values

As we have seen, it is important to know whether the association between the variables is linear or not. We have used the following to help determine whether the association is linear:

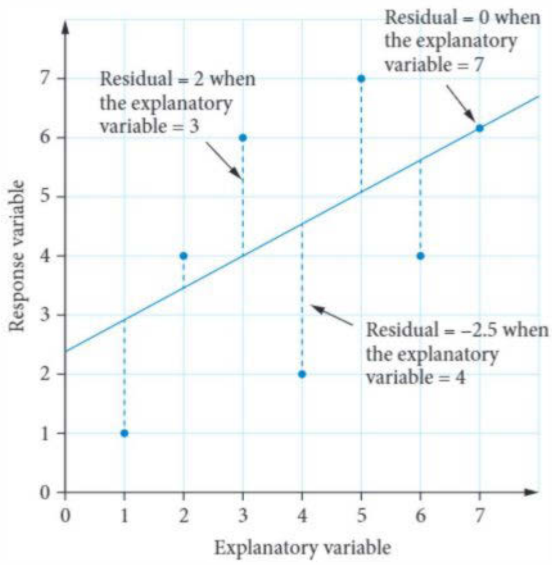
- the form of the scatterplot
- fitting a least squares line of best fit
- seeing how close the correlation coefficient, r, is to 1 or -1
- the value of the coefficient of determination, $r^2$.

However, one of the best ways to test the assumption that the association is linear is to calculate the residual values.

A residual is the vertical distance between each data point and the least squares line of best fit.

Video playlist Residual analysis

Worksheet Predictions and residuals




### Residual values and actual values

residual value = actual value - predicted value

The actual value can be read from a scatterplot or a table.

The predicted value needs to be calculated from the least squares line of best fit.

Actual values that lie

- *above* the least squares line of best fit will have a positive residual value
- *below* the least squares line of best fit will have a negative residual value
- *on* the least squares line of best fit will have a residual value of zero.

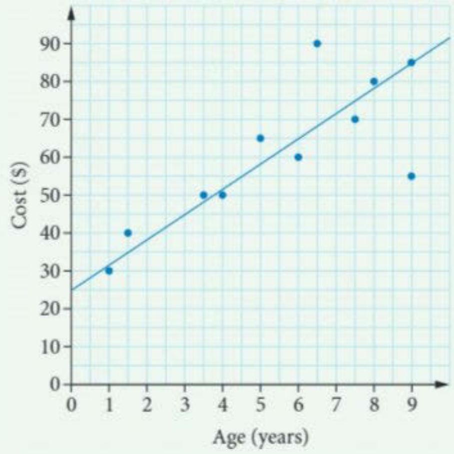
The further the actual value is from the least squares line of best fit, the larger the residual value and the less accurate the prediction.

WB

p. 39

**WORKED EXAMPLE 7** Finding residual values from the line of best fit

Find the residual values for each of the following.

| Steps | Working |
|---|---|
| a For a study done to establish the association between the age of a child (years) and the cost of their favourite toy ($), find the residual values for the following directly from the graph:<br>i Elke who is $1\frac{1}{2}$ years old.<br>ii Ethan who is 6 years old.<br>iii Cara, the nine-year old with the cheapest favourite toy.<br>iv Carson, the nine-year old with the most expensive favourite toy. | <br> |
| Read the vertical distance from the point to the line of best fit from the graph. The value is negative if it is below the line and zero if it is on the line. Include the response variables' units. | i Residual value for Elke = $5.00<br>ii Residual value for Ethan = -$5.00<br>iii Residual value for Cara = -$30.00<br>iv Residual value for Carson = $0.00 |

b For a study done to establish the association between the age of children (years) and their height (cm), find the residual values for the following from the least squares line of best fit equation, correct to one decimal place:

*height* = 100 + 2.5 x *age*

i Martha who is 12 years old and 142 cm tall.

ii Mo who is 11 years old and 118 cm tall.

| Steps | Working |
|---|---|
| 1 Calculate the predicted height by substituting the age into the line of best fit equation. | i Martha's predicted *height*<br>= 100 + 2.5 x *age*<br>= 100 +2.5 x 12<br>= 130cm<br>ii Mo's predicted *height*<br>= 100 + 2.5 x *age*<br>= 100 +2.5 x 11<br>= 127.5cm |
| 2 Find the residual value using the formula:<br>residual value = actual value - predicted value<br>Include the response variables' units. | i residual value = 142 - 130<br>= 12 cm<br>ii residual value =118- 127.5<br>= -9.5 cm |



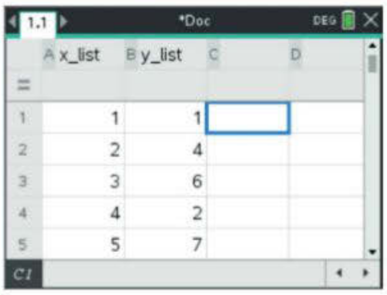
9780170448406

## USING CAS 4 Calculating residual values from a table

Create a table of residual values for the following set of data, stating all answers correct to one decimal place.

| $x$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $y$ | 1 | 4 | 6 | 2 | 7 | 4 | 5 |

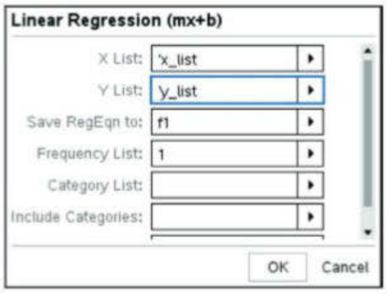
### TI-Nspire




1 Start a new document and add a **Lists & Spreadsheet** page.

2 Label the columns and enter the data from the table as shown above (avoid using $x$ and $y$ for list names).

3 Place the cursor in column **C**, which will be used later for the residual values.

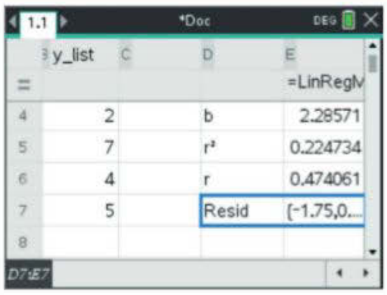



4 Press **menu** > **Statistics** > **Stat Calculations** > **Linear Regression (a+bx).**

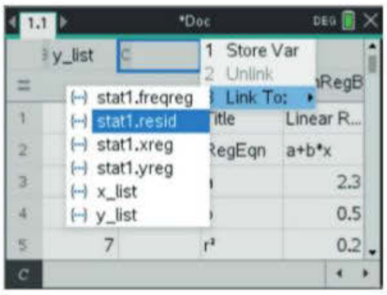
5 In the **X List:** field select **x_list**.

6 In the **Y List:** field select **y_list**.

7 Press **OK**.




8 Scroll down column **E** to view the residual values, which are now in a list. We will now link this list to column **C**.

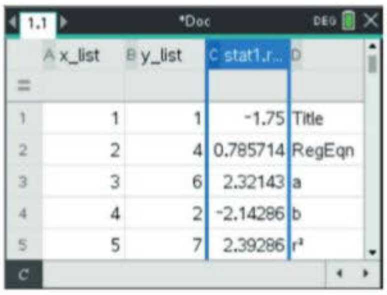



9 Place the cursor in the column **C** heading.

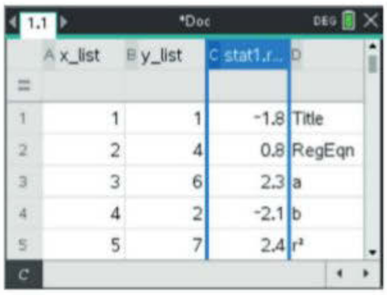
10 Press **var**.

11 Select **Link To:**.

12 From the drop down menu, select **stat1.resid**.



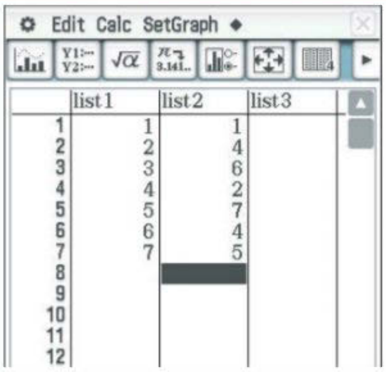

13 The residual values will be displayed in column **C**.




14 Change the **Document Settings** > **Display Digits** to **Fix 1** to display the residual values to one decimal place.

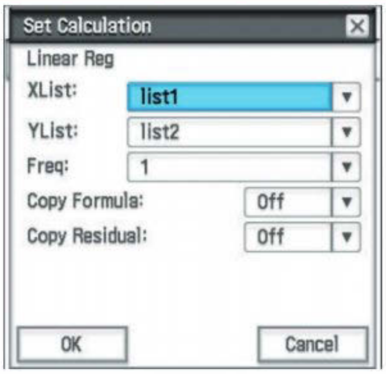
3.4

## ClassPad



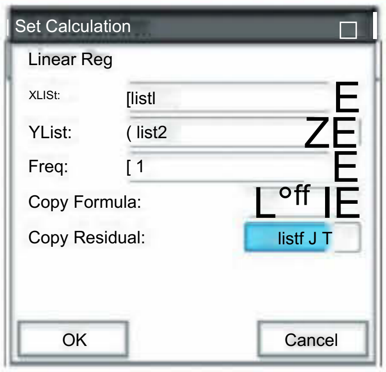

1 Open the Statistics application.

2 Clear all lists and enter the data from the table as shown.



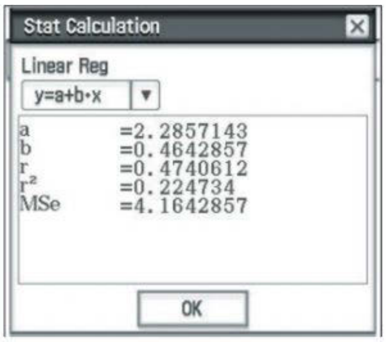

3 Tap Calc > Regression > Linear Reg.

4 Leave the XList: and YList: default settings of listl and Iist2.



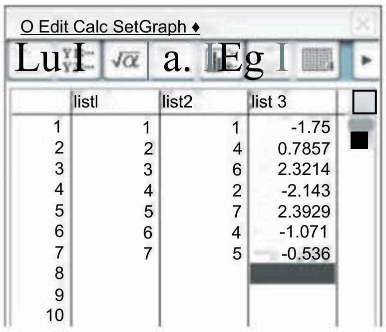

5 Change the Copy Residual: field from Off to Iist3.

6 Tap OK.




7 The linear regression results will appear as before.

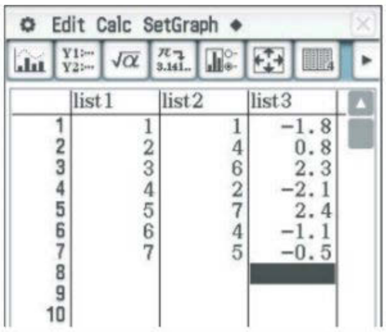
8 Tap OK.




9 Tap Menu and open the Statistics application again.

10 Tap in the upper window and select Resize to expand it.

11 Scroll left and the residual values will be displayed in Iist3.




12 Change the Basic Format > Number Format to Fix 1 to display the residual values to one decimal place.



9780170448406

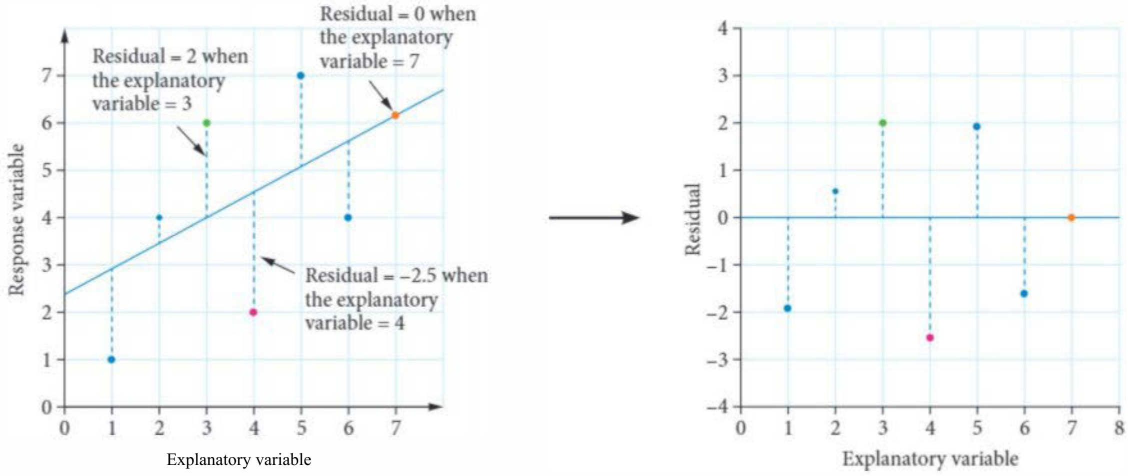
## Residual plots

Once all of the residual values have been found, a residual plot can be constructed. A residual plot is like a scatterplot with the explanatory variable on the x-axis and the residual values on the y-axis. Residual values are negative as well as positive so the residual plot will always have both positive and negative y-axes.

> **Exam hack**
>
> Think of the zero line in a residual plot as the line of best fit in a scatterplot made horizontal.

Worksheet
Residual plots

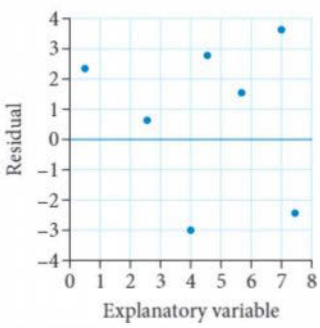
Scatterplot with line of best fit | Residual plot




Most residual plots will fall into one of the following three types:

### "Type 1: Randomly scattered

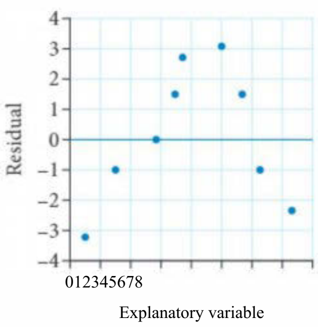
The residual values are randomly scattered above and below the x-axis. This lack of a clear pattern suggests that the association between the explanatory and response variables is *linear*.




> **Exam hack**
>
> For residual plots, a random scattering indicates that there is a *linear association* between explanatory and response variables. For a scatterplot, a random scattering indicates there is *no association* between the explanatory and response variables. Don’t get the two confused.

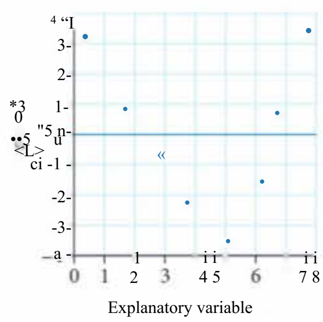
### Type 2: Hill

The residual values show a hill shape. This suggests that the association between the explanatory and response variables is *non-linear*.




### Type 3: Valley

The residual values show a V or V valley shape. This also suggests that the association between the explanatory and response variables is *non-linear*.

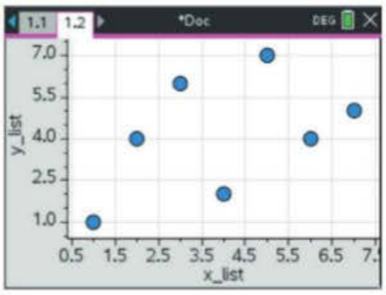


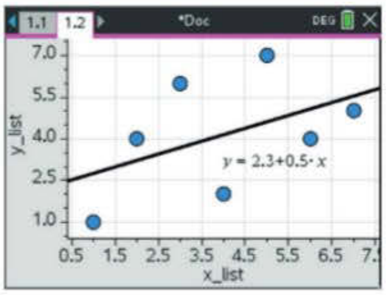
## USING CAS 5 Creating a residual plot

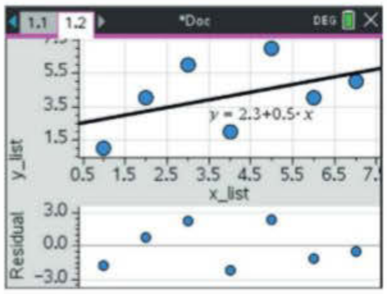
Graph a residual plot for the following data and use it to decide whether the data being investigated is linear or non-linear.

| $x$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $y$ | 1 | 4 | 6 | 2 | 7 | 4 | 5 |

### TI-Nspire




1 Insert a **Lists & Spreadsheet** application and enter the data into **x_list** and **y_list** (or use the previous data from Using CAS 4).

2 Insert a **Data & Statistics** application.

3 For the horizontal axis, select **x_list**.

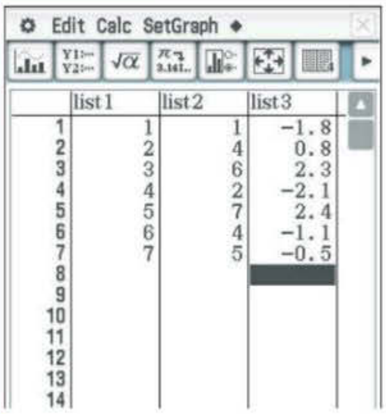
4 For the vertical axis, select **y_list**.

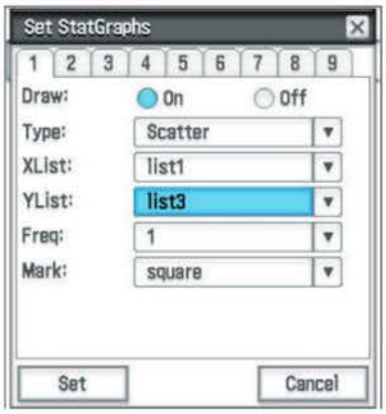
5 Press **menu** > **Analyze** > **Regression** > **Show Linear (a+bx)** (note the mode setting is **Fix 1**).

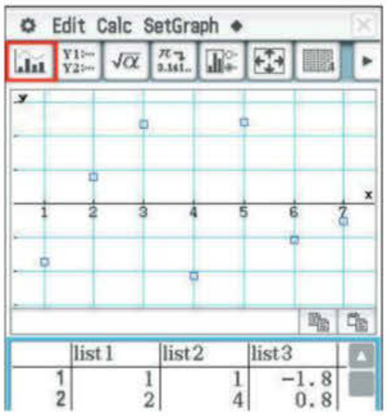
6 Press **menu** > **Analyze** > **Residuals** > **Show Residual Plot**

7 The residual plot will be displayed below the scatterplot.

### ClassPad




1 Open the **Statistics** application and enter the data (or use the previous data from Using CAS 4).

2 Insert the residual values into **list3**.

3 Tap **SetGraph** > **Setting**.

4 Keep the default settings except change the **YList:** field to **list3**.

5 Tap **Set**.

6 Tap **Graph**.

7 The graph of the residual values will appear in the lower window (note: the windows have been swapped).

The data is probably linear because the residual values appear randomly scattered above and below the $x$-axis.



9780170448406

3.4

## Recap

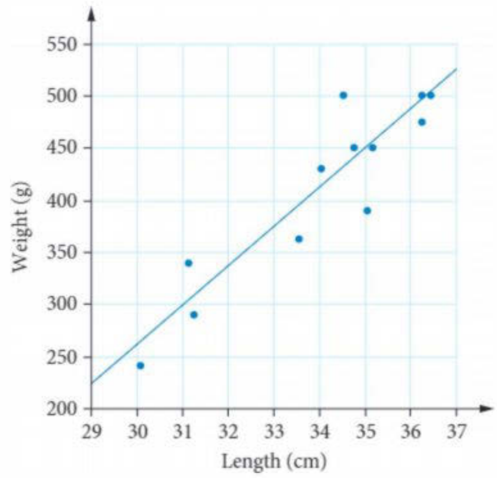
*Use the following information to answer the next two questions.*

The weights (in g) and lengths (in cm) of 12 fish were recorded and plotted in the scatterplot. The least squares line of best fit that enables the weight of these fish to be predicted from their length has also been plotted.

1 ©VCAA | 20081CQ8 62% The least squares line of best fit predicts that the weight (in g) of a fish of length 30 cm would be closest to

A 240 B 252 C 262

D 274 E 310




Data source: Journal of Statistics Education Data Archive (www.amstat.org/publications/jse)

2 ©VCAA 20081CQ9 | 55% The median weight (in g) of the 12 fish is closest to

A 346 B 375 C 440 D 450 E 475

## Mastery

3 WORKED EXAMPLE 7 Find the residual values for each of the following.

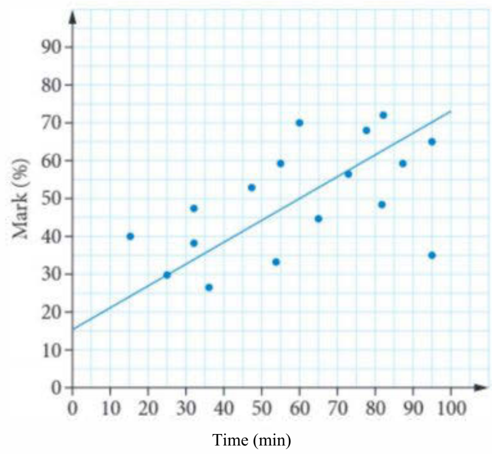
a For a study conducted to establish the association between the time (minutes) spent studying for a test and the percentage mark for the test, find the residual values directly from the graph for

i Selby who studied for 25 minutes for the test

ii Surinam who studied for 60 minutes for the test

iii Normie who studied over 90 minutes for the test and achieved 65%

iv Nissal who also studied for over 90 minutes for the test and achieved less than 50%.




b For a study conducted to establish the association between the height of a building and the number of levels, the least squares line of best fit equation was found to be

*height* = 5 + 0.57 x *number of levels*

Find the residual values, correct to one decimal place, using the equation for

i a building with 12 levels that is 12.3 metres tall

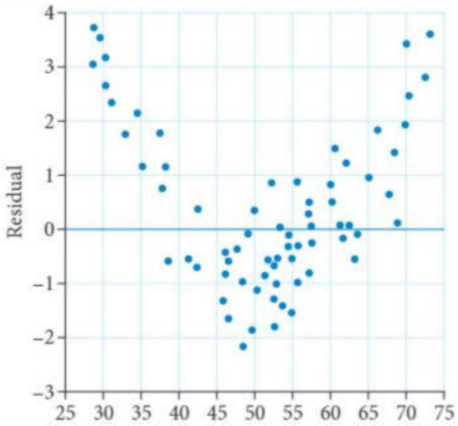
ii a building with 5 levels that is 7.2 metres tall.

**Exam hack**

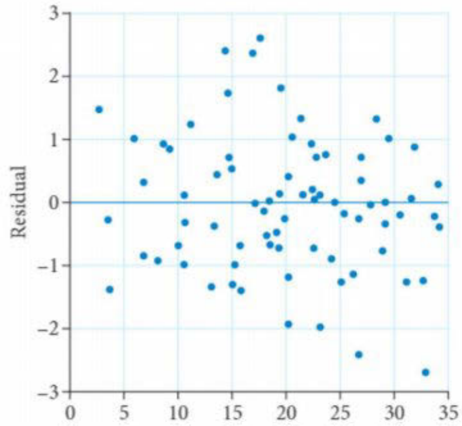
Don't lose easy marks. *Always* check whether the question has asked for a specific number of decimal places or significant figures.

4 For each of the following residual plots, state whether it suggests a linear or non-linear relationship. Provide a reason for your answer.

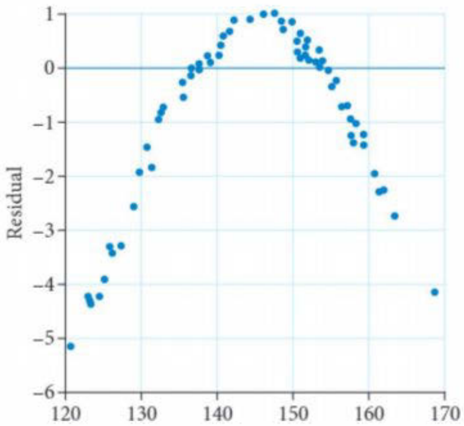
a



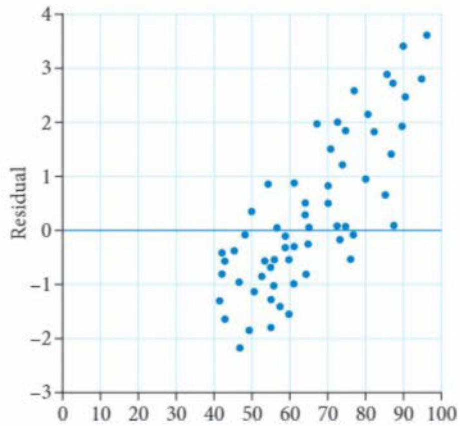

b




c



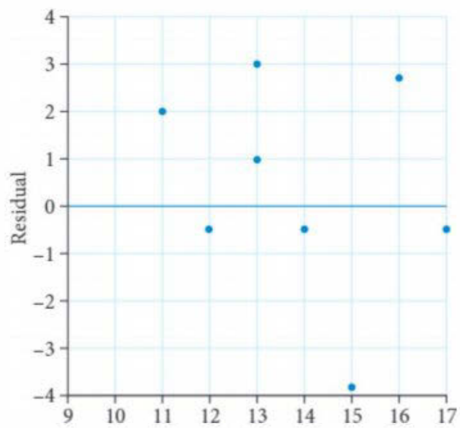

d

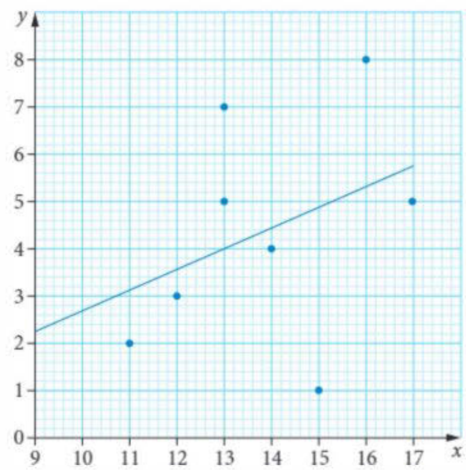



5 E3 CAS 4 J gjg Using CAS 5 J Create a table of residual values for the following set of data (assuming that *height* is the explanatory variable), giving all answers correct to one decimal place. Graph a residual plot and use it to decide whether the data being investigated is linear or non-linear.

| Height (cm) | Femur length (cm) |
|---|---|
| 178 | 50.2 |
| 173 | 48.4 |
| 165 | 45.1 |
| 164 | 44.6 |
| 168 | 45 |
| 165 | 42.6 |
| 155 | 39.9 |
| 155 | 38 |

6 Explain why the residual plot below does not match the scatterplot and least squares line of best fit, referring specifically to the data value at $x = 11$.





9780170448406

## Exam practice

80-100% 60-79% 0-59%

3.4

7 ©VCAA [2018 1CQ7] 73% The scatterplot displays the *resting pulse rate,* in beats per minute, and the *time spent exercising,* in hours per week, of 16 students. A least squares line has been fitted to the data.

Using this least squares line to model the association between *resting pulse rate* and *time spent exercising,* the residual for the student who spent four hours per week exercising is closest to

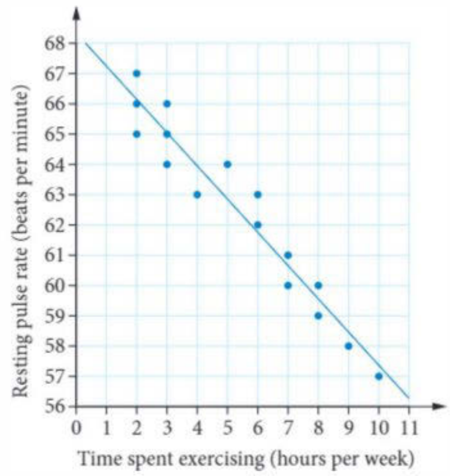
A -2.0 beats per minute.

B -1.0 beats per minute.

C -0.3 beats per minute.

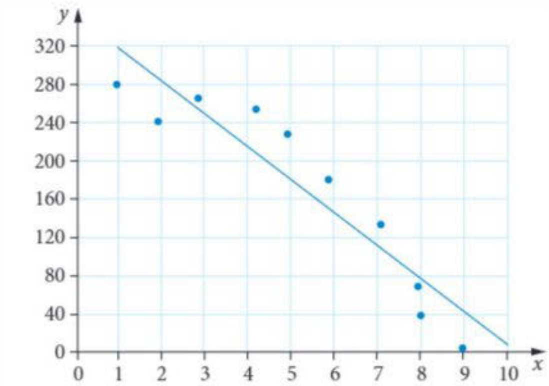
D 1.0 beats per minute.

E 2.0 beats per minute.

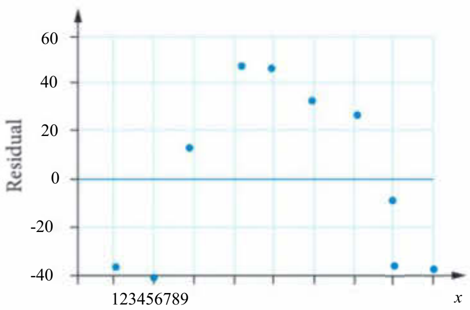



8 ©VCAA [2013 1CQ11] 69% A least squares line of best fit is fitted to data in a scatterplot, as shown.

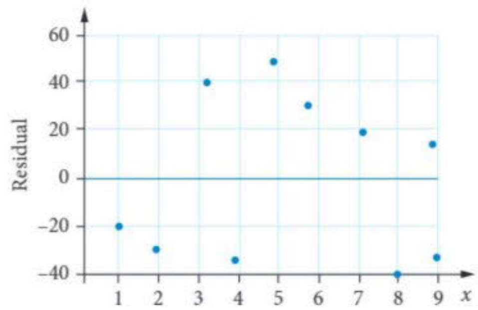
The corresponding residual plot is closest to



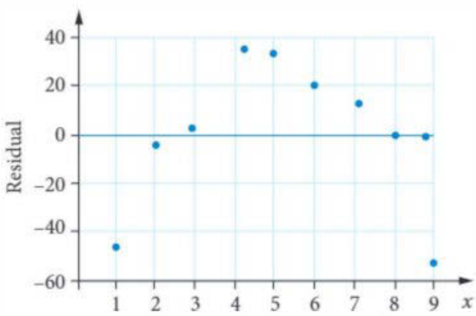
A



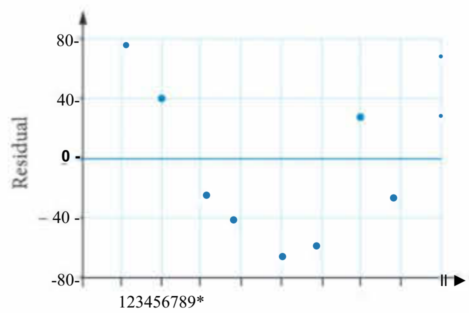

B



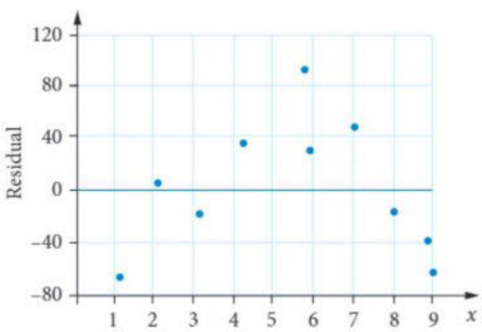

C




D

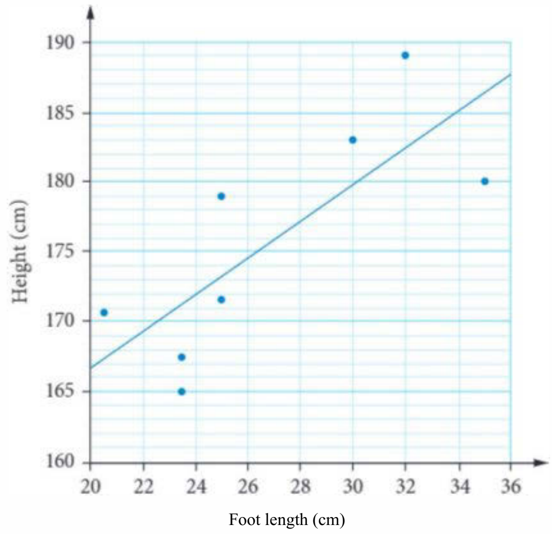



E



*Use the following information to answer the next three questions.*

The *height* (in cm) and *foot length* (in cm) for each of eight Year 12 students were recorded and displayed in the scatterplot. A least squares line of best fit has been fitted to the data as shown.




9 ©VCAA | 20101CQ7~ 73% By inspection, the value of the correlation coefficient (r) for this data is closest to

A 0.98 B 0.78 C 0.23 D -0.44 E -0.67

10 ©VCAA 2010 1CQ8 35% The explanatory variable *is foot length.* The equation of the least squares line of best fit is closest to

A *height* = -110 + 0.78 *xfoot length.*

B *height* =141 + 1.3 *xfoot length.*

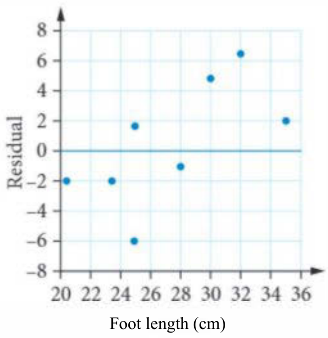
C *height* =167+1.3 *xfoot length.*

D *height* = 167 + 0.67 *xfoot length.*

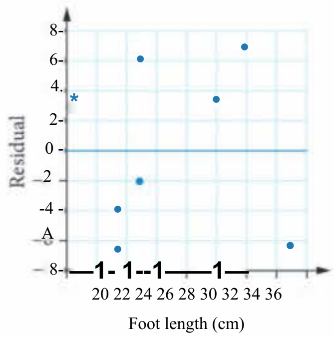
E *foot length* = 167 + 1.3 x *height.*

11 ©VCAA | 2010 1CQ9 61% The plot of the *residuals* against *foot length* is closest to

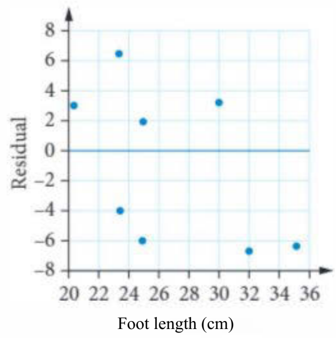
A



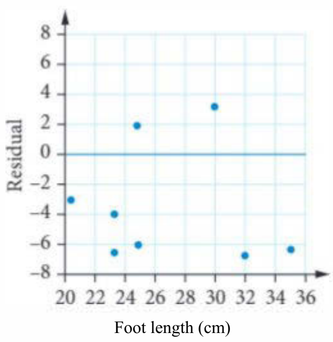

B



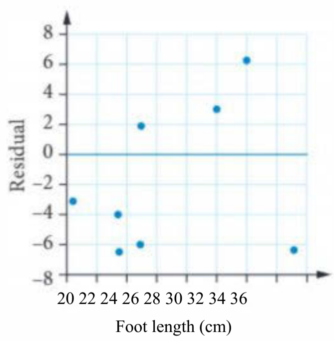

C




D




E





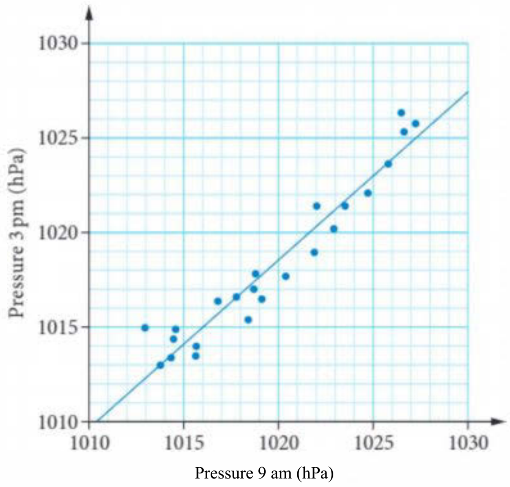

9780170448406

▶ 12 ©VCAA 2019 2CQ5 J (8 marks) The scatterplot shows the atmospheric pressure, in hectopascals (hPa), at 3 pm *(pressure 3pm)* plotted against the atmospheric pressure, in hectopascals, at 9 am *(pressure 9am)* for 23 days in November 2017 at a particular weather station.

A least squares line has been fitted to the scatterplot as shown.

The equation of this line is

*pressure 3pm* = 111.4 + 0.8894 x *pressure 9am*




Data: Australian Government, Bureau of Meteorology, www.bom.gov.au

3.4

a 44% Interpret the slope of this least squares line in terms of the atmospheric pressure at this weather station at 9 am and at 3 pm. 1 mark

b 84% Use the equation of the least squares line to predict the atmospheric pressure at 3 pm when the atmospheric pressure at 9 am is 1025 hPa.

Round your answer to the nearest whole number. 1 mark

c 87% Is the prediction made in part b an example of extrapolation or interpolation? 1 mark

d 40% Determine the residual when the atmospheric pressure at 9 am is 1013 hPa.

Round your answer to the nearest whole number. 1 mark

e The mean and the standard deviation of *pressure 9 am* and *pressure* 3 *pm* for these 23 days are shown in the table.

| | Pressure 9 am | Pressure 3 pm |
|---|---|---|
| Mean | 1019.7 | 1018.3 |
| Standard deviation | 4.5477 | 4.1884 |

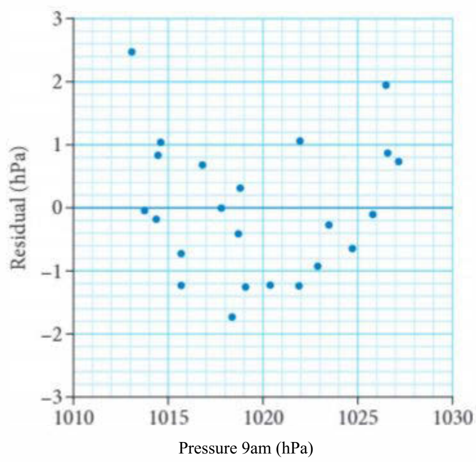
i 34% Use the equation of the least squares line and the information in the table to show that the correlation coefficient for this data, rounded to three decimal places, is r = 0.966. 1 mark

ii 46% What percentage of the variation in *pressure 3pm* is explained by the variation in *pressure 9 am*? Round your answer to one decimal place. 1 mark

## Exam hack

When the word 'use' appears in a question, make sure you use the method suggested even if it's possible to get the answer another way.

f The residual plot associated with the least squares line is shown.

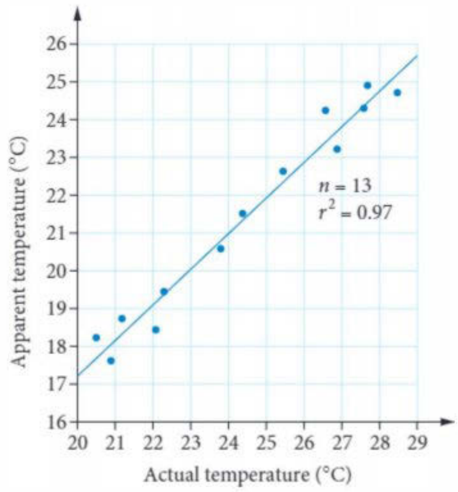



i 41 % The residual plot above can be used to test one of the assumptions about the nature of the association between the atmospheric pressure at 3 pm and the atmospheric pressure at 9 am. What is this assumption? 1 mark

ii 36% The residual plot above does not support this assumption. Explain why. 1 mark ▶

▶ 13 ©VCAA 2016 2CQ3 (8 marks) The data in the table shows a sample of actual temperatures and apparent temperatures recorded at a weather station. A scatterplot of the data is also shown. The data will be used to investigate the association between the variables *apparent temperature* and *actual temperature.*

| Apparent temperature (°C) | Actual temperature (°C) |
|---|---|
| 24.7 | 28.5 |
| 24.3 | 27.6 |
| 24.9 | 27.7 |
| 23.2 | 26.9 |
| 24.2 | 26.6 |
| 22.6 | 25.5 |
| 21.5 | 24.4 |
| 20.6 | 23.8 |
| 19.4 | 22.3 |
| 18.4 | 22.1 |
| 17.6 | 20.9 |
| 18.7 | 21.2 |
| 18.2 | 20.5 |




a 78% Use the scatterplot to describe the association between *apparent temperature* and *actual temperature* in terms of strength, direction and form. 1 mark

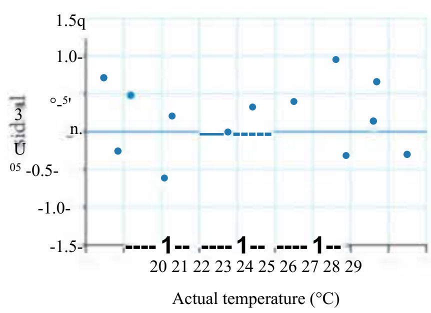
b i 53% Determine the equation of the least squares line that can be used to predict the *apparent temperature* from the *actual temperature.* Copy and complete the following by adding the values of the intercept and slope of this least squares line in the appropriate boxes. Round your answers to two significant figures.

*apparent temperature* = ☐ + ☐ x *actual temperature* 3 marks

ii 28% Interpret the intercept of the least squares line in terms of the variables *apparent temperature* and *actual temperature.* 1 mark

c 49% The coefficient of determination for the association between the variables *apparent temperature* and *actual temperature* is 0.97. Interpret the coefficient of determination in terms of these variables. 1 mark

d The residual plot obtained when the least squares line was fitted to the data is shown.




i 46% A residual plot can be used to test an assumption about the nature of the association between two numerical variables. What is this assumption? 1 mark

ii 46% Does the residual plot above support this assumption? Explain your answer. 1 mark ▶


9780170448406

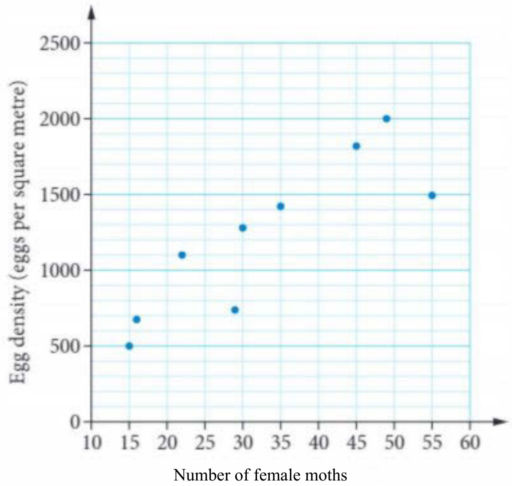
▶ 14 © VCAA 2017 2CQ3 , (6 marks) The *number of male moths* caught in a trap set in a forest and the egg *density* (eggs per square metre) in the forest are shown in the table.

| Number of male moths | 35 | 37 | 45 | 49 | 65 | 74 | 77 | 86 | 95 |
|---|---|---|---|---|---|---|---|---|---|
| Egg density (eggs per square metre) | 471 | 635 | 664 | 997 | 1350 | 1100 | 2010 | 1640 | 1350 |

a 74% Determine the equation of the least squares line that can be used to predict the egg *density* in the forest from the *number of male moths* caught in the trap. Copy and complete the following by writing the values of the intercept and slope of this least squares line in the appropriate boxes as shown. Round your answers to one decimal place.

egg *density* = ☐ + ☐ x *number of male moths* — 2 marks

b The *number of female moths* caught in a trap set in a forest and the egg *density* (eggs per square metre) in the forest can also be examined. A scatterplot of the data is shown.




## Exam hack

Take a ruler into the exam and use it for these sorts of questions. Watch out for horizontal axes that don't start with 0.

The equation of the least squares line is

egg *density* = 191 + 31.3 x *number of female moths*

i 26% Copy the above scatterplot and draw the graph of this least squares line. — 1 mark

ii 39% Interpret the slope of the least squares line in terms of the variables egg *density* and *number of female moths* caught in the trap. — 1 mark

iii 48% The egg *density* is 1500 when the *number of female moths* caught is 55. Determine the residual value if the least squares line is used to predict the egg *density* for this number of female moths. — 1 mark

## Exam hack

Make sure you get the sign right in residual questions. If rounding isn't asked for, give the full answer.

iv 47% The correlation coefficient is r = 0.862. Determine the percentage of the variation in egg *density* in the forest explained by the variation in the *number of female moths* caught in the trap. Round your answer to one decimal place. — 1 mark ▶

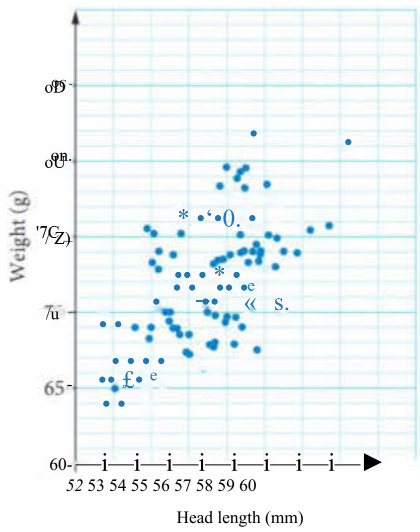
## Exam hack

When asked for rounding, do not round until the very last step.

3.4

▶ 15 ©VCAA 2015 N 2CQ5 (7 marks) The scatterplot shows the *weight,* in grams, and the *head length,* in millimetres, of 110 birds.

The equation of the least squares line fitted to this data is *weight* = -24.83 + 1.739 x *head length*




a Copy the above axes and draw this least squares line. 1 mark

b Use the equation to predict the *weight,* in grams, of a bird with a *head length* of 49.0 mm. Round your answer to one decimal place. 1 mark

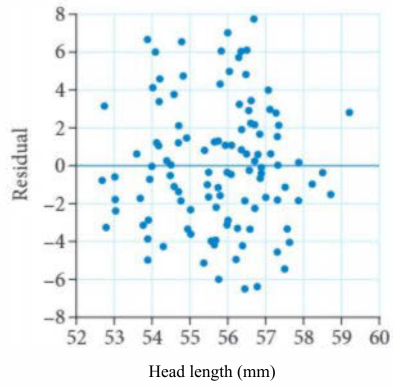
c Is the prediction made in part b an example of interpolation or extrapolation? Explain your answer briefly. 1 mark

d When the least squares line is used to predict the *weight* of a bird with a *head length* of 59.2 mm, the residual value is 2.78. Calculate the actual weight of this bird. Round your answer to one decimal place. 2 marks

e Pearsons correlation coefficient, *r,* is equal to 0.5957. Given this information, what percentage of the variation in the *weight* of these birds is not explained by the variation in *head length?* Round your answer to one decimal place. 1 mark

f The residual plot obtained when the least squares line is fitted to the data set is shown.

What does the residual plot indicate about the association between *head length* and *weight* for these birds? 1 mark



# @ Data transformations

3.5

## Types of data transformations

We now have a number of ways of dealing with linear associations; however, not all associations are linear. The way that we deal with non-linear associations is to apply a transformation to one of the variables so that the association between the two variables becomes closer to a straight line. This is called linearisation. We can then work with the data as if it were linear and then convert it back to its original form at the end.

Using a data transformation is similar to changing the units of measurement you work with. It is like converting centimetres to inches because you have a ruler with inches, working out your problem, and then converting your answer back to centimetres at the very end.

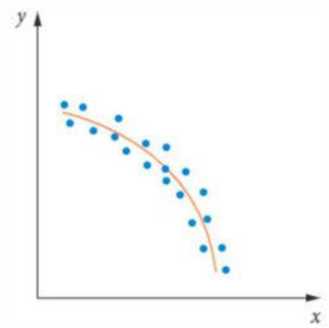
We will be looking at three types of data transformations that can be used with scatterplots that are consistently increasing or decreasing.

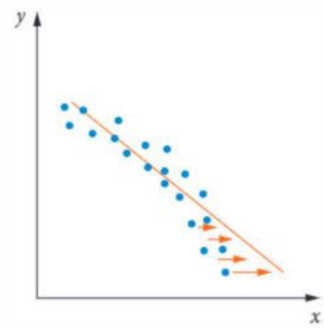
Video playlist
Data transformations

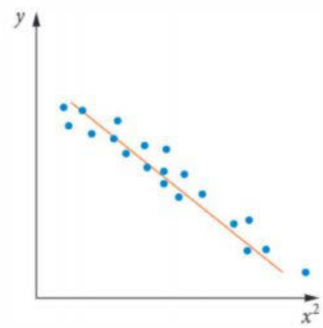
### 1 Squared transformation ($x^2$ ory$^2$)

Squared transformation involves squaring either the $x$ values or the $y$ values. So 5 becomes 25, 9 becomes 81,100 becomes 10000 and so on. Large values increase more than small values, and the effect of the transformation is that the data is *stretched* either horizontally (for an $x^2$ transformation) or vertically (for a $y^2$ transformation).

Original non-linear scatterplot → To stretch large $x$ values, use $x^2$ → Transformed linear scatterplot

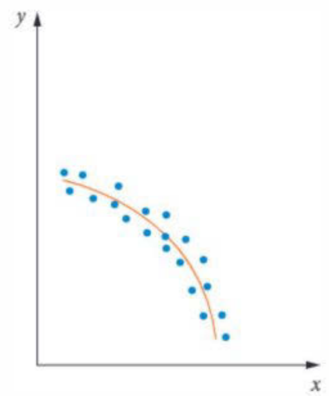


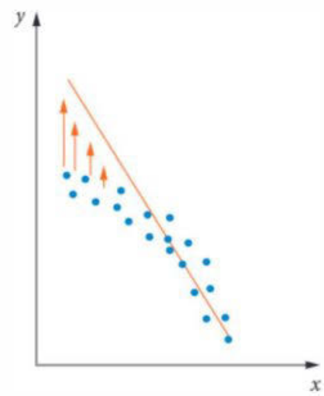



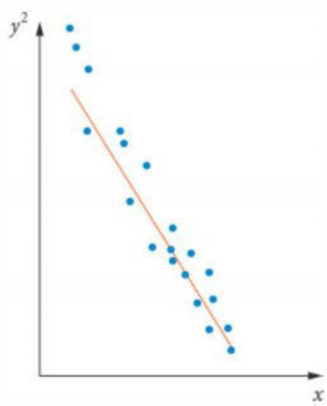




Line of best fit equation:
$y = a + bx^2$

Original non-linear scatterplot → To stretch large $y$ values, use $y^2$ → Transformed linear scatterplot



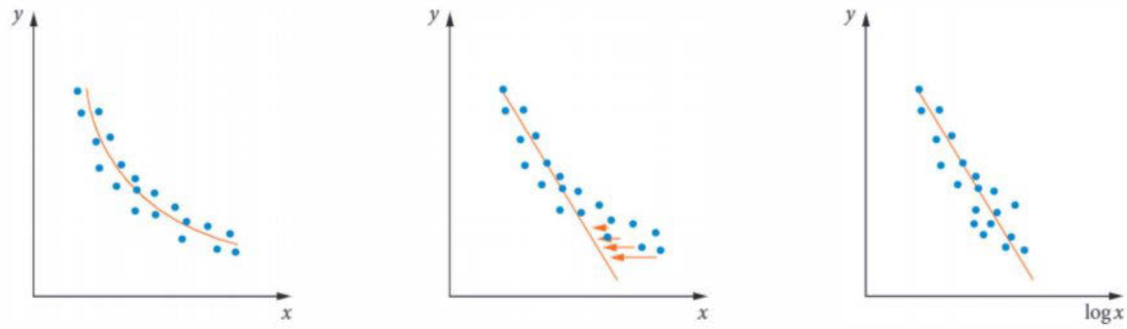




Line of best fit equation:
$y^2 = a + bx$

### 2 Logarithmic transformation (logx or logy)

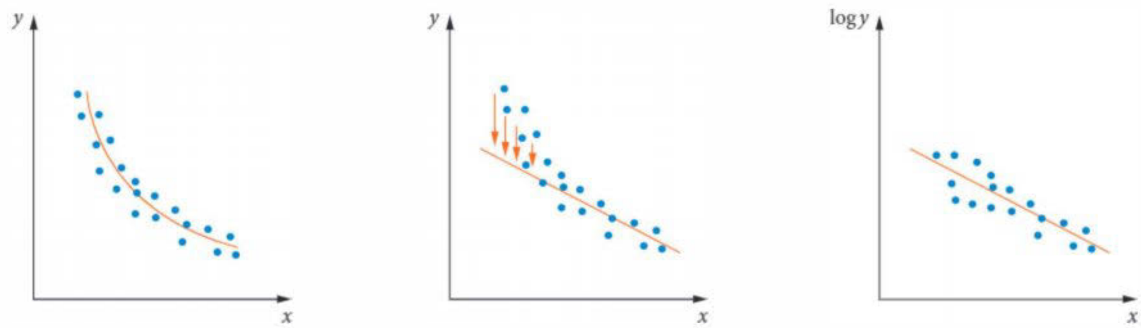
Logarithmic transformation (or log transformation) involves finding the log base 10 of either the $x$ values or the $y$ values. Large values are reduced more than small values, and the effect of the transformation is that the data is *compressed* either horizontally (for a logx transformation) or vertically (for a logy transformation).

Original non-linear scatterplot To compress large x values, use logx -> Transformed linear scatterplot




Line of best fit equation:
$y = a + b\log x$

Original non-linear scatterplot To compress large $y$ values, use logy Transformed linear scatterplot




Line of best fit equation:
$\log y = a + bx$

## Exam hack

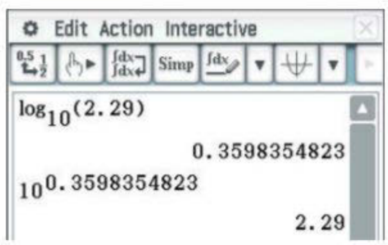
When dealing with logs, you will need to use CAS and round your answers.

If you know the value and need to find the log of it, use the $\log_{10}$ function (e.g. $\log_{10}(2.29) = 0.36$).

If you know the log but need to find the value, raise the log to the power of 10 (e.g. if $\log_{10}x = 0.36$, then $x = 10^{0.36} = 2.29$).

**TI-Nspire**


**ClassPad**



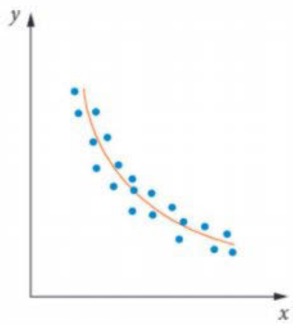


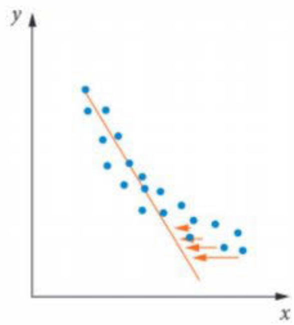
9780170448406

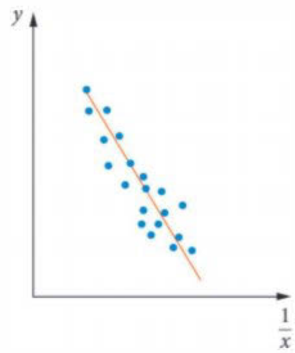
### 3 Reciprocal transformation ($\frac{1}{x}$ or $\frac{1}{y}$)

Reciprocal transformation involves taking the reciprocal of either the $x$ values or the $y$ values. So 5 becomes $\frac{1}{5} = 0.2$, 9 becomes $\frac{1}{9} = 0.11$, 100 becomes $\frac{1}{100} = 0.01$ and so on. As with the log transformation, large values are reduced more than small values, and the effect of the transformation is that the data is *compressed* either horizontally (for a $\frac{1}{x}$ transformation) or vertically (for a $\frac{1}{y}$ transformation). The compression of large values for a reciprocal transformation is greater than for a log transformation.

| Original non-linear scatterplot | -> To compress large $x$ values, use $\frac{1}{x}$ or $\log x$ → | Transformed linear scatterplot |
|---|---|---|



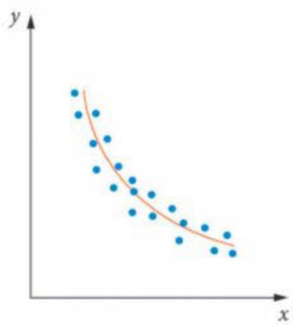



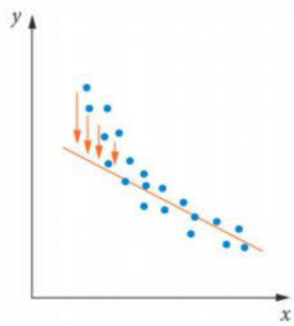



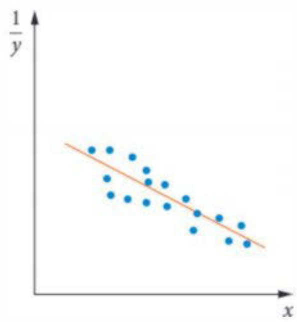

Line of best fit equation:

$y = a + b\frac{1}{x}$ or $y = a + b\log x$

| Original non-linear scatterplot | -> To compress large $y$ values, use $\frac{1}{y}$ or $\log y$ → | Transformed linear scatterplot |
|---|---|---|



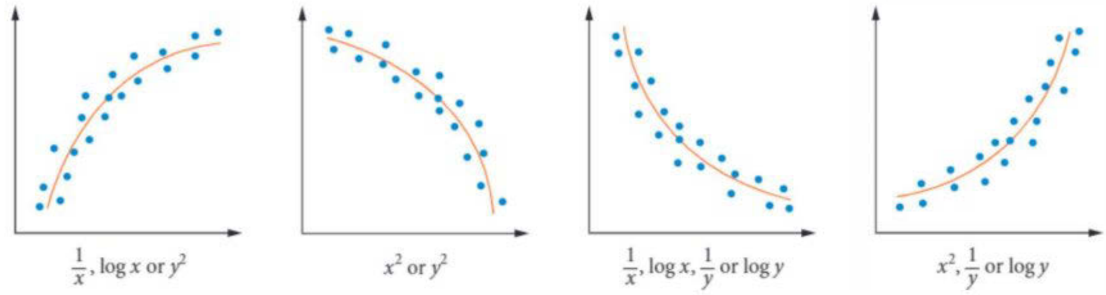

Line of best fit equation:

$\frac{1}{y} = a + bx$ or $\log y = a + bx$

## Choosing a transformation

The graphs show when transformations can be used to linearise non-linear data. For example,

the transformation options for data shaped like are : $x^2$, $\frac{1}{y}$ or $\log y$.



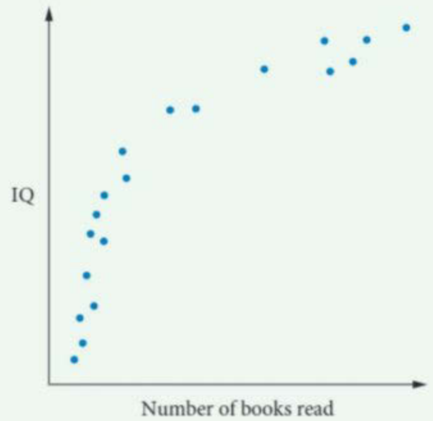

It is difficult to tell in advance which of the options would give you the most linear result, so if there is no information about which is the best option, calculate the coefficient of determination of each of them to see which is closest to 1.

WB p. 40

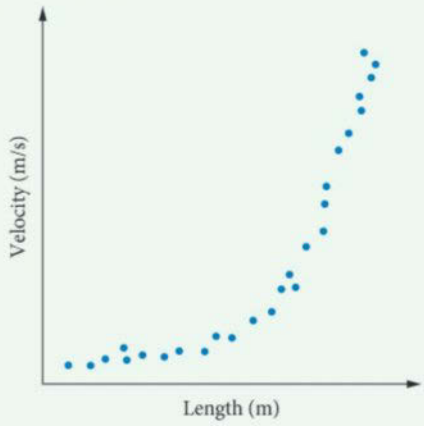
**WORKED EXAMPLE 8** Writing data transformation equations

For each of the following non-linear scatterplots, write the equations of the lines of best fit after applying the appropriate transformations (squared, log and reciprocal) that linearise the data.

**a**




**b**




| **Steps** | **Working** |
|---|---|
| **a** **1** Identify the options for the data shape. | $\frac{1}{x}$, $\log x$, $y^2$ |
| **2** Rewrite the line of best fit equation $y = a + bx$ in terms of the transformations, using the variable names. | $IQ = a + b \times \frac{1}{\textit{number of books read}}$<br>$IQ = a + b \times \log(\textit{number of books read})$<br>$(IQ)^2 = a + b \times \textit{number of books read}$ |
| **b** **1** Identify the options for the data shape. | $x^2$, $\frac{1}{y}$, $\log y$ |
| **2** Rewrite the line of best fit equation $y = a + bx$ in terms of the transformations, using the variable names. | $\textit{velocity} = a + b \times (\textit{length})^2$<br>$\frac{1}{\textit{velocity}} = a + b \times \textit{length}$<br>$\log(\textit{velocity}) = a + b \times \textit{length}$ |


9780170448406

**WORKED EXAMPLE 9** Working with data transformations

The data for the association between the maximum daily temperature (°C) and the daily number of boxes of hot pies sold at the school canteen has been linearised by applying two separate transformations, giving the following least squares line of best fit equations.

$$\textit{number of boxes of pies} = -0.8 + \frac{90}{\textit{temperature}}$$

$$\log(\textit{number of boxes of pies}) = 1.05 - 0.023 \times \textit{temperature}$$

a For each of these, use the equation to predict the number of boxes of pies sold when the temperature was

i 17°C ii 25°C iii 30°C

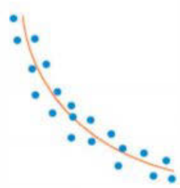
b What shape is the original data?

c The coefficient of determination was calculated to be 0.92 for the reciprocal transformation and 0.84 for the log transformation. Which of the two equations give the best fit to the data? Give a reason for your answer.

3.5

WB

p. 41

| Steps | Working |
|---|---|
| a Substitute the value into the equation.<br>Use CAS when dealing with logs.<br>$\log_{10}x = a$ is the same as $x = 10^{a}$.<br>$\log x$ means $\log_{10}x$.<br>Round the answer if necessary. | i $\textit{number of boxes of pies} = -0.8 + \frac{90}{17} = 4.49$<br>The prediction is 4 boxes of pies.<br>$\log(\textit{number of boxes of pies}) = 1.05 - 0.023 \times 17 = 0.659$<br>$\textit{number of boxes of pies} = 10^{0.659} = 4.56$<br>The prediction is 5 boxes of pies.<br>ii $\textit{number of boxes of pies} = -0.8 + \frac{90}{25} = 2.8$<br>The prediction is 3 boxes of pies.<br>$\log(\textit{number of boxes of pies}) = 1.05 - 0.023 \times 25 = 0.475$<br>$\textit{number of boxes of pies} = 10^{0.475} = 2.99$<br>The prediction is 3 boxes of pies.<br>iii $\textit{number of boxes of pies} = -0.8 + \frac{90}{30} = 2.2$<br>The prediction is 2 boxes of pies.<br>$\log(\textit{number of boxes of pies}) = 1.05 - 0.023 \times 30 = 0.36$<br>$\textit{number of boxes of pies} = 10^{0.36} = 2.29$<br>The prediction is 2 boxes of pies. |
| b 1 Which transformations have been used?<br>2 Which data shape matches both of these transformations? | $\frac{1}{x}$ and $\log y$<br> |
| c Which of the two coefficients of determination are closer to 1? | The coefficient of determination for the reciprocal transformation is closer to 1, so it gives the best fit to the data. |

**Exam hack**

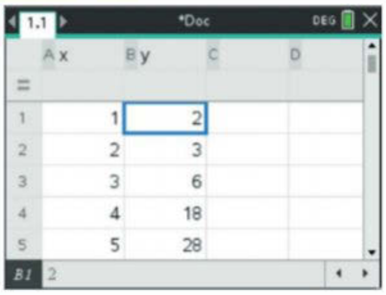
Sometimes you need to round to the nearest whole number, even though you are not specifically told to, because of the context of the question.

## USING CAS 6A Transforming non-linear data with TI-Nspire

Use the coefficient of determination to decide which transformation is the best choice for linearising the following data, and write down the equation of the least squares line of best fit in terms of the transformed variables, with the slope and intercept, correct to two decimal places.

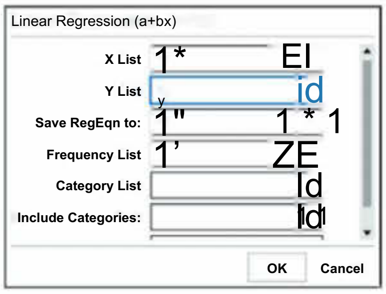
| $x$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $y$ | 2 | 3 | 6 | 18 | 28 |

[TI-Nspire J




1 Start a new document and add a Lists & Spreadsheet page.

2 Label the columns and enter the data from the table as shown above (note that normally we try to avoid using variables $x$ and $y$ as headings).

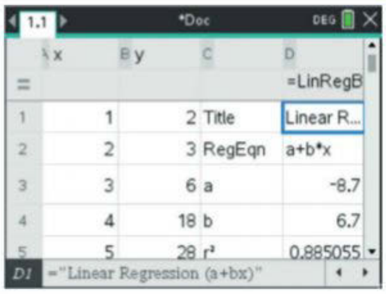



3 Press menu > Statistics > Stat Calculations > Linear Regression (a+bx).

4 For the X List: field, select x.

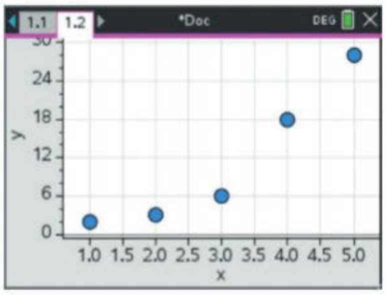
5 For the Y List: field, select y.

6 Select OK.




7 The linear regression labels and data will be displayed in columns C and D.

(Note that the mode setting is Float 6, not Fix 1.)



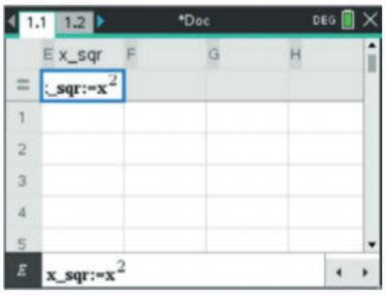

8 Insert a Data & Statistics page.

9 For the horizontal axis, select x.

10 For the vertical axis, select y.

The shape of the data is

so the options to linearise the data are $x^2$, $\frac{1}{y}$ or $\log y$.

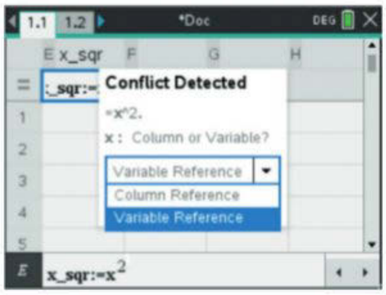



11 Return go the Lists & Spreadsheet page.

12 Label Column E as x_sqr.

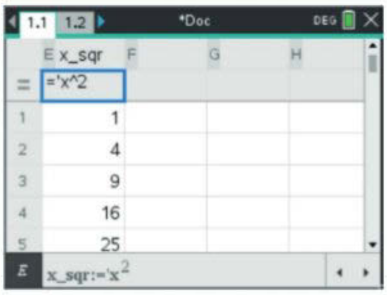
13 In the cell immediately under x_sqr, enter the formula x_sqr:=$x^2$.

14 Press enter.

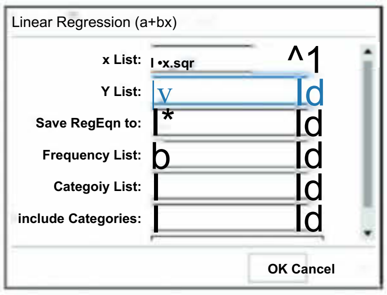



15 When prompted regarding a conflict on the next screen, select Variable Reference. This is because we are using the variable x instead of column X.

16 Select OK.




17 The $x^2$ data will be displayed in column E.

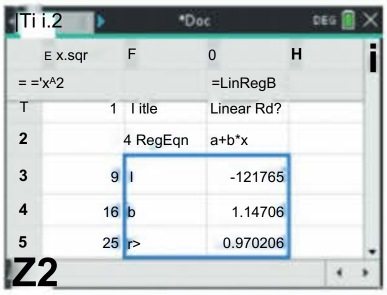



18 Press menu > Statistics > Stat Calculations > Linear Regresssion (a+bx).

19 For the X List: field, select x.sqr.

20 For the Y List: field, select y.

21 Select OK.

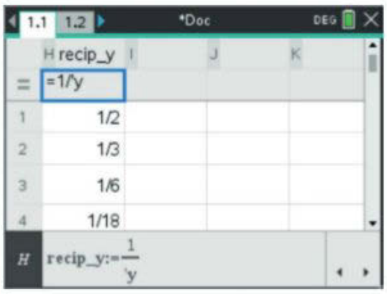



22 The linear regression $a_t$ $b$ and $r^2$ values will be displayed in columns F and G.

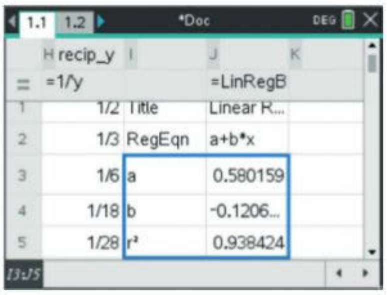
a = -1.22

$b$ = 1.15

$r^2$ = 0.97




23 Using the same method, insert the reciprocal $y$ values into column H.



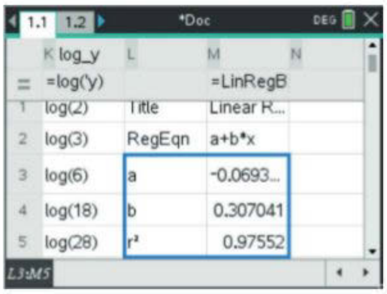

24 Calculate the linear regression for $x$ and $1/y$ (remember to select Variable Reference).

25 The $a$, $b$ and $r^2$ values will be displayed in columns I and J.

$a$ = 0.58

$b$ = -0.12

$r^2$ = 0.94




26 Using the same method, insert the logy values into column K.

27 Calculate the linear regression for $x$ and logy. The $a$, $b$ and $r^2$ values will be displayed in columns L and M.

$a$ = -0.07

6 = 0.31

$r^2$ = 0.98

The coefficient of determination of the logy transformation is closest to 1, so the logy transformation is the best choice for linearising the data.

Use the $a$ and $b$ values to write the equation of the least squares line of best fit for this option.

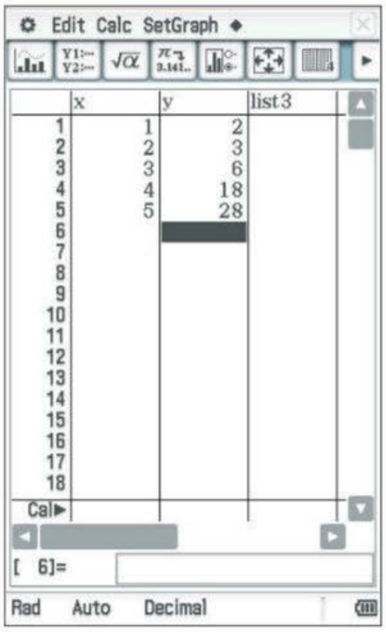
logy = -0.07 + 0.31x

## USING CAS 6B Transforming non-linear data with Casio ClassPad

Use the coefficient of determination to decide which transformation is the best choice for linearising the following data, and write down the equation of the least squares line of best fit in terms of the transformed variables, with the slope and intercept, correct to two decimal places.

| $x$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $y$ | 2 | 3 | 6 | 18 | 28 |

**ClassPad**

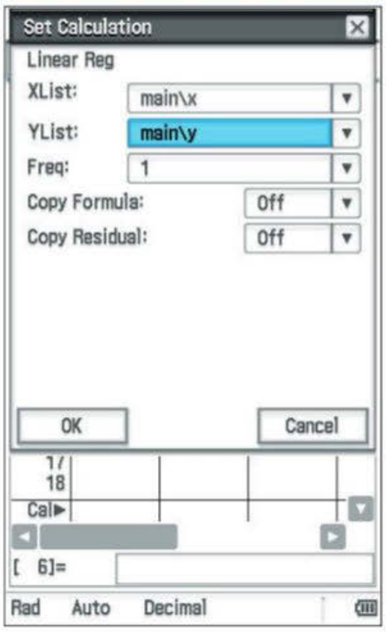



1 Open the **Statistics** application and clear all lists.

2 Rename **list1** as **x**.

3 Rename **list2** as **y**.

Note – use the letters x and y, not the variables $\boldsymbol{x}$ and $\boldsymbol{y}$.

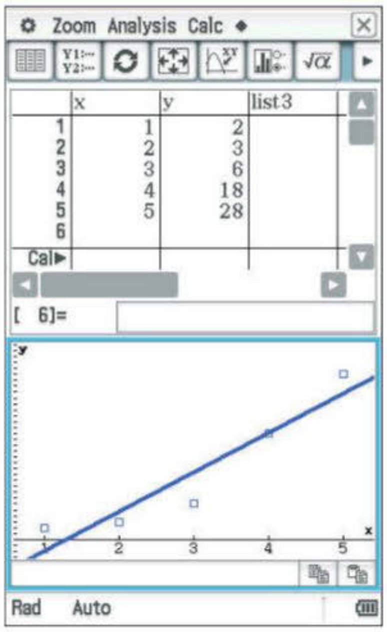



4 Tap **Calc** > **Regression** > **Linear Reg**.

5 Change the **XList:** field to **main\x**.

6 Change the **YList:** field to **main\y**.

7 Tap **OK**.




8 On the next screen displaying the linear regression values, tap **OK**.

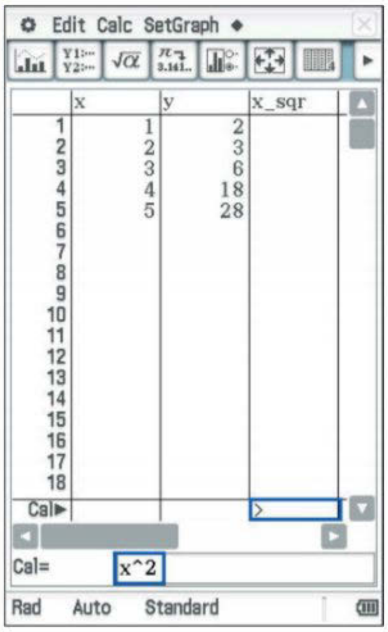
9 The shape of the data is

so the options to linearise the data are $x^2$, $\frac{1}{y}$ or $\log y$.



9780170448406

3.5

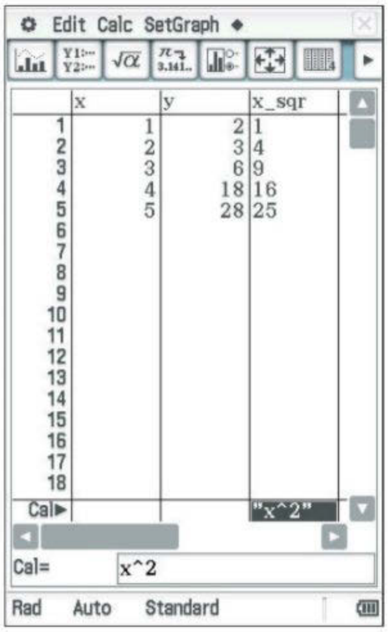



10 Tap to Resize the Statistics window.

11 Rename Iist3 as x_sqr.

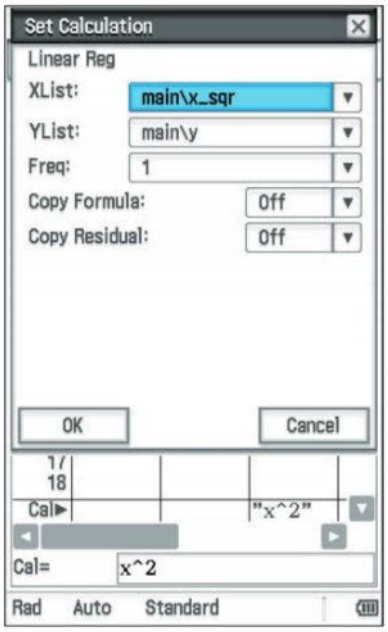
12 Tap on the Cal ▶ field at the bottom of the x_sqr list.

13 In the Cal= field at the bottom of the screen, enter the formula x^2.




14 Press EXE.

15 The $x^2$ values will appear in the x_sqr list.

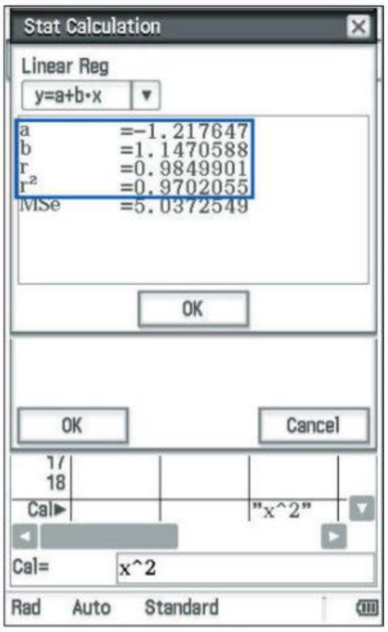



16 Tap Calc > Regression > Linear Reg.

17 Change the XList: field to main\x_sqr.

18 Change the YList: field to main\y.

19 Tap OK.

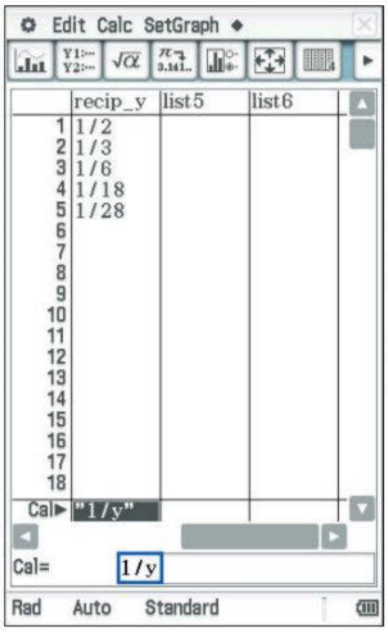



20 The linear regression *a*, *b*, *r* and $r^2$ values will be displayed.

a = -1.22

h-1.15

$r^2$ = 0.97

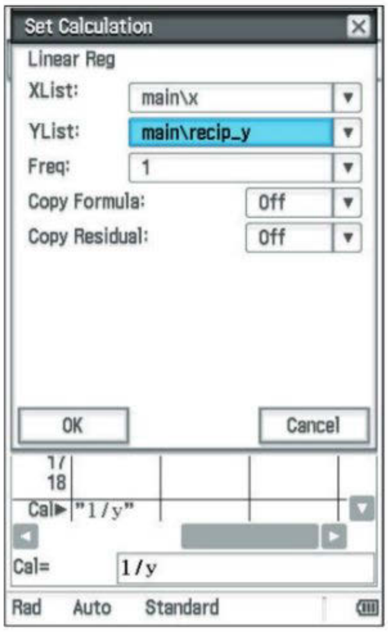



21 Tap OK.

22 Resize the Statistics window.

23 Rename Iist4 as recip.y.

24 Repeat the previous steps for the recip_y list so the Cal= field at the bottom of the screen equals 1/y.

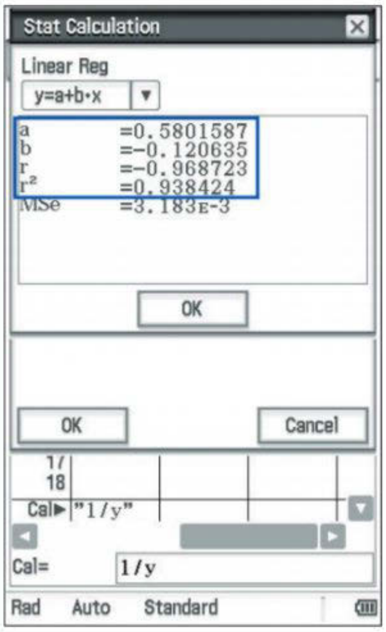


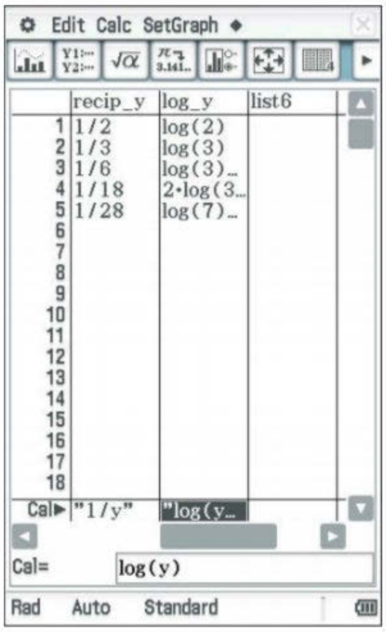

25 Tap Calc > Regression > Linear Reg.

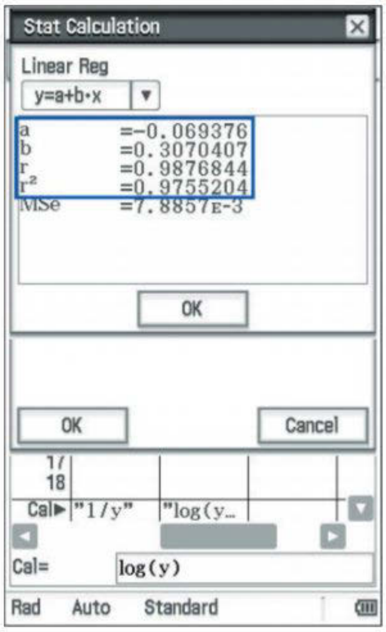
26 Change the XList: field to main\x.

27 Change the YList: field to main\recip_y.

28 Tap OK.








29 The linear regression a, $b_t$ $r$ and $r^2$ values will be displayed.

$a = 0.58$

$b = -0.12$

$r^2 = 0.94$

30 Repeat by changing Iist5 to log_y.

31 When calculating the linear regression, change the XList field to main\x and the YList: field to main\log_y.

32 The linear regression *a,* $b_t$ $r$ and $r^2$ values will be displayed.

$a = -0.07$

$b = 0.31$

$r^2 = 0.98$

The coefficient of determination of the logy transformation is closest to 1, so the logy transformation is the best choice for linearising the data.

Use the *a* and *b* values to write the equation of the least squares line of best fit for this option.

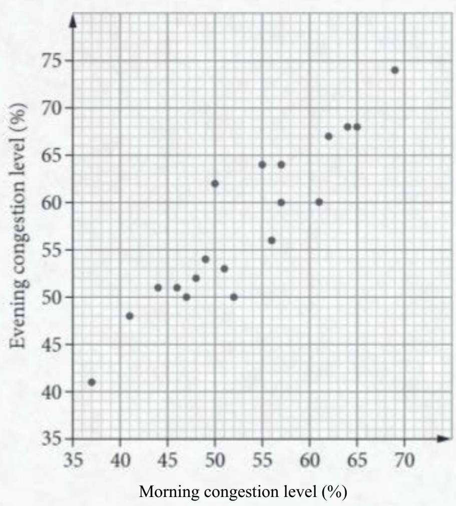
logy = -0.07 + 0.31*

Video VCE question analysis: Linear associations

## VCE QUESTION ANALYSIS

©VCAA 2018 2CQ2 2018 Examination 2 Core Question 2 (7 marks)

The congestion level in a city can be recorded as the percentage increase in travel time due to traffic congestion in peak periods (compared to non-peak periods). This is called the percentage congestion level. The percentage congestion levels for the morning and evening peak periods for 19 large cities are plotted on the scatterplot.



a Determine the median percentage congestion level for the morning peak period and for the evening peak period. 2 marks

A least squares line is to be fitted to the data with the aim of predicting evening congestion level from morning congestion level. The equation of this line is

*evening congestion level* = 8.48 + 0.922 x *morning congestion level*

b Name the response variable in this equation. 1 mark

c Use the equation of the least squares line to predict the evening congestion level when the morning congestion level is 60%. 1 mark

d Determine the residual value when the equation of the least squares line is used to predict the evening congestion level when the morning congestion level is 47%. Round your answer to one decimal place. 2 marks

e The value of the correlation coefficient r is 0.92. What percentage of the variation in the evening congestion level can be explained by the variation in the morning congestion level? Round your answer to the nearest whole number. 1 mark

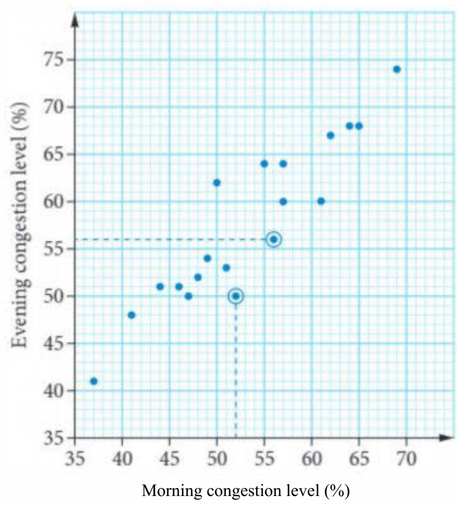
**Reading the question**

- Key concepts you need to be clear on are the median, response variable, least squares line, residual value and correlation coefficient.
- Take note of the 2-mark questions.
- Be aware of the rounding required for each part, particularly if it changes.

**Thinking about the question**

- You will need to know how to calculate median for *both* variables from a scatterplot.
- Some 2-mark questions require a single answer with two steps, while others require two separate answers.
- Remember the formulas for the residual value and the least squares line are on the formula sheet.
- When you see a reference to what percentage the variation ...' you should realise that the coefficient of determination needs to be calculated.

**Worked solution** (/ = 1 mark)




a There are 19 large cities, so the median occurs at the tenth dot on the scatterplot.

Count the dots from the left for the morning congestion period.

Count the dots from the bottom for the evening congestion period.

The median for the morning peak period is 52%. /

The median for the evening peak period is 56%. /

b The y-axis variable is always the response variable.

The response variable is *evening congestion level. V*

c Substitute *morning congestion level* = 60% into the equation.

*Evening congestion level* = 8.48 + 0.922 x 60 = 63.8% /

**Exam hack**

If no rounding is asked for and you round the answer, it will be marked as incorrect.

d From the scatterplot, when the *morning congestion level* is 47%, the actual *evening congestion level* is 50%. /

From the least squares line equation, the predicted

*evening congestion level* = 8.48 + 0.922 x 47

= 51.814%.

residual value = actual value - predicted value

= 50-51.814

= -1.814

The residual value to one decimal place is -1.8. /

e The coefficient of determination $r^2 = 0.92^2 = 0.8464$, which as a percentage is 84.64%, giving 85% when rounded to the nearest whole number.

This means 85% / of the variation in the evening congestion level can be explained by the variation in the morning congestion level.

**Student performance** 80-100% eo-79% 0-59%

a 56% This wasn't answered well. Some students gave the correct morning peak median, but incorrectly said evening peak median was they value of the morning peak median.

b 90%

c 78% No rounding was asked for, so 64% was not accepted.

d 53% Many students did not find the actual value of 50 from the graph and used 47 instead. Some found 51.814 but did not know how to proceed further.

e 59% 92% was a very common incorrect answer, indicating students didn't realise they had to square the correlation coefficient to find the coefficient of determination.

## EXERCISE 3.5 Data transformations

ANSWERS p. 702

### Recap

1 ©VCAA 2012 1CQ10 51% Which one of the following statistics is never negative?

A a median B a residual C a standardised score

D an interquartile range E a correlation coefficient

2 ©VCAA 20091CQ11 47% The table lists the average *body weight* (in kg) and average *brain weight* (in g) of nine animal species.

A least squares line of best fit is fitted to the data using *body weight* as the explanatory variable.

The equation of the least squares line of best fit is

*brain weight* = 49.4 + 2.68 x *body weight*

This equation is then used to predict the *brain weight* (in g) of the baboon. The residual value (in g) for this prediction will be closest to

A -351 B -102

C -78 D 78

E 102

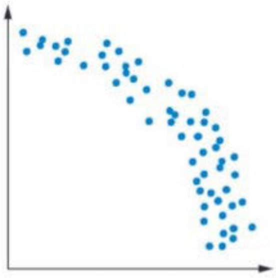
| Species | Body weight (kg) | Brain weight (g) |
|---|---|---|
| baboon | 10.55 | 179.5 |
| cat | 3.30 | 25.6 |
| goat | 27.70 | 115.0 |
| guinea pig | 1.04 | 5.5 |
| rabbit | 2.50 | 12.1 |
| rat | 0.28 | 1.9 |
| red fox | 4.24 | 50.4 |
| rhesus monkey | 6.80 | 179.0 |
| sheep | 55.50 | 175.0 |

 9780170448406

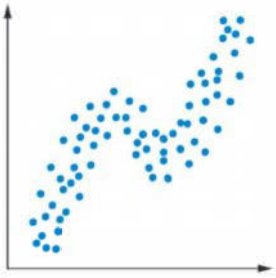
## Mastery

3 For each of the following scatterplots, state which of the transformations $x^2$,^2, logx, logy, $\frac{1}{x}$ or $\frac{1}{y}$ (if any) you would use to linearise the data.

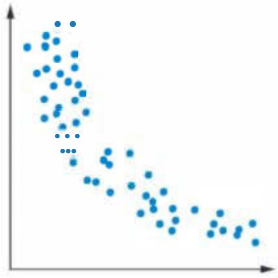
a



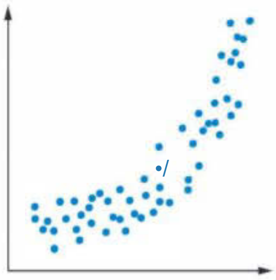
b



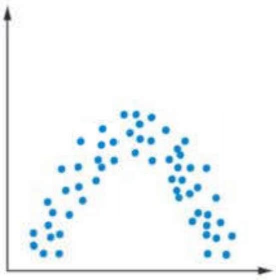
c



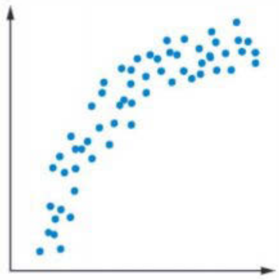
d



e

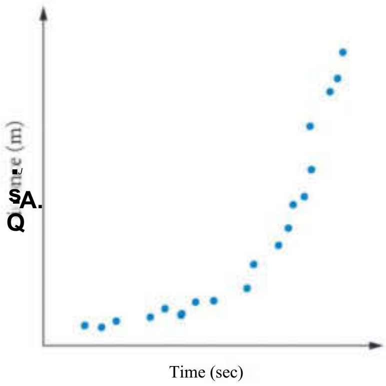


f

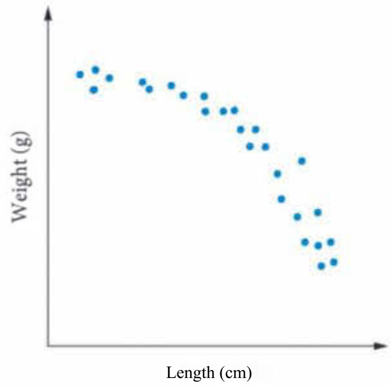


4 Ea WORKED EXAMPLE 8 I For each of the following scatterplots, write the equations of the lines of best fit after applying the appropriate transformations (squared, log and reciprocal) that linearise the data.

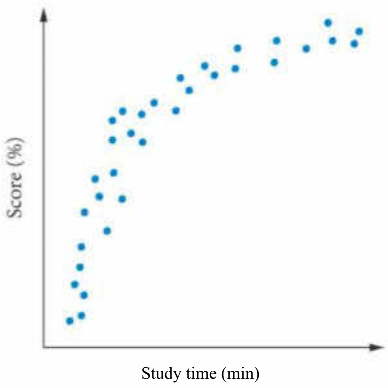
a



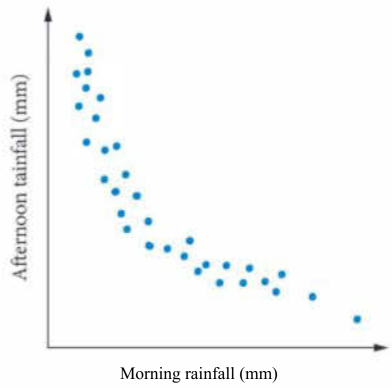

b




c




d



5S WORKED EXAMPLE 9 The data for the association between the diameter of a particular fruit (cm) and the number of seeds it contains has been linearised by applying two separate transformations, giving the following least squares line of best fit equations.

*(number of seeds)*$^2$ = 36 + 8 x *diameter*

*number of seeds* = 4 + 6.6 x log *(diameter)*

a For each of the following, use the equation to predict the number of seeds in a fruit of diameter

i 8 cm   ii 4 cm   iii 2 cm

b What shape is the original data?

c The coefficient of determination was calculated to be 0.79 for the squared transformation and 0.88 for the log transformation. Which of the two equations give the best fit to the data? Give a reason for your answer.

6 a & Using CAS 6A JS. using CAS 6EF Identify the shape of the graphed data of the values in the table shown and hence list suitable transformations that could be used to linearise the data.

| $x$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $y$ | 28 | 18 | 6 | 3 | 2 |

b For each of the transformations from part a, write a table of values.

c Use the tables of values in part b to calculate the coefficient of determination for each of the transformations.

d Use the coefficients of determination from part c to decide which of the transformations is the best choice for linearising the data.

e For your best transformation choice in part d, write the equation of the least squares line in terms of the transformed variables, with the slope and intercept correct to two decimal places.

## Exam practice

80-100% 60-79% 0-59%

7 ©VCAA 20201CQ14J 74% In a study, the association between the *number of tasks* completed on a test and the *time* allowed for the test, in hours, was found to be non-linear. The data can be linearised using a $\log_{10}$ transformation applied to the variable *number of tasks.*

The equation of the least squares line for the transformed data is

$\log_{10}$ (number *of tasks)* = 1.160 + 0.03617 x *time*

This equation predicts that the *number of tasks* completed when the *time* allowed for the test is three hours is closest to

A 13   B 16   C 19   D 25   E 26

8 ©VCAA 20151CQ11J 69% A log transformation is used to linearise the relationship between the *weight* of a mouse, in grams, and its *age,* in weeks. When a least squares line of best fit is fitted to the transformed data, its equation is

*weight* = -7 + 30$\log_{10}$(a#e)

This equation predicts that a mouse aged five weeks has a weight, in grams, that is closest to

A 14   B 21   C 23   D 41   E 143

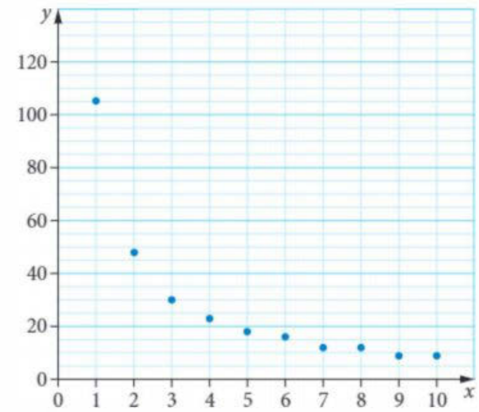


9780170448406

3.5

9 ©VCAA 2018ICQ11 , 58% Freya uses the following data to generate the scatterplot.

| X | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| y | 105 | 48 | 35 | 23 | 18 | 16 | 12 | 12 | 9 | 9 |



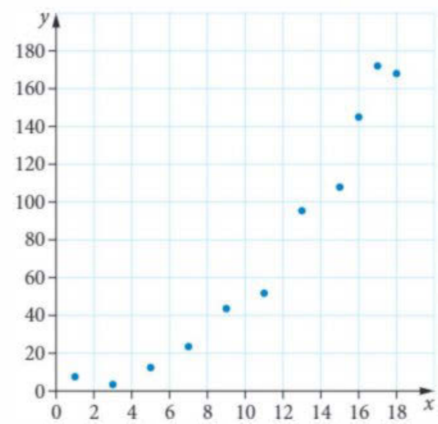

The scatterplot shows that the data is non-linear. To linearise the data, Freya applies a reciprocal transformation to the variable *y*. She then fits a least squares line to the transformed data. With *x* as the explanatory variable, the equation of this least squares line is closest to

A $\frac{1}{y}$ = -0.0039+ 0.012x     B $\frac{1}{y}$ = -0.025+ 1.lx     C $\frac{1}{y}$ = 7.8 - 0.082x

D *y* = 45.3 + 59.7 x- $\frac{1}{X}$     E *y* - 59.7 + 45.3 x - $\frac{1}{x}$

10 ©VCAA 2019 1CQ12 57% The table below shows the values of two variables *x* and *y*. The associated scatterplot is also shown. The explanatory variable is *x*.

| X | y |
|---|---|
| 1 | 7.6 |
| 3 | 3.4 |
| 5 | 12.1 |
| 7 | 23.4 |
| 9 | 43.6 |
| 11 | 51.8 |
| 13 | 95.4 |
| 15 | 108 |
| 16 | 145 |
| 17 | 172 |
| 18 | 168 |




The scatterplot is non-linear. A squared transformation applied to the variable *x* can be used to linearise the scatterplot. The equation of the least squares line fitted to the linearised data is closest to

A *y* = -1.34 + 0.546X     B *y* = -1.34 + 0.546$X^2$     C *y* = 3.93 - 0.008 64$X^2$

D y =34.6-10.5*     E *y* = 34.6 - 10.5$X^2$

11 ©VCAA | 2018ICQ12J 54% A $\log_{10}(y)$ transformation was used to linearise a set of non-linear bivariate data. A least squares line was then fitted to the transformed data. The equation of this least squares line is $\log_{10}(y)$ = 3.1 - 2.3*. This equation is used to predict the value ofy when *x* = 1.1. The value ofy is closest to

A -0.24     B 0.57     C 0.91     D 1.6     E 3.7

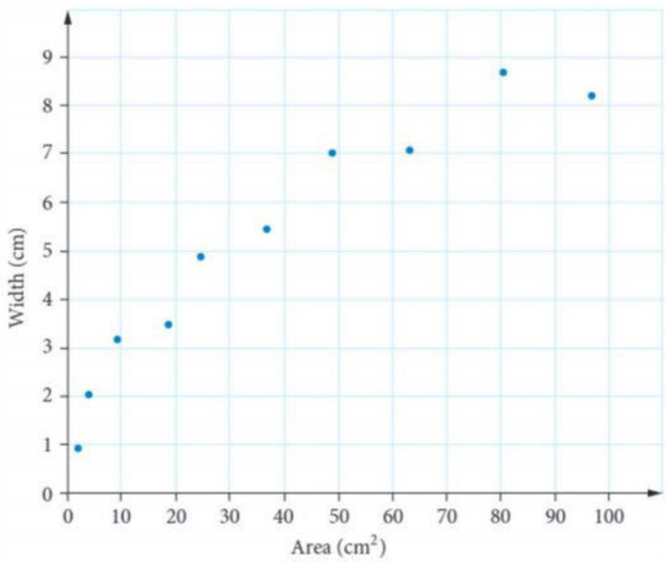
12 ©VCAA | 20131CQ10I 53% The data in the scatterplot shows the *width,* in cm, and the surface *area,* in cm$^2$, of leaves sampled from 10 different trees. The scatterplot is non-linear.

To linearise the scatterplot, $(width)^2$ is plotted against *area* and a least squares line of best fit is then fitted to the linearised plot.

The equation of this least squares line of best fit is

$$(width)^2 = 1.8 + 0.8 \times area$$

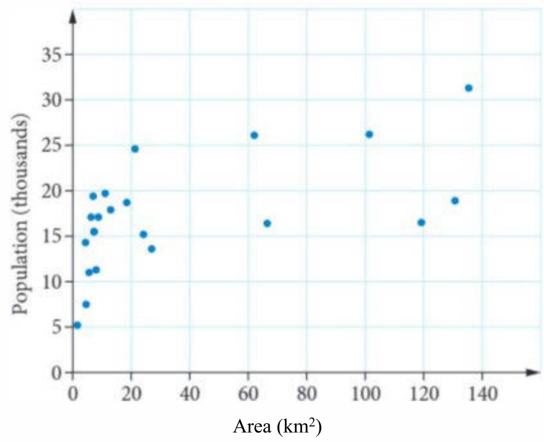
Using this equation, a leaf with a surface area of 120 cm$^2$ is predicted to have a width, in cm, closest to




A 9.2 B 9.9 C 10.6 D 84.6 E 97.8

13 ©VCAA 2016S 2CQ4 (3 marks) The scatterplot and table show the *population,* in thousands, and the *area,* in square kilometres, for a sample of 21 outer suburbs of the same city.

| Area (km$^2$) | Population (thousands) |
|---|---|
| 1.6 | 5.2 |
| 4.4 | 14.3 |
| 4.6 | 7.5 |
| 5.6 | 11.0 |
| 6.3 | 17.1 |
| 7.0 | 19.4 |
| 7.3 | 15.5 |
| 8.0 | 11.3 |
| 8.8 | 17.1 |
| 11.1 | 19.7 |
| 13.0 | 17.9 |
| 18.5 | 18.7 |
| 21.3 | 24.6 |
| 24.2 | 15.2 |
| 27.0 | 13.6 |
| 62.1 | 26.1 |
| 66.5 | 16.4 |
| 101.4 | 26.2 |
| 119.2 | 16.5 |
| 130.7 | 18.9 |
| 135.4 | 31.3 |





9780170448406

In the outer suburbs, the relationship between *population* and *area* is non-linear. A log transformation can be applied to the variable *area* to linearise the scatterplot.

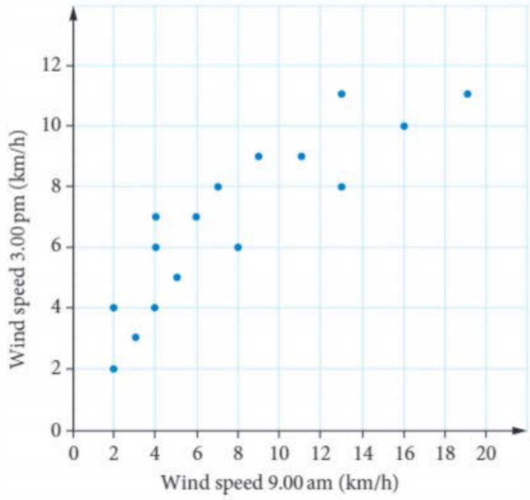
a Apply the log transformation to the data and determine the equation of the least squares line of best fit that allows the population of an outer suburb to be predicted from the logarithm of its area. Copy and complete the following by adding the slope and intercept in the boxes of this least squares line of best fit. Round your answers to two significant figures.

*population* = ☐ + ☐ log *(area)* 2 marks

b Use the equation of the least squares of best fit in part a to predict the population of an outer suburb with an area of 90 km$^2$. Round your answer to the nearest one thousand people. 1 mark

14 ©VCAA 2012 2CQ4 J (3 marks) The wind speeds (in km/h) that were recorded at the weather station at 9.00 am and 3.00 pm respectively on 18 days in November are given in the table. A scatterplot has been constructed from this data set.

| Wind speed (km/h) | | | |
|---|---|---|---|
| 9.00 am | 3.00 pm | 9.00 am | 3.00 pm |
| 2 | 2 | 13 | 8 |
| 4 | 6 | 11 | 9 |
| 4 | 7 | 2 | 4 |
| 4 | 4 | 7 | 8 |
| 13 | 11 | 5 | 5 |
| 6 | 7 | 8 | 6 |
| 3 | 3 | 6 | 7 |
| 16 | 10 | 19 | 11 |
| 6 | 7 | 9 | 9 |




Let the wind speed at 9.00 am be represented by the variable *ws9.00am* and the wind speed at 3.00 pm be represented by the variable *ws3.00pm.*

The association between *ws9.00am* and *ws3.00pm* shown in the scatterplot is non-linear.

A squared transformation can be applied to the variable *ws3.00pm* to linearise the data in the scatterplot.

a Apply the squared transformation to the variable *ws3:00pm* and determine the equation of the least squares line of best fit that allows $(ws3.00ptn)^2$ to be predicted from *ws9:00am.*

Copy and complete the equation below by writing the coefficients for this equation, correct to one decimal place.

$(ws3.00pm)^2$ = ☐ + ☐ x *ws9.00am* 2 marks

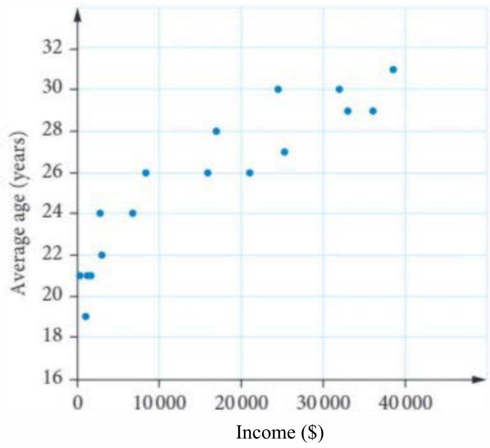
b Use this equation to predict the wind speed at 3.00 pm on a day when the wind speed at 9.00 am is 24 km/h. Write your answer correct to the nearest whole number. 1 mark

15 ©VCAA 2011 2CQ4 34% (3 marks) The average age of women at first marriage in years *(average age)* and average yearly income in dollars per person *(income)* were recorded for a group of 17 countries.

The results are displayed in the table. A scatterplot of the data is also shown.

| Average age (years) | Income ($) |
|---|---|
| 21 | 1750 |
| 22 | 3200 |
| 26 | 8600 |
| 26 | 16000 |
| 28 | 17000 |
| 26 | 21000 |
| 30 | 24500 |
| 30 | 32000 |
| 31 | 38500 |

| Average age (years) | Income ($) |
|---|---|
| 29 | 33000 |
| 27 | 25 500 |
| 29 | 36000 |
| 19 | 1300 |
| 21 | 600 |
| 24 | 3050 |
| 24 | 6900 |
| 21 | 1400 |




The association between *average age* and *income* is non-linear.

A log transformation can be applied to the variable *income* and used to linearise the scatterplot.

a Apply this log transformation to the data and determine the equation of the least squares line of best fit that allows *average age* to be predicted from *log(income)*.

Copy and complete the coefficients for this equation, correct to two decimal places, in the spaces provided.

*average age* = [ ] + [ ] x log (income) 2 marks

**Exam hack**

The first column in a data table isn't always the explanatory variable.

b Use this equation to predict the average age of women at first marriage in a country with an average yearly income of $20000 per person.

Write your answer correct to one decimal place. 1 mark

**Exam hack**

Always check the real-life context for your answer. If your answer looks to be impossible in real life, then it's almost certainly wrong.


9780170448406

# (£) Chapter summary



### Line of best fit

- A line of best fit is a straight line that is the best approximation for a set of data.
- The equation for the least squares line of best fit is $y = a + bx$ where
  - the slope or gradient of the line is $b = r\frac{s_y}{s_x}$
  - the y-intercept of the line is $a = \bar{y} - b\bar{x}$
  - r is the Pearson correlation coefficient
  - $s_x$ and $s_y$, are the sample standard deviations of $x$ and $y$ respectively
  - $\bar{x}$ and $\bar{y}$ are the sample means of $x$ and $y$ respectively.
- They-intercept is *a*. This means they variable is *a* units when the *x* variable is zero units.
- The slope is *b*. This means they variable on average increases/decreases by *b* units for every 1-unit increase in the *x* variable.

### The coefficient of determination ($r^2$)

- The coefficient of determination ($r^2$)
  - can be calculated by squaring the Pearson correlation coefficient (r)
  - is a value between 0 and 1
  - is a measure of how useful a line of best fit is as a linear model for a particular set of data (0 means it is a totally useless measure and 1 means its a perfect measure)
  - is usually converted to a percentage.
- $r^2 \times 100\%$ of the variation in the response variable can be explained by the variation in the explanatory variable.
- A high coefficient of determination (70% and above) indicates that the line of best fit is an appropriate model for the data.
- When calculating r from $r^2$, use the slope of the line of best fit or the scatterplot to see if *r* is positive or negative.

### Making predictions

- Values can be predicted by substituting values into the equation of the least squares line of best fit and solving.
- Predicting *within* the original data range is called interpolation.
- Predicting *outside* the original data range is called extrapolation.
- Predictions based on extrapolation are not as reliable as those based on interpolation.

## Residual analysis

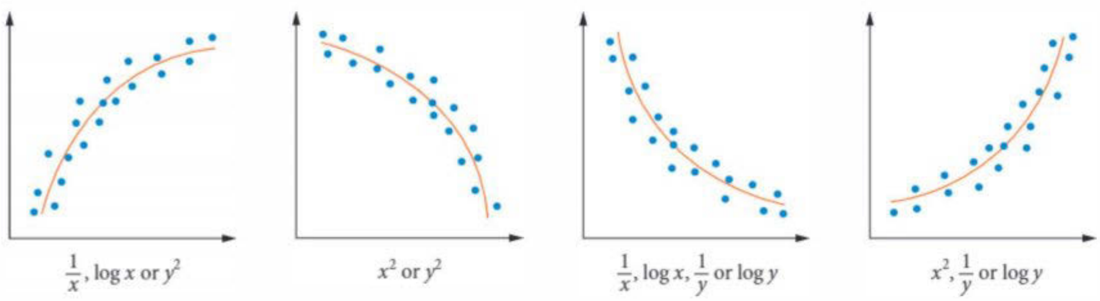
- A residual is the vertical distance between each data point and the least squares line of best fit.
  residual value = actual value - predicted value
  - The actual value can be read from a scatterplot or a table.
  - The predicted value needs to be calculated from the equation of the least squares line of best fit.
- Actual values that lie
  - above the least squares line of best fit will have a positive residual value
  - below the least squares line of best fit will have a negative residual value
  - on the least squares line of best fit will have a residual value of zero.
- The further the actual value is from the least squares line of best fit, the larger the residual value.

## Residual plot

- A residual plot is like a scatterplot with the explanatory variable on the x-axis and the residual values on the y-axis.
- When residual values
  - are randomly scattered above and below the x-axis, the association between the original two variables, x and y, is probably *linear*
  - show a hill or a valley shape, the association between the original two variables, x and y, is probably *non-linear*.

## Data transformations

- Data transformations are used to linearise data that isn't linear.
- The three data transformations we use are:
  - squared transformation
  - log transformation
  - reciprocal transformation.
- Choose the transformation options depending on the shape of the data.
- To choose the best option, calculate the coefficient of determination of each to see which is the closest to 1.





9780170448406

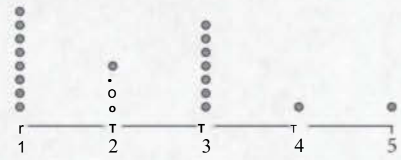
# Cumulative examination 1



Total number of marks: 11 Reading time: 5 minutes Writing time: 25 minutes

*Use the following information to answer the next two questions.*

The dot plot shows the distribution of the number of bedrooms in each of 21 apartments advertised for sale in a new high-rise apartment block.




1 ©VCAA 2007 1CQ1 The mode of this distribution is

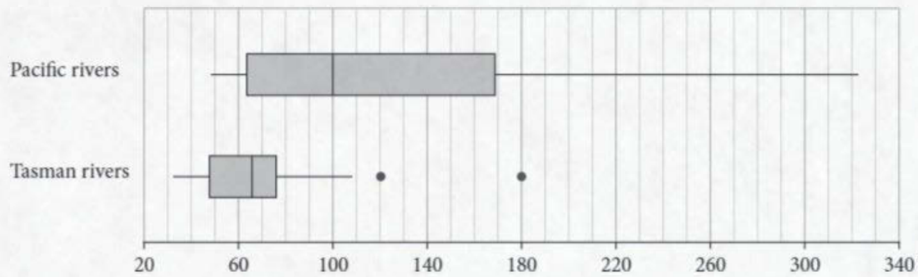
A 1 B 2 C 3 D 7 E 8

2 ©VCAA 2007 1CQ2 The median of this distribution is

A 1 B 2 C 3 D 4 E 5

*Use the following information to answer the next two questions.*

In New Zealand, rivers flow into either the Pacific Ocean (the Pacific rivers) or the Tasman Sea (the Tasman rivers). The boxplots shown can be used to compare the distribution of the lengths of the Pacific rivers and the Tasman rivers.




Source: *The New Zealand Yearbook*, 1982

3 ©VCAA 2015 1CQ6 The five-number summary for the lengths of the Tasman rivers is closest to

A 32,48,64,76,108 B 32,48,64,76,180 C 32,48,64,76,322
D 48,64,97,169,180 E 48,64,97,169,322

4 ©VCAA 2015 1CQ7 Which one of the following statements is ***not*** true?

A The lengths of two of the Tasman rivers are outliers.

B The median length of the Pacific rivers is greater than the length of more than 75% of the Tasman rivers.

C The Pacific rivers are more variable in length than the Tasman rivers.

D More than half of the Pacific rivers are less than 100 km in length.

E More than half of the Tasman rivers are greater than 60 km in length.

5 ©VCAA 20081CQ10 A large study of Year 12 students shows that there is a negative association between the time spent doing homework each week and the time spent watching television. The correlation coefficient is $r = -0.6$.

From this information it can be concluded that

A the time spent doing homework is 60% lower than the time spent watching television.

B 36% of students spend more time watching television than doing homework.

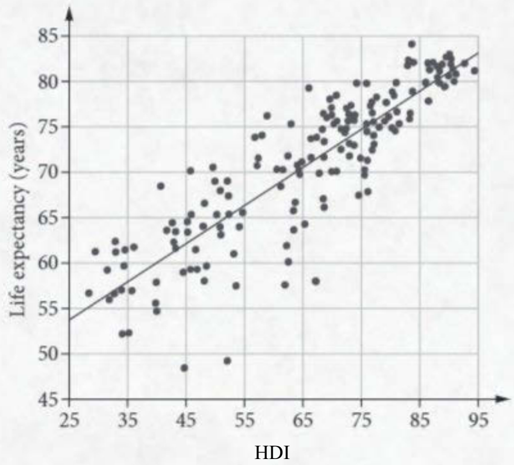
C the slope of the line of best fit is 0.6.

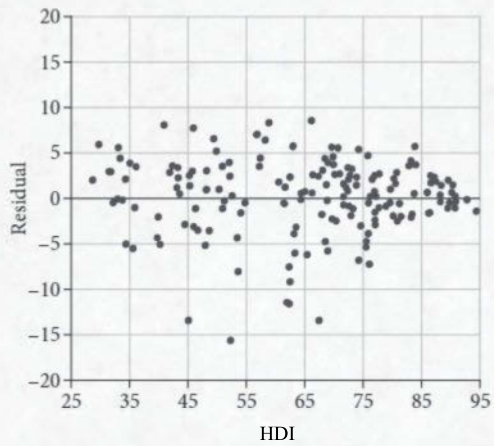
D if a student spends less time watching television, they will do more homework.

E an increased time spent watching television is associated with a decreased time doing homework.

*Use the following information to answer the next two questions.*

The scatterplot below shows life expectancy in years *(life expectancy)* plotted against the Human Development Index *(HDI)* for a large number of countries in 2011. A least squares line has been fitted to the data and the resulting residual plot is also shown.




Data: Gapminder

The equation of this least squares line is

*life expectancy* = 43.0 + 0.422 x *HDI*

The coefficient of determination is $r^2 = 0.875$.

6 ©VCAA 2016 1CQ9 Given the information above, which one of the following statements is not true?

A The value of the correlation coefficient is close to 0.94.

B 12.5% of the variation in life expectancy is not explained by the variation in the Human Development Index.

C On average, life expectancy increases by 43.0 years for each 10-point increase in the Human Development Index.

D Ignoring any outliers, the association between life expectancy and the Human Development Index can be described as strong, positive and linear.

E Using the least squares line to predict the life expectancy in a country with a Human Development Index of 75 is an example of interpolation.

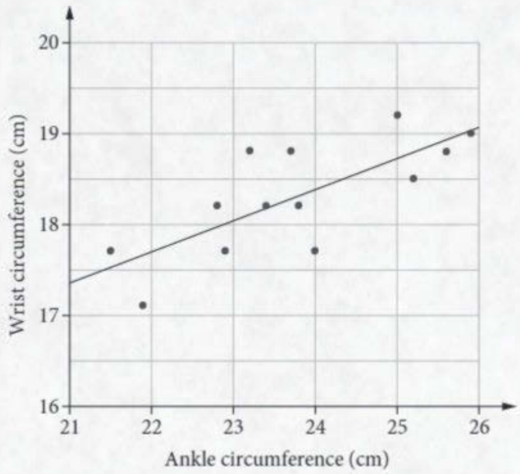
7 ©VCAA 2016ICQ10J In 2011, life expectancy in Australia was 81.8 years and the Human Development Index was 92.9. When the least squares line is used to predict life expectancy in Australia, the residual is closest to

A -0.6 B -0.4 C 0.4 D 11.1 E 42.6


9780170448406

*Use the following information to answer the next three questions.*

The scatterplot shows the *wrist* circumference and *ankle* circumference, both in centimetres, of 13 people. A least squares line has been fitted to the scatterplot with *ankle* circumference as the explanatory variable.




8 ©VCAA 2017 1CQ8 The equation of the least squares line is closest to

A *ankle* = 10.2 + 0.342 x *wrist*

B *wrist* = 10.2 + 0.342 x *ankle*

C *ankle* = 17.4 + 0.342 x *wrist*

D *wrist* = *17.4* + 0.342 x *ankle*

E *wrist* = 17.4 + 0.731 x *ankle*

9 ©VCAA 20171CQ9 I When the least squares line on the scatterplot is used to predict the wrist circumference of the person with an ankle circumference of 24 cm, the residual will be closest to

A -0.7 B -0.4 C -0.1 D 0.4 E 0.7

10 ©VCAA | 2OI7ICQIO~] The residuals for the least squares line have a mean of 0.02 cm and a standard deviation of 0.4 cm. The value of the residual for one of the data points is found to be -0.3 cm. The standardised value of this residual is

A -0.8 B -0.7 C -0.3 D 0.7 E 0.8

11 ©VCAA 2021 1CQ11 The table shows the *weight,* in kilograms, and the *height,* in centimetres, of 10 adults.

| Weight (kg) | Height (cm) |
|---|---|
| 59 | 173 |
| 67 | 180 |
| 69 | 184 |
| 84 | 195 |
| 64 | 173 |
| 74 | 180 |
| 76 | 192 |
| 56 | 169 |
| 58 | 164 |
| 66 | 180 |

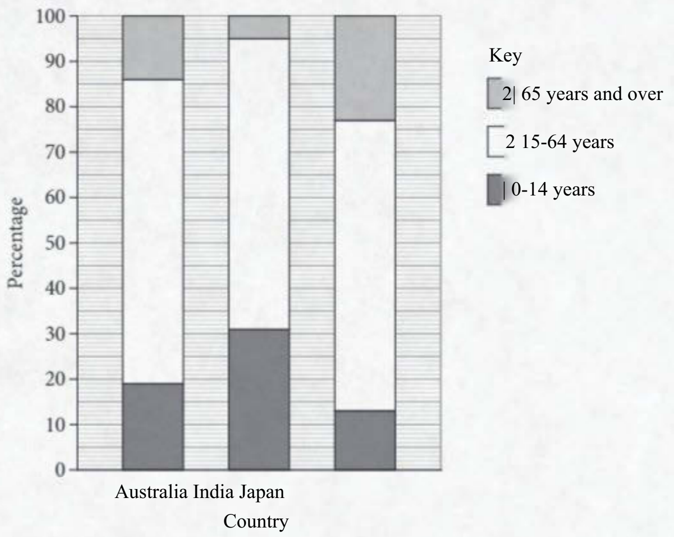
A least squares line is fitted to the data. The least squares line enables an adults *weight* to be predicted from their *height.* The number of times that the predicted value of an adult's *weight* is greater than the actual value of their *weight* is

A3 B4 C 5 D6 E7

# Cumulative examination 2

Total number of marks: 18 Reading time: 5 minutes Writing time: 27 minutes

1 © VCAA 2016S 2CQ1 (3 marks) The segmented bar chart shows the age distribution of people in three countries, Australia, India and Japan, for the year 2010.



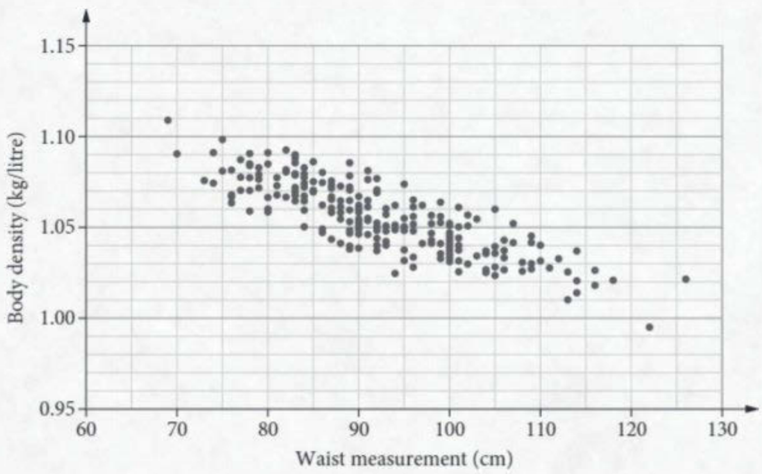

Source: Australian Bureau of Statistics, 3201.0 - *Population by Age and Sex, Australian States and Territories*, June 2010

a Write down the percentage of people in Australia who were aged 0-14 years in 2010. 1 mark

b In 2010, the population of Japan was 128 000 000. How many people in Japan were aged 65 years and over in 2010? 1 mark

c From the graph, it appears that there is no association between the percentage of people in the 15-64 age group and the country in which they live. Explain why, quoting appropriate percentages to support your explanation. 1 mark

2 © VCAA 2020 2CQ5 (7 marks) The scatterplot shows *body density,* in kilograms per litre, plotted against *waist measurement,* in centimetres, for 250 men.




Data: RW Johnson, 'Fitting percentage of body fat to simple body measurements', *Journal of Statistics Education*, 4:1,1996, https://doi.Org/10.1080/10691898.1996.11910505

When a least squares line is fitted to the scatterplot, the equation of this line is

*body density* = 1.195 - 0.001512 x *waist measurement*

a Copy the scatterplot and draw the graph of this least squares line on the scatterplot, 1 mark

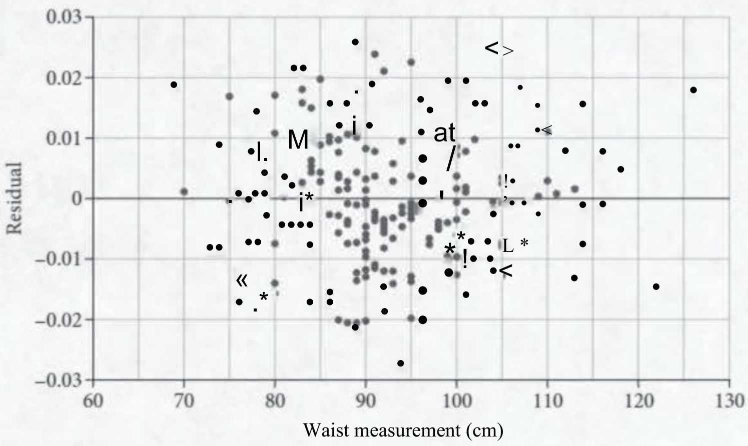
b Use the equation of this least squares line to predict the *body density* of a man whose *waist measurement* is 65 cm. Round your answer to two decimal places. 1 mark c When using the equation of this least squares line to make the prediction in part b, are you extrapolating or interpolating? 1 mark

d Interpret the slope of this least squares line in terms of a mans *body density* and *waist measurement.* 1 mark

*e* In this study, the body density of the man with a waist measurement of 122 cm was 0.995 kg/litre. Show that, when this least squares line is fitted to the scatterplot, the residual, rounded to two decimal places, is -0.02. 1 mark

f The coefficient of determination for this data is 0.6783. Write down the value of the correlation coefficient r. Round your answer to three decimal places. 1 mark

g The residual plot associated with fitting a least squares line to this data is shown below.




Does this residual plot support the assumption of linearity that was made when fitting this line to this data? Briefly explain your answer. 1 mark

3 ©VCAA 2021 2CQ2 (3 marks) In the sport of heptathlon, athletes compete in seven events. The two running events are the 200 m run and the 800 m run. The times taken by the athletes in these two events, *time200* and *time800,* are linearly related. When a least squares line is fitted to the data, the equation of this line is found to be

*time800* = 0.03931 + 5.2756 x *time200*

a Round the values for the intercept and the slope to three significant figures. Copy and complete the following and write your answers in the boxes.

*time800* = ☐ + ☐ x *time200* 1 mark

b The mean and the standard deviation for each variable, *time200* and *time800,* are shown in the table.

| Statistic | Time200 (seconds) | Time800 (seconds) |
|---|---|---|
| Mean | 24.6492 | 136.054 |
| Standard deviation | 0.96956 | 8.2910 |

The equation of the least squares line is

*time800* = 0.03931 + 5.2756 x *time200*

Use this information to calculate the coefficient of determination as a percentage. Round your answer to the nearest percentage. 2 marks

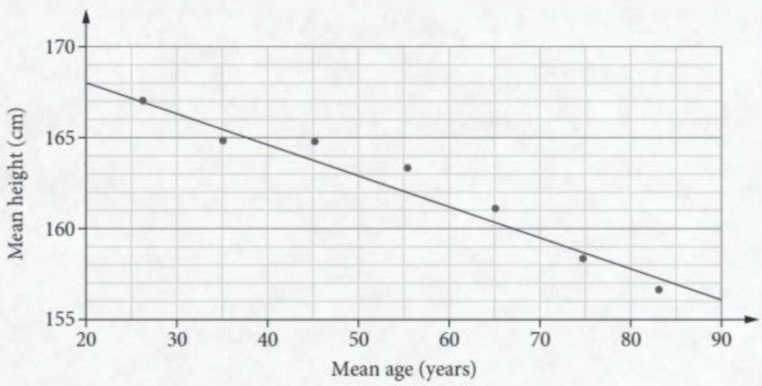
4 ©VCAA J 2020 2CQ6 (5 marks) The table shows the *mean age,* in years, and the *mean height,* in centimetres, of 648 women from seven different age groups.

Data: J Sorkln et al., 'Longitudinal change In height of men and women: Implications for Interpretation of the body mass Index', *American Journal of Epidemiology*, vol. 150, no. 9,1999, p. 971

| | Age group | | | | | | |
|---|---|---|---|---|---|---|---|
| | Twenties | Thirties | Forties | Fifties | Sixties | Seventies | Eighties |
| Mean age (years) | 26.3 | 35.2 | 45.2 | 55.3 | 65.1 | 74.8 | 83.1 |
| Mean height (cm) | 167.1 | 164.9 | 164.8 | 163.4 | 161.2 | 158.4 | 156.7 |

a What was the difference, in centimetres, between the *mean height* of the women in their twenties and the *mean height* of the women in their eighties? 1 mark

A scatterplot displaying this data shows an association between the *mean height* and the *mean age* of these women. In an initial analysis of the data, a line is fitted to the data by eye, as shown.

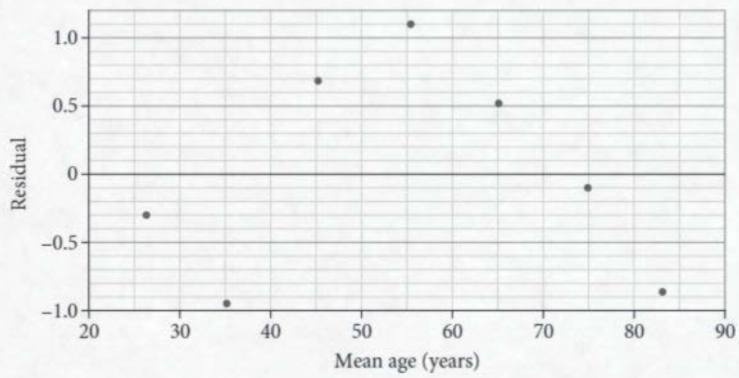



b Describe this association in terms of strength and direction. 1 mark

c The line on the scatterplot passes through the points (20,168) and (85,157). Using these two points, determine the equation of this line. Copy and complete the equation shown by writing the values of the intercept and the slope in the appropriate boxes. Round your answers to three significant figures.

*mean height* = ☐ + ☐ x *mean age* 1 mark

d In a further analysis of the data, a least squares line was fitted. The associated residual plot that was generated is shown below.




The residual plot indicates that the association between the *mean height* and the *mean age* of women is non-linear. It can be linearised by applying an appropriate transformation to the variable *mean age.*

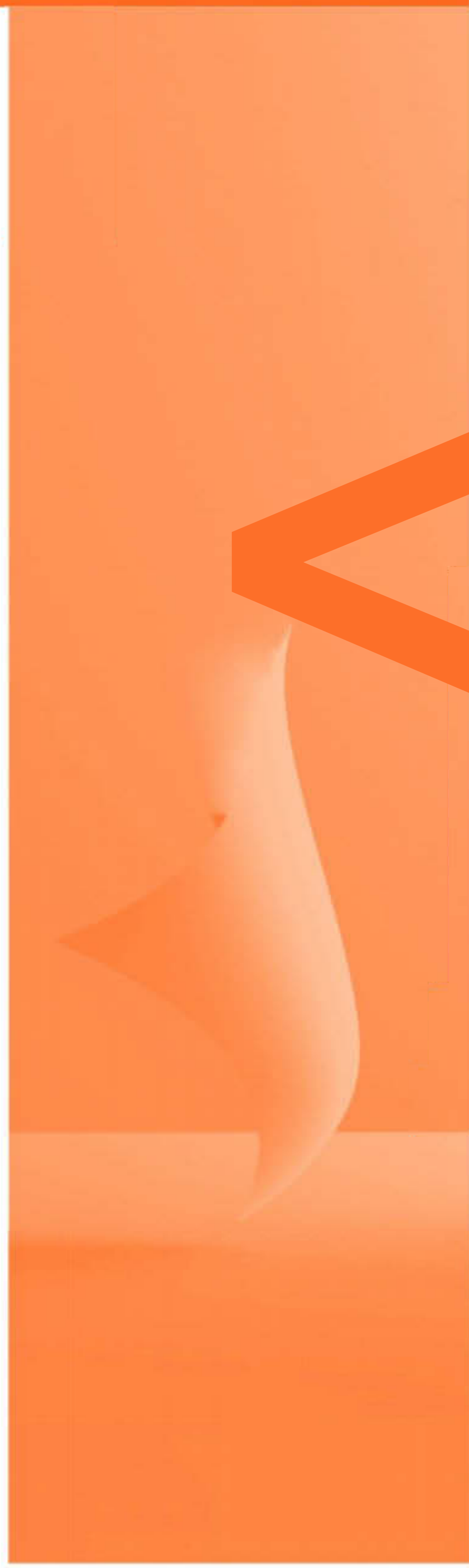
| Mean age (years) | 26.3 | 35.2 | 45.2 | 55.3 | 65.1 | 74.8 | 83.1 |
|---|---|---|---|---|---|---|---|
| Mean height (cm) | 167.1 | 164.9 | 164.8 | 163.4 | 161.2 | 158.4 | 156.7 |

Apply an appropriate transformation to the variable *mean age* to linearise the data. Fit a least squares line to the transformed data and write its equation. Round the values of the intercept and the slope to four significant figures. 2 marks

CHAPTER 4

# TIME SERIES





## Study Design coverage

### AREA OF STUDY 1: DATA ANALYSIS, PROBABILITY AND STATISTICS

#### Investigating and modelling time series data

- qualitative features of time series plots; recognition of features such as trend (long-term direction), seasonality (systematic, calendar related movements) and irregular fluctuations (unsystematic, short-term fluctuations); possible outliers and their sources, including one-off real-world events, and signs of structural change such as a discontinuity in the time series
- numerical smoothing of time series data using moving means with consideration of the number of terms required (using centring when appropriate) to help identify trends in time series plot with large fluctuations
- graphical smoothing of time series plots using moving medians (involving an odd number of points only) to help identify long-term trends in time series with large fluctuations
- seasonal adjustment including the use and interpretation of seasonal indices and their calculation using seasonal and yearly means
- modelling trend by fitting a least squares line to a time series with time as the explanatory variable (data de-seasonalised where necessary), and the use of the model to make forecasts (with re-seasonalisation where necessary) including consideration of the possible limitations of fitting a linear model and the limitations of extending into the future.



**Video playlists (6):**

- **4.1** Time series plots
- **4.2** Numerical smoothing
- **4.3** Graphical smoothing
- **4.4** Seasonal adjustment
- **4.5** Least squares trend lines
- **VCE question analysis** Time series

**Worksheets (7):**

- **4.1** Time series plots
- **4.2** Smoothing time series data • Moving means
- **4.3** Moving medians
- **4.4** Seasonal adjustment • De-seasonalisation • Seasonality

9

^Nelson MindTap

To access resources above, visit cengage.com.au/nelsonmindtap



9780170448406

# @ Time series plots

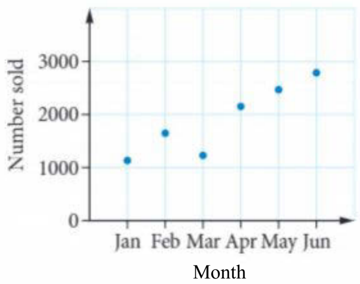
4.1

## Time series data and plots

A time series involves data where the explanatory variable is time, usually measured at equally spaced intervals such as hours, days, weeks, months, seasons or years.

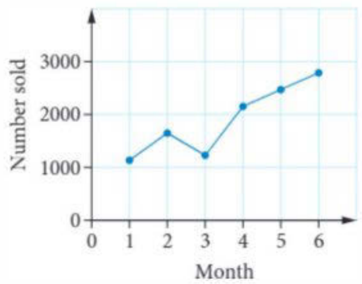
A time series plot is a scatterplot of a time series where the data points are joined in order by straight lines. The time divisions on the horizontal axis are often simplified into numbers 1, 2, 3 ... to make them easier to work with.

Scatterplot




Time series plot


Simplified time series plot

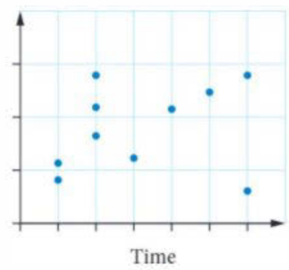



Video playlist Time series plots

Worksheet Time series plots

### Exam hack

In a scatterplot with *time* as the explanatory variable, you can have more than one data point for a specific time value. In a time series plot you can't. For example, you can't draw a time series plot for the data shown on this scatterplot:



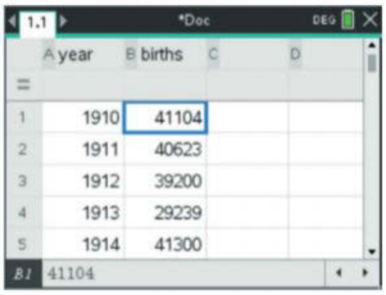

**USING CAS 1** Constructing time series plots

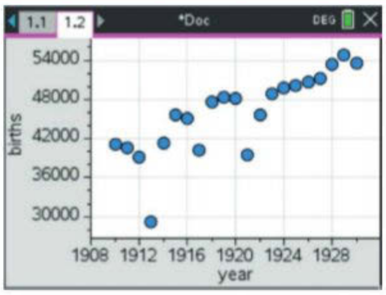
Create a time series plot using this table showing births in a state from 1910 to 1930.

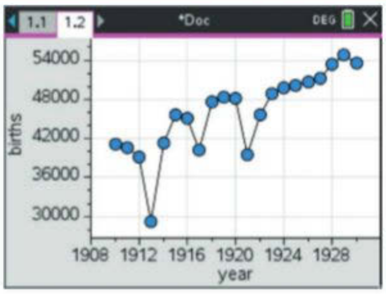
| Year | Total live births |
|---|---|
| 1910 | 41 104 |
| 1911 | 40 623 |
| 1912 | 39 200 |
| 1913 | 29 239 |
| 1914 | 41 300 |
| 1915 | 45 601 |
| 1916 | 45 122 |
| 1917 | 40 123 |
| 1918 | 47 563 |
| 1919 | 48 323 |
| 1920 | 48 124 |

| Year | Total live births |
|---|---|
| 1921 | 39 576 |
| 1922 | 45 623 |
| 1923 | 48 789 |
| 1924 | 49 792 |
| 1925 | 50 124 |
| 1926 | 50 678 |
| 1927 | 51 237 |
| 1928 | 53 456 |
| 1929 | 54 789 |
| 1930 | 53 627 |

## TI-Nspire




1 Start a new document and add a Lists & Spreadsheet page.

2 Label the columns and enter the data from the table as shown above.

3 Insert a Data & Statistics page.

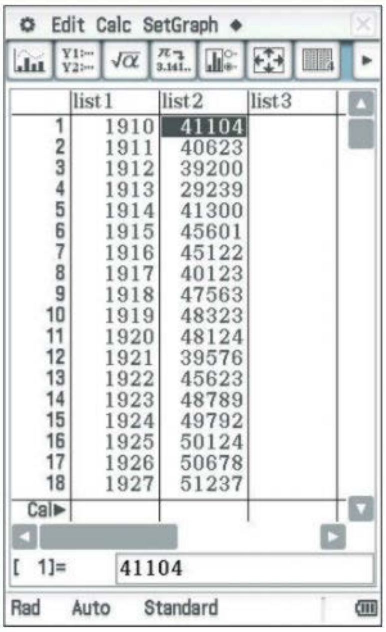
4 For the horizontal axis, select year.

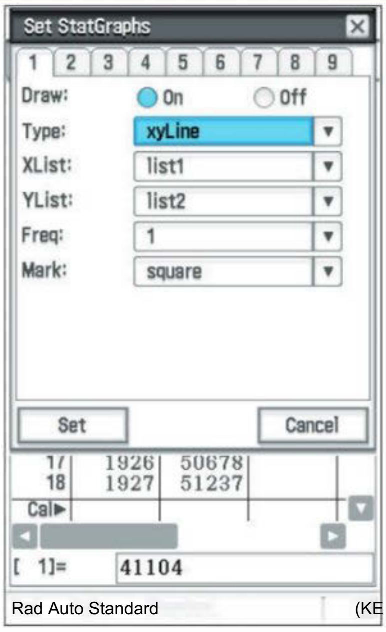
5 For the vertical axis, select births.

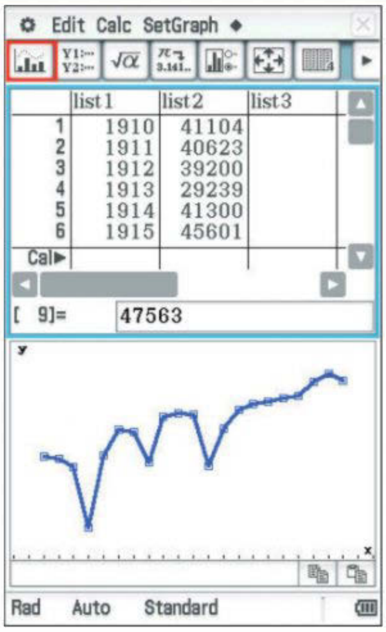
6 Press menu > Plot Type > XY Line Plot.

7 The time series plot will be displayed.

## ClassPad




1 Open the Statistics application.

2 Clear all lists and enter the data from the table as shown.

3 Tap SetGraph > Setting.

4 Change the Type: field to xyLine.

5 Tap Set.

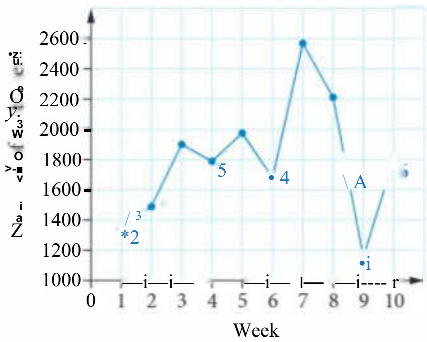
6 Tap Graph.

7 The time series plot will be displayed in the lower window.

4.1

## WORKED EXAMPLE 1 Finding the median of a time series

This time series plot shows the number of customers in a store each week over a ten-week period. Find the median number of weekly customers during this period.


| Steps | Working |
| --- | --- |
| 1 Find the number of data values, n, and note if the number is odd or even. | $n = 10$; even |
| 2 Find the position of the median. | $\frac{n+1}{2} = \frac{10+1}{2} = \frac{11}{2} = 5.5$ <br> The median is between the 5th and 6th value after the values have been ordered from smallest to largest. |
| 3 Count up from the lowest value until you get to the position of the median. Read the values from the graph. <br> If $n$ is odd, find the middle value. <br> If $n$ is even, average the two middle values. |  <br> The 5th value is 1800. The 6th value is 1800. <br> median $= \frac{1800 + 1800}{2} = 1800$ customers |



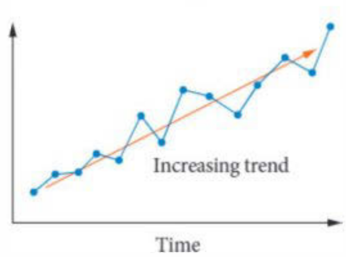
## Features of time series plots

Time series plots can be described by identifying the following features.

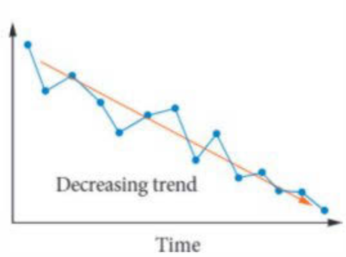
### Trend

The trend is the direction of the data over a significant period of time. Time series can have multiple trends.

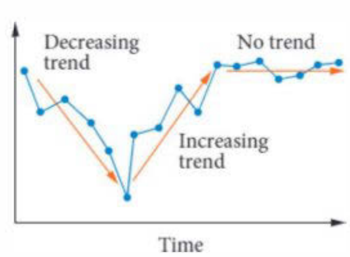
Increasing trend




Decreasing trend



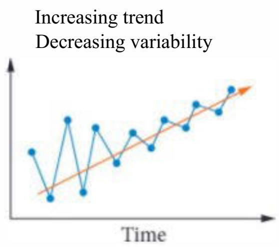

Multiple trends

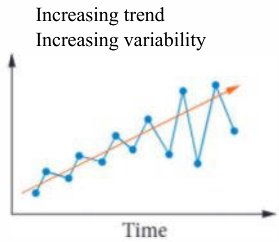


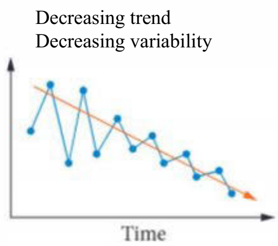

WB
p. 42

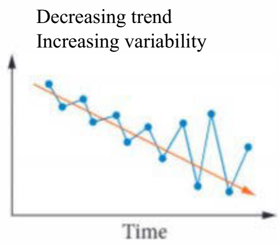
## Variability

Variability of a time series is how much the data points differ from each other. Variability can be increasing or decreasing regardless of the trend.



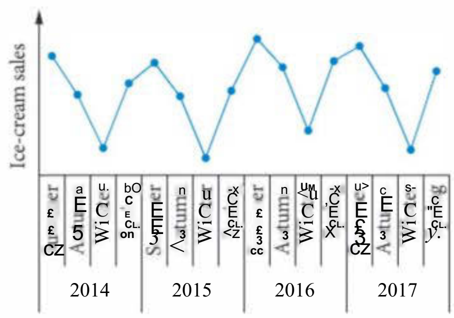



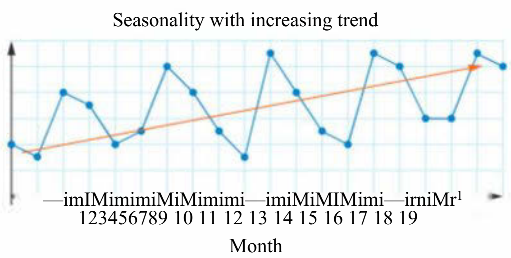

## Seasonality

Seasonality describes data that has regular and predictable changes in time periods of a year or less, such as days, weeks, months, quarters, or actual seasons. Retail stores, for example, would regularly and predictably sell more on Saturday and Sunday than the other days of the week. Ice-cream sales each year would regularly and predictably peak during summer and dip during winter.



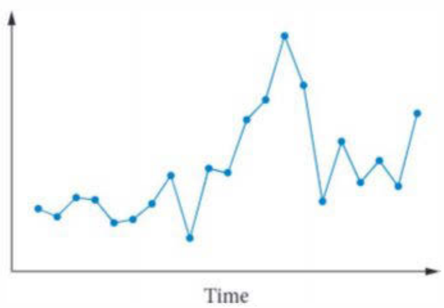

> **© Exam hack**
>
> To decide if a graph is showing seasonality, check to see if the peaks always occur at the same time. For example, always on a Sunday or always in summer. If they don't, then it's not showing seasonality.

It's possible for data to show both seasonality and a trend at the same time. The time series on the right has regular peaks every four months, as well as having an increasing trend:




## Irregular fluctuations

Irregular fluctuations (also called irregular variations) are the sudden random changes that occur in a time series that cant be explained by trends and seasonality. All real-world time series plots will have some irregular fluctuations.



**WORKED EXAMPLE 2** Identifying seasonality in time series

For each of the following, state whether a time series of the data is likely to show seasonality. Explain your reasoning.

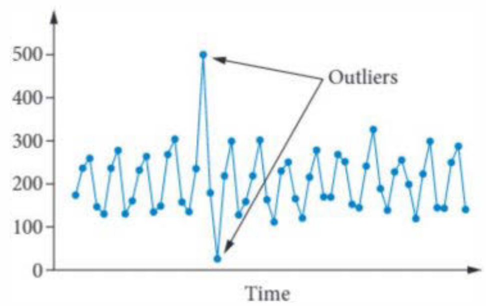
a sales of air conditioners
b sales of bread
c sales of dishwashers
d number of interstate visitors with school-age children in Melbourne

| Steps | Working |
| --- | --- |
| Is there likely to be a pattern throughout the year?<br>If so, what time period would the pattern be based on? | a Sales of air conditioners are likely to show seasonality because more would be bought in summer than in winter.<br>b Sales of bread are not likely to show seasonality because bread is eaten regularly each day.<br>c Sales of dishwashers are likely to show seasonality because more people would have the time to shop for them on weekends rather than on weekdays.<br>d The number of interstate visitors with school-age children to Melbourne are likely to show seasonality because there would be a larger number during school holidays and few otherwise. |

WB
p. 43

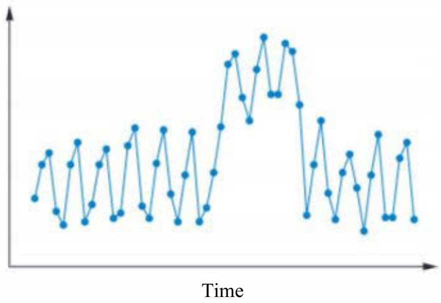
## Outliers in time series

Outliers in a time series can be easy to identify, as in the example shown below. The difficulty is deciding whether to keep the outliers in the data analysis or take them out. For example, if the outlier was the result of a one-off freak storm over a weather station, we would remove it, but if it was a measurement on a heart monitor we wouldn't.

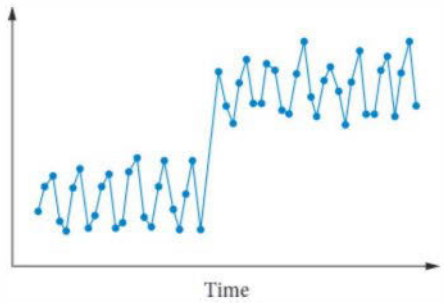


## Structural changes in time series

A structural change in a time series occurs when there is an unexpected shift in the data pattern. It is more significant than the appearance of a few outliers, as shown in this example:




Discontinuities are structural changes that are clear breaks in the time series. This plot shows structural change involving an obvious discontinuity:




Structural changes, and discontinuities in particular, are common in the real world and can lead to incorrect conclusions. As with outliers, it is important to work out the reasons for the changes.

Possible sources of structural change include significant extended events or technological change. For example, the introduction of geostationary satellite imagery in the 1970s affected weather forecasting and climate data. Similarly, a war or a pandemic can cause a disruption to economic data over time. Discontinuities can be caused by a change of measuring equipment, new measurement techniques, or inventions such as vaccines.

### EXERCISE 4.1 Time series plots

ANSWERS p. 704

#### Mastery

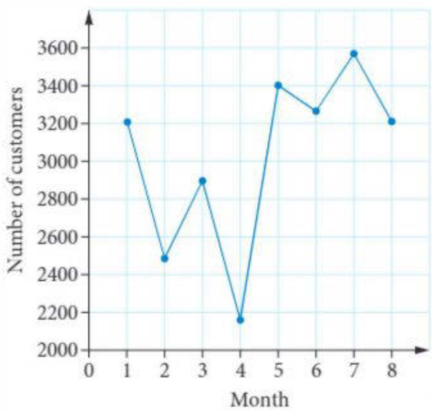
1 s using CAS 1J Create a time series plot using this table showing births by year in a state from 1910 to 1920.

Births by year in a state from 1910 to 1920

| Year | Total live births |
|---|---|
| 1910 | 11124 |
| 1911 | 12000 |
| 1912 | 12200 |
| 1913 | 13329 |
| 1914 | 13 500 |
| 1915 | 14005 |

| Year | Total live births |
|---|---|
| 1916 | 14100 |
| 1917 | 14127 |
| 1918 | 15163 |
| 1919 | 15323 |
| 1920 | 16224 |

20 WORKED EXAMPLE 1 This time series plot shows the number of customers in a service station each month over an eight-month period. Find the median number of monthly customers during this period.





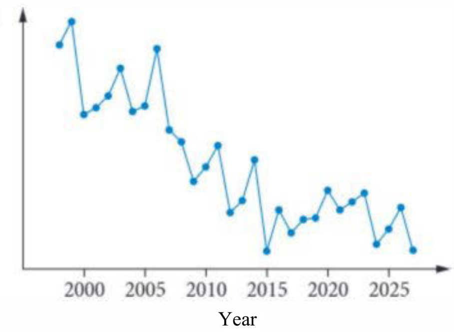

9780170448406

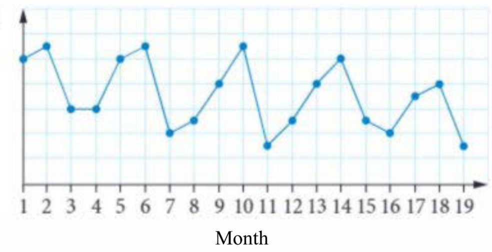
▶ 3H WORKED EXAMPLE 2 | For each of the following, state whether a time series of the data is likely to show seasonality. Explain your reasoning.

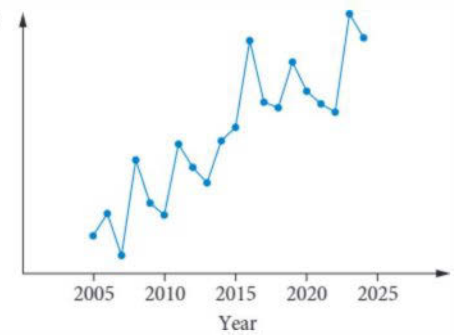
a sales of umbrellas
b sales of shorts
c sales of television sets
d consumption of milk
e occupancy rates at a holiday resort
f money earned by teachers
g money earned by fruit pickers
h number of people attending AFL matches.

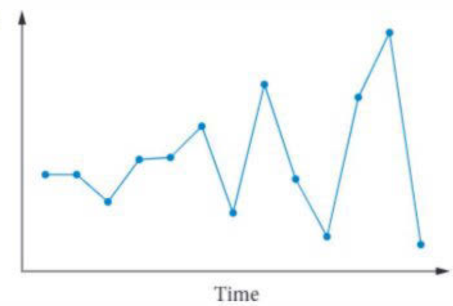
4 State which of the following features are shown in each of the time series plots below.

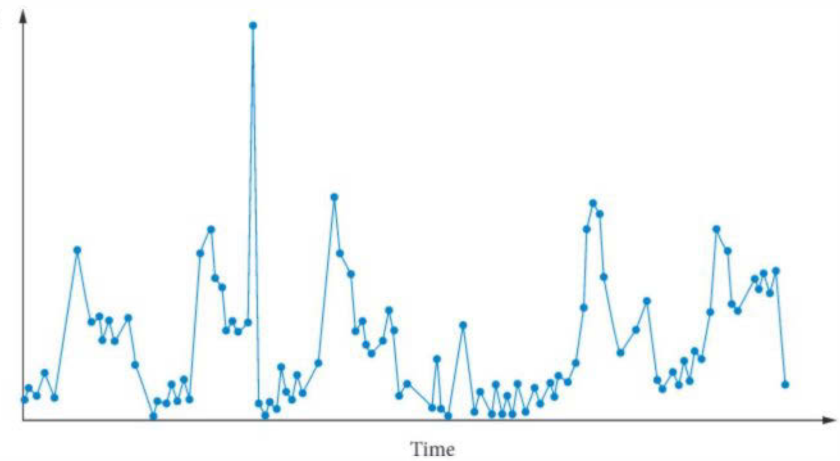
- increasing trend
- decreasing trend
- increasing variability
- decreasing variability
- outliers
- structural change
- seasonality
- irregular fluctuations

a



b



c


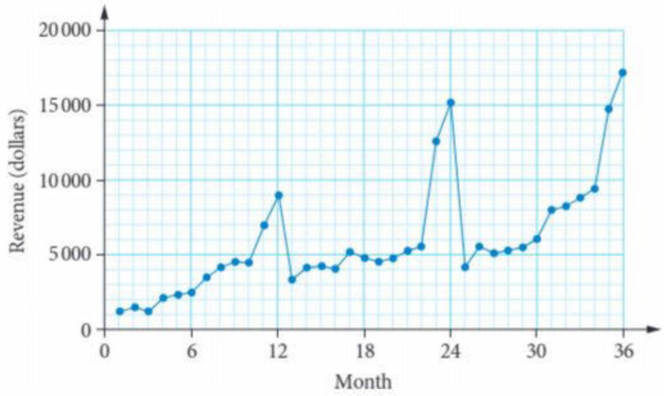

d



e


## Exam practice

80-100% 60-79% 0-59%

5 ©VCAAJ 2007 ICQITJ 91% The time series plot shows the revenue from sales (in dollars) each month made by a Queensland souvenir shop over a three-year period.




This time series plot indicates that, over the three-year period, revenue from sales each month showed

A no overall trend.

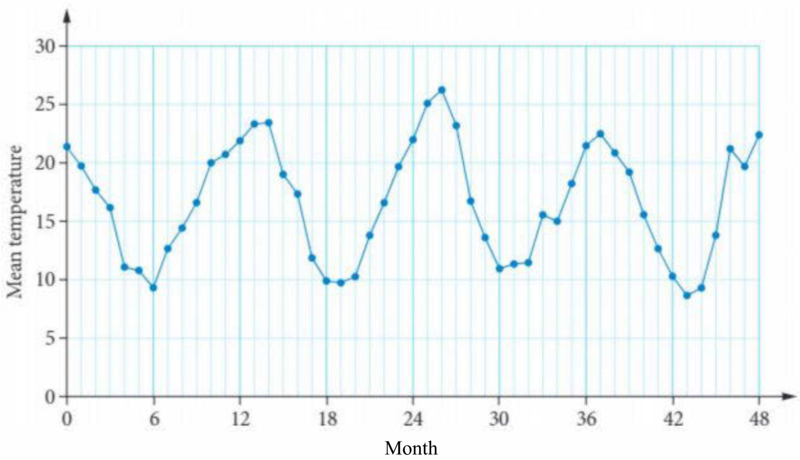
B no correlation.

C positive skew.

D an increasing trend only.

E an increasing trend with seasonal variation.

6 CVCAA 20161CQ13 I 72% Consider the following time series plot.




Data: Commonwealth of Australia, Bureau of Meteorology

The pattern in the time series plot shown above is best described as having

A irregular fluctuations only.

B an increasing trend with irregular fluctuations.

C seasonality with irregular fluctuations.

D seasonality with an increasing trend and irregular fluctuations.

E seasonality with a decreasing trend and irregular fluctuations.

7 ©VCAA 2019N1CQ13 The association between *amount of protein* consumed (in grams/day) and *family income* (in dollars) is best displayed using

A a scatterplot.

B a time series plot.

C parallel boxplots.

D back-to-back stem plots.

E a two-way frequency table.

8 ©VCAA J 2012 ICQT I 59% The temperature of a room is measured at hourly intervals throughout the day. The most appropriate graph to show how the temperature changes from one hour to the next is a

A boxplot. B stem plot. C histogram.

D time series plot. E two-way frequency table.

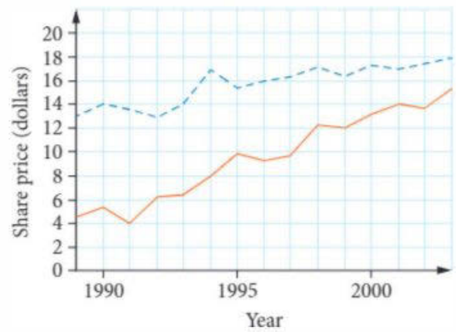
9 ©VCAA 20201CQ19 J 45% The time series plot below displays the number of airline *passengers,* in thousands, each month during the period January to December 1960.


Data: GEP Box and GM Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, 1970, p. 531

During 1960, the median number of monthly airline *passengers* was closest to

A 461000 B 465000 C 471000 D 573000 E 621000

10 ©VCAA 20041CQ12J 31% The time series plot shows the share price of two companies over a period of time. From the plot, it can be concluded that over the interval 1990-2000, the difference in share price between the two companies has shown



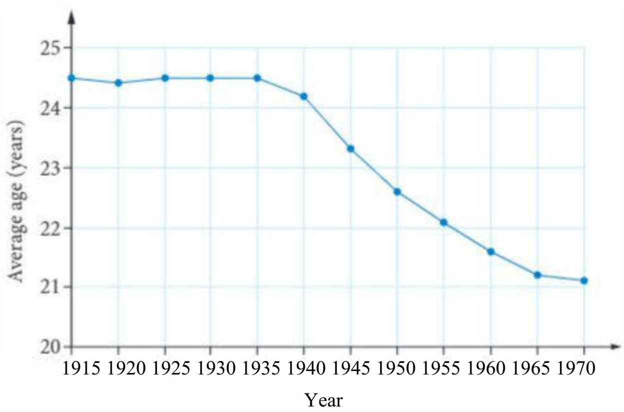

## Exam hack

When a word is in bold in a question, pay special attention to it and make sure you're clear why it's important.

A a decreasing trend. B an increasing trend. C seasonal variation.

D a five-year cycle. E no trend.

**11** ©VCAA [2011 2CQ3a J (1 mark) The time series plot shows the average age of women at first marriage in a particular country during the period 1915 to 1970. Use this plot to describe the way in which the average age of women at first marriage in this country has changed during the period 1915 to 1970.



Video playlist Numerical smoothing

Worksheets Smoothing time series data

Moving means

# 9.3 Numerical smoothing

Smoothing time series data is a technique for levelling out fluctuations to produce a smoother graph that lets us see the underlying trend more clearly. One way of achieving this is through numerical smoothing techniques.

## Smoothing using moving means

Moving means is the most common numerical smoothing technique. It involves finding a series of means of a fixed number of data points. The simplest method is to smooth using an odd number of data points, such as three, five or seven.

WB p. 44

### WORKED EXAMPLE 3 Moving mean smoothing with an odd number of points

The following table shows the number of students in a Year 12 General Mathematics class over the last 10 years.

| Year | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Number of students | 25 | 18 | 23 | 21 | 19 | 20 | 18 | 16 | 17 | 15 |

a Use a table with three columns to calculate the smoothed data using the method of three-point moving means.

b What are the smoothed number of students for the fifth and tenth years?

c Graph the original data and the smoothed data on the same set of axes.

d What does the graph of the smoothed data indicate about the trend in the original data?

**Steps**

a 1 Set up a table with three columns with Year in column 1, Number of students in column 2, and the calculations for three-point moving means in column 3.

2 Find the mean for each group of three consecutive values for the number of students.

3 Write the mean in the row of the middle number of the three consecutive values.

Note: You cannot find a moving mean for the first and last value for the number of students.

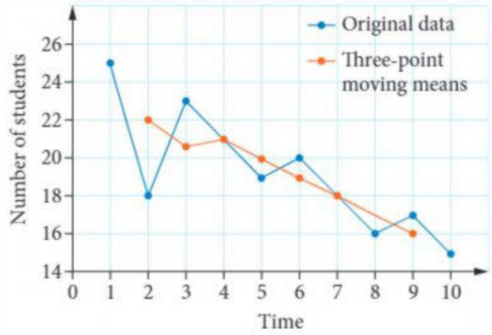
**Working**

| Year | Number of students | Three-point moving means |
|---|---|---|
| 1 | 25 | |
| 2 | 18 | $\frac{25+18+23}{3} = 22$ |
| 3 | 23 | $\frac{18+23+21}{3} = 20.6$[illegible] |
| 4 | 21 | $\frac{23+21+19}{3} = 21$ |
| 5 | 19 | $\frac{21+19+20}{3} = 20$ |
| 6 | 20 | $\frac{19+20+18}{3} = 19$ |
| 7 | 18 | $\frac{20+18+16}{3} = 18$ |
| 8 | 16 | $\frac{18+16+17}{3} = 17$ |
| 9 | 17 | $\frac{16+17+15}{3} = 16$ |
| 10 | 15 | |

b Read from the table.

The smoothed number of students for the fifth year is 20. There was not enough data to calculate the smoothed number of students for the tenth year.

c Graph the original data and the smoothed data on the same set of axes.




d Identify whether the smoothed graph shows an increasing or decreasing trend.

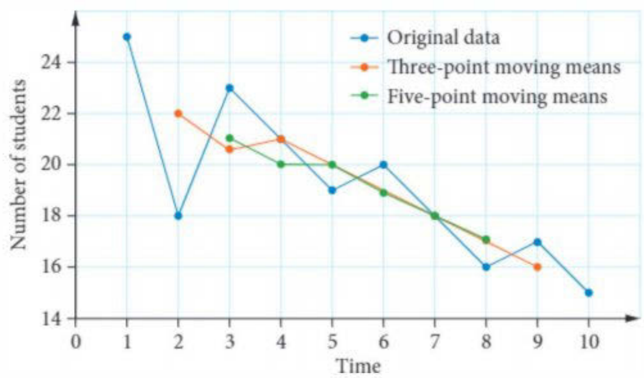
The graph of the smoothed data indicates a decreasing trend.

In three-point moving mean smoothing, the first and last points are lost because there isn't enough data to calculate them. In five-point moving mean smoothing, the first two and the last two points are lost. This pattern continues as the number of odd points increases.

Using the data in the previous worked example, five-point moving means smoothing is shown in the table on the right.

| Year | Number of students | Five-point moving means |
|---|---|---|
| 1 | 25 | |
| 2 | 18 | |
| 3 | 23 | $\frac{25+18+23+21+19}{5} = 21.2$ |
| 4 | 21 | $\frac{18+23+21+19+20}{5} = 20.2$ |
| 5 | 19 | $\frac{23+21+19+20+18}{5} = 20.2$ |
| 6 | 20 | $\frac{21+19+20+18+16}{5} = 18.8$ |
| 7 | 18 | $\frac{19+20+18+16+17}{5} = 18$ |
| 8 | 16 | $\frac{20+18+16+17+15}{5} = 17.2$ |
| 9 | 17 | |
| 10 | 15 | |

The result of the smoothing using five-point moving means can be seen in the graph. The larger the number of points used for the moving mean, the greater the smoothing effect.




## Smoothing with an even number of points

Smoothing with an even number of points is more complicated than smoothing with an odd number of points because the centre of an even number of points isn't at one of the original time values. We deal with this using a process called centring, which involves finding the mean of each of the two consecutive smoothed values.

| WORKED EXAMPLE 4 | Moving mean smoothing with an even number of points |
|---|---|

WB

p. 46

The following table shows the number of births per month over a calendar year in a country hospital.

| Month | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of births | 10 | 12 | 6 | 5 | 22 | 18 | 13 | 7 | 9 | 10 | 8 | 15 |

a Use a table with four columns to calculate the smoothed data using the method of four-point moving means with centring.

b What is the smoothed number of births for April and November?

c Graph the smoothed data.

d What does the graph of the smoothed data indicate about the trend in the original data?

e Find the smoothed data values for February and March using the two-point moving means method with centring.

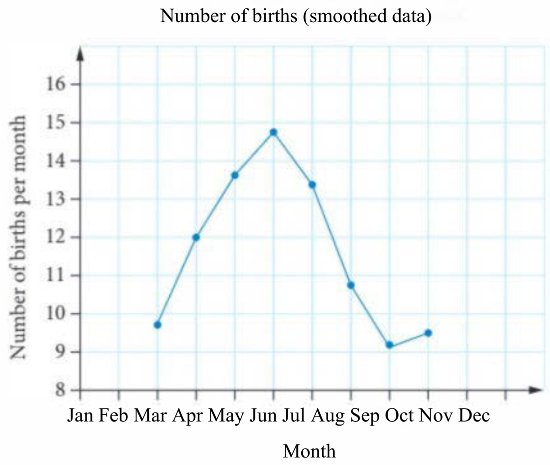
**Steps**

**Working**

a Set up a table with four columns and extra rows between time intervals for centring of four-point moving means.

| Month | Number of births | Four-point moving means | Four-point moving means with centring |
|---|---|---|---|
| Jan | 10 | | |
| Feb | 12 | | |
| | | $\frac{10+12+6+5}{4} = 8.25$ | |
| Mar | 6 | | $\frac{8.25+11.25}{2} = 9.75$ |
| | | $\frac{12+6+5+22}{4} = 11.25$ | |
| Apr | 5 | | $\frac{11.25+12.75}{2} = 12$ |
| | | $\frac{6+5+22+18}{4} = 12.75$ | |
| May | 22 | | $\frac{12.75+14.5}{2} = 13.625$ |
| | | $\frac{5+22+18+13}{4} = 14.5$ | |
| June | 18 | | $\frac{14.5+15}{2} = 14.75$ |
| | | $\frac{22+18+13+7}{4} = 15$ | |
| July | 13 | | $\frac{15+11.75}{2} = 13.375$ |
| | | $\frac{18+13+7+9}{4} = 11.75$ | |
| Aug | 7 | | $\frac{11.75+9.75}{2} = 10.75$ |
| | | $\frac{13+7+9+10}{4} = 9.75$ | |
| Sept | 9 | | $\frac{9.75+8.5}{2} = 9.125$ |
| | | $\frac{7+9+10+8}{4} = 8.5$ | |
| Oct | 10 | | $\frac{8.5+10.5}{2} = 9.5$ |
| | | $\frac{9+10+8+15}{4} = 10.5$ | |
| Nov | 8 | | |
| Dec | 15 | | |


9780170448406

| | |
|---|---|
| b Read the values from the table. | The smoothed number of births for April is 12. There are not enough data to calculate the smoothed number of births for November. |
| c Graph the smoothed data by hand. |  |
| d Identify whether the smoothed graph shows an increasing or decreasing trend, | The graph of the smoothed data indicates an increasing trend until June and a decreasing trend from June to September. |
| e Set up a table with four columns and extra rows between time intervals for centring of two-point moving means. Read the values from the table. | see table below |



| Month | Number of births | Two-point moving means | Two-point moving means with centring |
|---|---|---|---|
| Jan | 10n | | |
| | | $\frac{10+12}{2} = \frac{22}{2} =$ \|\| | |
| Feb | 12 = | | $\frac{11+9}{2} = \frac{20}{2} =$ \|Q |
| | | $\frac{12+6}{2} = \frac{18}{2} = 9$ | |
| Mar | 6 | | $\frac{9^\wedge}{2} = \frac{145}{2} = ?25$ |
| | | $\frac{6+5}{2} = \frac{U}{2} = 5.5$ | |
| Apr | 5 | | |

The smoothed data value for February is 10 and for March is 7.25.

We also lose data points when smoothing with an even number of points. In two-point moving mean smoothing, the first and last points are lost, and in four-point moving mean smoothing, the first two and the last two points are lost and so on.

## Moving mean smoothing with an odd number of points

Three-point moving means

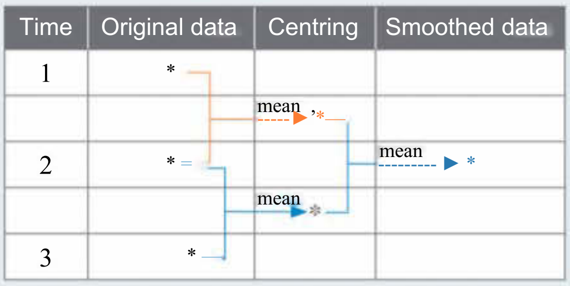
| Time | Original data | Smoothed data |
|---|---|---|
| 1 | * | |
| 2 | * | mean → * |
| 3 | * | mean → * |
| 4 | * | |

Five-point moving means

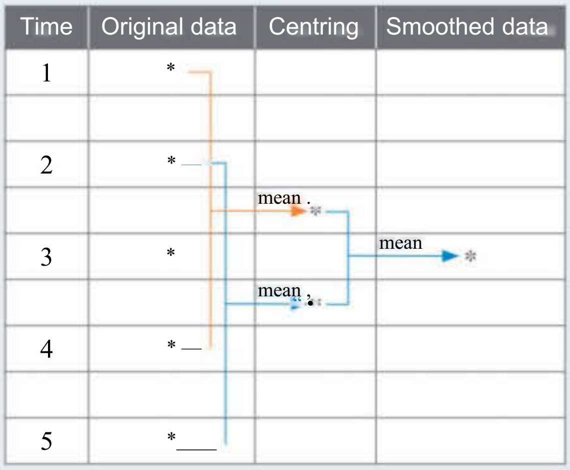
| Time | Original data | Smoothed data |
|---|---|---|
| 1 | * | |
| 2 | * | |
| 3 | * | mean → * |
| 4 | * | mean → * |
| 5 | * | |
| 6 | * | |

**Moving mean smoothing with an even number of points**

Two-point moving means with centring




Four-point moving means with centring




## Choosing the number of points for moving means

### How high can you go?

There are a number of factors to take into account when choosing how many points to use for moving mean smoothing. The larger the number of points, the greater the smoothing effect and the clearer the underlying trend, but the more data points that are lost.

- With two-point and three-point moving means, we lose two data points (one at the start and one at the end).
- With four-point and five-point moving means, we lose four data points (two at the start and two at the end).
- With six-point and seven-point moving means, we lose six data points and so on.

If we go too high, we could end up losing nearly all of the data points.

### Matching the moving means to natural cycles

The number of moving mean points is usually chosen by looking at the natural cycle of the data being considered.

| | Number of points for moving means |
|---|---|
| Daily sales figures for stores open Monday to Sunday | 7 |
| Daily sales figures for stores open only Monday to Friday | 5 |
| Monthly figures | 12 |
| Quarterly accounts | 4 |

Unless there is an even-numbered natural cycle involved, such as monthly and quarterly, it is usually better to use an odd number of points.

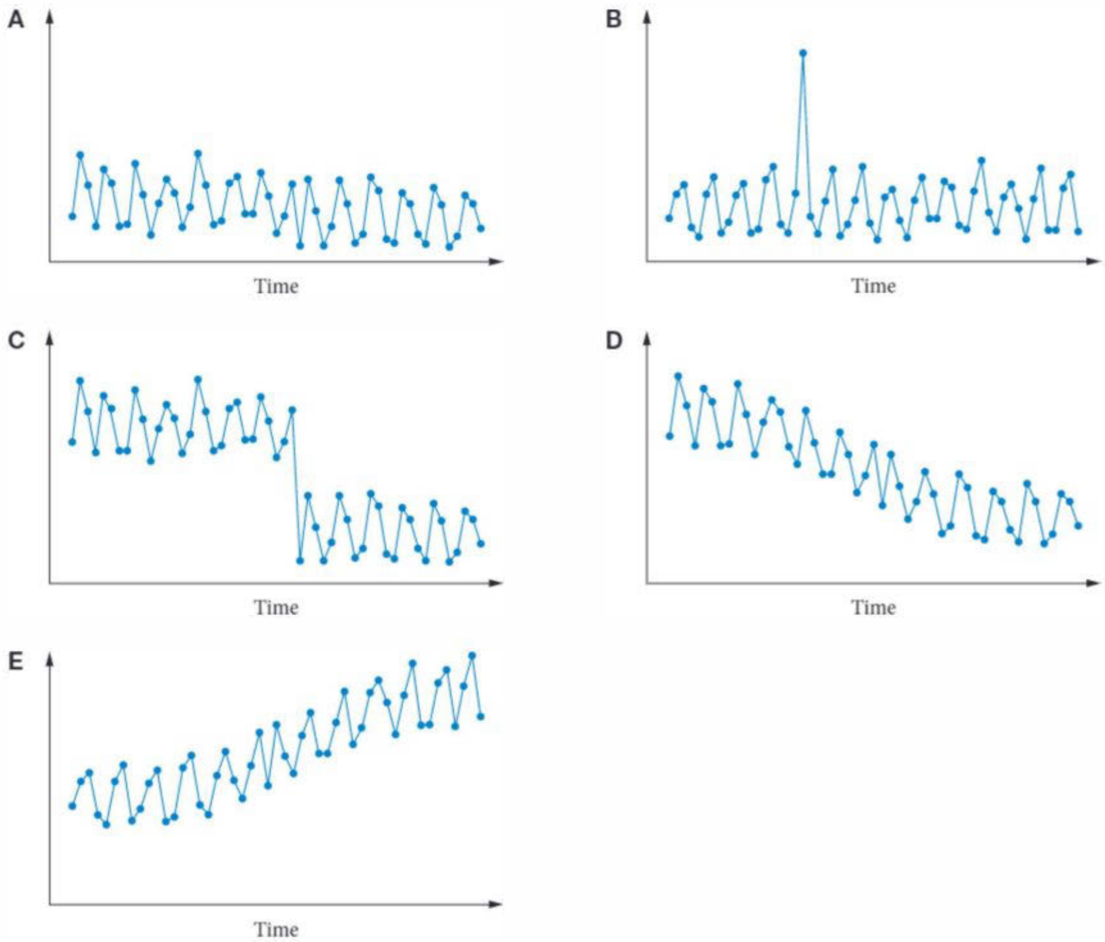

9780170448406

## EXERCISE 4.2 Numerical smoothing

ANSWERS p. 704

### Recap

1 Which one of the following time series plots shows a discontinuity?




2 A rain gauge collecting data of the amount of rain falling in a particular region was found to be affected by the wind. A wind shield was installed to minimise the impact of the wind, thus ensuring the readings were more accurate. As a result, the time series of this data could be affected by which of the following?

A outliers B seasonality C decreasing trend

D irregular fluctuations E discontinuity

## Mastery

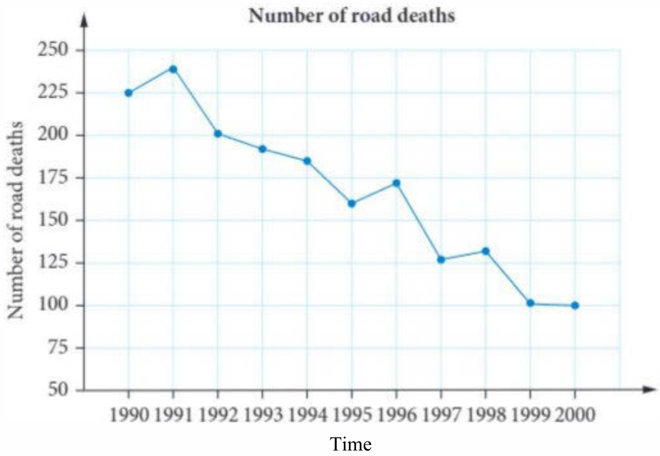
**30** WORKED EXAMPLE 3 The introduction of speed cameras in Victoria helped to reduce the number of deaths in road accidents over the period from 1990 to 2000. This data is shown below.

| Year | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of road deaths | 225 | 240 | 201 | 192 | 185 | 160 | 172 | 127 | 132 | 101 | 100 |

a Use a table with three columns to calculate the smoothed data using the method of three-point moving means.

b What are the smoothed number of road deaths in 1994 and 2000?

c Copy the following graph and add the time series plot for the smoothed data.




d What does the graph of the smoothed data indicate about the trend in the original data?

4 S WORKED EXAMPLE 4 The table shows the yearly sales of a mathematics textbook from 2010 to 2022.

| Year | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of textbook sales | 2250 | 2230 | 2000 | 2010 | 1990 | 3000 | 2045 | 2989 | 3000 | 1950 | 2120 | 2255 | 2297 |

a Use a table with four columns to calculate the smoothed data using the method of four-point moving means with centring.

b What is the smoothed number of sales for 2011 and 2019?

c Graph the smoothed data by hand.

d What does the graph of the smoothed data indicate about the trend in the original data?

e Find the smoothed data values for 2011 and 2012 using the two-point moving means method with centring.

## Exam practice 80-ioo<y0 6o-79% o-59%

5 © VCAA 2018 1CQ15 80% The table shows the monthly profit, in dollars, of a new coffee shop for the first nine months of 2018.

| Month | Jan | Feb | Mar | Apr | May | June | July | Aug | Sept |
|---|---|---|---|---|---|---|---|---|---|
| Profit ($) | 2890 | 1978 | 2402 | 2456 | 4651 | 3456 | 2823 | 2678 | 2345 |

Using four-mean smoothing with centring, the smoothed profit for May is closest to

A $2502 B $3294 C $3503 D $3804 E $4651

4.2

6 ©VCAA 20191CQ16 74% The time series plot shows the *monthly rainfall* at a weather station, in millimetres, for each *month* in 2017.


Data: adapted from © Commonwealth of Australia 2020, Bureau of Meteorology, <www.bom.gov.au>

If seven-mean smoothing is used to smooth this time series plot, the number of smoothed data points would be

A 3 B 5 C 6 D 8 E 10

7 ©VCAA 2020 1CQ18 MODIFIED The table shows the monthly rainfall for 2019, in millimetres, recorded at a weather station for each month of the year.

| | Jan | Feb | Mar | Apr | May | June | July | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Monthly rainfall (mm) | 18.4 | 17.6 | 46.8 | 23.6 | 92.6 | 77.2 | 80.0 | 86.8 | 93.8 | 55.2 | 97.3 | 69.4 |

The six-mean smoothed monthly rainfall with centring for August 2019 is closest to

A 67.8 mm B 75.9 mm C 81.3 mm D 83.4 mm E 86.4 mm

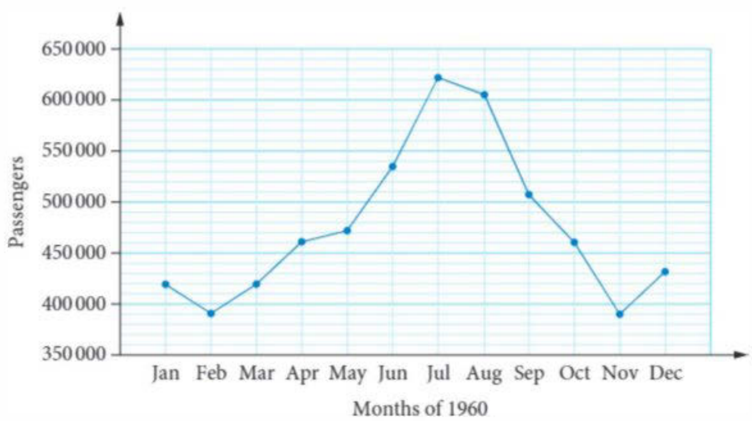
8 ©VCAA 20061CQ10J 64% The table below displays the total monthly rainfall (in mm) in a reservoir catchment area over a one-year period.

| Month | Jan | Feb | Mar | April | May | June | July | Aug | Sept | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Rainfall | 9 | 65 | 35 | 99 | 75 | 90 | 133 | 196 | 106 | 56 | 76 | 76 |

Using three-mean moving average smoothing, the smoothed value for the total rainfall in April is closest to

A 65 B 66 C 70 D 75 E 88

9 ©VCAA 20201CQ20 57% The time series plot below displays the number of airline *passengers,* in thousands, each month during the period January to December 1960.




Data: GEP Box and GM Jenkins, Time Series Analysis: Forecasting and Control, Holden-Day, San Francisco, 1970, p. 531

During the period January to May 1960, the total number of airline *passengers* was 2 160000. The five-mean smoothed number of passengers for March 1960 is

A 419000 B 424000 C 430000 D 432000 E 434000

10 ©VCAA 2011 1CQ13 55% The table below shows the number of broadband users in Australia for each of the years from 2004 to 2008.

| Year | 2004 | 2005 | 2006 | 2007 | 2008 |
|---|---|---|---|---|---|
| Number | 1012000 | 2016000 | 3900000 | 4830000 | 5140000 |

Based on data obtained from: www.data.worldbank.org

A two-point moving mean, with centring, is used to smooth the time series. The smoothed value for the number of broadband users in Australia in 2006 is

A 2958000 B 3379600 C 3455 500 D 3661500 E 3900000

11 ©VCAA | 20031CQ12J 37% The data gives the number of accidents recorded at a city intersection each year from 1993 to 2002.

Using a four-point moving average (mean) with centring, the smoothed value of the number of accidents in 1995 is

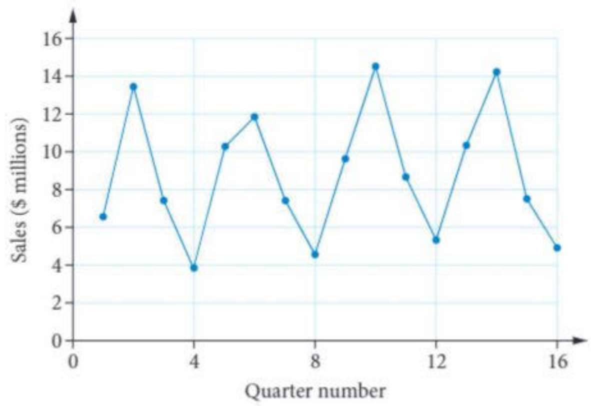
A 7.25 B 7.375 C 7.5

D 7.625 E 8

| Year | Number of accidents |
|---|---|
| 1993 | 13 |
| 1994 | 7 |
| 1995 | 3 |
| 1996 | 9 |
| 1997 | 10 |
| 1998 | 8 |
| 1999 | 7 |
| 2000 | 6 |
| 2001 | 10 |
| 2002 | 11 |

*Use the following information to answer the next two questions.*

The time series plot charts the quarterly sales figures, in millions of dollars, of a small manufacturing business over a period of four years.




12 © VCAA The time series plot is best described as having

A no variability.

B seasonality only.

C irregular variation only.

D a decreasing trend with seasonality.

E an increasing trend with seasonality.

13 ^^^5~2OI7N 1 CQI4 J The sales figures used to generate this time series plot are displayed in the table.

| Year | Quarter 1 | Quarter 2 | Quarter 3 | Quarter 4 |
|---|---|---|---|---|
| 2013 | 6.5 | 13.4 | 7.4 | 3.8 |
| 2014 | 10.2 | 11.8 | 7.4 | 4.5 |
| 2015 | 9.6 | 14.5 | 8.6 | 5.3 |
| 2016 | 10.3 | 14.2 | 7.5 | 4.9 |

The four-mean smoothed sales with centring for Quarter 3 in 2015, in millions of dollars, was closest to

A 8.6 B 9.3 C 9.5 D 9.6 E 9.7

▶ 14 (4 marks) The table shows the number of rescues on a stretch of beach each year over an 8-year period.

| Year | Number of rescues |
|---|---|
| 1 | 11 |
| 2 | 7 |
| 3 | 5 |
| 4 | 15 |
| 5 | 10 |
| 6 | 20 |
| 7 | 17 |
| 8 | 22 |

Find the following, rounding your answers to one decimal place if necessary,

a The two-mean smoothed number of rescues for Year 3. 1 mark

b The three-mean smoothed number of rescues for Year 3. 1 mark

c The four-mean smoothed number of rescues for Year 3. 1 mark

d How many smoothed values would there be if a six-point moving mean smoothing was used? 1 mark

# @ Graphical smoothing

Video playlist Graphical smoothing

Worksheet Moving medians

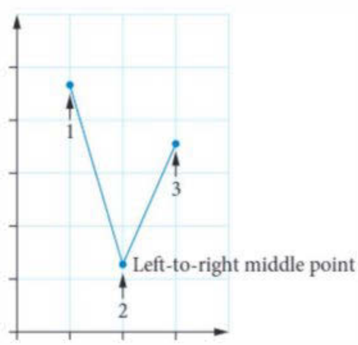
## Smoothing using moving medians

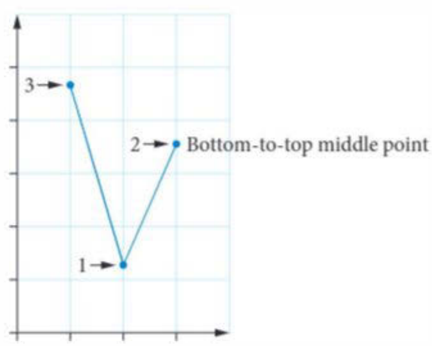
Graphical smoothing is an alternative to numerical smoothing and is based on working directly from a time series plot. The most common graphical smoothing method is moving medians, which involves finding a series of medians of a fixed number of data points. Although it is possible to use an even number of data points with this method, we will only be using an odd number.

The advantages of moving median smoothing over moving mean smoothing are

- moving median smoothing can be done directly from the plot without any calculations
- extreme data values are eliminated more quickly.

To smooth using moving medians, find both the left-to-right middle point and the bottom-to-top middle point:

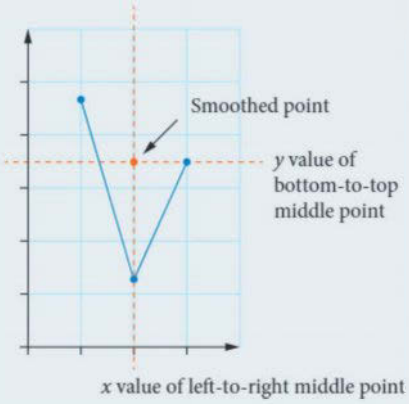


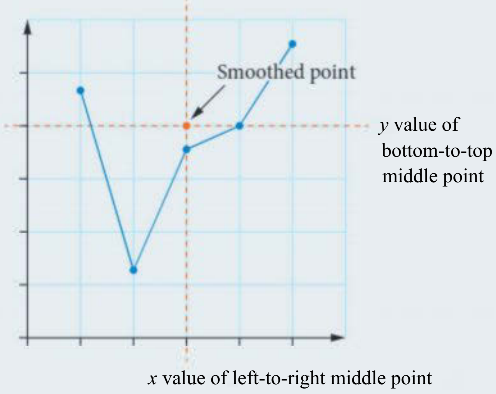
## Graphical smoothing using moving medians

The left-to-right middle point gives the $x$ coordinate of the smoothed point and the bottom-to-top middle point gives the $y$ coordinate of the smoothed point.

Three-point moving medians




Five-point moving medians




WB

p. 48

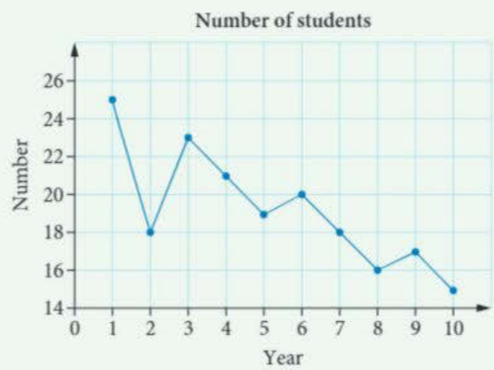
### WORKED EXAMPLE 5 Graphical smoothing using moving medians

For this time series plot showing the number of students in a Year 12 General Mathematics class over the last 10 years, smooth the data using the

**a** three-point moving median method

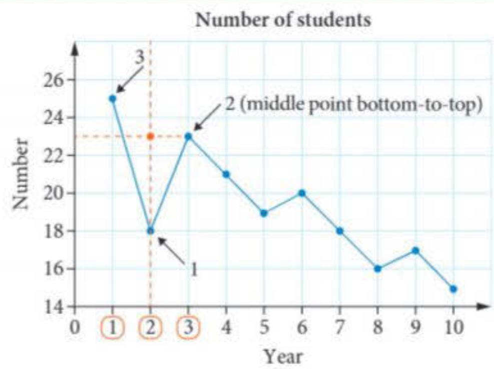
**b** five-point moving median method.

**c** What do the graphs of the smoothed data indicate about the trend in the original data?

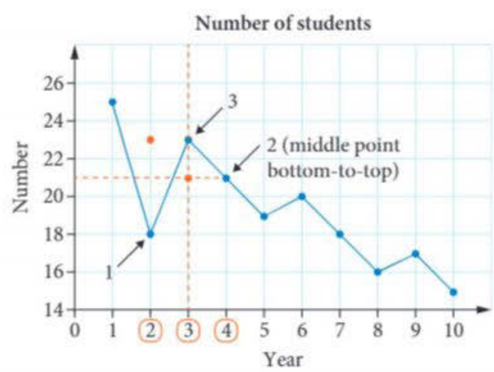



**Steps** | **Working**

**a** **1** Look at the first three points and find the middle point both left-to-right and bottom-to-top, giving you the coordinates of the first smoothed point.



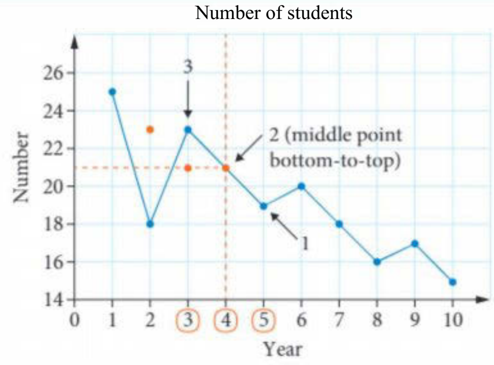

**2** Look at the next three points and repeat.





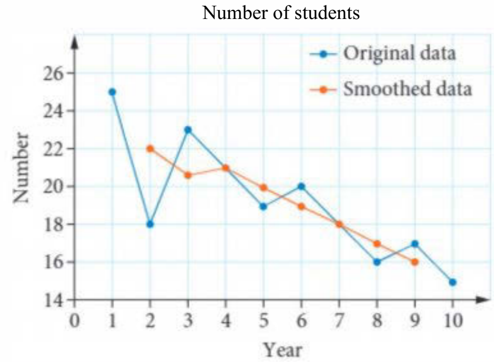
9780170448406

3 Repeat for the next three points. Sometimes the smoothed point is the same as the original point.

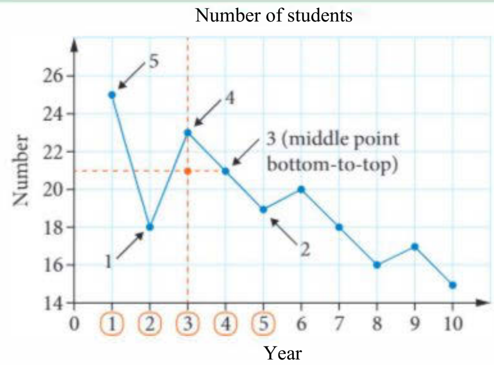



4 Repeat the process until you reach the last median and then join the points.

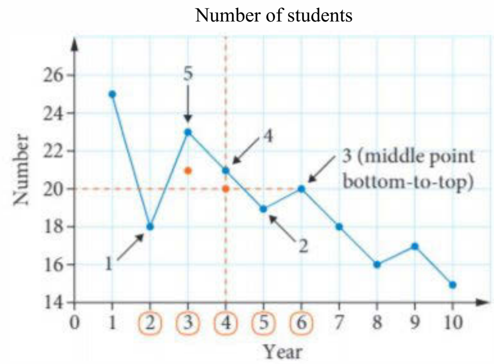
Three-point median smoothing




b 1 Look at the first five points and find the middle point both left-to-right and bottom-to-top, giving you the coordinates of the first smoothed point.



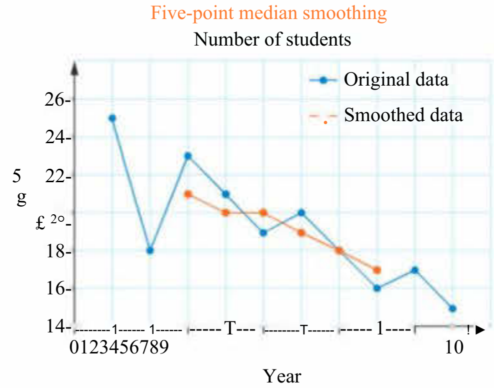

2 Look at the next five points and repeat.




3 Repeat the process until you reach the last median and then join the points.

As with three-point moving mean smoothing, the first and last data points are lost with three-point moving median smoothing. Similarly, the first two and last two data points are lost with five-point moving median smoothing and so on.

Five-point median smoothing




c Is there an increasing or decreasing trend?

The graphs of the smoothed data indicate a decreasing trend.

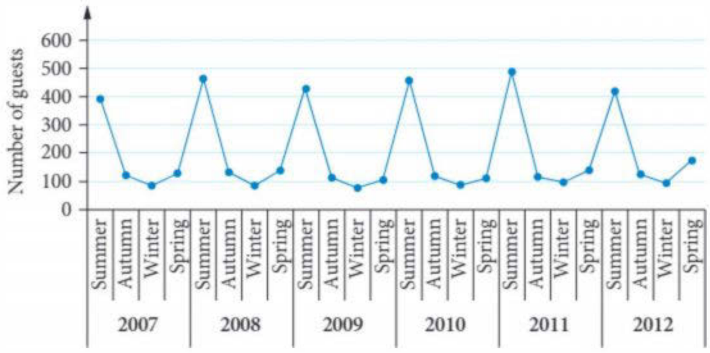
## EXERCISE 4.3 Graphical smoothing

ANSWERS p. 706

### Recap

*Use the following information to answer the next two questions.*

The time series plot displays the number of guests staying at a holiday resort during summer, autumn, winter and spring for the years 2007 to 2012 inclusive.




1 ©VCAA 2013 1CQ12 80% Which one of the following best describes the pattern in the time series?

A random variation only

B decreasing trend with seasonality

C seasonality only

D increasing trend only

E increasing trend with seasonality

2 ©VCAA 20131CQ13 J 51% The table shows the data from the time series plot for the years 2007 and 2008.

| Year | Season | Number of guests |
|---|---|---|
| 2007 | summer | 390 |
| | autumn | 126 |
| | winter | 85 |
| | spring | 130 |
| 2008 | summer | 460 |
| | autumn | 136 |
| | winter | 86 |
| | spring | 142 |

Using four-mean smoothing with centring, the smoothed number of guests for winter 2007 is closest to

A 85 B 107 C 183 D 192 E 200

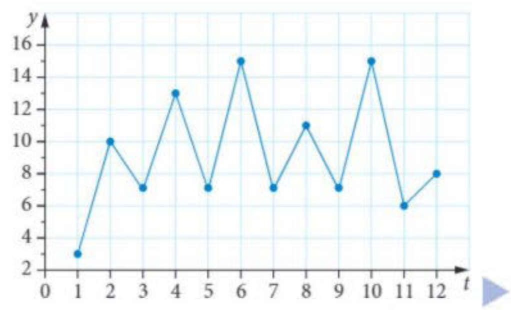
### Mastery

3 a WORKED EXAMPLE 5 For this time series plot, smooth the data using the

a three-point moving median method

b five-point moving median method.

c What do the graphs of the smoothed data indicate about the trend in the original data?





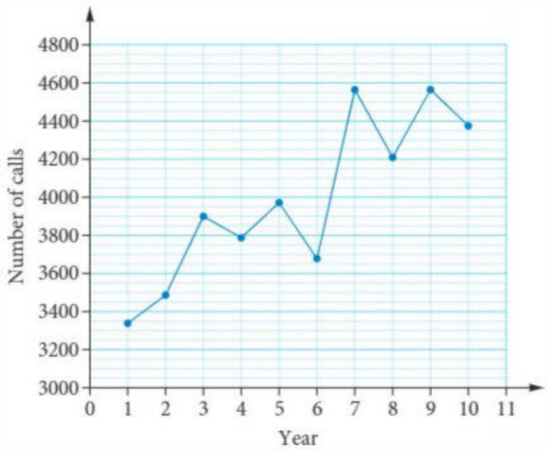

9780170448406

## Exam practice

80-100% 60-79% 0-59%

4.3

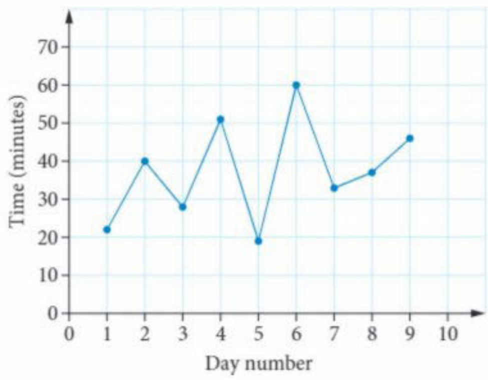
4 ©VCAA 20151CQ12J 63% The time series plot charts the number of calls per year to a computer help centre over a 10-year period.




Using five-median smoothing, the smoothed number of calls in Year 6 was closest to

A 3500 B 3700 C 3800 D 4000 E 4200

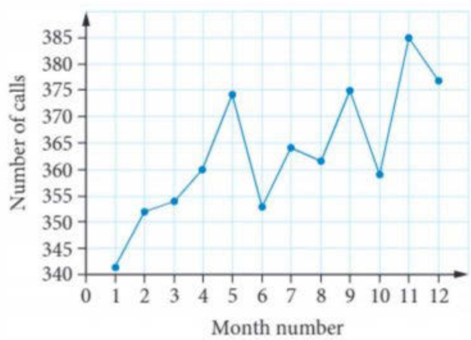
5 ©VCAA 20191CQ13] 45% The *time,* in minutes, that Liv ran each day was recorded for nine days. The time series plot was generated from this data.




Both three-median smoothing and five-median smoothing are being considered for this data. Both of these methods result in the same smoothed value on *day number*

A3 B4 C 5 D6 E7

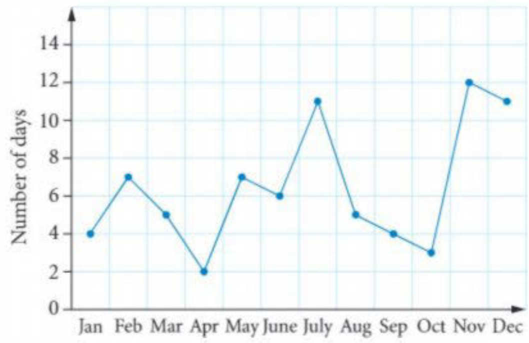
6 ©VCAA 20101CQ12 I 43% The time series plot shows the number of calls each month to a call centre over a twelve-month period. The plot is to be smoothed using five-point median smoothing.




The smoothed number of calls for month number 10 is closest to

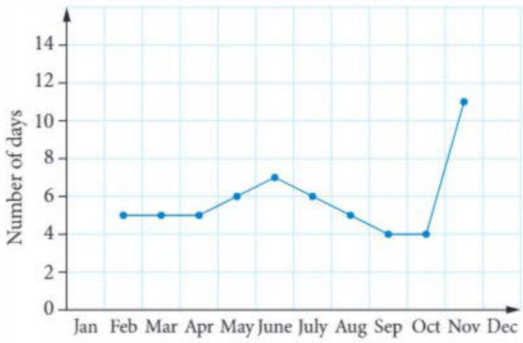
A 358 B 364 C 371 D 375 E 377

▶ 7 ©VCAA 20121009 42% The time series plot shows the number of days that it rained in a town each month during 2011.

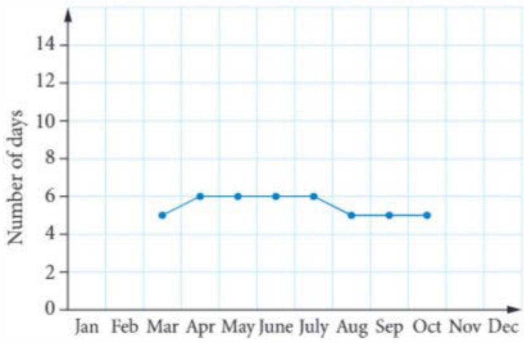



Using five-median smoothing, the smoothed time series plot will look most like

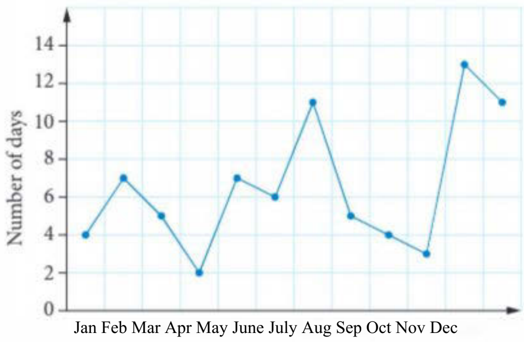
**A**



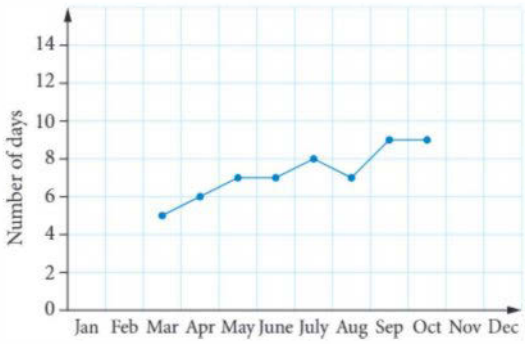
**B**



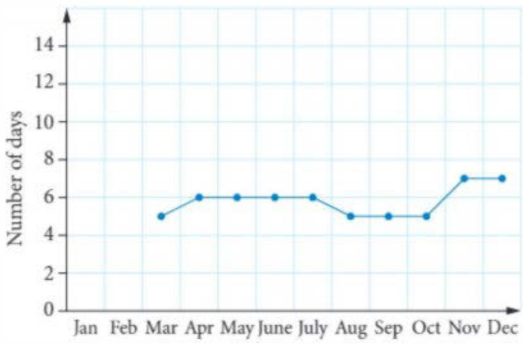
**C**



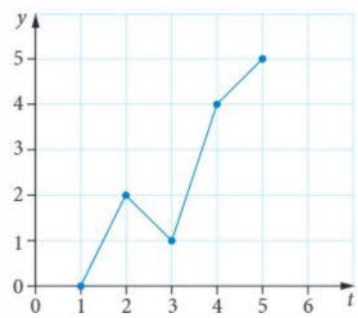
**D**



**E**



8 ©VCAA 2005ICQ10 30% The five points shown on the grid have been taken from a time series plot that is to be smoothed using median smoothing. The coordinates of the median of these five points are




A (3,1) B (3,2) C (3,2.4)

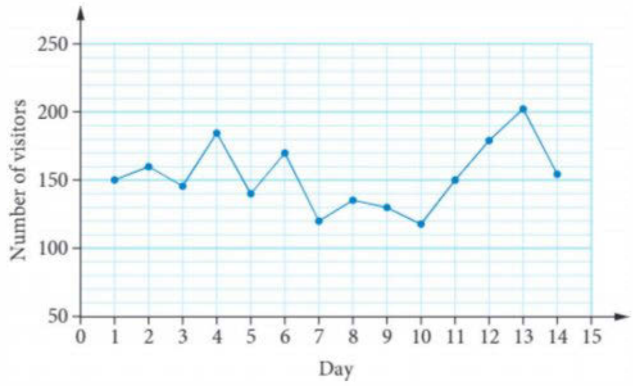
D (3,2.5) E (3,3)

▶

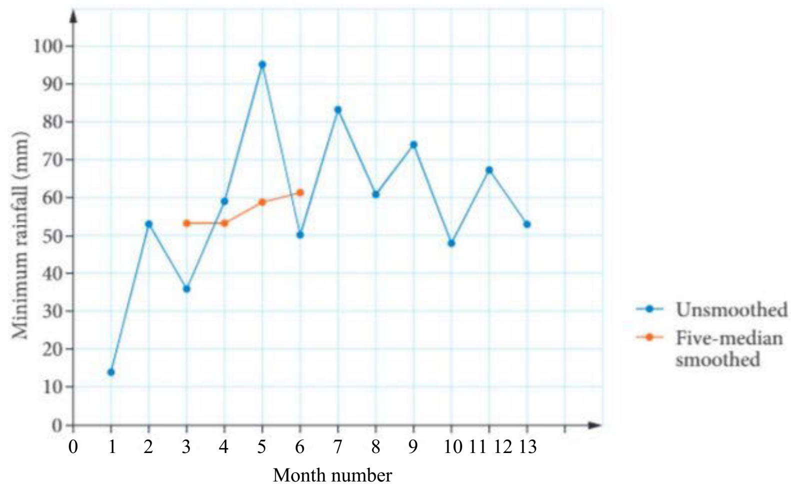
▶ 9 ©VCAA 2019N 1CQ14 The time series plot shows the daily number of visitors to a historical site over a two-week period.

This time series plot is to be smoothed using seven-median smoothing. The smoothed number of visitors on day 4 is closest to

A 120　　B 140　　C 145　　D 150　　E 160




10 ©VCAA [ 2016 2CQ4 J (4 marks) The time series plot shows the *minimum rainfall* recorded at a weather station each month plotted against the *month number* (1 = January, 2 = February, and so on). Rainfall is recorded in millimetres. The data was collected over a period of one year.

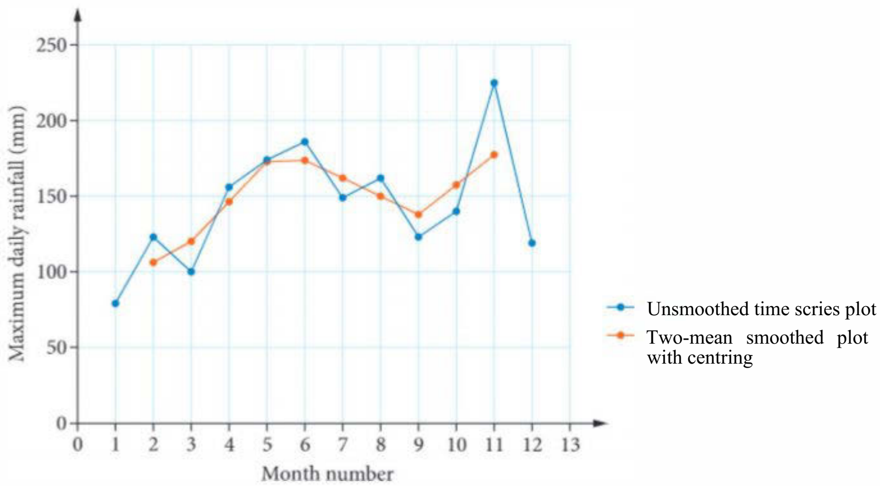



a 49% Five-median smoothing has been used to smooth the time series plot. The first four smoothed points are shown. Copy and complete the five-median smoothing by marking smoothed values on the time series plot. 2 marks ▶

The maximum daily rainfall each month was also recorded at the weather station. The table shows the *maximum daily rainfall* each month for a period of one year.

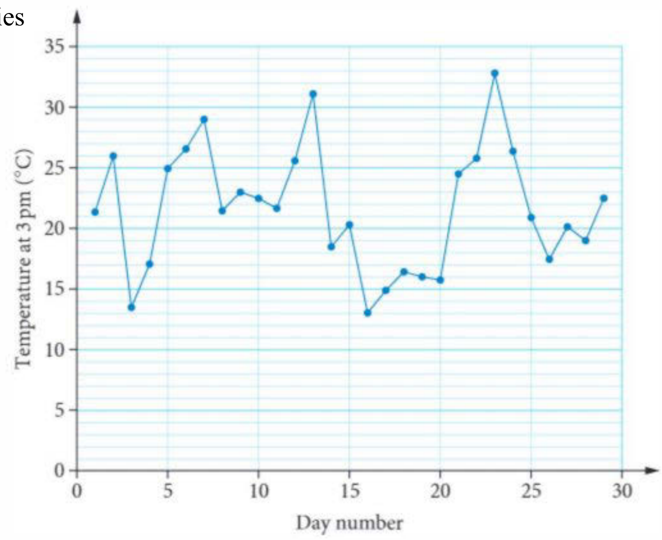
| Month | Jan | Feb | Mar | Apr | May | June | July | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Month number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Maximum daily rainfall (mm) | 79 | 123 | 100 | 156 | 174 | 186 | 149 | 162 | 124 | 140 | 225 | 119 |

The data in the table has been used to plot *maximum daily rainfall* against *month number* in the time series plot shown.




b 50% Two-mean smoothing with centring has been used to smooth this time series plot. The smoothed values are marked. Using the data given in the table, show that the two-mean smoothed rainfall centred on October is 157.25 mm. 2 marks

11 ©VCAA 2017N 2CQ4 , (4 marks) The time series plot shows the *temperature* (°C) recorded at a weather station at 3 pm for the 29 days of February in a particular leap year.

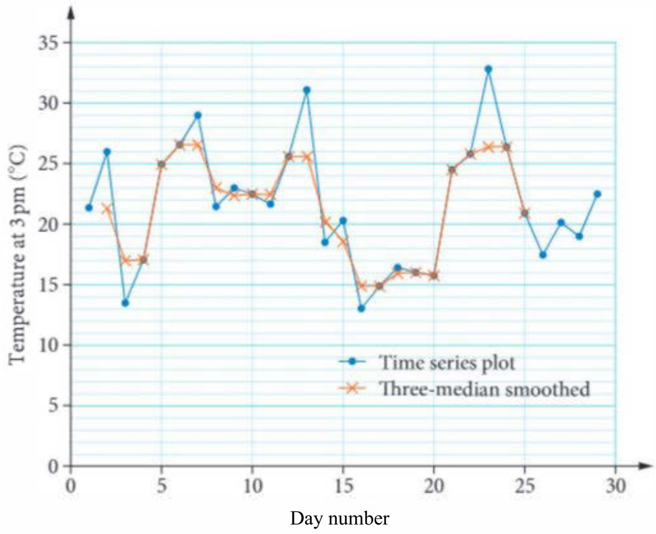



a Write down the range for the variable *temperature.* Round your answer to the nearest whole number. 1 mark

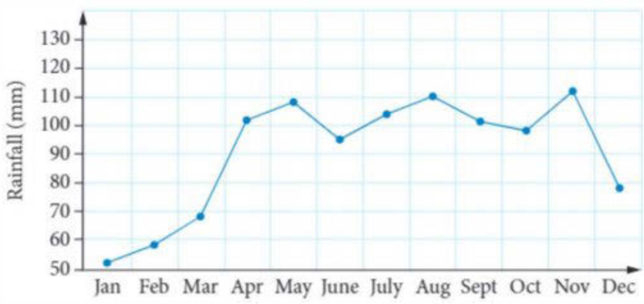
b Determine the five-median smoothed *temperature* at 3 pm on day 14. Round your answer to the nearest whole number. 1 mark


9780170448406

c Three-median smoothing has now been used to smooth the time series plot up to day 25. Copy the time series plot given below and complete the three-median smoothing by marking each remaining smoothed point with a cross (x). Draw only the section of the time series plot from day 25 to 29. 2 marks




12 ©VCAA 2009 2CQ2 (4 marks) The time series plot shows the rainfall (in mm) for each month during 2008.




a Which month had the highest rainfall? 1 mark

b Copy the plot and use three-median smoothing to smooth the time series. Draw the smoothed time series on the copied plot. Mark each smoothed data point with a cross (x). 2 marks

c Describe the general pattern in rainfall that is revealed by the smoothed time series plot. 1 mark

## Exam hack

When a question asks for general' patterns or trend, refer to what the data is showing, but don't give specific data values in your answer.

Video playlist Seasonal adjustment

Worksheet Seasonal adjustment

# @ Seasonal adjustment

## Interpreting seasonal indices

As we have already seen, some data has regular and predictable changes that repeat during a year or less, which we describe as seasonality. Remember a season can be a day, a week, a month, a quarter, or an actual season.

It's not effective to smooth seasonal data using the moving mean or moving median methods because smoothing two consecutive points might connect two data points with significant seasonal differences. You could end up smoothing out important information rather than an irregular fluctuation. To deal effectively with this sort of data, we need to make seasonal adjustments. Seasonal indices are used to make seasonal adjustments.

### Seasonal indices

- Seasonal indices compare each season to an average season.
- A season that is exactly average has a seasonal index of 1.
- An above average season has a seasonal index greater than 1.
- A below average season has a seasonal index less than 1.
- The mean of seasonal indices is 1.
- To interpret seasonal indices, convert them to percentages.

WB p. 50

**WORKED EXAMPLE 6** Interpreting seasonal indices

The following table shows the seasonal indices for the number of pizzas sold on each day of the week by a pizza chain:

| Mon | Tue | Wed | Thu | Fri | Sat | Sun |
|---|---|---|---|---|---|---|
| 0.3 | 1.05 | 0.4 | 0.85 | 1.4 | 1.8 | 1.2 |

| Steps | Working |
|---|---|
| **a** How many days had below average sales? | |
| How many indices are less than 1? | 3 days have below average sales |
| **b** Which day had the closest to average sales? | |
| Which of the indices is closest to 1? | Tuesday |
| **c** Which day had sales furthest away from the average? | |
| Which of the indices is furthest from 1? | Saturday |
| **d** What do the seasonal indices add to? | |
| Add the indices. | $0.3 + 1.05 + 0.4 + 0.85 + 1.4 + 1.8 + 1.2 = 7$ |
| **e** What is the mean of the seasonal indices? | |
| Divide the sum of the indices by the number of indices. | $\frac{7}{7} = 1$ |
| **f** Rewrite the table so that the seasonal indices are converted to percentages. | |
| Convert each decimal to a percentage. | see table below |
| **g** What is the mean of the percentaged seasonal indices? | |
| Add the percentages and divide by 7. | $\frac{30 + 105 + 40 + 85 + 140 + 180 + 120}{7} = \frac{700}{7} = 100\%$ |
| **h** What percentage below average were Wednesday's sales? | |
| Compare the percentage to 100%. | Wednesday's pizza sales were 60% *below* average. |

| Mon | Tue | Wed | Thu | Fri | Sat | Sun |
|---|---|---|---|---|---|---|
| 30% | 105% | 40% | 85% | 140% | 180% | 120% |



9780170448406

The seasonal indices in the table in Worked example 6 add up to 7. This is because the season we are looking at is the number of days in a week. If the seasonal indices were for months, they would add up to 12, and if they were for quarters, they would add up to 4.

4.4

### The sum of seasonal indices

The sum of the seasonal indices = the number of seasons

| Type of data | No. of seasons | Cycle | Sum of seasonal indices |
|---|---|---|---|
| Daily figures for data from Monday to Sunday | 7 | full week | 7 |
| Daily figures for data from Monday to Friday | 5 | working week | 5 |
| Monthly figures | 12 | year | 12 |
| Quarterly accounts | 4 | year | 4 |

## Calculating seasonal indices

The seasonal indices can be calculated from one seasons data using the formula:

$$\text{seasonal index} = \frac{\text{actual figure}}{\text{average for the season}}$$

### WORKED EXAMPLE 7 EEHEGS2ESH

WB p. 51

The quarterly sales figures for the number of cars sold were recorded by a car sales yard for 2023.

| Quarter 1 | Quarter 2 | Quarter 3 | Quarter 4 |
|---|---|---|---|
| 5 | 7 | 9 | 3 |

a Find the seasonal indices for each of the quarters, correct to two decimal places,

b Rewrite the table so that the seasonal indices are converted to percentages,

c Which of the quarters' sales were 50% below average?

| Steps | Working |
|---|---|
| a 1 Find the average for the season. | $\frac{5+7+9+S}{4} = \frac{24}{4} = 6$ |
| 2 For each of the actual figures, use the formula: seasonal index $= \frac{\text{actual figure}}{\text{average for the season}}$ rounding your answer to two decimal places. | Quarter 1: $\frac{5}{6} = 0.83$; Quarter 2: $\frac{7}{6} = 1.17$; Quarter 3: $\frac{9}{6} = 1.50$; Quarter 4: $\frac{3}{6} = 0.50$ |
| b Convert each decimal to a percentage. | Quarter 1: 83%; Quarter 2: 117%; Quarter 3: 150%; Quarter 4: 50% |
| c Read from the percentage table. | Quarter 4 |

## De-seasonalising time series data

The most common seasonal adjustment is to de-seasonalise data. To do this we use seasonal indices to remove the seasonal component of the time series. De-seasonalisation is a form of smoothing, which takes out the seasonal effects of the data so that a line of best fit can be fitted and long-term trends can be predicted.

Worksheets
De-seasonal-isation

Seasonality

## De-seasonalising time series data

$$\text{de-seasonalised figure} = \frac{\text{actual figure}}{\text{seasonal index}}$$

This version of the formula appears on the examination formula sheet with seasonal index as the subject:

$$\text{seasonal index} = \frac{\text{actual figure}}{\text{de-seasonalised figure}}$$

Realistically, a seasonal index would be calculated from several years' data. The following worked example shows how to calculate the seasonal index from a table of actual figures across several years.

WB p. 52

**WORKED EXAMPLE 8** Dealing with seasonalised time series data

The quarterly sales figures for the number of cars sold were recorded by a car sales yard for 2023 to 2025.

| Year | Q1 | Q2 | Q3 | Q4 |
|---|---|---|---|---|
| 2023 | 5 | 7 | 9 | 3 |
| 2024 | 4 | 8 | 9 | 4 |
| 2025 | 5 | 9 | 10 | 5 |

**a** Calculate the seasonal index for each quarter, correct to four decimal places.

**b** Use the seasonal indices to de-seasonalise the data, correct to two decimal places.

**c** Plot the original and the de-seasonalised data on the same set of axes.

**Steps** | **Working**

**a 1** Calculate the yearly mean. Calculate the totals for each year, and find the mean for each year by dividing each total by 4.

| Year | Q1 | Q2 | Q3 | Q4 | Yearly mean |
|---|---|---|---|---|---|
| 2023 | 5 | 7 | 9 | 3 | $\frac{5+7+9+3}{4} = 6$ |
| 2024 | 4 | 8 | 9 | 4 | $\frac{4+8+9+4}{4} = 6.25$ |
| 2025 | 5 | 9 | 10 | 5 | $\frac{5+9+10+5}{4} = 7.25$ |

**2** Calculate the quarterly proportions. Divide each quarterly sales figure by the corresponding yearly mean to obtain quarterly proportions. Give answers correct to four decimal places.

| Year | Q1 | Q2 | Q3 | Q4 |
|---|---|---|---|---|
| 2023 | $\frac{5}{6} = 0.8333$ | $\frac{7}{6} = 1.1667$ | $\frac{9}{6} = 1.5000$ | $\frac{3}{6} = 0.5000$ |
| 2024 | $\frac{4}{6.25} = 0.6400$ | $\frac{8}{6.25} = 1.2800$ | $\frac{9}{6.25} = 1.4400$ | $\frac{4}{6.25} = 0.6400$ |
| 2025 | $\frac{5}{7.25} = 0.6897$ | $\frac{9}{7.25} = 1.2414$ | $\frac{10}{7.25} = 1.3793$ | $\frac{5}{7.25} = 0.6897$ |


9780170448406

| | |
|---|---|
| 3 Calculate the seasonal indices by finding the mean of the quarterly proportions. Give answers correct to four decimal places. | see table below |

| Year | Q1 | Q2 | Q3 | Q4 |
|---|---|---|---|---|
| 2023 | 0.8333 | 1.1667 | 1.5000 | 0.5000 |
| 2024 | 0.6400 | 1.2800 | 1.4400 | 0.6400 |
| 2025 | 0.6897 | 1.2414 | 1.3793 | 0.6897 |
| Total | 2.1630 | 3.6881 | 4.3193 | 1.8297 |
| [illegible] | $\frac{2.1630}{3}$ = 0.7210 | $\frac{3.6881}{3}$ = 1.2294 | $\frac{4.3193}{3}$ = 1.4398 | $\frac{1.8297}{3}$ = 0.6099 |

4 Add the four seasonal index values to check the sum is 4. Note that the values may not add exactly to 4 because of rounding errors.

0.7210 + 1.2294 + 1.4398 + 0.6099 = 4.0001

5 Write a summary table for the seasonal indices.

| Year | Qi | Q2 | Q3 | Q4 |
|---|---|---|---|---|
| Seasonal index | 0.7210 | 1.2294 | 1.4398 | 0.6099 |

b 1 De-seasonalise the original time series data using the formula:

de-seasonalised figure

$$= \frac{\text{actual figure}}{\text{seasonal index}}$$

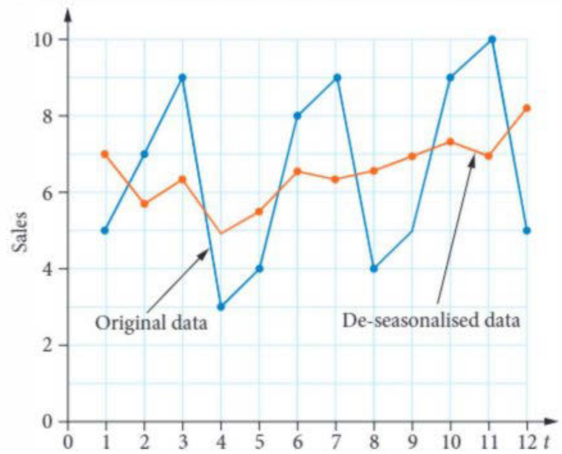
| Year | Q1 | Q2 | Q3 | Q4 |
|---|---|---|---|---|
| 2023 | $\frac{5}{0.7210}$ = 6.94 | $\frac{7}{1.2294}$ = 5.69 | $\frac{9}{1.4398}$ = 6.25 | $\frac{3}{0.6099}$ = 4.92 |
| 2024 | $\frac{4}{0.7210}$ = 5.55 | $\frac{8}{1.2294}$ = 6.51 | $\frac{9}{1.4398}$ = 6.25 | $\frac{4}{0.6099}$ = 6.56 |
| 2025 | $\frac{5}{0.7210}$ = 6.94 | $\frac{9}{1.2294}$ = 7.32 | $\frac{10}{1.4398}$ = 6.95 | $\frac{5}{0.6099}$ = 8.20 |

2 Write a summary table for the de-seasonalised values, correct to two decimal places.

| Year | Q1 | Q2 | Q3 | Q4 |
|---|---|---|---|---|
| 2023 | 6.94 | 5.69 | 6.25 | 4.92 |
| 2024 | 5.55 | 6.51 | 6.25 | 6.56 |
| 2025 | 6.94 | 7.32 | 6.95 | 8.20 |

c Plot the original and the de-seasonalised time series data on the same set of axes.

Use $t = 1$ to represent the first quarter of 2023, $t = 2$ to represent the second quarter of 2023, etc.



## Re-seasonalising time series data

Sometimes we are asked to re-seasonalise data. Re-seasonalisation involves finding the actual figure given the de-seasonalised figure and seasonal index. To do this we rearrange the seasonal index formula so that the actual figure is the subject.

**Re-seasonalising time series data**

actual figure = de-seasonalised figure x seasonal index

WB p. 54

**WORKED EXAMPLE 9** De-seasonalising and re-seasonalising time series data

The following table shows the quarterly seasonal indices for revenue to a publishing company from the sales of mathematics textbooks.

| Quarter | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Seasonal index | 0.7 | | 0.6 | 1.9 |

a What is the missing Quarter 2 seasonal index?

b To correct for seasonality, by what percentage should the sales for Quarter 2 be increased?

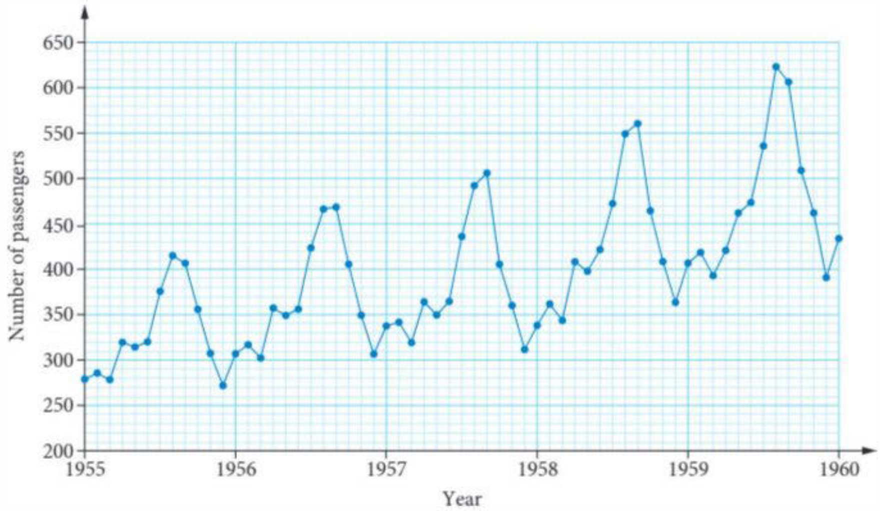
c The company predicts that its de-seasonalised quarterly sales will be $1000 000 for each quarter. Based on this, what would you predict the actual sales for Quarter 2 to be?

| Steps | Working |
|---|---|
| a Use the fact that the seasonal indices need to add to 4 for quarterly data. | 4 - 0.7 - 0.6 - 1.9 = 0.8<br>The Quarter 2 seasonal index is 0.8. |
| b Use the de-seasonalising formula:<br>$\text{de-seasonalised figure} = \frac{\text{actual figure}}{\text{seasonal index}}$ | $\text{de-seasonalised figure} = \frac{\text{actual figure}}{0.8}$<br>$= \frac{1}{0.8} \times \text{actual figure}$<br>= 1.25 x actual figure<br>= 125% x actual figure<br>So to correct for seasonality, the sales for Quarter 2 should be increased by 25%. |
| c Use the re-seasonalising formula:<br>actual figure<br>= de-seasonalised figure x seasonal index | actual figure = $1000000 x 0.8<br>=$800000<br>So the actual sales for Quarter 2 are $800 000. |

### Recap

*Use the following information to answer the next two questions.*

The time series plot shows the number of passengers who flew with an airline each month over the period 1955-1960.




Data: GEP Box and GM Jenkins, *Time Series Analysis, Forecasting and Control*, revised edition, Holden-Day, 1976, p. 531

1 ©VCAA 201 SN icon J The pattern shown in the time series plot above is best described as having

A no trend.

B irregular fluctuations only.

C an increasing trend with seasonality only.

D an increasing trend with irregular fluctuations only.

E an increasing trend with seasonality and irregular fluctuations.

2 ©VCAA 2018N 1CQ12 The dot above 1957 corresponds to the number of passengers in January 1957.

The five-median smoothed number of passengers for August 1957 is closest to

A 400 B 435 C 480 D 495 E 510

### Mastery

3H WORKED EXAMPLE 6 I The following table shows the seasonal indices for the number of ice creams sold each month by an ice cream franchise:

| Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2.5 | 2.1 | 1.6 | 0.7 | 0.2 | 0.2 | 0.1 | 0.0 | 0.6 | 0.9 | 1.2 | 1.9 |

a How many months had below average sales?

b Which month had the closest to average sales?

c Which month had sales furthest away from the average?

d What do the seasonal indices add to?

e What is the mean of the seasonal indices?

f Rewrite the table so that the seasonal indices are converted to percentages,

g What is the mean of the percentaged seasonal indices?

h What percentage above average were January’s sales?

4 & WORKED EXAMPLE 7 The quarterly sales figures for the number of king-sized beds sold were recorded by a furniture shop for 2023.

| Q1 | Q2 | Q3 | Q4 |
|---|---|---|---|
| 11 | 8 | 1 | 3 |

a Find the seasonal indices for each of the quarters, correct to two decimal places,

b Rewrite the table so that the seasonal indices are converted to percentages,

c Which of the quarters* sales were 39% above average?

5B WORKED EXAMPLE 8 The quarterly sales figures for the number of king-sized beds sold were recorded by a furniture shop for 2023 to 2025.

| Year | Q1 | Q2 | Q3 | Q4 |
|---|---|---|---|---|
| 2023 | 11 | 8 | 1 | 3 |
| 2024 | 9 | 9 | 3 | 1 |
| 2025 | 4 | 9 | 1 | 7 |

**a** Calculate the seasonal index for each quarter, correct to three decimal places.

**b** Use the seasonal indices to de-seasonalise the data.

c Plot the original and the de-seasonalised data on the same set of axes.

6 S WORKED EXAMPLE 9 The following table shows the quarterly seasonal indices for revenue to a company from the sales of a brand of soft drink.

| Quarter | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Seasonal index | 1.4 | 0.8 | | 1.1 |

a What is the missing Quarter 3 seasonal index?

b To correct for seasonality, by what percentage should the sales for Quarter 3 be increased? Round your answer to the nearest percentage.

c The company predicts that its de-seasonalised quarterly sales will be $100 000 for each quarter. Based on this, what would you predict the actual sales for Quarter 3 to be?

## Exam practice

a*_100* 6o_79% o_59%

7 © VCAA 2020 1CQ17 88% The table shows the monthly rainfall for 2019, in millimetres, recorded at a weather station, and the associated long-term seasonal indices for each month of the year.

| | Jan | Feb | Mar | Apr | May | June | July | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Monthly rainfall (mm) | 18.4 | 17.6 | 46.8 | 23.6 | 92.6 | 77.2 | 80.0 | 86.8 | 93.8 | 55.2 | 97.3 | 69.4 |
| Seasonal index | 0.728 | 0.734 | 0.741 | 0.934 | 1.222 | 0.973 | 1.024 | 1.121 | 1.159 | 1.156 | 1.138 | 1.072 |

Data: adapted from © Commonwealth of Australia 2020, Bureau of Meteorology, www.bom.gov.au

The de-seasonalised rainfall for May 2019 is closest to

A 71.3 mm B 75.8 mm C 86.1mm D 88.1 mm E 113.0 mm


9780170448406

*Use the following information to answer the next three questions.*

The table shows the long-term average of the number of meals served each day at a restaurant. Also shown is the daily seasonal index for Monday through to Friday.

| | Day of the week | | | | | | |
|---|---|---|---|---|---|---|---|
| | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday | Sunday |
| Long-term average | 89 | 93 | 110 | 132 | 145 | 190 | 160 |
| Seasonal index | 0.68 | 0.71 | 0.84 | 1.01 | 1.10 | | |

8 ©VCAA J 2016 1CQ14 79% The seasonal index for Wednesday is 0.84. This tells us that, on average, the number of meals served on a Wednesday is

A 16% less than the daily average.

B 84% less than the daily average.

C the same as the daily average.

D 16% more than the daily average.

E 84% more than the daily average.

9 ©VCAA | 2016 1CQ15 J 72% Last Tuesday, 108 meals were served in the restaurant. The de-seasonalised number of meals served last Tuesday was closest to

A 93 B 100 C 110 D 131 E 152

10 ©VCAA | 2016 1CQ16 J 54% The seasonal index for Saturday is closest to

A 1.22 B 1.31 C 1.38 D 1.45 E 1.49

11 ©VCAA 2007 1CQ13 53% The revenue from sales (in dollars) each month made by a Queensland souvenir shop for the first year is shown in the table.

| Month | Revenue ($) |
|---|---|
| January | 1236 |
| February | 1567 |
| March | 1240 |
| April | 2178 |
| May | 2308 |
| June | 2512 |
| July | 3510 |
| August | 4234 |
| September | 4597 |
| October | 4478 |
| November | 7034 |
| December | 8978 |

If this information is used to determine the seasonal index for each month, the seasonal index for September will be closest to

A 0.80 B 0.82 C 1.16 D 1.22 E 1.26

▶ 12 ©VCAA 20181CQ16 J 51% The quarterly sales figures for a large suburban garden centre, in millions of dollars, for 2016 and 2017 are displayed in the table.

| Year | Quarter 1 | Quarter 2 | Quarter 3 | Quarter 4 |
|---|---|---|---|---|
| 2016 | 1.73 | 2.87 | 3.34 | 1.23 |
| 2017 | 1.03 | 2.45 | 2.05 | 0.78 |

Using these sales figures, the seasonal index for Quarter 3 is closest to

A 1.28 B 1.30 C 1.38 D 1.46 E 1.48

*Use the following information to answer the next two questions.*

The table shows the long-term mean rainfall, in millimetres, recorded at a weather station, and the associated long-term seasonal indices for each month of the year. The long-term mean rainfall for December is missing.

| | Jan | Feb | Mar | Apr | May | June | July | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mean rainfall (mm) | 51.9 | 52.3 | 52.8 | 66.6 | 87.1 | 69.4 | 73.0 | 79.9 | 82.6 | 82.4 | 81.1 | |
| Seasonal index | 0.728 | 0.734 | 0.741 | 0.934 | 1.222 | 0.973 | 1.024 | 1.121 | 1.159 | 1.156 | 1.138 | 1.072 |

Data: adapted from © Commonwealth of Australia 2020, Bureau of Meteorology, www.bom.gov.au

13 ©VCAA 2020 1CQ15 43% To correct the rainfall in March for seasonality, the actual rainfall should be, to the nearest per cent

A decreased by 26% B increased by 26% C decreased by 35%

D increased by 35% E increased by 74%

14 ©VCAA 2020 1CQ16 67% The long-term mean rainfall for December is closest to

A 64.7 mm B 65.1mm C 71.3 mm D 76.4 mm E 82.0 mm

15 ©VCAA 32% The seasonal index for the sales of cold drinks in a shop in January is 1.6. To correct the January sales of cold drinks for seasonality, the actual sales should be

A reduced by 37.5% B reduced by 40% C reduced by 62.5%

D increased by 60% E increased by 62.5%

16 ©VCAA 2009 2CQ4 (3 marks) This table shows the seasonal indices for rainfall in summer, autumn and winter.

a Calculate the seasonal index for spring. 1 mark

Seasonal indices

| Summer | Autumn | Winter | Spring |
|---|---|---|---|
| 0.78 | 1.05 | 1.07 | |

b In 2008, a total of 188 mm of rain fell during summer.

Using the appropriate seasonal index in the table, determine the de-seasonalised figure for the summer rainfall in 2008. Write your answer correct to the nearest millimetre, 1 mark

c What does a seasonal index of 1.05 tell us about the rainfall in autumn? 1 mark ▶

> 17 **Ei J** 2019 2CQ6 ©VCAA (3 marks) The total rainfall, in millimetres, for each of the four seasons in 2015 and 2016 is shown in Table 1.

Table 1

| | Total rainfall (mm) | | | |
|---|---|---|---|---|
| Year | Summer | Autumn | Winter | Spring |
| 2015 | 142 | 156 | 222 | 120 |
| 2016 | 135 | 153 | 216 | 96 |

a 45% The seasonal index for winter is shown in Table 2. Use the values in Table 1 to find the seasonal indices for summer, autumn and spring. Copy and complete Table 2, rounding your answers two decimal places. 2 marks

Table 2

| | Summer | Autumn | Winter | Spring |
|---|---|---|---|---|
| Seasonal index | | | 1.41 | |

b 58% The total rainfall for each of the four seasons in 2017 is shown in Table 3.

Table 3

| | Total rainfall (mm) | | | |
|---|---|---|---|---|
| Year | Summer | Autumn | Winter | Spring |
| 2017 | 141 | 156 | 262 | 120 |

Use the appropriate seasonal index from Table 2 to de-seasonalise the total rainfall for winter in 2017. Round your answer to the nearest whole number. 1 mark

# 4.5 Least squares trend lines

Video playlist Least squares trend lines

As with other associations between two numerical variables we have looked at, we can use a least squares line of best fit (often called a trend line for time series) to model time series trends, as long as the data appears to be linear. If it isn't linear, then we would need to use transformation techniques from Chapter 3 to linearise the data first.

## Seasonality and forecasting

If there is seasonality in the time series, then we usually need to go through the extra step of de-seasonalising the data before fitting the least squares line. The least squares line based on the de-seasonalised data can be used to make predictions; however, the result will give a de-seasonalised figure. This value then needs to be re-seasonalised to give the actual figure using:

actual figure = de-seasonalised figure x seasonal index

When we use the least squares line to make predictions outside of the data range, the same issues of extrapolation that we have discussed in Chapter 3 apply. In the case of time series, this involves extending into the future, which is called trend line forecasting. We can never be certain that the equation of the line will apply in the future, and the further into the future we are trying to predict, the less reliable the equation of the least squares line will be.

WB p. 55

**WORKED EXAMPLE 10** Working with trend lines for de-seasonalised data

The following table lists the de-seasonalised number of sales of a particular joke coffee mug in a novelty store for each quarter in 2022-2023 and the seasonal indices.

| Quarter | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| De-seasonalised number of sales in 2022 | 5 | 10 | 12 | 28 |
| De-seasonalised number of sales in 2023 | 26 | 25 | 31 | 27 |
| Seasonal index | 1.8 | 1 | 0.5 | 0.7 |

a Find the equation of the least squares trend line for the de-seasonalised time series data for 2022-2023. Round the slope and intercept to three significant figures.

b Plot the time series and draw the trend line for the de-seasonalised data on the same axes. Comment on the trend by interpreting the slope of the trend line equation,

c Use the trend line equation to forecast the de-seasonalised number of sales for Quarter 3 2024.

d Use the trend line equation to forecast the actual number of sales for Quarter 3 2024.

**Steps** | **Working**

a 1 Rewrite the de-seasonalised number of sales in a table that represents the quarters from 1 to 8.

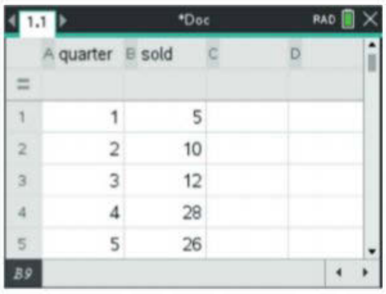
| Quarter number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| De-seasonalised number of sales 2022-2023 | 5 | 10 | 12 | 28 | 26 | 25 | 31 | 27 |

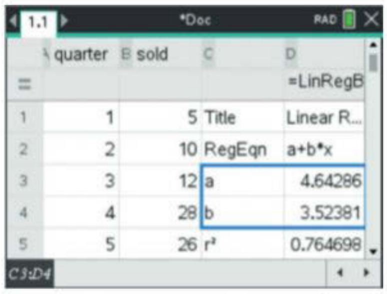
2 Use CAS to find $a$ and $b$ for the least squares line of best fit equation, rounding to the required significant figures.

$a = 4.642\ 86$ $\quad b = 3.523\ 81$

*de-seasonalised number of sales* $= 4.64 + 3.52$ x *quarter number*

**TI-Nspire**



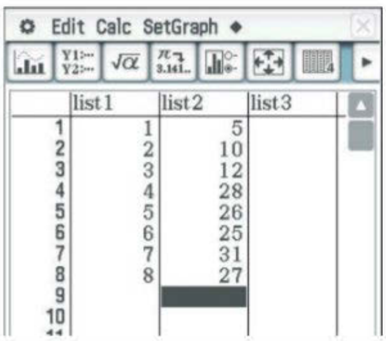

1 Open a Lists & Spreadsheet page.

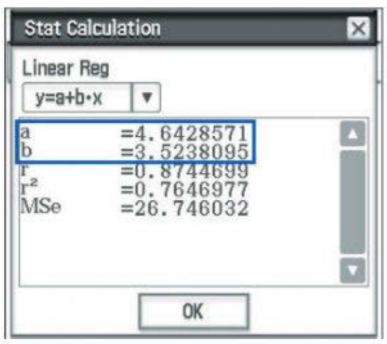
2 Enter the appropriate headings and the values in columns A and B.

3 Press menu > Statistics > Stat Calculations > Linear Regression (a+bx).

4 Select the headings then select OK.

**ClassPad**




1 Tap Menu > Statistics.

2 Enter the values into listl and Iist2.

3 Tap Calc > Regression > Linear Reg.

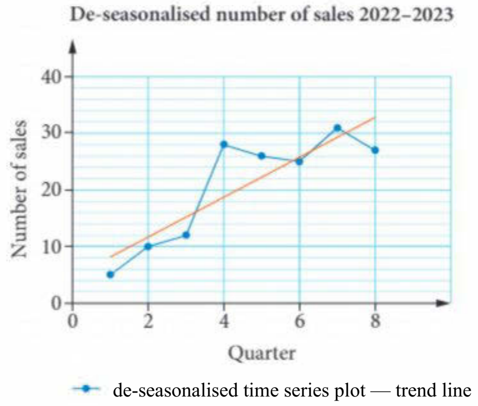
4 On the next screen, keep the default settings of XList: listl and YList: list 2 and tap OK.

5 Select y=a+bx from the dropdown menu.


9780170448406

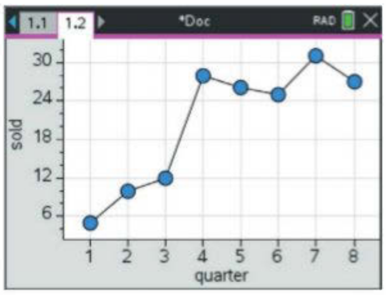
b Use CAS to plot the time series and draw the line of best fit.

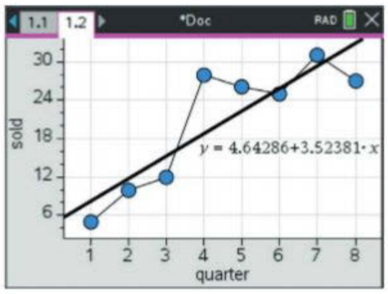
Refer to the slope from the trend line equation in your comment.




During 2022-2023 the sales of the mug increased on average by 3.52 per quarter.

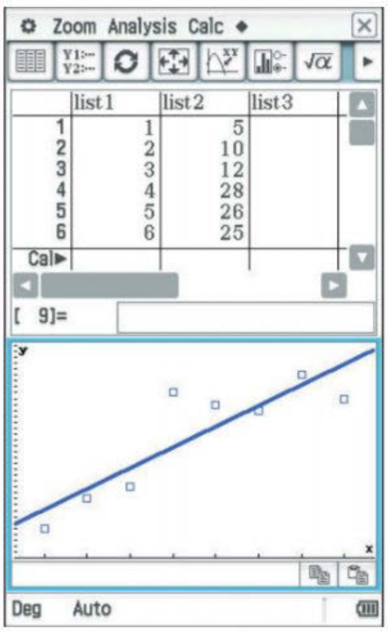
**TI-Nspire**

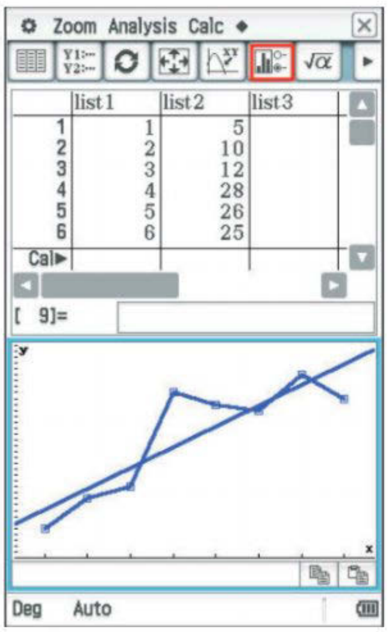


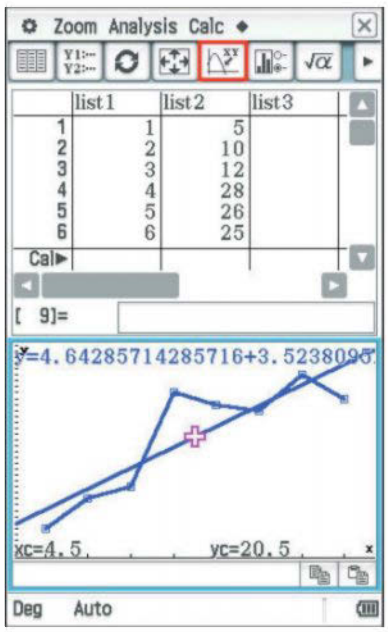




1 Add a Data & Statistics page.
2 Select the headings from the Lists & Spreadsheet columns.
3 Press menu > Plot Properties > Connect Data Points.
4 Press menu > Analyze > Regression > Show Linear (a+bx).
5 The time series and trend line will both be displayed.

**ClassPad**




1 Tap OK from the previous screen to display the scatterplot and least squares regression line in the lower window.
2 Tap the Set StatGraphs tool.
3 Change the Type: from Scatter to xyLine and tap Set.
4 The time series and trend line will both be displayed.
5 Tap the Equation tool to display the equation.

| | |
|---|---|
| c Find the number of the quarter and use the trend line equation to forecast the de-seasonalised number of sales.<br>Round the answer to whole mugs sold. | If Q4 2023 is quarter number 8 then<br>Q3 2024 is quarter number 8 + 3 = 11.<br>*de-seasonalised number of sales* = 4.64 + 3.52x11<br>= 43.36<br>De-seasonalised number forecast to be sold in Q3 2024 is 43 mugs. |
| d Re-seasonalise by using:<br>actual figure<br>= de-seasonalised figure x seasonal index<br>Use the unrounded de-seasonalised figure, then round the answer to whole mugs sold. | actual figure = 43.36 x 0.5<br>= 21.68<br>Actual number forecast to be sold in Q3 2024 is 22 mugs. |

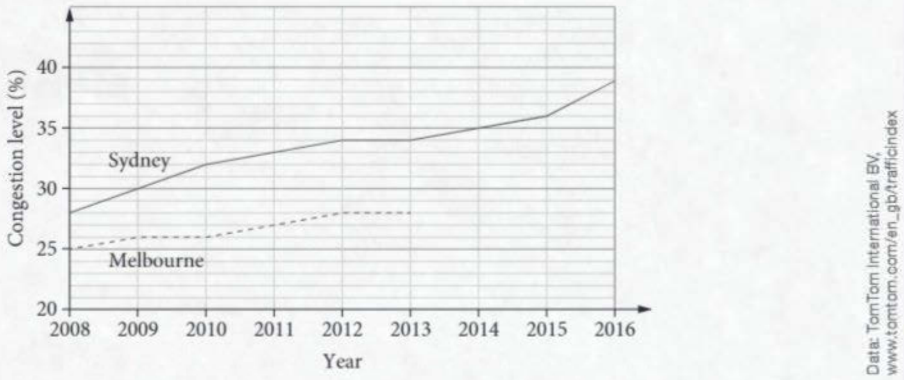
Video
VCE question analysis: Time series

## VCE QUESTION ANALYSIS

©VCAA 2018 2CQ3 2018 Examination 2 Core Question 3 (9 marks)

The table shows the yearly average traffic congestion levels in two cities, Melbourne and Sydney, during the period 2008 to 2016. Also shown is a time series plot of the same data. The time series plot for Melbourne is incomplete.

| | Congestion level (%) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Year | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 |
| Melbourne | 25 | 26 | 26 | 27 | 28 | 28 | 28 | 29 | 33 |
| Sydney | 28 | 30 | 32 | 33 | 34 | 34 | 35 | 36 | 39 |




a Copy the time series plot for Melbourne and use the data in the table to complete it. 1 mark

b A least squares line is used to model the trend in the time series plot for Sydney.

The equation is

*congestion level* = -2280 + 1.15 x *year*

i Draw this least squares line on the time series plot. 1 mark

ii Use the equation of the least squares line to determine the average rate of increase in percentage congestion level for the period 2008 to 2016 in Sydney. 1 mark

iii Use the least squares line to predict when the percentage congestion level in Sydney will be 43%. 1 mark

c When a least squares line is used to model the trend in the data for Melbourne, the intercept of this line is approximately -1514.755 56. Round this value to four significant figures. 1 mark



9780170448406

d Use the data in the table to determine the equation of the least squares line that can be used to model the trend in the data for Melbourne. The variable *year* is the explanatory variable. Copy the equation below and write the values of the intercept and the slope of this least squares line in the appropriate boxes. Round both the intercept and slope values to four significant figures. 2 marks

*congestion level* = ☐ + ☐ x *year*

e Since 2008, the equations of the least squares lines for Sydney and Melbourne have predicted that future traffic congestion levels in Sydney will always exceed future traffic congestion levels in Melbourne. Explain why, quoting the values of appropriate statistics. 2 marks

## Reading the question

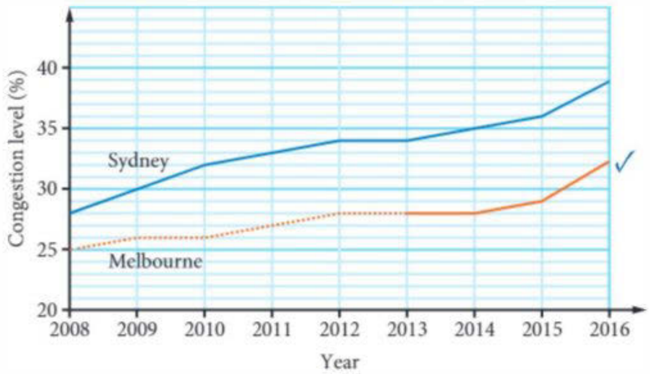
- Note that rounding to significant figures is asked for several times.
- The word 'use' tells you how to answer the question.
- There is a 2-mark 'Explain' question at the end that asks you to quote statistics.

## Thinking about the question

- You need to complete a time series plot and draw a least squares line.
- You need to find the average rate of increase from a least squares line.
- Make sure you are clear whether the question is referring to Melbourne or Sydney.

## Worked solution (/ = 1 mark)

a Use the table to plot the missing points.




b i Substitute into the line of best fit equation, using the two endpoints to draw the line of best fit.

For *year* = 2008:

*congestion level* = -2280 + 1.15 x *year*

= -2280+ 1.15x2008

= 29.2

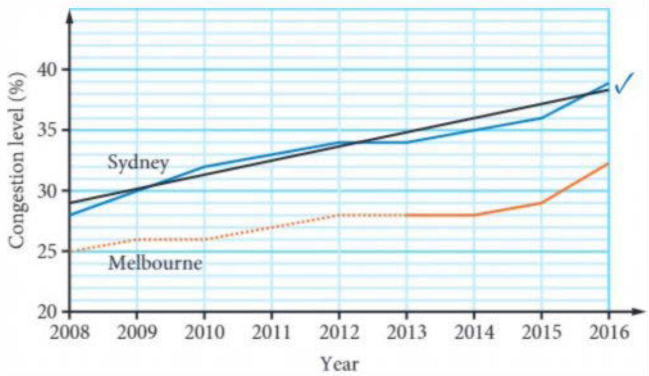
For *year* = 2016:

*congestion level* = -2280 + 1.15 x *year*

= -2280 + 1.15 x 2016

= 38.4

Draw the line joining (2008,29.2) and (2016,38.4).




**Exam hack**

Make sure you bring a ruler to the exam. Accuracy matters!

ii For the equation

*congestion level* = -2280 + 1.15 x *year*

the slope is 1.15. *Congestion level* is measured as a percentage and time is measured in years, so the average rate of increase in percentage congestion level is 1.15% per year. /

4.5

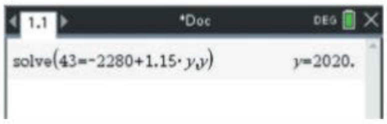
iii Substituting into *congestion level* = -2280 + 1.15 x *year*

43 = -2280 + 1.15 x *year*

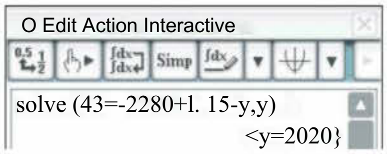
$$year = \frac{43 + 2280}{1.15} = 2020 \checkmark$$

or use the CAS solve function:

**TI-Nspire**




**ClassPad**

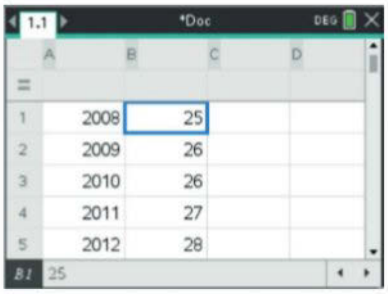


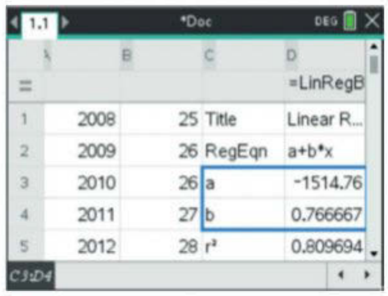

The congestion level in Sydney will be 43% in the year 2020.

c -1514.755 56 rounded to four significant figures is -1515. ✓

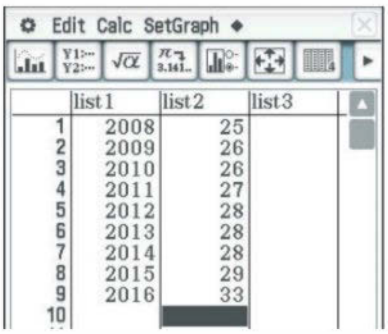
d Using CAS to find *a* and *b* for the least squares line:

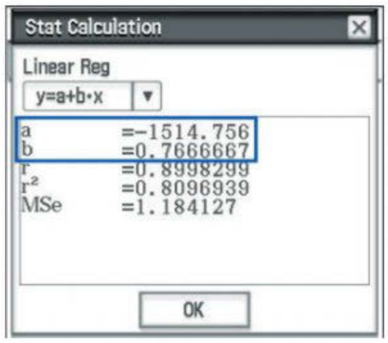
**TI-Nspire**




**ClassPad**




Rounding both the intercept *a* and the slope *b* to four significant figures gives the equation of the least squares line for Melbourne as *congestion level* = -1515 ✓ + 0.7667 ✓ x *year*

*e* The Sydney least squares line is: *congestion level* = -2280 + 1.15 x *year*

The Melbourne least squares line is: *congestion level* = -1515 + 0.7667 x *year*

Using these equations, in 2008:

Sydney's *congestion level* = 29.2% and Melbourne's *congestion level* = 24.5%.

Also, the slope of Sydney's *congestion level* is 1.15 and the slope of Melbourne's *congestion level* is 0.7667.

Since Sydney has the higher congestion level in 2008 (29.2 > 24.5) ✓ and the slope of Sydney's line is also greater than Melbourne's (1.15 > 0.7667) ✓, the least squares lines predict that Sydney's future traffic congestion levels will *always* exceed Melbourne's.



9780170448406

**Student performance** 80-100% 60-79% 0-59%

a 88%

b i 51% A significant number of students could not draw a least squares line. Successful students used a ruler and often wrote down the coordinates of the endpoints (2008,29.2) and (2016,38.4). A common wrong endpoint was (2008,28).

ii 36% Most students did not realise that the slope of the line measured the average rate of increase. Some seemed to get confused with the $R$ value in financial mathematics and wrote 15%.

iii 74%

c 66% Common errors were -1514, -1514.7556 and -1515.0000.

d 68% Some students did not recognise the link with part c and wrote a different number as the intercept. Some who correctly rounded in part c used -1514.7556 in part d. A method mark was awarded if the students' answer from part c was given as the intercept (even if incorrect),

e 18% Some students mentioned medians, which were irrelevant. Many focused on the starting values for Sydney and Melbourne from the table rather than the graphs. Some gave slope figures for both cities but did not say that Sydneys' slope was greater. Again, some confused the $R$ value with financial mathematics and described Melbourne's gradient of 0.7667 as a decrease of about 23%.

4.5

## EXERCISE 4.5 Least squares trend lines

ANSWERS p. 707

### Recap

*Use the following information to answer the next two questions.*

The seasonal indices for the first 11 months of the year for sales in a sporting equipment store are shown in the table.

| Month | Jan | Feb | Mar | Apr | May | June | July | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Seasonal index | 1.23 | 0.96 | 1.12 | 1.08 | 0.89 | 0.98 | 0.86 | 0.76 | 0.76 | 0.95 | 1.12 | |

1 ©VCAA 2016S1CQ14 The seasonal index for December is

A 0.89 B 0.97 C 1.02 D 1.23 E 1.29

2 ©VCAA 2016S1CQ15 In May, the store sold $213 956 worth of sporting equipment. The de-seasonalised figure of these sales was closest to

A $165857 B $190420 C $209677 D $218322 E $240400

### Mastery

3 WORKED EXAMPLE 10 The following table lists the de-seasonalised number of sales of a particular costume in a fancy-dress store for each quarter in 2022-2023 and the seasonal indices.

| Quarter | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| De-seasonalised number of sales in 2022 | 33 | 33 | 25 | 28 |
| De-seasonalised number of sales in 2023 | 22 | 21 | 11 | 7 |
| Seasonal index | 0.9 | 0.6 | 1.3 | 1.2 |

a Find the equation of the least squares trend line for the de-seasonalised time series data for 2022-2023. Round the slope and intercept to two significant figures.

b Plot the time series and draw the trend line for the de-seasonalised data on the same axes. Comment on the trend by interpreting the slope of the trend line equation.

c Use the trend line equation to forecast the de-seasonalised number of sales for Quarter 1 2024.

d Use the trend line equation to forecast the actual number of sales for Quarter 1 2024.

4 The following de-seasonalised data represents the monthly sales, in dollars, of a market stall over a period of 2 years.

| | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2021 sales ($) | 185 | 286 | 199 | 177 | 178 | 256 | 211 | 172 | 181 | 180 | 177 | 287 |
| 2022 sales ($) | 194 | 288 | 198 | 192 | 197 | 295 | 200 | 195 | 183 | 191 | 212 | 195 |
| Seasonal index | 1.81 | 0.70 | 0.77 | 0.73 | 0.86 | 0.89 | 0.76 | 1.13 | 1.07 | 0.97 | 1.22 | 1.09 |

a Find the equation of the least squares trend line for the de-seasonalised time series data for 2021-2022. Round the slope and intercept to three significant figures.

b Use the trend line equation to forecast the de-seasonalised sales for May of 2023, correct to the nearest dollar.

c Use the trend line equation to forecast the actual sales for May of 2023, correct to the nearest dollar.

## Exam practice

8(M(K,% 6o_79% o_59%

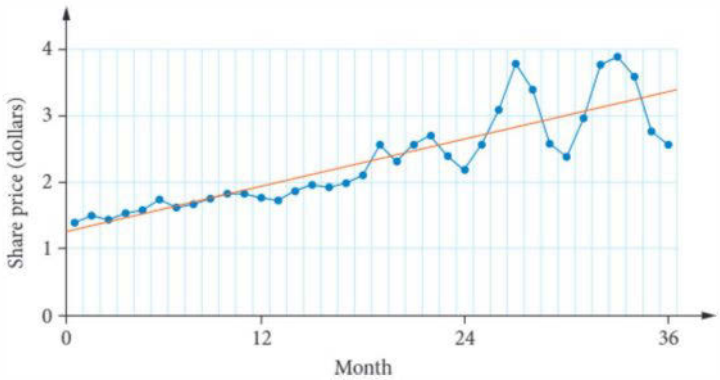
*Use the following information to answer the next two questions.*

The month-by-month price of a share listed on the Australian Stock Exchange is shown in the time series plot for a 36-month period. Also shown is a least squares line of best fit that has been fitted to the data. The equation of the least squares line of best fit is

*share price* = 1.24 + 0.06 x *month*

5 ©VCAA 20051CQ12J 90% The least squares line of best fit predicts that the price of the share after 48 months will be

A $1.96 B $4.12 C $5.04 D $28.80 E $62.40




6 ©VCAA 2005 1CQ13 70% Which one of the following statements best describes the time series plot for the period shown?

A The share price shows no trend and no change in variability.

B Theshare price shows no trend and increases in variability.

C Theshare price shows an increasing linear trend with constant variability.

D Theshare price shows an increasing linear trend with decreasing variability.

E Theshare price shows an increasing linear trend with increasing variability.

▶ 7 © VCAA 2019 1CQ14 69% The *time,* in minutes, that Liv ran each day was recorded for nine days. These times are shown in the table.

| Day number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Time (minutes) | 22 | 40 | 28 | 51 | 19 | 60 | 33 | 37 | 46 |

A least squares line is to be fitted to this time series. The equation of this least squares line, with *day number* as the explanatory variable, is closest to

A *day number* = 23.8 + 2.29 x *time*

B *day number* = 28.5 + 1.77 x *time*

C *time* = 23.8 + 1.77 x *day number*

D *time* = 23.8 + 2.29 x *day number*

E *time* = 28.5 + 1.77 x *day number*

*Use the following information to answer the next three questions.*

| Month | Jan | Feb | Mar | April | May | June | July | Aug | Sept | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Seasonal index | 1.30 | 1.21 | 1.00 | 0.95 | 0.95 | 0.86 | 0.86 | 0.89 | 0.94 | | 0.99 | 1.07 |

The table shows the seasonal indices for the monthly unemployment numbers for workers in a regional town.

8 © VCAA 2006 1CQ11 61% The seasonal index for October is missing from the table. The value of the missing seasonal index for October is

A 0.93 B 0.95 C 0.96 D 0.98 E 1.03

9 © VCAA 2006 1CQ12 66% The actual number of unemployed in the regional town in September is 330. The de-seasonalised number of unemployed in September is closest to

A 310 B 344 C 351 D 371 E 640

10 © VCAA 2006 1CQ13 29% A trend line that can be used to forecast the de-seasonalised number of unemployed workers in the regional town for the first nine months of the year is given by

*de-seasonalised number of unemployed* = *3733* - 3.38 x *month number*

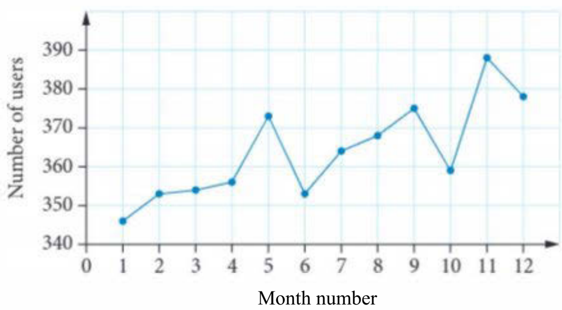
where month 1 is January, month 2 is February, and so on.

The actual number of unemployed for June is predicted to be closest to

A 304 B 353 C 376 D 393 E 410

*Use the following information to answer the next three questions.*

The time series plot shows the number of users each month of an online help service over a twelve-month period.




11 © VCAA 2008 1CQ11 39% The time series plot has

A no trend.

B no variability.

C seasonality only.

D an increasing trend with seasonality.

E an increasing trend only.

12 ©VCAA 20081CQ12 56% The data values used to construct the time series plot are given below.

| Month number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of users | 346 | 353 | 354 | 356 | 373 | 353 | 364 | 368 | 375 | 359 | 388 | 378 |

A four-point moving mean with centring is used to smooth the time series. The smoothed value of the number of users in month number 5 is closest to

A 357 B 359 C 360 D 365 E 373

13 ©VCAA 20081CQ13 63% A least squares line of best fit is fitted to the time series plot.

The equation of this least squares line of best fit is

*number of users* = 346 + 2.77 x *month number*

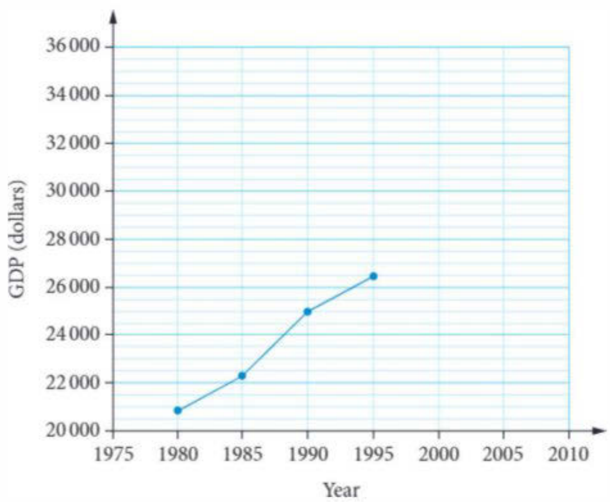
Let month number 1 = January 2007, month number 2 = February 2007, and so on.

Using the above information, the line of best fit predicts that the number of users in December 2009 will be closest to

A 379 B 412 C 443 D 446 E 448

14 ©VCAA 2010 2CQ3 57% (6 marks) The table shows the Australian gross domestic product (GDP) per person, in dollars, at five yearly intervals for the period 1980 to 2005.

| Year | 1980 | 1985 | 1990 | 1995 | 2000 | 2005 |
|---|---|---|---|---|---|---|
| GDP | 20900 | 22300 | 25000 | 26400 | 30900 | 33800 |




**@1 Exam hack**

For questions involving graphs, always check carefully what the intermediate intervals on the vertical scale are.

a Copy and complete the time series plot above by plotting the GDP for the years 2000 and 2005. 1 mark

b Briefly describe the general trend in the data. 1 mark

In the table below, the variable *year* has been rescaled using 1980 = 0,1985 = 5 and so on. The new variable is *time.*

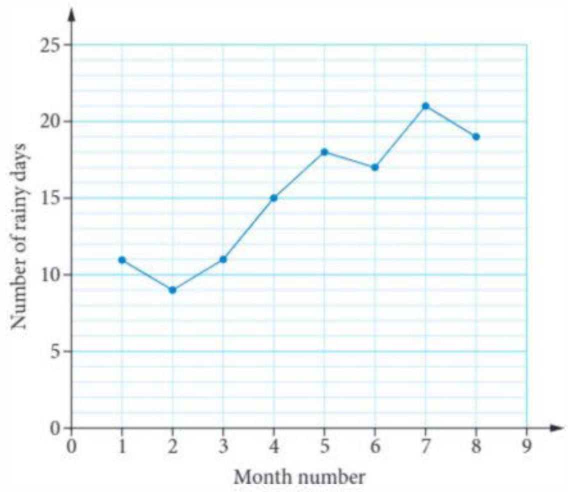
| Year | 1980 | 1985 | 1990 | 1995 | 2000 | 2005 |
|---|---|---|---|---|---|---|
| Time | 0 | 5 | 10 | 15 | 20 | 25 |
| GDP | 20900 | 22 300 | 25000 | 26400 | 30900 | 33800 |

c Use the variables *time* and *GDP* to write down the equation of the least squares line of best fit that can be used to predict *GDP* from *time.* Take *time* as the explanatory variable. 2 marks

d In the year 2007, the *GDP* was $34900. Find the error in the prediction if the least squares line of best fit calculated in part c is used to predict *GDP* in 2007. 2 marks


9780170448406

▶ 15 ©VCAA | 2Q17N2CQ3 (5 marks) The number of rainy days per month is recorded at a weather station. In the time series plot below, the *number of rainy days* per month is plotted for January (Month 1) to August (Month 8) in the same year.




a Describe the trend in the time series plot. 1 mark

The trend in the time series plot is to be modelled using a least squares line. The data used to construct this plot is given in the table.

| Month number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Number of rainy days | 11 | 9 | 11 | 15 | 18 | 17 | 21 | 19 |

b Use the data in the table to determine the equation of the least squares line. Round each of the numbers to three significant figures. 3 marks

c Copy the time series plot above and draw the least squares line on it. 1 mark

4.5

# (7) Chapter summary

## Time series plots

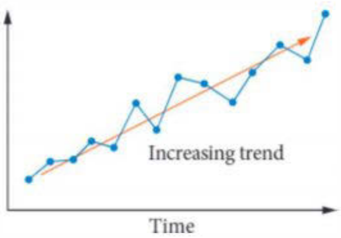
A time series involves data where the explanatory variable is time measured at equally spaced intervals such as hours, days, weeks, months, seasons and years.

A time series plot is a scatterplot of a time series where the data points are joined by straight lines.

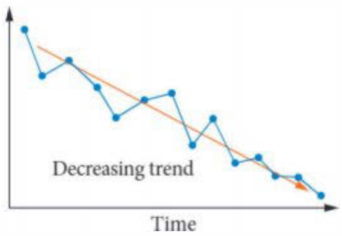
## Time series features

### Trend

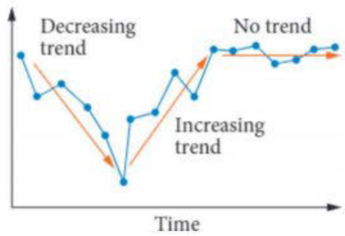
Increasing trend



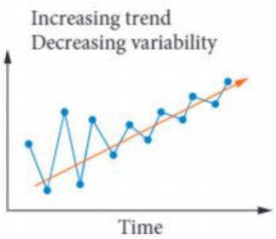

Decreasing trend

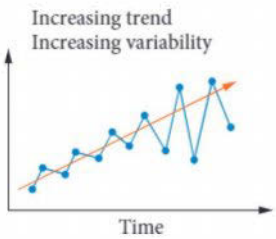


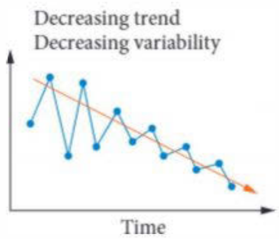

Multiple trends

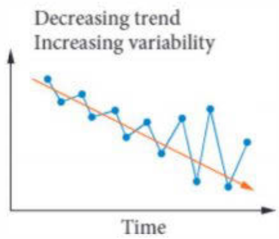



### Trend and variability

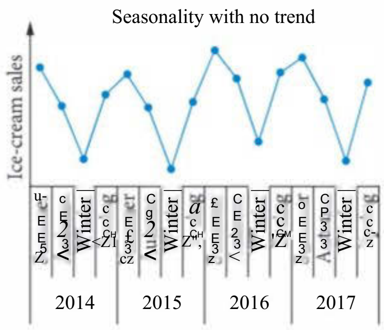


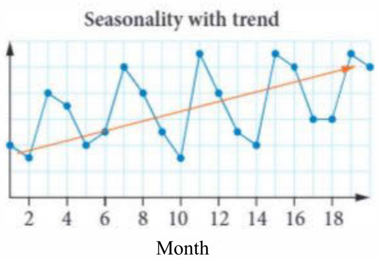



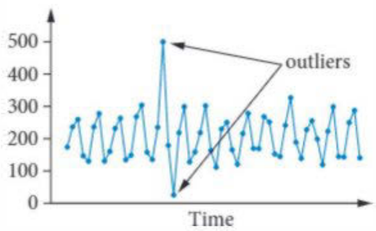

### Seasonality



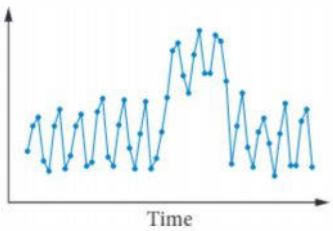



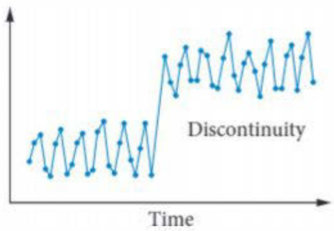

### Irregular fluctuations


### Outliers




### Structural change





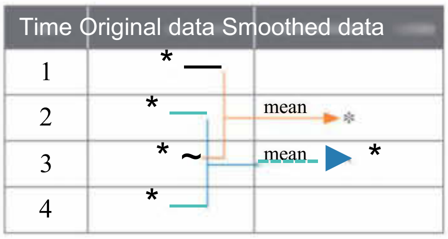

9780170448406

## Smoothing

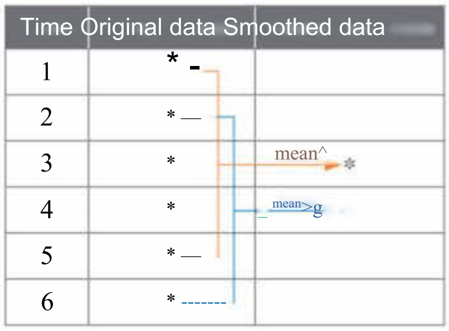
- Smoothing levels out fluctuations in time series to produce a smoother graph so we can see the underlying trends more clearly.
- Moving means involves finding a series of means of the data points. An odd number of data points involves one-step smoothing while an even number of data points involves a second step called centring.

## Moving mean smoothing with an odd number of points

Three-point moving means

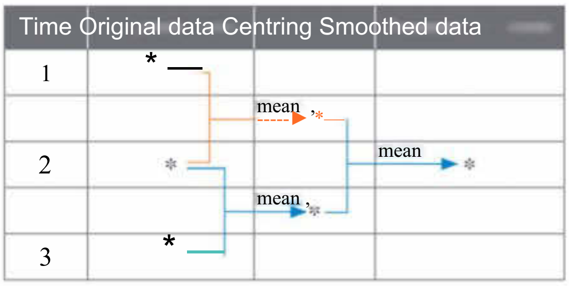



Five-point moving means

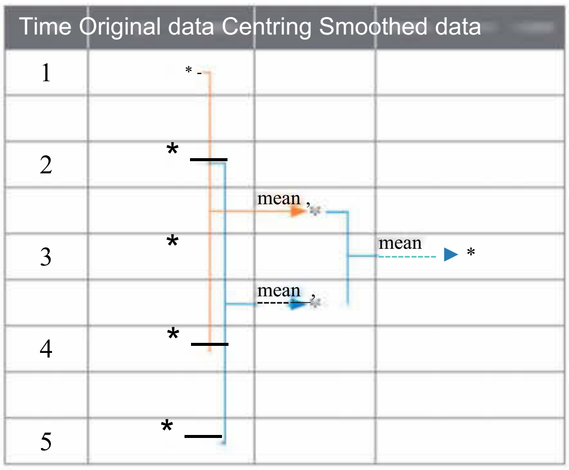



## Moving mean smoothing with an even number of points

Two-point moving means with centring

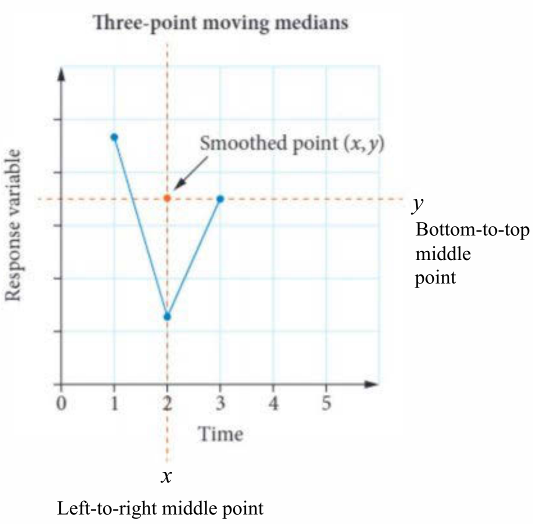


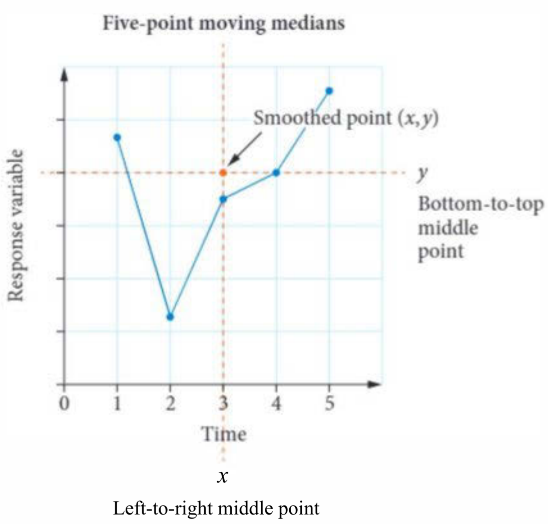

Four-point moving means with centring




- Moving medians involves finding a series of medians of the data points. We only need to consider using an odd number of data points for this method.
- The left-to-right middle point gives the $x$ coordinate of the smoothed point and the bottom-to-top middle point gives the $y$ coordinate of the smoothed point.

## Moving median smoothing



The advantages of moving median smoothing over moving means smoothing are

- moving median smoothing can be done directly from the plot without any calculations
- extreme data values are eliminated more quickly.

The number of moving points for both moving means and moving medians is usually chosen by looking at the natural cycle of the data being considered.

## Seasonal adjustment

- Seasonal indices are used to make seasonal adjustments.
- Seasonal indices compare each season to an average season.
- A season that is exactly average has a seasonal index of 1.
- An above average season has a seasonal index greater than 1.
- A below average season has a seasonal index less than 1.
- The mean of seasonal indices is 1.
- To interpret seasonal indices, convert them to percentages.
- The sum of the seasonal indices = the number of seasons.

| Type of data | No. of seasons | Cycle | Sum of seasonal indices |
|---|---|---|---|
| Daily figures for data from Monday to Sunday | 7 | full week | 7 |
| Daily figures for data from Monday to Friday | 5 | working week | 5 |
| Monthly figures | 12 | year | 12 |
| Quarterly accounts | 4 | year | 4 |

- Use the following versions of the same formula depending on what you are asked to find,

  the seasonal index:

  $$\text{seasonal index} = \frac{\text{actual figure}}{\text{de-seasonalised figure}}$$

  de-seasonalise the data:

  $$\text{de-seasonalised figure} = \frac{\text{actual figure}}{\text{seasonal index}}$$

  re-seasonalise the data:

  actual figure = de-seasonalised figure x seasonal index

## Fitting a least squares line

- We can use a least squares line of best fit to model time series trend as long as the data appears to be linear, but if there is seasonality then we usually need to de-seasonalise the data first before fitting the least squares line.
- The least squares line can be used to forecast, but the result will give a de-seasonalised figure. This value needs to be re-seasonalised to give the actual figure.

# Cumulative examination 1



Total number of marks: 11 Reading time: 5 minutes Writing time: 25 minutes

*Use the following information to answer the next two questions.*

The *blood pressure* (low, normal, high) and the *age* (under 50 years, 50 years or over) of 110 adults were recorded. The results are displayed in the two-way frequency table.

| Blood pressure | Age | |
|---|---|---|
| | Under 50 years | 50 years or over |
| low | 15 | 5 |
| normal | 32 | 24 |
| high | 11 | 23 |
| Total | 58 | 52 |

1 ©VCAA FrnKri The percentage of adults under 50 years of age who have high blood pressure is closest to

A 11% B 19% C 26% D 44% E 58%

2 ©VCAA 2016 1CQ2 The variables *blood pressure* (low, normal, high) and *age* (under 50 years, 50 years or over) are

A both nominal variables.

B both ordinal variables.

C a nominal variable and an ordinal variable respectively.

D an ordinal variable and a nominal variable respectively.

E a continuous variable and an ordinal variable respectively.

3 ©VCAA 2018 1CQ6 Data was collected to investigate the association between the following two variables:

- *age* (29 and under, 30-59,60 and over)
- *uses public transport* (yes, no)

Which one of the following is appropriate to use in the statistical analysis of this association?

A a scatterplot

B parallel box plots

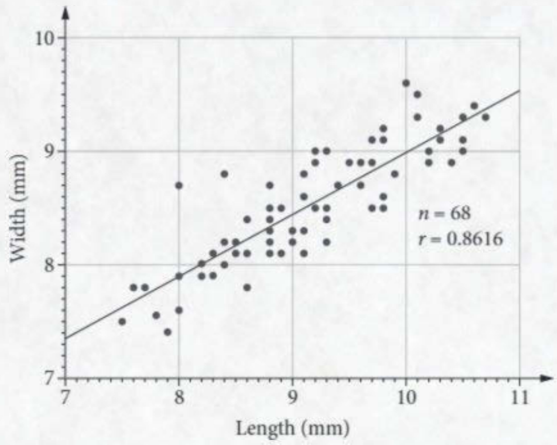
C a least squares line

D a segmented bar chart

E the correlation coefficient *r*

*Use the following information to answer the next two questions.*

The scatterplot displays the beak *length* and beak *width* of 68 birds of the same species. A least squares line has been fitted to the data.




Data: Howard Hughes Medical Institute

The correlation coefficient is $r = 0.8616$. The least squares line has been fitted to the scatterplot using beak *length* as the explanatory variable.

4 ©VCAA 2017N1CQ10 The equation of this line is closest to

A *width* = 0.56 + 3.5 x *length*

B *width* = 3.5 + 0.56 x *length*

C *width* = 7.4 + 0.56 x *length*

D *length* = 7.4 + 0.56 x *width*

E *length* = 3.5 + 0.56 x *width*

5 ©VCAA 2017N 1CQ11 Which one of the following statements is **not** true?

A The slope of the least squares line is positive.

B Birds with longer beaks tend to have wider beaks.

C There is a strong positive linear association between beak *width* and beak *length* for these birds.

D Approximately 74% of the variation in beak *width* is explained by the variation in beak *length.*

E Using the least squares line to predict the beak *width* of a bird with a beak *length* of 8.1 mm would be an example of extrapolation.

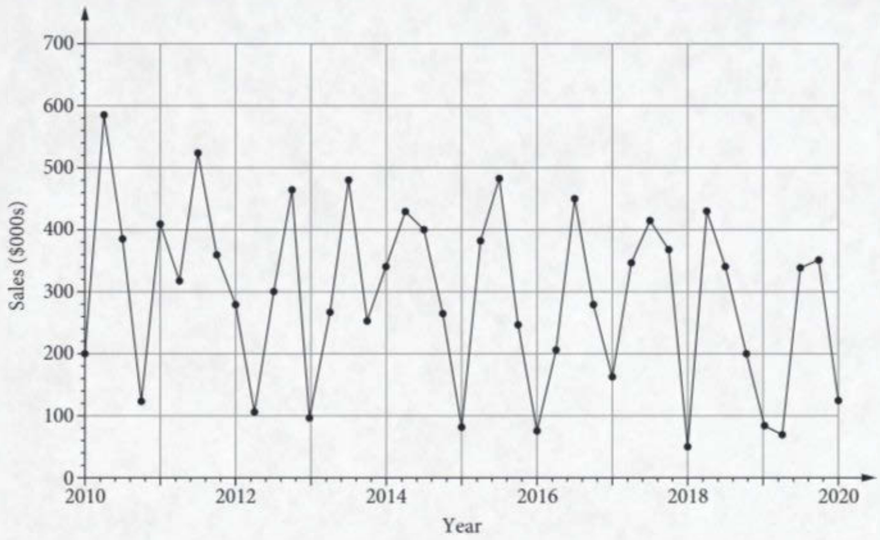
6 © VCAA 2Q18N1CQ14 A company sells central heating systems. The table shows the quarterly seasonal indices for sales in the last three quarters of a year.

| | Quarter 1 | Quarter 2 | Quarter 3 | Quarter 4 |
|---|---|---|---|---|
| Seasonal index | | 1.33 | 1.45 | 0.58 |

The de-seasonalised sales for Quarter 1 in 2018 were \$2.45 million. The actual sales, in millions of dollars, were closest to

A 0.64 B 1.57 C 2.16 D 2.45 E 2.83

7 ©VCAA | 202110012^ The time series plot shows the quarterly *sales,* in thousands of dollars, of a small business for the years 2010 to 2020.




The time series plot is best described as having

A seasonality only.

B irregular fluctuations only.

C seasonality with irregular fluctuations.

D a decreasing trend with irregular fluctuations.

E a decreasing trend with seasonality and irregular fluctuations.

*Use the following information to answer the next two questions.*

The table below shows the long-term average rainfall (in mm) for summer, autumn, winter and spring. Also shown are the seasonal indices for summer and autumn. The seasonal indices for winter and spring are missing.

| | Season | | | |
|---|---|---|---|---|
| | Summer | Autumn | Winter | Spring |
| Long-term average rainfall (mm) | 52.0 | 54.5 | 48.8 | 61.3 |
| Seasonal index | 0.96 | 1.01 | | |

8 ©VCAA 2012 1CQ11 The seasonal index for spring is closest to

A 0.90 B 1.03 C 1.13 D 1.15 E 1.17

9 ©VCAA 2012 1CQ12 In 2011, the rainfall in autumn was 48.9 mm.

The de-seasonalised rainfall (in mm) for autumn is closest to

A 48.4 B 48.9 C 49.4 D 50.9 E 54.0

10 ©VCAA 20171CQ11 Which one of the following statistics can never be negative?

A the maximum value in a data set

B the value of a Pearson correlation coefficient

C the value of a moving mean in a smoothed time series

D the value of a seasonal index

E the value of the slope of a least squares line fitted to a scatterplot

11 ©VCAA ^Ttwrm The quarterly seasonal indices for tractor sales for a supplier are displayed in Table 1.

Table 1

| Quarter number | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Seasonal index | 1.6 | 0.6 | 0.7 | 1.1 |

The quarterly tractor sales in 2014 for this supplier are displayed in Table 2.

Table 2

| Quarter number | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Sales (tractors sold) | 2800 | 1032 | 875 | 759 |

The sales data in Table 2 is to be de-seasonalised before a least squares line of best fit is fitted.
The equation of this least squares line of best fit is closest to

A *de-seasonalised sales* = 0.32 + 910 x *quarter number*

B *de-seasonalised sales* = 370 - 2300 x *quarter number*

C *de-seasonalised sales* = 910 + 0.32 x *quarter number*

D *de-seasonalised sales* = 2300 - 370 x *quarter number*

E *de-seasonalised sales* = 2300 - 0.32 x *quarter number*



9780170448406

# Cumulative examination 2



Total number of marks: 16 Reading time: 5 minutes Writing time: 24 minutes

**1** ©VCAA | 2002 2C J (15 marks) Over recent years, the salaries of Patagonian cricketers have increased rapidly. The following data gives the average salaries in dollars of a large group of these cricketers over the period 1991-2000.

| Year | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 |
|---|---|---|---|---|---|---|---|---|---|---|
| Average salary | 45000 | 47000 | 50000 | 58000 | 70000 | 78000 | 93000 | 105000 | 126000 | 142000 |

a This data will be used to predict future average salaries of Patagonian cricketers.

In this analysis, what is the explanatory variable? 1 mark

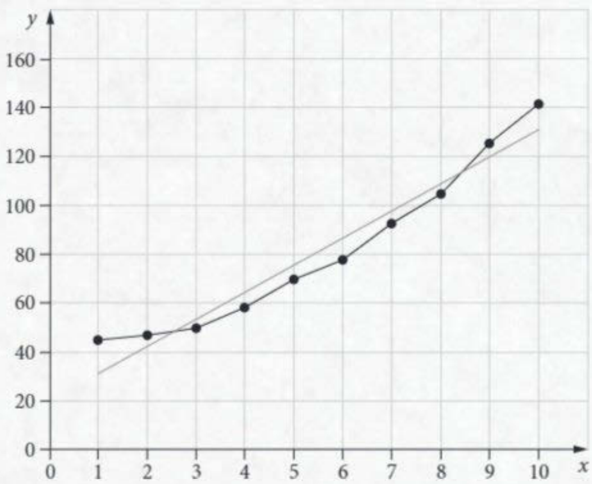
To begin the analysis, the years have been rescaled as $x = 1$ to $x = 10$(1991 = 1,1992 = 2, and so on), and average salary rescaled in thousands of dollars as the variable $y$.

| Year | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Average salary ($000s) | 45 | 47 | 50 | 58 | 70 | 78 | 93 | 105 | 126 | 142 |

This rescaled data is displayed as a time series plot. Also displayed is the least squares line of best fit which has been determined for this rescaled data.

The equation of the least squares line of best fit is

$y = 20.9 + 10.99x$

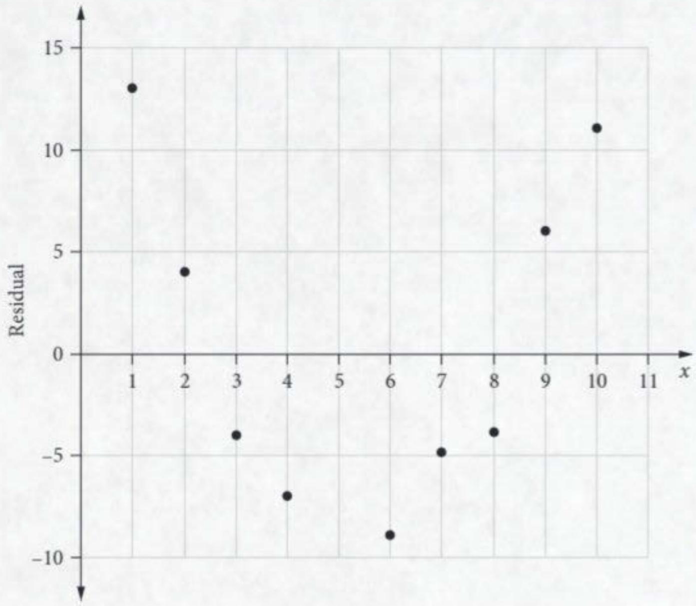



b Using the line of best fit equation $y = 20.9 + 10.99x$ to model the increase in these cricketers' average salaries, copy and complete the following

i The average salary increase for the cricketers was $ ☐ per year. 2 marks

ii The average cricketers' salary in the year 2005 will be about $ ☐. 2 marks

From the time series plot, the increase in Patagonian cricketers' salaries over time appears non-linear. This can be confirmed by constructing the corresponding residual plot as shown.




c Copy and complete the plot by

i calculating the value of the residual for 1995 — 1 mark

ii plotting this residual as a point on the graph. — 1 mark

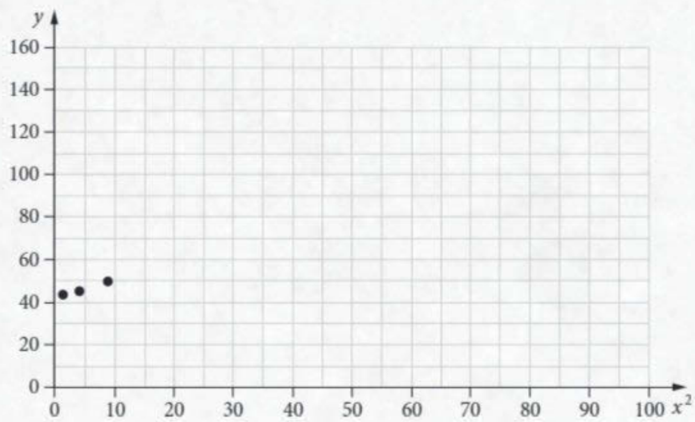
The time series, along with the residual plot, shows that the growth of salaries with time is non-linear. Inspecting the time series, it would appear that it would be appropriate to use an $x^2$ transformation to transform the data to linearity.

The original data has been reproduced in the table below and an extra row has been added for the transformed variable, $x^2$.

| Year (x) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Year$^2$ ($x^2$) | | | | | | | | | | |
| Average salary (y) ($000s) | 45 | 47 | 50 | 58 | 70 | 78 | 93 | 105 | 126 | 142 |

d i Copy and complete the table. — 1 mark

ii Copy and complete the time series plot below of the transformed data to show that the $x^2$ transformation has produced a more nearly linear plot. (Note the first three points have been plotted.) — 2 marks





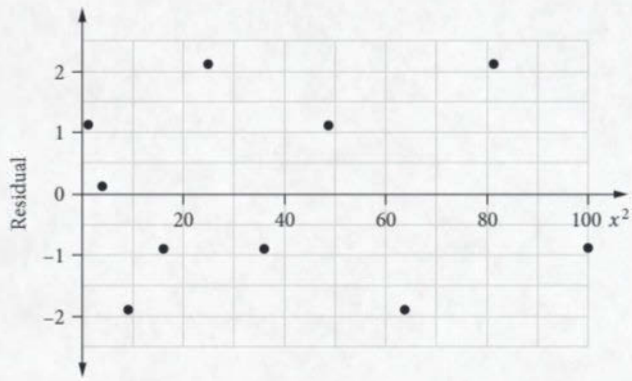
9780170448406

iii Find the equation of the least squares line of best fit for the transformed data. Copy the equation below and write the coefficients, correct to two decimal places, in the boxes. 2 marks

$y = \square + \square\, x^2$

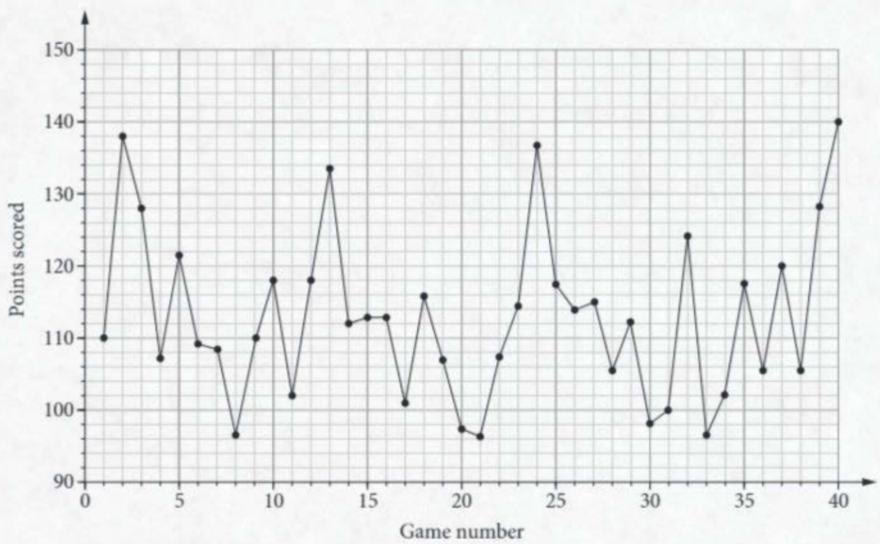
iv Use this line of best fit equation to predict the average salary of this group of Patagonian cricketers in 2005. 2 marks

The residual plot that results from fitting a least squares line of best fit to this transformed data is shown.




v This plot suggests that the $x^1$ transformation has been successful in linearising the time series plot. What feature of this residual plot supports this conclusion? 1 mark

2 ©VCAA 2021 1CQ13 MODIFIED (1 mark) The time series plot below shows the *points scored* by a basketball team over 40 games. What is the nine-median smoothed *points scored* for game number 10?



CHAPTER

# 5 INTEREST AND DEPRECIATION

## Study Design coverage

### AREA OF STUDY 2: DISCRETE MATHEMATICS

#### Depreciation of assets

- use of a first-order linear recurrence relation of the form: $u_0 = a,\ u_{n+}i = Ru_n + d$ where $a,\ R$ and $d$ are constants to generate the terms of a sequence
- use of a recurrence relation to model and compare (numerically and graphically) flat rate, unit cost and reducing balance depreciation of the value of an asset with time, including the use of a recurrence relation to determine the depreciating value of an asset after $n$ depreciation periods for the initial sequence
- use of the rules for the future value of an asset after $n$ depreciation periods for flat rate, unit cost and reducing balance depreciation and their application.

#### Compound interest investments and loans

- the concepts of simple and compound interest
- use of a recurrence relation to model and analyse (numerically and graphically) a compound interest investment or loan, including the use of a recurrence relation to determine the value of the compound interest loan or investment after $n$ compounding periods for an initial sequence from first principles
- the difference between nominal and effective interest rates and the use of effective interest rates to compare investment returns and the cost of loans when interest is paid or charged, for example, daily, monthly, quarterly
- the future value of a compound interest investment or loan after $n$ compounding periods and its use to solve practical problems.



STUDY DESIGN COVERAGE

**Video playlists (8):**

5.1 Recurrence relations
5.2 Simple interest
5.3 Flat rate depreciation
5.4 Unit cost depreciation
5.5 Compound interest
5.6 Effective interest rates
5.7 Reducing balance depreciation
**VCE question analysis** Interest and depreciation

**Worksheets (1):**

5.6 Effective interest rates

**Puzzles (1):**

5.5 Compound interest puzzle

# Nelson MindTap

To access resources above, visit
cengage.com.au/nelsonmindtap

Video playlist
Recurrence relations

# 5.1 Recurrence relations

## Recurrence relations and sequences

A sequence is a list of numbers called terms or values, separated by commas. The sequences we will be dealing with follow a pattern and continue forever.

For example:

3,6,12,24,48 ... follows the pattern that we start at 3 and multiply each term by 2 to find the next term.

A sequence can be written as a recurrence relation, which shows how a particular term in a sequence can be found from the previous term in the same sequence. It consists of two equations:

1 An equation telling us the first term of the series.

2 An equation telling us how to find a term if we know the previous term.

Calculations that continually use the previous answer to find the next answer are called recursive computation (or recursion).

For example:

The recurrence relation that generates the sequence 3, 6,12,24,48 ... is

$^{W}0 = {}^{3}{>}\ {}^{W}n_{+}l = {}^{2w}n$

where $u_0$ is the starting term and $u_n$ is the term before $u_{n+1}$.

The calculations required to generate the sequence for this recurrence relation are:

1 Start with 3.

2 Multiply each term by 2 to find the next term.

3 Keep going.

It generates the sequence in this way:

$w_0 = 3$
$Uj = 2u_0 = 2x3 = 6$
$u_2 = 2u_f = 2 \text{ x } 6 = 12$
$u_3 = 2U_2 = 2 \text{ x } 12 = 24$
$w_4 = 2M_3 = 2 \text{ x } 24 = 48$

and so on.

We will be looking at recurrence relations in the form $u_0 = a$, $u_{M+}\ j = Ru_n + d$. The starting value will always be $u_0$ (not iq).

**Recurrence relations**

A recurrence relation is a rule that generates a sequence by connecting each value to previous values.

It consists of

- the starting value, e.g. $u_0 = 3$
- a rule linking each value to the one before it, e.g. $u_{n+}\ j = u_n + 5$

## WORKED EXAMPLE 1 Generating a sequence using a recurrence relation

A sequence has the recurrence relation $u_Q = 2$, «n+i - 3u,, - 1.

a Describe in words the calculations required to generate the sequence.

b Find the first four terms of the sequence generated by this recurrence relation, showing all the steps.

c Find $u_5$.

| Steps | Working | |
|---|---|---|
| a State the starting value and the rule for the recursive relation. | Start with 2. Multiply each value by 3 and then subtract 1 to find the next value. | |
| b 1 Write the first value $u_0$. | «o = 2 | |
| 2 Substitute $u_0$ into the rule to find $u_1$. | $u_1 = 3u_0 - 1$<br>=3x2-1<br>= 5 | |
| 3 Repeat for the next two values, $u_2$ and $u_3$. | $M_2$ = 3u, - 1<br>=3x5-1<br>= 14 | $M_3$ - 3$M_2$ - 1<br>= 3 x 14-1<br>= 41 |
| 4 List the first four terms of the sequence. | 2,5,14,41 | |
| c Use the rule to find $u_4$ first and then use it to find $u_5$. | $u_4 = 3U_3 - 1$<br>= 3x41-1<br>= 122 | $u_5 = 3u_4 - 1$<br>= 3 x 122 - 1<br>= 365 |

5.1

WB

p. 58

### Exam hack

Although we will usually be using *u* as the variable for the recurrence relation, other variables are often used.

## USING CAS 1 Generating a sequence using recursive computation

Use recursive computation to find the first six terms of the sequences defined by each of the following recurrence relations.

**a** $u_0 = 12, \quad u_{n+1} = u_n - 4$

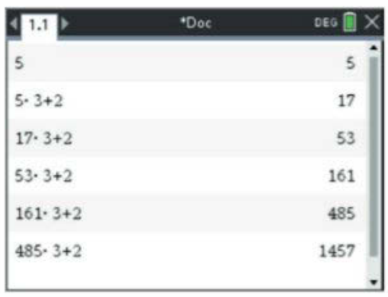
**b** $V_0 = 5, \quad V_{n+1} = 3V_n + 2$

**TI-Nspire**

a


1 Start a new document and add a **Calculator** page.
2 Enter **12** then press **enter**.
3 Enter **–4** then press **enter**.
4 Continue to press **enter** until the first six terms are displayed.

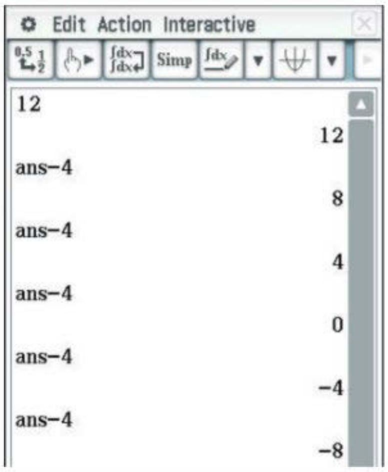
b




1 Press **menu** > **Actions** > **Clear History** to remove the previous calculations.
2 Enter **5** then press **enter**.
3 Enter **×3 + 2** then press **enter**.
4 Continue to press **enter** until the first six terms are displayed.

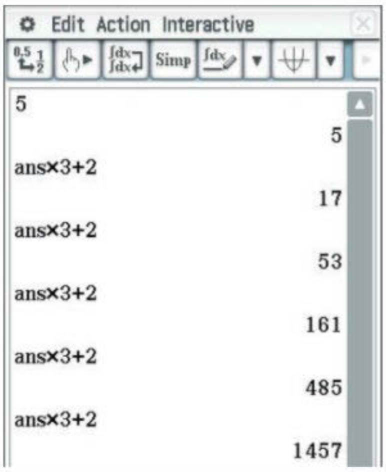
## ClassPad

a




1 Tap Main and clear all entries.
2 Enter 12 then press EXE.
3 Enter -4 then press EXE.
4 Continue to press EXE until the first six terms are displayed.

b




1 Clear all entries.
2 Enter 5 then press EXE.
3 Enter x3 + 2 then press EXE.
4 Continue to press EXE until the first six terms are displayed.

WB
p. 59

## WORKED EXAMPLE 2 Writing recurrence relations

Write the recurrence relation for each of the following.

a Start with 6. Multiply each value by 0.8 and then add 5.3 to find the next value.

b 4,40,400,4000...

c -11,-6,-1,4...

| Steps | Working |
|---|---|
| a What is the starting value? | $M_0$-6 |
| What is the rule connecting each value to the value before it? | “M+i-○_8“n +5_3 |
| b 1 What is the starting value? | The starting value is 4. |
| 2 What is the rule connecting each value to the value before it? | Multiply each value by 10 to find the next value. |
| 3 Write the recurrence relation. | uo - 4> «,,+1 = 10u,, |
| c 1 What is the starting value? | The starting value is -11. |
| 2 What is the rule connecting each value to the value before it? | Add 5 to each value to find the next value. |
| 3 Write the recurrence relation. | »0--lb M,,+1 = U,, + 5 |

### Exam hack

Make sure you write $u_{n+i}$ not $u_n$ + 1 or *un* + 1.

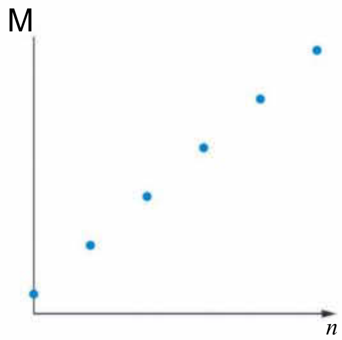
## Graphs of recurrence relations

Recurrence relations can be graphed as a series of points, where 0,1,2, 3 ... are plotted on the horizontal axis and the matching recurrence relation values are plotted on the vertical axis. We can identify four types of graphs from their recurrence relations.

### Linear growth

- Increases by same amount each time.
- Points in an increasing straight line.
- Addition is involved.
- No multiplication is involved.

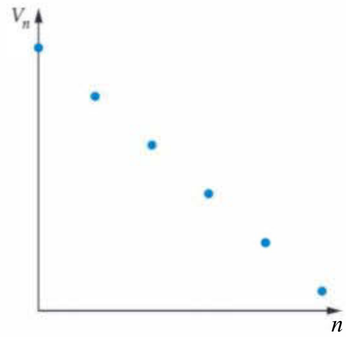
e.g. $V_o = 2$, $V_{n+l} = V$,, + 4




### Linear decay

- Decreases by same amount each time.
- Points in a decreasing straight line.
- Subtraction is involved.
- No multiplication is involved.

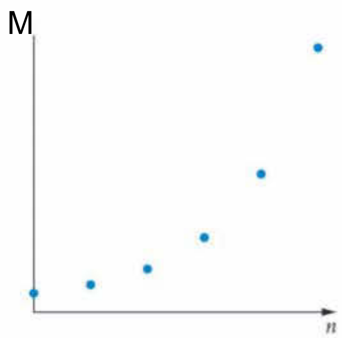
e.g. $V_o = 22$, $V_{,,+1} = V_{,,}-4$




### Geometric growth

- Increases get larger each time.
- Points in an increasing curve.
- No addition or subtraction is involved.
- $V_n$ multiplied by a number greater than 1

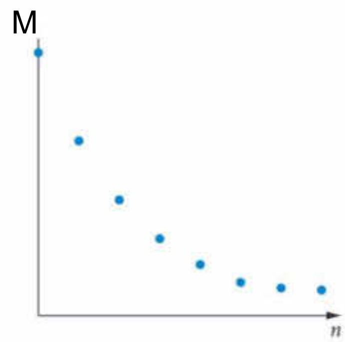
e.g. $V_0 = 35$, $V_{M+1} = 1.62V_n$




### Geometric decay

- Decreases get smaller each time.
- Points in a decreasing curve that never reaches zero.
- No addition or subtraction is involved.
- $V_n$ multiplied by a number between 0 and 1

e.g. $V_o = 35O$, $V_{n+1} = 0.62V$,,



WB

p. 60

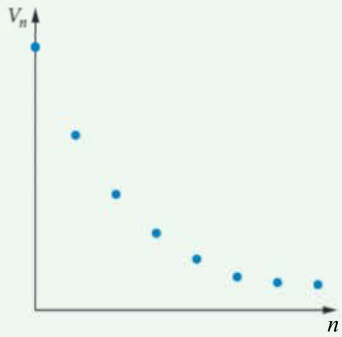
## WORKED EXAMPLE 3 Identifying graphs of recurrence relations

State whether each of the following is true or false and give a reason for your answer in each case.

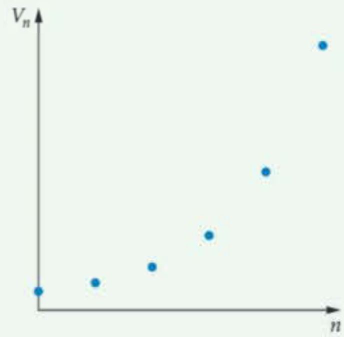
a The graph of the recurrence relation $V_o = 12000$, $V_{,,+1}$ = V,, - 2500 consists of points in an increasing straight line,

b The graph of the recurrence relation $V_o = 6000$, $V_{,,+1} = 6V_n$ consists of points in an increasing straight line.

c The graph of the recurrence relation $V_o = 5400$, $V_{n+1} = 0.9\ V_n$ will look like this.




d The following graph of a recurrence relation shows geometric growth.

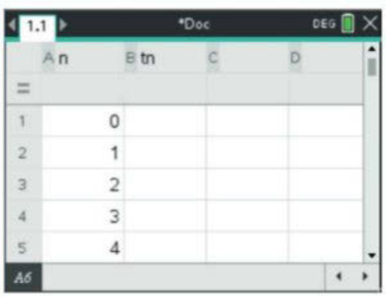


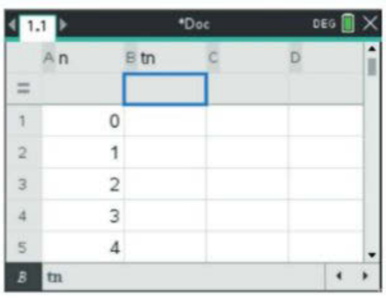

| Steps | Working |
|---|---|
| a Is addition or subtraction involved?<br>Is $V_n$ multiplied by a number greater than 1 or by a number between 0 and 1? | False. Subtraction is involved. No multiplication is involved. So the graph consists of points in a decreasing straight line. |
| b Is addition or subtraction involved?<br>Is $V_n$ multiplied by a number greater than 1 or by a number between 0 and 1? | False. No addition or subtraction is involved. $V_n$ is multiplied by a number greater than 1. So the graph consists of points in an increasing curve. |
| c Is addition or subtraction involved?<br>Is $V_n$ multiplied by a number greater than 1 or by a number between 0 and 1? | True. No addition or subtraction is involved. $V_n$ is multiplied by a number between 0 and 1. So the graph consists of points in a decreasing curve that never reaches zero. |
| d Is the graph a straight line or a curve?<br>Is the graph increasing or decreasing? | True. The graph consists of points in an increasing curve. |

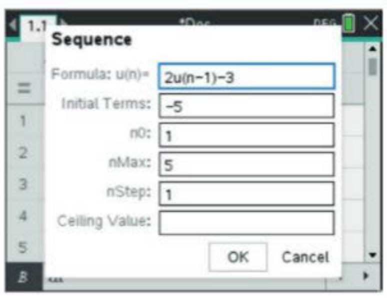
## USING CAS 2 Plotting recurrence relations

Plot the first five terms of the sequence generated by the recurrence relation $t_0 = -5, \quad t_{n+1} = 2t_n - 3$.

**TI-Nspire**




1 Start a new document and add a **Lists & Spreadsheet** page.

2 Label the first two columns n and tn.

3 Enter the digits **0** to **4** in column **A.**

4 Move the cursor to the cell immediately under the column **B** heading (do not click in the cell).

5 Press **menu** > **Data** > **Generate Sequence**.

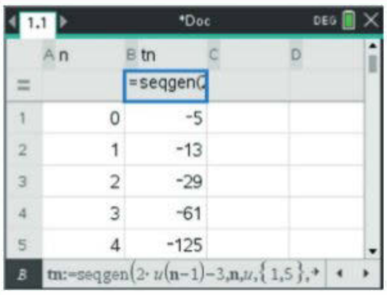
6 Enter the formula and values as shown above. Note that we must use the letter **u** instead of **t** for the recurrence relation.

7 Press **enter**.

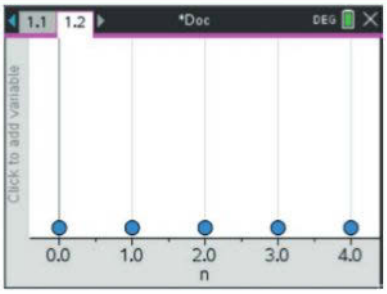


9780170448406

5.1

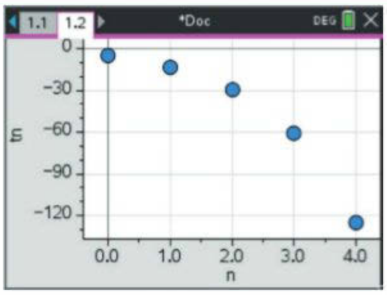



8 The tn terms will be displayed in column B.




9 Open a Data & Statistics page.

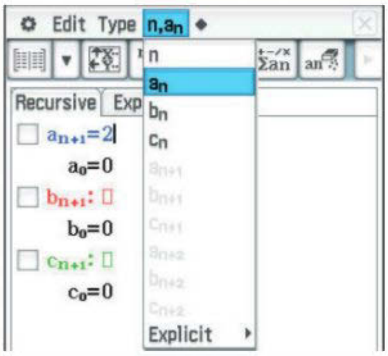
10 For the horizontal axis, select n.




11 For the vertical axis, select tn.

12 The plot of the first five terms in the sequence will be displayed.

**ClassPad**

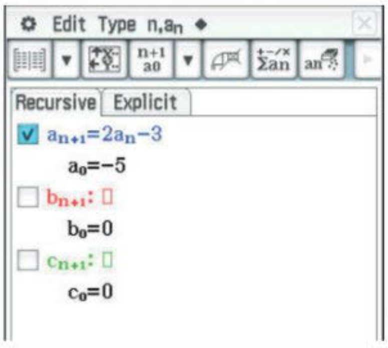



1 Tap Main and open the Sequence application.

2 Clear all sequences.

3 After $a_{n+1}$=, enter the digit 2.

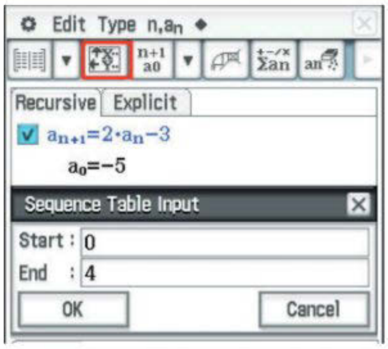
4 Tap on the n,$a_n$ menu and select $a_n$.




5 Enter -3 to complete the sequence formula.

6 After $a_0$=, enter -5.

7 Tap the box to the left of $a_{n+1}$= to ensure it is ticked.

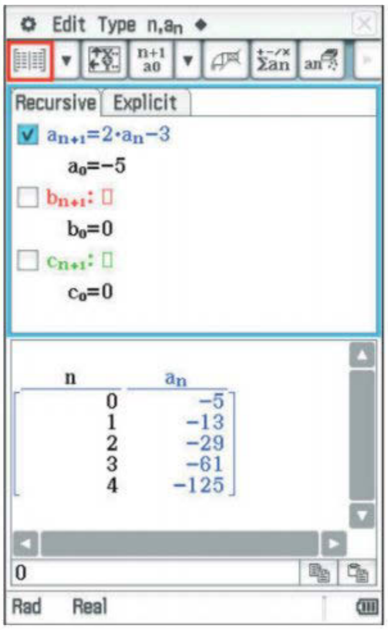



8 Tap the Sequence Table Input tool.

9 Set the Start: value to 0.

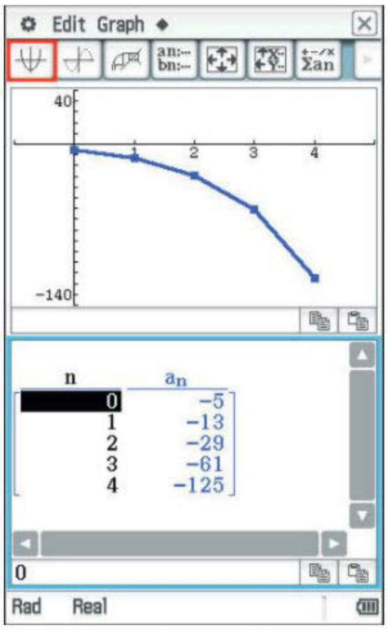
10 Set the End: value to 4.

11 Tap OK.




12 Tap the Table tool.

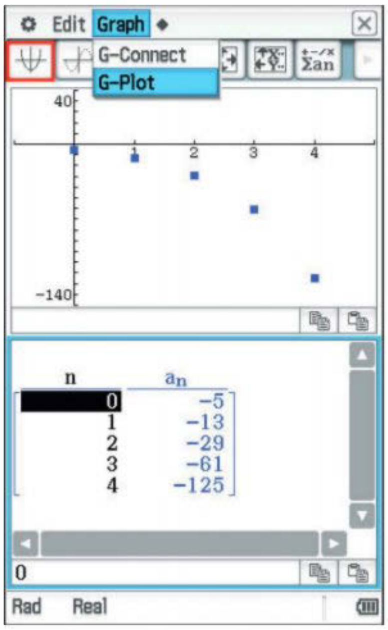
13 The $n$ and corresponding $a_n$ values will be displayed in the lower window.




14 Tap Graph.

15 The graph of the sequence will appear in the upper window.

16 Adjust the window settings to suit.




17 Tap in the lower window to highlight the table.

18 Tap Graph > G-Plot.

## EXERCISE 5.1 Recurrence relations

ANSWERS p. 708

### Mastery

1 0 WORKED EXAMPLE 1 Consider each of the following recurrence relations.

a $u_0 = 3$, $u_{n+1} = 2u_n + 1$

b UQ=12, $U_{n+1} = U_n - 10$

i Describe in words the calculations required to generate the sequence.

ii Find the first three terms of the sequence generated by this recurrence relation, showing all the steps.

iii Find $u_4$.

20 using CAS 1J Use recursive computation to find the first six terms of the sequences defined by each of the following recurrence relations.

a «0 = 15, $u_{n+1} = u_n - 2$

b «o = 2, $«_{n+}i = 4«_n$

c $V_o = -8$, $V_{n+1} - V_n + 7$

d ^0 = 64, $V_{M+}I$--0.5V,,

e $P_0$- 50, ?n+l =^n~^

f ^0 = 3, ^n+l " $2L_n + 1$

9 Q - 1, $C_{W+1} = 3 - 5C_n$

h $S_0 = 40$, $S_{B+}i = 0.5S_n + 2$

30 WORKED EXAMPLE 2 Write the recurrence relation for each of the following.

a Start with 120. Multiply each value by 0.7 and then add 30 to find the next value.

b 2,10,50,250...

c —6, —3, 0, 3 ...

d 60,56,52,48...

e 40,20,10,5...

f 1,12,122,1222 ...

#### © Exam hack

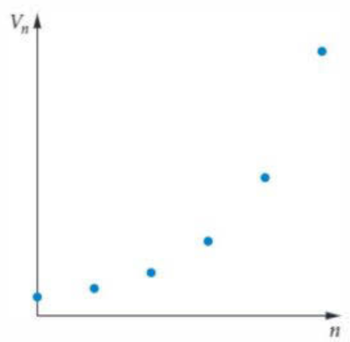
You don’t have to always use $u$ as the variable in a recurrence relation, but make sure you use the same variable for all parts of a recurrence relation. Don’t write, for example: $u_0 = 40$, $V_{n+}| = 0.5\ V_n + 2$.

40 WORKED EXAMPLE 3 State whether each of the following is true or false and give a reason for your answer in each case.

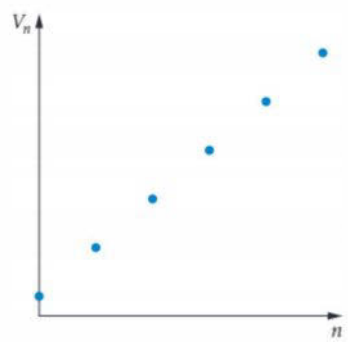
a The graph of the recurrence relation $V_o = 6000$, $V_{n+}1 = V_n + 400$ consists of points in an increasing straight line,

b The graph of the recurrence relation $V_o = 300$, $V_{n+1} = 0.7\ V_n$ consists of points in a decreasing straight line.

c The graph of the recurrence relation $V_o = 2800$, $V_{M+1} = 2.1\ V_n$ will look like this.



d The following graph of a recurrence relation shows geometric growth.



5 Using CAS 2] Plot the first six terms of the sequence generated by the recurrence relation $V_0$=l, $V_{n+1} = 1.3V_n - 5$.



9780170448406

## Exam practice

80-100% 60-79% 0-59%

5.1

6 ©VCAA 2019N1 CQI 7 1 A sequence of numbers is generated by the recurrence relation $P_o = 2$, $P_{n+1} = 3P_W - 1$.
What is the value of $P_3$?

A 2 B 5 C 11 D 41 E 122

7 ©VCAA 2017 1 CQI 8 J 92% The first five terms of a sequence are 2, 6,22, 86, 342 ... The recurrence relation that generates this sequence could be

A $P_o$ - 2, $P_{M+}$ j - $P_n$ + 4

B $P_o = 2$, $P_{n+1} = 2P_{,,} + 2$

C $P_o = 2$, $P_{M+1} = 3P_n$

D $P_o = 2$, $P_{w+1} = 4P_{,,}-2$

E $P_o = 2$, $P_{n+1} = 5P_{,,}-4$

8 ©VCAA J 2016ICQ1T] 89% Consider the recurrence relation $A_o = 2$, $A_{n+1} = 3A_n + 1$. The first four terms of this recurrence relation are

A 0,2,7,22...

B 1,2,7,22...

C 2,5,16,49...

D 2,7,18,54...

E 2,7,22,67...

9 ©VCAA 2020 1CQ21 85% The following recurrence relation can generate a sequence of numbers.

$T_o=10$, $T_{,,+1}$-r,, $_+$ 3

The number 13 appears in this sequence as

A Tj B $T_2$ C $T_3$ D $T_{10}$ E $T_{13}$

10 ©VCAA 2016S 1CQ17 The first three terms of a sequence generated by the recurrence relation below are

$P_o = 2000$, $P_{,,+1} = 1.5P_n - 500$

A 500,2500,2000...

B 2000,1500,1000...

C 2000,2500,3000...

D 2000,2500,3250...

E 2000,3000,4500...

11 ©VCAA Which of the following recurrence relations will generate a sequence whose values decay geometrically?

A $L_o = 2000$, $L_{,,+l} = L_{,,}-100$

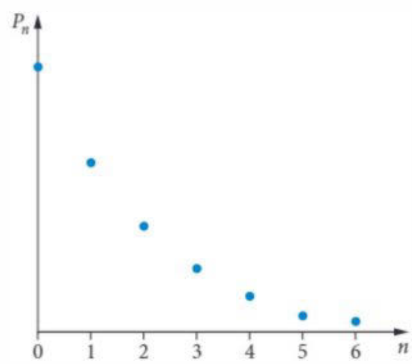
B $L_o$ = 2OOO, £„+, = L„+100

C $Lg = 2000$, ^$n+l$ = $0.65L_{,,}$

D $L_o$ = 2OOO, ^$n+l$ = $6.5L_n$

E $Lg$ - 2000, $L_{n+l} = 0.85L_n-100$

12 The first seven values of a recurrence relation are plotted on the graph shown.




The recurrence relation that could describe the sequence is

A $P_o$ = 4O, $P_{n+l}$- -$P_{n +}0.75$

B $P_o$ = 4O, $P/i+I$-$7$-$5P_n$

C $P_o$ = 4O, $P_{n+l}$- ^$P_{,,}$-0.75

D $P_o$ = 4O, $P_{n+l}$- - 0.75$P_{,,}$ - 7.5

E $P_o$ = 40, $P_{,,+i}$'- 0.75$P_{,,}$

13 ©VCAA 20191CQ17J 68% Consider the recurrence relation shown below.

$A_o$ - 3, $A_{n+1}$ - $2A_n + 4$

The value of $A_3$ in the sequence generated by this recurrence relation is given by

A 2x3+4

B 2x4+4

C 2x10+4

D 2x24 + 4

E 2x52 + 4

14 VCAA 62% The recurrence relation $u_0 = ?$, $u_{n+1} = 4u_n - 2$ generates a sequence. If $u_2 = 2$, then $u_4$ will be equal to

A 4 B 8 C 22 D 40 E 42

15 (4 marks) For the sequence 1, 2, 3.5, 5.75 ...

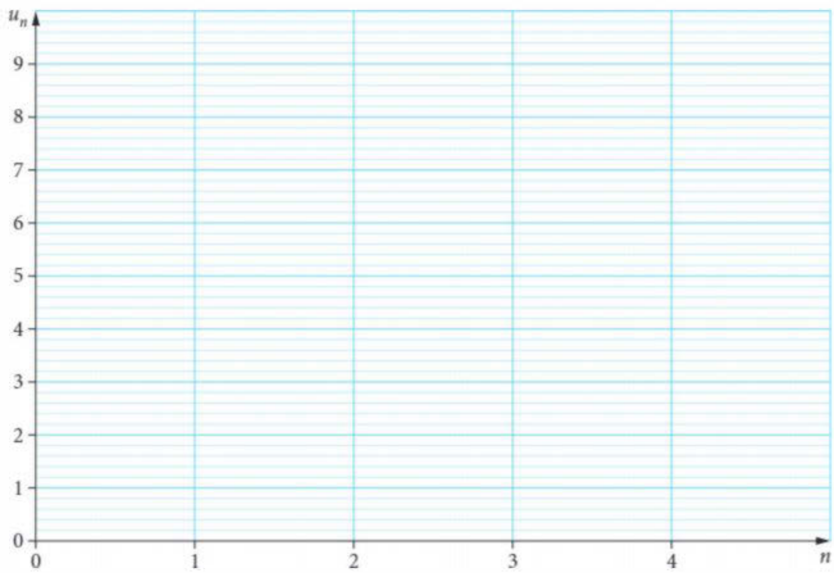
a Copy and complete the following describing the rule in words:

Start withMultiply each value by 1.5 and then add____ to find the next value. 1 mark

b Write the recurrence relation that generates the sequence using the variable $u$. 1 mark

c What is the value of $u_4$ in the recurrence relation? 1 mark

d Copy the axes below and sketch the graph of the first five values of the recurrence relation. 1 mark




# 5.2 Simple interest

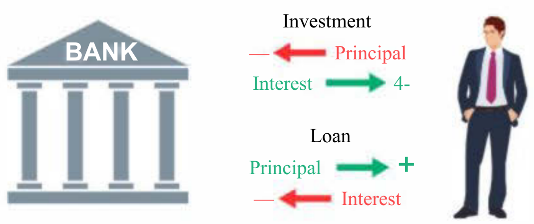
Video playlist Simple interest

## Simple interest recurrence relations and graphs

Interest is the fee for using someone else's money. It applies to both investments and loans.

- For investments, the bank uses our money and pays us the interest.
- For loans, we use the banks' money and we pay the bank the interest.

The amount that is invested or borrowed is called the principal.




Money coming *to the person* is positive.

Money going away *from the person* is negative.

Simple interest is a fixed amount of interest that is paid at regular time periods. When these time periods are years, we use the term per annum (or p.a.), which means per year.

We can use a recurrence relation to show how simple interest works. The starting value in the recurrence relation is the principal. The value of the investment or loan at any time is called the balance.

Loans involve extra payments to the bank in addition to the interest payments. We will focus on investments in this chapter.

5.2

## Simple interest recurrence relations

The recurrence relation for the value $V_n$ of a simple interest investment is:

$V_0$ = principal

$V_{n+1} = V_n + d$

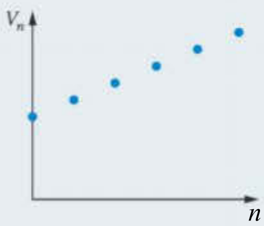
where

$d = \frac{r}{100} \times V_0$ is the fixed amount of interest paid per year

$r$ = the percentage interest rate per year

$n$ = the number of years.

The graph of $V_n$ would look like this.




$V_n - V_0$ = total amount of interest earned on an investment after $n$ years

### WORKED EXAMPLE 4 Using simple interest recurrence relations

WB p. 61

Shari invests $3500 in a bank account earning 3% per annum simple interest.

| Steps | Working |
|---|---|
| **a** What is the fixed amount of interest paid for each year? | |
| Use $d = \frac{r}{100} \times V_0$ to find the fixed amount of interest for each year. | $r = 3, V_0 = 3500$   $d = \frac{r}{100} \times V_0 = \frac{3}{100} \times 3500 = \$105$ |

**b** Copy and complete the following table to find

i Shari's bank account balance after three years

ii the first year when her balance is greater than $4000

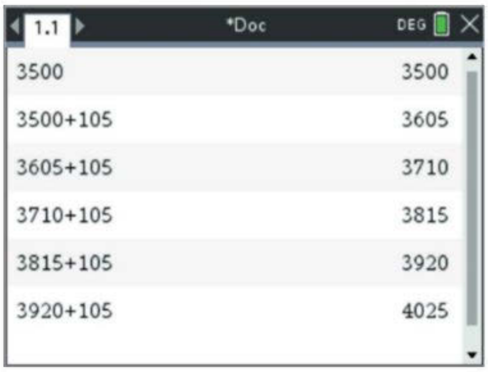
iii the total amount of interest earned after five years.

| $n$ | Account balance after $n$ years ($) |
|---|---|
| 0 | 3500 |
| 1 | 3500 + = |
| 2 | + = |
| 3 | + = |
| 4 | + = |
| 5 | + = |

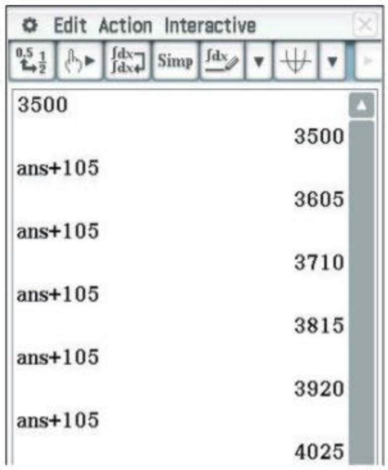
Complete the table by using CAS recursive computation to find the bank account balance after five years.

| $n$ | Account balance after $n$ years ($) |
|---|---|
| 0 | 3500 |
| 1 | 3500 + 105 = 3605 |
| 2 | 3605 + 105 = 3710 |
| 3 | 3710 + 105 = 3815 |
| 4 | 3815 + 105 = 3920 |
| 5 | 3920 + 105 = 4025 |

**TI-Nspire**




**ClassPad**




| | |
|---|---|
| i Read the answer from the table. | Shari's bank account balance after three years is \$3815. |
| ii Read the answer from the table. | Shari's balance is first greater than \$4000 after five years. |
| iii Total amount of interest earned on an investment after $n$ years = $V_n - V_0$ | Total amount of interest earned after five years is $V_5 - V_0 = 4025 - 3500$ $= \$525$ |

c Write a recurrence relation for the account balance.

| | |
|---|---|
| Identify the starting value. Each value is calculated by adding $d$ to the previous value. | Let $V_n$ = the account balance after $n$ years, The recurrence relation is $V_0 = 3500, \quad V_{n+1} = V_n + 105$ |

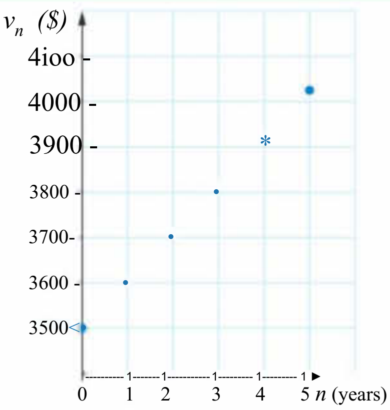
d Describe the sort of growth or decay modelled by the recurrence relation.

| | |
|---|---|
| Is addition or subtraction involved? | Addition is involved. |
| Is multiplication by a number greater than 1 or between 0 and 1 involved? | No multiplication is involved. So, the recurrence relation models linear growth, |

e Sketch the graph of the recurrence relation up to $n = 5$.

The horizontal axis is $n$ (years) and the vertical axis is $V_n$ (\$).

Plot the values from the table.






9780170448406

## WORKED EXAMPLE 5 Using recursive computation for simple interest

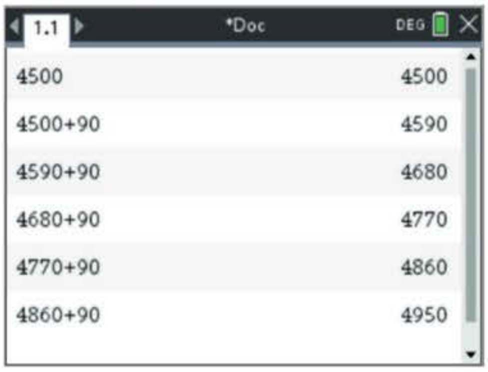
Find the following balances using CAS recursive computation.

**a** Phoebe invests \$4500 in a bank account earning 2% per annum simple interest. What is the balance after five years?

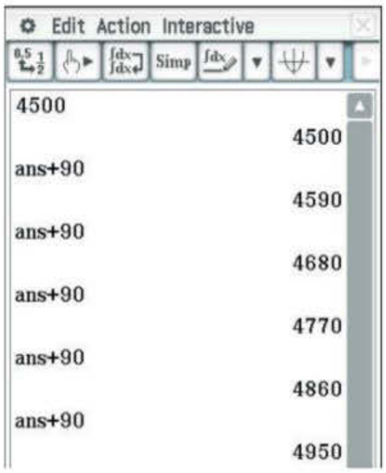
**b** Phil takes out a simple interest loan of \$4500 from a bank and makes monthly payments of \$90. What is the balance after five months?

| Steps | Working |
|---|---|
| **a** 1 Use $d = \frac{r}{100} \times V_0$ to find the fixed amount of interest for each year. Add for an investment and subtract for a loan. | $r = 2, V_0 = 4500$ $\quad d = \frac{r}{100} \times V_0 = \frac{2}{100} \times 4500 = \$90$ This is an investment, so we *add* \$90. |
| 2 Use CAS recursive computation to find the balance after five years. | |

**TI-Nspire**

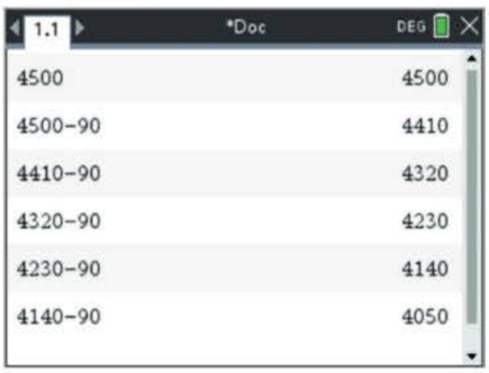



**ClassPad**

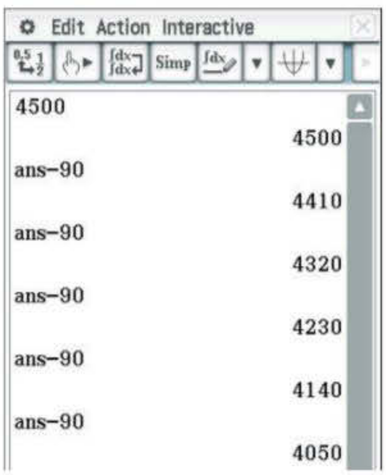



| Steps | Working |
|---|---|
| 3 Write the answer. | The balance after five years is \$4950. |
| **b** 1 Add for an investment and subtract for a loan. | This is a loan, so we *subtract* \$90. |
| 2 Use CAS recursive computation to find the balance after five months. | |

**TI-Nspire**




**ClassPad**




| Steps | Working |
|---|---|
| 3 Write the answer. | The balance after five months is \$4050. |

5.2

WB

p. 63

## Simple interest general rule

Recurrence relations can help us to understand how financial modelling works but the following general rules can be used to solve problems more quickly.

**Simple interest general rule**

The general rule for the value $V_n$ of a simple interest investment is

$$V_n = V_0 + nd$$

where

$V_0$ = principal

$d = \frac{r}{100} \times V_0$ is the fixed amount of interest paid each year

$n$ = the number of years.

When solving for $n$, always round *up*, never down, to the nearest whole number.

To find $r$ use:

$$r = \frac{d}{V_0} \times 100\%$$

where

$r$ = the percentage interest rate per year.

WB p. 65

**WORKED EXAMPLE 6** Using the simple interest general rule

Sasha invests $4000 in an account earning 2% p.a. simple interest.

| **Steps** | **Working** |
|---|---|
| **a** Find the fixed amount of interest paid each year. | |
| Find the value of $d$. | $r = 2$, $V_0 = 4000$<br>$d = \frac{r}{100} \times V_0$<br>$= \frac{2}{100} \times 4000$<br>$= 80$<br>The fixed amount of interest paid each year is $80. |
| **b** Write a rule that will calculate the balance of the account after $n$ years. | |
| Substitute the values of $d$ and $V_0$ into the simple interest general rule. | $d = 80$, $V_0 = 4000$<br>$V_n = V_0 + nd$<br>$V_n = 4000 + 80n$ |
| **c** Use the rule to find the balance of the account after 10 years. | |
| Substitute the value of $n$ into your rule. | $n = 10$<br>$V_n = 4000 + 80n$<br>$V_{10} = 4000 + 80 \times 10$<br>$= 4800$<br>The balance of Sasha's account after 10 years is $4800. |


9780170448406

d Use the rule to find how many years it would take for the investment to double.

1 Substitute the known values into $V_n = V_0 + nd$.

$V_n = 8000$, $V_0 = 4000$, $d = 80$

$V_n = V_0 + nd$

$8000 = 4000 + 80n$

2 Solve for n, using CAS if necessary. Round *up* to the nearest year if necessary.

$80n = 8000 - 4000$

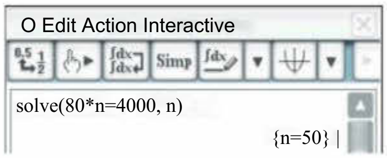
$80n = 4000$

$n = \frac{4000}{80}$

$n = 50$

**TI-Nspire**


**ClassPad**




3 Write the answer.

It would take 50 years for the investment to double.

## Exam hack

The general rule is *not* a recurrence relation. It *doesn't* link one term to the next. If a question asks you to answer using a recurrence relation, you can't answer it with a rule.

## EXERCISE 5.2 Simple interest

ANSWERS p. 708

### Recap

1 © VCAA 2017N1CQ17 A sequence is generated by the recurrence relation

$A_0 = 2$, $A_{n+1} = 3A_n - 3$

The sequence is

A 2,1,0,-3...  
B 2, 3, 0,-3...  
C 2, 3,2, 3...  
D 2, 3, 3, 3...  
E 2,3,6,15...

2 © VCAA 2018N 1CQ18 Consider the recurrence relation

$B_0 = 12$, $B_{n+1} = 2B_n - 14$

Which term of the sequence generated by this recurrence relation is the first to be negative?

A $B_1$  
B $B_2$  
C $B_3$  
D $B_4$  
E $B_5$

## Mastery

**30** WORKED EXAMPLE 4 Dylan invests $6000 in a bank account earning 4% per annum simple interest,

a What is the fixed amount of interest paid for each year?

**b** Copy and complete the following table to find:

i Dylan's bank account balance after three years

ii the first year in which the balance is greater than $7000

iii the total amount of interest earned after five years.

| $n$ | Account balance after $n$ years ($) |
|---|---|
| 0 | 6000 |
| 1 | 6000 + = |
| 2 | + = |
| 3 | + = |
| 4 | + = |
| 5 | + = |

c Write a recurrence relation, $V_n$, for the account balance after $n$ years,

d Describe the sort of growth or decay modelled by the recurrence relation,

e Sketch the graph of the recurrence relation up to $n = 5$.

**40** WORKED EXAMPLE 5 Find the following balances by using CAS recursive computation.

a Cary invests $3800 in a bank account earning 3% per annum simple interest.

b Cerise takes out a simple interest loan of $3800 from a bank and makes monthly payments of $114. What is the balance after six months?

**50** WORKED EXAMPLE 6 For each of the following

i find the fixed amount of interest paid each year

ii write a rule that will calculate the balance of the account after $n$ years

iii use the rule to find the balance of the account after 10 years

iv use the rule to find how many years it would take for the investment to double.

a Stephen invests $10000 in an account earning 4% p.a. simple interest.

b Bee invests $8500 in an account earning 5% p.a. simple interest.

c Novak invests $2000 in an account earning 2.5% p.a. simple interest.

6 If $V_n$ is the account balance after $n$ years for each of the following simple interest investments

i write the recurrence relation for the account balance

ii write the rule for the account balance

iii find the balance after eight years using the rule.

a $8000 invested at 14% p.a.

b $7500 invested at 5% p.a.

c 9200 invested at 4% p.a.

d $6000 invested at 12% p.a.

## Exam practice

80-100% 60-79% 0-59%

7 $48000 is invested at a simple interest rate of 4% per annum. Which of the following recurrence relations can be used to model this investment?

A A. = 48000, •^n+l = A,,+ 192

B $A_o$ = 48000, A,,+1-A,,-1920

C ^0 = 48000, Ai+1 = $A_n$ + 1920

D $A_o$ = 48000, $A_{n+1}$=A,,+ 19200

E ^0 = 48000, Ai+i = A,,-192

8 Which of the following rules could be used to find the balance of $20 000 invested at a simple interest rate of 4% p.a.?

A $V_n$ = 20000 + 8000 x *n* B V,, = 20000 + 8xn C $V_n$ = 20000 - 800 x *n*

D *V,,* = 20000 + 800 x *n* E $V_o$ = 20000, $V_{w+1}$ = $V_n$ + 800

9 © VCAA [illegible] 91% $4000 is invested at a simple interest rate of 5% per annum. The amount of interest earned in the first year is

A $20 B $200 C $220 D $420 E $2000

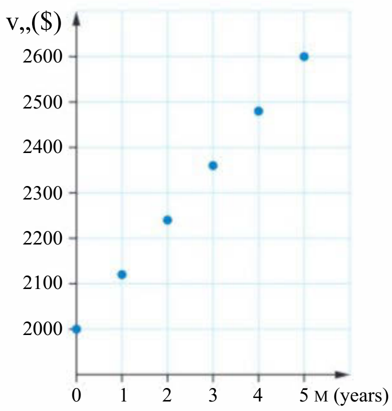
10 © VCAA 2009 1BRMQ1 91% An amount of $800 is invested for two years at a simple interest rate of 4% per annum. The total amount of interest earned by the investment is

A $32 B $64 C $160 D $320 E $640

11 ©VCAA 20041BRMQ1 I 90% Sarah invests $37 000 at a simple interest rate of 4% per annum. The total amount of interest earned in two years is

A $1480 B $2960 C $5920 D $38480 E $39960

12a (5 marks) The following graph shows the value of an investment at 6% p.a. simple interest.

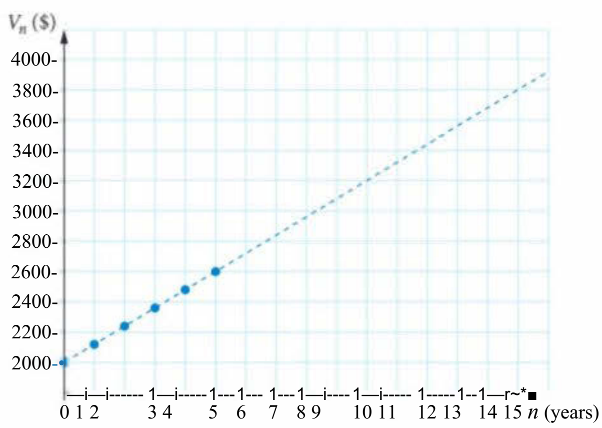



Use the graph to find

i the principal 1 mark

ii the value of the investment after five years. 1 mark

b The following graph has been extended to show the value of the investment for 15 years.




Use the graph to find:

i $V_{l0}$ 1 mark

ii $V_{1S}$ 1 mark

iii the first year in which the value of the investment is $3000. 1 mark

Video playlist
Flat rate depreciation

# @ Flat rate depreciation

## Appreciation and depreciation

When things such as property, gold, antiques and collectibles increase in value over time, it's called appreciation. On the other hand, items purchased by businesses to help them function decrease in value over time. These items, such as computers and machines, are called assets. We use the term depreciation to describe this decrease in value. Depreciation occurs due to age, amount of use or lack of demand. The estimate of the value of an item at any point in time is called the future value.

There are three ways of depreciating assets:

1 Flat rate
2 Unit cost
3 Reducing balance

## Flat rate depreciation recurrence relations and graphs

Flat rate depreciation calculates the future value of an asset by reducing the value every year by a fixed amount. The amount is usually given as a fixed percentage of the purchase price. Flat rate depreciation is the same as a simple interest loan where we subtract a fixed amount each time period.

**Flat rate depreciation recurrence relation**

The recurrence relation for the value of an asset $V_n$ being depreciated using flat rate depreciation is

$V_0$ = initial value of the asset, $\quad V_{n+1} = V_n - d$

where

$d = \frac{r}{100} \times V_0$ is the fixed amount of depreciation each year

$r$ = the percentage depreciation rate per year

$n$ = the number of years.

Its graph will looks like this.

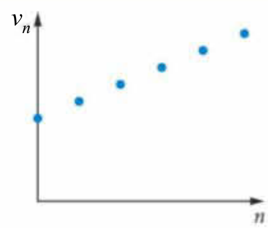

$V_n - V_0$ = total amount of depreciation after $n$ years

**Exam hack**

For an item *appreciating* at a flat rate, *add* the $d$ in the recurrence relation:

$V_o$ = initial value, $V_{n+1} = V_n + d$,

giving a graph like this.



## WORKED EXAMPLE 7 Using flat rate depreciation recurrence relations

5.3

A small forklift is purchased by a business for \$9500. Its value depreciates at a flat rate of 20% each year.

**Steps** | **Working**

**a** What is the fixed amount of depreciation each year?

Use $d = \frac{r}{100} \times V_0$ to find the fixed amount of depreciation each year.

$r = 20$, $V_0 = 9500$

$$d = \frac{r}{100} \times V_0 = \frac{20}{100} \times 9500 = \$1900$$

**b** Copy and complete the following table to find

- **i** the value of the forklift after three years
- **ii** when the value of the forklift first falls below \$2000
- **iii** when the forklift depreciates to zero.

| $n$ | Value after $n$ years (\$) |
|---|---|
| 0 | 9500 |
| 1 | 9500 – = |
| 2 | – = |
| 3 | – = |
| 4 | – = |
| 5 | – = |

WB p. 66

**1** Complete the table by using CAS recursive computation to find the value of the asset after five years.

| $n$ | Value after $n$ years (\$) |
|---|---|
| 0 | 9500 |
| 1 | 9500 – 1900 = 7600 |
| 2 | 7600 – 1900 = 5700 |
| 3 | 5700 – 1900 = 3800 |
| 4 | 3800 – 1900 = 1900 |
| 5 | 1900 – 1900 = 0 |

**2 i** Read the answer from the table.
The value of the forklift after three years is \$3800.

**ii** Read the answer from the table.
The value of the forklift first falls below \$2000 after four years.

**iii** Read the answer from the table.
The forklift depreciates to zero after five years.

**c** Write a recurrence relation for the value of the forklift.

Identify the initial value of the asset. Each value is calculated by subtracting $d$ from the previous value.

Let $V_n$ = value of the forklift after $n$ years.
The recurrence relation is
$V_0 = 9500, \quad V_{n+1} = V_n - 1900$

**d** Describe the sort of growth or decay modelled by the recurrence relation.

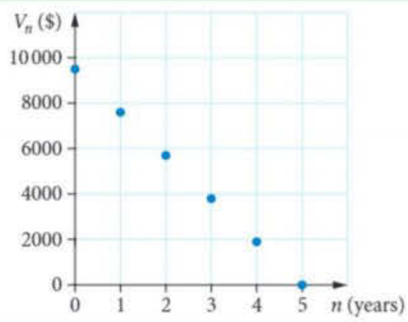
Is addition or subtraction involved?
Subtraction is involved.

Is multiplication by a number greater than 1 or between 0 and 1 involved?
No multiplication is involved.
So, the recurrence relation models linear decay.

**e** Sketch the graph of the recurrence relation up to $n = 5$.

The horizontal axis is $n$ (years) and the vertical axis is $V_n$ (\$).

Plot the values from the table.



## Flat rate depreciation general rule

The flat rate depreciation general rule is the same as the simple interest loan rule, where we subtract a fixed amount each year.

**Flat rate depreciation general rule**

The general rule for the value $V_n$ of a depreciated asset using flat rate depreciation is

$$V_n = V_0 - nd$$

where

$V_0$ = initial value of the asset

$d = \frac{r}{100} \times V_0$ is the fixed amount of depreciation each year

$r$ = the percentage depreciation rate per year

$n$ = the number of years.

When solving for $n$, always round *up*, never down, to the nearest whole number.

The last year of depreciation often involves a partial amount.

To find $r$ use:

$$r = \frac{d}{V_0} \times 100\%$$

**Exam hack**

For an item *appreciating* at a flat rate, *add* the *nd* in the rule:

$$V_n = V_0 + nd$$

WB

p. 68

**WORKED EXAMPLE 8** Using the flat rate depreciation general rule

A company buys a commercial building for \$1 500 000 and depreciates it at a flat rate of 4% per year.

| Steps | Working |
|---|---|
| **a** Find the fixed amount of depreciation each year. | |
| Find the value of $d$. | $r = 4$, $V_0 = 1\,500\,000$<br>$d = \frac{r}{100} \times V_0$<br>$= \frac{4}{100} \times 1\,500\,000$<br>$= 60\,000$<br>The fixed amount of depreciation paid each year is \$60 000. |
| **b** Write a rule that will calculate the value of the building after $n$ years. | |
| Substitute the values of $d$ and $V_0$ into the flat rate depreciation general rule. | $d = 60\,000$, $V_0 = 1\,500\,000$<br>$V_n = V_0 - nd$<br>$V_n = 1\,500\,000 - 60\,000n$ |
| **c** Use the rule to find the value of the building after 10 years. | |
| Substitute the value of $n$ into your rule. | $n = 10$<br>Substituting into<br>$V_n = 1\,500\,000 - 60\,000n$<br>$V_{10} = 1\,500\,000 - 10 \times 60\,000 = 900\,000$<br>The value of the building after 10 years is \$900 000. |



9780170448406

d Use the rule to find how many years it would take for the building to depreciate to zero.

| | |
|---|---|
| 1 Substitute the known values into V,,-$V_0$-*nd*. Let *V,,* - 0. | $V_o$ - 1 500000, *d* - 60000, V,, - 0, M - ?<br>From part b the rule is:<br>V,,= 1500 000-60000M<br>0= 1500000-60000M |
| 2 Solve for M, using CAS if necessary. Round *up* to the nearest year if necessary. | 60000M = 1500000<br>$n = \dfrac{1500000}{60000}$<br>M = 25 |
| 3 Write the answer. | It would take 25 years for the building to depreciate to zero. |

**WORKED EXAMPLE 9** Finding the rate for flat rate depreciation

Steve Gates paid $50 000 for new computer equipment. After four years of flat rate depreciation the equipment was valued at $28 000.

**a** What is the fixed amount of depreciation each year?

**b** What was the annual flat rate of depreciation he used, as a percentage of the purchase price?

| Steps | Working |
|---|---|
| **a 1** Identify what we know and what we need to find from the general rule for flat rate depreciation after *n* years. | Number of years: $n = 4$<br>Value after four years: $V_4 = 28\,000$<br>Initial value: $V_0 = 50\,000$<br>Fixed amount of depreciation each year: $d = ?$ |
| **2** Substitute into the rule and solve, using CAS if necessary. | $V_n = V_0 - nd$<br>$V_4 = V_0 - 4d$<br>$28\,000 = 50\,000 - 4d$<br>$4d = 50\,000 - 28\,000$<br>$4d = 22\,000$<br>$d = \dfrac{22\,000}{4}$<br>$d = 5500$ |
| **3** Write the answer. | The fixed amount of depreciation each year is $5500. |
| **b 1** Identify what we know and what we need to find. | $d = 5500$, $V_0 = 50\,000$, $r = ?$ |
| **2** Substitute the values into $r = \dfrac{d}{V_0} \times 100\%$ and evaluate. | $r = \dfrac{5500}{50\,000} \times 100\%$<br>$= 11$ |
| **3** Write the answer. | The annual flat rate of depreciation is 11%. |



WB p. 69

## EXERCISE 5.3 Flat rate depreciation

ANSWERS p. 709

### Recap

1 Sally-Ann invests $64000 at a simple interest rate of 6% per annum. Which of the following recurrence relations could be used to model this investment?

A Vj = 64000, V,,+J - V,, +38400 B $V_o$-64000, V,,_, = V,,-3840

C $V_o$-64000, $V_{fl+}$i-$V_n$ + 384 D $V_o$- 64000, $V_{,,+1}$ = V,,-38400

E $V_o$ = 64000, ^1 = ^ + 3840

2 @vcaa 89% $6000 is invested in an account that earns simple interest at the rate of 3.5% per annum. The total interest earned in the first four years is

A $70 B $84 C $210 D $840 E $885

### Mastery

30 WORKED EXAMPLE 7 An industrial vacuum cleaner is purchased by a cleaning business for $10 500. Its value depreciates at a flat rate of 25% each year.

a What is the fixed amount of depreciation each year?

b Copy and complete the following table to find

i the value of the vacuum cleaner after two years

ii when the value of the vacuum cleaner first falls below $3000

iii when the vacuum cleaner depreciates to zero.

| $n$ | Value after $n$ years ($) |
|---|---|
| 0 | 10 500 |
| 1 | 10 500 – = |
| 2 | – = |
| 3 | – = |
| 4 | – = |

c Write a recurrence relation for the value of the vacuum cleaner,

d Describe the sort of growth or decay modelled by the recurrence relation,

e Sketch the graph of the recurrence relation up to $n$ = 4.

40 WORKED EXAMPLE 8 For each of the following

i find the fixed amount of depreciation each year

ii write a rule that will calculate the value of the asset after $n$ years

iii use the rule to find the value of the asset after eight years

iv use the rule to find how many years it would take for the asset to depreciate to zero.

a A company buys a commercial property for $2 250000 and depreciates it at a flat rate of 5% per year,

b A business buys a car for $40000 and depreciates it at a flat rate of 10% per year.

c A commercial coffee machine is purchased for $20 500 and is depreciated at a flat rate of 8% per year.

5H WORKED EXAMPLE 9 I Melinda Jobs paid $60 000 for new office furniture. After five years of flat rate depreciation the furniture was valued at $24000.

a What is the fixed amount of depreciation each year?

b What was the annual flat rate of depreciation she used, as a percentage of the purchase price?

5.3

## Exam practice

80-100% 60-79% 0-59%

6 ©VCAA 20191CQ19 I 76% Geoff purchased a computer for $4500. He will depreciate the value of his computer by a flat rate of 10% of the purchase price per annum. A recurrence relation that Geoff can use to determine the value of the computer after *n* years, V,,, is

A $V_o$ = 4500, $V_{,,+1}$- V,, - 450 B $V_o$ = 45OO, $V_{n+1}$ = *V,,* + 450

C $V_o$ - 4500, $V_{,,+}$i - 0.9V,, D Vo-4500, $V_{,,+1}$-1.1V,,

E $V_o$ - 4500, $V_{,,+1}$ - 0.1(V,, - 450)

7 mJ 20201CQ22 J 55% An asset is purchased for $2480. The value of this asset after *n* time periods, V,,, can be determined using the rule

*V,,* - 2480 + 45M

A recurrence relation that also models the value of this asset after *n* time periods is

A VQ-2480, $K_+$i - $V_n$ + 45M B *V,,* = 2480, $^v{}_{n+}$i = V,, + 45M

C $V_o$-248O, Vn+l - $V_{,,}$ + 45 D V, - 2480, $V_{,,+1}$-V,, + 45

E V,, - 2480, $K_+$l - *Vn* + *45*

8 A photocopier bought for $20000 is depreciated at a rate of $3000 every year. Which of the following is *not* true?

A The depreciation method used is flat rate depreciation.

B The recurrence relation that can be used to model this is

$V_o$ = 20000, $V_{M+1}$ = K-3000,

where $V_n$ is the value of the photocopier after *n* years.

C The rule for finding the value after the nth year is $V_n = V_o$ - 3000n, where $V_n$ is the value of the photocopier after *n* years.

D The value of the photocopier after two years is $ 17 000.

E The graph representing the relation is a series of points in a straight line.

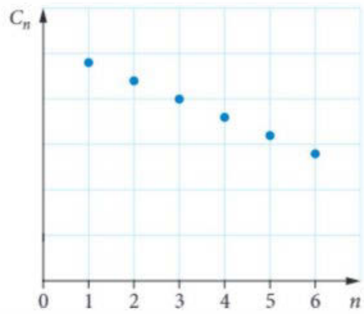
9 An asset purchased for $10000 is depreciated by a flat rate of 15% in the first year and a flat rate of 30% of its purchase price each year thereafter. Which calculation gives the value of the asset at the end of the third year?

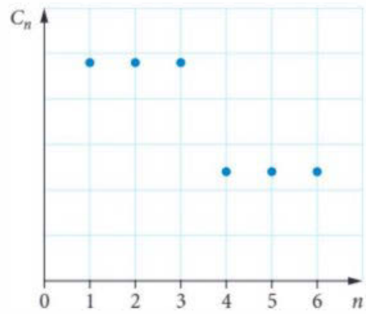
A 10 000 - 1500 - 3000 - 3000 B 10000- 1500-3000

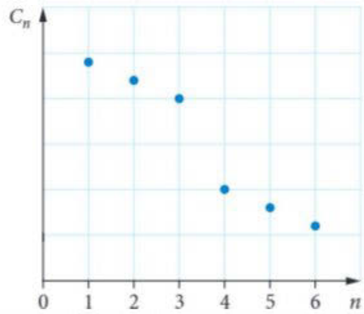
C 10 000 - 3000 - 3000 - 3000 D 10000- 1500- 1500-3000

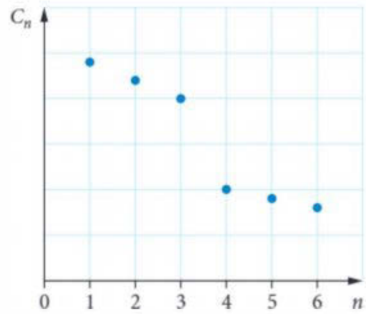
E 10000 - 1500 - 1500 - 1500

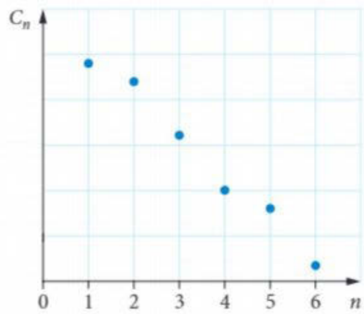
▶ 10 ©VCAA 2019N1CQ20 Marty has been depreciating the value of his car each year using flat rate depreciation. After three years of ownership, the value of the car was halved due to an accident. Marty continued to depreciate the value of his car by the same amount each year after the accident. Which one of the following graphs could show the value of Martys car after *n* years, C,,?

A



B



C



D



E



11 ©VCAA | 20i7 2CQ5abJ (4 marks) Alex is a mobile mechanic. He uses a van to travel to his customers to repair their cars. The value of Alex's van is depreciated using the flat rate method of depreciation. The value of the van, in dollars, after *n* years, $V_w$, can be modelled by the recurrence relation shown below.

$V_o$ = 75OOO, $V_{,,+1}$ = V,,-3375

a 97% Recursion can be used to calculate the value of the van after two years.
Copy and complete the calculations shown. 2 marks

$V_o$ = 75000

Vi = 75 000 -= 71625

$V_2$ =---- - -------=------

b i 96% By how many dollars is the value of the van depreciated each year? 1 mark

ii 69% Calculate the annual flat rate of depreciation in the value of the van.
Write your answer as a percentage. 1 mark ▶

▶ 12 ©VCAA | 2020 2CQ7 , (4 marks) Samuel owns a printing machine. The printing machine is depreciated in value by Samuel using flat rate depreciation. The value of the machine, in dollars, after $n$ years, $V_n$, can be modelled by the recurrence relation

$V_0 = 120000, \quad V_{n+1} = V_n - 15000$

a 94% By what amount, in dollars, does the value of the machine decrease each year? 1 mark

b 72% Showing recursive calculations, determine the value of the machine, in dollars, after two years. 1 mark

c 75% What annual flat rate percentage of depreciation is used by Samuel? 1 mark

d 44% The value of the machine, in dollars, after $n$ years, $V_n$, could also be determined using a rule of the form $V_n = a + bn$. Write down this rule for $V_n$. 1 mark

13 ©VCAA 2019 2CQ7d 52% (1 mark) Phil is a builder who has purchased a large set of tools. He depreciates the value of the tools by a flat rate of 8% of the purchase price per annum. Let $V_n$ be the value of the tools after $n$ years, in dollars. Write down a recurrence relation, in terms of $V_0$, $V_{n+1}$ and $V_n$, that could be used to model the value of the tools using this flat rate depreciation.

14 ©VCAA 2016 2CQ6ab (2 marks) Ken's caravan had a purchase price of $38 000. After eight years, the value of the caravan was $16000.

a 39% Show that the average depreciation in the value of the caravan per year was $2750. 1 mark

## Exam hack

The answer to a 'show that' question needs to be a calculation that *ends* with the number you've been asked to show.

b 33% Let $C_n$ be the value of the caravan $n$ years after it was purchased. Assume that the value of the caravan has been depreciated using the flat rate method of depreciation. Write down a recurrence relation, in terms of $C_{n+1}$ and $C_n$, that models the value of the caravan. 1 mark

## Exam hack

Often the result given in a 'Show that' question part can be used in the next part of the question. This means that it's possible for you to do the next part of the question even if you don't know how to do the 'Show that' part.

15 ©VCAA 2018N2CQ8aHii | (2 marks) Richard is selling his stereo system to help pay for a holiday. The stereo system was originally purchased for $8500. He will sell the stereo system at a depreciated value using a flat rate depreciation method. Let $S_n$ be the value, in dollars, of Richard's stereo system $n$ years after it was purchased. The value of the stereo system, $S_n$, can be modelled by the recurrence relation

$S_0 = 8500, \; S_{n+1} = S_n - 867$

a Using this depreciation method, what is the value of the stereo system four years after it was purchased? 1 mark

b Calculate the annual percentage flat rate of depreciation for this depreciation method. 1 mark ▶

16 ©VCJ 20i7N2CQ5a-c J (3 marks) The snooker table at a community centre was purchased for $3000. After purchase, the value of the snooker table was depreciated using the flat rate method of depreciation. The value of the snooker table, V,,, after *n* years, can be determined using the recurrence relation

Vo = 3000, $V_{n+1}$ = V,,-180

a What is the annual depreciation in the value of the snooker table? 1 mark

b Use recursion to show that the value of the snooker table after two years, $V_2$> is $2640. 1 mark

c After how many years will the value of the snooker table first fall below $2000? 1 mark

# @ Unit cost depreciation

Video playlist
Unit cost depreciation

## Unit cost depreciation recurrence relations and graphs

Unit cost depreciation occurs when an asset is depreciated according to the amount of use it has had, not according to its age. When using the unit cost method of depreciation, the amount of depreciation is determined by applying a rate per unit of use. For example, cars are often depreciated by the number of kilometres travelled, rather than by how old they are.

### Unit cost depreciation recurrence relation

The recurrence relation for the value of an asset $V_n$ being depreciated using unit cost depreciation is

$V_0$ = initial value of the asset

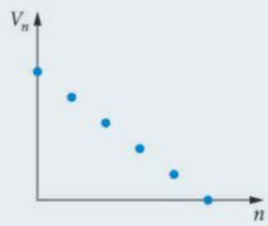
$V_{n+1} = V_n - d$

where

$d$ = cost per unit of use

$n$ = the number of units of use.

The graph of $V_n$ would look like this.




$V_n - V_0$ = total amount of depreciation after $n$ units of use

### Exam hack

The key difference between unit cost and flat rate depreciation is:

- for unit cost depreciation *n* is the number of units of use
- for flat rate depreciation *n* is the number of years of use.

| WORKED EXAMPLE 10 | Using unit cost depreciation recurrence relations |
|---|---|

A collector buys a limited edition original vinyl record of Queen's *Bohemian Rhapsody* for $25000, which depreciates by $3000 every time it's played.

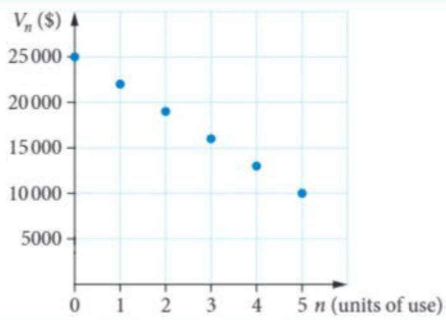
| **Steps** | **Working** |
|---|---|
| a Explain why this is unit cost depreciation not flat rate depreciation. | |
| 1 Refer to 'use' in your answer. | The amount of depreciation is determined by applying a rate per unit of use: $3000 every time the record is played. |
| b Copy and complete the following table to find<br>i the value of the record after three plays<br>ii how many plays it will take for the value of the record to first fall below $10000. | see table below |

| $n$ | Value after $n$ units of use ($) |
|---|---|
| 0 | 25000 |
| 1 | 25000- = |
| 2 | - = |
| 3 | - = |
| 4 | - = |
| 5 | - = |

1 Complete the table by using CAS recursive computation to find the value after five plays.

| $n$ | Value after $n$ units of use ($) |
|---|---|
| 0 | 25000 |
| 1 | 25000-3000 = 22000 |
| 2 | 22000-3000= 19000 |
| 3 | 19000-3000= 16000 |
| 4 | 16000-3000= 13000 |
| 5 | 13000-3000= 10000 |

| **Steps** | **Working** |
|---|---|
| 2 i Read the answer from the table. | The value of the record after three plays is $16 000. |
| ii Read the answer from the table. | The value of the record after five plays is $10 000 so it will fall below $10000 after six plays. |
| c Write a recurrence relation for the value of the record. | |
| Identify the initial value of the asset and the cost per unit of use. | Let $V_n$ = value of the record after $n$ plays.<br>Cost per play = $3000<br>The recurrence relation is<br>$V_0 = 25\ 000, \quad V_{n+1} = V_n - 3000$ |
| d Describe the sort of growth or decay modelled by the recurrence relation. | |
| Is addition or subtraction involved? | Subtraction is involved. |
| Is multiplication by a number greater than 1 or between 0 and 1 involved? | No multiplication is involved.<br>So, the recurrence relation models linear decay. |
| e Sketch the graph of the recurrence relation up to $n = 5$. | |
| The horizontal axis is $n$ (units of use) and the vertical axis is $V_n$ ($).<br>Plot the values from the table. | see graph below |




5.4

WB
p. 70

## Unit cost depreciation general rule

The unit cost depreciation general rule is similar to the flat rate general rule. The difference is for unit cost depreciation $n$ is the number of units of use, whereas for flat rate depreciation $n$ is the number of years.

**Unit cost depreciation general rule**

The general rule for the value $V_n$ of a depreciated asset using unit cost depreciation is

$V_n = V_0 - nd$

where

$V_0$ = initial value of the asset

$d$ = cost per unit of use

$n$ = the number of units of use.

When solving for $n$, always round *up*, never down, to the nearest whole number.

WB

p. 72

**WORKED EXAMPLE 11** Using the unit cost depreciation general rule

A delivery van was purchased for \$80 000. The van's value depreciates at a rate of 72 cents per kilometre.

| Steps | Working |
|---|---|
| **a** Write a rule that will calculate the value of the van after $n$ kilometres of travel. | |
| Substitute the values of $d$ and $V_0$ into the unit cost depreciation general rule $V_n = V_0 - nd$. | $d = 0.72$, $V_0 = 80\,000$<br>$V_n = 80\,000 - 0.72n$ |
| **b** Use the rule to find the value of the van after it has travelled a total distance of 50 000 kilometres. | |
| Substitute the value of $n$ into your rule. | $n = 50\,000$<br>Substituting into<br>$V_n = 80\,000 - 0.72n$<br>$V_{50\,000} = 80\,000 - 0.72 \times 50\,000$<br>$= 44\,000$<br>The value of the van after it has travelled a total distance of 50 000 km is \$44 000. |
| **c** How many kilometres has the van has travelled when it depreciates to \$20 000? Give your answer to the nearest kilometre. | |
| Use the given value of $V_n$ and solve the rule for $n$, using CAS if necessary. Round *up* to the nearest whole number. | $V_n = 20\,000$<br>$V_n = 80\,000 - 0.72n = 20\,000$<br>$0.72n = 80\,000 - 20\,000$<br>$n = \frac{60\,000}{0.72}$<br>$= 83\,333.333\ldots$<br>The van has travelled 83 334 km when it depreciates to \$20 000. |
| **d** How many kilometres has the van has travelled when it depreciates to zero? Give your answer to the nearest kilometre. | |
| Use the given value of $V_n$ and solve the rule for $n$, using CAS if necessary. Round *up* to the nearest whole number. | $V_n = 0$<br>$V_n = 80\,000 - 0.72n = 0$<br>$0.72n = 80\,000$<br>$n = 111\,111.111\ldots$<br>The van has travelled 111 112 km when it depreciates to zero. |

**WORKED EXAMPLE 12** Finding the cost per unit of use

The purchase price of a photocopier is $25000. After five years, the photocopier has a value of $12000. On average, 70 000 pages were copied every year during those five years. The value of the photocopier was depreciated using a unit cost method of depreciation.

Find

a the depreciation in the value of the photocopier, per page copied, to the nearest cent

b $V_w$, the value of the photocopier after *n* units are produced.

| Steps | Working |
|---|---|
| a 1 Find M, the number of units of use during the time period.<br>Note: *n* is *not* the number of years. | The number of pages copied during the five years is<br>*n* = 70 000 pages per year for five years<br>= 70000 x 5<br>= 350000 pages |
| 2 Identify what we know and what we need to find from the general rule for unit cost depreciation after *n* units of use. | Initial value: $V_o$ = 25 000<br>Number of units of use: *n* = 350000<br>Value after 350000 copies: V350000 - 12000<br>Cost per unit of use: d = ?<br>V,,=$V_0$-nd |
| 3 Substitute into the rule and solve, using CAS if necessary. | $V_{350}$QOO = 25000 - 350000 x *d*<br>12000 = 25000 - 350000 x *d*<br>350OOOd = 25000 - 12000<br>350 OOOd = 13000<br>d 13000<br>" 350000<br>= 0.0371... |
| 4 Write the answer to the nearest cent. | The depreciation in the value of the photocopier is $0.04 per page. |
| b Substitute the values of *d* and $V_o$ into the unit cost depreciation general rule $V_n = V_o - nd$ and simplify. | $V_n$ = 25000 - M x 0.04<br>$V_n$ = 25 000 -0.04M |

5.4

WB p. 73

## EXERCISE 5.4 Unit cost depreciation

ANSWERS p. 710

### Recap

*Use the following information to answer the next two questions.*

A company buys a commercial property for $600 000 and depreciates it at a flat rate of 5% per year.

1 Which of the following rules could be used to find the value of the property after the nth year?

A $V_n$ = 600 000 -5*n*

B $V_{,,+1}$ = $V_n$ - 3 000 000M

C $V_{M+1}$ = V,,-30000M

D V,, = 600000-30000M

E *V*,, = 600 000 - 3 000 000M

2 How many years will it take for the investment to depreciate to zero?

A2 B5 C 6 D10 E 20

## Mastery

**30** WORKED EXAMPLE io A collector buys the football used in the 1966 VFL Grand Final for $18000, which depreciates by $2000 every time its kicked.

a Explain why this is unit cost depreciation, not flat rate depreciation.

b Copy and complete the following table to find

i the value of the football after three kicks

ii how many kicks it will take for the value of the football to first fall below $10000.

| $n$ | Value after $n$ units of use ($) |
|---|---|
| 0 | 18 000 |
| 1 | 18 000 – = |
| 2 | – = |
| 3 | – = |
| 4 | – = |
| 5 | – = |

c Write a recurrence relation for the value of the football.

d Describe the sort of growth or decay modelled by the recurrence relation,

e Sketch the graph of the recurrence relation up to $n$ = 5.

40 WORKED EXAMPLE 11 A vehicle that is purchased for $75000 depreciates at a rate of 30 cents per kilometre.

a Write a rule that will calculate the value of the car after $n$ kilometres of travel.

b Use the rule to find the value of the car after it has travelled a total distance of 140000 kilometres,

c For how many kilometres has the car travelled when its value has depreciated to

i $15000? ii zero?

**50** WORKED EXAMPLE 12 The purchase price of a printer is $26 000. After six years the printer has a value of $14000. On average, 80000 pages were copied every year during those six years. The value of the printer was depreciated using a unit cost method of depreciation. Find

a the depreciation in the value of the printer, per page copied, to the nearest cent

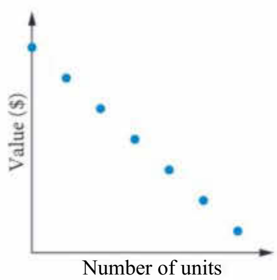
b $V_n$y the value of the printer after $n$ units are produced.

## Exam practice

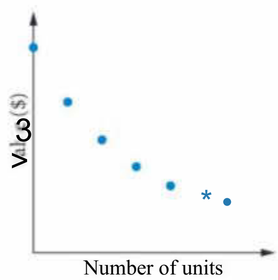
80-100% 60-79% 0-59%

6 © VCAA 2018N 1CQ20 The value of a photocopier is depreciated using a unit cost method. Which one of the following graphs could show the value of the photocopier as it depreciates?

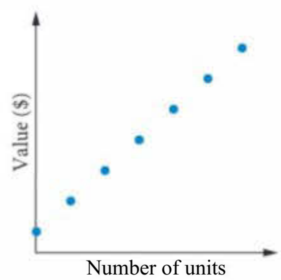
A



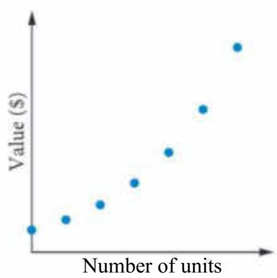

B



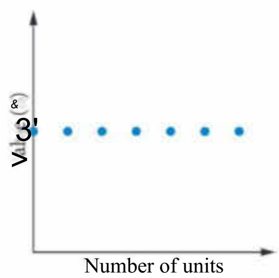

C




D




E



5.4

7 ©VCAA 2011 IBRMQ3 I 82% A van is purchased for $56 000. Its value depreciates at a rate of 42 cents for each kilometre that it travels. The value of the van after it has travelled 32000 km is

A $13400 B $26880 C $29120 D $32480 E $42 560

8 ©VCAA | 20091BRMQ4 I 76% A delivery truck when new was valued at $65 000. The trucks value depreciates at a rate of 22 cents per kilometre travelled. After it has travelled a total distance of 132 600 km, the value of the truck will be

A $14300 B $22100 C $22516 D $29172 E $35828

9 ©VCAA 2007 IBRMQ2 J 64% A car is valued at $30 000 when new. Its value is depreciated by 25 cents for each kilometre it travels. The number of kilometres the car travels before its value depreciates to $8000 is

A 32000 B 55000 C 88000 D 120000 E 550000

10 ©VCAA 20191CQ22 I 53% A machine is purchased for $30 000. It produces 24 000 items each year. The value of the machine is depreciated using a unit cost method of depreciation. After three years, the value of the machine is $18480. A rule for the value of the machine after $n$ units are produced, $V_n$ is

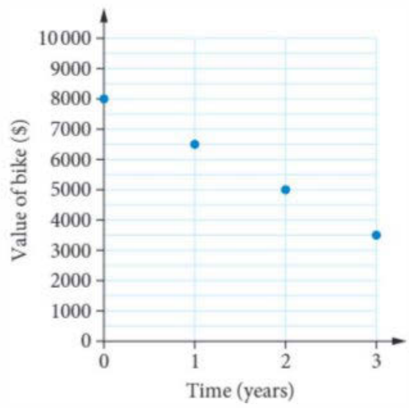
A $V_n = 0.872n$ B $V_n = 24000n - 3840$ C $V_n = 30000 - 24000n$

D $V_n = 30000 - 0.872n$ E $V_n = 30000 - 0.16n$

11 ©VCAA 20171CQ21 ] 53% A printer was purchased for $680. After four years the printer has a value of $125. On average, 1920 pages were printed every year during those four years. The value of the printer was depreciated using a unit cost method of depreciation. The depreciation in the value of the printer, per page printed, is closest to

A 3 cents. B 4 cents. C 5 cents. D 6 cents. E 7 cents.

12 ©VCAA | 2013 2BRMQ1 72% (4 marks) Hugo is a professional bike rider. The value of his bike will be depreciated over time using the flat rate method of depreciation. The graph shows his bike's initial purchase price and its value at the end of each year for a period of three years.




a What was the initial purchase price of the bike? 1 mark

b i Show that the bike depreciates in value by $1500 each year. 1 mark

ii Assume that the bike's value continues to depreciate by $1500 each year. Determine its value five years after it was purchased. 1 mark

The unit cost method of depreciation can also be used to depreciate the value of the bike. In a two-year period, the total depreciation calculated at $0.25 per kilometre travelled will equal the depreciation calculated using the flat rate method of depreciation as described previously.

c Determine the number of kilometres the bike travels in the two-year period. 1 mark

13 ©VCAA 2016 2CQ6c 29% (1 mark) Ken's caravan had a purchase price of $38000. After eight years, the value of the caravan was $16000. It has travelled an average of 5000 kin in each of the eight years since it was purchased. Assume that the value of the caravan has been depreciated using the unit cost method of depreciation. By how much is the value of the caravan reduced per kilometre travelled?

14 ©VCAA 2019N 2CQ6 J (5 marks) Marlon plays guitar in a band. He paid $3264 for a new guitar. The value of Marlon's guitar will be depreciated by a fixed amount for each concert that he plays. After 25 concerts, the value of the guitar will have decreased by $200.

a What will be the value of Marlon's guitar after 25 concerts? 1 mark

b Write a calculation that shows that the value of Marlon's guitar will depreciate by $8 per concert. 1 mark

c The value of Marlon's guitar after $n$ concerts, G,,, can be determined using a rule. Copy and complete the rule below by writing in the appropriate numbers. 1 mark

$G_n = \square - \square \times n$

d The value of the guitar continues to be depreciated by $8 per concert. After how many concerts will the value of Marlon's guitar first fall below $2500? 2 marks

# @ Compound interest

Video playlist Compound interest

Puzzle Compound interest puzzle

WB p. 74

## Compound interest vs simple interest

Most investments use compound interest where the interest is added onto the principal and then the interest for the next time period is calculated using this new amount. The interest is regularly calculated at the end of a certain time period, which is called a compounding period.

**WORKED EXAMPLE 13** Comparing compound and simple interest

Poh invests $3000 for four years and wants to compare an investment at 5% p.a. compounding yearly to 5% p.a. simple interest.

a Copy and complete the following table for $n = 3$ and $n = 4$.

| | Compound | | Simple | |
|---|---|---|---|---|
| $n$ | Interest ($) | Value of investment ($) | Interest ($) | Value of investment ($) |
| 0 | | 3000 | | 3000 |
| 1 | $\frac{5}{100} \times 3000 = 150$ | 3000+ 150 = 3150 | $\frac{5}{100} \times 3000 = 150$ | 3000+ 150 = 3150 |
| 2 | $\frac{5}{100} \times 3150 = 157.50$ | 3150+ 157.50 = 3307.50 | $\frac{5}{100} \times 3000 = 150$ | 3150+ 150 = 3300 |
| 3 | | | | |
| 4 | | | | |

b What is the value of the compound interest investment after four years?

c After four years, how much more is the value of the compound interest investment compared with the simple interest investment?


9780170448406

| Steps | Working |
|---|---|
| a | |

| | Compound | | Simple | |
|---|---|---|---|---|
| $n$ | Interest ($) | Value of investment ($) | Interest ($) | Value of investment ($) |
| 0 | | 3000 | | 3000 |
| 1 | $\frac{5}{100} \times 3000 = 150$ | 3000 + 150 = 3150 | $\frac{5}{100} \times 3000 = 150$ | 3000+ 150 = 3150 |
| 2 | $\frac{5}{100} \times 3150 = 157.50$ | 3150+ 157.50 = 3307.50 | $\frac{5}{100} \times 3000 = 150$ | 3150+ 150 = 3300 |
| 3 | $\frac{5}{100} \times 3307.50 = 165.38$ | 3307.50 + 165.38 = 3472.88 | $\frac{5}{100} \times 3000 = 150$ | 3300+ 150 = 3450 |
| 4 | $\frac{5}{100} \times 3472.88 = 173.64$ | 3472.88 + 173.64 = 3646.52 | $\frac{5}{100} \times 3000 = 150$ | 3450 + 150 = 3600 |

| Steps | Working |
|---|---|
| b Read from the table. | The value of the compound interest investment after four years is $3646.52. |
| c Compare the last entries in the two Value of investment ($) columns in the table. | 3646.52 - 3600 = 46.52<br>The compound interest investment has $46.52 more. |

## Compounding periods

Compound interest is always given as a rate per year, but *compounding periods* can vary.

- Daily compounding means the interest is calculated every day and added to the account.
- Weekly compounding means the interest is calculated every week and added to the account.

Further compounding periods are shown in the table.

| Compounding period | Number of compounding periods per year |
|---|---|
| Daily | 365 |
| Weekly | 52 |
| Fortnightly | 26 |
| Monthly | 12 |
| Quarterly | 4 |
| Six-monthly | 2 |
| Yearly | 1 |

**Interest rates per compounding period**

$$\text{percentage interest rate per compounding period} = \frac{\text{percentage interest rate per year}}{\text{number of compounding periods per year}}$$

**WORKED EXAMPLE 14** Working with compounding periods

For each of the following investments, find the

i number of compounding periods per year

ii number of compounding periods over eight years

iii percentage interest rate per compounding period

iv amount of interest earned in the first compounding period to the nearest cent,

a Karen invests $18000 at 7% compound interest per annum compounding weekly,

b Anton invests $60000 at 4% compound interest per annum compounding daily.

5.5

WB p. 75

9780170448406

| Steps | Working |
|---|---|
| a i How many compounding periods are there per year? | There are 52 weekly compounding periods per year. |
| ii Multiply the number of compounding periods per year by the number of years. | There are 52 x 8 = 416 weekly compounding periods over eight years. |
| iii Divide the percentage interest rate per year by the number of compounding periods per year. | The percentage interest rate per week = $\frac{7}{52}\%$ |
| iv Convert the compounding period interest rate to a decimal and multiply by the principal. Round to the nearest cent. | The amount of interest earned in the first week is: $\frac{7}{52} \times \frac{1}{100} \times 18000 = \$24.23$ |
| b i How many compounding periods are there per year? | There are 365 daily compounding periods per year. |
| ii Multiply the number of compounding periods per year by the number of years. | There are 365 x 8 = 2920 daily compounding periods over eight years. |
| iii Divide the percentage interest rate per year by the number of compounding periods per year. | The percentage interest rate per day = $\frac{4}{365}\%$ |
| iv Convert the compounding period interest rate to a decimal and multiply by the principal. Round to the nearest cent. | The amount of interest earned on the first day is: $\frac{4}{365} \times \frac{1}{100} \times 60\,000 = \$6.58$ |

### Exam hack

It's best to write percentage interest rate per compounding period as fractions rather than decimals to avoid rounding too early in calculations.

## Compound interest recurrence relations and graphs

### Compound interest investment recurrence relation

The recurrence relation for the value $V_n$ of a compound interest investment is

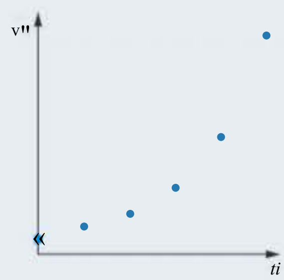
$$V_0 - \text{principal}, \quad V_{n+1} - \left(1 + \frac{r}{100}\right) V_n$$

where

$r$ = the percentage interest rate per compounding period

$n$ = the number of compounding periods.

The graph of a compound interest investment recurrence relation would look like this.



$V_n - V_0$ = total amount of interest earned after $n$ compounding periods


9780170448406

5.5

**WORKED EXAMPLE 15** Finding compound interest recurrence relations

a Write a recurrence relation for the account balance for each of the following savings accounts earning compound interest.

i Xanthi deposited $15 000 into an account at the rate of 2.4% per annum, compounding annually.

ii Elroy deposited $12 000 into an account at the rate of 3% per annum, compounding monthly.

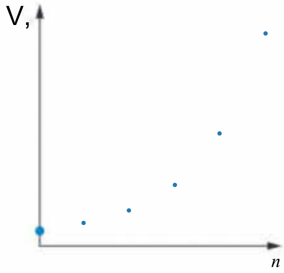
iii Juan deposited $10000 into an account at the rate of 4.8% per annum, compounding quarterly,

b Describe the sort of growth or decay modelled by the recurrence relations.

c Sketch the shape of the graph of the growth or decay modelled by the recurrence relations.

| Steps | Working |
|---|---|
| a i 1 Find the number of compounding periods per year. | There is one compounding period per year. |
| 2 Identify $V_w$, $V_o$ and *r*. | Let $V_n$ = the account balance after *n* compounding periods.<br>$V_o = 15000$, $r = \frac{2.4}{1} = 2.4\%$ |
| 3 Substitute the values into $V_o$ - principal, $V_{,,+}$ j - I $\left(1 + \frac{r}{100}\right) V_,$ and simplify. | $K_+ I = \left(I + \frac{2.4}{100}\right) K$<br>- (1 + 0.024) $V_{,,}$<br>= 1.024V,,<br>$V_o$- 15000, $V_{,,+1}$ - 1.024$V_{,,}$ |
| ii 1 Find the number of compounding periods per year. | There are 12 compounding periods per year. |
| 2 Identify $V_n > V_Q$ and r. | Let $V_n$ = the account balance after *n* compounding periods.<br>$V_o = 12000$, $r = \frac{3}{12} = 0.25\%$ |
| 3 Substitute the values into $V_o$ = principal, $V_{n+1} = \left(1 + \frac{r}{100}\right)V_n$ and simplify. | $V_{n+1} = \left(1 + \frac{0.25}{100}\right)V_n$<br>= (1 +0.0025) $V_n$<br>= 1.0025V,,<br>$V_o$= 12000, $V_{,,+1}$ = 1.0025V,, |
| iii 1 Find the number of compounding periods per year. | There are four compounding periods per year. |
| 2 Identify $V_n$, $V_o$ and r. | Let $V_n$ = the account balance after *n* compounding periods.<br>$V_o$- 10000, r- $\frac{4.8}{4}$ - 1.2% |
| 3 Substitute the values into $V_o$ = principal, $V_{n+}$, - $\left(1 + \frac{r}{100}\right) V_{,,}$ and simplify. | $V_{n+1} = \left(1 + \frac{1.2}{100}\right)V_n$<br>-(l+0.012)V,,<br>- 1.012V,,<br>$V_o$- 10000, $V_{,,+1}$ = 1.012 $V_{,,}$ |

WB p. 76

| | |
|---|---|
| b Is addition or subtraction involved?<br>Is multiplication by a number greater than 1 or between 0 and 1 involved? | No addition or subtraction is involved.<br>Multiplication by a number greater than 1 is involved.<br>So, the recurrence relations model geometric growth. |
| c Show the points forming the shape of the curve for geometric growth. |  |



WB

p. 78

## WORKED EXAMPLE 16 Working with compound interest recurrence relations

Dougie has invested his money in an account earning interest that compounds annually, according to the recurrence relation

$V_o$ = 8OOO, $V_{M+1}$ = 1.035 $V_n$

where $V_n$ is the account balance after $n$ compounding periods.

| Steps | Working |
|---|---|
| a How much money did Dougie initially invest? | |
| Identify $V_o$. | Dougie initially invested $8000. |
| b Use recursion to write down calculations that show that the amount of money in Dougies account after two years will be $8544.96. | |
| Step out the recurrence relation calculations to find $V_2$. | $V_o$ = 8000<br>V, = 1.035 $V_o$ = 1.035 x 8000 = 8280<br>$V_2$ - 1.035$V_1$ = 1.035 x 8280 = 8569.80 |
| c What is the annual percentage compound interest rate for this account? | |
| 1 Use the recurrence relation to find an equation for r, the percentage interest rate per compounding period. | Comparing:<br>$V_{n+1}$ = 1( H-------I \ $V_n$ (r / 100^)<br>$K_+$i - 1.035V,<br>we can see that<br>$1 + \frac{r}{100} = 1.035$ |
| 2 Solve for r, using CAS if necessary. | $\frac{r}{100}$ = 1.035-1<br>$\frac{r}{100}$ = 0.035<br>r=3.5 |
| 3 Multiply r by the compounding period to find the annual percentage compound interest rate. | The interest compounds annually, so the annual percentage compound interest rate is $r$ x 1 = 3.5 x 1 = 3.5% |



9780170448406

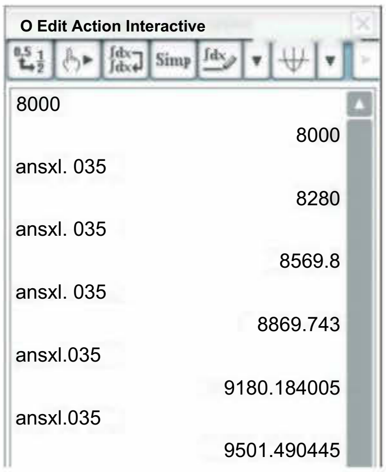
d After how many years will the balance of Dougies account first exceed $9400?

Use CAS recursive computation to continue the recurrence relation calculations until they are greater than the given amount.

The balance of Dougies account first exceeds $9400 after five years.

**TI-Nspire**


**ClassPad**




e What is the total interest, to the nearest cent, earned by Dougies investment after five years?

Total amount of interest earned after $n$ compounding periods = $V_n - V_o$.

Use CAS recursive computation to find $V_n$.

Round your answer to the nearest cent.

$n = 5$

$V_5 = 9501.49$

total interest = $V_5 - V_o$
= 9501.49-8000
= $1501.49

## Compound interest general rule

**Compound interest general rule**

The general rule for the value $V_n$ of a compound interest investment is

$$V_n = \left(1 + \frac{r}{100}\right)^n \times V_0$$

where

$V_o$ = principal
$r$ = the percentage interest rate per compounding period
$n$ = the number of compounding periods.

When solving for H, always round *up,* never down, to the nearest whole number.

WB

p. 80

## WORKED EXAMPLE 17 Using the compound interest rule

Alan inherits $35000 and decides to invest it in an account where he earns interest of 6% p.a. compounded monthly.

| Steps | Working |
|---|---|
| **a** Find r, the percentage interest rate per compounding period. | |
| Divide the yearly interest rate by the number of compounding periods per year. | The percentage interest rate per month = $\frac{6}{12}$% <br> = 0.5% |
| **b** Write a rule that will calculate the value of the investment after $n$ months. | |
| Substitute the values of $V_0$ and r into the compound interest general rule $V_n = \left(1 + \frac{r}{100}\right)^n \times V_0$ and simplify. | $V_n = \left(1 + \frac{0.5}{100}\right)^n \times 35\,000$ <br> $V_n = 1.005^n \times 35000$ |
| **c** Use the rule to find the value of the investment after seven years to the nearest cent. | |
| 1 Find the value of n, the number of compounding periods. | $n$ = the number of months in seven years <br> = 7 x 12 = 84 |
| 2 Substitute the value of $n$, the number of compounding periods, into your rule. | $V_{84} = 1.005^{84} \times 35\,000$ <br> = 53212.937... |
| 3 Write the answer, rounding to the nearest cent. | The value of the investment after seven years is $53212.94 |
| **d** Use a rule to find the value of Alans investment after seven years to the nearest cent if the interest was compounded quarterly instead of monthly. | |
| 1 Find $r$ by dividing the yearly interest rate by the number of compounding periods per year. | The percentage interest rate per quarter = $\frac{6}{4}$% <br> = 1.5% |
| 2 Find the value of n, the number of compounding periods. | $n$ = the number of quarters in seven years <br> = 7x4 <br> = 28 |
| 3 Substitute n, $r$ and $V_0$ into the rule. | $V_{28} = \left(1 + \frac{1.5}{100}\right)^{28} \times 35000$ <br> $V_{28} = 1.51722... \times 35000$ <br> = 53102.776... |
| 4 Write the answer, rounding to the nearest cent. | The value of the investment after seven years is $53 102.78. |
| **e** Which compounding period gives the larger balance after seven years? | |
| Compare the two values. | Monthly compounding gives a larger balance than quarterly compounding after seven years. |

### Exam hack

*Always* check how often compounding occurs. When interest is not compounded yearly, the values of $n$ and $r$ need to be adjusted accordingly.

## USING CAS 3 Creating interest graphs

Emily has $6000 to invest. She is comparing two different accounts. They offer:

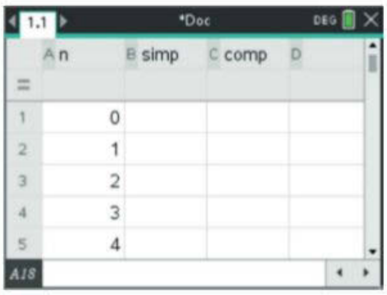
Account 1: simple interest at 6.5% p.a.
Account 2: compound interest at 6% p.a. compounded annually

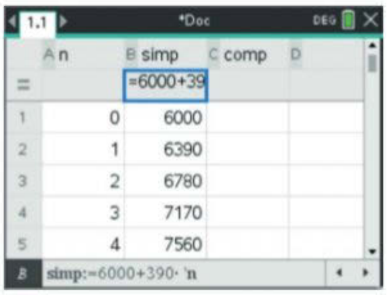
a Create a graph that shows the value of each investment over a 10-year period on the same set of axes,

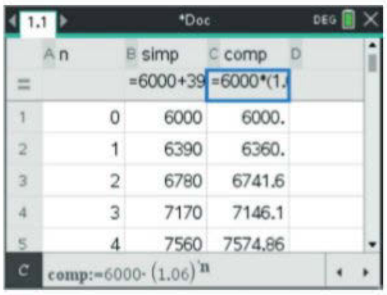
b What would be your advice to Emily?

**TI-Nspire**

a




1 Start a new document and add a Lists & Spreadsheet page.

2 Label the columns as shown above.

3 In column A, enter the values 0 to 10.

4 Click in the cell immediately under the column B heading.

5 Enter the formula =6000+390n (6.5% of $6000 = $390)

6 Press enter.

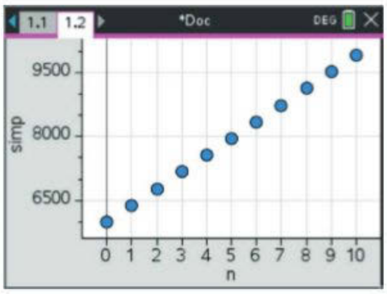
7 When prompted, select Variable Reference and press OK.

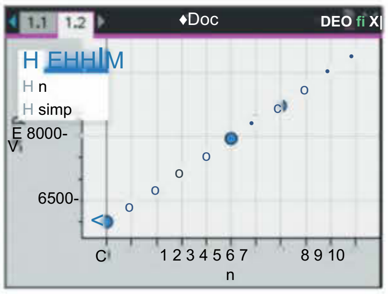
8 Click in the cell immediately under the column C heading.

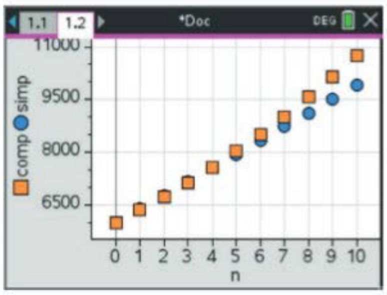
9 Enter the formula =6000x1.06“.

10 Press enter.

11 When prompted, select Variable Reference and press OK.




12 Insert a Data & Statistics page.

13 For the horizontal axis, select n.

14 For the vertical axis, select simp.

15 The graph of the simple interest will be displayed.

16 Press menu > Plot Properties > Add Y Variable.

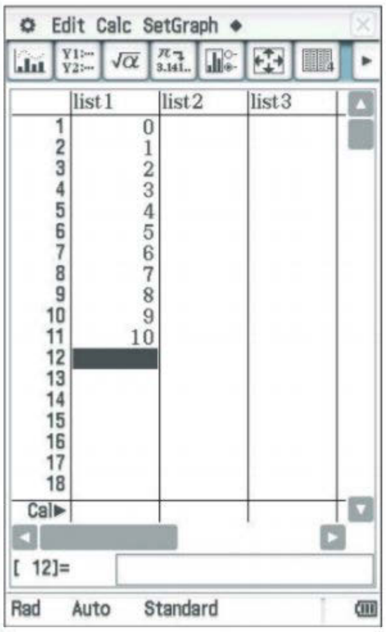
17 Select comp.

18 The graphs of both the simple interest and compound interest will be displayed on the same set of axes.

b Account 2 with compound interest is the better option as the value of the investment is greater after 10 years.

**ClassPad**

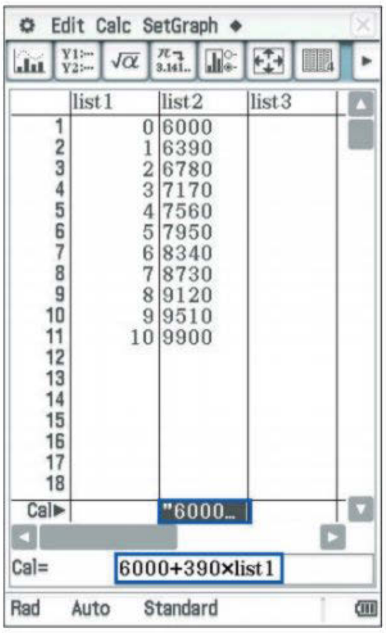
**a**




1 Open the Statistics application.

2 Clear all lists.

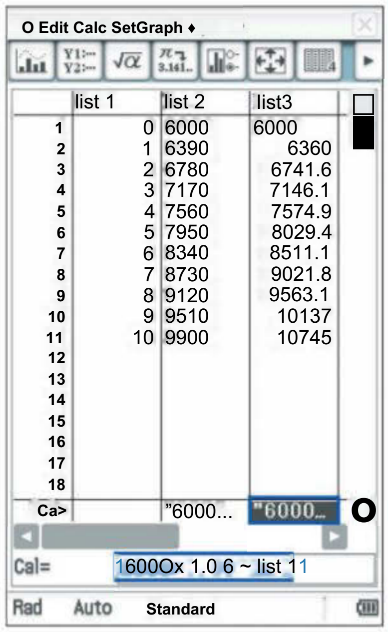
3 In list 1, enter the values 0 to 10.




4 Tap in the Cal cell at the bottom of Iist2.

5 In the Cal= field at the bottom of the screen, enter the formula 6000+390xlist1. (6.5% of \$6000 = \$390)

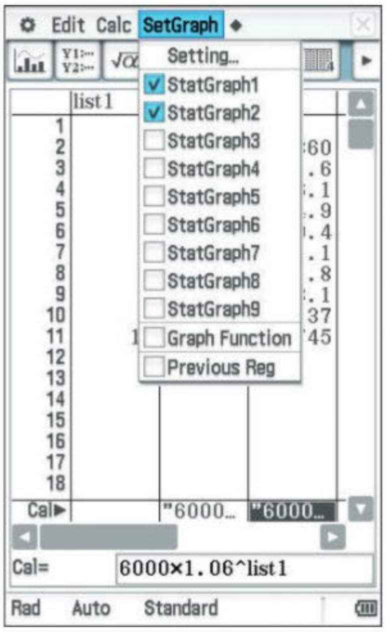
6 Press EXE.




7 Tap in the Cal cell at the bottom of Iist3.

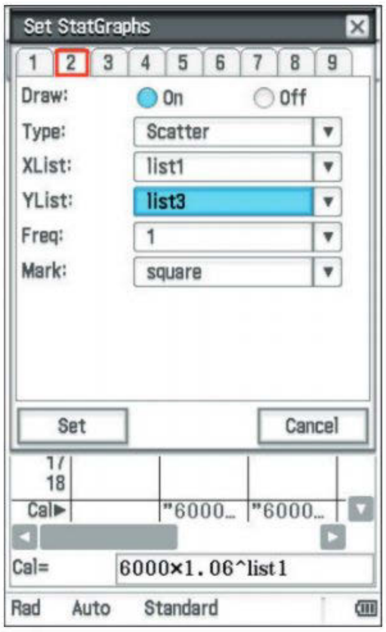
8 In the Cal= field at the bottom of the screen, enter the formula 6000x1.06^list1

9 Press EXE.




10 Tap SetGraph and tap to select both StatGraphl and StatGraph2.

11 Tap SetGraph > Setting.

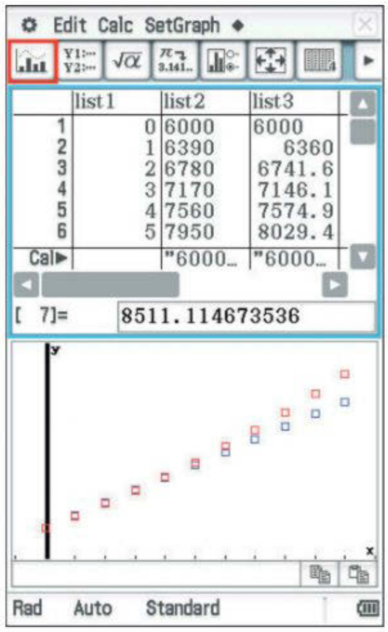



12 For graph 1, keep the default settings of listl and Iist2.

13 Tap tab 2 at the top of the screen.

14 Keep the other default settings but change the YList: field to Iist3.

15 Tap Set.




16 Tap Graph.

17 The graphs of both the simple interest and compound interest will be displayed on the same set of axes in the bottom window.

b Account 2 with compound interest is the better option as the value of the investment is greater after 10 years.



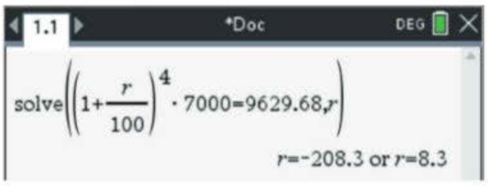
**WORKED EXAMPLE 18** Working with the compound interest rule

Find each of the following using the compound interest rule.

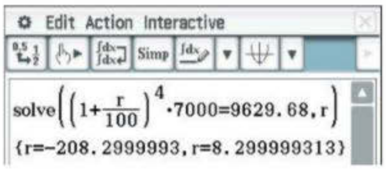
WB p. 81

| Steps | Working |
|---|---|
| **a** Soula invested \$7000 in an account earning interest compounding annually and after four years the balance is \$9629.68. What was the annual interest rate of the account to one decimal place? | |
| 1 Identify what we know and what we need to find from the compound interest rule. | $V_0 = 7000,\ n = 4,\ V_4 = 9629.68,\ r = ?$ <br> $V_n = \left(1 + \frac{r}{100}\right)^n \times V_0$ |
| 2 Substitute into the rule and solve using CAS. Ignore the negative solution. | $V_4 = \left(1 + \frac{r}{100}\right)^4 \times 7000$ <br> $= 9629.68$ |

**TI-Nspire**

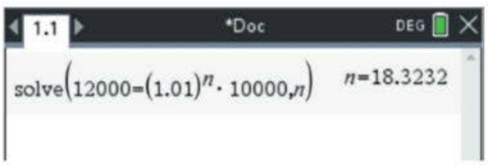



**ClassPad**

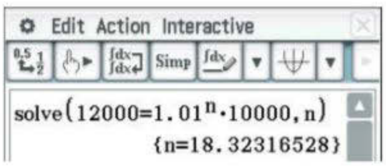



| Steps | Working |
|---|---|
| 3 Write the rounded answer. | $r = 8.3\%$ (the rate cannot be negative) <br> The interest rate was 8.3% p.a. |
| **b** Florencia invested \$10 000 in an account earning 12% interest per annum compounding monthly. How many months will it take to grow to \$12 000? | |
| 1 Identify what we know and what we need to find from the compound interest rule. | $V_0 = 10\,000,\ r = \frac{12}{12} = 1,\ V_n = 12\,000,\ n = ?$ <br> $V_n = \left(1 + \frac{r}{100}\right)^n \times V_0$ |
| 2 Substitute into the rule and solve using CAS. | $12\,000 = \left(1 + \frac{1}{100}\right)^n \times 10\,000$ <br> $12\,000 = 1.01^n \times 10\,000$ |

**TI-Nspire**




**ClassPad**




| Steps | Working |
|---|---|
| 3 Write the answer rounded *up* to the nearest month. | $n = 18.3$ <br> It will take 19 months for the investment to grow to \$12 000. |

c Jamie has a compound interest investment that earns 5% interest compounding quarterly. If the balance of Jamie's investment was \$7179.16 after two years, what amount did Jamie initially invest to the nearest dollar?

1 Identify what we know and what we need to find from the compound interest rule.

$r = \frac{5}{4}$, $n = 8$, $V_8 = 7179.16$, $V_0 = ?$

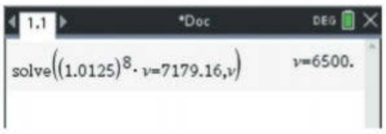
$$V_n = \left(1 + \frac{r}{100}\right)^n \times V_0$$

2 Substitute into the rule and solve using CAS.

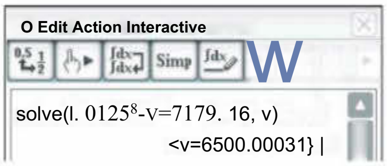
$$V_8 = \left(1 + \frac{\frac{5}{4}}{100}\right)^8 \times V_0 = 7179.16$$

$$1.0125^8 \times V_0 = 7179.16$$

TI-Nspire




ClassPad




3 Write the rounded answer.

\$6500

## EXERCISE 5.5 Compound interest

ANSWERS p. 711

### Recap

1 ©VCAA | 2005 1BRMQ4 73% A machine that makes boxes costs \$45 000. Its value depreciates by five cents for every box it makes. Each year it makes 120000 boxes. The depreciated value of this machine at the end of two years is

A \$33000 B \$38000 C \$39000 D \$45000 E \$115000

2 ©VCAA | 2006 1BRMQ5 67% A photocopier is depreciated by \$0.04 for each copy it makes. Three years ago the photocopier was purchased for \$48 000. Its depreciated value now is \$21000. The total number of copies made by the photocopier in the three years is

A 108000 B 192000 C 276000 D 525000 E 675000

### Mastery

3 WORKED EXAMPLE 13 Aisling invests \$4000 for four years and wants to compare an investment at 10% p.a. compounding yearly to 10% p.a. simple interest.

a Copy and complete the following table for $n = 3$ and $n = 4$.

| | Compound | | Simple | |
|---|---|---|---|---|
| $n$ | Interest (\$) | Value of investment (\$) | Interest (\$) | Value of investment (\$) |
| 0 | | 4000 | | 4000 |
| 1 | $\frac{10}{100} \times 4000 = 400$ | 4000 + 400 = 4400 | $\frac{10}{100} \times 4000 = 400$ | 4000 + 400 = 4400 |
| 2 | $\frac{10}{100} \times 4400 = 440$ | 4400 + 440 = 4840 | $\frac{10}{100} \times 4000 = 400$ | 4400 + 400 = 4800 |
| 3 | | | | |
| 4 | | | | |

b What is the value of the compound interest investment after four years?

c After four years how much more is the value of the compound interest investment compared with the simple interest investment?



9780170448406

4H WORKED EXAMPLE 14 I For each of the following compound interest investments, find

i the number of compounding periods per year

ii the number of compounding periods over nine years

iii the percentage interest rate per compounding period (as a fraction)

iv the amount of interest earned in the first compounding period to the nearest cent,

a Christos invests $10 000 at 5% p.a. compounding monthly.

b Jackson invests $35 000 at 3% p.a. compounding weekly.

c Fleur invests $22 000 at 8% p.a. compounding fortnightly

d Shirin invests $40000 at 7% p.a. compounding daily.

e Elsbeth invests $14000 at 6% p.a. compounding quarterly.

5.5

5 H WORKED EXAMPLE 15 J

a Write a recurrence relation for the account balance for each of the following compound interest savings accounts.

i Jed deposits $13 000 at the rate of 2.7% per annum, compounding annually.

ii Trey deposits $18000 at the rate of 3.6% per annum, compounding monthly.

iii Silvia deposits $11000 at the rate of 4.2% per annum, compounding quarterly,

b Describe the sort of growth or decay modelled by the recurrence relations.

c Sketch the shape of the graph of the recurrence relations.

6 S WORKED EXAMPLE 16 J Susie has invested her money in an account earning compound interest, compounding annually, according to the recurrence relation, where $V_n$ is the account balance after $n$ compounding periods.

$V_o = 6OOO, V_{M+1} = 1.028 V_n$

a How much money did Susie initially invest?

b Use recursion to write down calculations that show that the amount of money in Susies account after two years will be $6340.70.

c What is the annual percentage compound interest rate for this account?

d After how many years will the balance of Susies account first exceed $7000?

e What is the total interest, to the nearest cent, earned by Susies investment after six years?

7 Rj WORKED EXAMPLE 17J Michelle inherits $55000 and decides to invest it in an account that pays 9% p.a. interest compounded monthly.

a Find r, the percentage interest rate per compounding period.

b Write a rule that will calculate the value of the investment after $n$ months.

c Use the rule to find the value of the investment after five years to the nearest cent.

d Use a rule to find the value of Michelles investment after five years to the nearest cent if the interest was compounded quarterly instead of monthly.

e Which compounding period gives the larger balance after five years?

8H Using CAS 3^ Hamish has $1000 to invest and he compares two different accounts.

Account 1: simple interest at 5.8% p.a.

Account 2: compound interest at 5.4% p.a. compounded annually

a Create a graph that shows the value of each investment over a 10-year period on the same set of axes,

b What would your advice to Hamish be?

**9S** WORKED EXAMPLE 18 Find each of the following using the compound interest rule.

a Kenneth invested \$6000 in an account earning interest compounding annually and after five years the balance is \$8454.71. What was the annual interest rate of the account, correct to one decimal place?

b Stanley invested \$4000 in an account earning 6% interest compounding monthly. How many months will it take to grow to \$4500?

c Geraldine has a compound interest investment that earns 8% interest compounding quarterly. If the balance of Geraldine's investment was \$16473.43 after four years, what amount did Geraldine invest, correct to the nearest dollar?

## Exam practice

80-100% | 60-79% | 0-59%

10 ©VCAA 2O2O1CQ24~ 80% Manu invests \$3000 in an account that pays interest compounding monthly. The balance of his investment after $n$ months, $B_n$, can be determined using the recurrence relation

$B_0 = 3000, B_{n+1} = 1.0048 \times B_n$

The total interest earned by Manus investment after the first five months is closest to

A \$57.60 B \$58.02 C \$72.00 D \$72.69 E \$87.44

11 ©VCAA 20201CQ27 77% Gen invests \$10000 at an interest rate of 5.5% per annum, compounding annually. After how many years will her investment first be more than double its original value?

A 12 B 13 C 14 D 15 E 16

12 ©VCAA 66% The value of a compound interest investment, in dollars, after $n$ years, $V_n$, can be modelled by the recurrence relation

$V_0 = 100000, V_{n+1} = 1.01V_n$

The interest rate, per annum, for this investment is

A 0.01% B 0.101% C 1% D 1.01% E 101%

13 ©VCAA 20201CQ26, 20% Ray deposited \$5000 in an investment account earning interest at the rate of 3% per annum, compounding quarterly. A rule for the balance, $R_n$, in dollars, after $n$ years is given by

A $R_n = 5000 \times 0.03^n$ B $R_n = 5000 \times 1.03^n$ C $R_n = 5000 \times 0.03^{4n}$

D $R_n = 5000 \times 1.0075^n$ E $R_n = 5000 \times 1.0075^{4n}$

14 ©VCAA 2017N2CQ6 (4 marks) A community centre opened a savings account with Bank $P$. Let $P_n$ be the balance of the savings account $n$ years after it was opened. The value of $P_n$ can be determined using the recurrence relation model

$P_0 = A, P_{n+1} = 1.056 \times P_n$

The balance of the savings account one year after it was opened was \$1584.

**Exam hack**

Part a is a 'show that' question, so your answer must end with 'A = \$1500'.

a Show that the value of A is \$ 1500. 1 mark

b Write down the balance of the savings account four years after it was opened. 1 mark

c The balance of the savings account six years after it was opened was \$2080.05. This \$2080.05 was transferred into a savings account with Bank $Q$. This savings account pays interest at the rate of 5.52% per annum, compounding monthly. Let $Q_n$ be the balance of this savings account $n$ months after it was opened. The value of $Q_n$ can be determined from a rule. Copy and complete this rule by writing the missing values in the boxes. 2 marks

$Q_n = \square \times \square^n$

15 ©VCAA | 20162CQ5I (5 marks) Ken has opened a savings account to save money to buy a new caravan. The amount of money in the savings account after $n$ years, V,,, can be modelled by the recurrence relation shown below.

$V_0 = 15000, \; V_{n+1} = 1.04 \times V_n$

a 92% How much money did Ken initially deposit into the savings account? 1 mark

b 73% Use recursion to write down calculations that show that the amount of money in Kens savings account after two years, $V_2$, will be $16224. 1 mark

**© Exam hack**

When a question asks you to ‘use’ a method, there’s most likely an alternative method, so make sure you use the method you’re being asked to use.

c 65% What is the annual percentage compound interest rate for this savings account? 1 mark

d The amount of money in the account after $n$ years, $V_n$, can also be determined using a rule,

i 61% Copy and complete the rule by writing the appropriate numbers in the boxes.

$V_n = \square^n \times \square$ 1 mark

ii 47% How much money will be in Kens savings account after 10 years? 1 mark

16 ^^^J20182CQT] (5 marks) Julie deposits some money into a savings account that will pay compound interest every month. The balance of Julies account, in dollars, after $n$ months, V,,, can be modelled by the recurrence relation shown below.

$V_o = 12\,000, \; V_{n+1} = 1.0062\, V_n$

a 96% How many dollars does Julie initially invest? 1 mark

b Recursion can be used to calculate the balance of the account after one month.

i 89% Write down a calculation to show that the balance in the account after one month, Vj, is $12074.40. 1 mark

jj 82% After how many months will the balance of Julies account first exceed $12 300? 1 mark

c A rule of the form $V_n = ax\, b^n$ can be used to determine the balance of Julies account after $n$ months.

i 78% Copy and complete this rule for Julies investment after $n$ months by writing the appropriate numbers in the boxes.

balance = $\square \times \square^n$ 1 mark

ii 65% What would be the value of $n$ if Julie wanted to determine the value of her investment after three years? 1 mark

# 5.6 Effective interest rates

Video playlist Effective interest rates

Worksheet Effective interest rates

## Nominal vs effective interest rates

When deciding on an investment or loan, its important to be able to compare interest rates. The compound interest rates weve been dealing with so far are called nominal interest rates. They are given as a rate per year plus a compounding period.

If both the rates and the compounding periods are different, comparing rates becomes difficult. That’s when we calculate effective interest rates. These are interest rates after compounding has been taken into account.

- For an investment, the highest effective interest rate is the best option.
- For a loan, the lowest effective interest rate is the best option.

**Effective interest rate formula**

$$r_{effective} = \left[\left(1 + \frac{r}{100n}\right)^n - 1\right] \times 100\%$$

where

$r$ = the nominal interest rate *per year*

$n$ = the number of compounding periods *per year*.

**GF Exam hack**

The $r$ in the effective interest rate formula is per year, *not* per compounding period. The $n$ in the effective interest rate formula is also per year, *not* per compounding period. This formula appears on the formula sheet for the exam.

WB
p. 83

**WORKED EXAMPLE 19** Finding effective interest rates

Emma is looking to invest her money. She has done some research on interest rates and found the best offers from four different banks.

**Bank 1:** 8.45% p.a. compounding daily
**Bank 2:** 8.6% p.a. compounding monthly
**Bank 3:** 8.7% p.a. compounding six-monthly
**Bank 4:** 8.8% p.a. compounding annually

**a** Find the effective interest rate for each bank, rounding to two decimal places.

**b** Which bank should Emma choose if she wants to earn the most interest?

**c** Which bank would earn Emma the least interest?

**d** Why are the nominal and effective interest rates for Bank 4 the same?

| Steps | Working |
|---|---|
| **a** For each option, substitute the known variables into the effective interest rate formula and solve, rounding to two decimal places. $r_{effective} = \left[\left(1 + \frac{r}{100n}\right)^n - 1\right] \times 100\%$ | **Bank 1** $r_{effective} = ?, r = 8.45, n = 365$ $r_{effective} = \left[\left(1 + \frac{8.45}{36500}\right)^{365} - 1\right] \times 100\% = 8.82\%$ p.a. **Bank 2** $r_{effective} = ?, r = 8.6, n = 12$ $r_{effective} = \left[\left(1 + \frac{8.6}{1200}\right)^{12} - 1\right] \times 100\% = 8.95\%$ p.a. **Bank 3** $r_{effective} = ?, r = 8.7, n = 2$ $r_{effective} = \left[\left(1 + \frac{8.7}{200}\right)^{2} - 1\right] \times 100\% = 8.89\%$ p.a. **Bank 4** $r_{effective} = ?, r = 8.8, n = 1$ $r_{effective} = \left[\left(1 + \frac{8.8}{100}\right)^{1} - 1\right] \times 100\% = 8.8\%$ p.a. |
| **b** Compare the four results and choose the largest. | Emma should choose Bank 2 because it pays the higher effective rate of interest and will therefore pay more interest. |
| **c** Compare the four results and choose the smallest. | Bank 4 would earn Emma the least interest. |
| **d** Compare the nominal and effective interest rates. | The nominal and effective interest rates for Bank 4 are the same because the rate compounds annually. |

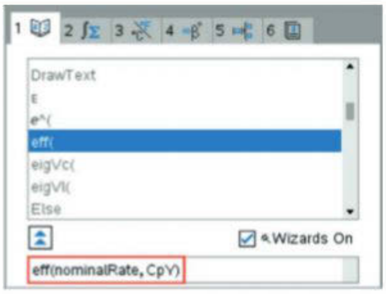
**Exam hack**

If a compound interest rate compounds annually, the effective interest rate is always the same as the nominal interest rate.

## USING CAS 4 Finding effective interest rates

Determine the effective interest rate for an investment that offers 3.6% interest compounding fortnightly, giving your answer correct to two decimal places.

**TI-Nspire**


1 Press menu > Finance > Interest Conversion > Effective Interest Rate.

2 For the first parameter, enter 3.6, which is the nominal interest rate.

3 For the second parameter, enter 26, which is the number of compounds per year.

4 Press enter.

5 The effective rate of interest of 3.66% will be displayed.

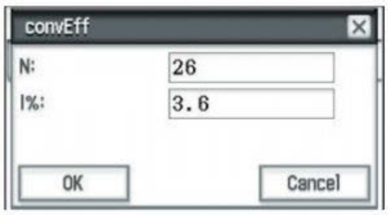
6 Alternatively, press catalog > E then scroll down to the eff function to display the parameters for the effective rate of interest function, shown in the red box above.

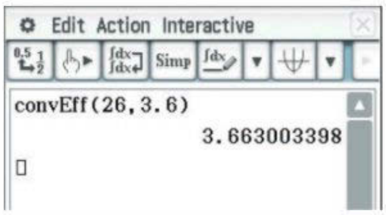
7 The two parameters are nominalRate and CpY.

8 Press enter to display the function in the Calculator page.

9 Enter the parameters and press enter.

**ClassPad**




1 Tap Interactive > Financial > Interest Conversion > ConvEff.

2 In the N: field, enter 26, which is the number of compounds per year.

3 In the 1%: field, enter 3.6, which is the nominal rate of interest.

4 Tap OK.

5 The effective rate of interest of 3.66% will be displayed.

## EXERCISE 5.6 Effective interest rates

ANSWERS p. 711

### Recap

1 ©VCAA | 2017N ICQIFI Andre deposited $20000 into a savings account earning compound interest at the rate of 3.1% per annum, compounding annually. Which one of the following recurrence relations can be used to determine the amount in the savings account, $S_w$, after *n* years?

A $S_o = 20000$, $S_{,,+1}$-$S_{,, +}$ 620

B $S_o$-2OOOO, $S_{n+1}$ - 1.031 x *S,,*

C $S_o = 20000$, $S_{n+}$i = 620xS,,

D $S_o$-2OOOO, $S_{M+1}$ = 3.1 x $S_n$ + 620

E $S_o = 20000$, $S_{,,+1}$ = $S_{n\ +}$3.1x620

▶ 2 $5000 is invested at 6% p.a. compounding yearly. Which of the following is *not* true, given $V_n$ is the value of the investment after *n* years?

A The recurrence relation that can be used to model this is $V_0 = 5000$, $V_{n+1} = 1.06V_n$

B The rule for finding the value after the nth year is $V_n = 1.06^n V_0$

C $V_5 = 1.06^5 \times 5000$

D The interest earned in the seventh year is $V_7 - V_6$

E $V_8 = 1.06 \times V_9$

**Mastery**

30 WORKED EXAMPLE 19 Georgio is looking to invest his money. He has done some research on interest rates and found the best offers from four different banks.

Bank 1: 7.4% p.a. compounding weekly
Bank 2: 7.5% p.a. compounding fortnightly
Bank 3: 7.6% p.a. compounding six-monthly
Bank 4: 7.7% p.a. compounding annually

a Find the effective interest rate for each bank, rounding to two decimal places.

b Which bank should Georgio choose if he wants to earn the most interest?

c Which bank would earn Georgio the least interest?

d What do you notice about the nominal and effective interest rates for Bank 4? Why is this the case?

4 S using CAS 4] For each of the following investments determine the effective interest rate, giving your answer correct to two decimal places.

a 9% p.a. compounding monthly

b 11% p.a. compounding weekly

c 12% p.a. compounding six-monthly

d 6% p.a. compounding daily

**Exam practice** 80-100% 60-79% 0-59%

5 ©VCAA 2016SICQ19 Eva has $1200 that she plans to invest for one year. One company offers to pay her interest at the rate of 6.75% per annum compounding daily. The effective annual interest rate for this investment would be closest to

A 6.75% B 6.92% C 6.96% D 6.98% E 6.99%

6 ©VCAA 2018N1CQ21 An amount of money is deposited into an account that earns compound interest. Which combination of interest rate and compounding period has the largest effective interest rate?

A 3.7% per annum, compounding weekly

B 3.7% per annum, compounding monthly

C 3.7% per annum, compounding quarterly

D 3.8% per annum, compounding monthly

E 3.8% per annum, compounding quarterly

7 Which has the lowest effective interest rate?

A 7.9% p.a. compounding quarterly

B 7.7% p.a. compounding fortnightly

C 7.5% p.a. compounding monthly

D 7.45% p.a. compounding daily

E 7.58% p.a. compounding six-monthly

8 Murray has chosen an investment with an interest rate of 5.4% p.a. compounded quarterly. What is the difference between the nominal interest rate and the effective interest rate?

A 0.1% B 0.3% C 5.4% D 5.5% E 5.6%

9 ©VCAA J 20201CQ28 58% The nominal interest rate for a loan is 8% per annum. When rounded to two decimal places, the effective interest rate for this loan is not

A 8.33% per annum when interest is charged daily.

B 8.32% per annum when interest is charged weekly.

C 8.31% per annum when interest is charged fortnightly.

D 8.30% per annum when interest is charged monthly.

E 8.24% per annum when interest is charged quarterly.

10 ©VCAA | 2018ICQ19J 54% Daniel borrows $5000, which he intends to repay fully in a lump sum after one year. The annual interest rate and compounding period for five different compound interest loans are given below:

- Loan I - 12.6% per annum, compounding weekly
- Loan II - 12.8% per annum, compounding weekly
- Loan III - 12.9% per annum, compounding weekly
- Loan IV - 12.7% per annum, compounding quarterly
- Loan V - 13.2% per annum, compounding quarterly

When fully repaid, the loan that will cost Daniel the least amount of money is

A Loan I. B Loan II. C Loan III. D Loan IV. E Loan V.

11 (8 marks) Jillian is choosing between the following investment options:

Bank of Victoria: 7.8% p.a. compounding quarterly
Power Bank: 8% p.a. compounding yearly
Aussie Bank: 8% p.a. compounding half-yearly

a What nominal interest rate is the Bank of Victoria offering? 1 mark

b Does Power Bank or Aussie Bank offer the higher effective interest rate? Explain why you don't need to do a calculation to decide. 2 marks

c What is the effective interest rate for Power Bank? Explain why you don't need to do a calculation. 2 marks

d Calculate the effective interest rates for Bank of Victoria and Aussie Bank, correct to two decimal places. Which should Jillian choose? 3 marks

# @ Reducing balance depreciation

Video playlist Reducing balance depreciation

## Reducing balance depreciation recurrence relations and graphs

We have previously looked at flat rate and unit cost depreciation. The third type of depreciation is reducing balance depreciation, which calculates the value of an asset by reducing it every year by a fixed percentage *of its value in the preceding year.*

Differences between flat rate and reducing balance depreciation

| Flat rate depreciation | Reducing balance depreciation |
|---|---|
| Similar to a *simple* interest investment except we *subtract* the interest amount. | Similar to *compound* interest investment except we *subtract* the interest amount. |
| Asset depreciates by the *same* amount each year. | Asset depreciates by *smaller amounts* each year. |
| Is an example of *linear* decay. | Is an example of *geometric* decay. |

## Reducing balance depreciation recurrence relation

The recurrence relation for the value of an asset $V_n$ being depreciated using reducing balance depreciation is

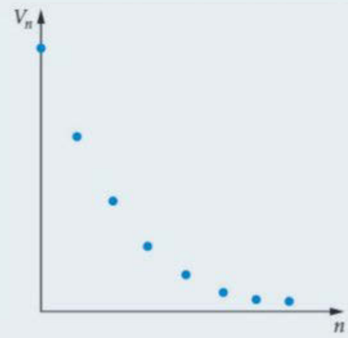
$$V_n = \text{initial value of the asset, } V_{n+1} = \left(1 - \frac{r}{100}\right) \times V_n$$

where

$r$ = the percentage depreciation rate per year

$n$ = the number of years.

The graph of a reducing balance depreciation recurrence relation would look like this.




$V_n - V_o$ = total amount of depreciation after $n$ years

### Exam hack

The main difference between reducing balance depreciation and compound interest investment is we subtract $\frac{r}{100}$ rather than add it. Also, we always depreciate per year, so we don't have to worry about different compounding periods.

WB

p. 84

## WORKED EXAMPLE 20 Using reducing balance depreciation recurrence relations

A business purchased a photocopier for \$10000. It is depreciated using reducing balance depreciation at a rate of 18% per annum. Give all answers to the nearest dollar.

**Steps** | **Working**

a Copy and complete the table to find
- i the value of the photocopier after five years
- ii the amount of depreciation in the third year
- iii when the photocopier first depreciates to under \$5000.

| $n$ | Depreciation after $n$ years (\$) | Value after $n$ years (\$) |
|---|---|---|
| 0 | | 10000 |
| 1 | $\frac{18}{100} \times 10000 = 1800$ | 10000 - 1800 = 8200 |
| 2 | $\frac{18}{100} \times 8200 = 1476$ | 8200 - 1476 = 6724 |
| 3 | | |
| 4 | | |
| 5 | | |

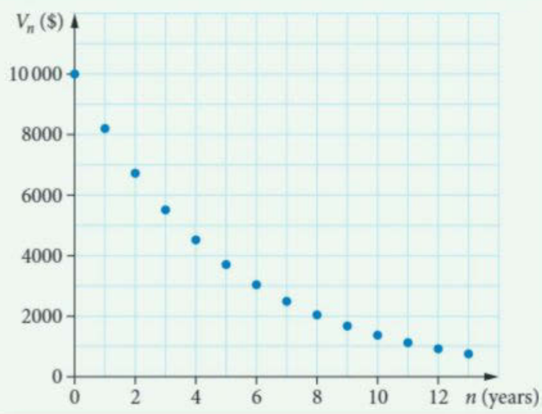
Calculate the percentage of successive values and subtract from the previous value.

Use CAS's recursive computation where possible.

Give all values to the nearest dollar, but don't round until after all the calculations have been done.

(Note: Answers can vary slightly depending on when values are rounded.)

| $n$ | Depreciation after $n$ years (\$) | Value after $n$ years (\$) |
|---|---|---|
| 0 | | 10000 |
| 1 | $\frac{18}{100} \times 10000 = 1800$ | 10000 - 1800 = 8200 |
| 2 | $\frac{18}{100} \times 8200 = 1476$ | 8200 - 1476 = 6724 |
| 3 | $\frac{18}{100} \times 6724 = 1210$ | 6724 - 1210 = 5514 |
| 4 | $\frac{18}{100} \times 5514 = 993$ | 5514 - 993 = 4521 |
| 5 | $\frac{18}{100} \times 4521 = 814$ | 4521 - 814 = 3707 |

| | |
|---|---|
| i Read the answer from the table. | The value of the photocopier after five years is \$3707. |
| ii Read the answer from the table. | The amount of depreciation in the third year is \$1210. |
| iii Read the answer from the table. | The photocopier first depreciates to under \$5000 after four years. |
| **b** Write down a recurrence relation that gives the value of the photocopier after $n$ years. | |
| 1 Identify $V_M$, $V_o$ and $r$. | Let $V_n$ = the value of the photocopier after $n$ years. <br> $V_o$= 10000, $r$= 18 |
| 2 Substitute the values into <br> $V_o$ = initial value of the asset, <br> $V_{n+1} = \left(1 - \frac{r}{100}\right)V_n$ <br> and simplify. | $V_o$- 10000, $V_{n+1} = \left(1 - \frac{18}{100}\right)V_n$ <br> $V_o$= 10000, $V_{n+1} = (1 - 0.18) V_n$ <br> $V_o$= 10000, $V_{n+1} = 0.82V_n$ |
| **c** What percentage of the previous value is each new value? | |
| Look at the decimal in front of $V_n$ and convert it to a percentage. | The value of the photocopier in any year is 82% of its value the previous year. |
| **d** Describe the sort of growth or decay modelled by the recurrence relation. | |
| Is addition or subtraction involved? <br> Is multiplication by a number greater than 1 or between 0 and 1 involved? | No addition or subtraction is involved. <br> Multiplication by a number between 0 and 1 is involved. <br> So, the recurrence relation models geometric decay. |
| **e** Use the graph to find the photocopier s' approximate value after eight years. |   |
| Read from the graph. | The value of the photocopier after eight years is around \$2000. |

WB

p. 86

## WORKED EXAMPLE 21 Working with reducing balance depreciation recurrence relations

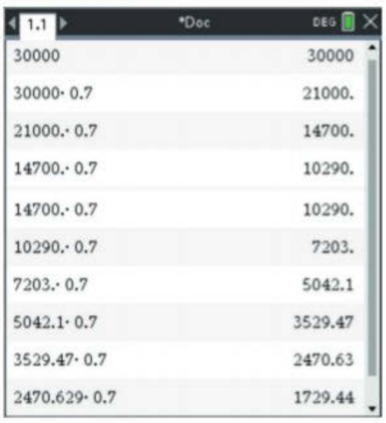
Herman the Handyman depreciates his power tools using the reducing balance method. The value of the tools, in dollars, after *n* years, can be modelled by the recurrence relation

Vo =30000, $V_{n+1} = 0.7V_n$

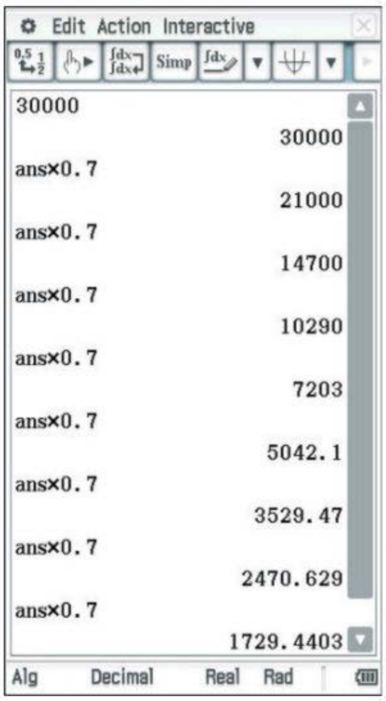
| **Steps** | **Working** |
|---|---|
| a Use recursion to show that the value of the tools after two years, $V_2$, is $14 700. | |
| Step out the recurrence relation working to find $V_2$. | $V_o = 30000$<br>^ = 0.7^ = 0.7 x 30000 = 21000<br>$V_2$ = 0.7VJ = 0.7x21000 = 14700 |
| b What is the annual percentage rate of depreciation used by Herman? | |
| 1 Use the recurrence relation to find an equation for r. | Comparing:<br>$V_{n+1} = \left(1 - \frac{r}{100}\right)V_n$<br>$V_{n+1} = 0.7V_n$<br>we can see that<br>$1 - \frac{r}{100} = 0.7$ |
| 2 Solve for r, using CAS if necessary. | $\frac{r}{100} = 1 - 0.7$<br>$\frac{r}{100} = 0.3$<br>r = 30 |
| 3 Write the answer. | The annual percentage rate of depreciation is 30%. |
| c If Herman plans to replace these tools when their value first falls below $2000, after how many years will Herman replace these tools? | |

1 Use CAS recursive computation to continue the recurrence relation calculations until they are less than the given amount.

**TI-Nspire**




**ClassPad**




2 Write the answer. Herman will replace these tools after eight years.

## Reducing balance depreciation general rule

**Reducing balance depreciation general rule**

The general rule for the value $V_n$ of a depreciated asset using reducing balance depreciation is

$$V_n = \left(1 - \frac{r}{100}\right)^n \times V_0$$

where

$V_0$ = initial value of the asset

$r$ = the percentage depreciation rate per year

$n$ = the number of years.

When solving for $n$, always round *up*, never down, to the nearest whole number.

$V_{n-1} - V_n$ = amount of depreciation in the $n$th year

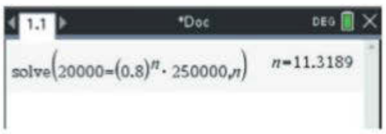
**WORKED EXAMPLE 22** Using the reducing balance depreciation rule

WB p. 87

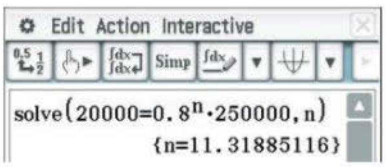
A truck bought for $250 000 is depreciated using the reducing balance method at a rate of 20% p.a.

| Steps | Working |
|---|---|
| **a** Write a rule that will calculate the value of the truck after $n$ years. | |
| Substitute the values of $V_0$ and $r$ into the reducing balance depreciation general rule and simplify. | $V_0 = 250\,000$, $r = 20$<br>$V_n = \left(1 - \frac{r}{100}\right)^n \times V_0$<br>$V_n = \left(1 - \frac{20}{100}\right)^n \times 250\,000$<br>$V_n = 0.8^n \times 250\,000$ |
| **b** Use the rule to find the value of the truck after nine years to the nearest dollar. | |
| Substitute the value of $n$ into your rule and solve. | $n = 9$<br>$V_9 = 0.8^9 \times 250\,000$<br>$= 33\,554.43\ldots$<br>The value of the truck after nine years is $33 554. |
| **c** Use the rule to find how many years it would take for the truck to depreciate to under $20 000. | |
| **1** Identify what we know and what we need to find from the reducing balance depreciation rule. | $V_0 = 250\,000$, $r = 20$, $V_n = 20\,000$, $n = ?$<br>$V_n = \left(1 - \frac{r}{100}\right)^n \times V_0$ |
| **2** Substitute into the rule and solve using CAS. | $20\,000 = 0.8^n \times 250\,000$ |
| **3** Write the answer, rounding *up* to the nearest year. | It will take the truck 12 years to depreciate to under $20 000. |

TI-Nspire




ClassPad



d How much is the truck depreciated by in the sixth year?

Amount of depreciation in the nth year $= V_{n-1} - V_n$

$n = 6$

$V_5 = 0.8^5 \text{ X } 250000 = 81920$

$V_6 = 0.8^6 \text{ x } 250\,000 = 65\,536$

$V_5 - V_6 = 81920 - 65\,536 = 16\,384$

The amount of depreciation in the sixth year is $16384.

WB p. 88

## WORKED EXAMPLE 23 Working with the reducing balance depreciation rule

Find each of the following using the reducing balance depreciation rule.

**Steps** | **Working**

**a** Saskia bought a car for $28 000. After two years of being depreciated on a reducing balance basis, the car is now valued at $20 230. Show that the annual rate of depreciation in the value of the car is 15%.

**1** Identify what we know and what we need to find from the reducing balance depreciation rule.

$V_0 = 28\,000,\ n = 2,\ V_2 = 20\,230,\ r = ?$

$$V_n = \left(1 - \frac{r}{100}\right)^n \times V_0$$

**2** Substitute into the rule and show the steps to solve for $r$.

If this wasn't a 'show' question, we could use CAS.

$$V_2 = \left(1 - \frac{r}{100}\right)^2 \times 28\,000 = 20\,230$$

$$\left(1 - \frac{r}{100}\right)^2 = \frac{20\,230}{28\,000} = 0.7225$$

$$1 - \frac{r}{100} = \sqrt{0.7225} = 0.85$$

$$\frac{r}{100} = 1 - 0.85 = 0.15$$

$$r = 15\%$$

**b** Lloyd is using the reducing balance method to depreciate the carpet in his investment property. The annual rate of depreciation in the value of the carpet is 3%. If after 10 years the value of the carpet is $20 647.88, what was the original price of the carpet to the nearest dollar?

**1** Identify what we know and what we need to find from the reducing balance depreciation rule.

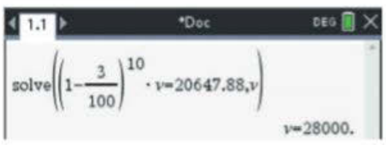
$r = 3,\ n = 10,\ V_{10} = 20\,647.88,\ V_0 = ?$

$$V_n = \left(1 - \frac{r}{100}\right)^n \times V_0$$

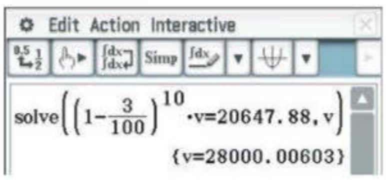
**2** Substitute into the rule and solve using CAS.

$$V_{10} = \left(1 - \frac{3}{100}\right)^{10} \times V_0 = 20\,647.88$$

**TI-Nspire**




**ClassPad**




**3** Write the rounded answer.

The original price of the carpet is $28 000.



9780170448406

## VCE QUESTION ANALYSIS

©VCAA 2017 2CQ6 2017 Examination 2 Core Question 6 (4 marks)

Alex the mobile mechanic sends a bill to his customers after repairs are completed. If a customer does not pay the bill by the due date, interest is charged. Alex charges interest after the due date at the rate of 1.5% per month on the amount of an unpaid bill. The interest on this amount will compound monthly,

a Alex sent Marcus a bill of $200 for repairs to his car. Marcus paid the full amount one month after the due date. How much did Marcus pay? 1 mark

Alex sent Lily a bill of $428 for repairs to her car. Lily did not pay the bill by the due date. Let $A_n$ be the amount of this bill $n$ months after the due date.

b Write down a recurrence relation, in terms of $A_0$, $A_{n+1}$ and $A_n$, that models the amount of the bill. 2 marks

c Lily paid the full amount of her bill four months after the due date. How much interest was Lily charged? Round your answer to the nearest cent. 1 mark

5.7

Video VCE question analysis: Interest and depreciation

### Reading the question

- Always note the compounding period.
- Highlight anything indicated with the words 'in terms of'. You must use these in your answer.
- Answers involving money generally ask for rounding to the nearest dollar or cent.

### Thinking about the question

- Be clear on what sort of financial situation this is. This is a compound interest problem.
- Make sure you know when to use the recurrence relation, when to use the general rule, and when to use a CAS.

### Worked solution (✓ = 1 mark)

a Use $V_n = \left(1 + \frac{r}{100}\right)^n \times V_0$, the compound interest rule where $n = 1$, $r = 1.5$, $V_0 = 200$.

$V_1 = \left(1 + \frac{1.5}{100}\right)^1 \times 200 = 1.015 \times 200 = 203$

Marcus pays $203. ✓

b Use $A_0$ = principal, $A_{n+1} = \left(1 + \frac{r}{100}\right)A_n$, the compound interest recurrence relation where

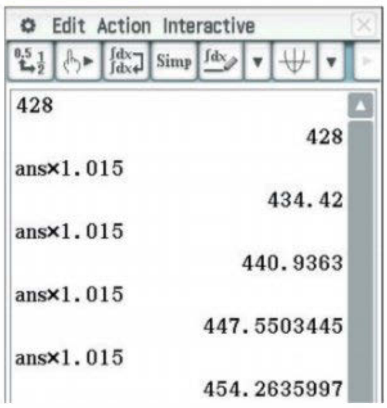
$r = 1.5$, $A_0 = 428$

$A_0 = 428$ ✓, $A_{n+1} = 1.015\,A_n$ ✓

c Use CAS recursive computation to find the value of the bill after 4 months.

TI-Nspire


ClassPad



The full amount of Lily's bill four months after the due date is \$454.26.

Use total amount of interest earned after $n$ compounding periods = $V_n - V_o$

$V_4 - V_o = 454.26 - 428 = 26.26$

Lily was charged S26.26 interest. /

**Student performance** so-100% eo-79% 0-59%

a 61% Many students read the interest rate as per annum, resulting in the incorrect calculation:

$200 \times \left(1 + \frac{1.5}{1200}\right) = \$200.25$. Another incorrect calculation was 200 x 1.5 = \$300.

b 47% Many students did not write a recurrence relation in the required form. Mistakes included not writing the $A_o$ value, not consistently using the same variable throughout the recurrence relation, and not recognising the difference between the recurrence relation and the rule $A_n = 428 \times 1.015$".

c 26% Many students gave the total amount Lily was charged (\$454.26) as the answer here, rather than doing the extra step of calculating the interest. Some rounded incorrectly to the nearest 5 or 10 cents.

## EXERCISE 5.7 Reducing balance depreciation

ANSWERS p. 712

### Recap

1 Shari wants to invest \$13 000. Which of the following nominal interest rates will give the highest effective interest rate?

A 4.9% per annum, compounding quarterly

B 4.9% per annum, compounding weekly

C 4.9% per annum, compounding yearly

D 4.9% per annum, compounding monthly

E 4.9% per annum, compounding daily

2 A bank offers a compound interest rate of 8% p. a. compounding weekly. Which of the following is closest to the difference between this rate and the effective interest rate?

A 0.08% B 0.15% C 0.3% D 5.2% E 8.3%

### Mastery

3 H WORKED EXAMPLE 20 A business purchased a workstation for \$12000. It is depreciated using reducing balance depreciation at a rate of 16% per annum. Give all answers to the nearest dollar.

a Copy and complete the table to find

i the value of the workstation after five years

ii the amount of depreciation in the fourth year

iii when the workstation first depreciates to less than \$8000.

| $n$ | Depreciation after $n$ years (\$) | Value after $n$ years (\$) |
|---|---|---|
| 0 | | 12 000 |
| 1 | $\frac{16}{100} \times 12\,000 = 1920$ | 12 000 – 1920 = 10 080 |
| 2 | $\frac{16}{100} \times 10\,080 = 1613$ | 10 080 – 1613 = 8467 |
| 3 | | |
| 4 | | |
| 5 | | |



9780170448406

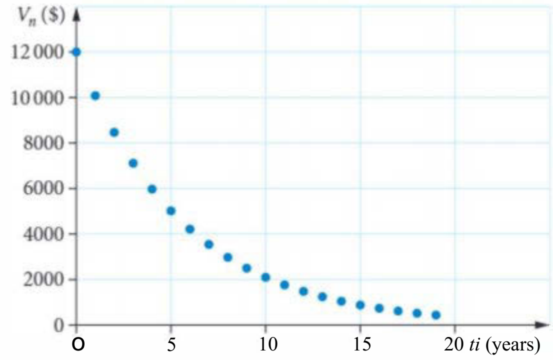
5.7

b Write down a recurrence relation that gives the value of the workstation after $n$ years,

c What percentage of the previous value is each new value?

d Describe the sort of growth or decay modelled by the recurrence relation,

e Use the graph to find the workstations approximate value after 14 years.




4 EJ WORKED EXAMPLE 21J Melanie runs a mowing service and she depreciates her ride-on mower using the reducing balance method. The value of the mower, in dollars, after $n$ years, $V_n$, can be modelled by the recurrence relation:

$V_o = 25000, V_{n+1} = 0.6V_n$,

a Use recursion to show that the value of the mower after two years, $V_2$, is $9000.

b What is the annual percentage rate of depreciation used by Melanie?

c If Melanie plans to replace the mower when its value first falls below $2000, after how many years will Melanie replace the mower?

5 s WORKED EXAMPLE 22 I A truck was bought for $200 000 and is being depreciated on a reducing balance basis rate of 25% per annum.

a Write a rule that will calculate the value of the truck after $n$ years.

b Use the rule to find the value of the truck after eight years to the nearest dollar.

c Use the rule to find how many years it would take for the truck to depreciate to under $40 000.

d How much is the truck depreciated by in the seventh year?

6 Ej WORKED EXAMPLE 23J Find each of the following using the reducing balance depreciation rule.

a Spencer bought a sports car for $64000. After two years of depreciation on a reducing balance basis, the car is now valued at $40 960. Show that the annual rate of depreciation in the value of the car is 20%.

b Talia’s Tree Lopping is using the reducing balance method to depreciate a mulcher. The annual rate of depreciation in the value of the mulcher is 16%. If after eight years the value of the mulcher is $1735.13, what was the original price of the mulcher to the nearest dollar?

## Exam practice

80-100% 60-79% 0-59%

7 A computer purchased for $4000 is depreciated using reducing balance depreciation at a rate of 15% per annum. What are the two values for $n = 2$ in the depreciation table?

| $n$ | Depreciation after $n$ years ($) | Value after $n$ years ($) |
|---|---|---|
| 0 | | 4000 |
| 1 | | |
| 2 | | |

A $\frac{15}{100}$ and 3400 B 15 and 4000 C 510 and 2890 D 600 and 510 E 600 and 3400

▶ 8 Which of the following recurrence relations best models a car purchased at $32 000 using reducing balance depreciation at a rate of 11% per annum?

A $V_{,,}$- 0.11" x 32000 B $V_{,,}$ = 0.89" x 32000

C $V_o$-32000, $V_{,,+1}$ = 0.11V,, D $V_o$-32000, $V_{,,+1}$-0.89V,,

E VQ-32000, V,,-0.89$V_{,,+1}$

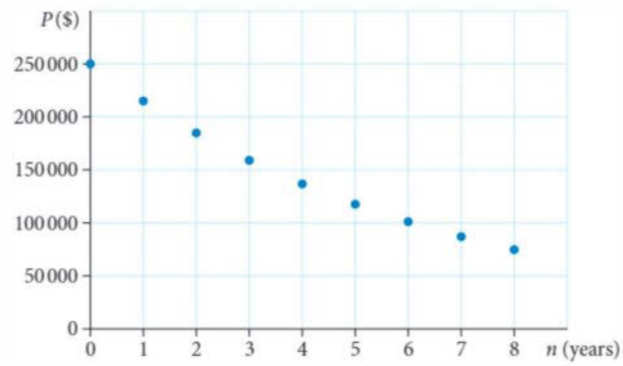
9 © VCAA 2019N1 CQI8 j A truck was purchased for $134000. Using the reducing balance method, the value of the truck is depreciated by 8.5% each year. Which one of the following recurrence relations could be used to determine the value of the truck after *n* years, V,,?

A $V_o$- 134000, $V_{,,+1}$ -0.915 x $V_{,,}$ B $V_o$- 134000, $V_{,,+1}$ - 1.085 x $V_{,,}$

C $V_o$= 134000, $V_{,,+1}$ = V,,-11390 D $V_o$- 134000, $V_{,,+1}$ = 0.915 x V,, - 8576

E $V_o$ - 134 000, $V_{n+}$, = 1.085 x V,, - 8576

10 © VCAA 2016S 1CQ24 The following graph shows the decreasing value of an asset over eight years.



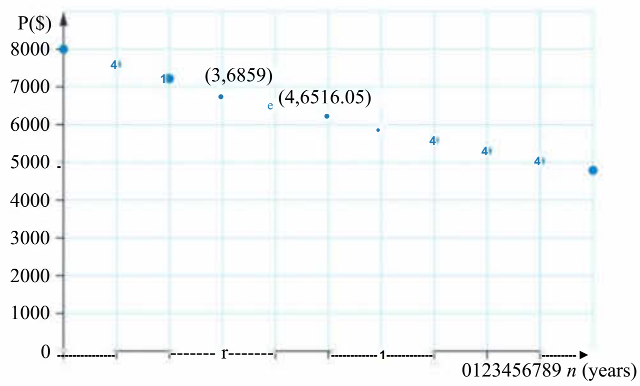

Let $P_n$ be the value of the asset after *n* years, in dollars. A rule for evaluating $P_n$ could be

A $P_n$ = 250000 x (1 +0.14)" B $P_n$ = 250000 x 1.14 x *n*

C $P_n$ = 250000 x (1 — 0.14) *xn* D $P_n$ = 250 000 x (0.14)"

E $P_n$ = 250000 x (1 -0.14)"

11 © VCAA 2017N1CQ23 J The value of a piano is depreciated using the reducing balance method. The graph shows the value of the piano as it depreciates over a period of 10 years.




Let $P_n$ be the value of the piano after *n* years. A recurrence relation that could be used to determine $P_n$ is

A $P_o$ = 8OOO, $P_{M+1}$ = 0.95 x $P_n$ B $P_o$ = 8000, $P_{,,+1}$ = 342.95 x P,,

C $P_o$ = 8000, $P_{M+1}$ = 1.05 x P,, - 3 D $P_o$ = 8OOO, $P_{w+1}$ = 0.95 xP,,-342.95

E $P_o$ = 8000, $P_{M+1}$ = $P_n$ - 342.95

▶

5.7

▶ 12 ©VCAA | 2016ICQ19 J 74% The purchase price of a car was \$26 000. Using the reducing balance method, the value of the car is depreciated by 8% each year. A recurrence relation that can be used to determine the value of the car after $n$ years, $C_n$, is

A $C_0 = 26000, \quad C_{n+1} = 0.92C_n$

B $C_0 = 26000, \quad C_{n+1} = 1.08C_n$

C $C_0 = 26000, \quad C_{n+1} = C_n + 8$

D $C_0 = 26000, C_{n+1} = C_n - 8$

E $C_0 = 26000, \quad C_{n+1} = 0.92C_n - 8$

13 ©VCAA 2019 2CQ7a-c J (3 marks) Phil is a builder who has purchased a large set of tools. The value of Phil's tools is depreciated using the reducing balance method. The value of the tools, in dollars, after $n$ years, $V_n$, can be modelled by the recurrence relation shown below.

$V_0 = 60000, V_{n+1} = 0.9V_n$

a 74% Use recursion to show that the value of the tools after two years, $V_2$, is \$48 600. 1 mark

b 66% What is the annual percentage rate of depreciation used by Phil? 1 mark

c 79% Phil plans to replace these tools when their value first falls below \$20000. After how many years will Phil replace these tools? 1 mark

14 ©VCAA 2017 2CQ5c J 58% (1 mark) Alex is a mobile mechanic. The value of his van is depreciated using the reducing balance method of depreciation. The value of the van, in dollars, after $n$ years, $R_n$, can be modelled by the recurrence relation

$R_0 = 75000, R_{n+1} = 0.943R_n$.

At what annual percentage rate is the value of the van depreciated each year?

15 ©VCAA | 2018 2CQ5 J (3 marks) Julie withdraws \$14000 from her account to purchase a car for her business. For tax purposes, she plans to depreciate the value of her car using the reducing balance method. The value of Julie's car, in dollars, after $n$ years, $C_n$, can be modelled by the recurrence relation shown below.

$C_0 = 14000, C_{n+1} = R \times C_n$

a 55% For each of the first three years of reducing balance depreciation, the value of R is 0.85. What is the annual rate of depreciation in the value of the car during these three years? 1 mark

b 41% For the next five years of reducing balance depreciation, the annual rate of depreciation in the value of the car is changed to 8.6%. What is the value of the car eight years after it was purchased? Round your answer to the nearest cent. 2 marks

**Exam hack**

Watch out for the word 'next' in questions. This means there are two stages to look at.

16 ©VCAA | 2017N 2CQ5d NTJ (2 marks) A snooker table at a community centre was purchased for \$3000. The value of the snooker table was depreciated using the reducing balance method of depreciation. After one year, the value of the snooker table is \$2760. After two years, the value of the snooker table is \$2539.20

a Show that the annual rate of depreciation in the value of the snooker table is 8%. 1 mark

b Let $S_n$ be the value of the snooker table after $n$ years. Write down a recurrence relation, in terms of $S_{n+1}$ and $S_n$, that can be used to determine the value of the snooker table after $n$ years using this reducing balance method. 1 mark

# (7) Chapter summary

## Recurrence relations

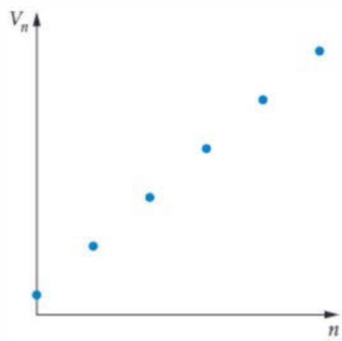
- A recurrence relation is a rule that generates a sequence by connecting each value to previous values.
- It consists of the starting value, e.g. $u_0 = 3$.
- A rule links each value to the one before it, e.g. $u$ $_{,,+1}$ = «„+ 5.
- $U_Q$ (not iq) is always the starting value.

## Graphs of recurrence relations

### Linear growth

- Increases by same amount each time.
- Points in an increasing straight line.
- Addition is involved.
- No multiplication is involved.

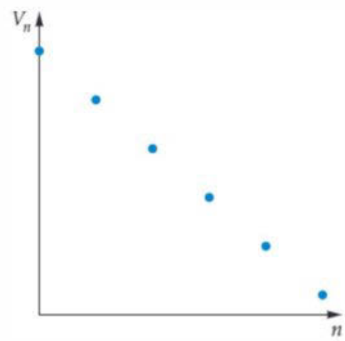
  e.g. $V_o = 2$, $V_{,,+1} = V_{,,} + 4$




### Linear decay

- Decreases by same amount each time.
- Points in a decreasing straight line.
- Subtraction is involved.
- No multiplication is involved.

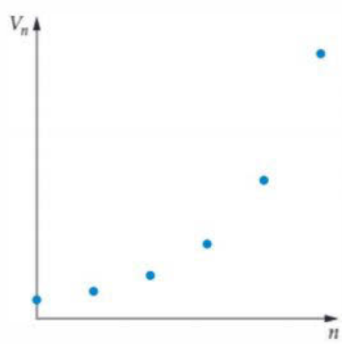
  e.g. $V_o = 22$, $V_{n+1}$ - $V_{,,}$ - 4




### Geometric growth

- Increases get larger each time.
- Points in an increasing curve.
- No addition or subtraction is involved.
- $V_n$ multiplied by a number greater than 1

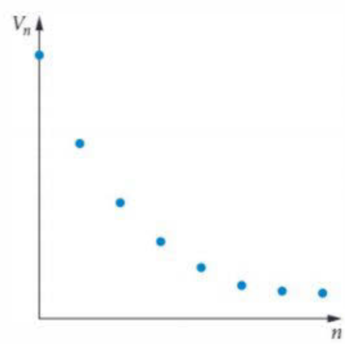
  e.g. VQ-35, $V_{n+1}$-1.62$V_{,,}$




### Geometric decay

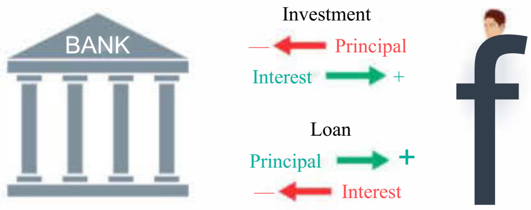
- Decreases get smaller each time.
- Points in a decreasing curve that never reaches zero.
- No addition or subtraction is involved.
- $V_n$ multiplied by a number between 0 and 1

  e.g. $V_o = 350$, $V_{n+1} = 0.62V_{,,}$



## Simple interest

- Interest is the fee for using someone elses money.
- Simple interest is a fixed amount of interest that is paid at regular time periods.
- When these time periods are years, we use the term per annum (or p.a.) which means per year.
- The amount that is borrowed or invested and is called the principal.
- The value of the investment or loan at any time is called the balance.




## Compound interest

- Compound interest involves adding interest to the principal and calculating the interest for the next time period using this new amount.
- The time period when the interest is added is called a compounding period.
- Compound interest is always given as a rate per year, but compounding periods can vary:

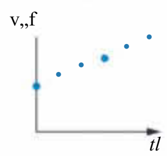
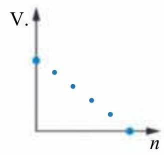
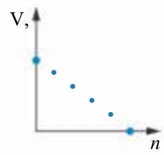
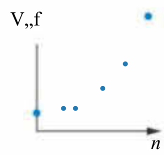
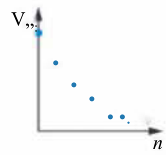
| Compounding period | Number of compounding periods per year |
|---|---|
| Daily | 365 |
| Weekly | 52 |
| Fortnightly | 26 |
| Monthly | 12 |
| Quarterly | 4 |
| Six-monthly | 2 |
| Yearly | 1 |

- percentage interest rate per compounding period = $\dfrac{\text{percentage interest rate per year}}{\text{number of compounding periods per year}}$

## Depreciation

- Depreciation is the decrease in value of something over time.
- The future value is an estimate of the reduced value of an asset at any point in time.
- There are three different ways of depreciating assets.
  - Flat rate depreciation calculates the future value of an asset by reducing the value every year by a fixed amount. It is similar to simple interest. The difference is we subtract rather than add the fixed amount each time period.
  - Unit cost depreciation occurs when an asset is depreciated according to the amount of use it has had, not according to its age. It uses a rate per unit of use rather than a fixed amount each year.
  - Reducing balance depreciation calculates the value of an asset by reducing it every year by a fixed percentage of its value in the preceding year. It is similar to compound interest. The difference is we subtract rather than add the rate.

## Interest and depreciation summary

| | Simple interest investment (and flat rate appreciation) | Aat rate depreciation | Unit cost depreciation | Compound interest investment | Reducing balance depreciation |
|---|---|---|---|---|---|
| Recurrence relation for balance | $V_0$ = principal, $V_{n+1} = V_n + d$ | $V_0$ = initial asset value, $V_{n+1} = V_n - d$ | $V_0$ = initial asset value, $V_{n+1} = V_n - d$ | $V_0$ = principal, $V_{n+1} = \left(1 + \frac{r}{100}\right) V_n$ | $V_0$ = initial value of the asset, $V_{n+1} = \left(1 - \frac{r}{100}\right) V_n$ |
| Type | add an amount | subtract an amount | subtract an amount | multiply by a number greater than 1 | multiply by a number between 0 and 1 |
| Growth/ Decay | linear growth: increase by same amount | linear decay: decrease by same amount | linear decay: decrease by same amount | geometric growth: increases get larger | geometric decay: decreases get smaller |
| Graph |  |  |  |  |  |
| Rule | $V_n = V_0 + nd$ | $V_n = V_0 - nd$ | $V_n = V_0 - nd$ | $V_n = \left(1 + \frac{r}{100}\right)^n \times V_0$ | $V_n = \left(1 - \frac{r}{100}\right)^n \times V_0$ |
| $d$ | fixed amount each year $d = \frac{r}{100} \times V_0$ | fixed amount each year $d = \frac{r}{100} \times V_0$ | cost per unit of use | — | — |
| $r\%$ | interest rate per year $r = \frac{d}{V_0} \times 100\%$ | depreciation rate per year $r = \frac{d}{V_0} \times 100\%$ | – | interest rate per compounding period | depreciation rate per year |
| $n$ | number of years | number of years | number of units of use | number of compounding periods | number of years |

- Total amount of interest/depreciation after $n$ years/compounding periods = $V_n - V_0$
- Amount of interest/depreciation *in* the nth year/compounding period = $V_{n-1} - V_n$
- The last year/compounding period of a loan/depreciation often involves a partial amount.

## Effective interest rates

- The nominal interest rate is the interest rate given for an investment or loan, which consists of a rate per year and a compounding period.
- The effective interest rate is the interest rate after the compounding periods have been taken into account, which allows us to compare different rates.
- $r_{effective} = \left[\left(1 + \frac{r}{100n}\right)^n - 1\right] \times 100\%$

  where

  $r$ = the nominal interest rate *per year*

  $n$ = the number of compounding periods *per year.*



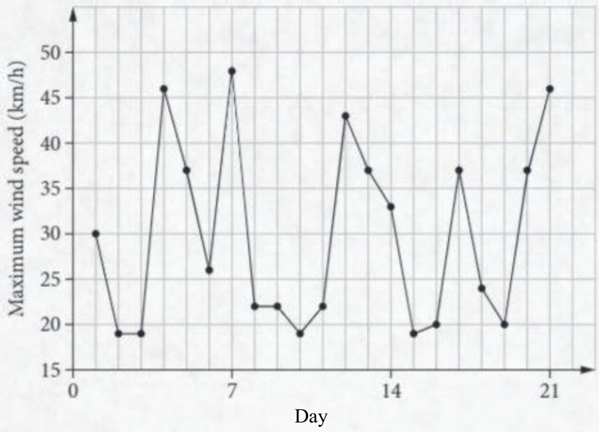
9780170448406

# Cumulative examination 1

Total number of marks: 13 Reading time: 6 minutes Writing time: 32 minutes

*Use the following information to answer the next three questions.*

The wind speed at a city location is measured throughout the day. The time series plot shows the daily *maximum wind speed,* in kilometres per hour, over a three-week period.




1 ©VCAA | 20171CQ13] The time series is best described as having

A seasonality only.

B irregular fluctuations only.

C seasonality with irregular fluctuations.

D a decreasing trend with irregular fluctuations.

E an increasing trend with irregular fluctuations.

2 ©VCAA 2017 1CQ14 The seven-median smoothed *maximum wind speed*, in kilometres per hour, for Day 4 is closest to

A 22 B 26 C 27 D 30 E 32

3 ©VCAA 2017 1CQ15 The table below shows the daily *maximum wind speed*, in kilometres per hour, for the days in week 2.

| Day | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|
| Maximum wind speed (km/h) | 22 | 22 | 19 | 22 | 43 | 37 | 33 |

A four-point moving mean with centring is used to smooth the time series data above. The smoothed *maximum wind speed,* in kilometres per hour, for Day 11 is closest to

A 22 B 24 C 26 D 28 E 30

4 ©VCAA 2011 1BRMQ2 An amount of $22000 is invested for three years at an interest rate of 3.5% per annum, compounding annually. The value of the investment at the end of three years is closest to

A $2310 B $9433 C $24040 D $24392 E $31433

5 ©VCAA 2020 1CQ23 Consider the following four recurrence relations representing the value of an asset after $n$ years, $V_n$.

. $V_o$-20000, $V_{,,+1}$ = V,, + 2500

. $V_o$-20000, $V_{,,+1}$-V,,-2500

. $V_o$- 20000, $V_{,,+1}$ = 0.875V,

. $V_o$ = 20000, $V_{,,+1}$ = 1.125 V,-2500

How many of these recurrence relations indicate that the value of an asset is depreciating?

A0 BI C 2 D3 E 4

6 ©VCAA | 20051 NPQ6 MODIFIED] The first term in a sequence is 3. The second term is obtained by multiplying the value of the first term by 5, then subtracting 2. The third term is obtained by multiplying the value of the second term by 5, then subtracting 2. This pattern continues. The recurrence relation which generates this sequence is

A MO = 3> U„ = 5«n+l -2    B UQ - 3, «n+l = 5“n -2    c $u_0$ = 3, -5u M+1 = u - 2

D MO = 3» 5u„+l " $u_n$ -2    E $u_0$ = 3, «n+l = -5«n -2

7 ©VCAA 2021 1CQ17 The recurrence relation $L_o = 37$, $L_{n+i} = L_n + C$ can generate a sequence of numbers. The value of $L_2$ is 25. The value of C is

A -6    B -4    C 4    D 6    E 37

8 ©VCAA 2017NICQ19J Consider the recurrence relation: $£_0 = 2000$, $L_{n+l} = L_n + 80$. This recurrence relation could be used to model a

A simple interest investment of $2000 with an annual interest rate of 4%.

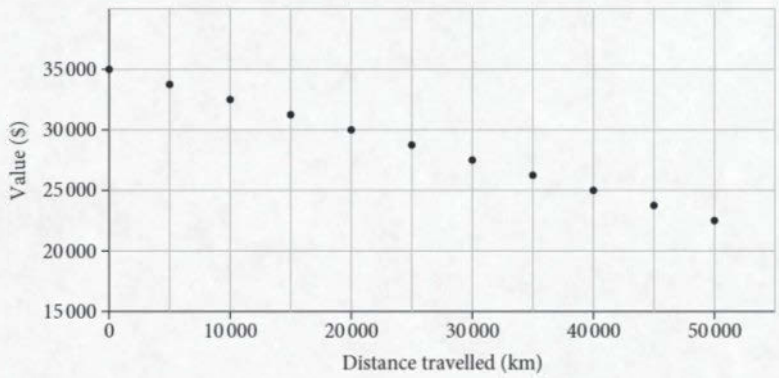
B simple interest investment of $2000 with an annual interest rate of 8%.

C simple interest investment of $2000 with an annual interest rate of 40%.

D compound interest investment of $2000 with an annual interest rate of 4%.

E compound interest investment of $2000 with an annual interest rate of 8%.

9 ©VCAA 2015 1BRMQ7 The following graph shows the depreciating value of a van.




The graph could represent the van being depreciated using

A flat rate depreciation with an initial value of $35 000 and a depreciation rate of $25 per year.

B flat rate depreciation with an initial value of $35 000 and a depreciation rate of 25 cents per year.

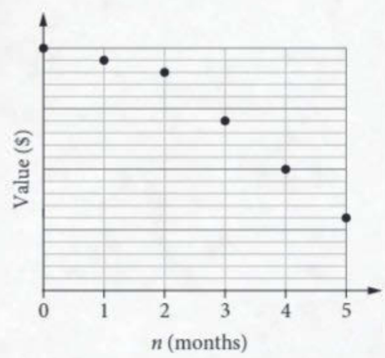
C reducing balance depreciation with an initial value of $35 000 and a depreciation rate of 2.5% per annum.

D unit cost depreciation with an initial value of $35 000 and a depreciation rate of 25 cents per kilometre travelled.

E unit cost depreciation with an initial value of $35 000 and a depreciation rate of $25 per kilometre travelled.

10 ©VCAA 20181CQ20 The graph shows the value, V„, of an asset as it depreciates over a period of five months.

Which one of the following depreciation situations does this graph best represent?



A flat rate depreciation with a decrease in depreciation rate after two months

B flat rate depreciation with an increase in depreciation rate after two months

C unit cost depreciation with a decrease in units used per month after two months

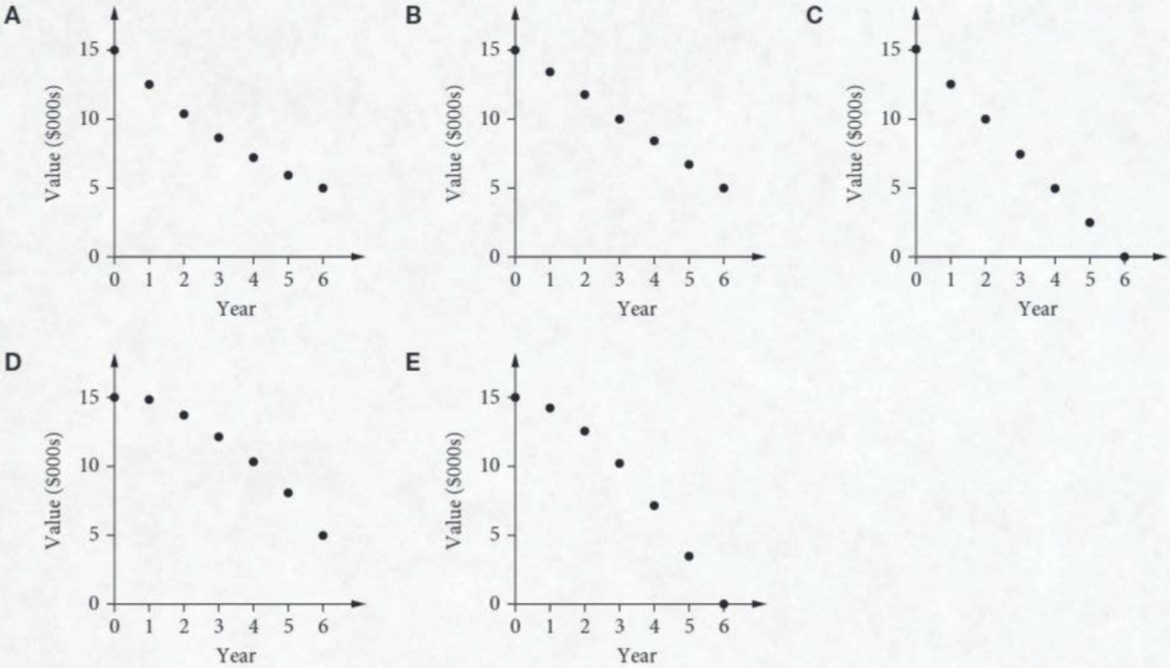
D reducing balance depreciation with an increase in the rate of depreciation after two months

E reducing balance depreciation with a decrease in the rate of depreciation after two months

11 ©VCAA [ 2003 IBRMQ5 I Zoltan is running a convenience store. He purchases equipment for $6500. It is anticipated that the equipment will last five years and have a depreciated value of $2000. Assuming the flat rate method of depreciation, the equipment is depreciated annually by

A $400 B $900 C $1027 D $1300 E $4500

12 ©VCAA 2004iBRMQ8n A machine is purchased for $15000. Using the reducing balance method of depreciation, its future value after six years will be $5000. The graph that best represents the value of the machine at the end of each year over the six-year period is




13 ©VCAA 2002 1BRMQ8 The following graph represents the growth of an investment over several years.

If *A* dollars is the value of the investment after *n* years, then a rule for describing the growth of this investment could be

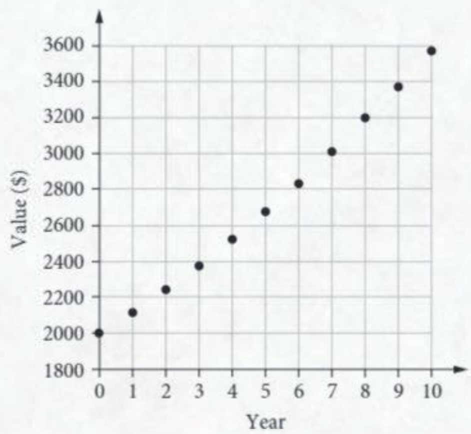
A A = 2000 x (1.06)”

B A = 2000 x (0.06)”

C A = 2000 x 1.06n

D A = 2000 x 0.06n

E A = 2000+ (1.06)”




14 Which of the following has the largest effective interest rate?

A 6.6% p.a. compounding yearly

B 6.53% p.a. compounding daily

C 6.5% p.a. compounding monthly

D 6.45% p.a. compounding weekly

E 6.57% p.a. compounding quarterly

# Cumulative examination 2

Total number of marks: 16 Reading time: 5 minutes Writing time: 24 minutes

1 ©VCAA 2011 2CQ2 J (3 marks) The following table shows information about a particular country. It shows the percentage of women, by age at first marriage, for the years 1986,1996 and 2006.

| | Year of marriage | | |
|---|---|---|---|
| Age of women at first marriage | 1986 | 1996 | 2006 |
| 19 years and under | 8.5% | 3.7% | 2.0% |
| 20 to 24 years | 42.1% | 31.3% | 21.5% |
| 25 to 29 years | 23.4% | 31.7% | 34.5% |
| 30 years and over | 26.0% | 33.3% | 42.0% |

a Of the women who first married in 1986, what percentage were aged 20 to 29 years inclusive? 1 mark

b Does the information in the table support the opinion that, for the years 1986,1996 and 2006, the age of women at first marriage was associated with year of marriage? Justify your answer by quoting appropriate percentages. It is sufficient to consider one age group only when justifying your answer. 2 marks

2 ©VCAA 2016S 2CQ5 J (2 marks) There is an association between the variables *population density,* in people per square kilometre, and *area,* in square kilometres, of 38 inner suburbs of the same city. For this association, $r^2 = 0.141$

a Write down the value of the correlation coefficient for this association between the variables *population density* and *area.* Round your answer to three decimal places, 1 mark

b The mean and standard deviation of the variables *population density* and *area* for these 38 inner suburbs are shown in the table.

| | Population density (people per km$^2$) | *Area* (km$^2$) |
|---|---|---|
| Mean | 4370 | 3.4 |
| Standard deviation | 1560 | 1.6 |

One of these suburbs has a population density of 3082 people per square kilometre. Determine the standard z-score of this suburbs population density. Round your answer to one decimal place. 1 mark

3 ©VCAA 2007 2BRMQ3b (2 marks) Khan will depreciate his $900 fax machine for taxation purposes. He considers two methods of depreciation.

Flat rate depreciation

Under flat rate depreciation the fax machine will be valued at $300 after five years,

a Calculate the annual depreciation in dollars. 1 mark

Unit cost depreciation

Suppose Khan sends 250 faxes a year. The $900 fax machine is depreciated by 46 cents for each fax it sends.

b Determine the value of the fax machine after five years. 1 mark

4 ©VCAA | 2004 2BRMQid J (1 mark) Remy borrows $650 to buy a digital camera. Remy uses his camera for work and he wants to depreciate its value over five years. He can either use a flat rate method of depreciation at the rate of 12% per annum or a reducing balance method of depreciation at the rate of 15% per annum. Which method gives the greater total depreciation over five years? Explain your answer.

5 © VCAA | 2O18N2CQ7 (4 marks) Roslyn invested some money in a savings account that earns interest compounding annually. The interest is calculated and paid at the end of each year. Let $V_n$ be the amount of money in Roslyn's savings account, in dollars, after $n$ years. The recursive calculations below show the amount of money in Roslyn's savings account after one year and after two years.

$V_o = 5000$

^ = 1.05x 5000 = 5250

$V_2 = 1.05x\ 5250 = 5512.50$

a How much money did Roslyn initially invest? 1 mark

b How much interest in total did she earn by the end of the second year? 1 mark

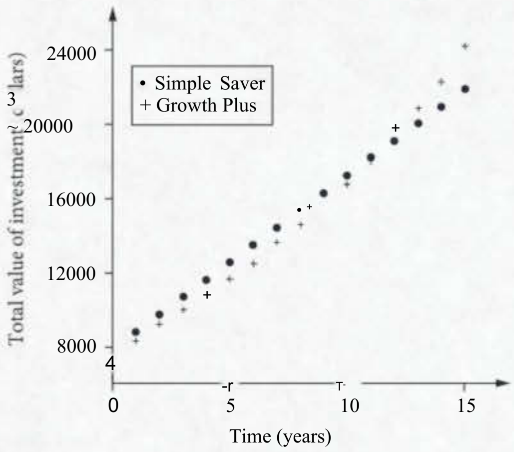
c Let $V_n$ be the amount of money in Roslyn's savings account, in dollars, after $n$ years.

Write down a recurrence relation, in terms of $V_o$. V,,$_{+1}$ and $V_n$, that can be used to model the amount of money, in dollars, in Roslyn's savings account. 1 mark

d Roslyn plans to use her savings to pay for a holiday. The holiday will cost $6000.

What minimum annual percentage interest rate would have been required for Roslyn to have saved this $6000 after two years? Round your answer to one decimal place. 1 mark

6 © VCAA | 2010 2BRMQ3 | (4 marks) Simple Saver is a simple interest investment in which interest is paid annually. Growth Plus is a compound interest investment in which interest is paid annually. Initially, $8000 is invested with both Simple Saver and Growth Plus. The graph shows the total value (principal and all interest earned) of each of these investments over a 15-year period.




The increase in the value of each investment over time is due to interest.

a Which investment pays the highest annual interest rate, Growth Plus or Simple Saver? Give a reason to justify your answer. 1 mark

b After 15 years, the total value (principal and all interest earned) of the Simple Saver investment is $21800. Find the amount of interest paid annually. 1 mark

c After 15 years, the total value (principal and all interest earned) of the Growth Plus investment is $24 000.

i Write down an equation that can be used to find the annual compound interest rate, $r$. 1 mark

ii Determine the annual compound interest rate. Write your answer as a percentage correct to one decimal place. 1 mark

CHAPTER

# 6 LOANS, INVESTMENTS AND FINANCE SOLVERS

STUDY DESIGN COVERAGE

## Study Design coverage

### AREA OF STUDY 2: DISCRETE MATHEMATICS

#### Reducing balance loans

- use of a first-order linear recurrence relation to model and analyse (numerically and graphically) the amortisation of a reducing balance loan, including the use of a recurrence relation to determine the value of the loan or investment after $n$ payments for an initial sequence from first principles
- use of a table to investigate and analyse the amortisation of a reducing balance loan on a step-by-step basis, the payment made, the amount of interest paid, the reduction in the principal and the balance of the loan
- use of technology with financial modelling functionality to solve problems involving reducing balance loans, such as repaying a personal loan or a mortgage, including the impact of a change in interest rate on repayment amount, time to repay the loan, total interest paid and the total cost of the loan.

#### Annuities and perpetuities

- use of a first-order linear recurrence relation to model and analyse (numerically and graphically) the amortisation of an annuity, including the use of a recurrence relation to determine the value of the annuity after $n$ payments for an initial sequence from first principles
- use of a table to investigate and analyse the amortisation of an annuity on a step-by-step basis, the payment made, the interest earned, the reduction in the principal and the balance of the annuity
- use of technology to solve problems involving annuities including determining the amount to be invested in an annuity to provide a regular income paid, for example, monthly, quarterly
- simple perpetuity as a special case of an annuity that lasts indefinitely.

#### Compound interest investment with periodic and equal additions to the principal

- use of a first-order linear recurrence relation to model and analyse (numerically and graphically) annuity investment, including the use of a recurrence relation to determine the value of the investment after $n$ payments have been made for an initial sequence from first principles
- use of a table to investigate and analyse the growth of an annuity investment on a step-by-step basis after each payment is made, the payment made, the interest earned and the balance of the investment
- use of technology with financial modelling functionality to solve problems involving annuity investments, including determining the future value of an investment after a number of compounding periods, the number of compounding periods for the investment to exceed a given value and the interest rate or payment amount needed for an investment to exceed a given value in a given time.



**Video playlists (8):**
- **6.1** Using finance solvers
- **6.2** Reducing balance loans
- **6.3** Using finance solvers for reducing balance loans
- **6.4** Changing the terms of reducing balance loans
- **6.5** Annuities
- **6.6** Perpetuities
- **6.7** Annuity investments
- **VCE question analysis** Loans, investments and finance solvers

**Worksheets (3):**
- **6.2** Reducing balance loans
- **6.3** Loan repayment problems
- **6.5** Annuity problems

**Puzzles (1):**
- **6.6** Perpetuities

9

Nelson MindTap

To access resources above, visit cengage.com.au/nelsonmindtap

Video playlist
Using finance solvers

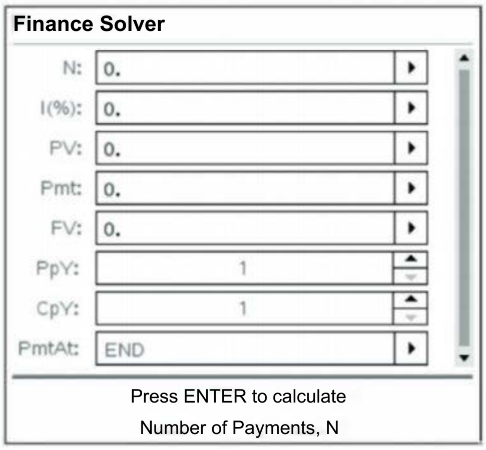
# Using finance solvers

## Introducing CAS finance solvers

CAS finance solvers are a fast way of solving complex financial problems. We will be dealing with more complex loan and investment problems in this chapter, but first we will look at how to use CAS finance solvers for the sorts of situations that were covered in the previous chapter.

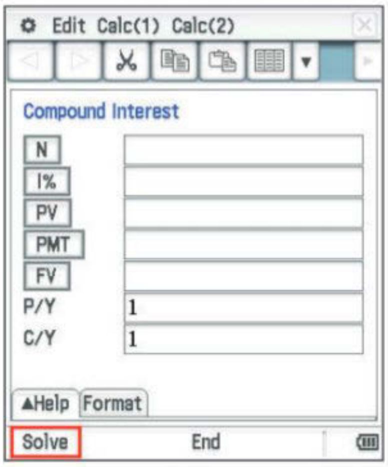
Follow these steps to access the finance solver on your CAS.

**TI-Nspire**




1 Start a new document and add a Calculator page.
2 Press menu > Finance > Finance Solver.
(Note: Scroll down to display the CpY: and PmtAt: fields.)
3 After entering the other values, place the cursor in the blank field and press enter to evaluate.

**ClassPad**




1 Tap Menu and open the Financial application.
2 If the list of Financial options does not display, tap Edit > Clear All.
3 Tap Compound Interest.
4 After entering the other values, place the cursor in the blank field and tap Solve to evaluate.

For the Using CAS examples in this chapter, the TI-Nspire and ClassPad screens will be combined as follows, with the value to be calculated highlighted in red:

| |
|---|
| N |
| I% |
| PV |
| Pmt or PMT |
| FV |
| PpY or P/Y |
| CpY or C/Y |


9780170448406

### Using finance solvers for compound interest problems

Values in a finance solver can be positive, negative or zero.

- Money coming *to the person* is positive.
- Money going away/rom *the person* is negative.

| Compound interest investments with no payments | |
|---|---|
| N | Total number of compounding periods |
| 1% | Interest rate per year |
| PV | Present value for an investment is negative because the money is going away from the person to the bank. |
| Pmt or PMT | Regular payments for the investments we have looked at so far are zero. |
| FV | Future value has the opposite sign of the present value, so it will be positive. |
| PpY or P/Y | Number of payments per year. This will always take the same value as CpY or C/Y. |
| CpY or C/Y | Number of compounding periods per year |

| Reducing balance depreciation | |
|---|---|
| N | Total number of compounding periods |
| 1% | Interest rate per year will be negative because the asset is losing money. |
| PV | Present value for reducing balance depreciation is negative because the person has spent money to buy the asset so the money is going away from them. |
| Pmt or PMT | Regular payments for depreciation are zero. |
| FV | Future value has the opposite sign of the present value, so it will be positive (or zero). |
| PpY or P/Y | Number of payments per year. This will always take the same value as CpY or C/Y. |
| CpY or C/Y | Number of compounding periods per year |

PV and FV always have opposite signs (except when FV is zero).

## Compound interest and finance solvers

Finance solvers can be used to solve problems involving compound interest investments and reducing balance depreciation like the questions in the previous chapter. The examples we have looked at so far have had no regular payments involved so Pmt or PMT is 0.

WB

p. 89

## WORKED EXAMPLE 1 Using finance solvers for compound interest investments

Collette invested $36 000 in a term deposit earning 8.1% p.a. interest compounding monthly. Give answers to the nearest cent.

**Steps** | **Working**

a What is the value of Collette's investment after five years?

TI-Nspire + ClassPad

| Steps | Working | |
|---|---|---|
| 1 Find FV after five years: 5 x 12 = 60 months. | | |
| 2 Total number of compounding periods | N | 60 |
| Annual interest rate | 1% | 8.1 |
| Present value for an investment is negative. | PV | -36000 |
| Payment amount | Pmt or PMT | 0 |
| Future value has the opposite sign to present value | FV | 53901.493347501 |
| Same as CpY or C/Y | PpYorP/Y | 12 |
| Number of compounding periods per year | CpY or C/Y | 12 |
| 3 Write the answer, rounding to nearest cent. | The value of Collette's investment after five years will be $53 901.49. | |

b How long will it take for her investment to grow to $46 000?

| Steps | Working | |
|---|---|---|
| 1 Enter the amount as FV and find N. | | |
| 2 Total number of compounding periods | N | 36.436862038112 |
| Annual interest rate | 1% | 8.1 |
| Present value for an investment is negative. | PV | -36000 |
| Payment amount | Pmt or PMT | 0 |
| Future value has the opposite sign to present value. | FV | 46000 |
| Same as CpY or C/Y | PpY or P/Y | 12 |
| Number of compounding periods per year | CpY or C/Y | 12 |
| 3 Write the answer, rounding up to the next whole number. | It will take 37 months for the investment to grow to $46 000. | |

c i How long will it take for her investment to grow to $46 000 if the interest compounded quarterly?

ii How much longer will it take for quarterly compounding to reach $46 000 compared to monthly compounding?

TI-Nspire + ClassPad

| | | |
|---|---|---|
| 1 Change CpY or C/Y to compounding quarterly and find N. | | |
| 2 Total number of compounding periods | N | 12.226964462433 |
| Annual interest rate | I% | 8.1 |
| Present value for an investment is negative. | PV | -36000 |
| Payment amount | Pmt or PMT | 0 |
| Future value has the opposite sign to present value. | FV | 46000 |
| Same as CpY or C/Y | PpY or P/Y | 4 |
| Number of compounding periods per year | CpY or C/Y | 4 |
| 3 i Write the answer, rounding up to the next whole number. | It will take 13 quarters for the investment to grow to $46 000. | |
| 4 ii Use the fact that there are three months in a quarter to compare. | It takes 13x2 = 39 months to reach $46 000 compounding quarterly.<br>So, compounding quarterly takes two months longer than compounding monthly. | |

## Exam hack

Finance solvers are often the quickest way to answer a question. You can use them if the question doesn't specifically say to use a different method.

6.1

## WORKED EXAMPLE 2 Using finance solvers for reducing balance depreciation

Fran has been depreciating her farm machinery on a reducing balance basis rate per annum.

**Steps** **Working**

a Fran bought a planter for $44 000 and after seven years its value is $22 390. What is the rate of depreciation, rounded to one decimal place?

TI-Nspire + ClassPad

| | | |
|---|---|---|
| 1 Find I from negative PV to positive FV. | | |
| 2 Total number of compounding periods | N | 7 |
| Annual interest rate for depreciation is negative | I% | -9.1999857397511 |
| Present value for depreciation is negative. | PV | -44000 |
| Payment amount | Pmt or PMT | 0 |
| Future value has the opposite sign to present value. | FV | 22390 |
| Same as CpY or C/Y | PpY or P/Y | 1 |
| Number of compounding periods per year | CpY or C/Y | 1 |
| 3 Write the answer, rounding to one decimal place. | The rate of depreciation is 9.2% p.a. | |

WB p. 91

b Fran has been depreciating her combine harvester at a rate of 12% p.a. After five years its value was $168874. What was the original cost of the combine harvester to the nearest dollar?

| | TI-Nspire | + ClassPad |
|---|---|---|
| 1 Enter the amount as FV and find PV. | | |
| 2 Total number of compounding periods | N | 5 |
| Annual interest rate for depreciation is negative. | I% | -12 |
| Present value for depreciation is negative. | PV | -319999.5956735 |
| Payment amount | Pmt or PMT | 0 |
| Future value has the opposite sign to present value. | FV | 168874 |
| Same as CpY or C/Y | PpY or P/Y | 1 |
| Number of compounding periods per year | CpY or C/Y | 1 |
| 3 Write the answer, rounding to the nearest dollar. | The combine harvester was originally bought for $320 000. | |

WB p. 92

**WORKED EXAMPLE 3** Using finance solvers for two-step compound interest investment problems

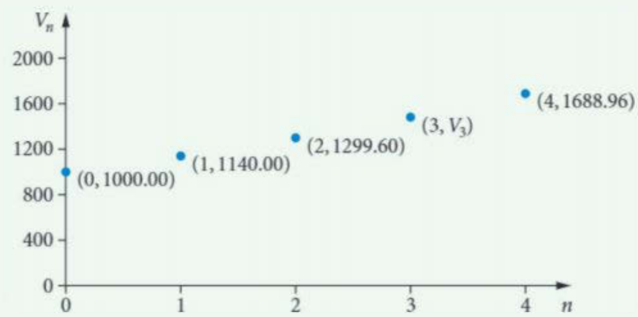
Solve each of the following by using a finance solver twice.

**Steps** **Working**

a The balance of Maddie's investment, which compounds monthly, was $5416 after four months and $5604 after a year. What was the amount of money that she initially invested to the nearest dollar?

| | TI-Nspire 1 | + ClassPad |
|---|---|---|
| 1 Find the interest rate from negative PV to positive FV. | | |
| 2 Total number of compounding periods = 12-4 = 8 months | N | 8 |
| Annual interest rate | I% | 5.1293955805141 |
| Present value for an investment is negative. | PV | -5416 |
| Payment amount | Pmt or PMT | 0 |
| Future value has the opposite sign to present value. | FV | 5604 |
| Same as CpY or C/Y | PpY or P/Y | 12 |
| Number of compounding periods per year | CpY or C/Y | 12 |
| 3 Use the unrounded answer to find the initial amount. | The interest rate is 5.129 ...%. | |
| 4 Total number of compounding periods | N | 4 |
| Annual interest rate from previous solver (unrounded) | I% | 5.1293955805141 |
| Present value for an investment is negative. | PV | -5234.306923626 |
| Payment amount | Pmt or PMT | 0 |
| Future value has the opposite sign to present value. | FV | 5416 |
| Same as CpY or C/Y | PpY or P/Y | 12 |
| Number of compounding periods per year | CpY or C/Y | 12 |
| 5 Write the answer, rounding to the nearest dollar. | The amount of money that Maddie initially invested was $5234. | |

b The graph shows the value, in dollars, of a compound interest investment after $n$ compounding periods, $V_n$, for a period of four compounding periods. What is the value of $V_3$, rounded to the nearest cent?




TI-Nspire + ClassPad

| | | | |
|---|---|---|---|
| 1 | Look at the first compounding period only. Find the interest rate using the value for $n$ = 1 as FV. | | |
| 2 | Total number of compounding periods | N | 1 |
| | Annual interest rate | I% | 14 |
| | Present value for an investment is negative. | PV | -1000 |
| | Payment amount | Pmt or PMT | 0 |
| | Future value has the opposite sign to present value. | FV | 1140 |
| | Same as CpY or C/Y | PpY or P/Y | 1 |
| | Number of compounding periods per year | CpY or C/Y | 1 |
| 3 | Use the unrounded answer to find the value for n = 3. | The rate of depreciation is 14%. | |
| 4 | Total number of compounding periods | N | 3 |
| | Annual interest rate from previous solver (unrounded) | 1% | 14 |
| | Present value for an investment is negative. | PV | -1000 |
| | Payment amount | Pmt or PMT | 0 |
| | Future value has the opposite sign to present value. | FV | 1481.544 |
| | Same as CpY or C/Y | PpY or P/Y | 1 |
| | Number of compounding periods per year | CpY or C/Y | 1 |
| 5 | Write the answer, rounding to the nearest cent. | $V_3$ = $1481.54 | |

6.1

## EXERCISE 6.1 Using finance solvers

ANSWERS p. 712

### Mastery

1 WORKED EXAMPLE 1 Sara invested $40000 in a term deposit earning 6.3% p.a. compounding monthly,

a What is the value of her investment after seven years? Give answers to the nearest cent,

b How long will it take for her investment to grow to $50 000?

c i How long will it take for her investment to grow to $50 000 if the interest is compounded quarterly?

ii How much longer will it take to reach $50 000 if compounding quarterly, rather than compounding monthly?

20 WORKED EXAMPLE 2 Eugen depreciates his printing machinery on a reducing balance basis rate per annum.

a He bought a paper guillotine for $33000 and after eight years its value is $17840. What is the rate of depreciation, rounded to one decimal place?

b He depreciates his printing press at a rate of 10%. After seven years its value is $64 570. What was the original cost of the printing press, rounded to the nearest dollar?

3 a Kirily invested $45 700 in an account earning 2.9% p.a. compounding weekly. To find the balance after three years, what values would you enter for N and CpY or C/Y in a CAS finance solver?

b Fletcher invested $22 000 at 4.1% p.a. compounding six-monthly. To find the value of his investment after five years, what values would you enter for N and PpY or P/Y in a CAS finance solver?

c Sinjins investment, which compounds monthly, has grown from $15000 to $17 000 in the last four years. To find the interest rate of the account, what values would you enter for N and PV in a CAS finance solver?

d Heidi invested $4000 in an account earning 3.6% interest compounding fortnightly. To find out how long it will take to grow to $5000, what values would you enter for 1%, PV and FV in a CAS finance solver?

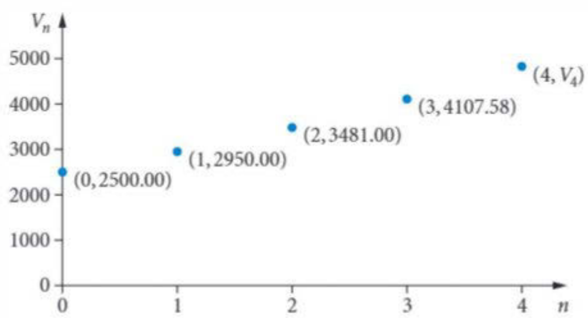
e A company depreciates a machine bought for $70 000 on a reducing balance basis rate per annum. If its value after six years is $3200, what would you enter for N, PV and CpY or C/Y in a CAS finance solver to find the rate of depreciation?

f A business has been depreciating its laptops at a rate of 12.5% p.a. After four years their value was $18600. To find the original value of the laptops, what would you enter for N, 1% and FV in a CAS finance solver?

4S WORKED EXAMPLE 3 Solve each of the following by using a finance solver twice.

a The balance of Arthurs investment, which compounds monthly, was $6234 after eight months and $7840 after one year. What was the amount of money that he initially invested to the nearest dollar?

b The graph below shows the value, in dollars, of a compound interest investment after *n* compounding periods, V,,, for a period of four compounding periods. What is the value of $V_4$, rounded to the nearest cent?




## Exam practice

80–100% 60–79% 0–59%

5 ©VCAA 2004 1BRMQ2 72% Ardy invests $150 000 for six years at an interest rate of 3.5% per annum, compounding annually. The value of the investment after six years is

A $31500.00 B $34388 30 C $178107.00
D $181500.00 E $184388.30

6 ©VCAA 2010 1BRMQ5 57% A file server costs $30 000. The file server depreciates by 20% of its value each year. After three years its value is

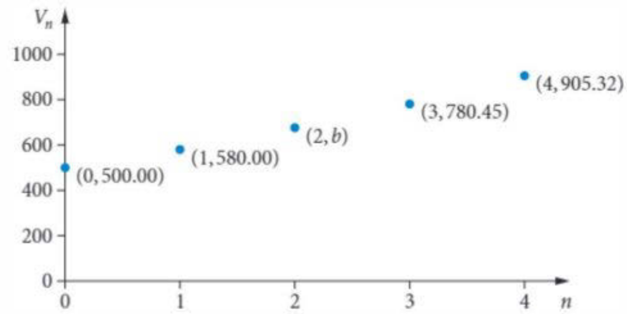
A $6000 B $12000 C $15360 D $19200 E $24000


9780170448406

7 ©VCAA | 20201CQ29 54% The value of a van purchased for $45 000 is depreciated by $k\%$ per annum using the reducing balance method. After three years of this depreciation, it is then depreciated in the fourth year under the unit cost method at the rate of 15 cents per kilometre. The value of the van after it travels 30000 km in this fourth year is $26166.24. The value of $k$ is

A 9 B 12 C 14 D 16 E 18

8 ©VCAA 20191CQ21J 54% The graph below shows the value, in dollars, of a compound interest investment after $n$ compounding periods, $V_w$, for a period of four compounding periods.




The coordinates of the point where $n = 2$ are $(2, b)$.

The value of $b$ is

A 660.00 B 670.00 C 672.80 D 678.40 E 685.60

9 ©VCAA 20191CQ24J 51% Millie invested $20 000 in an account at her bank with interest compounding monthly. After one year, the balance of Millies' account was $20732. The difference between the rate of interest per annum used by her bank and the effective annual rate of interest for Millies' investment is closest to

A 0.04% B 0.06% C 0.08% D 0.10% E 0.12%

10 ©VCAA | 20081BRMQ3 I 47% A computer originally purchased for $6000 is depreciated each year using the reducing balance method. If the computer is valued at $2000 after four years, then the annual rate of depreciation is closest to

A 17% B 24% C 25% D 28% E 33%

11 ©VCAA | 2017N1CQ24 J Geoff has a compound interest investment that earns interest compounding monthly. The balance of Geoff's compound interest investment was $4418.80 after six months. The balance of Geoff's compound interest investment was $4862.80 after two years. The amount of money that Geoff initially invested is closest to

A $4000 B $4015 C $4280 D $4370 E $4715

12 ©VCAA 2018N2CQ8bc , (3 marks) Richard is selling his stereo system to help pay for a holiday. The stereo system was originally purchased for $8500. He is using a reducing balance depreciation method, with an annual depreciation rate of 8%.

a Using this depreciation method, what is the value of the stereo system four years after it was purchased? Round your answer to the nearest cent. 1 mark

b Four years after it was purchased, Richard sold his stereo system for $4500. Assuming a reducing balance depreciation method was used, what annual percentage rate of depreciation did this represent? Round your answer to one decimal place. 2 marks

**Exam hack**

When answering extended answer questions with finance solvers, show your working by listing the values your entered for N, 1%, PV etc.

6.1

Video playlist Reducing balance loans

Worksheet Reducing balance loans

# @ Reducing balance loans

## Reducing balance loan recurrence relations

A reducing balance loan is a compound interest loan with regular payments. Home loans and personal loans are examples of reducing balance loans. For each compounding period, the balance of the loan changes in two ways. The balance:

- increases, because compound interest is being charged by the bank
- decreases, because a fixed amount is paid into the account.

The change in the amount owed is a combination of geometric growth and linear decay.

### Reducing balance loan recurrence relation

The recurrence relation for the value $V_n$ of a reducing balance loan is:

$$V_0 = \text{principal}, \quad V_{n+1} = \left(1 + \frac{r}{100}\right)V_n - d$$

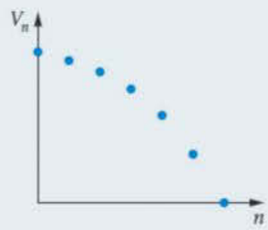
where

$r$ = percentage interest rate per compounding period

$n$ = number of compounding periods

$d$ = payment made per compounding period.

The graph of a reducing balance loan recurrence relation would look like this.




### Exam hack

Don't confuse reducing balance depreciation and reducing balance loans.

WB p. 94

### WORKED EXAMPLE 4 Finding reducing balance loan recurrence relations

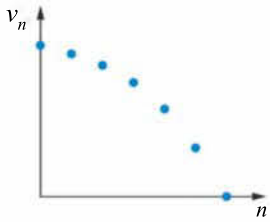
Freya has taken out a loan of $17 000 with an interest rate of 12.3% per annum compounding monthly, and she makes regular monthly payments of $540.

**a** Write a recurrence relation for the balance.

**b** Describe the sort of growth or decay modelled by the recurrence relation.

**c** Sketch the shape of the graph of the recurrence relation.

| Steps | Working |
|---|---|
| **a** 1 Find the number of compounding periods per year. | There are 12 compounding periods per year. |
| 2 Identify $V_n$, $V_0$, $r$ and $d$. | Let $V_n$ = the balance after $n$ compounding periods.<br>$V_0 = 17\,000$, $r = \frac{12.3}{12} = 1.025\%$, $d = 540$ |
| 3 Substitute the values into these formulas and simplify.<br>$V_0$ = principal<br>$V_{n+1} = \left(1 + \frac{r}{100}\right)V_n - d$ | $V_{n+1} = \left(1 + \frac{1.025}{100}\right)V_n - 540$<br>$= (1 + 0.010\,25)V_n - 540$<br>$= 1.010\,25V_n - 540$<br>$V_0 = 17\,000, \quad V_{n+1} = 1.010\,25V_n - 540$ |

| | |
|---|---|
| b Is there addition or subtraction involved? | There is subtraction involved. |
| Is there multiplication involved by a number greater than 1 or between 0 and 1? | There is multiplication involved by a number greater than 1. So, the recurrence relation models a combination of linear decay *and* geometric growth. |
| c Show the points forming the shape of the curve. |  |

6.2

## Exam hack

Questions don't always state that the loan is a reducing balance one. You know it's a reducing balance loan because it involves compound interest and a regular payment.

WB p. 95

## WORKED EXAMPLE 5 Working with reducing balance loan recurrence relations

Gary has taken out a reducing balance loan, compounding quarterly and with regular quarterly payments, according to the recurrence relation

$V_0 = 14000, \quad V_{n+1} = 1.0215V_n - 655,$

where

$V_n$ is the value of the loan after $n$ compounding periods.

| Steps | Working |
|---|---|
| **a** How much money did Gary borrow? | |
| Identify $V_0$. | Gary borrowed \$14 000. |
| **b** How much are the regular quarterly payments? | |
| Identify $d$. | The regular quarterly payments are \$655. |
| **c** Use recursion to write down calculations that show that the amount owing on his loan after three quarters will be \$12 915 when rounded to the nearest dollar. | |
| Step out the recurrence relation calculations to find $V_3$, giving answers to the nearest dollar. | $V_0 = 14000$<br>$V_1 = 1.0215V_0 - 655$<br>$= 1.0215 \times 14000 - 655$<br>$= 13646.00$<br>$V_2 = 1.0215V_1 - 655$<br>$= 1.0215 \times 13646.00 - 655$<br>$= 13284.39$<br>$V_3 = 1.0215V_2 - 655$<br>$= 1.0215 \times 13284.39 - 655$<br>$= 12915$ |

d What is the annual percentage compound interest rate for this loan?

1 Use the recurrence relation to find an equation for r, the percentage interest rate per compounding period.

Comparing

$$V_{n+1} = \left(1 + \frac{r}{100}\right)V_n - d$$

$$V_{n+1} = 1.0215V_n - 655$$

we can see that

$$1 + \frac{r}{100} = 1.0215$$

2 Solve for r, using CAS if necessary.

$$\frac{r}{100} = 1.0215 - 1$$

$$\frac{r}{100} = 0.0215$$

$r = 2.15\%$ per quarter.

3 Multiply $r$ by the compounding period to find the annual percentage compound interest rate.

The interest compounds quarterly, so the annual percentage compound interest rate is:

$r \times 4 = 2.15 \times 4 = 8.6\%$

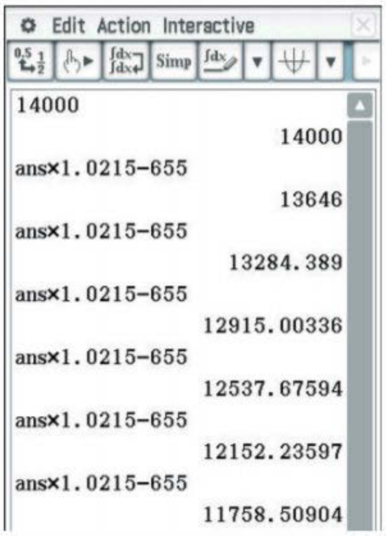
e After how many quarters will the balance of Garys loan fall below $12000?

Use CAS recursive computation to continue the recurrence relation calculations until they are less than the given amount.

The balance of Gary's loan will fall below $12000 after six quarters.

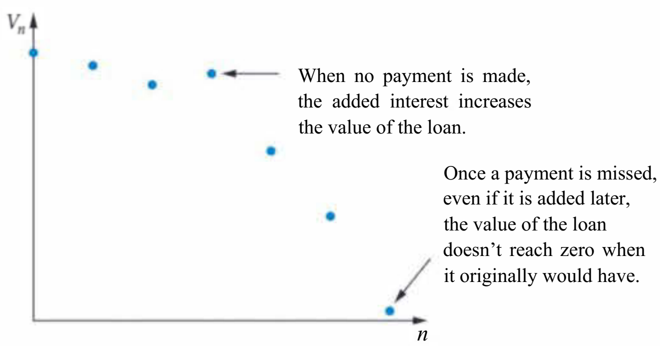
**TI-Nspire**


**ClassPad**




## Reducing balance loan graphs

If a payment is missed in a reducing balance loan, then extra interest needs to be paid on the next balance. This means the loan will take longer to repay even if the next payment is doubled to try to make up for the missed payment. The graph of a reducing balance loan where a payment was missed and the next payment was doubled would look like this.





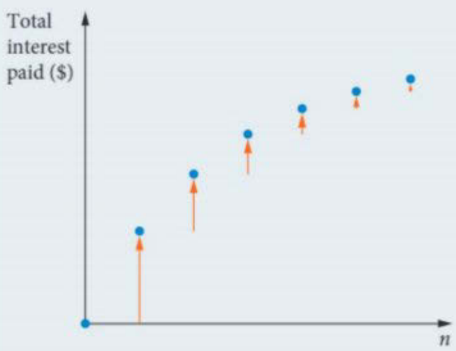
9780170448406

For a reducing balance loan, the regular payment is part interest and part principal.

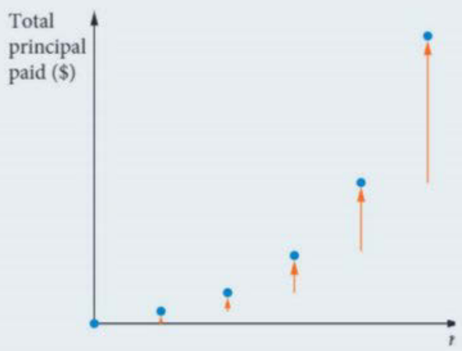
- The amount of interest ($) paid decreases with each compounding period. Towards the end of the life of the loan, the interest becomes a small part of the regular payment.
- The amount paid off the principal ($) increases with each compounding period. Towards the end of the life of the loan, the amount paid off the principal becomes a large part of the regular payment.

**Reducing balance loan payment graphs**

Graph shape of total interest paid ($)




Graph shape of total amount paid off principal ($)




## Reducing balance loan amortisation tables

Amortisation tables help us track the interest paid, the principal reduction and the balance of the loan for each compounding period.

**Reducing balance loan amortisation table**

$$r = \frac{\text{interest}}{\text{previous balance}} \times 100$$

| Payment number | Payment | Interest | Principal reduction | Balance |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | Vo |
| $n$ | $d$ | $\frac{r}{100}$ x previous balance | payment - interest | previous balance - principal reduction |

where

$n$ = payment number

$d$ = payment made per compounding period

$r$ = percentage interest rate per compounding period

$V_o$ = principal.

**Exam hack**

When entering values into an amortisation table, always round to the nearest cent. The usual rule of not rounding until the very end of your calculations doesn't apply.

WB

p. 97

**WORKED EXAMPLE 6** Using reducing balance loan amortisation tables

Reggie has taken out a housing loan of $450 000 with interest compounding monthly and is paying monthly instalments of $3797. The amortisation table below shows the first few calculations for his loan.

| Payment number | Payment | Interest | Principal reduction | Balance |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 450000.00 |
| 1 | 3797.00 | $\frac{r}{100}$ x 450000.00 = 2250.00 | 3797.00 - 2250.00 = 1547.00 | 450000.00- 1547.00 = 448453.00 |
| 2 | 3797.00 | $\frac{r}{100}$ x 448453.00 - 2242.27 | | |
| 3 | | | | |
| 4 | | | | |

**Steps** **Working**

a Use the table to show two calculations that will give r, the percentage interest rate per compounding period.

Use this formula:

$$r = \frac{\text{interest}}{\text{previous balance}} \times 100$$

Using interest for payment number 1:

$$r = \frac{2250.00}{450000.00} \times 100$$

= 0.005 x 100

= 0.5%

Using interest for payment number 2:

$$r = \frac{2242.27}{448453.00} \times 100$$

= 0.005 x 100

= 0.5%

b What is the nominal interest rate for the loan?

Use the compounding period to find the annual interest rate.

The compounding period is 12.

The nominal interest rate is 12 x 0.5 = 6% per annum compounding monthly.

c Complete the amortisation table, showing the interest paid, principal reduction and balance for the first four months of the loan, giving all values to the nearest cent.

Complete the table using:

$$\text{interest} = \frac{r}{100} \times \text{previous balance}$$

principal reduction = payment - interest

balance = previous balance - principal reduction

Give all values to the nearest cent.

| Payment number | Payment | Interest | Principal reduction | Balance |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 450000.00 |
| 1 | 3797.00 | 2250.00 | 1547.00 | 448453.00 |
| 2 | 3797.00 | 2242.27 | 1554.73 | 446898.27 |
| 3 | 3797.00 | 2234.49 | 1562.51 | 445335.76 |
| 4 | 3797.00 | 2226.68 | 1570.32 | 443765.44 |

**Exam hack**

When calculating r from an amortisation table, you can use any of the interest values except payment 0. Choose the option with the easiest calculations, which is nearly always the interest value for payment 1.



9780170448406

6.2

## WORKED EXAMPLE 7 Analysing reducing balance loan amortisation tables

Karl has taken out a short-term personal loan of $20 000 for a gap year trip to Europe. The interest rate is 18% p.a. compounding quarterly and it needs to be repaid by making four quarterly payments of $5570. Answer the following questions using the amortisation table of the loan given, below which has some missing entries.

| Payment number | Payment | Interest | Principal reduction | Balance |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 20000.00 |
| 1 | 5570.00 | 900.00 | 4670.00 | 15330.00 |
| 2 | 5570.00 | 689.85 | 4880.15 | |
| 3 | 5570.00 | 470.24 | | 5350.09 |
| 4 | 5570.00 | | 5329.25 | 20.84 |

WB p. 98

| Steps | Working |
|---|---|
| a What is the balance after two payments? | |
| balance = previous balance - principal reduction | The balance after two payments<br>= 15330.00-4880.15<br>= $10449.85 |
| b How much has the principal been reduced by at payment number 3? | |
| principal reduction = payment - interest | The principal reduction at payment number 3<br>= 5570.00 - 470.24<br>= $5099.76 |
| c How much interest was paid at payment 4? | |
| 1 Find r, the percentage interest rate per compounding period. | $r = \frac{18}{4} = 4.5\%$ |
| 2 interest - $\frac{r}{100}$ x previous balance | Interest for payment number 4<br>$= \frac{4.5}{100} \times 5350.09$<br>= 0.045 X 5350.09<br>= $240.75 |
| d How much of the principal has been paid after two payments? | |
| Add the values in the principal reduction column up to payment number 2. | 4670.00 + 4880.15 = 9550.15<br>$9550.15 of the principal has been paid after 2 payments. |
| e How much has the loan cost Karl? | |
| Add all the values in the interest column. | 900.00 + 689.85 + 470.24 + 240.75 = 2300.84<br>The loan has cost Karl $2300.84. |
| f Explain how you know the loan isn't paid out fully after the four payments. | |
| Look at the balance after the last payment. | The balance after the last payment is $20.84.<br>It should be $0 if the loan has been fully paid out. |
| g If Karl wants to fully pay the loan with payment 4, what should the fourth payment be adjusted to? | |
| Add the left over amount to the last payment. | 5570.00 + 20.84 = $5590.84<br>The last payment is $5590.84. |

## Interest-only loan recurrence relations and formula

For most reducing balance loans, the regular payments are greater than the interest at each compounding period. If the payments were less than the interest, then the balance of the loan would keep growing, and the borrower would owe an ever-increasing amount.

An interest-only loan is a reducing balance loan where the payments are *exactly* equal to the interest for each compounding period. This means the balance doesn't change. This sort of loan is used when the borrower expects what they've bought with the loan to increase in value so that it can be sold and the loan repaid in full from the sale.

### Interest-only loan

For an interest-only loan with value $V_n = V_o$ after $n$ compounding periods:

$$d = \frac{r}{100} \times V_0$$

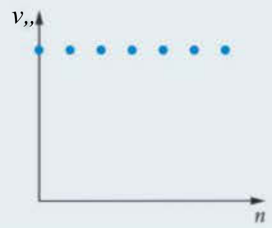
where

$d$ = the payment made per compounding period

$r$ = percentage interest rate per compounding period

$V_Q$ = principal.

The graph of an interest-only loan would look like this.




WB
p. 99

### WORKED EXAMPLE 8 Working with interest-only loans

**a** Elsa has taken out a loan with a recurrence relation for $V_n$, the value of the loan after $n$ compounding periods, of

$$V_0 = 14\,500, \quad V_{n+1} = 1.008V_n - 116$$

Find $V_1$ and $V_2$ and explain why this shows that this models an interest-only loan.

**b** Lewis has taken out an interest-only loan with an interest rate of 4.8% per annum compounding monthly with regular monthly payments of \$1200. How much did Lewis borrow?

**c** Jessie has taken out a \$500 000 interest-only loan. The interest rate is 3.5% per annum compounding quarterly. What are Jessie's regular quarterly repayments?

| Steps | Working |
|---|---|
| **a** 1 Find $V_1$. | $V_1 = 1.008V_0 - 116$<br>$= 1.008 \times 14\,500 - 116$<br>$= 14\,500$ |
| 2 Find $V_2$. | $V_2 = 1.008V_1 - 116$<br>$= 1.008 \times 14\,500 - 116$<br>$= 14\,500$ |
| 3 For interest-only loans, $V_n = V_0$. | The value of the loan stays at the principal value \$14 500 for all compounding periods. |
| **b** 1 Identify what we know and what we need to find from the interest-only loan formula. | $r = \frac{4.8}{12} = 0.4, d = 1200, V_0 = ?$<br>$d = \frac{r}{100} \times V_0$ |
| 2 Substitute into the formula and solve, using CAS if necessary. | $1200 = \frac{0.4}{100} \times V_0$<br>$V_0 = \frac{1200 \times 100}{0.4}$<br>$= 300\,000$ |
| 3 Write the answer. | Lewis borrowed \$300 000. |


9780170448406

| | |
|---|---|
| c 1 Identify what we know and what we need to find from the interest-only loan formula. | $r - \frac{3.5}{4} - 0.875$, $V_o$ - 500 000, *d* - ? <br> $<* = \frac{r}{100} x V_0$ |
| 2 Substitute into the formula and solve. | $d = \frac{0.875}{100} x\ 500000$ <br> $= 4375$ |
| 3 Write the answer. | Jessie's regular quarterly repayment for this loan is $4375. |

6.2

## EXERCISE 6.2 Reducing balance loans

ANSWERS p. 713

### Recap

1 An investment is made at 4% interest per annum compounding fortnightly. The value of the investment at the end of the tenth year is $20879.13. How much money was originally invested?

A $7395 B $14000 C $18821 D $20560 E $20561

2 A workstation originally purchased for $8000 is depreciated each year using the reducing balance method. If the computer is valued at $3000 after five years, then the annual rate of depreciation is closest to

A -89% B 17% C 18% D 22% E 23%

### Mastery

3 Ed WORKED EXAMPLE 4~ Aisling has taken out a loan of $30000 with an interest rate of 15.4% per annum compounding quarterly. She makes regular quarterly payments of $230.

a Write a recurrence relation for the balance.

b Describe the sort of growth or decay modelled by the recurrence relation.

c Sketch the shape of the graph of the recurrence relation.

4 H WORKED EXAMPLE 5 Ishmael has taken out a reducing balance loan, compounding monthly and with regular monthly payments, according to the recurrence relation:

$V_o = 15000, V_{n+1} = 1.006\ V_n - 465$

where $V_n$ is the value of the loan after *n* compounding periods.

a How much money did Ishmael borrow?

b How much are the regular monthly payments?

c Use recursion to write down calculations that show that the amount owing on his loan after three months will be $13868, rounded to the nearest dollar.

d What is the annual percentage compound interest rate for this loan?

e After how many months will the balance of Ishmaels loan fall below $12 500?

▶ 50 WORKED EXAMPLE 6 Praveen has taken out a housing loan of $380 000 with interest compounding monthly. He pays monthly instalments of $3854. The amortisation table below shows the first few calculations for his loan.

| Payment number | Payment | Interest | Principal reduction | Balance |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 380000.00 |
| 1 | 3854.00 | $\frac{r}{100}$ x 380000.00 = 2850.00 | 3854.00 - 2850.00 = 1004.00 | 380000.00 - 1004.00 = 378996.00 |
| 2 | 3854.00 | $\frac{r}{100}$ x 378996.00 = 2842.47 | | |
| 3 | | | | |
| 4 | | | | |

a Use the table to show two calculations that will give r, the percentage interest rate per compounding period.

b What is the nominal interest rate for the loan?

c Complete the amortisation table, showing the interest paid, the principal reduction and the balance for the first four months of the loan, giving all values to the nearest cent.

6 Ej WORKED EXAMPLE 7j Chester has taken out a short-term personal loan of $10000 for a car. The interest rate is 16% p.a. compounding quarterly and it needs to be repaid in four quarterly payments of $2750. Answer the following questions using the amortisation table of the loan given below, which has some missing entries.

| Payment number | Payment | Interest | Principal reduction | Balance |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 10000.00 |
| 1 | 2750.00 | 400.00 | 2350.00 | 7650.00 |
| 2 | 2750.00 | 306.00 | 2444.00 | 5206.00 |
| 3 | 2750.00 | 208.24 | 2541.76 | |
| 4 | 2750.00 | | | 20.81 |

a What is the balance after three payments?

b How much has the principal been reduced by at payment number 4?

c How much interest was paid at payment number 4?

d How much of the principal has been paid after three payments?

e How much has the loan cost Chester?

f Explain how you know the loan isn’t paid out fully after the four payments.

g If Chester wants to pay out the loan fully with payment number 4, what should be the value of this payment?

▶


9780170448406

▶ 7H WORKED EXAMPLE 8 J

a Venice has taken out a loan with a recurrence relation for $V_n$, the value of the loan after $n$ compounding periods, of

$V_0 = 80000, \quad V_{n+1} = 1.0036V_n - 288$

Find Vj and $V_2$ and explain why this shows that this models an interest-only loan.

b Jim has taken out an interest-only loan with an interest rate of 5.2% per annum, compounding fortnightly with regular fortnightly payments of $400. How much did he borrow?

c Myrtle has taken out a $650000 interest-only loan. The interest rate is 4.2% per annum compounded quarterly. What are her regular quarterly repayments?

## Exam practice

80-100% 60-79% 0-59%

8 ^^^^2OITICQI7^ 82% The value of a reducing balance loan, in dollars, after $n$ months, $V_n$, can be modelled by the recurrence relation shown.

$V_0 = 26\,000, \; V_{n+1} = 1.003\, V_n - 400$

What is the value of this loan after five months?

A $24380.31 B $24706.19 C $25031.10 D $25355.03 E $25678.00

*Use the following information to answer the next two questions.*

Shirley would like to purchase a new home. She will establish a loan for $225 000 with interest charged at the rate of 3.6% per annum, compounding monthly. Each month, Shirley will pay only the interest charged for that month.

9 ©VCAA 20171CQ19J 61% After three years, the amount that Shirley will owe is

A $73362 B $170752 C $225000 D $239605 E $245865

10 ©VCAA 20171CQ20 J 46% Let $V_n$ be the value of Shirley's loan, in dollars, after $n$ months. A recurrence relation that models the value of $V_n$ is

A $V_0 = 225000, \quad V_{n+1} = 1.003\,V_n$

B $V_0 - 225\,000, \quad V_{n+1} - 1.036V_n$

C $V_0 = 225000, \quad V_{n+1} - 1.003\,V_n - 8100$

D $V_0 - 225\,000, \quad V_{n+1} - 1.003\,V_n - 675$

E $V_0 = 225000, \quad V_{n+1} - 1.036V_n - 675$

11 ©VCAA 2019 1CQ20 58% Consider the following amortisation table for a reducing balance loan.

| Payment number | Payment | Interest | Principal reduction | Balance of loan |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 300000.00 |
| 1 | 1050.00 | 900.00 | 150.00 | 299850.00 |
| 2 | 1050.00 | 899.55 | 150.45 | 299699.55 |
| 3 | 1050.00 | 899.10 | 150.90 | 299548.65 |

The annual interest rate for this loan is 3.6%. Interest is calculated immediately before each payment. For this loan, the repayments are made

A weekly B fortnightly C monthly D quarterly E yearly

12 ©VCAA 20161CQ22 54% The first three lines of an amortisation table for a reducing balance home loan are shown. The interest rate for this home loan is 4.8% per annum compounding monthly. The loan is to be repaid with monthly payments of $1500.

| Payment number | Payment | Interest | Principal reduction | Balance of loan |
|---|---|---|---|---|
| 0 | 0 | 0.00 | 0.00 | 250000.00 |
| 1 | 1500 | 1000.00 | 500.00 | 249500.00 |
| 2 | 1500 | | | |

The amount of payment number 2 that goes towards reducing the principal of the loan is

A $486 B $502 C $504 D $996 E $998

13 ©VCAA | 20051BRMQ5 35% An investor borrows $200 000 for five years to buy an apartment. The interest rate is 8.5% per annum compounding monthly. It is an interest-only loan, that is, at the end of five years, the investor will still owe $200 000. He is required to make monthly repayments. Correct to the nearest cent, his monthly repayment will be

A $666.67 B $1416.67 C $1757.67 D $4103.31 E $6789.95

14 ©VCAA 2018 1CQ23 29% Five lines of an amortisation table for a reducing balance loan with monthly repayments are shown below.

| Repayment number | Repayment | Interest | Principal reduction | Balance of loan |
|---|---|---|---|---|
| 25 | $2200.00 | $972.24 | $1227.76 | $230256.78 |
| 26 | $2200.00 | $967.08 | $1232.92 | $229023.86 |
| 27 | $2200.00 | $961.90 | $1238.10 | $227785.76 |
| 28 | $2200.00 | $1002.26 | $1197.74 | $226588.02 |
| 29 | $2200.00 | $996.99 | $1203.01 | $225385.01 |

The interest rate for this loan changed immediately before repayment number 28. This change in interest rate is best described as

A an increase of 0.24% per annum.

B a decrease of 0.024% per annum.

C an increase of 0.024% per annum.

D a decrease of 0.0024% per annum.

E an increase of 0.000 24% per annum.

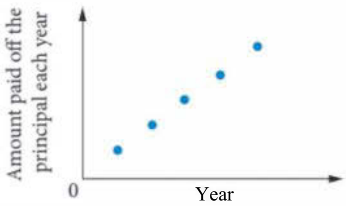
15 ^^^3 2019N1CQ21 Yazhen has a reducing balance loan. Six lines of the amortisation table for Yazhen's loan are shown below.

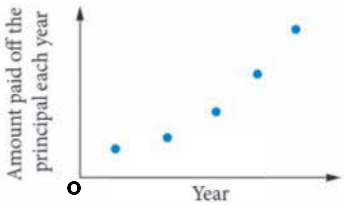
| Payment number | Payment | Interest | Principal reduction | Balance of loan |
|---|---|---|---|---|
| 15 | 1000.00 | 349.50 | 650.50 | 34299.50 |
| 16 | 1000.00 | 343.00 | 657.00 | 33642.50 |
| 17 | 1000.00 | 403.71 | 596.29 | 33046.21 |
| 18 | 1000.00 | 396.55 | 603.45 | 32442.76 |
| 19 | 1000.00 | 389.31 | 610.69 | 31832.07 |
| 20 | 1000.00 | 381.98 | 618.02 | 31214.05 |

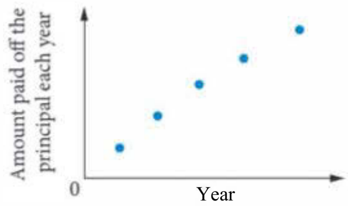
The interest rate for Yazhen's loan increased after one of these six repayments had been made. The first repayment made at the higher interest rate was repayment number

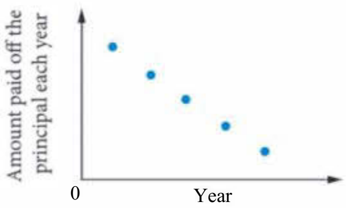
A 15 B 16 C 17 D 18 E 19

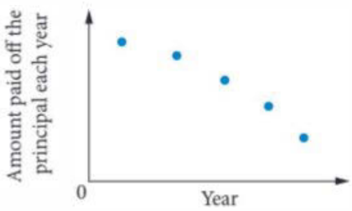
16 ©VCAA 2008 1BRMQ7 25% Ernie took out a reducing balance loan to buy a new family home. He correctly graphed the amount paid off the principal of his loan each year for the first five years. The shape of this graph (for the first five years of the loan) is best represented by

A


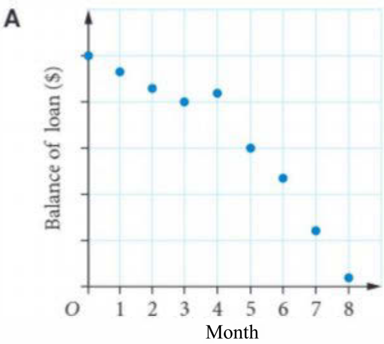

B


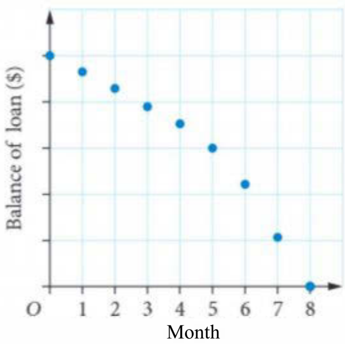

C


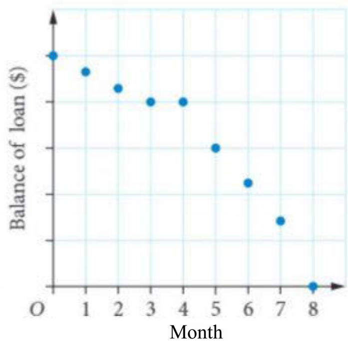

D


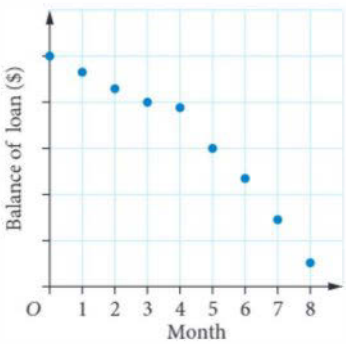

E


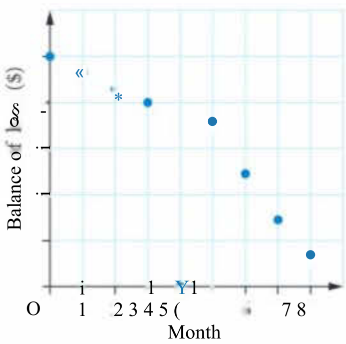

17 ©VCAA 2012 1BRMQ9 20% Peter took out a reducing balance loan where interest was calculated monthly. He planned to repay this loan fully, with eight equal monthly payments of $260. Peter missed the fourth payment, but made a double payment of $520 in the fifth month. He then continued to make payments of $260 for the remaining three months. Which graph could show the balance of the loan each month over the eight-month period?

A



B



C



D



E


18 © VCAA 2020 2CQ8 J (3 marks) Samuel has a reducing balance loan. The first five lines of the amortisation table for Samuels loan are shown below.

| Payment number | Payment | Interest | Principal reduction | Balance of loan |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 320000.00 |
| 1 | 1600.00 | 960.00 | 640.00 | 319360.00 |
| 2 | 1600.00 | 958.08 | 641.92 | 318718.08 |
| 3 | 1600.00 | 956.15 | | 318074.23 |
| 4 | 1600.00 | | | |

Interest is calculated monthly and Samuel makes monthly payments of $1600. Interest is charged on this loan at the rate of 3.6% per annum.

a Using the values in the amortisation table

i 85% calculate the principal reduction associated with payment number 3 — 1 mark

ii 39% calculate the balance of the loan after payment number 4 is made. Round your answer to the nearest cent. — 1 mark

b 30% Let $S_n$ be the balance of Samuel s loan after *n* months. Write down a recurrence relation, in terms of $S_0$, $S_{n+1}$ and $S_n$, that could be used to model the month-to-month balance of the loan. — 1 mark

# 6.3 Using finance solvers for reducing balance loans

Video playlist Using finance solvers for reducing balance loans

Worksheet Loan repayment problems

## Reducing balance loans and finance solvers

At the start of this chapter we used CAS finance solvers to answer questions about compound interest investments and reducing balance depreciation, which don't involve regular payments. We will now look at how to use finance solvers for financial questions where payments are involved, starting with reducing balance loans.

### Using finance solvers for reducing balance loans

| | |
|---|---|
| N | Total number of compounding periods |
| I% | Interest rate per year |
| PV | Present value for a loan is positive because the money has come to the person. |
| Pmt or PMT | Regular payments for a loan are negative because the money is going away from the person. |
| FV | Future value has the opposite sign of the present value so it will be negative. A future value of zero means the loan has been fully repaid. |
| PpY or P/Y | Number of payments per year. This will always take the same value as CpY or C/Y. |
| CpY or C/Y | Number of compounding periods per year |

When solving for N, always round *up,* never down, to the nearest whole number.

Ignore the negative sign in values when using the following formulas:

total interest paid = N x Pmt - (PV - FV)

total loan cost = N x Pmt

$$\text{percentage decrease in loan balance} = \frac{PV - FV}{PV} \times 100\%$$


9780170448406

6.3

WB p. 100

## WORKED EXAMPLE 9 Using finance solvers for annual loan interest rates

Shari's reducing balance loan of $200 000 is to be fully repaid over 25 years with quarterly repayments of $3985.14. Find the interest rate per annum correct to one decimal place and explain why you need to enter the payment as a negative.

| Steps | Working | |
|---|---|---|
| | TI-Nspire + ClassPad | |
| 1 Find I, given there are 25 x 4 = 100 quarters. | | |
| 2 Total number of compounding periods | N | 100 |
| Annual interest rate | I% | 6.2999921767059 |
| Present value for a loan is positive. | PV | 200000 |
| Money moving away from a person is negative. | Pmt or PMT | -3985.14 |
| Future value is zero when a loan is fully repaid. | FV | 0 |
| Same as CpY or C/Y | PpY or P/Y | 4 |
| Number of compounding periods per year | CpY or C/Y | 4 |
| 3 After entering all the other values, the *I%* field displays the annual interest rate. Round to the correct number of decimal places. | The annual interest rate is 6.3%. | |
| 4 Is the money moving to the person or away from the person? | The value is negative because Shari is making payments to the bank, so the money is moving *away* from her. | |

WB p. 101

## WORKED EXAMPLE 10 Using finance solvers for loan repayments, total loan cost and total interest paid

Dougal borrowed $312 000 at 5.5% per annum for 25 years with monthly repayments and interest compounding monthly. Find, to the nearest cent,

| Steps | Working | |
|---|---|---|
| **a** the repayment required to repay the loan in full. | | |
| | TI-Nspire + ClassPad | |
| **1** Find Pmt, given there are 25 × 12 = 300 months. | | |
| **2** Total number of compounding periods | N | 300 |
| Annual interest rate | I% | 5.5 |
| Present value for a loan is positive. | PV | 312 000 |
| Money moving away from a person is negative. | Pmt or PMT | −1915.9529759182 |
| Future value is zero when a loan is fully repaid. | FV | 0 |
| Same as CpY or C/Y | PpY or P/Y | 12 |
| Number of compounding periods per year | CpY or C/Y | 12 |
| **3** Write the answer correct to the nearest cent. | Dougal's monthly repayment is $1915.95. | |
| **b** the total cost of the loan. | | |
| Total loan cost = N × Pmt | Total loan amount repaid<br>= N × Pmt | |
| Ignore negative sign in the Pmt value.<br>Round to the nearest cent. | = 300 × $1915.95<br>= $574 785.00 | |

c the total interest paid on the loan.

| | |
|---|---|
| Total interest paid | Amount of interest paid |
| = N x Pmt - (PV-FV) | -NxPmt-(PV-FV) |
| Ignore the negative signs in values. | - 300 x $1915.95 - (312 000 - 0) |
| Round to the nearest cent. | -574785.00-312000 |
| | -$262785.00 |

WB

p. 102

## WORKED EXAMPLE 11 Using finance solvers for amount owed and number of loan repayments

Jordan takes out a loan of $30000 at a rate of 8.5% per annum, compounding monthly, with repayments of $650 per month.

| Steps | Working | |
|---|---|---|
| **a** How much is still owing on Jordan's loan after two years to the nearest cent? | | |
| | TI-Nspire | ClassPad |
| 1 Find FV after 2 x 12 = 24 months. | | |
| 2 Total number of compounding periods | N | 24 |
| Annual interest rate | I% | 8.5 |
| Present value for a loan is positive. | **PV** | 30000 |
| Money moving away from a person is negative. | Pmt or PMT | -650 |
| Future value has the opposite sign to present value. | **FV** | -18598.558692278 |
| Same as CpY or C/Y | PpY or P/Y | 12 |
| Number of compounding periods per year | CpY or C/Y | 12 |
| 3 Round to the nearest cent. | After two years, Jordan still owes $18598.56. | |
| **b** By what percentage has Jordan reduced the balance of his loan? Round to the nearest percentage. | | |
| Percentage decrease in loan balance $= \dfrac{PV - FV}{PV} \times 100\%$. | $\dfrac{30000-18598.56}{30000} \times 100\%$ | |
| Ignore the negative signs in values. | = 0.380048x100% | |
| Round to the nearest percentage. | = 38% | |
| **c** How many repayments are required in order for him to repay the loan in full? | | |
| 1 Find N when FV = 0. | | |
| 2 Total number of compounding periods | N | 56.08886655355 |
| Annual interest rate | I% | 8.5 |
| Present value for a loan is positive. | **PV** | 30000 |
| Money moving away from a person is negative. | Pmt or PMT | -650 |
| Future value is zero when a loan is fully repaid. | FV | 0 |
| Same as CpY or C/Y | PpY or P/Y | 12 |
| Number of compounding periods per year | CpY or C/Y | 12 |
| 3 When solving for N, always round *up,* never down, to the nearest whole number. | Jordan will need to make 57 repayments. So 56 repayments won't be enough. The 57th repayment will be less than the usual $650. | |



9780170448406

d How much is still owing on Jordan's loan after 56 repayments to the nearest cent?

| | | |
|---|---|---|
| 1 Find FV after 56 months. | | |
| 2 Total number of compounding periods | N | 56 |
| Annual interest rate | I% | 8.5 |
| Present value for a loan is positive. | PV | 30000 |
| Money moving away from a person is negative. | Pmt or PMT | -650 |
| Future value has the opposite sign to present value. | FV | -57.541 59411493 |
| Same as CpY or C/Y | PpY or P/Y | 12 |
| Number of compounding periods per year | CpY or C/Y | 12 |
| 3 After entering all the other values, the FV field displays the amount still owing. Round to the nearest cent. | Jordan still owes $57.54. | |

e Jordan decides to adjust the value of the 56th repayment so that the loan is fully repaid. What will the adjusted 56th payment be to the nearest cent?

For the last payment, add the amount still owing on the loan to the regular payment, rounding to the nearest cent.

$650 + $57.54 = $707.54

?o repay the loan in full the final payment will be $707.54.

## Interest-only loans and finance solvers

Finance solvers can also be used to answer questions about interest-only loans where the balance of the loan doesn't change because the payments are exactly equal to the interest for each compounding period.

### Using finance solvers for interest-only loans

For an interest-only loan:

- all compounding periods are the same so N = 1
- FV and PV have the same value but opposite signs so FV —PV.

### WORKED EXAMPLE 12 Using finance solvers for interest-only loans

WB p. 104

Use finance solvers to find each of the following.

| Steps | Working | |
|---|---|---|
| a Savaratnam has a $100 000 interest-only loan compounding monthly with monthly payments of $348. What is the annual interest rate correct to one decimal place? | | |
| | TI-Nspire + ClassPad | |
| 1 Find I when the value of the loan stays the same. | | |
| 2 All compounding periods are the same for interest-only loans. | N | 1 |
| Annual interest rate | I% | 4.176 |
| Present value for a loan is positive. | PV | 100000 |
| Money moving away from a person is negative. | Pmt or PMT | -348 |
| FV = -PV for interest-only loans | FV | -100000 |
| Same as CpY or C/Y | PpY or P/Y | 12 |
| Number of compounding periods per year | CpY or C/Y | 12 |
| 3 Round your answer to one decimal place. | The annual interest rate is 4.2%. | |

b Voula has taken out a $300 000 interest-only loan. The interest rate is 4.3% per annum compounding weekly. What are Voula's regular weekly repayments to the nearest cent?

| | | | |
|---|---|---|---|
| 1 | Find Pmt when the value of the loan stays the same. | | |
| 2 | All compounding periods are the same for interest-only loans. | N | 1 |
| | Annual interest rate | I% | 4.3 |
| | Present value for a loan is positive. | PV | 300000 |
| | Money moving away from a person is negative. | Pmt or PMT | -248.07692307692 |
| | FV = -PV for interest-only loans | FV | -300000 |
| | Same as CpY or C/Y | PpY or P/Y | 52 |
| | Number of compounding periods per year | CpY or C/Y | 52 |
| 3 | Round your answer to the nearest cent. | Voula's weekly repayments are $248.08. | |

## EXERCISE 6.3 Using finance solvers for reducing balance loans

ANSWERS p. 713

### Recap

1 Marina borrows $200000 at a rate of 6% per annum, calculated monthly on the reducing balance. Each month, she repays $1700. If $A_n$ is the amount owing after the nth repayment, a recurrence relation that models the amount owing on this loan is

A $A_Q$ = 200000, $A_{M+1}$ = 0.06A,,-1700

B $A_o$=17OO, $A_{w+1}$ = 0.06A,, - 200000

C $AQ$ = 200000, $A_{,,+1}$ = 0.005A,, - 1700

D $A_o$ = 200000, $A_{,,+1}$ = 1.005A,, - 1700

E $A_o$ = 1700, $A_{,,+}$! = 1.005A,, - 200 000

2 © VCAA 2018N 1CQ17 The first three lines of an amortisation table for a reducing balance loan are shown below.

| Repayment number | Payment | Interest | Principal reduction | Balance |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 250000.00 |
| 1 | 945.00 | 791.67 | 153.33 | 249846.67 |
| 2 | 945.00 | 791.18 | | 249692.85 |

What is the principal reduction from repayment number 2?

A $153.33 B $153.82 C $791.21 D $791.67 E $945.00

### Mastery

3 R3 WORKED EXAMPLE 9 Find the interest rate per annum for each of the following reducing balance loans. Answer correct to one decimal place.

a Arthurs loan of $500000 is fully repaid over 20 years with monthly repayments of $3355.27.

b Celestes loan of $150000 is fully repaid over 12 years with monthly repayments of $1300.34.

c Florences loan of $250 000 is fully repaid over 20 years with quarterly repayments of $5642.56.

d Yayas loan of $500 000 is fully repaid over 25 years with fortnightly repayments of $4045.55.

e Why do you need to enter the payments into the finance solver as a negative?



9780170448406

4 H WORKED EXAMPLE io I For each of the following, find to the nearest cent

i the repayment required to repay the loan in full

ii the total cost of the loan

iii the total interest paid on the loan.

a Fidel borrowed $8000 at 6% per annum for five years with monthly repayments and interest compounding monthly.

b Kim borrowed $300000 at 7.5% per annum for 30 years with monthly repayments and interest compounding monthly.

c Hector borrowed $35 000 at 12% per annum for 10 years with quarterly repayments and interest compounding quarterly.

d Marisol borrowed $75 000 at 9.25% per annum for 10 years with fortnightly repayments and interest compounding fortnightly.

5H WORKED EXAMPLE 11 J

a Chantelle takes out a loan of $58000 at a rate of 9.5% per annum, compounding monthly, with repayments of $835 per month.

i How much is still owing on Chantelles loan after four years, correct to the nearest cent?

ii By what percentage has Chantelle reduced the balance of her loan, correct to the nearest percentage?

iii How many repayments are required in order for her to repay the loan in full?

iv How much is still owing on Chantelles loan after 101 repayments, correct to the nearest cent?

v Chantelle decides to adjust the value of the 101st repayment so that the loan is fully repaid. What will be the adjusted 101st payment, correct to the nearest cent?

b Briony takes out a loan of $40000 at a rate of 11.2% per annum, compounding fortnightly, with repayments of $250 per fortnight.

i How much is still owing on Brionys loan after three years, correct to the nearest cent?

ii By what percentage has Briony reduced the balance of her loan, correct to the nearest percentage?

iii How many repayments are required in order for her to repay the loan in full?

iv How much is still owing on Briony's loan after 271 repayments, correct to the nearest cent?

v Briony decides to adjust the value of the 271st repayment so that the loan is fully repaid. What will be the adjusted 271st payment, correct to the nearest cent?

6H WORKED EXAMPLE 12 J Use finance solvers to find each of the following.

a Senad has a $92 000 interest-only loan compounding weekly with weekly payments of $80. What is the annual interest rate, correct to one decimal place?

b Cedric has taken out a $268 000 interest-only loan. The interest rate is 5.25% per annum compounding quarterly. What are Cedric's regular quarterly repayments, correct to the nearest cent?

c Taiko has a $173 000 interest-only loan compounding fortnightly with fortnightly payments of $210. What is the annual interest rate, correct to one decimal place?

d Luisa has taken out a $500000 interest-only loan. The interest rate is 3.7% per annum compounding six-monthly. What are Luisa's regular six-monthly repayments, correct to the nearest cent?

## Exam practice

80-100% 60-79% 0-59%

7 ©VCAA | 2009IBRMQ7~ 67% A loan of $17500 is to be paid back over four years at an interest rate of 6.25% per annum on a reducing monthly balance. The monthly repayment, correct to the nearest cent, will be

A $364.58 B $413.00 C $802.08 D $1156.77 E $5079.29

8 ©VCAA | 20101BRMQ7 64% A loan of $300000 is to be repaid over a period of 20 years. Interest is charged at the rate of 7.25% per annum compounding quarterly. The quarterly repayment to the nearest cent is

A $2371.13 B $5511.46 C $7113.39 D $7132.42 E $7156.45

9 © VCAA 2003l BRMQ4 1 45% Swee borrowed $150000 at 6.2% per annum compounding monthly. The repayments are $1100 per month. The balance of the loan at the end of five years is closest to

A $0 B $84000 C $127000 D $137000 E $148000

### Exam hack

This loan has compounding and regular repayments, so you should know it's a reducing balance loan even though this isn't mentioned.

10 ©VCAA | 20091BRMQ8 45% Robin takes out a reducing balance loan of $100 000 with quarterly repayments of $2150. After seven years of quarterly repayments, Robin still owes $80000. Correct to one decimal place, the interest rate per annum for this loan is

A 6.3% B 8.2% C 12.9% D 18.9% E 24.7%

11 ©VCAA | 20191CQ23 J 38% Joseph borrowed $50 000 to buy a new car. Interest on this loan is charged at the rate of 7.5% per annum, compounding monthly. Joseph will fully repay this loan with 60 monthly repayments over five years. Immediately after the 59th repayment is made, Joseph still owes $995.49. The value of his final repayment, to the nearest cent, will be

A $995.49 B $998.36 C $1001.71 D $1001.90 E $1070.15

12 ©VCAA | 20151BRMQ8 33% Cindy took out a reducing balance loan of $8400 to finance an overseas holiday. Interest was charged at a rate of 9% per annum, compounding quarterly. Her loan is to be fully repaid in six years, with equal quarterly payments. After three years, Cindy will have reduced the balance of her loan by approximately

A 9% B 35% C 43% D 50% E 57%



9780170448406

6.3

13 ©VCAA 2018ICQ22 I 29% Adam has a home loan with a present value of $175 260.56. The interest rate for Adams loan is 3.72% per annum, compounding monthly. His monthly repayment is $3200. The loan is to be fully repaid after five years. Adam knows that the loan cannot be exactly repaid with 60 repayments of $3200. To solve this problem, Adam will make 59 repayments of $3200. He will then adjust the value of the final repayment so that the loan is fully repaid with the 60th repayment. The value of the 60th repayment will be closest to

A $368.12 B $2831.88 C $3200.56 D $3557.09 E $3568.12

14 ©VCAA 2002 2BRMQ4b~] (2 marks) Sally wants to borrow $20 000 for four years. Interest is calculated quarterly on the reducing balance at an annual rate of 8%. Sally can afford to repay this loan at $1500 per quarter. Will this enable her to repay the loan in four years? Explain.

15 ©VCAA | 2012 2BRMQ3 MODIFIED I (4 marks) An area of a club needs to be refurbished. $40 000 is borrowed at an interest rate of 7.8% per annum. Interest on the unpaid balance is charged to the loan account monthly. Suppose the $40 000 loan is to be fully repaid in equal monthly instalments over five years.

a Determine the monthly payment, correct to the nearest cent. 1 mark

b If, instead, the monthly payment was $1000, how many months will it take to fully repay the $40 000? 1 mark

c Suppose no payments are made on the loan in the first 12 months.

i What will be the balance of the loan account after the first 12 months, correct to the nearest dollar? 1 mark

ii After the first 12 months, only the interest on the loan is paid each month. Determine the monthly interest payment, correct to the nearest cent. 1 mark

16 ©VCAA | 2O2O2CQ11~I 21% (2 marks) Samuel took out a reducing balance loan. The interest rate for this loan was 4.1% per annum, compounding monthly. The balance of the loan after four years of monthly repayments was $329 587.25. The balance of the loan after seven years of monthly repayments was $280875.15. Samuel will continue to make the same monthly repayment. To ensure the loan is fully repaid, to the nearest cent, the required final repayment will be lower. In the first seven years, Samuel made 84 monthly repayments. From this point on, how many more monthly repayments will Samuel make to repay the loan in full?

# Changing the terms of reducing balance loans

Video playlist
Changing the terms of reducing balance loans

The terms of a reducing balance loan can change mid-loan. Not all interest rates are fixed. Many types of loans have interest rates that increase or decrease due to economic factors. A borrower can also negotiate the frequency of their repayments or the amount of the repayment. Changes to terms have an effect on the length of a loan, the total interest paid, and the total cost of the loan. Finance solvers can be used to deal with these sorts of changes.

WB p. 105

**WORKED EXAMPLE 13** Changing the length of the loan

Allison wants to buy a house and has borrowed $225 000 at an interest rate of 4.23% per annum, fixed for 10 years. Interest is calculated monthly and monthly repayments are set at $1279. After 10 years, Allison renegotiates the conditions for the balance of her loan. The new interest rate will be 4.05% per annum. She will pay $1620 per month. How many years will it take her to repay the loan in full?

| Steps | Working | |
|---|---|---|
| | TI-Nspire | ClassPad |
| 1 Find FV after 10 x 12 = 120 months for first part of loan. | | |
| 2 Total number of compounding periods. | N | 120 |
| Annual interest rate for first part of loan. | I% | 4.23 |
| Present value for a loan is positive. | PV | 225000 |
| Money moving away from a person is negative. | Pmt or PMT | -1279 |
| Future value has the opposite sign to present value. | FV | -152580.60044351 |
| Same as CpY or C/Y | PpY or P/Y | 12 |
| Number of compounding periods per year | CpY or C/Y | 12 |
| 3 Use the unrounded FV value as the positive PV value to find N for second part of loan. | | |
| 4 Total number of compounding periods | N | 113.53759056301 |
| Annual interest rate for second part of loan. | I% | 4.05 |
| Present value for a loan is positive. | PV | 152580.60044351 |
| Money moving away from a person is negative. | Pmt or PMT | -1620 |
| Future value is zero when a loan is fully repaid. | FV | 0 |
| Same as CpY or C/Y | PpY or P/Y | 12 |
| Number of compounding periods per year | CpY or C/Y | 12 |
| 5 Always round up your answer for N. | The length of the renegotiated part of the loan is 114 months. | |
| Convert to years. | $\frac{114}{12} = 9.5$ years | |
| 6 Add the lengths of the two parts of the loan. | 10 + 9.5 = 19.5 | |
| | It will take 19½ years for Allison to repay the loan in full. | |



9780170448406

## WORKED EXAMPLE 14 Changing the interest rate

Yuki borrows $100 000 at 8% per annum compounding monthly, to be repaid over 10 years. The repayments for the first five years are $1213.28 each month. After five years, the interest rate is reduced to 7.5% per annum.

| Steps | Working | |
|---|---|---|
| a What is the new repayment to the nearest cent required for Yuki to repay the loan? | | |
| | ■ TI-Nspire ■ft Sgl ClassPad | |
| 1 Find FV after 5 x 12 = 60 months for first part of loan. | | |
| 2 Total number of compounding periods. | N | 60 |
| Annual interest rate | I% | 8 |
| Present value for a loan is positive. | PV | 100000 |
| Money moving away from a person is negative. | Pmt or PMT | -1213.28 |
| Future value has the opposite sign to present value. | FV | -59836.570684934 |
| Same as CpY or C/Y | PpY or P/Y | 12 |
| Number of compounding periods per year | CpY or C/Y | 12 |
| 3 Use the unrounded FV as the positive PV for the second part of the loan and find Pmt. | | |
| 4 Total number of compounding periods | N | 60 |
| Annual interest rate | I% | 7.5 |
| Present value for a loan is positive. | PV | 59836.570684934 |
| Money moving away from a person is negative. | Pmt or PMT | -1199.00 |
| Future value is zero when a loan is fully repaid. | FV | 0 |
| Same as CpY or C/Y | PpY or P/Y | 12 |
| Number of compounding periods per year | CpY or C/Y | 12 |
| 5 Write your answer to the nearest cent. | Over the last five years, the monthly repayments required to repay the loan are $1199.00. | |
| b How much total interest to the nearest cent has Yuki paid over the 10 years? | | |
| 1 Use the formula for both parts of the loan:<br>Total interest paid = N x Pmt - (PV - FV)<br>Ignore the negative sign in values. | Total interest paid for first five years<br>= 60 x 1213.28 - (100000 - 59836.57)<br>= 72796.80- 40163.43<br>= 32633.37<br>Total interest paid for second five years<br>= 60 x 1199.00 - (59836.57 - 0)<br>= 71940.00- 59836.57<br>= 12103.43 | |
| 2 Add the two totals. | Total interest over ten years<br>= 32633.37 + 12103.43<br>-$44736.80 | |

6.4

WB

p. 106

WB

p. 107

## WORKED EXAMPLE 15 Changing the repayments

Ryan has taken out a loan of $15 000 at the rate of 10.4% per annum, compounding monthly and with regular monthly payments, to pay for an overseas trip.

| Steps | Working | |
|---|---|---|
| a Ryan will make interest-only repayments for the first two years of this loan. How much is each interest-only repayment to the nearest dollar? | | |
| | TI-Nspire + ClassPad | |
| 1 Find Pmt when the value of the loan stays the same. | | |
| 2 All compounding periods are the same for interest-only loans. | N | 1 |
| Annual interest rate | I% | 10.4 |
| Present value for a loan is positive. | PV | 15000 |
| Money moving away from a person is negative. | Pmt or PMT | -130 |
| FV = -PV for interest-only loans | FV | -15000 |
| Same as CpY or C/Y | PpY or P/Y | 12 |
| Number of compounding periods per year | CpY or C/Y | 12 |
| 3 Write your answer to the nearest dollar. | Each interest-only repayment is $130. | |
| b For the next two years, Ryan will increase his monthly repayments so that the balance of the loan is $8320. What are Ryan's repayments, to the nearest cent, each month during these two years? | | |
| 1 Find Pmt for the next 2 x 12 = 24 months. | | |
| 2 Total number of compounding periods | N | 24 |
| Annual interest rate | I% | 10.4 |
| Present value for a loan is positive. | PV | 15000 |
| Money moving away from a person is negative. | Pmt or PMT | -381.58949056117 |
| Future value has the opposite sign to present value. | FV | -8320 |
| Same as CpY or C/Y | PpY or P/Y | 12 |
| Number of compounding periods per year | CpY or C/Y | 12 |
| 3 Write your answer to the nearest cent. | The repayments are $381.59. | |
| c Ryan will fully repay the outstanding balance of $8320 over the next three years. The first 35 monthly repayments will each be $270. The 36th repayment will have a different value to ensure the loan is repaid exactly to the nearest cent. What is the value of the 36th repayment, rounded to the nearest cent? | | |
| 1 Find FV after the next 3 x 12 = 36 months of the loan. | | |
| 2 Total number of compounding periods | N | 36 |
| Annual interest rate | I% | 10.4 |
| Present value for a loan is positive. | PV | 8320 |
| Money moving away from a person is negative. | Pmt or PMT | -270 |
| Future value has the opposite sign to present value. | FV | -1.18484779369 |
| Same as CpY or C/Y | PpY or P/Y | 12 |
| Number of compounding periods per year | CpY or C/Y | 12 |
| 3 For the last payment, add the amount still owing on the loan to the regular payment, rounding to the nearest cent. | $270 + $1.18 = $271.18<br>The adjusted value of the 36th repayment to exactly pay the loan is $271.18. | |



9780170448406

## EXERCISE 6.4 Changing the terms of reducing balance loans

ANSWERS p. 713

6.4

### Recap

1 ©VCAA | 2008IBRMQ8 J 53% A loan of $300 000 is taken out to finance a new business venture. The loan is to be repaid fully over 20 years with quarterly payments of $6727.80. Interest is calculated quarterly on the reducing balance. The annual interest rate for this loan is closest to

A 4.1% B 6.5% C 7.3% D 19.5% E 26.7%

2 ©VCAA 20021BRMQ4 I 50% Rho takes a 20-year loan of $172 000 at 6% per annum, compounding monthly and with monthly repayments. To repay the loan in full in 20 years, the amount he must repay each month is

A $716.67 B $1216.54 C $1232.26 D $9058.63 E $10320.00

### Mastery

3 EJ WORKED EXAMPLE 13 I Indira has borrowed $190000 to buy a house and she makes monthly repayments of $1470. The interest rate of 4.1% per annum is fixed for eight years and is calculated monthly. After eight years, Indira renegotiates the conditions for the balance of her loan. The new interest rate will be 3.9% per annum and she will pay $1530 per month. How many years will it take her to repay the loan in full?

4 WORKED EXAMPLE 14 J Jodie borrows $130000 at 6% per annum compounding monthly, to be repaid over 14 years. The repayments for the first seven years are $1352.73 each month. After seven years, the interest rate is reduced to 5.5% per annum.

a What are the new monthly repayments, to the nearest cent, required for Jodie to pay out the loan?

b How much total interest to the nearest cent has Jodie paid over the fourteen years?

5 WORKED EXAMPLE isj Saverio has taken out a loan of $19000 at the rate of 10.2% per annum, compounding monthly and with regular monthly payments, to pay for an overseas trip.

a Saverio will make interest-only repayments for the first two years of this loan. How much is each interest-only repayment to the nearest cent?

b For the next three years, Saverio will increase his monthly repayments so that the balance of the loan is $10200. What are Saverios repayments, to the nearest cent, each month during these three years?

c Saverio will fully repay the outstanding balance of $10200 over the next two years. The first 23 monthly repayments will each be $462. The 24th repayment will have a different value to ensure the loan is repaid exactly to the nearest cent. What is the value of the 24th repayment, rounded to the nearest cent?

### Exam practice

80-100% 60-79% 0-59%

6 ©VCAA | 20171CQ24] 58% Xavier borrowed $245 000 to pay for a house. For the first 10 years of the loan, the interest rate was 4.35% per annum, compounding monthly. Xavier made monthly repayments of $1800. After 10 years, the interest rate changed. If Xavier now makes monthly repayments of $2000, he could repay the loan in a further five years. The new annual interest rate for Xaviers loan is closest to

A 0.35% B 4.1% C 4.5% D 4.8% E 18.7%

7 ©VCAA 20091BRMQ9 I 45% To purchase a house Sam has borrowed $250000 at an interest rate of 4.45% per annum, fixed for 10 years. Interest is calculated monthly on the reducing balance of the loan. Monthly repayments are set at $1382.50. After 10 years, Sam renegotiates the conditions for the balance of his loan. The new interest rate will be 4.25% per annum. He will pay $1750 per month. The total time it will take him to pay out the loan fully is closest to

A 17 years. B 20 years. C 21 years. D 22 years. E 23 years.

8 ©VCAA | 20021BRMQ7 45% Ravi has a loan of $135 000 at 7% per annum interest, compounding monthly. The loan is to be repaid monthly over 20 years. The scheduled repayments are $1046.65 per month. However, he finds that he can afford to pay $1200 per month and decides to do so for the duration of the loan. The amount of time this will save in paying off the loan is closest to

A 6 months B 1 year C 5 years D 10 years E 15 years

9 ©VCAA | 2OIVIBRMQ8 , 43% Teresa borrowed $120 000 at an interest rate of 7.67% per annum, compounding monthly. The loan is to be repaid with equal monthly payments. She decides to repay the loan by making monthly payments of $1430. Which of the following statements is true?

A She will pay out the loan fully in less than 10 years.

B The amount of interest that she pays on the loan will increase each year.

C After four years the amount that she owes on the loan will be less than $80 000.

D Every monthly payment that she makes reduces the amount that she owes on the loan by the same amount.

E Monthly payments of $1560 (instead of $1430) will enable her to repay this loan in less than nine years.

10 ©VCAA | 2007 2BRMQ2 , 36% (3 marks) Khan decides to extend his home office and borrows $30000 for building costs. Interest is charged on the loan at a rate of 9% per annum compounding monthly. Assume Khan will pay only the interest on the loan at the end of each month.

a Calculate the amount of interest he will pay each month. 1 mark

Suppose the interest rate remains at 9% per annum compounding monthly and Khan pays $400 each month for five years.

b Determine the amount of the loan that is outstanding at the end of five years. Write your answer correct to the nearest dollar. 1 mark

Khan decides to repay the $30 000 loan fully in equal monthly instalments over five years. The interest rate is 9% per annum compounding monthly.

c Determine the amount of each monthly instalment. Write your answer correct to the nearest cent. 1 mark

11 ©VCAA 2018N 2CQ9 J (3 marks) Andrew borrowed $10 000 to pay for a holiday and other expenses. Interest on this loan will be charged at the rate of 12.9% per annum, compounding monthly. Immediately after the interest has been calculated and charged each month, Andrew will make a repayment.

a For the first year of this loan, Andrew will make interest-only repayments each month. What is the value of each interest-only repayment? 1 mark

b For the next three years of this loan, Andrew will make equal monthly repayments. After these three years, the balance of Andrew's loan will be $3776.15. What amount, in dollars, will Andrew repay each month during these three years? 1 mark

c Andrew will fully repay the outstanding balance of $3776.15 with a further 12 monthly repayments. The first 11 repayments will each be $330. The twelfth repayment will have a different value to ensure the loan is repaid exactly to the nearest cent. What is the value of the twelfth repayment? Round your answer to the nearest cent. 1 mark

▶ 12 ©VCAA 2011 2BRMQ4 J (3 marks) Tania takes out a reducing balance loan of $265 000 to pay for her house. Her monthly repayments will be $1980. Interest on the loan will be calculated and paid monthly at the rate of 7.62% per annum.

a i How many monthly repayments are required to repay the loan? Write your answer to the nearest month. 1 mark

ii Determine the amount that is paid off the principal of this loan in the first year. Write your answer to the nearest cent. 1 mark

Immediately after Tania made her twelfth payment, the interest rate on her loan increased to 8.2% per annum, compounding monthly. Tania decided to increase her monthly repayment so that the loan would be repaid in a further nineteen years.

b Determine the new monthly repayment. Write your answer to the nearest cent. 1 mark

13 ©VCAA 2016 2CQ7 J (4 marks) Ken has borrowed $70 000 to buy a new caravan. He will be charged interest at the rate of 6.9% per annum, compounding monthly.

a For the first year (12 months), Ken will make monthly repayments of $800.

i 46% Find the amount that Ken will owe on his loan after he has made 12 repayments. 1 mark

ii 26% What is the total interest that Ken will have paid after 12 repayments? 1 mark

b 11% After three years, Ken will make a lump sum payment of $L in order to reduce the balance of his loan. This lump sum payment will ensure that Kens loan is fully repaid in a further three years. Kens repayment amount remains at $800 per month and the interest rate remains at 6.9% per annum, compounding monthly. What is the value of Kens lump sum payment, $L? Round your answer to the nearest dollar. 2 marks

# 6.5 Annuities

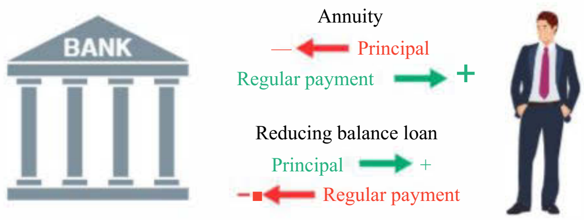
Video playlist Annuities

Worksheet Annuity problems

## Annuity recurrence relations

An annuity is a compound interest investment with regular payments to the investor called withdrawals. It is exactly the same as a reducing balance loan except the bank is borrowing from the person rather than the person borrowing from the bank.

The recurrence relation for an annuity is the same as the recurrence relation for a reducing balance loan.




### Annuity recurrence relation

The recurrence relation for the value $V_n$ of an annuity

$$V_0 = \text{principal}, \quad V_{n+1} = \left(1 + \frac{r}{100}\right)V_n - d$$

where

$r$ = percentage interest rate per compounding period

$n$ = number of compounding periods

$d$ = payment made per compounding period.

The graph of an annuity recurrence relation will look like this.

WB

p. 108

## WORKED EXAMPLE 16 Working with annuity recurrence relations

Serena plans to invest \$40 000 in an annuity at a rate of 8.4% per annum compounding quarterly where she makes regular quarterly withdrawals of \$300.

**a** Write a recurrence relation for the balance.

Serena changes her mind and decides to invest the \$40 000 in an annuity compounding six-monthly with six-monthly withdrawals that has the recurrence relation

$V_0 = 40\,000, \quad V_{n+1} = 1.06V_n - 300$

**b** What is the annual interest rate of this investment?

**c** How much has the investment increased between the third and fourth six-monthly periods, to the nearest cent?

| Steps | Working |
|---|---|
| **a** 1 Find the number of compounding periods per year. | There are four compounding periods per year. |
| 2 Identify $V_n$, $V_0$, $r$ and $d$. | Let $V_n$ = the balance after $n$ compounding periods. $V_0 = 40\,000,\ r = \frac{8.4}{4} = 2.1\%,\ d = 300$ |
| 3 Substitute the values into these formulas and simplify. $V_0$ = principal $V_{n+1} = \left(1 + \frac{r}{100}\right)V_n - d$ | $V_{n+1} = \left(1 + \frac{2.1}{100}\right)V_n - 300$ $= (1 + 0.021)V_n - 300$ $= 1.021V_n - 300$ $V_0 = 40\,000, \quad V_{n+1} = 1.021V_n - 300$ |
| **b** 1 Use the recurrence relation to find an equation for $r$, the percentage interest rate per compounding period. | Comparing $V_{n+1} = \left(1 + \frac{r}{100}\right)V_n - d$ $V_{n+1} = 1.06V_n - 300$ we can see that $1 + \frac{r}{100} = 1.06$ |
| 2 Solve for $r$, using CAS if necessary. | $\frac{r}{100} = 1.06 - 1$ $\frac{r}{100} = 0.06$ $r = 6\%$ per six months. |
| 3 Multiply $r$ by the compounding period to find the annual percentage compound interest rate. | The interest compounds six-monthly, so the annual percentage compound interest rate is $r \times 2 = 6 \times 2$ $= 12\%$ |



9780170448406

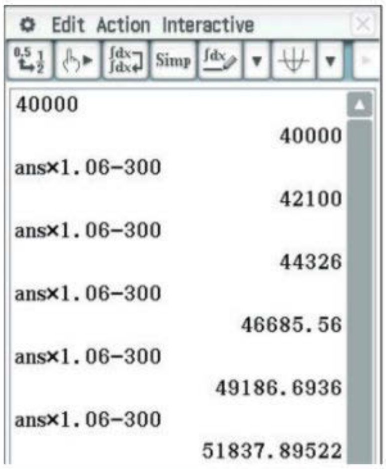
c Use CAS recursive computation to find the balances, and subtract to find the difference, rounding to the nearest cent.

$V_4 - V_3 = 49186.6936 - 46685.56$
$= 2501.1336$

The investment increased by $2501.13.

[TI-Nspire]


ClassPad




## Annuity amortisation tables

Amortisation tables work the same way for annuities as for reducing balance loans.

### Annuity amortisation table

$$r = \frac{\text{interest}}{\text{previous balance}} \times 100$$

| Payment number | Payment | Interest | Principal reduction | Balance |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | Vo |
| $n$ | $d$ | $\frac{r}{100}$ x previous balance | payment - interest | previous balance - principal reduction |

where

$n$ = payment number

$d$ = payment made per compounding period

$r$ = percentage interest rate per compounding period

$V_0$ = principal.

WB

p. 110

## WORKED EXAMPLE 17 Analysing annuity amortisation tables

The following amortisation table shows the first three payments of an annuity with monthly compounding interest and monthly withdrawals. Some of the entries are missing.

| Payment number | Payment | Interest | Principal reduction | Balance |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 450000.00 |
| 1 | 3797.00 | 2250.00 | 1547.00 | |
| 2 | 3797.00 | 2242.27 | 1554.73 | 446898.27 |
| 3 | 3797.00 | | | 445335.76 |

Find the following from the table:

| **Steps** | **Working** |
|---|---|
| a the amount invested | |
| Read the principal from the table, | amount invested = \$450 000.00 |
| b the regular monthly withdrawal | |
| Read the regular payment from the table. | regular monthly withdrawal = \$3797.00 |
| c the interest rate per compounding period | |
| $r = \frac{\text{interest}}{\text{previous balance}} \times 100$<br>Choose the option with the easiest calculations. | Interest rate per compounding period is<br>$r = \frac{2250.00}{450000.00} \times 100 = 0.5\%$ |
| d the interest rate per annum | |
| Multiply $r$ by the compounding period. | interest rate per annum<br>= 12 x 0.5<br>= 6% |
| e the interest paid at payment 3 | |
| interest $= \frac{r}{100} \times$ previous balance | $r = 0.5$<br>interest for payment 3<br>$= \frac{0.5}{100} \times 446898.27$<br>= 0.005 x 446898.27<br>= \$2234.95 |
| f the amount by which the principal has been reduced at payment 3 | |
| principal reduction = payment - interest | principal reduction at payment 3<br>- 3797.00 - 2234.95<br>= \$1562.05 |
| g the balance after one withdrawal. | |
| balance = previous balance - principal reduction | balance after one withdrawal<br>= 450000.00 - 1547.00<br>= \$448453 |



9780170448406

## Annuities and finance solvers

Annuities can be analysed using finance solvers in a similar way to reducing balance loans. The difference is:

- an annuity is an investment, not a loan, so the PV is always negative
- PMT is always positive because the payments are coming to the person.

6.5

### Using finance solvers for annuities

| | |
|---|---|
| N | Total number of compounding periods |
| 1% | Interest rate per year |
| PV | Present value for an annuity is negative because the money is going away from the person. |
| Pmt or PMT | Regular payments for an annuity are positive because the money is coming to the person. |
| FV | Future value has the opposite sign of the present value so it will be positive. A future value of zero means the investment has been fully paid out. |
| PpY or P/Y | Number of payments per year. This will always take the same value as CpY or C/Y. |
| CpY or C/Y | Number of compounding periods per year |

When solving for N, always round *up,* never down, to the nearest whole number.

Ignore the negative sign in values when using the following formula:

total interest paid = N x Pmt - (PV - FV)

### WORKED EXAMPLE 18 Calculating how long an annuity will last

WB p. in

Roger purchases a $400 000 annuity. Interest is paid at 8% per annum compounding monthly. If he receives monthly payments of $3500, how many years will the annuity last?

| Steps | Working | |
|---|---|---|
| | TI-Nspire + ClassPad | |
| 1 Find N when FV = 0. | | |
| 2 Total number of compounding periods | N | 215.97942643326 |
| Annual interest rate | I% | 8 |
| Present value for an investment is negative. | PV | -400000 |
| Money moving to a person is positive. | Pmt or PMT | 3500 |
| Future value is zero when an investment is fully paid out. | FV | 0 |
| Same as CpY or C/Y | PpYorP/Y | 12 |
| Number of compounding periods per year. | CpY or C/Y | 12 |
| 3 Always round N *up* to the next whole number. | The annuity will last for 216 months. | |
| 4 Convert to years and answer the question. | 216 months $= \frac{216}{12} = 18$ years<br>The annuity will last 18 years. | |

WB

p. 112

**WORKED EXAMPLE 19** Calculating how much to withdraw from an annuity

Caroline invests $350 000 in an annuity with interest paid at 7.5% per annum compounding monthly. She receives monthly payments from this investment. What monthly payment will she receive, to the nearest cent, if she wishes to receive payments for 15 years?

| Steps | Working | |
|---|---|---|
| | TI-Nspire + ClassPad | |
| 1 Find Pmt, given there are 15 x 12 = 180 months. | | |
| 2 Total number of compounding periods | N | 180 |
| Annual interest rate | I% | 7.5 |
| Present value for an investment is negative. | PV | -350000 |
| Money moving to a person is positive. | Pmt or PMT | 3244.5432600096 |
| Future value is zero when an investment is fully paid out. | FV | 0 |
| Same as CpY or C/Y | PpYorP/Y | 12 |
| Number of compounding periods per year. | CpY or C/Y | 12 |
| 3 Round your answer to the nearest cent. | The amount of each monthly payment is $3244.54. | |

## EXERCISE 6.5 Annuities

ANSWERS p. 713

### Recap

1 ©VCAA | 2018NICQ24 J Indira borrowed $29 000 to buy a car and was charged interest at the rate of 12.5% per annum, compounding monthly. For the first year of the loan, Indira made monthly repayments of $425. For the second year of the loan, Indira made monthly repayments of $500. The total amount of interest that Indira paid over this two-year period is closest to

A $2500 B $4300 C $5900 **D** $6800 **E** $7700

2 ©VCAA | 2011 IBRMQ9~ 44% Xavier borrows $45 000 from the bank to buy a car. He is offered a reducing balance loan for three years with an interest rate of 9.75% per annum, compounding monthly. He can repay this loan by making 36 equal monthly payments. Instead, Xavier decides to repay the loan in 18 equal monthly payments. If there are no penalties for repaying the loan early, the amount he will save is closest to

A $2697 B $3530 C $3553 **D** $6581 **E** $7083

### Mastery

3 a WORKED EXAMPLE 16 Ashleigh plans to invest $45000 in an annuity at a rate of 13% per annum compounding fortnightly where she makes regular fortnightly withdrawals of $250.

a Write a recurrence relation for the balance.

Ashleigh changes her mind and decides to invest the $45 000 in an annuity compounding weekly with weekly withdrawals that has the recurrence relation

$V_0 = 45\,000, \quad V_{n+1} = 1.002\, V_n - 25$

b What is the annual interest rate of this investment?

c How much has the investment increased between the third and fourth week, correct to the nearest cent?



9780170448406

▶ *S WORKED EXAMPLE 17 J The following amortisation table shows the first three payments of an annuity with monthly compounding interest and monthly withdrawals. Some of the entries are missing.

| Payment number | Payment | Interest | Principal reduction | Balance |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 380000.00 |
| 1 | 3854.00 | 2850.00 | | 378996.00 |
| 2 | 3854.00 | 2842.47 | 1011.53 | 377984.47 |
| 3 | 3854.00 | | 1019.11 | |

Find the following from the table:

a the amount invested

b the regular monthly withdrawal

c the interest rate per compounding period

d the interest rate per annum

e the interest paid at payment 3

f the amount by which the principal has been reduced at payment 1

g the balance after three withdrawals.

5H WORKED EXAMPLE 18 , Ashton purchases a $500000 annuity. Interest is paid at 7% per annum compounding monthly. If he receives monthly payments of $4200, for how many years will the annuity last?

60 WORKED EXAMPLE 19 , Anton invests $450000 in an annuity with interest paid at 6.5% per annum compounding monthly. He receives monthly payments from this investment. What will be the monthly payment, to the nearest cent, if he wishes to receive payments for 20 years?

## Exam practice

80-100% 60-79% 0-59%

7 ©VCAA 2018N 1CQ19 Cheryl invested $175000 in an annuity. This investment earns interest at the rate of 4.8% per annum, compounding quarterly. Immediately after the interest has been added to the account each quarter, Cheryl withdraws a payment of $3500. A recurrence relation that can be used to determine the value of Cheryls investment after $n$ quarters, $V_n$, is

A $V_o = 175000$, $V_{n+1} = 0.952\, V_n - 3500$

B $V_o = 175000$, $V_{n+1} = 0.988\, V_n - 3500$

C $V_o = 175000$, $V_{n+1} = 1.004\, V_n - 3500$

D $V_o - 175000$, $V_{n+1} = 1.012 V_n - 3500$

E $V_o - 175000$, $V_{n+1} - 1.048 V_n - 3500$

8 ©VCAA 2016 1CQ18 77% The value of an annuity, $V_n$, after $n$ monthly payments of $555 have been made, can be determined using the recurrence relation

$V_o = 100000$, $V_{M+1} = 1.0025\, V_n - 555$.

The value of the annuity after five payments have been made is closest to

A $97225 B $98158 C $98467 D $98775 E $110224 ▶

► *Use the following information to answer the next three questions.*

Kim invests $400 000 in an annuity paying 3.2% interest per annum. The annuity is designed to give her an annual payment of $47 372 for 10 years. The amortisation table for this annuity is shown below. Some of the information is missing.

| Payment number | Payment made | Interest earned | Reduction in principal | Balance of annuity |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 400000.00 |
| 1 | 47372.00 | 12800.00 | 34572.00 | |
| 2 | 47372.00 | 11693.70 | 35678.30 | 329749.70 |
| 3 | 47372.00 | 10551.99 | 36820.01 | 292929.69 |
| 4 | 47372.00 | 9373.75 | 37998.25 | 254931.44 |
| 5 | 47372.00 | 8157.81 | | 215717.24 |
| 6 | 47372.00 | 6902.95 | 40469.05 | 175248.19 |
| 7 | 47372.00 | 5607.94 | 41 764.06 | 133484.14 |
| 8 | 47372.00 | | | 90383.63 |
| 9 | 47372.00 | 2892.28 | 479.72 | 45903.90 |
| 10 | 47372.00 | 1468.92 | 45903.08 | 0.83 |

9 ©VCAA 2016S1CQ21 The balance of the annuity after one payment has been made is

A $339828.00 B $352628.00 C $365428.00

D $387200.00 E $400000.00

10 ©VCAA 2016S1CQ22 The reduction in the principal of the annuity after payment number 5 is

A $36820.01 B $37998.25 C $39214.19

D $40469.05 E $41764.06

11 © VCAA 2016S 1CQ23 The amount of payment number 8 that is the interest earned is closest to

A $3799.82 **B** $4074.67 **C** $4271.49 **D** $4836.57 **E** $5607.94

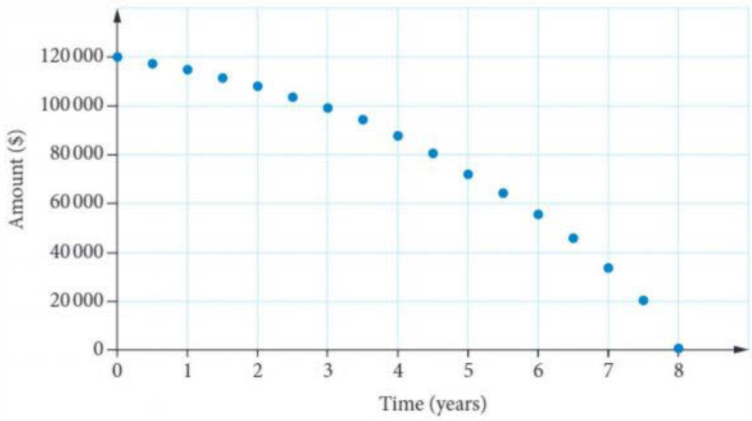
12 ©VCAA | 20161CQ24 I 30% Mai invests in an annuity that earns interest at the rate of 5.2% per annum compounding monthly. Monthly payments are received from the annuity. The balance of the annuity will be $130784.93 after five years. The balance of the annuity will be $66992.27 after 10 years. The monthly payment that Mai receives from the annuity is closest to

A $1270 B $1400 C $1500 D $2480 E $3460



9780170448406

▶ 13 Which of the following annuities could this graph model?




A An annuity of \$120 000 at 6% interest compounding annually with regular annual payments over eight years.

B An annuity of \$120 000 compounding annually over eight years with annual payments of \$23 000.

C An annuity of \$120000 at 5.5% interest compounding six-monthly with regular six-monthly payments over eight years.

D An annuity of \$ 120 000 compounding monthly over six years with monthly payments of \$2451.

E An annuity of \$120000 at 8% interest compounding six-monthly with regular six-monthly payments over seven years.

14 (6 marks) Ekaterina invests \$200000 at a rate of 9% p.a. compounding monthly. Each month, after interest is paid, she withdraws an income of \$2000. The value of her investment, $V_n$, at the end of month $n$ is given by a recurrence relation of the form

$V_0$ = a, $V_{,,+i}$-$RV_{n+}$ $d$

a Find the values of $a$, $R$ and $d$. 1 mark

b For how many months will Ekaterina receive an income from this investment? 1 mark

c If she withdrew \$3000 a month:

i how long would it last? 1 mark

ii what is the value of the final payment? 1 mark

d If the interest rate was 8% and her monthly withdrawals were \$2000:

i how long would it last? 1 mark

ii what is the value of the final payment? 1 mark

15 ©VCAA | 20i9N2CQ8n (3 marks) A record producer gave the band \$50 000 to write and record an album of songs. This \$50000 was invested in an annuity that provides a monthly payment to the band. The annuity pays interest at the rate of 3.12% per annum, compounding monthly. After six months of writing and recording, the band has \$32 667.68 remaining in the annuity.

a What is the value, in dollars, of the monthly payment to the band? 1 mark

b After six months of writing and recording, the band decided that it needs more time to finish the album. To extend the time that the annuity will last, the band will work for three more months without withdrawing a payment. After this, the band will receive monthly payments of \$3800 for as long as possible. The annuity will end with one final monthly payment that will be smaller than all of the others. Calculate the total number of months that this annuity will last. 2 marks

# @ Perpetuities

Video playlist Perpetuities

Puzzle Perpetuities

A perpetuity (or perpetuity investment) is a type of annuity where the payments are *exactly* equal to the interest for each compounding period. This means the balance doesn't change and investment provides regular payments that continue forever. Often scholarship funds are set up as perpetuities. Perpetuities are investment versions of interest-only loans.

## Perpetuity recurrence relations and formula

The perpetuity recurrence relations and formula work the same way as the ones for interest-only loans.

**Perpetuity**

For a perpetuity with value $V_n = V_0$ after $n$ compounding periods:

$$d = \frac{r}{100} \times V_0$$

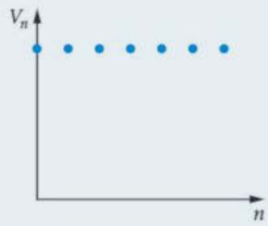
where

$d$ = payment made per compounding period

$r$ = percentage interest rate per compounding period

$V_0$ = principal.

The graph of perpetuity will look like this.




WB p. 113

**WORKED EXAMPLE 20** Working with perpetuities

a Marina has an investment whose recurrence relation for $V_n$, the value of the investment after $n$ compounding periods, is

$V_o = 24000$, $V_{,,+1} = 1.007V_{,,} - 168$

Find Vj and $V_2$ and explain why this shows that this is a perpetuity.

b Lucy wishes to set up a scholarship fund in her name so that each year an amount of $5000 is awarded to a promising cross-country runner at the school she attended. If interest on the investment is 4% per year, compounded annually, how much should she invest in this perpetuity?

c Jeremiah has $500 000 to set up a perpetuity for his granddaughter Sophie. He invests the money in bonds that return 3.5% per annum compounding quarterly. Use the perpetuity formula to find how much Sophie will receive each quarter from this investment, rounding to the nearest cent.

| Steps | Working |
|---|---|
| a 1 Find $V_r$ | $V_, = 1.OO7V_o - 168$<br>$= 1.007 \times 24000 - 168$<br>$= 24000$ |
| 2 Find $V_2$. | $V_2 = 1.007Vj - 168$<br>$= 1.007 \times 24000 - 168$<br>$= 24000$ |
| 3 For perpetuities $V_n = V_o$. | The value of the investment stays at the principal value $24000 for all compounding periods. |

**Exam hack**

Sometimes using the formula is quicker than using a finance solver. Compare Worked examples 20c and 21b.



9780170448406

6.6

b 1 Identify what we know and what we need to find from the perpetuity formula.

$r = \frac{4}{1} = 4$, $d = 5000$, $V_o = ?$

$d = \frac{r}{100} \text{x} V_0$

2 Substitute into the formula and solve, using CAS if necessary.

$5000 = \frac{4}{100} \text{x} V_0$

w 5000 x 100

yo= T~

=125000

3 Write the answer.

Lucy should invest $125000.

c 1 Identify what we know and what we need to find from the perpetuity formula.

$r = \frac{3.5}{4} = 0.875$, $V_o = 500000$, d=?

$d = \frac{r}{100} \times V_0$

2 Substitute into the formula and solve.

$d = \frac{0.875}{100} \text{x } 500000$

= 4375

3 Round your answer to the nearest cent.

Sophie will receive $4375.00 quarterly from this investment.

## Perpetuities and finance solvers

We can use finance solvers to answer questions about perpetuities in a similar way to how we used them for interest-only loans.

**Using finance solvers for perpetuities**

For a perpetuity

- all compounding periods are the same so N = 1
- FV and PV have the same value but opposite signs so FV = -PV.

**WORKED EXAMPLE 21** Using finance solvers for perpetuities

WB p. 114

Use finance solvers to find each of the following.

**Steps** | **Working**

a A magazine owner has $350 000 to invest in a perpetuity. The interest earned from this perpetuity will provide an annual literary prize of $10 000. What annual interest rate, correct to one decimal place, would be required for this investment?

TI-Nspire 1 ClassPad ligl

1 Find I when the value of the investment stays the same.

| Steps | | |
|---|---|---|
| 2 All compounding periods are the same for perpetuities. | N | 1 |
| Annual interest rate | I% | 2.8571428571429 |
| Present value for an investment is negative. | PV | -350000 |
| Money moving to a person is positive. | Pmt or PMT | 10000 |
| FV = -PV for perpetuities | FV | 350000 |
| Same as CpY or C/Y | PpY or P/Y | 1 |
| Number of compounding periods per year | CpY or C/Y | 1 |

3 Round your answer to one decimal place.

The annual interest rate is 2.9%.

b Jeremiah has $500 000 to set up a perpetuity for his granddaughter Sophie. He invests the money in bonds that return 3.5% per annum compounding quarterly. Use a finance solver to find how much Sophie will receive each quarter from this investment, rounding to the nearest cent.

TI-Nspire + ClassPad

| | | |
|---|---|---|
| 1 Find Pmt when the value of the investment stays the same. | | |
| 2 All compounding periods are the same for perpetuities. | N | 1 |
| Annual interest rate | I% | 3.5 |
| Present value for an investment is negative. | PV | -500000 |
| Money moving to a person is positive. | Pmt or PMT | 4375 |
| FV = -PV for perpetuities | FV | 500000 |
| Same as CpY or C/Y | PpY or P/Y | 4 |
| Number of compounding periods per year | CpY or C/Y | 4 |
| 3 Round your answer to the nearest cent. | Sophie will receive $4375.00 quarterly from this investment. | |

## EXERCISE 6.6 Perpetuities

ANSWERS p. 714

### Recap

1 The following amortisation table shows the first payment of an annuity with quarterly compounding interest and quarterly withdrawals.

| Payment number | Payment | Interest | Principal reduction | Balance |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 60000.00 |
| 1 | 1630.00 | 720.00 | 910.00 | 59090.00 |

Which of the following is closest to the annual interest rate of the annuity?

A 1.2% B 4.8% C 0.012% D 0.048% E 0.44%

2 ©VCAA 2017N 1CQ22 Vusa has invested $420 000 in an annuity that pays interest at the rate of 3.6% per annum, compounding monthly. After the interest has been added each month, Vusa immediately receives a payment from the annuity. The value of Vusas investment is $372934.71 after three years. The monthly payment that Vusa receives from the annuity is closest to

A $1260 B $1310 C $2500 D $15120 E $16900

### Mastery

3 WORKED EXAMPLE 20

a The recurrence relation for an investment, where $V_n$, is the value of the investment after $n$ compounding periods, is

$$V_0 = 31000,\ V_{n+1} = 1.006\,V_n - 186$$

Find $V_1$ and $V_2$ and explain why this shows that the investment is a perpetuity.

b Emil sets up a scholarship fund in his name so that each year an amount of $3000 is awarded to the top mathematics student in their final year at the school he attended. If interest on the investment is 5.6% per annum, compounded yearly, how much should he invest in this perpetuity, rounded to the nearest cent?

c Fleur invests $650000 in a perpetuity that returns 4.2% per annum compounding six-monthly. How much will Fleur receive each six months from this investment, rounded to the nearest dollar?

▶ 4H WORKED EXAMPLE 21J Use finance solvers to find each of the following.

a A performing arts school has $400 000 to invest in a perpetuity. The interest earned from this perpetuity will provide an annual scholarship of $25 000. What annual interest rate, correct to one decimal place, would be required for this investment?

b Felicity has $330000 in a perpetuity for her grandchildren, with an interest rate of 4.91% per annum compounding half yearly. Using a finance solver, find how much the grandchildren receive every six months from this investment, rounded to the nearest dollar.

c A billionaire has decided to use $2 000 000 to set up a perpetuity so that her only child will receive $14500 a month for life. What annual interest rate, correct to one decimal place, would be required for this investment?

d A college awards a prize each year to the top student. A donation from a former student of $866 000 is used as a perpetuity to fund the prize. The investment has an interest rate of 5.23% per annum, compounding yearly. How much is the prize, rounded to the nearest dollar?

6.6

## Exam practice

80-100% 60-79% 0-59%

5 ©VCAA EZ2ZEZ3U 51% Which one of the following recurrence relations could be used to model the value of a perpetuity investment, P,,, after *n* months?

A $P_o$- 120000, $P_{,,+1}$- 1.0029 x$P_{,,}$-356

B $P_o$- 180000, $P_{,,+1}$ = 1.0047 x $P_{,,}$ - 846

C $P_o$ = 210000, $P_{M+1}$ = 1.0071 x$P_{,,}$-1534

D $P_o$ = 240000, $P_{w+1}$ = 0.0047 x $P_n$ - 2232

E $P_o$ = 250000, $P_{n+(}$ = 0.0085 $xP_n$ - 2125

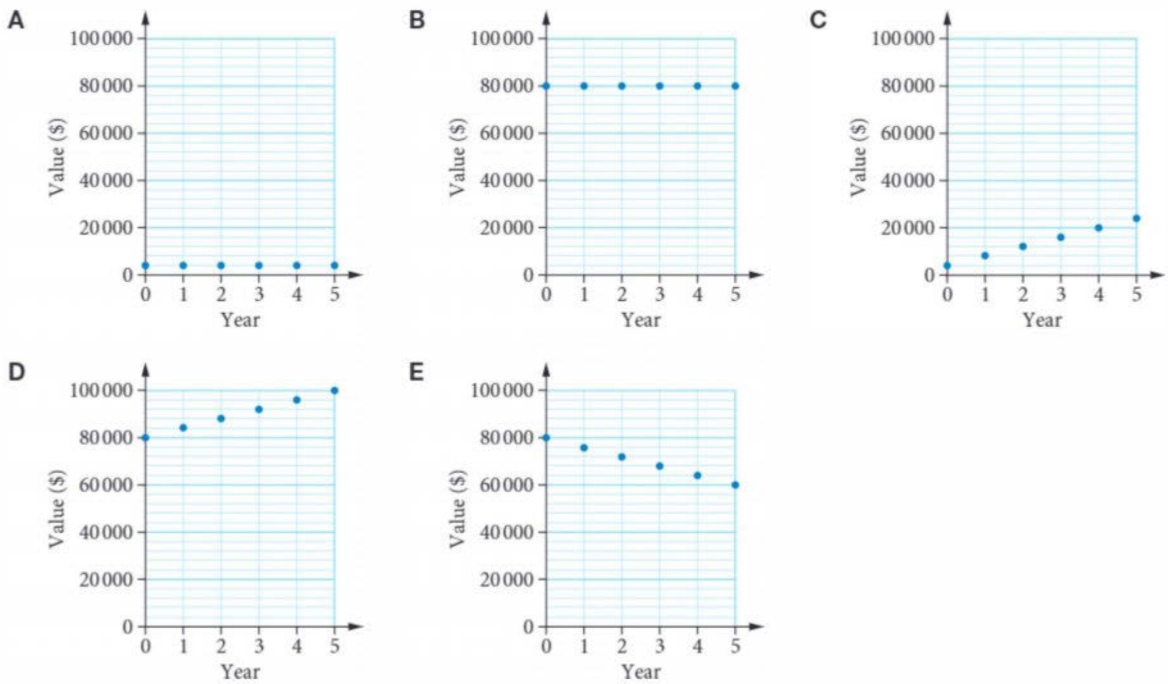
6 ©VCAA 2018N1CQ22 I Amir invested some money in a perpetuity from which he receives a monthly payment of $525. The perpetuity pays interest at an annual rate of 4.2%, paid monthly. How much money did Amir invest in the perpetuity?

A $12500 B $22 500 C $52500 D $120500 E $150000

7 ©VCAA 201 ZN 1CQ20 A music school has $80000 to invest in a perpetuity. The interest earned from this perpetuity will provide an annual prize of $3000 to a talented musician from the school. What annual interest rate would be required for this investment?

A 0.3125% B 3.75% C 3.90% D 41.92% E 45.00%

8 ©VCAA 20161CQ21 J 50% Juanita invests $80000 in a perpetuity that will provide $4000 per year to fund a scholarship at a university. The graph that shows the value of this perpetuity over a period of five years is



9 ©VCAA 2011 1BRMQ5 , 44% Jane invests in an ordinary perpetuity to provide her with a weekly payment of $500. The interest rate for the investment is 5.9% per annum. Assuming there are 52 weeks per year, the amount that Jane needs to invest in the perpetuity is closest to

A $26000 B $102000 C $154000 D $221000 E $441000

10 ©VCAA | 2014 2BRMQ2ab , (2 marks) A sponsor of a cricket club has invested $20 000 in a perpetuity. The annual interest from this perpetuity is $750. The interest from the perpetuity is given to the best player in the club every year, for a period of 10 years.

a What is the annual rate of interest for this perpetuity investment? 1 mark

b After 10 years, how much money is still invested in the perpetuity? 1 mark

11 ©VCAA 2020 2CQ10 | (3 marks) Samuel invests $500000 in an annuity from which he receives a regular monthly payment. The balance of the annuity, in dollars, after $n$ months, $A_n$, can be modelled by a recurrence relation of the form

$$A_0 = 500\,000,\ A_{n+1} = kA_n - 2000$$

a 60% Calculate the balance of this annuity after two months if $k = 1.0024$. 1 mark

b 48% Calculate the annual compound interest rate percentage for this annuity if $k = 1.0024$. 1 mark

c 36% For what value of $k$ would this investment act as a simple perpetuity? 1 mark

12 ©VCAA 2012 2BRMQ4 , 27% (3 marks) Arthur invested $80 000 in a perpetuity that returns $1260 per quarter. Interest is calculated quarterly.

a Calculate the annual interest rate of Arthurs investment. 1 mark

b After Arthur has received 20 quarterly payments, how much money remains invested in the perpetuity? 1 mark

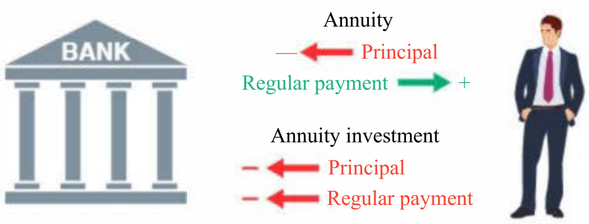
c Arthurs wife, Martha, invested a sum of money at an interest rate of 9.4% per annum, compounding quarterly. She will be paid $1260 per quarter from her investment. After 10 years, the balance of Marthas investment will have reduced to $7000. Determine the initial sum of money Martha invested. Write your answer, correct to the nearest dollar. 1 mark

# @ Annuity investments

Video playlist Annuity investments

## Annuity investment recurrence relations

An annuity investment is a compound interest investment where regular payments are made into the account. This is different to an annuity where the regular payments are taken *out of the* account. People often use annuity investments when saving for retirement or for the deposit of a house.




### @ Exam hack

Don't confuse annuities and annuity investments. Both are investments, but with annuities the regular payments are withdrawals and with annuity investments the regular payments are added.

The recurrence relation for an annuity investment is the same as the recurrence relation for an annuity except the payment is added rather than subtracted.

 9780170448406

## Annuity investment recurrence relation

The recurrence relation for the value $V_n$ of an annuity investment

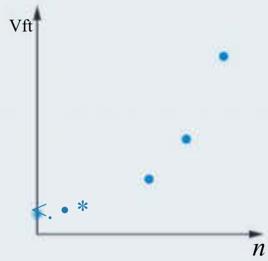
$$V_o \text{ - principal, } V_{n+1} = \left(1 + \frac{r}{100}\right)V_n + d$$

where

$r$ = percentage interest rate per compounding period

$n$ = number of compounding periods

$d$ = payment made per compounding period.




The graph of an annuity recurrence relation will look the same as a compound interest graph, except it will increase faster.



WB p. 115

**WORKED EXAMPLE 22** Working with annuity investment recurrence relations

Teofilo plans to invest $55000 in an annuity investment a rate of 3.6% per annum compounding monthly where he makes regular monthly additions of $260.

a Write a recurrence relation for the balance.

Teofilo then changes his mind and decides to invest the $55 000 in a different annuity investment compounding quarterly with quarterly additions that has the recurrence relation

$V_o = 55000,\ V_{n+1} = 1.01\ V_n + 260$

b What is the annual interest rate of this investment?

c How much has the investment increased between the third and fourth quarter, to the nearest cent?

| Steps | Working |
|---|---|
| a 1 Find the number of compounding periods per year. | There are 12 compounding periods per year. |
| 2 Identify $V_n$, $V_o$, r and $d$. | Let $V_n$ = the balance after $n$ compounding periods.<br>$V_o = 55\,000$, $r = \frac{3.6}{12} = 0.3\%$, $d = 260$ |
| 3 Substitute the values into<br>$V_o$ = principal, $V_{n+1} = \left(1 + \frac{r}{100}\right)V_n + d$<br>and simplify. | $V_{n+1} = \left(1 + \frac{0.3}{100}\right)V_n + 260$<br>$= (1 + 0.003)\ V_n + 260$<br>$= 1.003V_n + 260$<br>$V_o = 55000,\quad V_{n+1} = 1.003V_n + 260$ |
| b 1 Use the recurrence relation to find an equation for r, the percentage interest rate per compounding period. | Comparing:<br>$V_{n+1} = \left(1 + \frac{r}{100}\right)V_n + d$<br>$V_{n+1} = 1.01V_n + 260$<br>we can see that<br>$1 + \frac{r}{100} = 1.01$ |
| 2 Solve for r, using CAS if necessary. | $\frac{r}{100} = 1.01 - 1$<br>$\frac{r}{100} = 0.01$<br>r = 1% per quarter |
| 3 Multiply $r$ by the compounding period to find the annual percentage compound interest rate. | The interest compounds quarterly, so the annual percentage compound interest rate is<br>$r \times 4 = 1 \times 4 = 4\%$ |

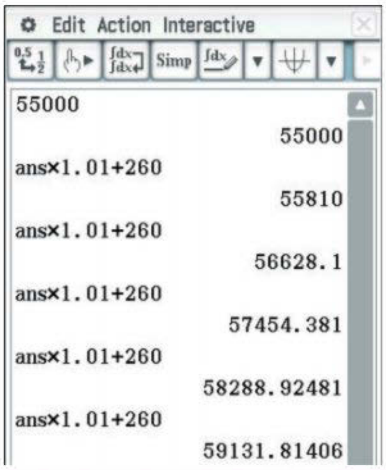
c Use CAS recursive computation to find the balances, and subtract to find the difference, rounding to the nearest cent.

$V_4 - V_3 = 58288.92481 - 57454.381$
$= 834.54381$

The investment increased by $834.54.

TI-Nspire


ClassPad




## Annuity investment amortisation tables

Amortisation tables work in a similar way for annuity investments as for annuities. The difference is annuity investments involve a principal *addition* rather than a principal reduction.

**Annuity investment amortisation tables**

$$r = \frac{\text{interest}}{\text{previous balance}} \times 100$$

| Payment number | Payment | Interest | Principal addition | Balance |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | Vo |
| $n$ | $d$ | $\frac{r}{100} \times$ previous balance | payment + interest | previous balance + principal addition |

where

$n$ = payment number

$d$ = payment made per compounding period

$r$ = percentage interest rate per compounding period

$V_o$ = principal.


9780170448406

6.7

WB

p. 117

**WORKED EXAMPLE 23** Using annuity investment amortisation tables

Vinh invests \$4000 in an annuity investment earning interest compounding annually. He deposits an extra \$1200 into the account each year after the initial deposit. The amortisation table below shows the first few calculations for his investment.

| Payment number | Payment | Interest | Principal addition | Balance |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 4000.00 |
| 1 | 1200.00 | $\frac{r}{100}$ x 4000 = 320.00 | 1200.00 + 320.00= 1520.00 | 4000.00 + 1520.00 = 5520.00 |
| 2 | 1200.00 | $\frac{r}{100}$ x 5520 = 441.60 | | |
| 3 | 1200.00 | | | |
| 4 | | | | |

**Steps** | **Working**

a Use the table to show two calculations that will give r, the percentage interest rate per compounding period.

Use $r = \frac{\text{interest}}{\text{previous balance}} \times 100$

Using interest for payment number 1

$r = \frac{320.00}{4000.00} \times 100$

$= 0.08 \times 100$

$= 8\%$

Using interest for payment number 2

$r = \frac{441.60}{5520.00} \times 100$

$= 0.08 \times 100$

$= 8\%$

b What is the nominal interest rate for the loan?

Use the compounding period to find the annual interest rate.

The compounding period is 1.

The nominal interest rate

= 1x8

= 8% per annum compounding annually

c Complete the amortisation table, showing the interest paid, principal addition and balance for the first four years of the loan, giving all values to the nearest cent.

Complete the table using:

interest $= \frac{r}{100} \times$ previous balance

principal addition = payment + interest

balance = previous balance + principal addition

Give all values to the nearest cent.

| Payment number | Payment | Interest | Principal addition | Balance |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 4000.00 |
| 1 | 1200.00 | 320.00 | 1520.00 | 5520.00 |
| 2 | 1200.00 | 441.60 | 1641.60 | 7161.60 |
| 3 | 1200.00 | 572.93 | 1772.93 | 8934.53 |
| 4 | 1200.00 | 714.76 | 1914.76 | 10849.29 |

**Exam hack**

As with reducing balance loans, choose payment 1 when calculating r from an amortisation table.

WB

p. 118

**WORKED EXAMPLE 24** Analysing annuity investment amortisation tables

Tobias has invested $2500.00 in an annuity investment with an interest rate of 10% p.a. compounding six-monthly and he is making additional six-monthly payments of $1000. Answer these questions using the amortisation table of the investment given below, which has some missing entries. Round your answers to the nearest cent where relevant.

| Payment number | Payment | Interest | Principal addition | Balance |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 2500.00 |
| 1 | 1000.00 | | 1125.00 | 3625.00 |
| 2 | 1000.00 | 181.25 | | 4806.25 |
| 3 | 1000.00 | 240.31 | 1240.31 | 6046.56 |
| 4 | | 302.33 | 1532.33 | 7578.89 |

| Steps | Working |
|---|---|
| a Find r, the interest rate per compounding period. | |
| Divide the percentage interest rate per year by the number of compounding periods per year. | $r = \frac{10}{2}$<br>$= 5\%$ |
| b What is the missing interest earned in the first six-month period? | |
| $\text{interest} = \frac{r}{100} \times \text{previous balance}$ | $\text{interest} = \frac{5}{100} \times 2500.00$<br>$= \$125.00$ |
| c What is the missing principal addition in the second six-month period? | |
| principal addition = payment + interest | principal addition = 1000 + 181.25<br>= $1181.25 |
| d Tobias paid a higher amount in the fourth six-month period. How much is the payment? | |
| principal addition = payment + interest | 1532.33 = payment + 302.33<br>payment = 1532.33 - 302.33<br>= $1230.00 |



9780170448406

## Annuity investments and finance solvers

Annuity investments can be analysed using finance solvers in a similar way to annuities. The difference is

- Pint is always negative because the payments are going away from the person
- FV will continue to increase and will never reach zero.

6.7

### Using finance solvers for annuity investments

| | |
|---|---|
| N | Total number of compounding periods |
| 1% | Interest rate per year |
| PV | Present value for an annuity investment is negative because the money is going away from the person. |
| Pmt or PMT | Regular payments for an annuity investment are negative because the money is going away from the person. |
| FV | Future value has the opposite sign of the present value so it will be positive. |
| PpY or P/Y | Number of payments per year. This will always take the same value as CpY or C/Y. |
| CpY or C/Y | Number of compounding periods per year |

When solving for N, always round *up,* never down, to the nearest whole number.

Ignore the negative sign in values when using the following formula:

total interest paid = N x Pint - (PV - FV)

**WORKED EXAMPLE 251** Using finance solvers for annuity investments

WB p. 119

Ruben's investment earns interest at the rate of 4.7% per annum, compounding monthly. Ruben initially invested $110 000 and adds monthly payments of $2000.

| Steps | Working | |
|---|---|---|
| a After how many months will the value of this investment first exceed $130 000? | | |
| | STr^spir^J + ClassPad | |
| 1 Enter the amount as FV and find N. | | |
| 2 Total number of compounding periods | N | 8.1136927001109 |
| Annual interest rate | I% | 4.7 |
| Present value for an investment is negative. | PV | -110000 |
| Money moving away from a person is negative. | Pmt or PMT | -2000 |
| Future value has the opposite sign to present value. | FV | 130000 |
| Same as CpY or C/Y | PpY or P/Y | 12 |
| Number of compounding periods per year | CpY or C/Y | 12 |
| 3 Round N *up* to the nearest whole number. | The value of the investment will first exceed $130 000 after nine months. | |

b Ruben wants to reach a target of $250 000 in four years. After one year, he increased his payments so that he would reach his target. What did Ruben increase his payments to? Give your answer to the nearest cent.

| | TI-Nspire | ClassPad |
|---|---|---|
| 1 Find FV after 12 months. | | |
| 2 Total number of compounding periods | N | 12 |
| Annual interest rate | I% | 4.7 |
| Present value for an investment is negative. | PV | -110000 |
| Money moving away from a person is negative. | Pmt or PMT | -2000 |
| Future value has the opposite sign to present value. | FV | 139806.646892 |
| Same as CpY or C/Y | PpY or P/Y | 12 |
| Number of compounding periods per year | CpY or C/Y | 12 |
| 3 Use FV after one year as PV for the next three years. | | |
| 4 Total number of compounding periods | N | 36 |
| Annual interest rate | I% | 4.7 |
| Present value for an investment is negative. | PV | -139806.646892 |
| Money moving away from a person is negative. | Pmt or PMT | -2308.605149083 |
| Future value has the opposite sign to present value. | FV | 250000 |
| Same as CpY or C/Y | PpY or P/Y | 12 |
| Number of compounding periods per year | CpY or C/Y | 12 |
| 5 Write your answer to the nearest cent. | Ruben increased his payments to $2308.61. | |

## Exam hack

Questions don't always state what type of compound interest investment you're dealing with. Look at the type of payment involved to decide.

- No payment means it's an ordinary compound interest investment.
- Withdrawals from the account mean it's an annuity.
- Additions to the account mean it's an annuity investment.

6.7

WB

p. 120

## WORKED EXAMPLE 26 Using finance solvers for different types of compound interest investments

Antionette started her savings by investing $20 000 for 10 years in an account earning 4.8% p.a. interest compounding monthly. After the 10 years she put the balance into a superannuation account for her retirement earning 5.6% p.a. interest compounding monthly and made regular monthly payments of $600 for 30 years until she retired. Since Antionette retired, she has been taking out regular monthly payments of $4000 from this account. How much money, to the nearest cent, is in her account after she has been retired for five years?

| Steps | Working | |
|---|---|---|
| | TI-Nspire 1 | ClassPad |
| 1 The first account is an ordinary compound interest investment. Find FV after 10 x 12 = 120 months. | | |
| 2 Total number of compounding periods | N | 120 |
| Annual interest rate | I% | 4.8 |
| Present value for an investment is negative. | PV | -20000 |
| Payment amount | Pmt or PMT | 0 |
| Future value has the opposite sign to present value. | FV | 32290.556720832 |
| Same as CpY or C/Y | PpY or P/Y | 12 |
| Number of compounding periods per year | CpY or C/Y | 12 |
| 3 The superannuation account is an annuity investment. Enter the unrounded FV from the first account as the new PV. Find FV after 30 x 12 = 360 months. | | |
| 4 Total number of compounding periods | N | 360 |
| Annual interest rate | I% | 5.6 |
| Present value for an investment is negative. | PV | -32290.556720832 |
| Money moving away from a person is negative. | Pmt or PMT | -600 |
| Future value has the opposite sign to present value. | FV | 731176.221 6121 |
| Same as CpY or C/Y | PpY or P/Y | 12 |
| Number of compounding periods per year | CpY or C/Y | 12 |
| 5 The superannuation account is now an annuity. Enter the unrounded FV from the annuity account as the new PV. Find FV after 5 x 12 = 60 months. | | |
| 6 Total number of compounding periods | N | 60 |
| Annual interest rate | I% | 5.6 |
| Present value for an investment is negative. | PV | -731176.2216121 |
| Money moving to a person is positive. | Pmt or PMT | 4000 |
| Future value has the opposite sign to present value. | FV | 690581.1645821 |
| Same as CpY or C/Y | PpY or P/Y | 12 |
| Number of compounding periods per year | CpY or C/Y | 12 |
| 7 Write your answer to the nearest cent. | Antionette has $690 581.16 in her account after she has been retired for five years. | |

Video
VCE question analysis: Loans, investments and finance solvers

## VCE QUESTION ANALYSIS

©VCAA | 2018 2CQ6 | 2018 Examination 2 Core Question 6 (4 marks)

Julie has retired from work and has received a superannuation payment of \$492 800.
She has two options for investing her money.

Option 1: Julie could invest the \$492 800 in a perpetuity. She would then receive \$887.04 each fortnight for the rest of her life.

a At what annual percentage rate is interest earned by this perpetuity? — 1 mark

Option 2: Julie could invest the \$492 800 in an annuity, instead of a perpetuity.
The annuity earns interest at the rate of 4.32% per annum, compounding monthly.
The balance of Julie's annuity at the end of the first year of investment would be \$480242.25.

b i What monthly payment, in dollars, would Julie receive? — 1 mark

ii How much interest would Julie's annuity earn in the second year of investment? Round your answer to the nearest cent. — 2 marks

### Reading the question

- Note part a is about a perpetuity and part b is about an annuity and you are comparing two options.
- Highlight whether regular payments and interest rate compounding are made daily, weekly, fortnightly, monthly or yearly.
- Highlight what sort of rounding is required.

### Thinking about the question

- This is the last question in the Recursion and financial modelling section, so expect it to be challenging.
- What is different about how perpetuities work compared to other annuities?
- Be aware you need to use a formula (which isn't on the formula sheet) to solve perpetuity problems.
- Remember that the *r* in the formula is the interest rate per compounding period not per year.

### Worked solution (/ = 1 mark)

Use the perpetuity formula,

$d = \frac{r}{100} \times V_0$, where

$d = 887.04$, $r = ?$, $V_0 = 492800$

$887.04 = \frac{r}{100} \times 492800$

Solve for *r* algebraically or use the CAS solve function:

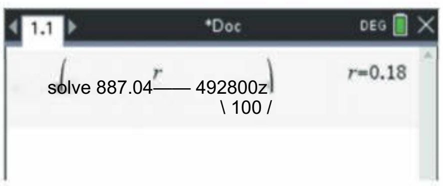
**Algebraic**

$\frac{r}{100} \times 492\,800 = 887.04$

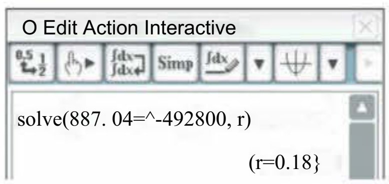
$r = \frac{100 \times 887.04}{492\,800}$

$= 0.18$

**TI-Nspire**




**ClassPad**




Interest rate per fortnight = 0.18%.

Annual percentage interest rate is 0.18% x 26 = 4.68% /


9780170448406

6.7

b i Use a finance solver to find Pint for the annuity, after 12 months.

| | | |
|---|---|---|
| Total number of compounding periods | N | 12 |
| Annual interest rate | I% | 4.32 |
| Present value for an investment is negative. | PV | -492800 |
| Money moving to a person is positive. | Pmt or PMT | 2800.000202 |
| Future value is zero when an investment is fully paid out. | FV | 480242.25 |
| Same as CpY or C/Y | PpY or P/Y | 12 |
| Number of compounding periods per year | CpY or C/Y | 12 |

The monthly payment Julie would receive (rounded to whole dollars) = S2800 /

ii Total interest paid = N x Pmt - (PV - FV) (ignoring the negative sign in values)

N = 12, Pmt = 2800

PV = value after first year = 480 242.25

FV = value after second year = 467131.14 / (from finance solver)

**Exam hack**

When answering extended answer questions with finance solvers, show your working by listing the values you entered for N, 1%, PV etc.

| | | |
|---|---|---|
| Total number of compounding periods | N | 24 |
| Annual interest rate | 1% | 4.32 |
| Present value for an investment is negative. | PV | -492800 |
| Money moving to a person is positive. | Pmt or PMT | 2800 |
| Future value has the opposite sign to present value. | FV | 467131.1389 |
| Same as CpY or C/Y | PpY or P/Y | 12 |
| Number of compounding periods per year | CpY or C/Y | 12 |

Total interest paid in second year = 12 x 2800 - (480242.25 - 467131.14)
= 12 x \$2800-\$13111.11
= \$33600-\$13111.11
= S20 488.89 /

**Student performance** 80-100% 60-79% 0-59%

a 41% A common error was forgetting about the 26 fortnightly payments, giving an incorrect answer of 0.18% p.a. Another common incorrect answer was 18%.

b i 48%

ii 16% This was too difficult for most students and there were many non-attempts. This was a 2-mark question, with 80% of students receiving 0 marks, 9% receiving 1 mark and 11% receiving 2 marks. This meant an average mark of only 0.31 out of 2 (i.e. 16%) for the question.

- Many found the future value after two years (\$467131) but couldn't go further.
- A few students tried unsuccessfully to use the annuities formula.
- An answer of \$20488.88, based on an unrounded value from Question b i, was also awarded full marks.

## EXERCISE 6.7 Annuity investments

ANSWERS p. 714

### Recap

1 Which one of the following recurrence relations could not be used to model the value of a perpetuity investment, P,,, after *n* months?

A PQ-30000, *P,,* $_{+i}$ - 1.0063 x *P,,* - 189

B $P_o$-50000, $P_{,,+1}$- 1.0074 x *P,,* - 370

C $P_o$-2OOOO, $P_{,,+1}$ - 1.0017 xP,,-64

D $P_o$ = 60000, $P_{,,+1}$- 1.0049 xP,,- 294

E $P_o$ = 10000, $P_{,,+1}$ = 1.0058 x P,, - 58

2 ©VCAA 2008 IBRMQ2 , 68% Pia invests $800 000 in an ordinary perpetuity to provide an ongoing fortnightly pension for her retirement. The interest rate for this investment is 5.8% per annum. Assuming there are 26 fortnights per year, the amount she will receive at the end of each fortnight is closest to

A $464 B $892 C $1422 D $1785 E $3867

### Mastery

30 WORKED EXAMPLE 22 Shane plans to invest $27 000 in an annuity investment, which has an interest rate of 7.2% per annum, compounding monthly. He makes regular monthly additional payments of $310.

a Write a recurrence relation for the balance in terms of $V_n$.

Shane then changes his mind and decides to invest the $27 000 in a different annuity investment, compounding quarterly with quarterly additions, according to the recurrence relation

$V_o$ = 27000, $V_{,,+1}$ = 1.018V,,+ 310

b What is the annual interest rate of this investment?

c How much has the investment increased between the third and fourth quarter, to the nearest cent?

4 H WORKED EXAMPLE 23 J Megan invests $500 in an annuity investment earning interest compounded annually. She deposits an extra $500 into the account each year after the initial deposit. The amortisation table below shows the first few calculations for her investment.

| Payment number | Payment | Interest | Principal addition | Balance |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 500.00 |
| 1 | 500.00 | $\frac{r}{100}$ x 500 = 40.00 | 500.00 + 40.00 = 540.00 | 500.00 + 540.00= 1040.00 |
| 2 | 500.00 | $\frac{r}{100}$ x 1040 = 83.20 | | |
| 3 | 500.00 | | | |
| 4 | 500.00 | | | |

a Use the table to show two calculations that will give r, the percentage interest rate per compounding period.

b What is the nominal interest rate for the loan?

c Complete the amortisation table, showing the interest paid, principal addition and balance for the first four years of the loan. Give all values correct to the nearest cent.



9780170448406

6.7

5Q WORKED EXAMPLE 24 J Sinjin has invested $6000.00 in an annuity investment with an interest rate of 12% p.a. compounding monthly and he is making additional monthly payments of $1200. Answer these questions using the amortisation table of the investment given below, which has some missing entries. Give your answers to the nearest cent where relevant.

| Payment number | Payment | Interest | Principal addition | Balance |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 6000.00 |
| 1 | 1200.00 | | 1260.00 | 260.00 |
| 2 | 1200.00 | 72.60 | | 8532.60 |
| 3 | 1200.00 | 85.33 | 1285.33 | 9817.93 |
| 4 | | 98.18 | 1698.18 | 11516.11 |

a Find r, the interest rate per compounding period.

b What is the missing interest earned in the first month?

c What is the missing principal addition in the second month?

d Sinjin paid a higher amount in the fourth month. How much is the payment?

60 WORKED EXAMPLE 25 J Phillips investment earns interest at the rate of 4.2% per annum, compounding monthly. Phillip initially invested $8000 and adds monthly payments of $500.

a After how many months will the value of this investment first exceed $12 000?

b Phillip wants to reach a target of $45 000 in five years. After one year, Phillip increased his payments so that he would reach his target. What did Phillip increase his payments to? Give your answer to the nearest cent.

7 H WORKED EXAMPLE 26 I Mei started her savings by investing $25000 for 12 years in an account earning 4.6% p.a. interest compounding monthly. After the 12 years she put the balance into a superannuation account for her retirement earning 5.9% p.a. interest compounding monthly and made regular monthly payments of $500 for 35 years until she retired. Since Mei retired, she has been taking out regular monthly payments of $5500 from this account. How much money, to the nearest cent, is in her account after she has been retired for four years?

**Exam practice** 80-100% 60-79% 0-59%

*Use the following information to answer the next two questions.*

The value of an annuity investment, in dollars, after *n* years, $V_n$, can be modelled by the recurrence relation

$V_0 = 46000, V_{n+1} = 1.0034V_n + 500$

8 ©VCAA 2018 1CQ17 79% What is the value of the regular payment added to the principal of this annuity investment?

A $34.00 B $156.40 C $466.00 D $500.00 E $656.40

9 ©VCAA 2018 1CQ18 61% Between the second and third years, the increase in the value of this investment is closest to

A $656 B $658 C $661 D $1315 E $1975

10 ©VCAA 2018N 1CQ23 An annuity investment earns interest at the rate of 3.8% per annum, compounding monthly. Cho initially invested $85000 and will add monthly payments of $1500. The value of this investment will first exceed $95 000 after

A five months. B six months. C seven months. D eight months. E nine months.

▶ 11 ©VCAA 2016ICQ23J 58% Sarah invests $5000 in a savings account that pays interest at the rate of 3.9% per annum compounding quarterly. At the end of each quarter, immediately after the interest has been paid, she adds $200 to her investment. After two years, the value of her investment will be closest to

A $5805 B $6600 C $7004 D $7059 E $9285

12 ©VCAA | 20181CQ24 45% Mariska plans to retire from work 10 years from now. Her retirement goal is to have a balance of $600 000 in an annuity investment at that time. The present value of this annuity investment is $265 298.48, on which she earns interest at the rate of 3.24% per annum, compounding monthly. To make this investment grow faster, Mariska will add a $1000 payment at the end of every month. Two years from now, she expects the interest rate of this investment to fall to 3.20% per annum, compounding monthly. It is expected to remain at this rate until Mariska retires. When the interest rate drops, she must increase her monthly payment if she is to reach her retirement goal. The value of this new monthly payment will be closest to

A $1234 B $1250 C $1649 D $1839 E $1854

13 ©VCAA | 2017 1CQ23 38% Four lines of an amortisation table for an annuity investment are shown. The interest rate for this investment remains constant, but the payment value may vary.

| Payment number | Payment | Interest | Principal addition | Balance of investment |
|---|---|---|---|---|
| 17 | 100.00 | 27.40 | 127.40 | 6977.50 |
| 18 | 100.00 | 27.91 | 127.91 | 7105.41 |
| 19 | 100.00 | 28.42 | 128.42 | 7233.83 |
| 20 | | | | 7500.00 |

The balance of the investment after payment number 20 is $7500. The value of payment number 20 is closest to

A $29 B $100 C $135 D $237 E $295

14 ©VCAA |_2O19N 1CQ24 Robyn has a current balance of $347 283.45 in her superannuation account. Robyn's employer deposits $350 into this account every fortnight. This account earns interest at the rate of 2.5% per annum, compounding fortnightly. Robyn will stop work after 15 years and will no longer receive deposits from her employer. The balance of her superannuation account at this time will be invested in an annuity that will pay interest at the rate of 3.6% per annum, compounding monthly. After 234 monthly payments there will be no money left in Robyns annuity. The value of Robyns monthly payment will be closest to

A $3993 B $5088 C $6664 D $8051 E $9045

15 ©VCAA 2017N 2CQ7 (3 marks) A community centre has received a donation of $5000. The donation is deposited into a savings account. This savings account pays interest compounding monthly. Immediately after the interest has been added each month, the community centre deposits a further $100 into the savings account. After five years, the community centre would like to have a total of $14000 in the savings account.

a What is the annual interest rate, compounding monthly, that is required to achieve this goal? Write your answer correct to two decimal places. 1 mark

b The interest rate for this savings account is actually 6.2% per annum, compounding monthly. After 36 deposits, the community centre stopped making the additional monthly deposits of $100. How much money will be in the savings account five years after it was opened? 2 marks ▶


9780170448406

▶ 16 ©VCAA 2017 2CQ7 J (3 marks) Alex sold his mechanics' business for $360 000 and invested this amount in a perpetuity. The perpetuity earns interest at the rate of 5.2% per annum. Interest is calculated and paid monthly.

6.7

a 50% What monthly payment will Alex receive from this investment? 1 mark

b 29% Later, Alex converts the perpetuity to an annuity investment. This annuity investment earns interest at the rate of 3.8% per annum, compounding monthly. For the first four years Alex makes a further payment each month of $500 to his investment. This monthly payment is made immediately after the interest is added. After four years of these regular monthly payments, Alex increases the monthly payment. This new monthly payment gives Alex a balance of $500000 in his annuity after a further two years. What is the value of Alex's new monthly payment? Round your answer to the nearest cent. 2 marks

17 ©VCAA 2019N2CQ7J (4 marks) Tisha plays drums in the same band as Marlon. She would like to buy a new drum kit and has saved $2500.

a Tisha could invest this money in an account that pays interest compounding monthly.

The balance of this investment after $n$ months, $T_n$, could be determined using the recurrence relation below.

$T_o = 2500, \quad T_{n+1} = 1.0036 \times T_n$

Calculate the total interest that would be earned by Tisha's investment in the first months. Round your answer to the nearest cent. 2 marks

Tisha could invest the $2500 in a different account that pays interest at the rate of 4.08% per annum, compounding monthly. She would make a payment of $150 into this account every month.

b Let $V_n$ be the value of Tisha's investment after $n$ months.

Write down a recurrence relation, in terms of $V_o$, $V_n$ and $V_{n+1}$, that would model the change in the value of this investment. 1 mark

c Tisha would like to have a balance of $4500, to the nearest dollar, after 12 months.

What annual interest rate would Tisha require? Round your answer to two decimal places. 1 mark

# Q Chapter summary

## Finance solvers

Values in a finance solver can be positive, negative or zero.

- Money coming *to the person* is positive.
- Money going away/rom *the person* is negative.
- PV and FV always have opposite signs (except when FV is zero).

| | |
|---|---|
| N | Total number of compounding periods |
| 1% | Annual interest rate |
| PV | Present value |
| Pmt or PMT | Regular payments |
| FV | Future value has the opposite sign of Present value (or can be zero). |
| PpY or P/Y | Number of payments per year. This will always take the same value as CpY or C/Y. |
| CpY or C/Y | Number of compounding periods per year |

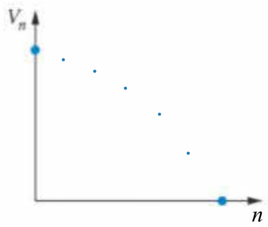
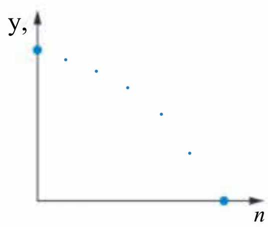
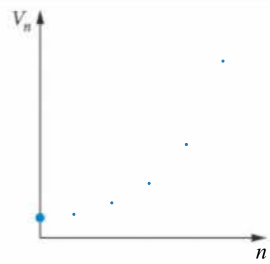
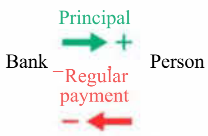
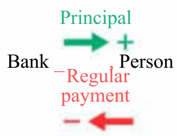
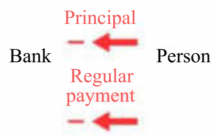
## Compound interest and finance solvers

| | Compound interest investment | Reducing balance depreciation |
|---|---|---|
| 1% | positive | negative |
| PV | negative | negative |
| Pmt or PMT | zero | zero |
| FV | positive | positive or zero |



9780170448406

### Loan and investment summary

| | Reducing balance loan | Annuity | Annuity investment |
|---|---|---|---|
| Recurrence relation | $V_0$ = principal,<br>$V_{n+1} = \left(1 + \frac{r}{100}\right)V_n - d$ | $V_0$ = principal,<br>$V_{n+1} = \left(1 + \frac{r}{100}\right)V_n - d$ | $V_0$ = principal,<br>$V_{n+1} = \left(1 + \frac{r}{100}\right)V_n + d$ |
| Type | multiply by a number greater than 1<br>subtract an amount | multiply by a number greater than 1<br>subtract an amount | multiply by a number greater than 1<br>add an amount |
| Growth/decay | geometric growth<br>linear decay | geometric growth<br>linear decay | geometric growth<br>linear growth |
| Graph |  |  |  |
| Principal / Payment |  |  |  |
| PV | positive | negative | negative |
| Pmt or PMT | negative | positive | negative |
| FV | negative or zero | positive or zero | positive |

$r$ = the percentage interest rate per compounding period

$n$ = the number of compounding periods

$d$ = the payment made per compounding period

### Working with finance solvers

When solving for N, always round *up,* never down, to the nearest whole number.

Ignore the negative sign in values when using the following formulas:

Total interest paid = N x Pmt - (PV - FV)

Total loan cost = N x Pmt

Percentage decrease in loan balance $\frac{\text{PV - FV}}{\text{PV}} \times 100\%$

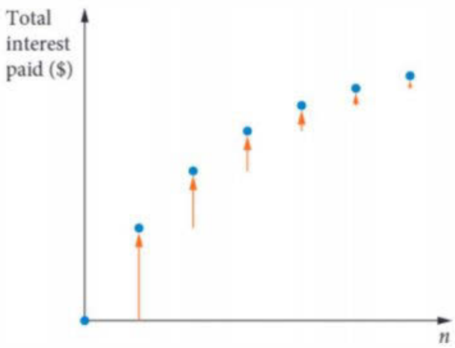
CHAPTER SUMMARY 6

### Reducing balance loan payment graphs

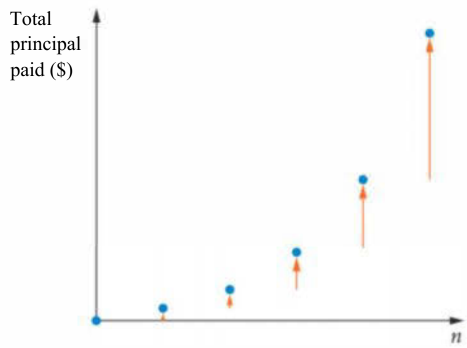
For a reducing balance loan the regular payment is part interest and part principal:

- the amount of interest ($) paid decreases with each compounding period. Towards the end of the life of the loan, the interest becomes a small part of the regular payment.
- the amount paid off the principal ($) increases with each compounding period. Towards the end of the life of the loan, the amount paid off the principal becomes a large part of the regular payment.

Graph shape of total interest paid ($)




Graph shape of total amount paid off principal ($)




### Amortisation tables

$$r = \frac{\text{interest}}{\text{previous balance}} \times 100$$

### Reducing balance loan amortisation table

| Payment number | Payment | Interest | Principal reduction | Balance |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | $V_o$ |
| $n$ | $d$ | $\frac{r}{100} \times$ previous balance | payment - interest | previous balance - principal reduction |

### Annuity amortisation table

| Payment number | Payment | Interest | Principal reduction | Balance |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | $V_o$ |
| $n$ | $d$ | $\frac{r}{100} \times$ previous balance | payment - interest | previous balance - principal reduction |

### Annuity investment amortisation table

| Payment number | Payment | Interest | Principal addition | Balance |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | $V_o$ |
| $n$ | $d$ | $\frac{r}{100} \times$ previous balance | payment + interest | previous balance + principal addition |

$n$ = payment number

$d$ = payment made per compounding period

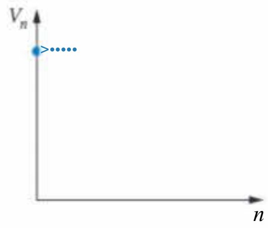
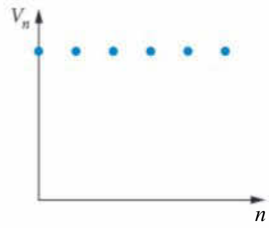
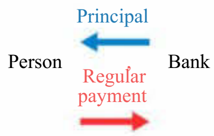
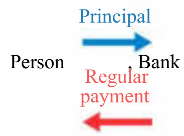
r = percentage interest rate per compounding period

$V_o$ = principal



9780170448406

## Interest-only loans and perpetuities

| | Interest-only loan | Perpetuity |
|---|---|---|
| Type | Reducing balance loan with $V_n = V_o$ | Annuity with $V_n = V_o$ |
| Formula | $d = \frac{r}{100} x V_o$ | $<* = \frac{r}{100} x V_0$ |
| Graph |   |   |
| Principal / payment |   |   |
| N | 1 | 1 |
| PV | positive | negative |
| Pmt or PMT | negative | positive |
| FV | Same as PV but with opposite sign | Same as PV but with opposite sign |

$d$ = the payment made per compounding period

$r$ = the percentage interest rate per compounding period

$V_o$ = principal

# Cumulative examination 1

Total number of marks: 13 Reading time: 5 minutes Writing time: 30 minutes

*Use the following information to answer the next two questions.*

The seasonal indices for the first 11 months of the year, for sales in a sporting equipment store, are shown in the table below.

| Month | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Seasonal index | 1.23 | 0.96 | 1.12 | 1.08 | 0.89 | 0.98 | 0.86 | 0.76 | 0.76 | 0.95 | 1.12 | |

1 ©VCAA 20141CQ10 The seasonal index for December is

A 0.89 B 0.97 C 1.02 D 1.23 E 1.29

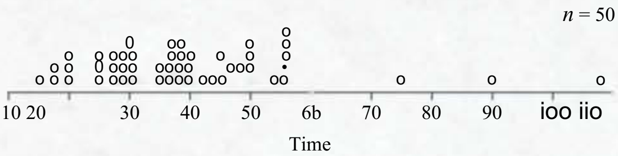
2 ©VCAA 20141CQ11 In May, the store sold $213956 worth of sporting equipment. The de-seasonalised value of these sales was closest to

A $165857 B $190420 C $209677 D $218322 E $240400

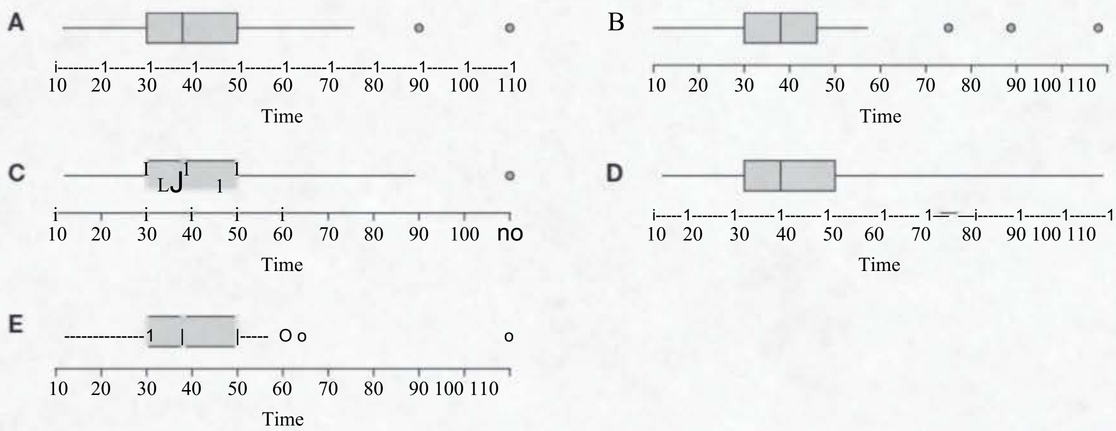
3 ©VCAA 20141CQ2 The time spent by shoppers at a hardware store on a Saturday is approximately normally distributed with a mean of 31 minutes and a standard deviation of 6 minutes. If 2850 shoppers are expected to visit the store on a Saturday, the number of shoppers who are expected to spend between 25 and 37 minutes in the store is closest to

A 16 B 68 C 460 D 1900 E 2400

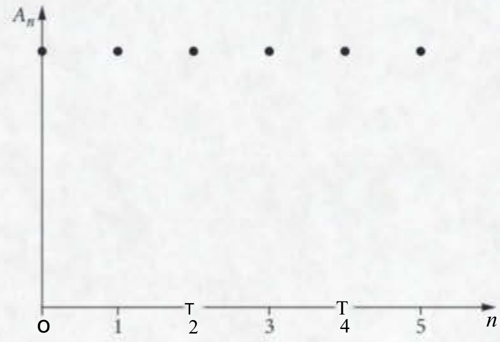
4 ©VCAA 20141CQ6 The dot plot shows the distribution of the time, in minutes, that 50 people spent waiting to get help from a call centre.




Which of the following boxplots best represents the data?



5 ©VCAA 20201CQ25 The graph below represents the value of an annuity, $A_w$, in dollars, after *n* time periods.




A recurrence relation that could match this graphical representation is

A $A_o = 200000$, $A_{n+1} = 1.015A_n - 2500$

B $A_o = 200000$, $A_{n+l} = 1.025A_n - 5000$

C $A_o = 200000$, $A_{w+1} = 1.03A_n - 5500$

D $A_o = 200000$, $A_{n+1} - 1.04A_n - 6000$

E $A_o = 200000$, $A_{n+1} = 1.05A_n - 8000$

6 ©VCAA 20161CQ20 Consider the recurrence relation

$V_o = 10000$, $V_{h+1} = 1.04\ V_n + 500$.

This recurrence relation could be used to model

A a reducing balance depreciation of an asset initially valued at $10000.

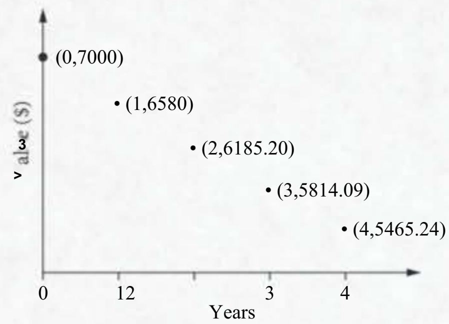
B a reducing balance loan with periodic repayments of $500.

C a perpetuity with periodic payments of $500 from the annuity.

D an annuity investment with periodic additions of $500 made to the investment.

E an interest-only loan of $ 10 000.

7 ©VCAA 2017 1CQ22 Consider the graph shown.




This graph could show the value of

A a piano depreciating at a flat rate of 6% per annum.

B a car depreciating with a reducing balance rate of 6% per annum.

C a compound interest investment earning interest at the rate of 6% per annum.

D a perpetuity earning interest at the rate of 6% per annum.

E an annuity investment with additional payments of 6% of the initial investment amount per annum.

8 ©VCAA 2019N 1CQ19 Consider the recurrence relation shown below.

$V_o = 125000, V_{M+1} = 1.013V_w - 2000$

This recurrence relation could be used to determine the value of

A a perpetuity with a payment of $2000 per quarter.

B an annuity with withdrawals of $2000 per quarter.

C an annuity investment with additional payments of $2000 per quarter.

D an item depreciating at a flat rate of 1.3% of the purchase price per quarter.

E a compound interest investment earning interest at the rate of 1.3% per annum.

9 ©VCAA | 2Q2oicQ3on Twenty years ago, Hector invested a sum of money in an account earning interest at the rate of 3.2% per annum, compounding monthly. After 10 years, he made a one-off extra payment of $10000 to the account. For the next 10 years, the account earned interest at the rate of 2.8% per annum, compounding monthly. The balance of his account today is $686 904.09. The sum of money Hector originally invested is closest to

A $355000 B $370000 C $377000 D $384000 E $385000

*Use the following information to answer the next two questions.*

Armand borrowed $12000 to pay for a holiday. He will be charged interest at the rate of 6.12% per annum, compounding monthly. This loan will be repaid with monthly repayments of $500.

10 © VCAA j 2019N1CQ22 J After four months, the total interest that Armand will have paid is closest to

A $231 B $245 C $255 D $734 E $1796

11 ©VCAA | 2019N1CQ23 J After eight repayments, Armand decided to increase the value of his monthly repayments. He will make a number of monthly repayments of $850 and then one final repayment that will have a smaller value. This final repayment has a value closest to

A $168 B $169 C $180 D $586 E $681

12 © VCAA 2°17N 1CQ21 The amortisation table below shows the repayment, interest, principal reduction and balance of a reducing balance loan after the first repayment.

| Repayment number | Repayment | Interest | Principal reduction | Balance of loan |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 180000.00 |
| 1 | 850.00 | 720.00 | 130.00 | 179870.00 |
| 2 | 850.00 | | | |

What amount of interest is paid with Repayment number 2?

A $608.56 B $609.44 C $717.12 D $719.48 E $720.00

13 ©VCAA 2021 1CQ23 Bimal has a reducing balance loan. The balance, in dollars, of the loan from month to month, $B_w$, is modelled by the recurrence relation

$B_o = 450\,000, B_{n+}\text{j} = RB_n - 2633$

Given that the loan will be fully repaid in 20 years, the value of $R$ is closest to

A 1.003 B 1.0036 C 1.03 D 1.036 E 1.36

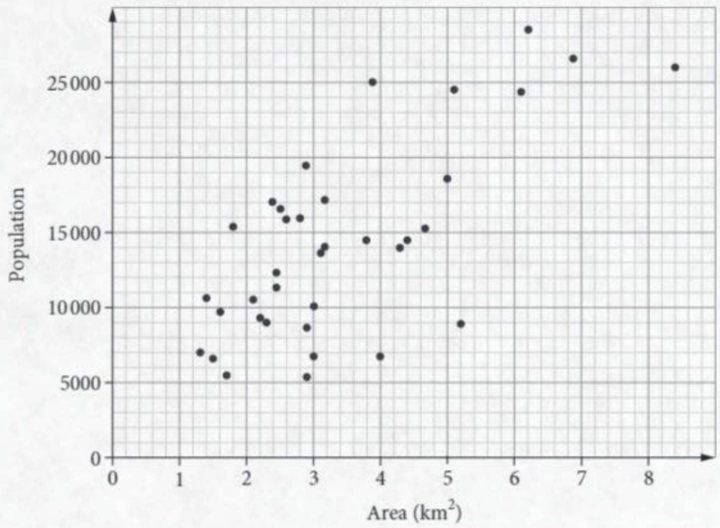
# Cumulative examination 2



Total number of marks: 20 Reading time: 6 minutes Writing time: 30 minutes

1 ©VCAA 2014 2CQ2 (6 marks)

The scatterplot below shows the *population* and *area* (in square kilometres) of a sample of inner suburbs of a large city.




The equation of the least squares line of best fit for the data in the scatterplot is

*population* = 5330 + 2680 x *area*

a Write down the dependent variable. 1 mark

b Draw the least squares line of best fit on the scatterplot. 1 mark

c Interpret the slope of this least squares line of best fit in terms of the variables *area* and *population.* 2 marks

d Wiston is an inner suburb. It has an area of 4 $km^2$ and a population of 6690.

The correlation coefficient, r, is equal to 0.668.

i Calculate the residual when the least squares line of best fit is used to predict the population of Wiston from its area. 1 mark

ii What percentage of the variation in the population of the suburbs is explained by the variation in area? Write your answer, correct to one decimal place. 1 mark

2 ©VCAA | 2020 2CQ9 J (3 marks) Samuel opens a savings account. Let $B_n$ be the balance of this savings account, in dollars, *n* months after it was opened. The month-to-month value of $B_n$ can be determined using the recurrence relation

$B_o$ = 5OOO, $B_{,, +l}$ = 1.003B,,

a Write down the value of $B_4$, the balance of the savings account after four months. Round your answer to the nearest cent. 1 mark

b Calculate the monthly interest rate percentage for Samuels savings account, 1 mark

c After one year, the balance of Samuels savings account, to the nearest dollar, is $5183. If Samuel had deposited an additional $50 at the end of each month immediately after the interest was added, how much extra money would be in the savings account after one year? Round your answer to the nearest dollar. 1 mark

3 ©VCAA 2019 2CQ9 J (4 marks) Phil would like to purchase a block of land. He will borrow $350000 to make this purchase. Interest on this loan will be charged at the rate of 4.9% per annum, compounding fortnightly. After three years of equal fortnightly repayments, the balance of Phils loan will be $262332.33.

a What is the value of each fortnightly repayment Phil will make? Round your answer to the nearest cent. 1 mark

b What is the total interest Phil will have paid after three years? Round your answer to the nearest cent. 1 mark

c Over the next four years of his loan, Phil will make monthly repayments of $3517.28 and will be charged interest at the rate of 4.8% per annum, compounding monthly. Let *B,,* be the balance of the loan *n* months after these changes apply. Write down a recurrence relation, in terms of $B_o$, $B_n + 1$ and B,,, that could be used to model the balance of the loan over these four years. 2 marks

4 ©VCAA | 2019 2CQ8] (4 marks) Phil invests $200 000 in an annuity from which he receives a regular monthly payment. The balance of the annuity, in dollars, after *n* months, $A_w$, can be modelled by the recurrence relation

$A_o = 200\ 000$, $A_{n+}$ j = 1.0035A,, - 3700

a What monthly payment does Phil receive? 1 mark

b Show that the annual percentage compound interest rate for this annuity is 4.2%. 1 mark

At some point in the future, the annuity will have a balance that is lower than the monthly payment amount.

c What is the balance of the annuity when it first falls below the monthly payment amount? Round your answer to the nearest cent. 1 mark

d If the payment received each month by Phil had been a different amount, the investment would act as a simple perpetuity. What monthly payment could Phil have received from this perpetuity? 1 mark

5 ©VCAA | 2021 2CQ8 (3 marks) For renovations to her coffee shop, Sienna took out a reducing balance loan of $570000 with interest calculated fortnightly. The balance of the loan, in dollars, after *n* fortnights, $S_M$, can be modelled by the recurrence relation

$S_o = 570000$, $S_{M+1}$ = 1.00lS,, - 1193

a Calculate the balance of this loan after the first fortnightly repayment is made, 1 mark

b Show that the compound interest rate for this loan is 2.6% per annum, 1 mark

c For the loan to be fully repaid, to the nearest cent, Siennas final repayment will be a larger amount. Determine this final repayment amount. Round your answer to the nearest cent. 1 mark

CHAPTER

7

# MATRICES AND THEIR APPLICATIONS

## Study Design coverage

### AREA OF STUDY 2: DISCRETE MATHEMATICS

#### Matrices and their applications

- matrix arithmetic: the order of a matrix, types of matrices (row, column, square, diagonal, symmetric, triangular, zero, binary and identity), the transpose of a matrix, and elementary matrix operations (sum, difference, multiplication of a scalar, product and power)
- inverse of a matrix, its determinant, and the condition for a matrix to have an inverse
- use of matrices to represent numerical information presented in tabular form, and the use of a rule for the flyth element of a matrix to construct the matrix
- binary and permutation matrices, and their properties and applications
- communication and dominance matrices and their use in analysing communication systems and ranking players in round-robin tournaments.



**Video playlists (8):**

**7.1** Matrix introduction
**7.2** Matrix elements
**7.3** Matrix addition, subtraction and scalar multiplication
**7.4** Matrix multiplication
**7.5** Inverse matrices
**7.6** Matrix applications
**7.7** Communication and dominance matrices
**VCE question analysis** Matrices and their applications

**Worksheets (2):**

**7.3** Addition and subtraction of matrices
**7.4** Multiplying matrices

9

# Nelson MindTap

To access resources above, visit cengage.com.au/nelsonmindtap

# 7.1 Matrix introduction

## The order of a matrix

Matrices are rectangular arrangements of numbers organised into rows and columns, usually presented in square brackets. The order of a matrix tells us how many rows and columns it has. We always write the order in the form *number of rows* x *number of columns,* so Matrix $A$ is a 4 x 2 (pronounced *four by two)* matrix. The numbers in the matrix are called elements. Matrix $A$ has 4x2 = 8 elements.

Column 1 Column 2

$$A = \begin{bmatrix} 32 & 18 \\ 78 & 38 \\ 27 & 12 \\ 9 & 31 \end{bmatrix} \begin{matrix} \leftarrow \text{Row 1} \\ \leftarrow \text{Row 2} \\ \leftarrow \text{Row 3} \\ \leftarrow \text{Row 4} \end{matrix}$$

### © Exam hack

The order of a matrix with $m$ rows and $n$ columns is $m$ x $n$ ('w by $n$). The number of elements in a matrix is $m$ x $n$ *Cm* times $n$). These are two different things even though they both use a x.

### Order of a matrix

Order of a matrix = *number of rows* x *number of columns*

A matrix with $m$ rows and $n$ columns has order $m$ x $n$.

7.1

Video playlist Matrix introduction

WB p. 121

**WORKED EXAMPLE 1** Understanding the order of matrices

The different types of tickets sold for a school production are shown in the following table.

| | Sold by school office | Sold online | Sold at theatre |
|---|---|---|---|
| Student ticket | 283 | 78 | 41 |
| Adult ticket | 4 | 140 | 5 |
| Concession ticket | 0 | 54 | 7 |
| Teacher ticket | 8 | 13 | 19 |

Find

| Steps | Working |
|---|---|
| **a** the matrix $P$ that could be used to show this information, stating its order and number of elements | |
| Rewrite the information in the table as a matrix. | $P = \begin{bmatrix} 283 & 78 & 41 \\ 4 & 140 & 5 \\ 0 & 54 & 7 \\ 8 & 13 & 19 \end{bmatrix}$ The order of $P$ is $4 \times 3$. $P$ has 12 elements. |
| **b** the matrix that could be used to show the number of adult tickets sold at the theatre and state its order | |
| Find the information in the table and write as a matrix. | $[5]$ The order is $1 \times 1$. |
| **c** the $4 \times 1$ matrix that could be used to show the number of tickets sold by the school office | |
| Find the information in the table and write as a matrix with the given order. | $\begin{bmatrix} 283 \\ 4 \\ 0 \\ 8 \end{bmatrix}$ |

d the 3x1 matrix that could be used to show the number of student tickets sold in each of the different ways

Find the information in the table and write as a matrix with the given order.

$$\begin{bmatrix} 283 \\ 78 \\ 41 \end{bmatrix}$$

e the 1x4 matrix that could be used to show the number of each type of ticket sold online

Find the information in the table and write as a matrix with the given order.

[ 78 140 54 13 ]

f the 4x1 matrix that could be used to show the total for each of the four types of tickets sold.

Find the information in the table and write as a matrix with the given order.

$$\begin{bmatrix} 283 + 78 + 41 \\ 4 + 140 + 5 \\ 0 + 54 + 7 \\ 8 + 13 + 19 \end{bmatrix} = \begin{bmatrix} 402 \\ 149 \\ 61 \\ 40 \end{bmatrix}$$

g Copy the labelled matrix showing the information from the table and fill in the missing numbers. The ticket types are shown by S = student, *A* = adult, C = concession and *T* = teacher.

| | S | *A* | C | *T* |
|---|---|---|---|---|
| Office | 283 | □ | 0 | □ |
| Online | 78 | 140 | □ | □ |
| Theatre | □ | □ | □ | □ |

Find the information in the table and complete the matrix.

| | *S* | *A* | *C* | *T* |
|---|---|---|---|---|
| Office | 283 | | | 408 |
| Online | 78 | 140 | 54 | 13 |
| Theatre | 41 | 5 | 7 | 19 |

## Types of matrices

| Type of matrix | Description | Examples » | Order of examples |
|---|---|---|---|
| Row matrix | A matrix with just one row. | [ -2 3 11 ]    [ 0 0 5 -3 1 | 1x3 and 1x4 |
| Column matrix | A matrix with just one column. | $\begin{bmatrix} 12 \\ -5 \end{bmatrix}$ $\begin{bmatrix} 20 \\ 35 \\ 63 \end{bmatrix}$ | 2x1 and 3x1 |
| Summing matrix | A row or column matrix where all the elements are T. | [ 1 1 1 1 1 1 1    i !d | 1x7 and 2x1 |
| Square matrix | A matrix that has the same number of rows as columns. | $\begin{bmatrix} 6 & 10 & 0 & 3 \\ -2 & 4 & 1 & -5 \\ 3 & 5 & 0 & 11 \\ -6 & 6 & 9 & 2 \end{bmatrix}$ $\begin{bmatrix} 5 & 16 \\ -3 & 0 \end{bmatrix}$ | 4x4 and 2x2 |
| Zero matrix | A matrix where all the elements are 0. | $\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$ 1    0 0 0 0 0 0 1 | 3x2 and 1x6 |
| Binary matrix | A matrix where every element is either 'O' or T. | $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$ $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ | 2x3 and 2x2 |
| Permutation matrix | A square matrix where every row and column has exactly one T, with zeros everywhere else. | $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ | 3x3 and 2x2 |



9780170448406

## WORKED EXAMPLE 2 Identifying types of matrices

For each of the following matrices, state the order and identify whether it is a row, column, summing, square, zero, binary or permutation matrix.

**a** $\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}$ **b** $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 8 \\ 1 \end{bmatrix}$ **c** $\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \end{bmatrix}$

**d** $\begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$ **e** $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

| Steps | Working |
|---|---|
| Does the matrix have<br>• just one row or just one column<br>• the same number of rows and columns<br>• all zeros<br>• just zeros and/or ones<br>• exactly one '1' in every row and column and zeros everywhere else? | **a** $4 \times 2$; binary matrix<br>**b** $5 \times 1$; column matrix<br>**c** $1 \times 5$; row matrix, summing matrix, binary matrix<br>**d** $6 \times 6$; square matrix, binary matrix, permutation matrix<br>**e** $3 \times 3$; square matrix, zero matrix, binary matrix |

7.1

WB
p. 123

## The transpose of a matrix

A transpose of a matrix is a new matrix formed by switching the rows and columns. The transpose of matrix $A$ is written as $A^T$.

The order of the transpose of a matrix is the reverse order of the original matrix. Here the original matrix was 2x3 and its transpose is 3 x 2.

$$\begin{bmatrix} 3 & 1 & 4 \\ 8 & 2 & 7 \end{bmatrix}^T = \begin{bmatrix} 3 & 8 \\ 1 & 2 \\ 4 & 7 \end{bmatrix}$$

For a square matrix, both the original matrix and its transpose have the same order. In this example, they are both 2x2.

$$\begin{bmatrix} 5 & 9 \\ 4 & 1 \end{bmatrix}^T = \begin{bmatrix} 5 & 4 \\ 9 & 1 \end{bmatrix}$$

### Transpose of a matrix

A transpose of matrix $A$ is $A^T$ where the rows and columns are switched.

If $A$ has order $m$ x H, then $A^T$ has order $n$ x $m$.

WB
p. 124

## WORKED EXAMPLE 3 Finding the transpose of matrices

Find the transpose of the following matrices and state the orders of both $A$ and $A^T$.

**a** $A = \begin{bmatrix} 0 & 3 \\ 6 & -3 \\ 2 & 9 \end{bmatrix}$  **b** $A = \begin{bmatrix} 1 & 9 & 3 & 8 \end{bmatrix}$  **c** $A = \begin{bmatrix} 7 & 1 & 0 \\ 12 & 6 & 2 \\ 3 & 0 & 10 \end{bmatrix}$

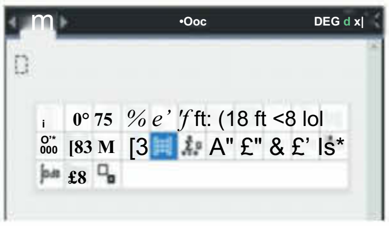
| Steps | Working | |
|---|---|---|
| **a** Switch the rows and columns of $A$. | $A^T = \begin{bmatrix} 0 & 6 & 2 \\ 3 & -3 & 9 \end{bmatrix}$ | The order of $A$ is $3 \times 2$. The order of $A^T$ is $2 \times 3$. |
| **b** Switch the rows and columns of $A$. | $A^T = \begin{bmatrix} 1 \\ 9 \\ 3 \\ 8 \end{bmatrix}$ | The order of $A$ is $1 \times 4$. The order of $A^T$ is $4 \times 1$. |
| **c** Switch the rows and columns of $A$. | $A^T = \begin{bmatrix} 7 & 12 & 3 \\ 1 & 6 & 0 \\ 0 & 2 & 10 \end{bmatrix}$ | The order of $A$ is $3 \times 3$. The order of $A^T$ is $3 \times 3$. |

## USING CAS 1 Working with matrices

For the following matrix, find its transpose.

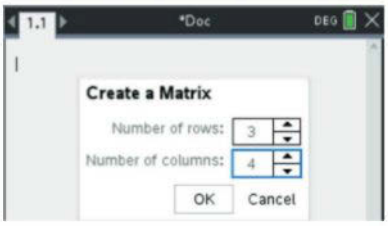
$\begin{bmatrix} 2 & -3 & 0 & 8 \\ 6 & 12 & -5 & 4 \\ -7 & 9 & 13 & 1 \end{bmatrix}$

### TI-Nspire



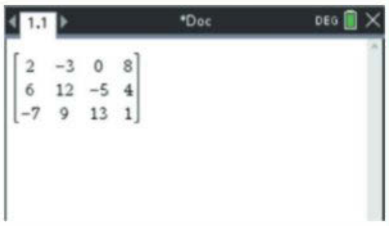
1 From a Calculator page, press the template key.

2 Select the 3x3 matrix template.



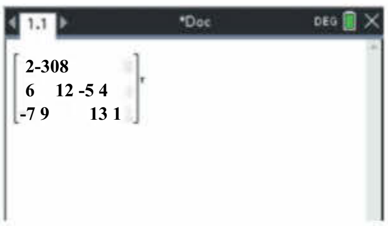

3 In the Create a Matrix screen, keep the default value of 3 rows and change the value to 4 columns.

4 Select OK.



5 Enter the values for the matrix.

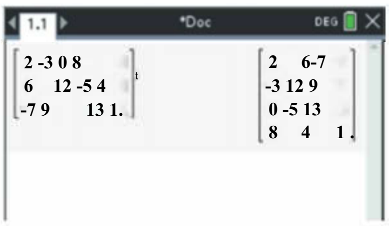
6 Move the cursor so it is outside the right bracket.



7 Press menu > Matrix & Vector > Transpose.

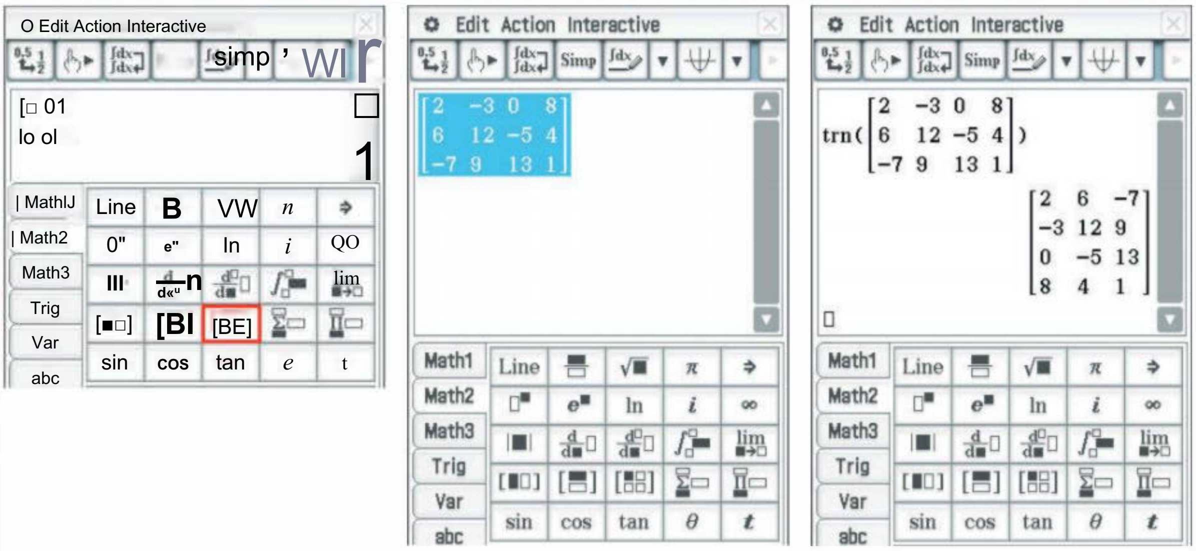
8 A small T will appear to the right of the matrix.

9 Press enter.



10 The transposed matrix will be displayed.

**ClassPad**



1 Open the Main application and clear all values.

2 Open the Keyboard and tap Math2.

3 Tap on the 2x2 matrix template twice to create a 3x3 matrix.

4 Tap on the 1x2 matrix template to add a column to create a 3x4 matrix.

5 Enter the values in the matrix.

6 Highlight the matrix.

7 Tap Interactive > Matrix > Create > trn.

8 Tap OK.

9 The transposed matrix will be displayed.

## Square matrices and leading diagonals

The leading diagonal in a square matrix is the diagonal running from the upper left to the lower right. The subdiagonal is the diagonal immediately under the leading diagonal.

Leading diagonal

$$\begin{bmatrix} 4 & 3 & -7 \\ -6 & 1 & 9 \\ 11 & 2 & 10 \end{bmatrix}$$

Subdiagonal

$$\begin{bmatrix} 4 & 3 & -7 \\ -6 & 1 & 9 \\ 11 & 2 & 10 \end{bmatrix}$$

There are special types of square matrices whose definition depends on the leading diagonal.

| Type of square matrix | Description | Examples | Order of examples |
|---|---|---|---|
| Symmetric matrix | A square matrix where the elements are symmetric around the leading diagonal. A symmetric matrix is the same as its transpose. | $\begin{bmatrix} 4 & 6 & -8 \\ 6 & 1 & 2 \\ -8 & 2 & 10 \end{bmatrix}$ $\begin{bmatrix} -8 & 11 \\ 11 & 3 \end{bmatrix}$ | 3x3 and 2x2 |
| Diagonal matrix | A square matrix where all the elements except the ones in the leading diagonal are 'O'. | $\begin{bmatrix} 9 & 0 & 0 & 0 \\ 0 & -13 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$ $\begin{bmatrix} -8 & 0 \\ 0 & 3 \end{bmatrix}$ | 4x4 and 2x2 |
| Identity matrix (Z) (also called the unit matrix) | A diagonal matrix where all the elements in the leading diagonal are '1'. We use / to indicate this matrix. | $I = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ | 4x4 and 2x2 |
| Upper triangular matrix | A square matrix where all the elements below the leading diagonal are 'O'. | $\begin{bmatrix} -3 & 8 & 2 \\ 0 & -4 & 7 \\ 0 & 0 & 5 \end{bmatrix}$ $\begin{bmatrix} 6 & -4 \\ 0 & 7 \end{bmatrix}$ | 3x3 and 2x2 |
| Lower triangular matrix | A square matrix where all the elements above the leading diagonal are 'O'. | $\begin{bmatrix} 2 & 0 & 0 & 0 \\ 7 & 6 & 0 & 0 \\ 0 & 4 & -3 & 0 \\ -6 & 1 & 15 & 8 \end{bmatrix}$ $\begin{bmatrix} 15 & 0 \\ 14 & 20 \end{bmatrix}$ | 4x4 and 2x2 |

WB
p. 125

**WORKED EXAMPLE 4** Identifying types of square matrices using leading diagonals

For each of the following matrices, list the numbers in the leading diagonal and state whether it is a symmetric, upper triangular, lower triangular, diagonal or an identity matrix. If the matrix is not any of these, explain why.

**a** $\begin{bmatrix} 3 & 0 & 0 \\ 4 & -7 & 0 \\ -2 & 6 & 8 \end{bmatrix}$ **b** $\begin{bmatrix} -9 & -5 & 0 \\ -5 & -2 & 0 \\ 0 & 0 & 6 \end{bmatrix}$ **c** $\begin{bmatrix} 8 & 0 & 0 \\ 3 & -5 & 0 \end{bmatrix}$

**d** $\begin{bmatrix} 23 & 31 \\ 0 & 18 \end{bmatrix}$ **e** $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

| Steps | Working |
|---|---|
| Is it a square matrix? | a 3, -7, 8; lower triangular matrix |
| Is the matrix the same as its transpose? | b -9, -2, 6; symmetric matrix |
| Does the matrix have<br>• all zero elements except the ones in the leading diagonal | c This is not a square matrix so it has no leading diagonal and cannot be any of the options, |
| • a leading diagonal made up of only ones | d 23,18; upper triangular matrix |
| • all zero elements below the leading diagonal<br>• all zero elements above the leading diagonal? | e 1,1,1; identity matrix, diagonal matrix, upper triangular matrix, lower triangular matrix, symmetric matrix |

## EXERCISE 7.1 Matrix introduction

ANSWERS p. 714

### Mastery

1 H WORKED EXAMPLE 1 I The quantity (in grains) of the main ingredients used to bake various are shown in the following table.

| | Type of cake | | | | |
|---|---|---|---|---|---|
| Ingredient | Chocolate | Fruit | Tea | Banana | Butter |
| Sugar | 100 | 80 | 80 | 75 | 100 |
| Flour | 225 | 125 | 150 | 175 | 150 |
| Butter | 125 | 100 | 150 | 150 | 175 |

Find

a the matrix C that could be used to show this information, stating its order and number of elements

b the matrix that could be used to show the quantity of butter used in a tea cake and state its order

c the 1x5 matrix that could beused to show the quantity of flour needed in each of the cakes

d the 5x1 matrix that could beused to show the quantity of sugar needed in each of the cakes

e the 1x3 matrix could be used to show the quantity of each ingredient needed in a banana cake

f the 3x1 matrix that could beused to show the total quantities of each main ingredient needed if one of every type of cake is made.

g Copy the following labelled matrix showing the information from the table and fill in the missing numbers. The ingredients are shown by S = sugar, $F$ = flour and $B$ = butter.

$$\begin{array}{l} \\ \text{Chocolate} \\ \text{Fruit} \\ \text{Tea} \\ \text{Banana} \\ \text{Butter} \end{array} \begin{array}{c} \begin{array}{ccc} S & F & B \end{array} \\ \begin{bmatrix} 100 & 225 & \square \\ \square & 125 & 100 \\ 80 & \square & 150 \\ \square & \square & 150 \\ \square & \square & \square \end{bmatrix} \end{array}$$

20 WORKED EXAMPLE 2 I For each of the following matrices, state the order and whether it is a row, column, summing, square, zero, binary or permutation matrix.

a $\begin{bmatrix} 0 \\ 0 \\ 1 \\ 3 \end{bmatrix}$ b $\begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}$ c $\begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}$

d $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$ e $\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$ f $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

3 WORKED EXAMPLE 3 I Find the transpose of the following matrices and state the orders of both $A$ and $A^T$.

a $A = \begin{bmatrix} 7 & 2 \\ 8 & -1 \\ 0 & 9 \\ 3 & 4 \end{bmatrix}$ b $A = \begin{bmatrix} 6 \\ 3 \\ 0 \end{bmatrix}$ c $A = \begin{bmatrix} 7 & 1 \\ 12 & 6 \end{bmatrix}$

4 S using CAS 1 For the following matrix, find its transpose. $\begin{bmatrix} -4 & 14 & 0 \\ 0 & 12 & -5 \\ -10 & 9 & 13 \\ 18 & 1 & -2 \end{bmatrix}$

▶ **5S** WORKED EXAMPLE 4 For each of the following matrices, list the numbers in the leading diagonal and state whether it is a symmetric, upper triangular, lower triangular, diagonal or an identity matrix. If the matrix is not any of these, explain why.

a
[ 5 3-9
0 8-6
0 0 2 ]

b
[ -5 0 0 0
3 -1 0 0
-4 0 7 0
6 -9 2 1 ]

c
$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

d
[ 5 0 0 0
-9 -3 0 0
8 4 2 0 ]

e
$\begin{bmatrix} 22 & 31 & 28 \\ 31 & 25 & 0 \\ 28 & 0 & 19 \end{bmatrix}$

6 State the order of each matrix, then answer the question.

a
[ 3 10
1 -6 ]
Is this a column, square or zero matrix?

b
$\begin{bmatrix} 2.1 \\ 3.7 \\ 1.5 \end{bmatrix}$
Is this a column, upper triangular or diagonal matrix?

c
$\begin{bmatrix} -3 & -9 \end{bmatrix}$
Is this a square, binary or row matrix?

d
[ 8 0
1 6 ]
Is this a permutation, lower triangular or diagonal matrix?

e
[ 1 0 1
1 1 0
0 0 1 ]
Is this a permutation, binary or symmetric matrix?

f
[ 0 0 10 6
0 0 14
10 1 8-2
6 4-2 0 ]
Is this a symmetric, lower triangular or upper triangular matrix?

7 Answer true or false to each statement.

a A 4 x 3 matrix has 43 elements.

b A 2 x 10 matrix has 20 elements.

c The transpose of a 5 x 2 matrix has 10 elements.

d A permutation matrix with three Ts has 6 elements.

e If a matrix has 5 elements, it has to be either a row matrix or a column matrix.

## Exam practice

80-100% 60-79% 0-59%

8 ©VCAA 2017 1MQ1 98% Kai has a part-time job. Each week, he earns money and saves some of this money. The matrix shows the amounts earned (E) and saved (S), in dollars, in each of three weeks.

| | E | S |
|---|---|---|
| Week 1 | 300 | 100 |
| Week 2 | 270 | 90 |
| Week 3 | 240 | 80 |

How much did Kai save in Week 2?

A $80 B $90 C $100 D $170 E $270

9 ©VCAA 2019NIMQ1 , The number of individual points scored by Rhianna (R), Suzy (S), Tina (T), Ursula (U) and Vicki (V) in five basketball matches (F, G, H, *I, J)* is shown in matrix *P* below.

Match

*F G H I J*

$$p = \begin{bmatrix} 2\ 0\ 3 \quad 1 \quad 8 \\ 4\ 7\ 2\ 5\ 3 \\ 6\ 4\ 0\ 0\ 5 \\ 16\ 14\ 5 \\ 0\ 5 \quad 3\ 2\ 0 \end{bmatrix} \begin{matrix} R \\ S \\ T \text{ Player} \\ U \\ V \end{matrix}$$

Who scored the highest number of points and in which match?

A Suzy in match *I*

B Tina in match *H*

C Vicki in match *F*

D Ursula in match G

E Rhianna in match *J*

10 ©VCAA 2007 1MQ2 93% The number of tourists visiting three towns, Oldtown, Newtown and Twixtown, was recorded for three years. The data is summarised in the table.

| | 2004 | 2005 | 2006 |
|---|---|---|---|
| Oldtown | 975 | 1002 | 1390 |
| Newtown | 2105 | 1081 | 1228 |
| Twixtown | 610 | 1095 | 1380 |

The 3x1 matrix that could be used to show the number of tourists visiting the three towns in the year 2005 is

A $\begin{bmatrix} 975\ 1002\ 1390 \end{bmatrix}$

$\begin{bmatrix} 1002 & 1081 & 1095 \end{bmatrix}$

C $\begin{bmatrix} 975 \\ 1002 \\ 1390 \end{bmatrix}$

D $\begin{bmatrix} 1002 \\ 1081 \\ 1095 \end{bmatrix}$

E $\begin{bmatrix} 975 & 1002 & 1390 \\ 2105 & 1081 & 1228 \\ 610 & 1095 & 1380 \end{bmatrix}$

11 ©VCAA 2010 1MQ1 92% The order of the matrix $\begin{bmatrix} 2 & 2 \\ 2 & 2 \\ 2 & 2 \end{bmatrix}$ is

A 2x2

B 2x3

C 3x2

D 4

E 6

12 ©VCAA 2020 1MQ1 83% The matrix $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$ is an example of

A a binary matrix.

B an identity matrix,

C a triangular matrix.

D a symmetric matrix.

E a permutation matrix.

13 ©VCAA 2016 1MQ1 81% The transpose of $\begin{bmatrix} 2 & 7\ 10 \\ 13 & 19 \quad 8 \end{bmatrix}$ is

A $\begin{bmatrix} 13\ 19 \quad 8 \\ 2 \quad 7\ 10 \end{bmatrix}$

**B** $\begin{bmatrix} 10 \quad 7 \quad 2 \\ 8\ 19\ 13 \end{bmatrix}$

C $\begin{bmatrix} 2\ 13 \\ 7\ 19 \\ 10\ 8 \end{bmatrix}$

D $\begin{bmatrix} 13 & 2 \\ 19 & 7 \\ 8\ 10 \end{bmatrix}$

E $\begin{bmatrix} 8\ 10 \\ 19 & 7 \\ 13 & 2 \end{bmatrix}$

▶ 14 ©VCAA 20121MQ7 J 72% A store has three outlets, A, *B* and C. These outlets sell dresses, jackets and skirts made by the fashion house Ocki. The table lists the number of Ocki dresses, jackets and skirts that are currently held at each outlet.

| | Size 10 | Size 12 | Size 14 | Size 16 |
|---|---|---|---|---|
| Outlet A | 2 dresses | 3 jackets | 1 skirt | 4 jackets |
| Outlet *B* | 1 skirt | 1 jacket | 3 jackets | 1 dress |
| Outlet C | 2 skirts | 2 dresses | 2 dresses | 1 jacket |

A matrix that shows the total number of Ocki dresses (D), jackets (/) and skirts (S) in each size held at the three outlets is given by

**A**
$$\begin{array}{r} \\ \text{Size 10} \\ \text{Size 12} \\ \text{Size 14} \\ \text{Size 16} \end{array} \begin{array}{c} \begin{matrix} D & / & S \end{matrix} \\ \begin{bmatrix} 2 & 1 & 2 \\ 3 & 1 & 2 \\ 1 & 3 & 2 \\ 4 & 1 & 1 \end{bmatrix} \end{array}$$

**B**
$$\begin{array}{r} \\ \text{Size 10} \\ \text{Size 12} \\ \text{Size 14} \\ \text{Size 16} \end{array} \begin{array}{c} \begin{matrix} D & J & S \end{matrix} \\ \begin{bmatrix} 2 & 0 & 3 \\ 2 & 4 & 0 \\ 2 & 3 & 1 \\ 1 & 5 & 0 \end{bmatrix} \end{array}$$

**C**
$$\begin{array}{r} \\ \text{Size 10} \\ \text{Size 12} \\ \text{Size 14} \\ \text{Size 16} \end{array} \begin{array}{c} \begin{matrix} D & / & S \end{matrix} \\ \begin{bmatrix} 2 & 3 & 1 \\ 4 & 1 & 1 \\ 3 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix} \end{array}$$

**D**
$$\begin{array}{r} \\ \text{Size 10} \\ \text{Size 12} \\ \text{Size 14} \\ \text{Size 16} \end{array} \begin{array}{c} \begin{matrix} D & J & S \end{matrix} \\ \begin{bmatrix} 7 & 0 & 0 \\ 0 & 12 & 0 \\ 0 & 0 & 4 \\ 0 & 0 & 0 \end{bmatrix} \end{array}$$

**E**
$$\begin{array}{r} \\ \text{Size 10} \\ \text{Size 12} \\ \text{Size 14} \\ \text{Size 16} \end{array} \begin{array}{c} \begin{matrix} D & J & s \end{matrix} \\ \begin{bmatrix} 1 & 0 & 2 \\ 2 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 0 \end{bmatrix} \end{array}$$

15 ©VCAA 2016S1MQ3 Consider the following four statements.

A permutation matrix is always:

i a square matrix

ii a binary matrix

iii a diagonal matrix

iv equal to the transpose of itself.

How many of the statements above are true?

A 0 B 1 C 2 D 3 E 4

16 ©VCAA 20131MQ5 MODIFIED I (4 marks) Five students, Richard (R), Brendon (B), Lee (L), Arif (A) and Karl (K), were asked whether they played each of the following sports: football (F), golf (G), soccer (S) or tennis (T). Their responses are displayed in the table.

| Student | Sport played | | | |
|---|---|---|---|---|
| | Football (F) | Golf (G) | Soccer (S) | Tennis (T) |
| R | yes | no | no | yes |
| *B* | yes | yes | yes | no |
| *L* | no | no | no | yes |
| *A* | no | yes | no | yes |
| *K* | yes | no | no | yes |

The incomplete matrix is designed to show which sport each student plays by using a T to indicate that the student plays a particular sport and a 'O' to indicate that the student does not play a particular sport.

$$\begin{array}{c} \begin{matrix} R & B & L & A & K \end{matrix} \\ \begin{bmatrix} 1 & & & & \\ 0 & & & & \\ & & & & \\ & & & & \end{bmatrix} \end{array} \begin{array}{l} \\ F \\ G \\ S \\ T \end{array}$$

a What is the order of the matrix? 1 mark

b Copy and complete the matrix. 1 mark

c What type of matrix is this? 1 mark

d Write the row matrix whose elements sum to tell us how many tennis players there are among the five students. 1 mark

# (TM) Matrix elements

7.2

## Element notation

We refer to the element of a matrix by the row and column in which it appears.

For example, in matrix $A = \begin{bmatrix} 5 & 9 \\ 8 & 4 \\ 3 & 7 \end{bmatrix}$:

a j! is the element in row 1 and column 1 $a_{11} = 5$ ← For 11 in a, p we say óne, one¹ not eleven'.

$A_{12}$ is the element in row 1 and column 2 $a_{12} = 9$

We use capital letters for matrices and matching lower case letters for elements.

**Matrix elements**

$a_{ij}$ is an element of matrix $A$, where $i$ is the row number and $j$ is the column number.

For example, the elements of a 3 × 3 matrix $A$ are written as $\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$.

Video playlist Matrix elements

**WORKED EXAMPLE 5** Using element notation

Runners' Paradise has completed a stocktake of its Run-fit running shoe brand. The matrix S gives the number of pairs of road-runners (R), trail-runners (T) and cross-trainers (C) that the store currently has in the four most popular sizes. The element in row $i$ and column $j$ of matrix S is $s$-. ij

| | R | T | C |
|---|---|---|---|
| Size 8 | 2 | 0 | 12 |
| Size 8 i | 5 | 2 | 4 |
| Size9² | 7 | 10 | 0 |
| Size $9\frac{1}{2}$ | 1 | 6 | 3 |

What information about the stock does each the following give?

a $s_{13}$ b $s_{42}$ c $s_{21} + s_{22} + s_{23}$ d $s_{11} + s_{21} + s_{31} + s_{41}$

| Steps | Working |
|---|---|
| a $s_{13}$ is the element in row 1 and column 3. | $s_{13}$ tells us there are 12 size 8 cross-trainers in stock. |
| b $s_{42}$ is the element in row 4 and column 2. | $s_{42}$ tells us there are 6 size $9\frac{1}{2}$ trail-runners in stock. |
| c $s_{21} + s_{22} + s_{23}$ is the sum of all the elements in row 2. | $s_{21} + s_{22} + s_{23}$ tells us there are a total of 5 + 2 + 4= 11 size $8\frac{1}{2}$ Run-fit running shoes in stock. |
| d $s_{11} + s_{21} + s_{31} + s_{41}$ is the sum of all the elements in column 1. | $s_{11} + s_{21} + s_{31} + s_{41}$ tells us there are a total of 2 + 5 + 7+1 = 15 road-runners in the four most popular sizes in stock. |

WB
p. 126

## Element rules

We can construct matrices using element rules.

WB p. 127

**WORKED EXAMPLE 6** Working with matrix elements

Construct the matrix, $A$, for each set of rules about the elements $a_{ij}$, where $i$ is the row number and $j$ is the column number.

| Steps | Working |
|---|---|
| **a** $A$ is a $3 \times 2$ matrix, where $a_{12} = 6$, $a_{31} = 9$, $a_{32} = 7$ and all the other elements are '2'. | |
| **1** List the elements of the matrix in $a_{ij}$ form. | $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$ |
| **2** Use the rule to find all the elements of $A$. | $a_{12} = 6, a_{31} = 9, a_{32} = 7$<br>$a_{11} = 2, a_{21} = 2, a_{22} = 2$<br>$A = \begin{bmatrix} 2 & 6 \\ 2 & 2 \\ 9 & 7 \end{bmatrix}$ |
| **b** $A$ is the transpose of $B$. $B$ is a $3 \times 2$ matrix, where $b_{11} = 7$, $b_{21} = 5$ and all the other elements are '0'. | |
| **1** List the elements of matrix $B$ in $b_{ij}$ form. | $B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}$ |
| **2** Use the rule to find all the elements of $B$. | $b_{11} = 7, b_{21} = 5$<br>$b_{12} = 0, b_{22} = 0, b_{31} = 0, b_{32} = 0$<br>$B = \begin{bmatrix} 7 & 0 \\ 5 & 0 \\ 0 & 0 \end{bmatrix}$ |
| **3** Find $A$, the transpose of $B$. | $A = \begin{bmatrix} 7 & 5 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ |
| **c** $A$ is a $1 \times 4$ matrix, where $a_{ij} = i - j$. | |
| **1** List the elements of the matrix in $a_{ij}$ form. | $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \end{bmatrix}$ |
| **2** Use the rule to find all the elements of $A$. | For the rule $a_{ij} = i - j$,<br>$A = \begin{bmatrix} 1-1 & 1-2 & 1-3 & 1-4 \end{bmatrix}$<br>$= \begin{bmatrix} 0 & -1 & -2 & -3 \end{bmatrix}$ |
| **d** $A$ is a $3 \times 3$ matrix, where $a_{ij} = 4$ when $i = j$ and $a_{ij} = 0$ when $i \neq j$. | |
| **1** List the elements of the matrix in $a_{ij}$ form. | $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$ |
| **2** Use the rule to find all the elements of $A$. | For $i = j$, $a_{ij} = 4$, so $a_{11} = a_{22} = a_{33} = 4$.<br>For $i \neq j$, $a_{ij} = 0$, so every other element is 0.<br>$A = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}$ |

## EXERCISE 7.2 Matrix elements

ANSWERS p. 715

### Recap

1 Which one of the following matrices is symmetric?

**A** $\begin{bmatrix} 3 & 4 & 15 \\ 4 & 6 & 2 & 5 \\ 1 & 2 & 6 & 10 \\ 5 & 4 & 10 & 3 \end{bmatrix}$ **B** $\begin{bmatrix} 0 & 4 & 0 & 4 \end{bmatrix}$ **C** $\begin{bmatrix} 21 & 12 \\ 12 & 17 \end{bmatrix}$

**D** $\begin{bmatrix} 3 & 3 & 3 \\ 3 & 3 & 3 \end{bmatrix}$ **E** $\begin{bmatrix} 5 & 0 \\ 0 & 5 \\ 5 & 0 \end{bmatrix}$

2 ©VCAA 20061MQ4 I 88% Three teams, Blue (B), Green (G) and Red (R), compete for three different sporting competitions. The table shows the competition winners for the past three years.

| | Athletics | Cross country | Swimming |
|---|---|---|---|
| 2004 | Green | Green | Blue |
| 2005 | Green | Red | Blue |
| 2006 | Blue | Green | Blue |

A matrix that shows the total number of competitions won by each of the three teams in each of these three years could be

**A** $\begin{matrix} & \begin{matrix} B & G & R \end{matrix} \\ \begin{matrix} 2004 \\ 2005 \\ 2006 \end{matrix} & \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \end{matrix}$ **B** $\begin{matrix} & \begin{matrix} B & G & R \end{matrix} \\ \begin{matrix} 2004 \\ 2005 \\ 2006 \end{matrix} & \begin{bmatrix} 1 & 2 & 0 \\ 1 & 1 & 1 \\ 2 & 1 & 0 \end{bmatrix} \end{matrix}$ **C** $\begin{matrix} & \begin{matrix} B & G & R \end{matrix} \\ \begin{matrix} 2004 \\ 2005 \\ 2006 \end{matrix} & \begin{bmatrix} 1 & 2 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \end{matrix}$

**D** $\begin{matrix} & \begin{matrix} B & G & R \end{matrix} \\ \begin{matrix} 2004 \\ 2005 \\ 2006 \end{matrix} & \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{matrix}$ **E** $\begin{matrix} & \begin{matrix} B & G & R \end{matrix} \\ \begin{matrix} 2004 \\ 2005 \\ 2006 \end{matrix} & \begin{bmatrix} 2 & 1 & 0 \\ 2 & 0 & 1 \\ 0 & 3 & 0 \end{bmatrix} \end{matrix}$

### Mastery

3 Ej WORKED EXAMPLE 5~] Three schools compete in a School Zone Sport Competition: Valley College (V), Forest SC (F) and Heights High (H). The matrix S shows the winning schools in four sports during a ten-year period. The element in row *i* and column *j* of matrix S is *Sy*.

$$S = \begin{matrix} & \begin{matrix} V & F & H \end{matrix} \\ \begin{matrix} \text{Football} \\ \text{Netball} \\ \text{Soccer} \\ \text{Basketball} \end{matrix} & \begin{bmatrix} 0 & 4 & 6 \\ 5 & 2 & 3 \\ 1 & 7 & 2 \\ 1 & 1 & 8 \end{bmatrix} \end{matrix}$$

What information do each the following give about the winning schools?

a $s_{32}$ b $s_{43}$ c $s_{13} + s_{23} + s_{33} + s_{43}$ d $5_{31} + s_{32} + s_{33}$

4 EJ WORKED EXAMPLE 6 I Construct the matrix, A, for each set of rules about the elements $a_{ij}$, where *i* is the row number and *j* is the column number.

a A is a 2 x 3 matrix, where $a_{13} = 8$, $a_{21} = 5$, $a_{23} = 7$ and all the other elements areT.

b A is the transpose of *B*. *B* is a 2 x 3 matrix, where $fe_{12} = 6$, $b_{23} = 9$ andall the other elements are T.

c A is a 3 x 1 matrix, where $a_{ij} = i + j$.

d A is a 3 x 3 matrix, where $a_{ij} = 0$ when $i = j$ and $= 1$ when $i * j$.

## Exam practice

80-100% 60-79% 0-59%

5 ©VCAA 2015 IMQI 85% Matrix $B$ below shows the number of photography (P), art (A) and cooking (C) books owned by Steven (S), Trevor (T), Ursula (U), Veronica (V) and William (W).

$$B = \begin{bmatrix} 8 & 5 & 4 \\ 1 & 4 & 5 \\ 3 & 3 & 4 \\ 4 & 2 & 2 \\ 1 & 4 & 1 \end{bmatrix} \begin{matrix} S \\ T \\ U \\ V \\ W \end{matrix} \quad (\text{columns } P\ A\ C)$$

The element in row $i$ and column $j$ of matrix $B$ is $by$. The element $b_{32}$ is the number of

A art books owned by Trevor.

B art books owned by Ursula,

C art books owned by Veronica.

D cooking books owned by Ursula.

E cooking books owned by Trevor.

6 IfK = $\begin{bmatrix} 9 & 12 & 0 \\ 3 & 15 & 2 \end{bmatrix}$, what is the value of $k_{13} + k_{i}$ j?

A 3 B 5 C 9 D 12 E 18

7 ©VCAA 2019N 1MQ3 In matrix A, the element $a_{21} = 7$. The order of matrix A is 3 x 2. Matrix $B$ is the transpose of matrix A. Matrix $B$ could be

A $\begin{bmatrix} 3 & 9 & 4 \\ 2 & 7 & 1 \end{bmatrix}$

B $\begin{bmatrix} 4 & 7 & 2 \\ 1 & 0 & 5 \end{bmatrix}$

C $\begin{bmatrix} 4 & 1 & 3 \\ 7 & 8 & 0 \end{bmatrix}$

D $\begin{bmatrix} 5 & 6 \\ 7 & 1 \\ 4 & 2 \end{bmatrix}$

E $\begin{bmatrix} 9 & 7 \\ 3 & 4 \\ 0 & 2 \end{bmatrix}$

8 ©VCAA 2016S 1MQ7 The order of matrix X is 3 x 2. The element in row $i$ and column $j$ of matrix $X$ is $Xy$ and it is determined by the rule $Xy = i + j$. The matrix $X$ is

A $\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$

B $\begin{bmatrix} 2 & 3 \\ 4 & 5 \\ 6 & 7 \end{bmatrix}$

C $\begin{bmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}$

D $\begin{bmatrix} 1 & 2 \\ 3 & 3 \\ 4 & 4 \end{bmatrix}$

E $\begin{bmatrix} 2 & 3 \\ 3 & 4 \\ 4 & 5 \end{bmatrix}$

9 ©VCAA 20201MQ6 73% The element in row $i$ and column $j$ of matrix $M$ is $M_{ij}$ is $M$ is a 3 x 3 matrix. It is constructed using the rule $my$- $3i + 2j$. $M$ is

A $\begin{bmatrix} 5 & 7 & 9 \\ 7 & 9 & 11 \\ 11 & 13 & 15 \end{bmatrix}$

B $\begin{bmatrix} 5 & 7 & 9 \\ 8 & 10 & 12 \\ 9 & 13 & 15 \end{bmatrix}$

C $\begin{bmatrix} 5 & 7 & 10 \\ 8 & 10 & 13 \\ 11 & 13 & 16 \end{bmatrix}$

D $\begin{bmatrix} 5 & 8 & 11 \\ 7 & 10 & 13 \\ 9 & 12 & 15 \end{bmatrix}$

E $\begin{bmatrix} 5 & 8 & 11 \\ 8 & 11 & 14 \\ 11 & 14 & 17 \end{bmatrix}$

10 ©VCAA Consider the matrix P, where P = $\begin{bmatrix} 3 & 2 & 1 \\ 5 & 4 & 3 \end{bmatrix}$. The element in row $i$ and column $j$ of matrix P is $py$. The elements in matrix P are determined by the rule

A $py$-4-$j$

B $py = 2i+l$

C $py=i+j+l$

D $p_{ir}i + 2j$

E p,-,-2«-j + 2

11 ©VCAA 20161MQ5 I 51 % Let $M = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 5 & 6 \end{bmatrix}$. The element in row *i* and column *j* of *M* is *my*.

The elements of *M* are determined by the rule

A *my* = *i* + *j* - 1 B *my* = 2*i* -*j* + 1 C *my* = 2*i*+*j* -2

D *my* = *i* + 2*j* -2 E *my-i+j+1*

12 H2 2017N1MQ4 I The element in row *i* and column *j* of matrix *M* is *my*. The elements in matrix *M* are determined using the rule *my* = 2*i* + *j*. Matrix *M* could not be

A $\begin{bmatrix} 3 \end{bmatrix}$ B [ 3 4 5 ] C $\begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix}$ D $\begin{bmatrix} 3 \\ 4 \\ 5 \\ 6 \end{bmatrix}$ E $\begin{bmatrix} 3 & 4 & 5 \\ 5 & 6 & 7 \\ 7 & 8 & 9 \end{bmatrix}$

7.3

# @ Matrix addition, subtraction and scalar multiplication

Video playlist Matrix addition, subtraction and scalar multiplication

Worksheet Addition and subtraction of matrices

## Equal matrices

For two matrices to be equal, they must have the same order *and* have all the same elements in the same places.

So if $\begin{bmatrix} 3 & 1 \\ 9 & c \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 9 & 2 \end{bmatrix}$ then *c* = 2.

But $\begin{bmatrix} 3 & 3 & 3 \\ 3 & 3 & 3 \\ 3 & 3 & 3 \end{bmatrix} \neq \begin{bmatrix} 3 & 3 & 3 \\ 3 & 3 & 3 \end{bmatrix}$

## Addition and subtraction of matrices

Only matrices that have the same order can be added or subtracted. To do this, we add or subtract each pair of corresponding elements. The answer we get has the same order as the original two matrices. For example:

$$a_{11} + b_{11} = c_{11}$$

$$\begin{bmatrix} 2 & -7 \\ 11 & 3 \end{bmatrix} + \begin{bmatrix} 1 & 3 \\ 5 & 12 \end{bmatrix} = \begin{bmatrix} 2+1 & -7+3 \\ 11+5 & 3+12 \end{bmatrix} = \begin{bmatrix} 3 & -4 \\ 16 & 15 \end{bmatrix}$$

| | | | | | | |
|---|---|---|---|---|---|---|
| | *A* | + | *B* | | = | *C* |
| Order: | 2x2 | + | 2x2 | | = | 2x2 |

$$a_{11} - b_{11} = c_{11}$$

$$\begin{bmatrix} 2 & -7 \\ 11 & 3 \end{bmatrix} - \begin{bmatrix} 1 & 3 \\ 5 & u \end{bmatrix} = \begin{bmatrix} 2-1 & -7-3 \\ 11-■ & 3-12 \end{bmatrix} = \begin{bmatrix} 1 & -10 \\ 6 & »9 \end{bmatrix}$$

| | | | | | | |
|---|---|---|---|---|---|---|
| | *A* | – | *B* | | = | *C* |
| Order: | 2x2 | – | 2x2 | | = | 2x2 |

### Adding and subtracting matrices

When adding or subtracting matrices, add or subtract pairs of corresponding elements.

For addition or subtraction of matrices to be defined, they must have the same order.

The answer has the same order as the matrices being added or subtracted.

#### Exam hack

Matrix questions often use the word 'defined'. If something is defined, it means it's possible. If it's not defined, then it's impossible.

## Scalar multiplication

A scalar is a regular number that is not in a matrix. When we multiply a matrix by a scalar, we multiply each element by the scalar. Scalar multiplication can be done to any matrix. The answer we get has the same order as the original matrix. For example:

4xan = Cii

$$4 \times \begin{bmatrix} 5 & 2 \\ 7 & -3 \\ 12 & 0 \end{bmatrix} = \begin{bmatrix} 4 \times 5 & 4 \times 2 \\ 4 \times 7 & 4 \times (-3) \\ 4 \times 12 & 4 \times 0 \end{bmatrix} = \begin{bmatrix} 20 & 8 \\ 28 & -12 \\ 48 & 0 \end{bmatrix}$$

$4A = C$

Order: $3 \times 2 = 3 \times 2$

### Multiplying a matrix by a scalar

When multiplying a matrix by a regular number called a scalar, multiply each element by the scalar.

Scalar multiplication is defined for any matrix.

The answer has the same order as the original matrix.

#### Exam hack

We don't always write the multiplication sign between the scalar and the matrix.



9780170448406

## WORKED EXAMPLE 7 Adding and subtracting matrices, and multiplying matrices by a scalar

If $A = \begin{bmatrix} 3 \\ 0 \\ -5 \end{bmatrix}$, $B = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$, $C = \begin{bmatrix} 3 & -5 & 0 \\ -1 & 6 & 4 \end{bmatrix}$ and $D = \begin{bmatrix} 1 & 0 & 7 \\ 2 & 3 & 4 \end{bmatrix}$, calculate each of the following, giving a reason if the addition or subtraction is not defined.

**a** $A - B$ **b** $C + D$ **c** $2C$

**d** $\frac{1}{5}B$ **e** $A - 4B$ **f** $2A + D$

### Steps

1 Check that the matrices have the same order.

2 Add or subtract corresponding elements.

3 Multiply each element by the scalar.

### Working

**a** $A - B = \begin{bmatrix} 3 \\ 0 \\ -5 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3-1 \\ 0-1 \\ -5-2 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ -7 \end{bmatrix}$

**b** $C + D = \begin{bmatrix} 3 & -5 & 0 \\ -1 & 6 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 7 \\ 2 & 3 & 4 \end{bmatrix} = \begin{bmatrix} 3+1 & -5+0 & 0+7 \\ -1+2 & 6+3 & 4+4 \end{bmatrix} = \begin{bmatrix} 4 & -5 & 7 \\ 1 & 9 & 8 \end{bmatrix}$

**c** $2C = 2 \times \begin{bmatrix} 3 & -5 & 0 \\ -1 & 6 & 4 \end{bmatrix} = \begin{bmatrix} 2\times 3 & 2\times -5 & 2\times 0 \\ 2\times -1 & 2\times 6 & 2\times 4 \end{bmatrix} = \begin{bmatrix} 6 & -10 & 0 \\ -2 & 12 & 8 \end{bmatrix}$

**d** $\frac{1}{5}B = \frac{1}{5} \times \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{5}\times 1 \\ \frac{1}{5}\times 1 \\ \frac{1}{5}\times 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{5} \\ \frac{1}{5} \\ \frac{2}{5} \end{bmatrix}$

**e** $A - 4B = \begin{bmatrix} 3 \\ 0 \\ -5 \end{bmatrix} - 4 \times \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ -5 \end{bmatrix} - \begin{bmatrix} 4 \\ 4 \\ 8 \end{bmatrix} = \begin{bmatrix} 3-4 \\ 0-4 \\ -5-8 \end{bmatrix} = \begin{bmatrix} -1 \\ -4 \\ -13 \end{bmatrix}$

**f** $2A + D$ is not defined because the order of $2A$ is $3 \times 1$ and the order of $D$ is $2 \times 3$. Matrices must have the same order for addition to be possible.

WB p. 129

WB
p. 130

## WORKED EXAMPLE 8 — Solving matrices using addition, subtraction and scalar multiplication

Solve each of the following.

| Steps | Working |
|---|---|
| **a** Find $c$ and $d$ if $\begin{bmatrix}6 & c\\3 & 7\end{bmatrix} + 2 \times \begin{bmatrix}2 & 5\\d & 2\end{bmatrix} = \begin{bmatrix}10 & 19\\5 & 11\end{bmatrix}$ | |
| 1 Simplify by doing the matrix addition, subtraction and scalar multiplication until there is one matrix on either side of the equal sign. | $\begin{bmatrix}6 & c\\3 & 7\end{bmatrix} + 2 \times \begin{bmatrix}2 & 5\\d & 2\end{bmatrix} = \begin{bmatrix}10 & 19\\5 & 11\end{bmatrix}$<br>$\begin{bmatrix}6 & c\\3 & 7\end{bmatrix} + \begin{bmatrix}4 & 10\\2d & 4\end{bmatrix} = \begin{bmatrix}10 & 19\\5 & 11\end{bmatrix}$<br>$\begin{bmatrix}10 & c+10\\3+2d & 11\end{bmatrix} = \begin{bmatrix}10 & 19\\5 & 11\end{bmatrix}$ |
| 2 Use the fact that two equal matrices must have all the same elements in the same places and solve, using CAS if necessary. | $c + 10 = 19$<br>$c = 19 - 10$<br>$c = 9$<br>$3 + 2d = 5$<br>$2d = 5 - 3$<br>$2d = 2$<br>$d = 1$ |
| **b** Find $M$ if $3 \times \begin{bmatrix}5 & -2\\3 & 0\end{bmatrix} + M = \begin{bmatrix}8 & 0\\4 & 1\end{bmatrix}$ | |
| 1 Simplify by doing the matrix addition, subtraction and scalar multiplication. | $3 \times \begin{bmatrix}5 & -2\\3 & 0\end{bmatrix} + M = \begin{bmatrix}8 & 0\\4 & 1\end{bmatrix}$<br>$\begin{bmatrix}15 & -6\\9 & 0\end{bmatrix} + M = \begin{bmatrix}8 & 0\\4 & 1\end{bmatrix}$ |
| 2 Solve for the unknown matrix by using matrix addition, subtraction and scalar multiplication. | $M = \begin{bmatrix}8 & 0\\4 & 1\end{bmatrix} - \begin{bmatrix}15 & -6\\9 & 0\end{bmatrix}$<br>$M = \begin{bmatrix}-7 & 6\\-5 & 1\end{bmatrix}$ |
| **c** $M$ and $N$ are both 2x3 matrices. $M$ has elements that follow the rule $m_{ij} = i - j$ and $N$ has elements that follow the rule $n_{ij} = i + 2j$. Find $M + 10N$. | |
| 1 Find $M$ using the element rule. | $M = \begin{bmatrix}m_{11} & m_{12} & m_{13}\\m_{21} & m_{22} & m_{23}\end{bmatrix}$ where $m_{ij} = i - j$.<br>$M = \begin{bmatrix}1-1 & 1-2 & 1-3\\2-1 & 2-2 & 2-3\end{bmatrix} = \begin{bmatrix}0 & -1 & -2\\1 & 0 & -1\end{bmatrix}$ |
| 2 Find $N$ using the element rule. | $N = \begin{bmatrix}n_{11} & n_{12} & n_{13}\\n_{21} & n_{22} & n_{23}\end{bmatrix}$ where $n_{ij} = i + 2j$.<br>$N = \begin{bmatrix}1+2 & 1+4 & 1+6\\2+2 & 2+4 & 2+6\end{bmatrix} = \begin{bmatrix}3 & 5 & 7\\4 & 6 & 8\end{bmatrix}$ |
| 3 Use matrix addition, subtraction and scalar multiplication. | $M + 10N = \begin{bmatrix}0 & -1 & -2\\1 & 0 & -1\end{bmatrix} + 10 \times \begin{bmatrix}3 & 5 & 7\\4 & 6 & 8\end{bmatrix}$<br>$= \begin{bmatrix}0 & -1 & -2\\1 & 0 & -1\end{bmatrix} + \begin{bmatrix}30 & 50 & 70\\40 & 60 & 80\end{bmatrix}$<br>$= \begin{bmatrix}30 & 49 & 68\\41 & 60 & 79\end{bmatrix}$ |

**WORKED EXAMPLE 9** Working with matrices using addition, subtraction and scalar multiplication

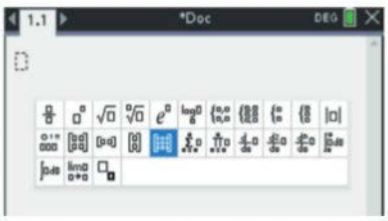
The cost prices of four different mobile phones in a store are \$740, \$960, \$1050 and \$1200. The selling price of each of these four mobile phones is 1.2 times the cost price.

| Steps | Working |
|---|---|
| a Show a matrix calculation involving a column matrix that will give the selling price of each phone. | |
| Use scalar multiplication. | $1.2 \times \begin{bmatrix} 740 \\ 960 \\ 1050 \\ 1200 \end{bmatrix}$ |
| b If the selling price also needs to allow for a \$40 commission for the salesperson, show how this can be included in the matrix calculation. | |
| Use matrix addition. | $1.2 \times \begin{bmatrix} 740 \\ 960 \\ 1050 \\ 1200 \end{bmatrix} + \begin{bmatrix} 40 \\ 40 \\ 40 \\ 40 \end{bmatrix}$ |
| c The store has a sale where mobile phones with a cost price less than \$1000 have their selling price reduced by \$50, and mobile phones with a cost price greater than \$1000 have their selling price reduced by \$100. Show how this can be included in the matrix calculation of the selling price. | |
| Use matrix subtraction. | $1.2 \times \begin{bmatrix} 740 \\ 960 \\ 1050 \\ 1200 \end{bmatrix} + \begin{bmatrix} 40 \\ 40 \\ 40 \\ 40 \end{bmatrix} - \begin{bmatrix} 50 \\ 50 \\ 100 \\ 100 \end{bmatrix}$ |

**USING CAS 2** Addition, subtraction and scalar multiplication of matrices

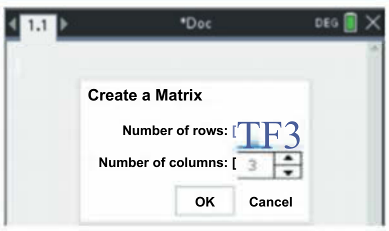
Given that $Q = \begin{bmatrix} 3 & -5 & 10 \\ 11 & 6 & 17 \\ -2 & 15 & 0 \end{bmatrix}$ and $R = \begin{bmatrix} 12 & 2 & 1 \\ -3 & 4 & 7 \\ 22 & 14 & -7 \end{bmatrix}$, evaluate $4Q - 2R$ and $3Q + 7R$.

**TI-Nspire**



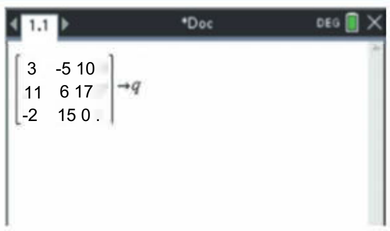
1 From a Calculator page, press the template key.

2 Select the 3x3 matrix template.




3 In the Create a Matrix screen, keep the default values of 3 rows and 3 columns.

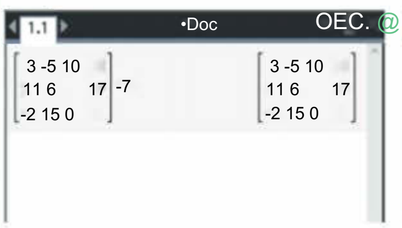
4 Select OK.



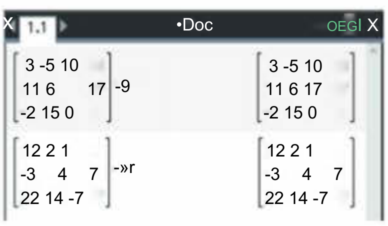
5 Enter the values for matrix Q.

6 Press Ctrl > var to store the matrix as the variable *q*.

7 Press enter.

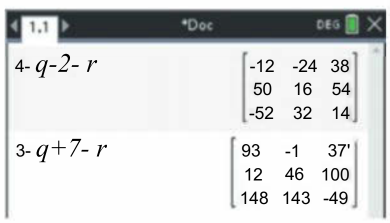


8 The matrix is now stored as the variable *q*.



9 Create a new 3x3 matrix and enter the values for matrix *R*.

10 Press Ctrl > var to store the matrix as the variable r.

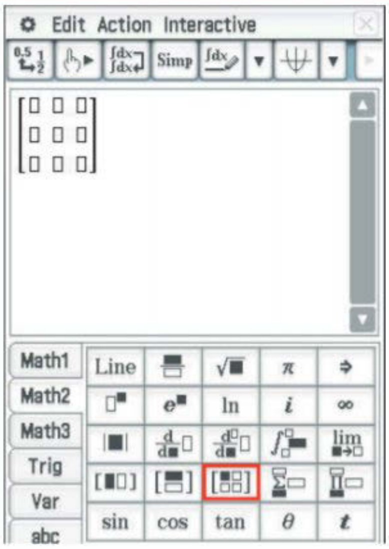


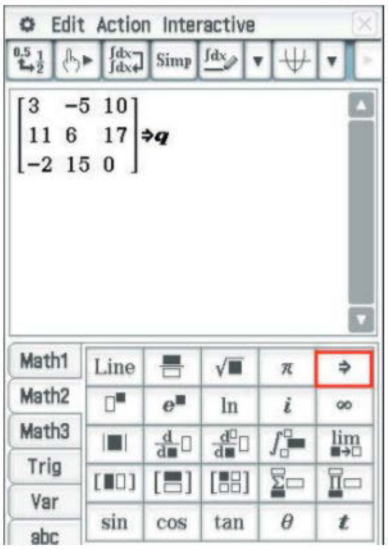
11 Use the matrices stored in *q* and *r* to complete the matrix operations.

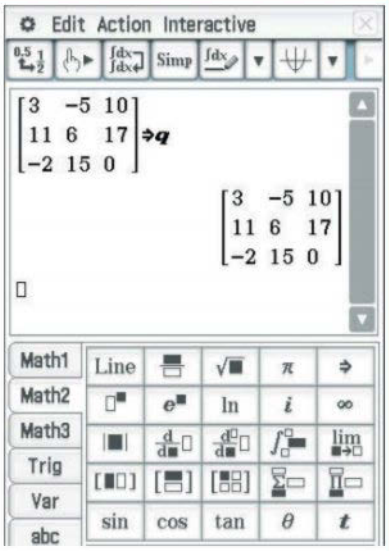
7.3

WB p. 132

ClassPad




1 Open the Main application and clear all values.

2 Open the Keyboard and tap Math2.

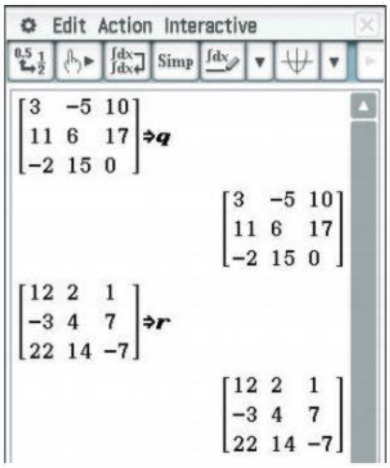
3 Tap on the 2x2 matrix template twice to create a 3x3 matrix.

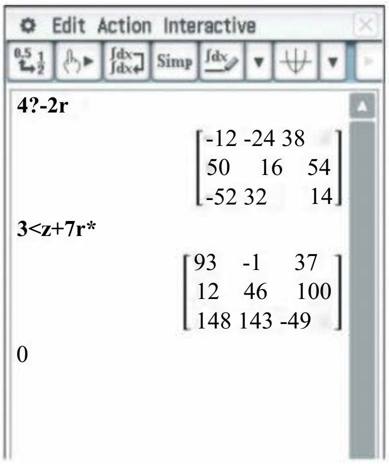
4 Enter the values for matrix Q.

5 Tap the store arrow to store the matrix as the variable *q*.

6 Press EXE.

7 The matrix is now stored as the variable *q*.




8 Create a new 3x3 matrix and enter the values for matrix *R*.

9 Tap the store arrow to store the matrix as the variable *r*.

10 Use the matrices stored in *q* and *r* to complete the matrix operations.



9780170448406

## EXERCISE 7.3 Matrix addition, subtraction and scalar multiplication

ANSWERS p. 715

7.3

**Recap**

1 If $M = \begin{bmatrix} 6 & 9 & 2 \\ 4 & 10 & 1 \\ 3 & 5 & 7 \end{bmatrix}$, what is the value of $m_{23}$ + AM$_{31}$?

A 4 B 7 C 8 D 9 E 12

2 For $Q = \begin{bmatrix} 3 & 5 \\ 4 & 6 \\ 5 & 7 \end{bmatrix}$, the element in row *i* and column *j* of matrix Q is *q*^. The elements in matrix Q are determined by the rule

A «i, = 4-y B <?y = 2i+; C $_{qij}$ = *i*+*j*+*i*

D fy = i + 2j E <7y = 2'-; + 2

**Mastery**

3 Kj WORKED EXAMPLE 7] Given the matrices $A = \begin{bmatrix} -4 & 5 \\ 0 & 6 \end{bmatrix}$, $B = C = \begin{bmatrix} 1 & 6 \end{bmatrix}$ and $D = \begin{bmatrix} 3 & 5 \\ 0 & 2 \end{bmatrix}$, calculate each of the following, giving a reason if the addition or subtraction is not defined.

a *A-D* b *C+D* c 4C

d $\frac{1}{7}B$ e C- 6B f *3A* + *D*

4H WORKED EXAMPLE & 1 Solve each of the following.

a Find *e* and/if $\begin{bmatrix} e & 6 \\ 5 & 3 \end{bmatrix} + 3 \times \begin{bmatrix} -4 & -1 \\ 1 & f \end{bmatrix} = \begin{bmatrix} -10 & 3 \\ 8 & -6 \end{bmatrix}$

b Find *M* if $4 \times \begin{bmatrix} -1 & 5 \\ -3 & 2 \end{bmatrix} + M = \begin{bmatrix} 6 & 15 \\ 1 & 8 \end{bmatrix}$

c *M* and *N are* two 3x3 matrices. *M* has elements that follow the rule = *i*+*j* and *N* has elements that follow the rule = 3*i* -*j*. Find *M* + *5N.*

5 H WORKED EXAMPLE 9 The cost prices of three different Bluetooth speakers in a store are $50, $120 and $840. The selling price of each of these three Bluetooth speakers is 1.5 times the cost price,

a Show a matrix calculation involving a column matrix that will give the selling price of each speaker,

b If the selling price also needs to allow for a $20 commission for the salesperson, show how this can be included in the matrix calculation.

c The store has a sale where Bluetooth speakers with a cost price less than $500 have their prices reduced by $10 and Bluetooth speakers with a cost price greater than $500 have their prices reduced by $40. Show how this can be included in the matrix calculation of the selling price.

6 using CAS 2 J Given that $M = \begin{bmatrix} 5 & 9 \\ 3 & 8 \end{bmatrix}$ and $N = \begin{bmatrix} 3 & -2 \\ 5 & 4 \end{bmatrix}$, evaluate the following using CAS.

a *2M* + *2N* b *3N-2M* c *5M* + *4N* d *6M* - *4N*

## Exam practice

80-100% 60-79% 0-59%

7 ©VCAA 2007 1MQ1 I 97% The matrix sum $\begin{bmatrix} 0 & -4 \\ 2 & 5 \end{bmatrix} + \begin{bmatrix} 5 & 4 \\ -2 & 2 \end{bmatrix}$ is equal to

A $\begin{bmatrix} 5 & 0 \\ 0 & 7 \end{bmatrix}$ B $\begin{bmatrix} 0 & 0 \\ 0 & 7 \end{bmatrix}$ C $\begin{bmatrix} 5 & -4 \\ 0 & 7 \end{bmatrix}$

D $\begin{bmatrix} 0 & 5 & -4 & 4 \\ 2 & -2 & 5 & 2 \end{bmatrix}$ E $\begin{bmatrix} 0-4 & 5\ 4 \\ 2\ \ 5 & -2\ 2 \end{bmatrix}$

8 © VCAA 2008 1MQ1 94% If $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 3 & 7 \\ 8 & d \end{bmatrix} = \begin{bmatrix} 4 & 7 \\ 8 & 11 \end{bmatrix}$, then $d$ is equal to

A −11 B -10 C 7 D 10 E 11

9 © VCAA 2019 1MQ1 87% Consider the following four matrix expressions.

$\begin{bmatrix} 8 \\ 12 \end{bmatrix} + \begin{bmatrix} 4 \\ 2 \end{bmatrix}$ $\begin{bmatrix} 8 \\ 12 \end{bmatrix} + \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}$ $\begin{bmatrix} 8 & 0 \\ 12 & 0 \end{bmatrix} + \begin{bmatrix} 4 \\ 2 \end{bmatrix}$ $\begin{bmatrix} 8 & 0 \\ 12 & 0 \end{bmatrix} + \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}$

How many of these four matrix expressions are defined?

A0 BI C 2 D 3 E 4

10 ©VCAA 2017N1MQ1 $\begin{bmatrix} 4 & 2 \\ 1 & 5 \end{bmatrix} + 3\begin{bmatrix} 4 & 2 \\ 1 & 5 \end{bmatrix}$ is equal to

A $\begin{bmatrix} 8 & 4 \\ 2 & 10 \end{bmatrix}$ B $\begin{bmatrix} 11 & 1 \\ 5 & 13 \end{bmatrix}$ C $\begin{bmatrix} 16 & 8 \\ 4 & 20 \end{bmatrix}$ D $\begin{bmatrix} 24 & 12 \\ 6 & 30 \end{bmatrix}$ E $\begin{bmatrix} 38 & 34 \\ 31 & 40 \end{bmatrix}$

11 ©VCAA | 20081MQ3 72% The cost prices of three different electrical items in a store are $230, $290 and $310 respectively. The selling price of each of these three electrical items is 1.3 times the cost price plus a commission of $20 for the salesperson. A matrix that lists the selling price of each of these three electrical items is determined by evaluating

A $1.3 \times \begin{bmatrix} 230 \\ 290 \\ 310 \end{bmatrix} + [20]$ B $1.3 \times \begin{bmatrix} 230 \\ 290 \\ 310 \end{bmatrix} + 1.3 \times 20$ c $1.3 \times \begin{bmatrix} 230 \\ 290 \\ 310 \end{bmatrix} + \begin{bmatrix} 20 \\ 20 \\ 20 \end{bmatrix}$

D $1.3 \times \begin{bmatrix} 230 \\ 290 \\ 310 \end{bmatrix} + 1.3 \times \begin{bmatrix} 20 \\ 20 \\ 20 \end{bmatrix}$ E $1.3 \times \begin{bmatrix} 230 + 20 \\ 290 + 20 \\ 310 + 20 \end{bmatrix}$

12 ©VCAA 2017 1MQ6 58% The table shows information about two matrices, $A$ and $B$.

| Matrix | Order | Rule |
|---|---|---|
| $A$ | $3 \times 3$ | $a_{ij} = 2i + j$ |
| $B$ | $3 \times 3$ | $b_{ij} = i - j$ |

The element in row $i$ and column; of matrix $A$ is $a$^. The element in row $i$ and column; of matrix $B$ is *by*. The sum $A + B$ is

A $\begin{bmatrix} 5 & 7 & 9 \\ 8 & 10 & 12 \\ 11 & 13 & 15 \end{bmatrix}$ B $\begin{bmatrix} 5 & 8 & 11 \\ 7 & 10 & 13 \\ 9 & 12 & 15 \end{bmatrix}$ C $\begin{bmatrix} 3 & 6 & 9 \\ 3 & 6 & 9 \\ 3 & 6 & 9 \end{bmatrix}$ D $\begin{bmatrix} 3 & 3 & 3 \\ 6 & 6 & 6 \\ 9 & 9 & 9 \end{bmatrix}$ E $\begin{bmatrix} 3 & 6 & 3 \\ 6 & 3 & 9 \\ 3 & 9 & 3 \end{bmatrix}$

▶ 13 ©VCAA 2018N 1MQ2 Consider the matrix equation $2 \times \begin{bmatrix} 3 & 0 \\ 4 & -1 \end{bmatrix} + W = \begin{bmatrix} 6 & 2 \\ 7 & 0 \end{bmatrix}$. Which one of the following is matrix W?

A $\begin{bmatrix} 0 & 2 \\ 1 & -2 \end{bmatrix}$ B $\begin{bmatrix} 0 & 2 \\ -1 & 2 \end{bmatrix}$ C $\begin{bmatrix} 0 & -2 \\ 1 & -2 \end{bmatrix}$ D $\begin{bmatrix} 12 & 2 \\ 15 & -2 \end{bmatrix}$ E $\begin{bmatrix} 12 & -2 \\ 15 & -2 \end{bmatrix}$

14 ©VCAA | 20151MQ8 43% The order of matrix X is 2 x 3. The element in row $i$ and column $j$ of matrix X is $x_{ij}$ and it is determined by the rule $x_{ij} = i - j$. Which one of the following calculations would result in matrix X?

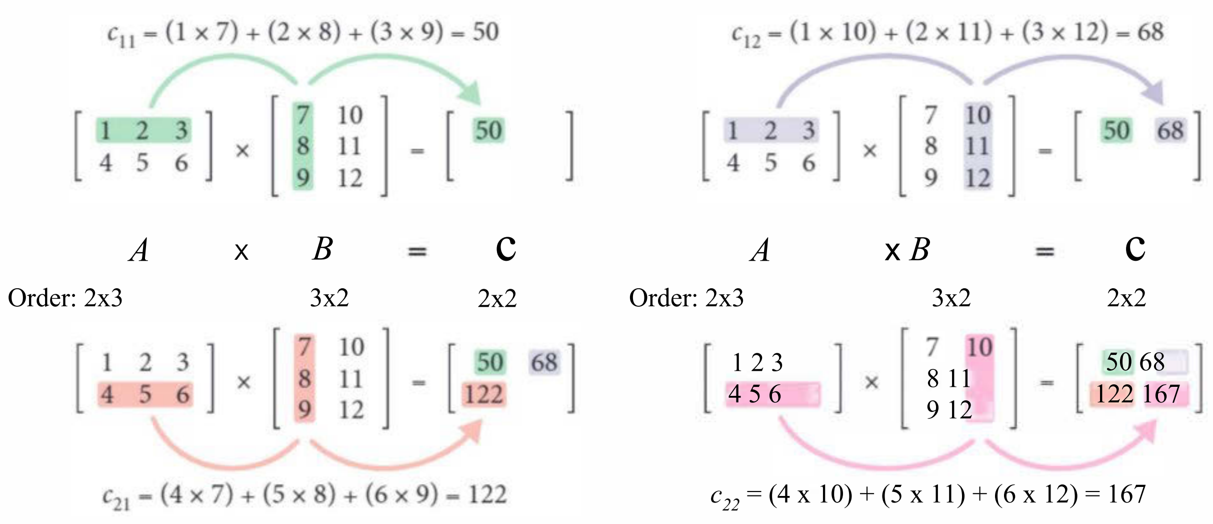
A $\begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \end{bmatrix} - \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}$ B $\begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix} - \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \end{bmatrix}$ C $\begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix} - \begin{bmatrix} 3 & 3 & 3 \\ 3 & 3 & 3 \end{bmatrix}$

D $\begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{bmatrix} - \begin{bmatrix} 1 & 1 \\ 2 & 2 \\ 3 & 3 \end{bmatrix}$ E $\begin{bmatrix} 1 & 1 \\ 2 & 2 \\ 3 & 3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{bmatrix}$

# @ Matrix multiplication

## Multiplying matrices

Matrix multiplication involves multiplying the elements of each row in the first matrix by the elements of each column in the second matrix, and then adding them. For example:




**© Exam hack**

We don't always write the multiplication sign between matrices.

7.4

Video playlist Matrix multiplication

Worksheet Multiplying matrices

## Matrix multiplication order

Matrix multiplication $AB = C$ is only defined if the number of columns in $A$ is the same as the number of rows in $B$. The product matrix C has the same number of rows as $A$ and the same number of columns as $B$. We can show this with a matrix order equation.

$$AB = C$$

Order: $(m \times n)\ (n \times q) = (m \times q)$

Product is defined

> **Exam hack**
>
> Order is important when multiplying matrices! Except in some special cases, $AB \neq BA$.

## Powers of matrices

Matrices can be raised to powers in a similar way to regular numbers. For a matrix $A$:

$A^2 = A \times A$, $A^3 = A \times A \times A$ and so on

Only square matrices can be raised to a power. For non-square matrices the product is not defined.

$$AA = A^2$$

Order: $(m \times m)\ (m \times m) = (m \times m)$

Product is defined for square metrices

$$AA$$

Order: $(m \times n)\ (m \times n)$

Product is not defined for non-square matrices

> **Exam hack**
>
> The only time it's correct to put brackets around the order of a matrix is in a matrix order equation.

**Multiplying matrices**

When multiplying matrices $AB = C$:

- multiply the elements of each row in $A$ by the elements of each column in B, and then add them
- the product is defined if the number of columns in $A$ equals the number of rows in $B$
- C has the same number of rows as $A$ and the same number of columns as B.

When raising matrices to powers:

- only powers of square matrices are defined
- the resulting matrix will always have the same order as the original matrix.

WB p. 133

**WORKED EXAMPLE 10** Multiplying matrices

If $A = \begin{bmatrix} 2 & 6 \\ 0 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 3 & 5 & 1 \\ 2 & 4 & 7 \end{bmatrix}$, $C = \begin{bmatrix} 3 & 7 & 2 \end{bmatrix}$ and $D = \begin{bmatrix} 5 \\ 9 \\ 7 \end{bmatrix}$, for each of the following

**a** $AB$ **b** $BA$ **c** $BC$ **d** $BD$

**e** $CD$ **f** $B^3$ **g** $A^2 + 5A$

**i** state whether or not the expression is defined, giving a reason.

For those that are defined

**ii** state the order of the answer before performing the calculation

**iii** compute the matrix expressions to find the answer.

| **Steps** | **Working** |
|---|---|
| a i Do the number of columns in A equal the number of rows in B? | A has order 2 x 2, *B* has order 2x3.<br>number of columns in A = number of rows in *B*<br>So *AB* is defined. |
| ii How many rows does *A* have?<br>How many columns does *B* have? | (2 x 2)(2 x 3) = (2 x 3)<br>*AB* has order 2x3.<br>Note: Giving the answer as (2 x 3) is incorrect. You can only use brackets in the working. |
| iii Calculate *AB*. | $AB = \begin{bmatrix} 2 & 6 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 3 & 5 & 1 \\ 2 & 4 & 7 \end{bmatrix}$<br>$= \begin{bmatrix} (2\text{x}3)+(6\text{x}2) = 18 & (2 \text{ x } 5) + (6 \text{ x } 4) = 34 & (2 \text{ x } 1) + (6 \text{ x } 7) = 44 \\ (0 \text{ x } 3) + (1 \text{ x } 2) = 2 & (0 \text{ x } 5) + (1 \text{ x } 4) = 4 & (0 \text{ x } 1) + (1 \text{ x } 7) = 7 \end{bmatrix}$<br>$= \begin{bmatrix} 18 & 34 & 44 \\ 2 & 4 & 7 \end{bmatrix}$ |
| b i Do the number of columns in *B* equal the number rows in A? | *B* has order 2 x 3, A has order 2x2.<br>number of columns in *B* * number of rows in A<br>So *BA* is not defined. |
| c i Do the number of columns in *B* equal the number rows in C? | *B* has order 2 x 3, C has order 1x3.<br>number of columns in *B* * number of rows in C<br>So *BC* is not defined. |
| d i Do the number of columns in *B* equal the number rows in D? | *B* has order 2 x 3, *D* has order 3x1.<br>number of columns in B = number of rows in *D*<br>So *BD* is defined. |
| ii How many rows does *B* have?<br>How many columns does *D* have? | (2 x 3)(3 x 1) = (2 x 1)<br>*BD* has order 2x1. |
| iii Calculate *BD*. | $BD = \begin{bmatrix} 3 & 5 & 1 \\ 2 & 4 & 7 \end{bmatrix}\begin{bmatrix} 5 \\ 9 \\ 7 \end{bmatrix} = \begin{bmatrix} (3 \text{ x } 5) + (5 \text{ x } 9) + (1 \text{ x } 7) \\ (2\text{x}5)+(4\text{x}9)+(7\text{x}7) \end{bmatrix} = \begin{bmatrix} 67 \\ 95 \end{bmatrix}$ |
| *e* i Do the number of columns in C equal the number rows in D? | C has order 1 x 3, D has order 3x1.<br>number of columns in C = number of rows in *D*<br>So *CD* is defined. |
| ii How many rows does C have?<br>How many columns does *D* have? | (1 x3)(3x 1) = (1 x 1)<br>*CD* has order lxl. |
| iii Calculate *CD*. | CD-[ 3 7 2 ] $\begin{bmatrix} 5 \\ 9 \\ 7 \end{bmatrix}$ -[ (3 x 5) + (7 x 9) + (2 x 7) ] = [64] |

| | |
|---|---|
| f i Is *B* square matrix? | *B* is a not square matrix.<br>Only powers of square matrices are defined, so $B^3$ is not defined. |
| g i Is *A* square matrix? | A is a square matrix.<br>Powers of square matrices are always defined, so $A^2$ is defined. |
| ii How many rows does *A* have?<br>How many columns does *A* have?<br>Is the sum possible? | (2 x 2)(2 x 2) = (2 x 2)<br>$A^2$ has order 2x2.<br>5A has order 2x2.<br>Matrices must have the same order to be added, so $A^2 + 5A$ is defined. |
| iii Calculate $A^2 + 5A$. | $A^2 = \begin{bmatrix} 2 & 6 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 2 & 6 \\ 0 & 1 \end{bmatrix}$<br>$= \begin{bmatrix} (2 \times 2) + (6 \times 0) & (2 \times 6) + (6 \times 1) \\ (0 \times 2) + (1 \times 0) & (0 \times 6) + (1 \times 1) \end{bmatrix}$<br>$= \begin{bmatrix} 4 & 18 \\ 0 & 1 \end{bmatrix}$<br>$A^2 + 5A = \begin{bmatrix} 4 & 18 \\ 0 & 1 \end{bmatrix} + 5\begin{bmatrix} 2 & 6 \\ 0 & 1 \end{bmatrix}$<br>$= \begin{bmatrix} 4 & 18 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 10 & 30 \\ 0 & 5 \end{bmatrix}$<br>$= \begin{bmatrix} 14 & 48 \\ 0 & 6 \end{bmatrix}$ |

WB
p. 135

**WORKED EXAMPLE 11** Working with matrix multiplication

Show each of the following.

| Steps | Working |
|---|---|
| **a** Show that for the matrix equation $\begin{bmatrix} 2 & 1 \\ 5 & 1 \end{bmatrix}\begin{bmatrix} 4 \\ a \end{bmatrix} = \begin{bmatrix} 3 & -2 \\ 6 & -1 \end{bmatrix}\begin{bmatrix} b \\ 2 \end{bmatrix}$ to be true, the equations $a = 3b - 12$ and $a = 6b - 22$ must be true. | |
| **1** Multiply the matrices on both sides of the equation. | $\begin{bmatrix} 2 & 1 \\ 5 & 1 \end{bmatrix}\begin{bmatrix} 4 \\ a \end{bmatrix} = \begin{bmatrix} (2 \times 4) + (1 \times a) \\ (5 \times 4) + (1 \times a) \end{bmatrix} = \begin{bmatrix} 8 + a \\ 20 + a \end{bmatrix}$<br>$\begin{bmatrix} 3 & -2 \\ 6 & -1 \end{bmatrix}\begin{bmatrix} b \\ 2 \end{bmatrix} = \begin{bmatrix} (3 \times b) + (-2 \times 2) \\ (6 \times b) + (-1 \times 2) \end{bmatrix} = \begin{bmatrix} 3b - 4 \\ 6b - 2 \end{bmatrix}$ |
| **2** Equate the two matrix products and use the fact that corresponding elements must be equal.<br>Rearrange each of the equations to make *a* the subject. | If $\begin{bmatrix} 8 + a \\ 20 + a \end{bmatrix} = \begin{bmatrix} 3b - 4 \\ 6b - 2 \end{bmatrix}$ then<br>$8 + a = 3b - 4$ $\quad$ $20 + a = 6b - 2$<br>$a = 3b - 12$ $\quad$ $a = 6b - 22$ |


9780170448406

b Show that if S is a square matrix, matrix $R$ is a row matrix, and matrix C is a column matrix, then $RSC$ is a 1 x 1 matrix.

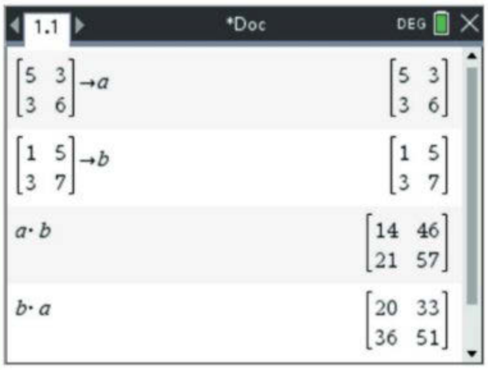
| | |
|---|---|
| 1 Write the orders of each matrix. | Let S have order $m \times m$.<br>For RS to be defined R needs to have order $1 \times m$.<br>For $SC$ to be defined C needs to have order $m \times 1$. |
| 2 Use the fact that the product matrix has the same number of rows as the first matrix and the same number of columns as the second matrix. | $RSC$ has order<br>$RSC$<br>$(1 \times m)(m \times m)(m \times 1)$<br>RS C<br>$(1 \times m)(m \times 1)$<br>RSC<br>$(1 \times 1)$<br>So, RSC is a 1 x 1 matrix. |

## USING CAS 3 Multiplication and powers of matrices

If $A = \begin{bmatrix} 5 & 3 \\ 2 & 6 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & -2 \\ 0 & 3 \end{bmatrix}$, find the following.

**a** $AB$ **b** $B^8$ **c** $3A^2 - B^3A$

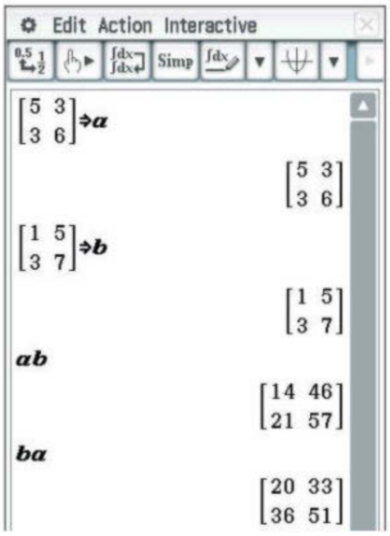
**TI-Nspire**




1 Create the two matrices and store them as variables ***a*** and ***b***.

2 Perform the calculations as shown above. Make sure you insert a multiplication sign between the ***a*** and ***b***.

**ClassPad**




1 Create the two matrices and store them as variables ***a*** and ***b***.

2 Perform the calculations as shown above. When you use variables instead of letters, there is no need to insert multiplication signs.

## Multiplying summing matrices

As we have seen, summing matrices are row or column matrices where all the elements areT. If we multiply them with other matrices, they sum the rows or columns of the other matrix. We can also use them to calculate the means of the rows or columns of the other matrix.

## Row summing matrix

A *row* summing matrix on the *left* sums each of the *columns* of another matrix.

$$\begin{bmatrix} 1 & 1 \end{bmatrix}\begin{bmatrix} 10 & 2 & 1 \\ 3 & 1 & 5 \end{bmatrix} \quad [(lx\ 10)+(1x3) \quad (lx\ 2)+(1x7)\ (lxl)+(1x5)] - [\ 13\ 9\ b]$$

Order: (1x2) (2x3) (1x3)

Multiply by $\frac{1}{2}$ to calculate the mean of each column:

$$\frac{1}{2}\begin{bmatrix} 1 & 1 \end{bmatrix}\begin{bmatrix} 10 & 2\ 1 \\ 3 & 7\ 5 \end{bmatrix} - \begin{bmatrix} \frac{13}{2} & \frac{9}{2} & \frac{6}{2} \end{bmatrix} - \begin{bmatrix} 6.5 & 4.5\ 3 \end{bmatrix}$$

## Column summing matrix

A *column* summing matrix on the *right* sums each of the *rows* of another matrix.

$$\begin{bmatrix} 2 & 4\ 3 \\ 5 & 1\ 9 \\ 6\ 20\ 7 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} (2x\ l)+(4x\ l)+(3x\ 1) \\ (5x\ 1)+(1\ x\ l)+(9x\ 1) \\ (6x\ 1)+(20\ x\ l)+(7x\ 1) \end{bmatrix} - \begin{bmatrix} 9 \\ 15 \\ 33 \end{bmatrix}$$

Order: (3 x 3) (3 x 1) (3x1)

Multiply by $\frac{1}{3}$ to calculate the mean of each row:

$$\frac{1}{3}\begin{bmatrix} 2 & 4 & 3 \\ 5 & 1 & 9 \\ 6 & 20 & 7 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} \frac{9}{3} \\ \frac{15}{3} \\ \frac{33}{3} \end{bmatrix} - \begin{bmatrix} 3 \\ 5 \\ 11 \end{bmatrix}$$

**Summing matrices**

| | To sum each of the *columns* of an $m$ x $n$ matrix | To sum each of the *rows* of an $m$ x $n$ matrix |
|---|---|---|
| Type of summing matrix | row | column |
| Multiply on the | left | right |
| Order of summing matrix | lxm | wx 1 |
| Order of multiplication | (1 x *m)(m* x *n)* | *(m* x n)(n x 1) |
| Order of answer | 1 x *n* | *mxl* |
| To find means multiply by | $\frac{1}{m}$ | $\frac{1}{n}$ |


9780170448406

7.4

WB p. 136

**WORKED EXAMPLE 12** Multiplying summing matrices

For the matrices $X = \begin{bmatrix} 3 & 2 \\ 4 & 6 \\ 5 & 10 \end{bmatrix}$ and $Y = \begin{bmatrix} 10 & 2 & 1 & 4 \\ 3 & 5 & 0 & 6 \end{bmatrix}$, show how to use a summing matrix S to calculate

a the sums and means of the columns of $X$

b the sums and means of the rows of $Y$.

| Steps | Working |
|---|---|
| a 1 Find $m$ and $n$ for $X$. | $X$ has order 3 x 2, so $m = 3$ and $n = 2$. |
| 2 What sort of summing matrix is needed? | We want to sum columns, so S is a 1 x 3 row summing matrix multiplied on the left. |
| 3 Multiply the matrices to find the sums of each of the columns, using CAS if necessary. | $\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 4 & 6 \\ 5 & 10 \end{bmatrix}$ $= \begin{bmatrix} (1 \times 3) + (1 \times 4) + (1 \times 5) & (1 \times 2) + (1 \times 6) + (1 \times 10) \end{bmatrix}$ $= \begin{bmatrix} 12 & 18 \end{bmatrix}$ |
| 4 What do we need to multiply by to find the means of each of the columns? | $\frac{1}{3}$ |
| 5 Perform the scalar multiplication to find the means of each of the columns, using CAS if necessary. | $\frac{1}{3}\begin{bmatrix} 12 & 18 \end{bmatrix} = \begin{bmatrix} \frac{12}{3} & \frac{18}{3} \end{bmatrix} = \begin{bmatrix} 4 & 6 \end{bmatrix}$ |
| b 1 Find $m$ and $n$ for $Y$. | $Y$ has order 2 x 4, so $m = 2$ and $n = 4$. |
| 2 What sort of summing matrix is needed? | We want to sum rows, so S is a 4 x 1 column summing matrix multiplied on the right. |
| 3 Multiply the matrices to find the sums of each of the rows, using CAS if necessary. | $\begin{bmatrix} 10 & 2 & 1 & 4 \\ 3 & 5 & 0 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} (10 \times 1) + (2 \times 1) + (1 \times 1) + (4 \times 1) \\ (3 \times 1) + (5 \times 1) + (0 \times 1) + (6 \times 1) \end{bmatrix}$ $= \begin{bmatrix} 17 \\ 14 \end{bmatrix}$ |
| 4 What do we need to multiply by to find the means of each of the rows? | $\frac{1}{4}$ |
| 5 Perform the scalar multiplication to find the means of each of the rows, using CAS if necessary. | $\frac{1}{4}\begin{bmatrix} 17 \\ 14 \end{bmatrix} = \begin{bmatrix} \frac{17}{4} \\ \frac{14}{4} \end{bmatrix} = \begin{bmatrix} 4.25 \\ 3.5 \end{bmatrix}$ |

## Multiplying identity and permutation matrices

As we have seen, the identity matrix $I$ is a square matrix with all 'Is on the leading diagonal and zeros elsewhere. Multiplying the identity matrix with another matrix leaves the other matrix unchanged. It acts like a T does when multiplying regular numbers. For example:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 3 \\ 5 & 7 \\ S & 0 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 5 & 7 \\ 9 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 15 & 10 & 13 \\ 12 & 18 & 11 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 15 & 10 & 13 \\ 12 & 18 & 11 \end{bmatrix}$$

The identity matrix is a special kind of permutation matrix, which are square matrices where every row and column has exactly one T, with zeros elsewhere.

Multiplying a permutation matrix with another matrix rearranges the rows or columns of the other matrix. The position of the 1 in the permutation matrix tells us where the row or column moves to. For example,

- the position of 1 0 0 ... tells where the first row or column in the other matrix will move to
- the position of 0 1 0 ... tells where the second row or column in the other matrix will move to
- the position of 0 0 1 ... tells where the third row or column in the other matrix will move to
- and so on.

A permutation matrix on the *left* rearranges the *rows* of another matrix.

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 3 \\ 5 & 7 \\ 9 & 0 \end{bmatrix} = \begin{bmatrix} 5 & 7 \\ 9 & 0 \\ 1 & 3 \end{bmatrix}$$

A permutation matrix on the *right* rearranges the *columns* of another matrix.

$$\begin{bmatrix} 15 & 10 & 13 \\ 12 & 18 & 11 \end{bmatrix}\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 13 & 15 & 10 \\ 11 & 12 & 18 \end{bmatrix}$$

WB
p. 138

**WORKED EXAMPLE 13** Multiplying permutation matrices

Find the following matrices.

| **Steps** | **Working** |
|---|---|
| a What is this matrix product equal to? | $\begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} T \\ H \\ I \\ N \\ G \end{bmatrix}$ |
| 1 Locate where the 1 appears in each row to identify where the letter has moved. | First row: the 1 is in the fourth position. The fourth letter is N.<br>Second row: the 1 is in the third position. The third letter is I.<br>Third row: the 1 is in the fifth position. The fifth letter is G.<br>Fourth row: the 1 is in the second position. The second letter is H.<br>Fifth row: the 1 is in the first position. The second letter is T. |
| 2 Complete the matrix product. | $\begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} T \\ H \\ I \\ N \\ G \end{bmatrix} = \begin{bmatrix} N \\ I \\ G \\ H \\ T \end{bmatrix}$ |


9780170448406

b Matrix P is a 4 x 4 permutation matrix and matrix $A$ is another matrix such that the matrix product $Ax\ P$ is defined. This matrix product results in the entire second and fourth columns of matrix $A$ being swapped. Find the permutation matrix P.

1 Locate where the 1 appears in each column.

First column: unchanged so the 1 is in the first position.
Second column: swapped so the 1 is in the fourth position.
Third column: unchanged so the 1 is in the third position.
Fourth column: swapped so the 1 is in the second position.

2 Write the matrix.

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

7.4

## EXERCISE 7.4 Matrix multiplication

ANSWERS p. 716

### Recap

1 ©VCAA 2012 1MQ1 95% $2 \times \begin{bmatrix} 2 & 8 \\ 4 & -1 \\ 3 & 5 \end{bmatrix} - \begin{bmatrix} 3 & 7 \\ 4 & 2 \\ 2 & 3 \end{bmatrix}$ equals

**A** $\begin{bmatrix} 1 & 1 \\ 0 & -3 \\ 4 & 2 \end{bmatrix}$ B $\begin{bmatrix} -2 & 2 \\ 0 & -6 \\ 2 & 4 \end{bmatrix}$ C $\begin{bmatrix} 1 & 9 \\ 12 & 0 \\ 8 & 13 \end{bmatrix}$ D $\begin{bmatrix} 1 & 9 \\ 4 & -4 \\ 4 & 7 \end{bmatrix}$ E $\begin{bmatrix} -1 & 1 \\ 0 & -3 \\ 1 & 2 \end{bmatrix}$

2 How many of these four matrix expressions are defined?

$\begin{bmatrix} -7 & 1 \\ -2 & 8 \end{bmatrix} - \begin{bmatrix} 3 & 10 \\ -9 & 0 \end{bmatrix}$ $7\begin{bmatrix} 5 \\ 2 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ $\begin{bmatrix} 1 \\ 0 \end{bmatrix} - 6\begin{bmatrix} 3 & 0 \\ 2 & 0 \end{bmatrix}$ $\begin{bmatrix} 4 & 9 \\ 1 & 0 \\ 3 & 2 \end{bmatrix} - \frac{1}{5}\begin{bmatrix} 8 & 5 \\ 2 & 3 \\ 1 & 7 \end{bmatrix}$

A 0 B 1 C 2 D 3 E 4

### Mastery

3 WORKED EXAMPLE 10 If $A = [2\ 1\ 5]$, $B = \begin{bmatrix} 3 \\ 4 \\ 10 \end{bmatrix}$, $C = \begin{bmatrix} 2 & -1 \\ 3 & 0 \end{bmatrix}$ and $D = \begin{bmatrix} 6 & 4 \\ 1 & 3 \\ 5 & 2 \end{bmatrix}$, for each of the following

a $AB$ b $BA$ c $BC$ d $BD$

e $DC$ f $C^2-2C$ g $D^4$

i state whether or not the expression is defined, giving a reason.

For those that are defined

ii state the order of the answer before performing the calculation

iii compute the matrix expressions to find the answer.

4 WORKED EXAMPLE 11

a Show that for the matrix equation $\begin{bmatrix} -1 & 2 \\ -2 & 3 \end{bmatrix}\begin{bmatrix} 2 \\ a \end{bmatrix} = \begin{bmatrix} -2 & 4 \\ 6 & -1 \end{bmatrix}\begin{bmatrix} b \\ 1 \end{bmatrix}$ to be true, the equations $a = 3 - b$ and $a = 1 + 2b$ must be true.

b Show that if the matrix $P^2$ is defined, matrix R is a row matrix, and matrix C is a column matrix, then $RP^2C$ is a $1 \times 1$ matrix.

5 ES Using CAS 3 J If M = $\begin{bmatrix} 3 & 7 \\ 9 & 4 \end{bmatrix}$ and $N = \begin{bmatrix} 2.5 & 6.1 \\ 3.9 & 2.7 \end{bmatrix}$, find the following.

a $MN$

b $M^3$

c $4M-NM^2$

6 WORKED EXAMPLE 12 J For the matrices $X = \begin{bmatrix} 9 & 20 \\ 3 & 9 \end{bmatrix}$ and $Y = \begin{bmatrix} 6 & 4 \\ 1 & 3 \\ 5 & 2 \end{bmatrix}$, show how to use a summing matrix S to calculate

a the sums and means of the rows of $X$

b the sums and means of the columns of $Y$.

7 WORKED EXAMPLE 13 Find the following matrices.

a What is this matrix product equal to? $\begin{bmatrix} 0\,0\,0\,1 \\ 0\,0\,1\,0 \\ 0\,1\,0\,0 \\ 1\,0\,0\,0 \end{bmatrix}\begin{bmatrix} R \\ A \\ T \\ S \end{bmatrix}$

b Matrix P is a 5 x 5 permutation matrix and matrix $Q$ is another matrix such that the matrix product $P$ x Q is defined. This matrix product results in the entire first and third rows of matrix $Q$ being swapped. Find the permutation matrix P.

## Exam practice

80-100% 60-79% 0-59%

8 ©VCAA 20151MQ2 ] 88% Four matrices are shown.

$W = \begin{bmatrix} 3 \\ 6 \\ 2 \end{bmatrix}$ $X = \begin{bmatrix} 4 & 1\,5 \\ 2 & 0\,6 \end{bmatrix}$ $Y = \begin{bmatrix} 7 & 1 \end{bmatrix}$ $Z = \begin{bmatrix} 8\,5 & 0 \\ 1\,9 & 3 \\ 4\,2 & 7 \end{bmatrix}$

Which one of the following matrix products is not defined?

A $WxY$ B $XxW$ C $YxX$ D $ZxW$ E $ZxY$

9 ©VCAA 20161MQ2 J 86% The matrix product $\begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0\,0\,1\,0\,0 \\ 1 & 0 & 0 & 0 & 0 \\ 0\,1\,0\,0\,0 \\ 0\,0\,0\,0\,1 \end{bmatrix} \times \begin{bmatrix} L \\ E \\ A \\ P \\ S \end{bmatrix}$ is equal to

A $\begin{bmatrix} L \\ A \\ P \\ S \\ E \end{bmatrix}$ B $\begin{bmatrix} L \\ E \\ A \\ P \\ S \end{bmatrix}$ C $\begin{bmatrix} p \\ L \\ E \\ A \\ S \end{bmatrix}$ D $\begin{bmatrix} p \\ A \\ L \\ E \\ S \end{bmatrix}$ E $\begin{bmatrix} p \\ E \\ A \\ L \\ S \end{bmatrix}$

10 ©VCAA 2006 1MQ3 I 84% Let $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 2\,1 \\ 1\,0 \end{bmatrix}$ and $C = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$.

Then $A^3(B - C)$ equals

A $\begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix}$ B $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ C $\begin{bmatrix} 3 & 6 \\ 6 & -3 \end{bmatrix}$ D $\begin{bmatrix} 3 & 0 \\ 0 & -3 \end{bmatrix}$ E $\begin{bmatrix} 5 & 10 \\ 10 & “5 \end{bmatrix}$



9780170448406

7.4

▶ **11** ©VCAA 20201MQ2 I 79% Matrix $A = \begin{bmatrix} 1 & 2 \\ 0 & 3 \\ 1 & 0 \\ 4 & 5 \end{bmatrix}$ and matrix $B = \begin{bmatrix} 2 & 0 & 3 & 1 \\ 4 & 5 & 2 & 0 \end{bmatrix}$.

Matrix Q = *Ax B*. The element in row *i* and column *j* of matrix Q is *q*-. Element $g_{41}$ is determined by the calculation

A 0x0+3x5 B lxl+2x0 C Ix2+2x4

D 4xl+5x0 E 4x2+5x4

12 ©VCAA 2019N1MQ6 Matrix W is a 3 x 2 matrix. Matrix Q is a matrix such that *Q x W = W.* Matrix Q could be

A I 1 I L J

B $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

C $\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$

D $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

E $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

13 ©VCAA | 20181MQ4 69% Matrix P is a 4 x 4 permutation matrix. Matrix IV is another matrix such that the matrix product *PxWis* defined. This matrix product results in the entire first and third rows of matrix *W* being swapped. The permutation matrix *P* is

A $\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

B $\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

C $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

D $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

E $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

14 ©VCAA 2018 1MQ2 68% The matrix product $\begin{bmatrix} 4 & 2 & 0 \end{bmatrix}$« $\begin{bmatrix} 4 \\ 12 \\ 8 \end{bmatrix}$ is equal to

A [144]

B $\begin{bmatrix} 16 \\ 24 \\ 0 \end{bmatrix}$

C $4 \times \begin{bmatrix} 1 & 2 & 0 \end{bmatrix} \times \begin{bmatrix} 1 \\ 12 \\ 8 \end{bmatrix}$

D $2 \times \begin{bmatrix} 2 & 1 & 0 \end{bmatrix} \times \begin{bmatrix} 2 \\ 6 \\ 4 \end{bmatrix}$

E $4 \times \begin{bmatrix} 2 & 1 & 0 \end{bmatrix} \times \begin{bmatrix} 2 \\ 6 \\ 4 \end{bmatrix}$

15 ©VCAA 2020 1MQ8 57% The table below shows information about three matrices: $A_y$ *B* and C.

| Matrix | Order |
|---|---|
| *A* | 2 × 4 |
| *B* | 2 × 3 |
| *C* | 3 × 4 |

The transpose of matrix A, for example, is written as $A^T$. What is the order of the product $C^T$ *x* $(A^T$ *x* $B)^T$*

A 2x3 B 3x4 C 4x2 D 4x3 E 4x4

16 ©VCAA 201 o 1MQ4 57% Which matrix expression results in a matrix that contains the sum of the numbers 2, 5,4,1 and 8?

A $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \times \begin{bmatrix} 2 & 5 & 4 & 1 & 8 \end{bmatrix}$

B $\begin{bmatrix} 2 & 5 & 4 & 1 & 8 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$

C $\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 8 \end{bmatrix}$

D $\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 2 \\ 5 \\ 4 \\ 1 \\ 8 \end{bmatrix}$

E $\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 8 \end{bmatrix}$

17 ©VCAA | 20121MQ91 39% Which set of equations below could be used to determine the values of *a* and *b* that are shown in the matrix equation shown?

$$\begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix} \times \begin{bmatrix} a \\ 3 \end{bmatrix} = \begin{bmatrix} 6 & 3 \\ 2 & -i \end{bmatrix} \times \begin{bmatrix} 2 \\ b \end{bmatrix}$$

A $a-b=2$, $a+b=0$

B $a+b=-2$, $a-b=0$

C $a+b=2$, $a-b=0$

D $a-b=8$, $a+b=2$

E $a-b=8$, $a+b=-2$

# Q-s) Inverse matrices

Video playlist Inverse matrices

## The inverse matrix

We have seen that multiplying the identity matrix $I$ with another matrix A, leaves A unchanged. It acts like a T when multiplying regular numbers.

So $AI = IA = A$ is like saying 5 x 1 = 1 x 5 = 5.

The inverse matrix A”l is the matrix which when multiplied by A results in the identity matrix $I$. It acts like the inverse of a regular number.

So $AA^{-1} = A^{-1}A = I$ is like saying $5 \times \frac{1}{5} = \frac{1}{5} \times 5 = 1$.

**The inverse matrix**

If the matrix $A^{-1}$ is the inverse of the matrix $A$ then

$$AA^{-1} = A^{-1}A = I$$

Only square matrices have inverses.

Not *all* square matrices have inverses.

$A$ and $A^{-1}$ always have the same order.

**WORKED EXAMPLE 14** Showing whether matrices are inverses of each other

Show that the matrices $\begin{bmatrix} 3 & 7 \\ 2 & 5 \end{bmatrix}$ and $\begin{bmatrix} 5 & -7 \\ -2 & 3 \end{bmatrix}$ are inverses of each other.

| Steps | Working |
|---|---|
| **1** Show that the product of the two matrices equals the identity matrix. | $\begin{bmatrix} 3 & 7 \\ 2 & 5 \end{bmatrix}\begin{bmatrix} 5 & -7 \\ -2 & 3 \end{bmatrix}$ $= \begin{bmatrix} 3\times 5+7\times(-2) & 3\times(-7)+7\times 3 \\ 2\times 5+5\times(-2) & 2\times(-7)+5\times 3 \end{bmatrix}$ $= \begin{bmatrix} 15-14 & -21+21 \\ 10-10 & -14+15 \end{bmatrix}$ $= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ |
| **2** Reverse the order of the two matrices and show that their product also equals the identity matrix. | $\begin{bmatrix} 5 & -7 \\ -2 & 3 \end{bmatrix}\begin{bmatrix} 3 & 7 \\ 2 & 5 \end{bmatrix}$ $= \begin{bmatrix} 5\times 3+(-7)\times 2 & 5\times 7+(-7)\times 5 \\ -2\times 3+3\times 2 & -2\times 7+3\times 5 \end{bmatrix}$ $= \begin{bmatrix} 15-14 & 35-35 \\ -6+6 & -14+15 \end{bmatrix}$ $= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ |

7.5

WB

p. 139

## Finding the determinant and the inverse of a matrix

To find the inverse of a matrix, we first need to find the determinant. We will start with 2x2 matrices. The calculations involved for finding determinants and inverses of higher order square matrices are more complex and require CAS.

**The determinant**

For a matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, the determinant is $\det(A) = ad - bc$ and

$$A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{\det(A)}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

The inverse does not exist when $\det(A) = 0$.

**WORKED EXAMPLE 15** Finding the determinant and inverse of a matrix

For each of the following matrices, find

**i** the determinant **ii** the inverse (if it exists).

**a** $A = \begin{bmatrix} 6 & 2 \\ 8 & 3 \end{bmatrix}$ **b** $B = \begin{bmatrix} 2 & 5 \\ 3 & 7 \end{bmatrix}$ **c** $C = \begin{bmatrix} 8 & 2 \\ 8 & 2 \end{bmatrix}$

| Steps | Working |
|---|---|
| **a i** For $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, use $\det(A) = ad - bc$. | $\det(A) = 6\times 3 - 2\times 8 = 18 - 16 = 2$ |
| **ii** Find $A^{-1} = \frac{1}{\det(A)}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$ (if it exists). | $A^{-1} = \frac{1}{2}\begin{bmatrix} 3 & -2 \\ -8 & 6 \end{bmatrix} = \begin{bmatrix} \frac{3}{2} & -1 \\ -4 & 3 \end{bmatrix}$ |

WB

p. 140

| | |
|---|---|
| b i For B = $\begin{vmatrix} a & b \\ e & d \end{vmatrix}$, use det (B) = *ad* - *be*. | det (B) = 2 x 7 - 5 x 3 = 14 - 15 = -1 |
| ii Find $B^{-1} = \frac{1}{\det(B)}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$ (if it exists). | $B^{-1} = -1\begin{bmatrix} 7 & -5 \\ -3 & 2 \end{bmatrix} = \begin{bmatrix} -7 & 5 \\ 3 & -2 \end{bmatrix}$ |
| c i For C = $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, use det (C) = *ad* - *be*. | det(C) = 8 x 2 - 2 x 8 = 16-16 = 0 |
| ii Find C“* = $\frac{1}{\det(C)}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$ (if it exists). | det (C) = 0, so C7¹ does not exist. |

WB p. 141

## WORKED EXAMPLE 16 Working with determinants and inverses of matrices

**a** If the determinant of $\begin{bmatrix} 5 & -3 \\ x & 3 \end{bmatrix}$ is equal to 21, what is the value of $x$?

**b** For the matrix $A = \begin{bmatrix} 2 & 3 \\ x & -4 \end{bmatrix}$, what is the value of $x$ for which $A = A^{-1}$?

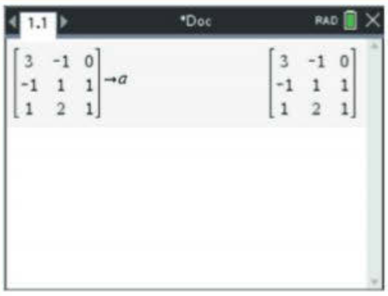
**c** Show that if $A$ and $B$ are both square matrices and $A + B$ is defined, then $A^{-1}B + A^4$ is defined.

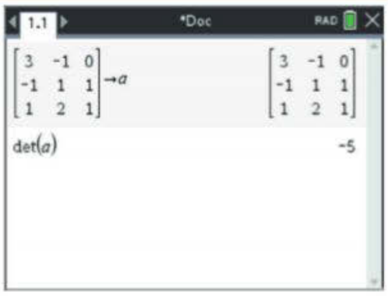
| Steps | Working |
|---|---|
| **a** **1** Find the determinant of the matrix. | $\det = 5 \times 3 - (-3) \times x = 15 + 3x$ |
| **2** Let the determinant equal the number given and solve for $x$, using CAS if necessary. | $15 + 3x = 21$<br>$3x = 6$<br>$x = 2$ |
| **b** **1** Find the determinant of $A$. | $\det(A) = 2 \times (-4) - 3 \times x = -8 - 3x$ |
| **2** Find $A^{-1}$. | $A^{-1} = \frac{1}{-8-3x}\begin{bmatrix} -4 & -3 \\ -x & 2 \end{bmatrix} = \begin{bmatrix} \frac{-4}{-8-3x} & \frac{-3}{-8-3x} \\ \frac{-x}{-8-3x} & \frac{2}{-8-3x} \end{bmatrix}$ |
| **3** Let $A = A^{-1}$.<br>Every pair of corresponding elements have to be equal, so set a pair of corresponding elements equal to each other.<br>Solve for $x$, using CAS if necessary. | $A = A^{-1}$<br>$\begin{bmatrix} 2 & 3 \\ x & -4 \end{bmatrix} = \begin{bmatrix} \frac{-4}{-8-3x} & \frac{-3}{-8-3x} \\ \frac{-x}{-8-3x} & \frac{2}{-8-3x} \end{bmatrix}$<br>So, $2 = \frac{-4}{-8-3x}$<br>$x = -2$ |
| **c** **1** State the orders of $A$ and $B$. | $A$ and $B$ are both square matrices, so let $A$ be of order $m \times m$ and $B$ be of order $n \times n$. |
| **2** Use the information given about $A$ and $B$. | $A + B$ is defined, so $A$ and $B$ must be of the same order, so $m = n$.<br>This means both $A$ and $B$ are of order $m \times m$. |
| **3** Use the orders of $A$ and $B$ to find the orders of the other matrices in the question. | Show that $A^{-1}B + A^4$ is defined.<br>If $A$ is $m \times m$, then $A^{-1}$ is $m \times m$ and $A^4$ is $m \times m$. |
| **4** Write a matrix order equation and answer the question. | The matrix order equation for $A^{-1}B + A^4$ is<br>$(m \times m) \times (m \times m) + (m \times m) = (m \times m) + (m \times m)$<br>The two matrices being added together are of the same order, so $A^{-1}B + A^4$ is defined. |

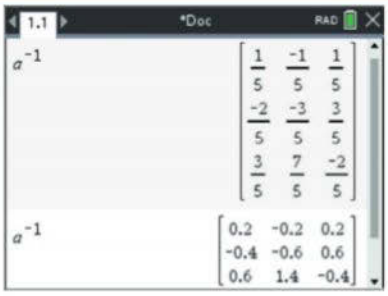
**USING CAS 4** Finding the determinant and inverse of a matrix

For $A = \begin{bmatrix} 3 & -1 & 0 \\ -1 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix}$, find **i** $\det(A)$ **ii** $A^{-1}$

**TI-Nspire**




1 In a **Calculator** page, create a **3×3** matrix and store it in as the variable ***a***.

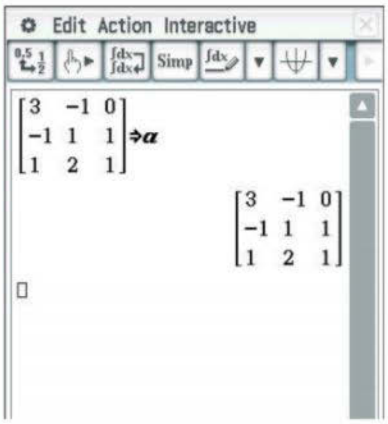
2 Press **menu > Matrix & Vector > Determinant**.

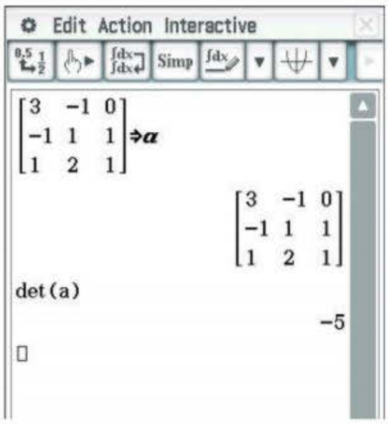
3 Enter ***a*** and press **enter** to find the determinant.

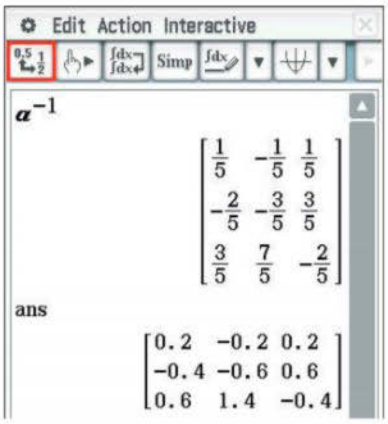
4 Find $\boldsymbol{a}^{-1}$ by pressing **A^(–1)** **enter** to calculate the inverse matrix.

5 Press **ctrl** + **enter** for the approximate answer.

**ClassPad**




1 In the **Main** application, create the **3×3** matrix and store it as the variable ***a***.

2 Tap **Interactive** > **Matrix** > **Calculation** > **det**.

3 In the dialogue box, enter ***a***.

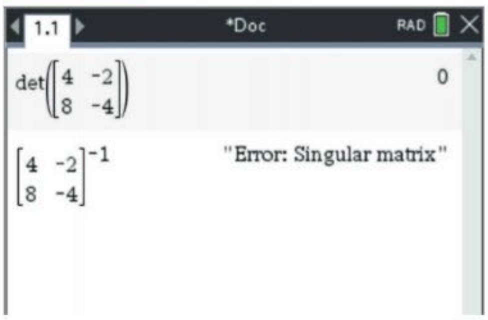
4 Tap **OK**.

5 Find $\boldsymbol{a}^{-1}$ by typing ***a*****^(–1)** to calculate the inverse matrix.

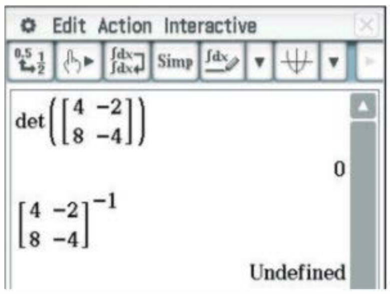
6 Tap the **Convert** tool to display the matrix in decimal form.

When the determinant of a matrix is equal to 0, it is known as a singular matrix. The inverse of a singular matrix does not exist. When attempting to find the inverse of a singular matrix, CAS will display an error message.

**TI-Nspire**




**ClassPad**



## EXERCISE 7.5 Inverse matrices

ANSWERS p. 716

### Recap

1 ©VCAA | 20131MQ2 80% Matrix $A$ has three rows and two columns. Matrix $B$ has four rows and three columns. Matrix C = $B$ x $A$ has

A two rows and three columns.
B three rows and two columns.
C three rows and three columns.
D four rows and two columns.
E four rows and three columns.

2 ©VCAA 2006 1MQ2 75% $\begin{bmatrix} -2 \\ 0 \end{bmatrix}$, $B$ = [ 0 9 j and C = [ 2 j. Using these matrices, the matrix product that is not defined is

A AB B AC C BA D $BC$ E $CB$

### Mastery

3 E2 WORKED EXAMPLE 14 J Show that the matrices $\begin{bmatrix} 3 & -4 \\ -2 & 3 \end{bmatrix}$ and $\begin{bmatrix} 3 & 4 \\ 2 & 3 \end{bmatrix}$ are inverses of each other.

4 S WORKED EXAMPLE 15 J For each of the following matrices find

i the determinant
ii the inverse (if it exists).

**a** $A = \begin{bmatrix} 2 & 8 \\ 2 & 7 \end{bmatrix}$ **b** $B = \begin{bmatrix} 5 & 10 \\ 1 & 1 \end{bmatrix}$ c $C = \begin{bmatrix} 5 & 2 \\ 10 & 4 \end{bmatrix}$

5 WORKED EXAMPLE 16 J

a If the determinant of $\begin{bmatrix} 6 & -12 \\ 3 & x \end{bmatrix}$ is equal to 12, what is the value of x?

b For the matrix A = $\begin{bmatrix} 5 & x \\ -4 & -5 \end{bmatrix}$, what is the value of x for which A = A *?

c Show that if A and B are both square matrices and A + B is defined, then $A^2 A$ *$B^3$ - B is defined.

6 0 using CAS 4~J For each of the following matrices, find

i det (A)
ii $A^{-1}$

a A = $\begin{bmatrix} 4 & 3 & 6 \\ 2 & -10 & 7 \\ -1 & -8 & 1 \end{bmatrix}$ b A = $\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix}$ c A = $\begin{bmatrix} 5 & 1 & 4 \\ 0 & -2 & 1 \\ -1 & -4 & 1 \end{bmatrix}$

### Exam practice

80–100% 60–79% 0–59%

7 | ©VCAA | 2008 1MQ5 J 80% The determinant of $\begin{bmatrix} 3 & 2 \\ 6 & x \end{bmatrix}$ is equal to 9. The value of x is

A -7 B -4.5 C 1 D 4.5 E 7

3 ©VCAA | 20181MQ1 78% Which one of the following matrices has a determinant of zero?

A $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ B $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ C $\begin{bmatrix} 1 & 2 \\ -3 & 6 \end{bmatrix}$ D $\begin{bmatrix} 3 & 6 \\ 2 & 4 \end{bmatrix}$ E $\begin{bmatrix} 4 & 0 \\ 0 & -2 \end{bmatrix}$


9780170448406

9 How many of the following matrices do not have an inverse?

$\begin{bmatrix} 2 \\ 3 \\ -6 \end{bmatrix}$ $\begin{bmatrix} 3 & 1 \\ -2 & 4 \end{bmatrix}$ $\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$ $\begin{bmatrix} 4 & 1 \\ 12 & 3 \end{bmatrix}$

A 0 B 1 C 2 D 3 E 4

10 ©VCAA 67% Matrix $A$ is a 1 × 3 matrix. Matrix $B$ is a 3 × 1 matrix. Which one of the following matrix expressions involving $A$ and $B$ is defined?

A $A + \frac{1}{3}B$ B $2Bx3A$ C $A^2B$ D $B'^1$ E $B-A$

11 How many of the following matrices have an inverse?

$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$ $\begin{bmatrix} 4 & 4 & 4 \\ 5 & 5 & 5 \\ 6 & 6 & 6 \end{bmatrix}$ $\begin{bmatrix} 20 & 2 \\ 30 & 3 \end{bmatrix}$ $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$

A 0 B 1 C 2 D 3 E 4

12 ©VCAA 20111MQ4 57% Matrix A is a 3 x 4 matrix. Matrix B is a 3 x 3 matrix. Which one of the following matrix expressions is defined?

A $BA^2$ B BA-2A C A + 2B D $B^2\sim AB$ E $A'^1$

## © Exam hack

Some multiple-choice questions require separate calculations for each of the options. Once you find the correct option, you don't need to work out the remaining options. However, to be safe, make a note to come back to check the remaining options if you have time.

13 ©VCAA 20111MQ8 49% Consider the matrix A = $\begin{bmatrix} 3 & k \\ -4 & -3 \end{bmatrix}$. A is equal to its inverse $A^{-1}$ for a particular value of $k$. This value of $k$ is

A -4 B -2 C 0 D 2 E 4

14 ©VCAA 2014 1MQ9 40% A and B are square matrices such that AB = BA = Z, where $I$ is an identity matrix. Which one of the following statements is not true?

A ABA = A

B $AB^2A = I$

C B must equal A

D B is the inverse of A

E both A and B have inverses

# 7.6 Matrix applications

Video playlist Matrix applications

WB p. 142

## Dealing with data in table form

Matrix multiplication can be used to do calculation with a large amount of data in table form.

**WORKED EXAMPLE 17** Solving sport problems using matrices

Each week, the coach of the Little Diggers basketball team awards the Best and Fairest Player Award to the player who scores the most game points. The results of their last game were as follows.

| | 3 pointers | 2 pointers | 1 pointer |
|---|---|---|---|
| Asma | 0 | 5 | 0 |
| Sarah | 1 | 3 | 2 |
| Lucy | 0 | 7 | 0 |
| Matt | 2 | 5 | 0 |
| Kelly | 1 | 7 | 0 |
| Min-Lee | 0 | 2 | 2 |
| Sophie | 1 | 8 | 0 |

**Steps** | **Working**

a Write a column matrix, P, to represent the three different scores possible.

When a player scores, they can either get 3 points, 2 points or 1 point.

$$p = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$$

b Write a matrix, G, to represent the data from the table.

Matrix $G$ will have 7 rows and 3 columns.

$$G - \begin{bmatrix} 0 & 5 & 0 \\ 1 & 3 & 2 \\ 0 & 7 & 0 \\ 2 & 5 & 0 \\ 1 & 7 & 0 \\ 0 & 2 & 2 \\ 1 & 8 & 0 \end{bmatrix}$$

c Use matrix multiplication to find a score matrix, S, that represents the total scored by each of the players.

To find S, we must find $GP$.

This will give us a (7 x 3)(3 x 1), which will result in a 7 x 1 matrix that represents the personal total for each of the seven players.

$$S = \begin{bmatrix} 0 & 5 & 0 \\ 1 & 3 & 2 \\ 0 & 7 & 0 \\ 2 & 5 & 0 \\ 1 & 7 & 0 \\ 0 & 2 & 2 \\ 1 & 8 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 10 \\ 11 \\ 14 \\ 16 \\ 17 \\ 6 \\ 19 \end{bmatrix}$$

d Which player won the Best and Fairest Award?

The largest element in S is the highest personal score. | Sophie won with 19 points.

e The opposition team scored a total of 87 points. Did the Little Diggers win the game?

1 Find the total score for Little Diggers by adding all of the elements in S. | Score = 10 + 11 + 14+16+17 + 6+19 = 93

2 Write the answer. | 93 is greater than 87, so Little Diggers won.



9780170448406

## WORKED EXAMPLE 18 Solving problems using summing matrices

Matrix $B$ shows the number of sightings of rare birds made by members of a birdwatching club.

| | Flitter | Bitwing | Redwing |
|---|---|---|---|
| Harry | 13 | 3 | 4 |
| Esther | 2 | 5 | 2 |
| Aniela | 1 | 1 | 3 |
| Gordon | 5 | 7 | 10 |

$$B = \begin{bmatrix} 13 & 3 & 4 \\ 2 & 5 & 2 \\ 1 & 1 & 3 \\ 5 & 7 & 10 \end{bmatrix} \quad P = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \quad Q = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

| Steps | Working |
|---|---|
| **a** i Calculate the product $A = PB$ and describe the information given by this matrix. <br> ii What information does the element $a_{13}$ give? | |
| i Multiply the matrices. The matrix order equation tells us the order of the product: <br> $(1 \times 4)(4 \times 3) = 1 \times 3$ | $A = PB = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 13 & 3 & 4 \\ 2 & 5 & 2 \\ 1 & 1 & 3 \\ 5 & 7 & 10 \end{bmatrix}$ <br> $= \begin{bmatrix} 21 & 16 & 19 \end{bmatrix}$ |
| The product involves summing the columns of $B$. | $A$ gives the totals for each type of bird sighted by the club members. |
| ii $a_{ij}$ is the element in the $i$th row and $j$th column in $A$. | $a_{13}$ tells us that a total of 19 Redwing birds were sighted by the club members. |
| **b** i Calculate $C = BQ$ and describe the information given by $C$. <br> ii What information does the element $c_{21}$ give? | |
| i Multiply the matrices. The matrix order equation tells us the order of the product: <br> $(4 \times 3)(3 \times 1) = 4 \times 1$ | $C = BQ = \begin{bmatrix} 13 & 3 & 4 \\ 2 & 5 & 2 \\ 1 & 1 & 3 \\ 5 & 7 & 10 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 20 \\ 9 \\ 5 \\ 22 \end{bmatrix}$ |
| The product involves summing the rows of $B$. | $C$ gives the total number of birds sighted by each birdwatcher. |
| ii $c_{ij}$ is the element in the $i$th row and $j$th column in $C$. | $c_{21}$ tells us that Esther sighted a total of 9 birds. |
| **c** Calculate $\frac{1}{4}PBQ$ and describe the information given by this matrix. | |
| Perform the multiplication. The matrix order equation tells us the order of the product: <br> $(1 \times 4)(4 \times 1) = 1 \times 1$ | $\frac{1}{4}PBQ = \frac{1}{4}PC = \frac{1}{4}\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 20 \\ 9 \\ 5 \\ 22 \end{bmatrix}$ <br> $= \frac{1}{4}[56]$ <br> $= [14]$ |
| The product involves adding elements and dividing by the number of elements. | $\frac{1}{4}PBQ$ tells us that the mean number of birds sighted by the birdwatchers is 14. |

### Exam hack

Even if your answer for a matrix multiplication is a single number, you must give your answer with matrix brackets around it, or it will not be marked as correct.

7.6

WB p. 143

## Costing and pricing matrices

Matrix multiplication can be used to solve costing or pricing problems involving different product categories.

WB
p. 144

**WORKED EXAMPLE 19** Solving costing and pricing problems using matrices

The manager of a local hardware store purchases small metal fasteners for \$3 each and large metal fasteners for \$5 each. In the last two weeks, he purchased the number of fasteners shown in the table.

| | Small metal fasteners | Large metal fasteners |
|---|---|---|
| Week 1 | 121 | 112 |
| Week 2 | 95 | 84 |

**Steps** **Working**

a Find the two matrices that can be multiplied to give the total purchase cost of metal fasteners in each of the two weeks and complete the multiplication.

We need a matrix product that calculates number of small metal fasteners x cost of small metal fasteners + number of large metal fasteners x cost of large metal fasteners.

$$\begin{matrix}\text{Weekl}\\ \text{Week 2}\end{matrix}\begin{bmatrix}121 & 112\\ 95 & 84\end{bmatrix}\begin{bmatrix}3\\ 5\end{bmatrix}=\begin{bmatrix}923\\ 705\end{bmatrix}$$

b The manager sells goods at a 55% mark-up. He recorded his purchase costs over the last two weeks for metal fasteners and three other items in the following table.

| | Weekl | Week 2 |
|---|---|---|
| Paint | \$1060 | \$1555 |
| Timber | \$3029 | \$1124 |
| Metal fasteners | | |
| Gardening tools | \$896 | \$2130 |

i Represent these costs in a 4 x 2 cost matrix, C.

ii Using scalar multiplication, represent the selling prices of these goods in a 4 x 2 matrix, S.

i The table already has 4 rows and 2 columns. Fill in the missing information from part a.

$$C=\begin{bmatrix}1060 & 1555\\ 3029 & 1124\\ 923 & 705\\ 896 & 1230\end{bmatrix}$$

ii 1 State what the mark-up means for the cost price in terms of the selling price.

A 55% mark-up means the selling price is 155% of the cost price.

2 Write an equation connecting the selling price and cost price by converting the percentage to a decimal.

selling price = 1.55 x cost price

3 Show this as a matrix equation and give the answer using scalar multiplication.

$$S=1.55C=1.55\begin{bmatrix}1060 & 1555\\ 3029 & 1124\\ 923 & 705\\ 896 & 1230\end{bmatrix}=\begin{bmatrix}1643.00 & 2410.25\\ 4694.95 & 1742.20\\ 1430.65 & 1092.75\\ 1388.80 & 1906.50\end{bmatrix}$$


9780170448406

c i Create a profit matrix.

ii Calculate the total profit to be made if all the goods purchased over these two weeks are sold.

i To create a profit matrix:

profit = selling price - cost price

profit = S-C

$$= \begin{bmatrix} 1643.00 & 2410.25 \\ 4694.95 & 1742.20 \\ 1430.65 & 1092.75 \\ 1388.80 & 1906.50 \end{bmatrix} - \begin{bmatrix} 1060 & 1555 \\ 3029 & 1124 \\ 923 & 705 \\ 896 & 1230 \end{bmatrix}$$

$$= \begin{bmatrix} 583.00 & 855.25 \\ 1665.95 & 618.20 \\ 507.65 & 387.75 \\ 492.80 & 676.50 \end{bmatrix}$$

ii The total profit can be found by adding all of the elements in the profit matrix.

total profit = 583.00 + 855.25 + 1665.95 + 618.20 + 507.65 + 387.75 + 492.80 + 676.50
= $5787.10

## EXERCISE 7.6 Matrix applications

ANSWERS p. 717

### Recap

1 Which of the following matrices has no inverse?

A $\begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$  B $\begin{bmatrix} 10 & 4 \\ 5 & 2 \end{bmatrix}$  C $\begin{bmatrix} 4 & 10 \\ 5 & 2 \end{bmatrix}$  D $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$  E $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$

2 If A is a 2 x 3 matrix, *B* is a 3 x 2 matrix and C is a 3 x 3 matrix, which one of the following matrices has no inverse?

*A BA*  B C  C $C^2$  D *AB*  E *CB*

### Mastery

**30** WORKED EXAMPLE 17 I Each week, the coach of the Little Clunkers cricket team awards the Best Batter to the player who makes the most runs. The results of their last game are shown in the table.

| | Single runs | Fours | Sixes |
|---|---|---|---|
| Heshan | 26 | 5 | 2 |
| Sam | 12 | 2 | 0 |
| Ahmat | 18 | 7 | 2 |
| Toni | 9 | 2 | 0 |
| Vishna | 30 | 3 | 0 |
| Jordan | 5 | 0 | 0 |
| Yasara | 3 | 0 | 0 |

a Write a column matrix, W, to represent the three different ways of scoring runs.

b Write a matrix, *R,* to represent the data from the table.

c Use matrix multiplication to find a score matrix, S, that represents the total runs made by each of the players.

d Which player won the Best Batter award?

e The opposition team made 207 runs. Did the Little Clunkers win the game?

4 WORKED EXAMPLE 18 Members of a trainspotting club recorded the number of different trains they saw over a weekend.

$$B = \begin{array}{l} \\ \text{Kristin} \\ \text{Amy} \\ \text{Seth} \\ \text{Steve} \\ \text{Judd} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Redrattler} & \text{Blueghost} & \text{Silverstreak} \end{array} \\ \begin{bmatrix} 6 & 7 & 0 \\ 2 & 6 & 3 \\ 7 & 10 & 1 \\ 6 & 9 & 2 \\ 2 & 5 & 4 \end{bmatrix} \end{array}$$

Let $P = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \end{bmatrix}$ and $Q = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

a i Calculate the product $A = PB$ and describe the information given by this matrix,

ii What information does the element $u_{12}$ give?

b i Calculate C = $BQ$ and describe the information is given by C.

ii What information does the element $c_{41}$ give?

c Calculate | $PBQ$ and describe the information given by this matrix.

5& WORKED EXAMPLE 19 A chemist purchases small bottles of vitamin C tablets for $2 each and large bottles of vitamin C tablets for $3 each. In the last two weeks, he purchased the number of vitamin C bottles shown in the table.

| | Small bottles of vitamin C | Large bottles of vitamin C |
|---|---|---|
| Week 1 | 75 | 60 |
| Week 2 | 47 | 82 |

a Find the two matrices that can be multiplied to give the total purchase cost of vitamin C bottles in each of the two weeks and complete the multiplication.

b The chemist sells goods at a 75% mark-up. He recorded his purchase costs over the last two weeks for vitamin C tablets and three other items in the following table.

| | Week 1 | Week 2 |
|---|---|---|
| Vitamin C bottles | | |
| Vitamin D bottles | $473 | $542 |
| Multivitamin bottles | $628 | $745 |
| Calcium bottles | $263 | $220 |

i Represent these costs in a 4 x 2 cost matrix, C.

ii Using scalar multiplication, represent the selling prices of these goods in a 4 x 2 matrix, S.

c i Create a profit matrix.

ii Calculate the total profit to be made if all goods purchased over these two weeks are sold.


9780170448406

6 ©VCAA 2017N 1MQ2 The cost of fruit at a stall, in dollars per kilogram, is shown in the table.

| | |
|---|---|
| Apples | $2.50 |
| Pears | $3.20 |
| Bananas | $1.90 |

Sean wants to buy 2 kg of apples, 1 kg of pears and 3 kg of bananas. Which one of the following matrix products will result in a matrix that contains the total cost of Sean's fruit purchase, in dollars?

**A** $\begin{bmatrix} 2 & 1 & 3 \end{bmatrix}\begin{bmatrix} 2.50 \\ 3.20 \\ 1.90 \end{bmatrix}$

**B** $\begin{bmatrix} 2 & 1 & 3 \end{bmatrix}\begin{bmatrix} 2.50 & 3.20 & 1.90 \end{bmatrix}$

**C** $\begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}\begin{bmatrix} 2.50 \\ 3.20 \\ 1.90 \end{bmatrix}$

**D** $\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}\begin{bmatrix} 2.50 \\ 3.20 \\ 1.90 \end{bmatrix}$

**E** $\begin{bmatrix} 2.50 \\ 3.20 \\ 1.90 \end{bmatrix}\begin{bmatrix} 2 & 1 & 3 \end{bmatrix}$

7 ©VCAA 2019 1MQ2 68% There are two rides called The Big Dipper and The Terror Train at a carnival. The cost, in dollars, for a child to ride on each ride is shown in the table.

| Ride | Cost ($) |
|---|---|
| The Big Dipper | 7 |
| The Terror Train | 8 |

Six children ride once only on The Big Dipper and once only on The Terror Train. The total cost of the rides, in dollars, for these six children can be determined by which one of the following calculations?

**A** $\begin{bmatrix} 6 \end{bmatrix} \times \begin{bmatrix} 7 & 8 \end{bmatrix}$

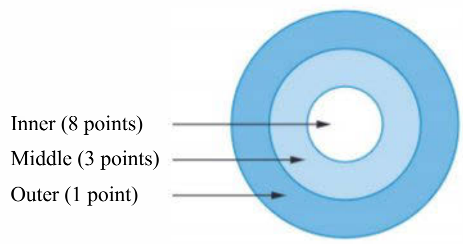
**B** $\begin{bmatrix} 6 \end{bmatrix} \times \begin{bmatrix} 7 \\ 8 \end{bmatrix}$

**C** $\begin{bmatrix} 6 & 6 \end{bmatrix} \times \begin{bmatrix} 7 & 8 \end{bmatrix}$

**D** $\begin{bmatrix} 6 & 6 \end{bmatrix} \times \begin{bmatrix} 7 \\ 8 \end{bmatrix}$

**E** $\begin{bmatrix} 6 \\ 6 \end{bmatrix} \times \begin{bmatrix} 7 & 8 \end{bmatrix}$

8 ©VCAA 2018N1MQ5 The target in a game is a circle divided into three regions: inner, middle and outer. The diagram shows these regions and the number of points awarded for hitting each region with a ball.




The number of inner, middle and outer regions hit by Mustafa and Neville in one game is shown in the table.

| Player | Inner | Middle | Outer |
|---|---|---|---|
| Mustafa | 6 | 7 | 2 |
| Neville | 5 | 9 | 7 |

Neville was the winner of this game. Which one of the following matrix calculations shows the difference between the winning and losing scores?

**A** $\begin{bmatrix} 8 \\ 3 \\ 1 \end{bmatrix} \times \left(\begin{bmatrix} 5 & 9 & 7 \end{bmatrix} - \begin{bmatrix} 6 & 7 & 2 \end{bmatrix}\right)$

**B** $\begin{bmatrix} 5 & 9 & 7 \end{bmatrix} - \begin{bmatrix} 6 & 7 & 2 \end{bmatrix} \times \begin{bmatrix} 8 \\ 3 \\ 1 \end{bmatrix}$

**C** $\left(\begin{bmatrix} 5 & 9 & 7 \end{bmatrix} - \begin{bmatrix} 6 & 7 & 2 \end{bmatrix}\right) \times \begin{bmatrix} 8 \\ 3 \\ 1 \end{bmatrix}$

**D** $\left(\begin{bmatrix} 5 \\ 9 \\ 7 \end{bmatrix} - \begin{bmatrix} 6 \\ 7 \\ 2 \end{bmatrix}\right) \times \begin{bmatrix} 8 & 3 & 1 \end{bmatrix}$

**E** $\begin{bmatrix} 8 \\ 3 \\ 1 \end{bmatrix} \times \begin{bmatrix} 5 & 9 & 7 \end{bmatrix} - \begin{bmatrix} 8 \\ 3 \\ 1 \end{bmatrix} \times \begin{bmatrix} 6 & 7 & 2 \end{bmatrix}$

9 ©VCAA 2013 1MQ7 43% A school has three computer classes, A, *B* and C. There are 15 students in each class. Each student is given a mark out of 100 based on their performance in a test. Matrix *M* below displays the marks obtained by these 45 students, listed by class.

$$M = \begin{bmatrix} 56 & 78 & 79 & 43 & 67 & 56 & 80 & 85 & 75 & 89 & 55 & 64 & 95 & 34 & 63 \\ 90 & 45 & 56 & 65 & 76 & 79 & 27 & 45 & 69 & 73 & 70 & 63 & 65 & 34 & 59 \\ 76 & 76 & 89 & 47 & 50 & 66 & 68 & 89 & 88 & 90 & 45 & 67 & 78 & 45 & 87 \end{bmatrix} \begin{matrix} A \\ B \text{ Class} \\ C \end{matrix}$$

Two other matrices, S and *R, are* defined as follows.

$$S = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \text{ and } R = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$$

Which one of the following matrix expressions can be used to generate a matrix that displays the mean mark obtained for each class?

A $\frac{1}{45}M$ B $\frac{1}{3}R \times M$ C $\frac{1}{3}R \times M \times S$ D $\frac{1}{15}M \times S$ E $\frac{1}{15}S \times R \times M$

10 ©VCAA 2016 1MQ4 34% The table shows the number of each type of coin saved in a moneybox.

| Coin | 5 cent | 10 cent | 20 cent | 50 cent |
|---|---|---|---|---|
| Number | 15 | 32 | 48 | 24 |

The matrix product that displays the total number of coins and the total value of these coins is

A $\begin{bmatrix} 5 & 10 & 20 & 50 \end{bmatrix} \begin{bmatrix} 15 \\ 32 \\ 48 \\ 24 \end{bmatrix}$

B $\begin{bmatrix} 15 & 32 & 48 & 24 \end{bmatrix} \begin{bmatrix} 1 & 5 \\ 1 & 10 \\ 1 & 20 \\ 1 & 50 \end{bmatrix}$

C $\begin{bmatrix} 5 & 10 & 20 & 50 \end{bmatrix} \begin{bmatrix} 1 & 15 \\ 1 & 32 \\ 1 & 48 \\ 1 & 24 \end{bmatrix}$

D $\begin{bmatrix} 15 & 32 & 48 & 24 \end{bmatrix} \begin{bmatrix} 5 \\ 10 \\ 20 \\ 50 \end{bmatrix}$

E $\begin{bmatrix} 5 & 10 & 20 & 50 \\ 15 & 32 & 48 & 24 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$


9780170448406

▶ 11 ©VCAA 2018 2MQI J (3 marks) A toll road is divided into three sections, E, *F* and *G*. The cost, in dollars, to drive one journey on each section is shown in matrix C.

$$C- \begin{bmatrix} 3.58 \\ 2.22 \\ 2.87 \end{bmatrix} \begin{matrix} E \\ F \\ G \end{matrix}$$

a 99% What is the cost of one journey on section G? 1 mark

b 94% Write down the order of matrix C. 1 mark

c 66% One day Kim travels once on section *E* and twice on section G. His total toll cost for this day can be found by the matrix product *M* x C. Write down the matrix *M*. 1 mark

12 ©VCAA 2Q15 2MQI 86% (5 marks) Students in a music school are classified according to three ability levels: beginner (B), intermediate (Z) or advanced (A). Matrix $S_o$ lists the number of students at each level in the school for a particular week.

$$S_0 = \begin{bmatrix} 20 \\ 60 \\ 40 \end{bmatrix} \begin{matrix} B \\ I \\ A \end{matrix}$$

a How many students in total are in the music school that week? 1 mark

b The music school has four teachers, David (D), Edith (E), Flavio (E) and Geoff (G). Each teacher will teach a proportion of the students from each level, as shown in matrix *P*.

$$\begin{matrix} & D & E & E & G \\ P - & [\,0.25 & 0.5 & 0.15 & 0.1\,] \end{matrix}$$

The matrix product, Q = $S_0$P, can be used to find the number of students from each level taught by each teacher.

i Copy and complete matrix Q by writing the missing elements in the boxes. 1 mark

$$Q = \begin{bmatrix} 5 \; 3 \; 2 \\ 15 \; 30 \;\; 6 \\ 10 \; 20 \; 6 \; 4 \end{bmatrix}$$

ii How many intermediate students does Edith teach? 1 mark

c The music school pays the teachers $15 per week for each beginner student, $25 per week for each intermediate student and $40 per week for each advanced student. These amounts are shown in matrix C.

$$\begin{matrix} & B\ I\ A \\ C- & [\ 15\ 25\ 40\ ] \end{matrix}$$

The amount paid to each teacher each week can be found using a matrix calculation.

i Write down a matrix calculation in terms of *Q* and C that results in a matrix that lists the amount paid to each teacher each week. 1 mark

ii How much is paid to Geoff each week? 1 mark ▶

▶ 13 ©VCAA 2016 2MQ1 , (3 marks) A travel company arranges flight (F), hotel (H), performance (P) and tour (T) bookings. Matrix C contains the number of each type of booking for a month.

$$C = \begin{bmatrix} 85 \\ 38 \\ 24 \\ 43 \end{bmatrix} \begin{matrix} F \\ H \\ P \\ T \end{matrix}$$

a 91 % Write down the order of matrix C. 1 mark

A booking fee, per person, is collected by the travel company for each type of booking. Matrix *G* contains the booking fees, in dollars, per booking.

$$\begin{matrix} F & H & P & T \end{matrix}$$
$$G - [\ 40\ 25\ 15\ 30\ ]$$

b i 79% Calculate the matrix product $J = G \times C$. 1 mark

ii 42% What does matrix *J* represent? 1 mark

14 ©VCAA | 2020 2MQ1J (6 marks) The three major shopping centres in a large city, Eastmall (E), Grandmall (G) and Westmall (W), are owned by the same company. The total number of shoppers at each of the centres at 1.00 pm on a typical day is shown in matrix *V*.

$$\begin{matrix} E & G & W \end{matrix}$$
$$V = [\ 2300\ 2700\ 2200\ ]$$

a 95% Write down the order of matrix *V*. 1 mark

Each of these centres has three major shopping areas: food (F), clothing (C) and merchandise (M). The proportion of shoppers in each of these three areas at 1.00 pm on a typical day is the same at all three centres and is given in matrix *P* below.

$$P = \begin{bmatrix} 0.48 \\ 0.27 \\ 0.25 \end{bmatrix} \begin{matrix} F \\ C \\ M \end{matrix}$$

b 45% Grandmalls management would like to see 700 shoppers in its merchandise area at 1.00 pm. If this were to happen, how many shoppers, in total, would be at Grandmall at this time? 1 mark

c The matrix $Q = P \times V$ is shown below. Two of the elements of this matrix are missing.

$$Q = \begin{matrix} & E & G & W & \\ & \left[\begin{matrix} 1104 \\ 621 \\ 575 \end{matrix}\right. & \begin{matrix} \_\_\_\_ \\ \_\_\_\_ \\ 675 \end{matrix} & \left.\begin{matrix} 1056 \\ 594 \\ 550 \end{matrix}\right] & \begin{matrix} F \\ C \\ M \end{matrix} \end{matrix}$$

i 86% Find the missing elements in the matrix Q. 1 mark

ii 66% The element in row *i* and column *j* of matrix *Q* is q^. What does the element $q_{23}$ represent? 1 mark

The average daily amount spent, in dollars, by each shopper in each of the three areas at Grandmall in 2019 is shown in matrix $A_{2019}$.

$$A_{2019} = \begin{bmatrix} 21.30 \\ 34.00 \\ 14.70 \end{bmatrix} \begin{matrix} F \\ C \\ M \end{matrix}$$

On one particular day, 135 shoppers spent the average daily amount on food, 143 shoppers spent the average daily amount on clothing and 131 shoppers spent the average daily amount on merchandise.

▶

d | 60% Write a matrix calculation, using matrix $A_{2019}$, showing that the total amount spent by all these shoppers is $9663.20. 1 mark

e 24% In 2020, the average daily amount spent by each shopper was expected to change by the percentage shown in the table below.

| Area | Food | Clothing | Merchandise |
|---|---|---|---|
| Expected change | increase by 5% | decrease by 15% | decrease by 1% |

The average daily amount, in dollars, expected to be spent in each area in 2020 can be determined by forming the matrix product $A_{2020} = K * A_{2019}$. Write down matrix $K$. 1 mark

15 ©VCAA | 2018 2MQ2 | (2 marks) The Westhorn Council must prepare roads for expected population changes in each of three locations: main town (M), villages (V) and rural areas *(R).* The population of each of these locations in 2018 is shown in matrix $P_{2018}$.

$$P_{2018} = \begin{bmatrix} 2100 \\ 1800 \\ 1700 \end{bmatrix} \begin{matrix} M \\ V \\ R \end{matrix}$$

The expected annual change in population in each location is shown in the table.

| Location | Main town | Villages | Rural areas |
|---|---|---|---|
| Annual change | increase by 4% | decrease by 1% | decrease by 2% |

a 62% Write down matrix $P_{2019}$, which shows the expected population in each location in 2019. 1 mark

b 40% The expected population in each of the three locations in 2019 can be determined from the matrix product $P_{2019} = F \times P_{2018}$, where F is a diagonal matrix. Write down matrix F. 1 mark

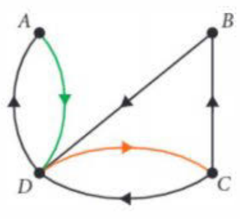
# @ Communication and dominance matrices

Video playlist Communication and dominance matrices

## Communication diagrams and matrices

Matrices are useful when investigating how people and systems communicate. Communication links can be shown either by a diagram or a matrix. Communication diagrams show the direct links with arrows. Communication matrices are square binary matrices that show the direct links with Ts. Links with the same sender and receiver are called redundant links. These do not count as communication, so the leading diagonals of communication matrices are always all '0 s.

Communication diagram




Communication matrix

Receiver

*A B C D*

| | | | |
|---|---|---|---|
| | *A* | OOOl' | A communicates with *D* |
| *M* - Sender £ | | 0 0 0 1 | |
| | C | 0 10 1 | |
| | *D* | 10 10 | *D* communicates with C |

Direct links shown by T s in a communication matrix are called one-step communications. Two-step communications are communication that occur via another link. For the matrix *M,* although *A* does not have a one-step communication with C, *A* does have a two-step communication with C: A —* D —> C.

## Exam hack

Communications can be one-way (A to B but not *B* to *A)* or two-way (A to B *and Bio A).* Don’t confuse this with one-step and two-step communications.

## Communication matrices

A communication matrix

- is a square binary matrix where T’ indicates direct one-step communication and ‘O’ indicates non-communication
- has a leading diagonal of all zeros indicating redundant links where the sender and receiver are the same
- is symmetric if every communication goes in both directions
- can be used to find two-step communications where communication occurs via another link.

WB
p. 146

### WORKED EXAMPLE 20 Working with communication matrices and diagrams

a The communication matrix *M* shows how direct messages can be sent between four people: Andrew (A), Bella (B), Corinna (C) and Davida (D).

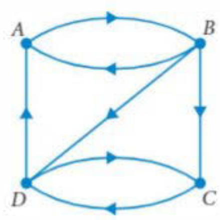
$$M = \text{Sender} \begin{matrix} \\ A \\ B \\ C \\ D \end{matrix} \overset{\text{Receiver}}{\begin{matrix} A & B & C & D \end{matrix}} \hspace{-3.5cm} \begin{matrix} \\ \end{matrix} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

i List who each person can send direct messages to.

ii Explain why the leading diagonal is all zeros.

iii Draw a communication diagram showing the communication links given in the matrix.

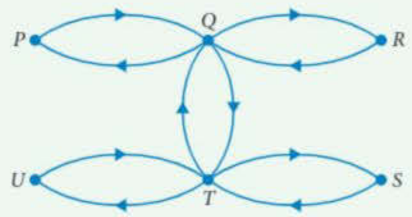
iv How could Corinna get a message to Andrew in two steps?

| Steps | Working |
|---|---|
| i Look at each row in order. T means the person can send a direct message. | Andrew can send direct messages to Bella. Bella can send direct messages to Andrew, Corinna and Davida. Corinna can send direct messages to Davida. Davida can send direct messages to Andrew and Corinna. |
| ii Refer to redundant links. | The leading diagonal represents links where the sender and receiver are the same. This is not considered communication, so they are redundant links. |
| iii Draw a diagram with arrows that match the list of possible direct messages. |  |
| iv Find how the message could be passed on in two steps. | Corinna -> Davida —> Andrew |




9780170448406

b i Construct the communication matrix for the following communication diagram.




ii Explain why the matrix is symmetric by referring to the communication diagram.

i Set up a 6 x 6 binary matrix where T indicates communication and 'O' indicates non-communication.

| | | Receiver |
|---|---|---|
| | | P Q R S T U |
| Sender | P | 0 1 0 0 0 0 |
| | Q | 10 10 10 |
| | R | 0 1 0 0 0 0 |
| | S | 0 0 0 0 1 0 |
| | T | 0 10 10 1 |
| | u | 0 0 0 0 1 0 |

ii Refer to the direction of the arrows in the communication diagram.

The matrix is symmetric because all the communications go both ways.

## Using two-step communication

Squaring a communication matrix gives us the number of two-step communications. The values that appear in the leading diagonal of the squared matrix give the number of two-step communications where the sender and receiver are the same. These are redundant links that should be ignored. Adding the communication matrix and its square gives us the total number of one-step and two-step communications.

| $M$ | $M^2$ | $M + M^2$ |
|---|---|---|
| Number of one-step communications | Number of two-step communications | Total number of one-step and two-step communications |
| Receiver<br>A B C D<br>Sender A: 0 0 0 1<br>B: 0 0 0 1<br>C: 0 10 1<br>D: 10 10 | Receiver<br>A B C D<br>Sender A: 10 10<br># C: 10 10<br>C: 10 11<br>D: 0 10 2 | Receiver<br>A B C D<br>Sender A: 10 11<br>B: 10 11<br>C: 1112<br>D: 1112 |
| No one-step communication from A to C. | 1 two-step communications from A to C. | Total of 1 one-step and two-step communications from A to C. |

This pattern extends to three-step and higher communications where Af³ gives the number of three-step communications and so on.

**Exam hack**

While $M^2$ gives the *number* of two-step communications, we need to look at $M$ to find what the two-step communications are.

**Using two-step communication**

For a communication matrix $M$:

- $M^2$ gives the number of two-step communications
- the values in the leading diagonal of $M^2$ are redundant two-step links
- $M + M^2$ gives the total number of one-step and two-step communications.

WB

p. 148

**WORKED EXAMPLE 21** Working with two-step communication

For the communication matrix representing the connections between four computers, find the following.

$$M = \text{From} \begin{matrix} & \text{To} \\ & \begin{matrix} A & B & C & D \end{matrix} \\ \begin{matrix} A \\ B \\ C \\ D \end{matrix} & \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \end{matrix}$$

**a** The number of ways $B$ can connect with $A$ by connecting directly to one other computer.

**b** The list of all the two-step connections from $B$ to $A$.

**c** The total number of redundant two-step connections.

**d** The list of redundant two-step connections from $C$ to $C$.

**e** The total number of one-step and two-step connections from $C$ to $D$.

| **Steps** | **Working** |
|---|---|
| **a** Find $M^2$ using CAS and read the number of two-step connections from the matrix. | $M^2 = \text{From} \begin{matrix} & \text{To} \\ & \begin{matrix} A & B & C & D \end{matrix} \\ \begin{matrix} A \\ B \\ C \\ D \end{matrix} & \begin{bmatrix} 1 & 1 & 0 & 1 \\ 2 & 1 & 0 & 1 \\ 1 & 0 & 2 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \end{matrix}$<br>There are 2 ways $B$ can connect with $A$ by connecting directly to one other computer. |
| **b** Use $M$ to find the two-step connections. | $B \rightarrow C \rightarrow A$ and $B \rightarrow D \rightarrow A$ |
| **c** Add the values in the leading diagonal of $M^2$. | The total number of redundant two-step connections is $1 + 1 + 2 + 0 = 4$ |
| **d** Use $M^2$ to find the number of redundant two-step connections.<br>Use $M$ to find the redundant two-step connections. | There are 2 redundant two-step connections from $C$ to $C$:<br>$C \rightarrow A \rightarrow C$ and $C \rightarrow B \rightarrow C$ |
| **e** Find $M + M^2$, using CAS if necessary, and read the number from the matrix. | $M + M^2 = \text{From} \begin{matrix} & \text{To} \\ & \begin{matrix} A & B & C & D \end{matrix} \\ \begin{matrix} A \\ B \\ C \\ D \end{matrix} & \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & 1 & 1 \\ 2 & 1 & 2 & 2 \\ 1 & 0 & 1 & 0 \end{bmatrix} \end{matrix}$<br>The total number of one-step and two-step connections from $C$ to $D$ is 2. |

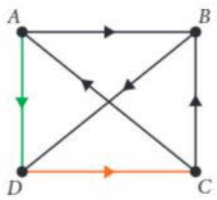


9780170448406

## Dominance diagrams and matrices

Dominance involves situations where there are clear winners and losers, for example, round-robin tournaments where everyone plays everyone else once and defeating someone is dominance. Dominance is similar to communication, except *all* the arrows in a dominance diagram are *one-way*. An arrow from *A* to *B* in a dominance diagram shows *A* has dominated *B*. A dominance matrix is a square binary matrix where dominance is shown by a ‘1’. Summing each row of a dominance matrix gives the one-step dominance scores or how many wins each participant has.

Dominance diagram




Dominance matrix

A dominates *D*

Loser

$$M = \text{Winner}\begin{matrix} A \\ B \\ C \\ D \end{matrix}\overset{\begin{matrix} A & B & C & D \end{matrix}}{\begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}\ \begin{matrix} \text{Scores} \\ 2 \\ 1 \\ 2 \\ 1 \end{matrix}$$

*D* dominates C

### Exam hack

As with communication matrices, the leading diagonals in dominance matrices are all ‘Os’. Also, dominance is one-way, so:

- if A to B is T, then B to A is ‘0’
- if A to B is ‘O’, then B to A is ‘ 1’.

## Using two-step dominance

The overall winner is not always clear from the one-step dominance scores. We also need to consider two-step dominance where *A* dominates C indirectly by dominating D, who dominates C. The dominance matrix *M* gives us the number of one-step dominances. $M^2$ gives us the number of two-step dominances. $M + M^2$ gives us the sum of one-step and two-step dominances. Summing each row of $M + M^2$ gives the total dominance scores, which can be used to rank all the participants and find an overall winner.

| $M$ | $M^2$ | $M + M^2$ |
|---|---|---|
| Number of one-step dominances | Number of two-step dominances | Total number of one-step and two-step dominances |
| Loser<br>$\text{Winner}\begin{matrix} \\ A \\ B \\ C \\ D \end{matrix}\begin{matrix} A\ B\ C\ D \\ \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\end{matrix}\begin{matrix} \text{Scores} \\ 2 \\ 1 \\ 2 \\ 1 \end{matrix}$ | Loser<br>$\text{Winner}\begin{matrix} \\ A \\ B \\ C \\ D \end{matrix}\begin{matrix} A\ B\ C\ D \\ \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 2 \\ 1 & 1 & 0 & 0 \end{bmatrix}\end{matrix}\begin{matrix} \text{Scores} \\ 2 \\ 1 \\ 3 \\ 2 \end{matrix}$ | Loser<br>$\text{Winner}\begin{matrix} \\ A \\ B \\ C \\ D \end{matrix}\begin{matrix} A\ B\ C\ D \\ \begin{bmatrix} 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 1 \\ 1 & 2 & 0 & 2 \\ 1 & 1 & 1 & 0 \end{bmatrix}\end{matrix}\begin{matrix} \text{Total scores} \\ 4 \\ 2 \\ 5 \\ 3 \end{matrix}\begin{matrix} \text{Rank} \\ 2 \\ 4 \\ 1 \\ 3 \end{matrix}$ |
| Two with equal highest score. No overall winner. | Unlike communication matrices, the leading diagonal has all ‘0’s. | *C* is overall winner and rankings are clear. |

## Dominance matrices

A dominance matrix $M$ is a square binary matrix where

- winners and losers such as in a round-robin tournament are indicated
- '1' indicates one-step dominance where $A$ dominates $B$
- the leading diagonal has all zeros
- if $A$ to $B$ is T then $B$ to $A$ is 'O', and if $A$ to $B$ is 'O' then $B$ to $A$ is '1'
- $M^2$ gives the number of two-step dominances where $A$ dominates $B$ and $B$ dominates C
- $M + M^2$ gives the number of one-step and two-step dominances
- the sum of each row in $M$ gives a one-step dominance score for each participant
- the sum of each row in $M + M^2$ gives a total dominance score for each participant which can be used to rank the participants and find the overall winner.

WB p. 149

### WORKED EXAMPLE 22 Using dominance matrices

For each of the following, find

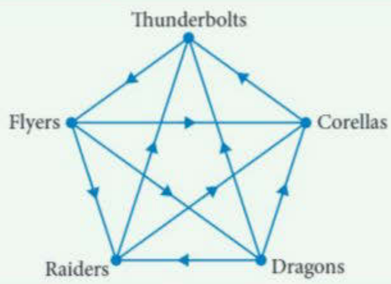
i the dominance matrix

ii the total dominance scores

iii the ranking of the participants and overall winner.

**Steps** / **Working**

**a** Five netball teams played off in a round-robin tournament. The results are shown in the following dominance diagram where an arrow indicates which team defeated the other.




**i** Use the direction of the arrows in the diagram to construct a square binary matrix $M$, where dominance is shown by a '1'.

$M$ = Winner (rows), Loser (columns)

| Winner \ Loser | $F$ | $T$ | $C$ | $D$ | $R$ |
|---|---|---|---|---|---|
| $F$ | 0 | 0 | 1 | 1 | 1 |
| $T$ | 1 | 0 | 0 | 0 | 0 |
| $C$ | 0 | 1 | 0 | 0 | 0 |
| $D$ | 0 | 1 | 1 | 0 | 1 |
| $R$ | 0 | 1 | 1 | 0 | 0 |

**ii** Find $M + M^2$ using CAS and add each row to find the total dominance scores.

$M + M^2$ = Winner (rows), Loser (columns)

| Winner \ Loser | $F$ | $T$ | $C$ | $D$ | $R$ | Total scores |
|---|---|---|---|---|---|---|
| $F$ | 0 | 3 | 3 | 1 | 2 | 9 |
| $T$ | 1 | 0 | 1 | 1 | 1 | 4 |
| $C$ | 1 | 1 | 0 | 0 | 0 | 2 |
| $D$ | 1 | 3 | 2 | 0 | 1 | 7 |
| $R$ | 1 | 2 | 1 | 0 | 0 | 4 |

**iii** Rank from 1 to 5. Ties are given equal ranking, and the ranking of positions below the ties is not affected by the ties.

Ranking: Flyers (1), Dragons (2), Thunderbolts (3), Raiders (3), Corellas (5)

The overall winners were the Flyers.


9780170448406

**7.7**

b Abbie, Bess, Cath and Denise played off against each other in a series of one-on-one basketball games. In these games

- Abbie beat Bess and Cath
- Bess beat Cath
- Cath beat Denise
- Denise beat Abbie and Bess.

i Construct a square binary matrix $M$, where dominance is shown by a '1'.

$$M = \text{Winner} \begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \text{Loser} \\ \begin{array}{cccc} A & B & C & D \end{array} \\ \left[\begin{array}{cccc} 0 & & 110 & \\ 0 & 0 & 10 & \\ 0 & 0 & 0 & 1 \\ 110\ 0 & & & \end{array}\right] \end{array}$$

ii Find $M + M^2$ using CAS and add each row to find the total dominance scores.

$$M+M^2\text{- Winner} \begin{array}{c} \\ A \\ £ \\ C \\ D \end{array} \begin{array}{c} \text{Loser} \\ \begin{array}{cccc} A & B & C & D \end{array} \\ \left[\begin{array}{cccc} 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 2 & 2 & 0 \end{array}\right] \end{array} \begin{array}{c} \text{Total scores} \\ 4 \\ 2 \\ 3 \\ 5 \end{array}$$

iii Rank from 1 to 4. Ties are given equal ranking, and the ranking of positions below the ties isn't affected by the ties.

Ranking: Denise (1), Abbie (2), Cath (3), Bess (4)

The overall winner was Denise.

## WORKED EXAMPLE 23 Working with dominance matrices

WB p. 151

The dominance matrix shows the result of each match between four teams, $A$, $B$, $C$ and $D$ in a round-robin tournament.

$$M = \text{Winner} \begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \text{Loser} \\ \begin{array}{cccc} A & B & C & D \end{array} \\ \left[\begin{array}{cccc} 0 & 0 & 1 & y \\ 1 & x & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & z & 0 \end{array}\right] \end{array}$$

| Steps | Working |
|---|---|
| **a** Complete the dominance matrix by including the values for $x$, $y$ and $z$. | |
| The leading diagonals in dominance matrices are all zeros.<br>Dominance is one-way, so:<br>• if A to B is '1', then B to A is '0'<br>• if A to B is '0', then B to A is '1'. | $x$ is in the leading diagonal, so $x = 0$.<br>$D$ to $A$ is 0, so $A$ to $D = y = 1$.<br>$C$ to $D$ is 1, so $D$ to $C = z = 0$.<br>$$M = \text{Winner} \begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \text{Loser} \\ \begin{array}{cccc} A & B & C & D \end{array} \\ \left[\begin{array}{cccc} 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{array}\right] \end{array}$$ |

b The result of one game not involving C in the tournament is in dispute. If the result of the game is reversed, team C would be declared the clear overall winner. Which game is it?

1 List the games that could be in dispute.

The games not involving C are: A vs B, *B* vs *D* and A vs *D*.

2 For each of the reversed result options:

- state the reversed result
- find the dominance matrices
- calculate the total dominance scores
- state the overall winner.

Option 1: Reversed result is A defeated *B*.

$$M = \text{Winner}\ \begin{matrix} A \\ B \\ C \\ D \end{matrix} \overset{\displaystyle\text{Loser}}{\overset{A\ B\ C\ D}{\begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}}} \qquad M + M^2 \sim \text{Winner}\ \begin{matrix} A \\ B \\ C \\ D \end{matrix} \overset{\displaystyle\text{Loser}}{\overset{A\ B\ C\ D}{\begin{bmatrix} 0 & 3 & 1 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}}} \ \overset{\text{Total scores}}{\begin{matrix} 6 \\ 0 \\ 3 \\ 1 \end{matrix}}$$

*A* is the overall winner.

Option 2: Reversed result is *B* defeated *D*.

$$M = \text{Winner}\ \begin{matrix} A \\ B \\ C \\ D \end{matrix} \overset{\displaystyle\text{Loser}}{\overset{A\ B\ C\ D}{\begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}}} \qquad M + M^2 - \text{Winner}\ \begin{matrix} A \\ B \\ C \\ D \end{matrix} \overset{\displaystyle\text{Loser}}{\overset{A\ B\ C\ D}{\begin{bmatrix} 0 & 1 & 1 & 2 \\ 1 & 0 & 1 & 2 \\ 1 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}}} \ \overset{\text{Total scores}}{\begin{matrix} 4 \\ 4 \\ 4 \\ 0 \end{matrix}}$$

A, *B* and C have the same rank, so C is not the clear overall winner.

Option 3: Reversed result is *D* defeated A.

$$M = \text{Winner}\ \begin{matrix} A \\ B \\ C \\ D \end{matrix} \overset{\displaystyle\text{Loser}}{\overset{A\ B\ C\ D}{\begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}}} \qquad M + M^2 = \text{Winner}\ \begin{matrix} A \\ B \\ C \\ D \end{matrix} \overset{\displaystyle\text{Loser}}{\overset{A\ B\ C\ D}{\begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 2 & 2 & 0 & 1 \\ 2 & 1 & 1 & 0 \end{bmatrix}}} \ \overset{\text{Total scores}}{\begin{matrix} 3 \\ 2 \\ 5 \\ 4 \end{matrix}}$$

*C* is the clear overall winner.

3 State the answer.

*C* would be declared the clear overall winner if the result of the game between A and *D* was reversed.

## VCE QUESTION ANALYSIS

©VCAA | 2019 2MQ1 J 2019 Examination 2 Matrices Question 1 (5 marks)

The car park at a theme park has three areas, A, *B* and C. The number of empty (£) and full (F) parking spaces in each of the three areas at 1 pm on Friday are shown in matrix *Q* below

$$Q = \begin{bmatrix} 70 & 50 \\ 30 & 20 \\ 40 & 40 \end{bmatrix} \begin{matrix} A \\ B \\ C \end{matrix} \text{ Area}$$

(columns: *E*, *F*)

a What is the order of matrix Q? 1 mark

b Write down a calculation to show that 110 parking spaces are full at 1 pm. 1 mark

Drivers must pay a parking fee for each hour of parking. Matrix P, below, shows the hourly fee, in dollars, for a car parked in each of the three areas.

Area

*A* *B* *C*

*P* - [ 1.30 3.50 1.80 ]

c The total parking fee, in dollars, collected from these 110 parked cars if they were parked for one hour is calculated as follows.

P x *L* = [207.00] where matrix *L* is a *3* x 1 matrix.

Write down matrix L. 1 mark

The number of whole hours that each of the 110 cars had been parked was recorded at 1 pm. Matrix P, below, shows the number of cars parked for one, two, three or four hours in each of the areas A, *B* and C.

Area

$$R = \begin{bmatrix} 3 & 1 & 1 \\ 6 & 10 & 3 \\ 22 & 7 & 10 \\ 19 & 2 & 26 \end{bmatrix} \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} \text{ Hours}$$

(columns: *A*, *B*, *c*)

d Matrix $R^T$ is the transpose of matrix *R*. Copy and complete the matrix $R^1$. 1 mark

$$R^T = \begin{bmatrix} \\ \\ \\ \end{bmatrix}$$

e Explain what the element in row 3, column 2 of matrix $R^T$ represents. 1 mark

### Reading the question

- You are being asked in part b to write a calculation, not an answer.
- Note when a question refers to an *hourly* fee and when it refers to a *total* fee.
- 'Explain' in part e means a worded answer is needed.

### Thinking about the question

- 'Showing a calculation means you need to include at least one of +, -, x or ÷.
- For part c, first think how you would do this calculation without matrices.
- The word 'represents' in part e is asking you to describe what the number is telling us about the real-life situation.

7.7

Video VCE question analysis: Matrices and their applications

**Worked solution** (/ = 1 mark)

a Matrix $Q$ has 3 rows and 2 columns, so it has order 3x2./

b Area $A$ has 50 full car parks, area $B$ has 20 full car parks, and area C has 40 full car parks.

So, a calculation showing that 110 parking spaces are full at 1 pm is 50 + 20 + 40 = 110 / or

$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 50 \\ 20 \\ 40 \end{bmatrix} = [110]$$

c The total parking fee calculation without using matrices is 1.3 x 50 + 3.5 x 20 + 1.8 x 40 = 207.

If $P$ x $L$ = [207.00] where matrix $L$ is a 3 x 1 matrix, then the matrix equation is

$$\begin{bmatrix} 1.30 & 3.50 & 1.80 \end{bmatrix}\begin{bmatrix} 50 \\ 20 \\ 40 \end{bmatrix} = [207.00]. \text{ So } L = \begin{bmatrix} 50 \\ 20 \\ 40 \end{bmatrix} ✓$$

d To find the transpose, swap the rows and columns of $R$.

$$R^T = \begin{bmatrix} 3 & 6 & 22 & 19 \\ 1 & 10 & 7 & 2 \\ 1 & 3 & 10 & 26 \end{bmatrix} ✓$$

e The element in row 3, column 2 of matrix $R^T$ is the same as the element in row 2, column 3 of matrix $R$.

The element represents the number of cars parked in area C for two hours. /

**Student performance** 80-100% 60-79% 0-59%

a 95%

b 77% Both the non-matrix and matrix calculation were accepted as correct.

c 64% This question was mostly well done.

d 76% Some responses had transcription errors.

e 53% Some students recognised area C and 2 hours but did not refer to 'the number of cars parked'.

## EXERCISE 7.7 Communication and dominance matrices

ANSWERS p. 717

### Recap

1 ©VCAA 20141MQ5 69% Students from Year 7 and Year 8 in a school sold trees to raise funds for a school trip. The number of large, medium and small trees that were sold by each year group is shown in the table below.

| Year group | Large | Medium | Small |
|---|---|---|---|
| 7 | 52 | 78 | 61 |
| 8 | 45 | 56 | 81 |

The large trees were sold for \$32 each, the medium trees were sold for \$26 each and the small trees were sold for \$18 each. A matrix product that can be used to calculate the amount, in dollars, raised by each year group by selling trees is

A $\begin{bmatrix} 52 & 78 & 61 \\ 32 & 26 & 18 \end{bmatrix}\begin{bmatrix} 45 \\ 56 \\ 81 \end{bmatrix}$

B $\begin{bmatrix} 7 & 52 & 78 & 61 \\ 8 & 45 & 56 & 81 \end{bmatrix}\begin{bmatrix} 32 \\ 26 \\ 18 \\ 0 \end{bmatrix}$

C $\begin{bmatrix} 32 & 26 & 18 \end{bmatrix}\begin{bmatrix} 52 & 45 \\ 78 & 56 \\ 61 & 81 \end{bmatrix}$

D $\begin{bmatrix} 32 & 26 & 18 \end{bmatrix}\begin{bmatrix} 52 & 78 & 61 \\ 45 & 56 & 81 \\ 0 & 0 & 0 \end{bmatrix}$

E $\left(\begin{bmatrix} 52 \\ 78 \\ 61 \end{bmatrix} + \begin{bmatrix} 45 \\ 56 \\ 81 \end{bmatrix}\right) \times \begin{bmatrix} 32 & 26 & 18 \end{bmatrix}$

▶ 2 © VCAA 2015 1MQ4 59% The numbers of adult and child tickets purchased for five performances of a stage show are shown in the table.

| Performance | Adult | Child |
|---|---|---|
| 1 | 142 | 24 |
| 2 | 128 | 31 |
| 3 | 89 | 24 |
| 4 | 104 | 18 |
| 5 | 115 | 23 |

Which one of the following matrix calculations can be used to determine both the total number of adult tickets and the total number of child tickets purchased for all five performances?

**A** $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 142 & 128 & 89 & 104 & 115 \\ 24 & 31 & 24 & 18 & 23 \end{bmatrix}$

**B** $\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 142 & 128 & 89 & 104 & 115 \\ 24 & 31 & 24 & 18 & 23 \end{bmatrix}$

**C** $\begin{bmatrix} 142 & 128 & 89 & 104 & 115 \\ 24 & 31 & 24 & 18 & 23 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$

**D** $\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 142 & 128 & 89 & 104 & 115 \\ 24 & 31 & 24 & 18 & 23 \end{bmatrix}$

**E** $\begin{bmatrix} 142 & 128 & 89 & 104 & 115 \\ 24 & 31 & 24 & 18 & 23 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$

## Mastery

3S WORKED EXAMPLE 20

a The communication matrix $M$ shows how direct messages can be sent between four people: Allie (A), Ben (B), Cassie (C) and Debra (D).

Receiver

$$M = \text{Sender} \begin{array}{c} A \\ B \\ C \\ D \end{array} \begin{bmatrix} 0 & 10 & 0 \\ 0 & 0 & 11 \\ 0 & 10 & 1 \\ 10\,10 & & \end{bmatrix}$$

(columns: $A \quad B \quad C \quad D$)

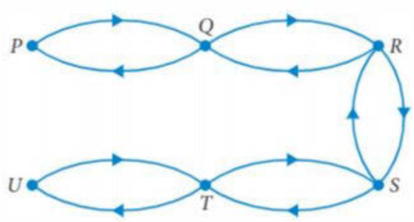
i List who each person can send direct messages to.

ii Explain why the leading diagonal is all zeros.

iii Draw a communication diagram showing the communication links given in the matrix.

iv How could Debra get a message to Ben in two steps?

b i Construct the communication matrix for the following communication diagram.




ii Explain why the matrix is symmetric by referring to the communication diagram. ▶

4E WORKED EXAMPLE 21J For the communication matrix representing the connections between four computers, find the following.

$$M = \text{From} \begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \overset{\displaystyle\text{To}}{\begin{array}{cccc} A & B & C & D \end{array}} $$

$$M = \text{From}\; \begin{matrix} A \\ B \\ C \\ D \end{matrix} \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

a The number of ways C can connect with A by connecting directly to one other computer,

b The list of all the two-step connections from $A$ to C.

c The total number of redundant two-step connections.

d The list of redundant two-step connections from $D$ to $D$.

e The total number of one-step and two-step connections from C to $A$.

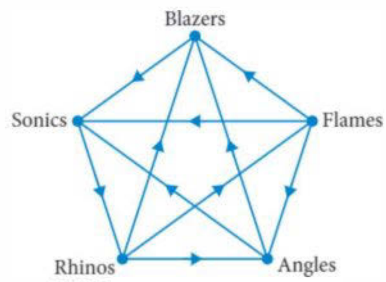
50 WORKED EXAMPLE 22 For each of the following, find

i the dominance matrix

ii the total dominance scores

iii the ranking of the participants and overall winner.

a Five softball teams played off in a round-robin tournament. The results are shown in the following dominance diagram where an arrow indicates which team defeated the other.




b Jackie, Kat, Lydia and Maisie played off against each other in a series of one-on-one basketball games. In these games

- Lydia beat Maisie
- Maisie beat Jackie
- Kat beat Lydia and Maisie
- Jackie beat Kat and Lydia.

6 S WORKED EXAMPLE 23 The dominance matrix shows the result of each match between four teams A, B, C and D in a round-robin tournament.

Loser (columns $A$, $B$, C, $D$)

$$M = \text{Winner}\; \begin{matrix} A \\ B \\ C \\ D \end{matrix} \begin{bmatrix} 0 & x & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ y & 1 & 1 & z \end{bmatrix}$$

a Complete the dominance matrix by including the values for x, $y$ and $z$.

b The result of one game not involving $A$ in the tournament is in dispute. If the result of the game is reversed, team $A$ would be declared the clear overall winner. Which game is it?


9780170448406

**Exam practice** 80-100% 60-79% 0-59%

7.7

7 ©VCAA 2017 1MQ2 I 90% The matrix shows how five people, Alan (A), Bevan (B), Charlie (C), Drew (D) and Esther (E), can communicate with each other.

| | Receiver A B C D E |
|---|---|
| *A* | 0 10 10 |
| *B* | 1 0 0 0 0 |
| Sender C | 0 0 0 1 1 |
| *D* | 10 10 0 |
| *E* | 0 0 10 0 |

A T in the matrix shows that the person named in that row can send a message directly to the person named in that column. For example, the T in row 3 and column 4 shows that Charlie can send a message directly to Drew. Esther wants to send a message to Bevan. Which one of the following shows the order of people through which the message is sent?

A Esther - Bevan

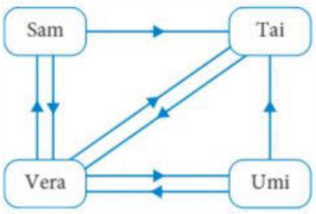
B Esther - Charlie - Bevan

C Esther - Charlie - Alan - Bevan

D Esther - Charlie - Drew - Bevan

E Esther - Charlie - Drew - Alan - Bevan

8 ©VCAA | 20201MQ5 86% The diagram below shows the direct communication links that exist between Sam (S), Tai (T), Umi (D) and Vera (V)- For example, the arrow from Umi to Vera indicates that Umi can communicate directly with Vera.




A communication matrix can be used to convey the same information. In this matrix:

- a T indicates that a direct communication link exists between a sender and a receiver
- a 'O' indicates that a direct communication link does not exist between a sender and a receiver.

The communication matrix could be

A

| | Receiver *S T U V* |
|---|---|
| s | 0 10 1 |
| Sender ? | 0 0 0 1 |
| *U* | 0 10 1 |
| *V* | 10 10 |

B

| | Receiver *S T U V* |
|---|---|
| S | 0 10 1 |
| Sender *T* | 10 0 1 |
| *U* | 0 10 1 |
| *V* | 1110 |

C

| | Receiver *S T U V* |
|---|---|
| S | 0 10 1 |
| Sender *T* | 0 0 0 1 |
| *U* | 0 10 0 |
| *V* | 1110 |

D

| | Receiver S *T u V* |
|---|---|
| S | 0 1 0 1 |
| Sender *T* | 0 0 0 1 |
| *U* | 0 1 0 1 |
| V | 1 1 1 0 |

E

| | Receiver *S T U V* |
|---|---|
| S | 0 10 2 |
| Sender *T* | 0 0 0 2 |
| *U* | 0 10 2 |
| *V* | 2 2 2 0 |

▶ 9 ©VCAA 2017 1MQ5 74% Four teams, *A*, *B*, *C* and *D*, competed in a round-robin competition where each team played each of the other teams once. There were no draws. The results are shown in the matrix.

| | | Loser | | | |
|---|---|---|---|---|---|
| | | *A* | *B* | C | *D* |
| Winner | *A* | 0 | 0 | *f* | 1 |
| | *B* | 1 | 0 | 0 | 0 |
| | *C* | 1 | *g* | 0 | 1 |
| | *D* | 0 | 1 | 0 | *h* |

A T in the matrix shows that the team named in that row defeated the team named in that column. For example, the T' in row 2 shows that team *B* defeated team *A*. In this matrix, the values of/, *g* and *h* are

A /=0,£=l,/i = 0 B /=0,£=l,h=l C /=l,£=0,h = 0

D /-1,^-l.h-O E *f*=*l*,*g*=*l*,*h*=*l*

10 ©VCAA | 20201MQ9 63% Five competitors, Andy (A), Brie (B), Cleo (C), Della (D) and Eddie (E), participate in a darts tournament. Each competitor plays each of the other competitors once only, and each match results in a winner and a loser. The matrix below shows the results of this darts tournament. There are still two matches that need to be played.

| | | Loser | | | | |
|---|---|---|---|---|---|---|
| | | *A* | *B* | C | *D* | E |
| | *A* | 0> ■ • | | 0 | 1 | 0 |
| | *B* | ... | 0 | 1 | 0 | 1 |
| Winner | C | 1 | 0 | 0 | ... | 1 |
| | *D* | 0 | 1 | ... | 0 | 0 |
| | *E* | 1 | 0 | 0 | 1 | 0 |

A T in the matrix shows that the competitor named in that row defeated the competitor named in that column.

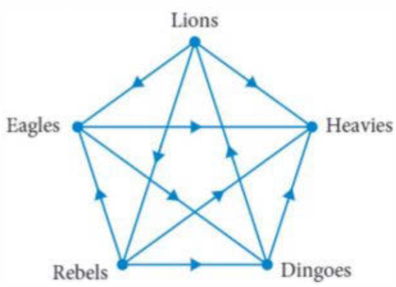
For example, the T in row 2, column 3 shows that Brie defeated Cleo.

A in the matrix shows that the competitor named in that row has not yet played the competitor named in that column.

The winner of this darts tournament is the competitor with the highest sum of their one-step and two-step dominances. Which player, by winning their remaining match, will ensure that they are ranked first by the sum of their one-step and two-step dominances?

A Andy B Brie C Cleo D Della E Eddie

11 ©VCAA 2015 1NQ8 59% There are five teams in a table tennis competition. Every team played one match against every other team, and each match had a winner and a loser. The results of the matches are summarised in the diagram shown. For example, an arrow from Lions to Eagles indicates that Lions defeated Eagles.




In determining the ranking of these teams, the total of each teams one-step dominances and two-step dominances will be calculated. The team with the highest total will be ranked first. The team with the next highest total will be ranked second, and so on. The ranking of these five teams from first to last is

A Lions, Rebels, Dingoes, Eagles, Heavies B Lions, Rebels, Eagles, Dingoes, Heavies

C Rebels, Lions, Dingoes, Eagles, Heavies D Rebels, Lions, Eagles, Dingoes, Heavies

E Eagles, Lions, Rebels, Dingoes, Heavies ▶

7.7

▶ 12 ©VCAA J 20191MQ7 54% The communication matrix below shows the direct paths by which messages can be sent between two people in a group of six people, 17 to Z.

Receiver

| Sender | u | V | W | X | Y | Z |
|---|---|---|---|---|---|---|
| u | 0 | 1 | 1 | 0 | 1 | 1 |
| V | 1 | 0 | 1 | 0 | 1 | 0 |
| W | 1 | 1 | 0 | 1 | 0 | 1 |
| X | 0 | 1 | 0 | 0 | 1 | 1 |
| y | 0 | 0 | 1 | 1 | 0 | 1 |
| Z | 1 | 1 | 0 | 1 | 1 | 0 |

A T in the matrix shows that the person named in that row can send a message directly to the person named in that column. For example, the T in row 4, column 2 shows that *X* can send a message directly to *V*. In how many ways can *Y* get a message to W by sending it directly to one other person?

AO BI C 2 D 3 E4

13 ©VCAA J 2016IMQ8 42% The matrix shows the result of each match between four teams, A, B, C and D, in a bowling tournament. Each team played each other team once and there were no draws.

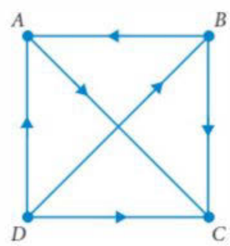
Loser

| Winner | *A* | *B* | *C* | *D* |
|---|---|---|---|---|
| *A* | 0 | 0 | 10 | |
| *B* | 10 | 0 | 1 | |
| *C* | 0 | 10 | 1 | |
| *D* | 10 0 0 | | | |

In this tournament, each team was given a ranking that was determined by calculating the sum of its one-step and two-step dominances. The team with the highest sum was ranked number one (1). The team with the second-highest sum was ranked number two (2), and so on. Using this method, team C was ranked number one (1). Team *A* would have been ranked number one (1) if the winner of one match had lost instead. That match was between teams

A A and B. B A and *D*. C B and C. D B and *D*. E C and *D*.

14 ©VCAA 2018N 1MQ4 The diagram shows the results of a chess competition between four players: Asha (A), Bai (B), Cam (C) and Drika (D).




Each competitor played each of the other competitors only once. The arrows in the diagram indicate the winner of each match. For example, the arrow from A to C shows that Asha defeated Cam. The two-step dominances in this competition can be shown as a matrix where the *winner* is the person who has two-step dominance over the *loser*. The matrix that shows the two-step dominances is

A Winner (Loser *A B C D*)

| | |
|---|---|
| A | 0 0 0 0 |
| *B* | 0 0 10 |
| C | 0 0 0 0 |
| *D* | 10 2 0 |

B Winner (Loser *A B C D*)

| | |
|---|---|
| A | 0 0 0 0 |
| *B* | 0 0 10 |
| C | 0 0 0 0 |
| *D* | 1110 |

C Winner (Loser *A B C D*)

| | |
|---|---|
| A | 0 0 10 |
| B | 10 10 |
| C | 0 0 0 0 |
| D | 1110 |

D Winner (Loser *A B C D*)

| | | | | |
|---|---|---|---|---|
| *A* | 0 | 0 | 10 | |
| # | 10 | 2 | 0 | |
| *C* | 0 | 0 | 0 | 0 |
| *D* | 2 13 0 | | | |

E Winner (Loser *A B C D*)

| | | | | |
|---|---|---|---|---|
| A | 0 | 0 | 10 | |
| *B* | 10 | 2 | 0 | |
| *C* | 0 | 0 | 0 | 0 |
| *D* | 2 12 0 | | | |

▶

15 ©VCAA | 2Q16 2MQ2H (2 marks) The travel company has five employees, Amara (A), Ben (B), Cheng (C), Dana (D) and Elka (E). The company allows each employee to send a direct message to another employee only as shown in the communication matrix $G$. The matrix $G^2$ is also shown below.

$$G = \text{Sender} \begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix} \overset{\text{Receiver} \quad A \; B \; C \; D \; E}{\begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}}$$

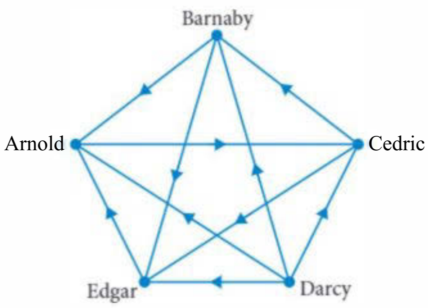
$$G^2 - \text{Sender} \begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix} \overset{\text{Receiver} \quad A\,B\,C\,D\,E}{\begin{bmatrix} 2\ 2\ 12\ 1 \\ 12\ 12\ 1 \\ 12\ 2\ 12 \\ 10\ 110 \\ 0\ 10\ 0\ 1 \end{bmatrix}}$$

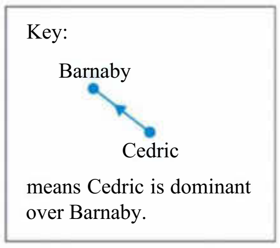
The T in row E, column $D$ of matrix $G$ indicates that Elka (sender) can send a direct message to Dana (receiver). The 'O' in row E, column C of matrix $G$ indicates that Elka cannot send a direct message to Cheng.

a 89% To whom can Dana send a direct message? 1 mark

b 71% Cheng needs to send a message to Elka, but cannot do this directly. Write down the names of the employees who can send the message from Cheng directly to Elka. 1 mark

16 ©VCAA | 2008 2NQ4 49% (5 marks) The children are taken to the zoo where they observe the behaviour of five young male lion cubs. The lion cubs are named Arnold, Barnaby, Cedric, Darcy and Edgar. A dominance hierarchy has emerged within this group of lion cubs. In the dominance diagram, the directions of the arrows show which lions are dominant over others.




a Name the two pairs of lion cubs who have equal totals of one-step dominances. 2 marks

b Over which lion does Cedric have both a one-step dominance and a two-step dominance? 1 mark

In determining the final order of dominance, the number of one-step dominances and two-step dominances are added together.

c Copy and complete the table for the final order of dominance. 1 mark

| Final order of dominance | Lion |
|---|---|
| 1st | Darcy |
| 2nd | |
| 3rd | |
| 4th | |
| 5th | |

Over time, the pattern of dominance changes until each lion cub has a one-step dominance over two other lion cubs.

d Determine the total number of two-step dominances for this group of five lion cubs. 1 mark

# (7) Chapter summary

## Matrix introduction

- Matrices are rectangular arrangements of numbers organised into rows and columns, usually presented in square brackets.
- Order of a matrix = *number of rows* x *number of columns*
- A matrix with *m* rows and *n* columns has order *m* x *n*.
- The leading diagonal is the diagonal running from the upper left to the lower right.
- The subdiagonal is the diagonal immediately under the leading diagonal.
- $a_{ij}$ is an element of matrix *A* where *i* is the row number and *j* is the column number.

For example, the elements of a 3 x 3 matrix *A* are written as [ «11 «12 % / «21 «22 «23 / «31 °32 «33 ].

## Types of matrices

| Type of matrix | Description | Examples | Order of examples |
|---|---|---|---|
| Binary matrix | A matrix where every element is either *0’ or T. | [ 1 1 0 / 101 ] [ 1 1 / 1 1 ] | 2x3 and 2x2 |
| Column matrix | A matrix with just one column. | [ 12 / -5 ] [ 20 / 35 / 63 ] | 2x1 and 3x1 |
| Diagonal matrix | A square matrix where all the elements except the ones in the leading diagonal are 0.’ | [ 9 0 0 0 / 0 -13 0 0 / 0 0 4 0 / 0 0 0 -1 ] [ -8 0 / 0 3 ] | 4x4 and 2x2 |
| Identity matrix (/) (also called the unit matrix) | A diagonal matrix where all the elements in the leading diagonal are T. We use *I* to indicate this matrix. | / - [ 10 0 0 / 0 10 0 / 0 0 10 / 0 0 0 1 ] I=r 1 0 / 0 1 ] | 4x4 and 2x2 |
| Lower triangular matrix | A square matrix where all the elements above the leading diagonal are ‘O’. | [ 2 0 0 0 / 7 6 0 0 / 0 4-30 / -6 1 15 8 ] [ 15 0 / 14 20 ] | 4x4 and 2x2 |
| Permutation matrix | A square matrix where every row and column has exactly one T» with zeros everywhere else. | [ 0 1 0 / 1 0 0 / 0 0 1 ] [ 1 0 / 0 1 ] | 3x3 and 2x2 |
| Row matrix | A matrix with just one row. | [ -2 3 11 ] [ 0 0 5 -3 1 | 1x3 and 1x4 |
| Square matrix | A matrix that has the same number of rows as columns. | [ 6 10 0 3 / -2 4 1-5 / 3 5 0 11 / -6 6 9 2 ] [ 5 16 / -3 0 ] | 4x4 and 2x2 |

| Type of matrix | Description | Examples | Order of examples |
|---|---|---|---|
| Summing matrix | A row or column matrix where all the elements are T. | [ 1 1 1 1 1 1 1 ]   [ 1 ; 1 ] | 1x7 and 2x1 |
| Symmetric matrix | A square matrix where the elements are symmetric around the leading diagonal. A symmetric matrix is the same as its transpose. | [ 4 6 -8 ; 6 1 2 ; -8 2 10 ]   [ -8 11 ; 11 3 ] | 3x3 and 2x2 |
| Transpose matrix | A matrix formed by switching the rows and columns of the original matrix. | $\begin{bmatrix} 3 & 1 & 4 \\ 8 & 2 & 7 \end{bmatrix}^T = \begin{bmatrix} 3 & 8 \\ 1 & 2 \\ 4 & 7 \end{bmatrix}$ | Original is 2 x 3. Transpose is 3 x 2. |
| Upper triangular matrix | A square matrix where all the elements below the leading diagonal are 'O'. | [ -3 8 2 ; 0 -4 7 ; 0 0 5 ]   [ 6 -4 ; 0 7 ] | 3x3 and 2x2 |
| Zero matrix | A matrix where all the elements are 'O'. | [ 0 0 ; 0 0 ; 0 0 ]   [ 0 0 0 0 0 0 ] | 3x2 and 1x6 |

## Matrix addition and subtraction

- When adding or subtracting matrices, add or subtract pairs of corresponding elements.
- For addition or subtraction of matrices to defined, they must have the same order.
- The answer has the same order as the matrices being added or subtracted.

## Scalar multiplication

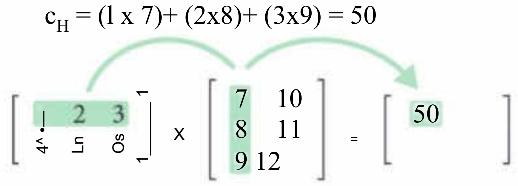
- A scalar is a regular number that is not in a matrix.
- When multiplying a matrix by a scalar, multiply each element by the scalar.
- Scalar multiplication is defined for any matrix.
- The answer has the same order as the original matrix.

## Matrix multiplication

When multiplying matrices $AB = C$

- multiply the elements of each row in $A$ by the elements of each column in B, and then add them:

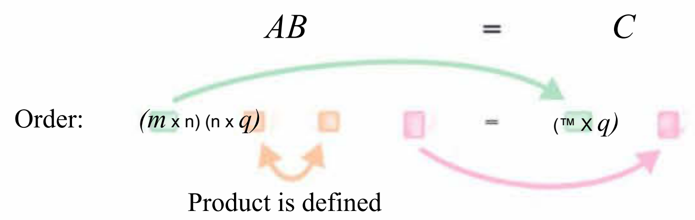
$c_H = (1 \times 7) + (2 \times 8) + (3 \times 9) = 50$




$A$ x $B$ = C

Order: 2x3   3x2   2x2

- the product is defined if the number of columns in $A$ = the number of rows in $B$. $C$ has the same number of rows as $A$ and the same number of columns as $B$.




When raising matrices to powers:

- only powers of square matrices are defined
- the power of a matrix will always have the same order as the original matrix.

## Multiplying summing matrices

If we multiply summing matrices (row or column matrices where all the elements areT') by another matrix, the result sums the rows or columns of the other matrix.

| | To sum each of the columns of an $m$ x $n$ matrix | To sum each of the rows of an $m$ x $n$ matrix |
|---|---|---|
| Type of summing matrix | row | column |
| Multiply on the | left | right |
| Order of summing matrix | 1 x $m$ | nx 1 |
| Order of multiplication | (1 x $m$)($m$ x $n$) | ($m$ x n)(n x 1) |
| Order of answer | lxn | $m$x 1 |
| To find means multiply by | $\frac{1}{m}$ | $\frac{1}{n}$ |

## Multiplying identity and permutation matrices

- Multiplying the identity matrix with another matrix, leaves the other matrix unchanged.
- The identity matrix is a special kind of permutation matrix.
- Multiplying a permutation matrix with another matrix rearranges the rows or columns of the other matrix.
- The position of the 1 in the permutation matrix tells us where the row or column of the other matrix moves to.
- A permutation matrix on the *left* rearranges the *rows* of another matrix.
- A permutation matrix on the *right* rearranges the *columns* of another matrix.

## The inverse matrix

- If matrix $A^{-1}$ is the inverse of the matrix A, then $AA^{-1} = A^{-1}A = I$.
- Only square matrices have inverses.
- Not *all* square matrices have inverses.
- A and $A^{-1}$ always have the same order.
- For a matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, the determinant is det (A) = $ad - be$.
- $A^{-1} = \frac{1}{ad - be}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{\det(A)}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$
- The inverse of A does not exist when det (A) = 0.
- Use CAS to find the determinant of higher order square matrices.
- When the determinant of a matrix is equal to 0, it is known as a singular matrix.

## Communication matrices

A communication matrix $M$ is a square binary matrix where

- T indicates direct one-step communication and 'O' indicates non-communication
- the leading diagonal has all zeros, indicating redundant links where the sender and receiver are the same
- $M^2$ gives the number of two-step communications where communication occurs via another link
- $M + M^2$ gives the total number of one-step and two-step communications.

## Dominance matrices

A dominance matrix $M$ is a square binary matrix where

- winners and losers such as in a round-robin tournament are indicated
- T indicates one-step dominance where $A$ dominates $B$
- the leading diagonal has all zeros
- if A to $B$ is T then $B$ to $A$ is 'O', and if $A$ to $B$ is 'O' then $B$ to $A$ is '1'
- $Al^2$ gives the number of two-step dominances where $A$ dominates $B$ and $B$ dominates C
- $M + M^2$ gives the number of one-step and two-step dominances
- the sum of each row in $M$ gives a one-step dominance score for each participant
- the sum of each row in $M + M^2$ gives a total dominance score for each participant which can be used to rank the participants and find the overall winner.



9780170448406

# Cumulative examination 1

Total number of marks: 19 Reading time: 8 minutes Writing time: 43 minutes

*Use the following information to answer the next four questions.*

The table shows the lean body mass (IBM), percentage body fat *(PBF)* and body mass index *(BMI)* of a sample of 12 professional athletes.

| LBM (kg) | PBF (%) | BMI (kg/m$^{2\ 2}$) |
|---|---|---|
| 63.3 | 19.8 | 20.6 |
| 58.6 | 21.3 | 20.7 |
| 55.4 | 19.9 | 21.9 |
| 57.2 | 23.7 | 21.9 |
| 53.2 | 17.6 | 19.0 |
| 53.8 | 15.6 | 21.0 |
| 60.2 | 20.0 | 21.7 |
| 48.3 | 22.4 | 20.6 |
| 54.6 | 18.0 | 22.6 |
| 53.4 | 15.1 | 19.4 |
| 61.9 | 18.1 | 21.2 |
| 48.3 | 23.3 | 22.0 |

1 ©VCAA 2018N1CQ7 J The mean, x, and the standard deviation, $s_x$, for the lean body mass *(LBM)* of these athletes, in kilograms, are closest to

A $\bar{x} = 48.3$ $s_x = 4.6$ B $\bar{x} = 55.0$ $s_x = 4.6$ C $\bar{x} = 55.0$ $s_x = 4.8$

D $\bar{x} = 55.7$ $s_x = 4.6$ E $\bar{x} = 55.7$ $s_x = 4.8$

2 ©VCAA | 2018N1CQ8 J A least squares line is fitted to the data using percentage body fat *(PBF)* as the response variable and body mass index *(BMI)* as the explanatory variable. The equation of the least squares line is closest to

A *PBF* = -4.7+ 1.2 *xBMI* B *BMI* = -4.7+ 1.2 x *PBF* C *PBF* = 17.8 + 1.7 x *BMI*

D *BMI*= 17.8+ 1.7 x *PBF* E *PBF* = 23.6- 0.1 x *BMI*

3 ©VCAA 201 SN 1 co9 I The Pearson correlation coefficient, r, between lean body mass *(LBM)* and percentage body fat *(PBF)* is closest to

A -0.235 B -0.124 C 0.124 D 0.235 E 0.352

4 ©VCAA 2018N1 CQI o J A least squares line is fitted to the data using lean body mass *(LBM)* as the response variable and body mass index *(BMI)* as the explanatory variable. The equation of this line is

LBM = 48.9 +0.320 xBMf

When this line is used to predict the lean body mass *(LBM)* of an athlete with a body mass index *(BMI)* of 22.0, the residual will be closest to

A -7.6 kg B -1.5kg C 1.5 kg D 33.9 kg E 55.9 kg

*Use the following information to answer the next three questions.*

A reducing balance loan has the following amortisation table.

| Payment number | Payment | Interest | Principal reduction | Balance |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 12000.00 |
| 1 | 3300.00 | 480.00 | 2820.00 | 9180.00 |
| 2 | 3300.00 | 367.20 | 2932.80 | 6247.20 |
| 3 | 3300.00 | 249.89 | 3050.11 | 3197.09 |
| 4 | 3300.00 | 127.88 | 3172.12 | 24.97 |
| Total | 13200.00 | 1224.97 | 11975.03 | |

5 The principal is

A \$3300.00 B \$11975.03 C \$12000.00

D \$13200.00 E Not shown in the table

6 Which one of the following is not true?

A The regular payments are \$3300.00.

B The interest paid reduces with every payment.

C The principal increases with every payment.

D The balance reduces with every payment.

E The balance after the first payment is \$9180.00.

7 After the first three payments have been made, if the loan is to be paid out fully in four payments, what would the last payment be?

A \$24.97 B \$2000.00 C \$3300.00 D \$3324.97 E \$13200.00

8 ©VCAA 2011 1MQ2 If $A = \begin{bmatrix} 0\,1 \\ 1\,0 \end{bmatrix}$, $B = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $C = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, then $AB + 2C$ equals

A $\begin{bmatrix} 0 \\ 3 \end{bmatrix}$ B $\begin{bmatrix} 3 \\ 0 \end{bmatrix}$ C $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ D $\begin{bmatrix} 2 \\ 0 \end{bmatrix}$ E $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$

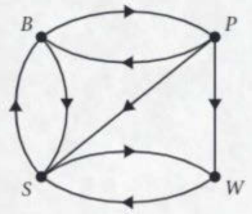
9 ©VCAA 2021 1MQ4 Ramon and Norma are names that contain the same letters but in a different order.

The permutation matrix that can change $\begin{bmatrix} R \\ A \\ M \\ O \\ N \end{bmatrix}$ into $\begin{bmatrix} N \\ 0 \\ R \\ M \\ \text{A} \end{bmatrix}$ is

A $\begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$ B $\begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$ C $\begin{bmatrix} 1\,0\,0\,0\,0 \\ 0\,0\,0\,1\,0 \\ 0\,0\,0\,0\,1 \\ 0\,0\,1\,0\,0 \\ 0\,1\,0\,0\,0 \end{bmatrix}$

D $\begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$ E $\begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$

10 ©VCAA I 2017N1MQ3 Peter (P), Sally (S) and Ben (B) are managers in a business. The diagram shows the direct communication that is possible between these managers and a worker, Whitney (IV). For example, the arrow from S to W indicates that Sally is able to communicate directly with Whitney.




A communication matrix can represent the direct communication that is possible between the managers and Whitney. The elements in the matrix are such that:

- T' indicates that direct communication from one person to another is possible
- 'O' indicates that direct communication is not possible.

This communication matrix could be

**A** From $\begin{array}{c} \\ p \\ \$ \\ B \\ W \end{array}\begin{array}{c} \text{To} \\ P\ S\ B\ W \\ \left[\begin{array}{c} 0\ 111 \\ 0\ 0\ 11 \\ 110\ 0 \\ 0100 \end{array}\right]\end{array}$

B From $\begin{array}{c} \\ P \\ S \\ B \\ W \end{array}\begin{array}{c} \text{To} \\ P\ S\ B\ W \\ \left[\begin{array}{c} 0\ 12\ 1 \\ 10\ 2\ 2 \\ 2\ 2\ 0\ 1 \\ 1210 \end{array}\right]\end{array}$

C From $\begin{array}{c} \\ p \\ S \\ B \\ W \end{array}\begin{array}{c} \text{To} \\ P\ S\ B\ W \\ \left[\begin{array}{c} 0\ 111 \\ 0\ 0\ 0\ 1 \\ 110\ 0 \\ 0\ 10\ 0 \end{array}\right]\end{array}$

**D** From $\begin{array}{c} \\ P \\ S \\ B \\ W \end{array}\begin{array}{c} \text{To} \\ P\ S\ B\ W \\ \left[\begin{array}{c} 0\ 111 \\ 10\ 11 \\ 110\ 1 \\ 1110 \end{array}\right]\end{array}$

E From $\begin{array}{c} \\ P \\ S \\ B \\ W \end{array}\begin{array}{c} \text{To} \\ P\ S\ B\ W \\ \left[\begin{array}{c} 0\ 0\ 10 \\ 10\ 11 \\ 110\ 0 \\ 110\ 0 \end{array}\right]\end{array}$

**11** ©VCAA | 2006IMQ5 I A company makes Regular (R), Queen (Q) and King (K) size beds. Each bed comes in either the Classic style or the more expensive Deluxe style. The price of each style of bed, in dollars, is listed in a price matrix P, where

$$P = \begin{array}{c} R\ Q\ K \\ \left[\begin{array}{c} 145\ 210\ 350 \\ 185\ 270\ 410 \end{array}\right] \end{array} \begin{array}{l} \text{Classic} \\ \text{Deluxe} \end{array}$$

The company wants to increase the price of all beds. A new price matrix, listing the increased prices of the beds, can be generated from P by forming a matrix product with the matrix, Al, where

$$Al = \left[\begin{array}{cc} 1.2\ 0 & \\ 0 & 1.35 \end{array}\right]$$

This new price matrix is

**A** $\left[\begin{array}{c} 145\ 210\ 350 \\ 185\ 270\ 410 \end{array}\right]$

B $\left[\begin{array}{ccc} 234.90 & 340.20 & 567 \\ 299.70 & 437.40 & 664.20 \end{array}\right]$

**C** $\left[\begin{array}{c} 174\ 252\ 420 \\ 222\ 324\ 492 \end{array}\right]$

**D** $\left[\begin{array}{ccc} 174 & 252 & 420 \\ 249.75 & 364.50 & 553.50 \end{array}\right]$

E $\left[\begin{array}{ccc} 195.75 & 283.50 & 472.50 \\ 249.75 & 364.50 & 553.50 \end{array}\right]$

12 ©VCAA 20181MQ3 Five people, India (/), Jackson (J), Krishna (K), Leanne (L) and Mustafa (M), competed in a table tennis tournament. Each competitor played every other competitor once only. Each match resulted in a winner and a loser. The matrix shows the tournament results.

$$\text{Winner}\begin{matrix} I \\ J \\ K \\ L \\ M \end{matrix}\overset{\text{Loser}}{\overset{\begin{matrix} I & J & K & L & M \end{matrix}}{\begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}}}$$

A T in the matrix shows that the competitor named in that row defeated the competitor named in that column. For example, the T' in the fourth row shows that Leanne defeated Jackson. A 'O' in the matrix shows that the competitor named in that row lost to the competitor named in that column. There is an error in the matrix. The winner of one of the matches has been incorrectly recorded as a 'O'. This match was between

A India and Mustafa.
B India and Krishna.
C Krishna and Leanne.
D Leanne and Mustafa.
E Jackson and Mustafa.

13 laWjJ 2020 1MQ3 ] Matrices $P$ and W are defined as $P = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$ $W = \begin{bmatrix} A \\ S \\ T \\ 0 \\ R \end{bmatrix}$

IfP"x W- $\begin{bmatrix} A \\ S \\ T \\ 0 \\ R \end{bmatrix}$, the value of $n$ could be

A 1
B 2
C 3
D 4
E 5

14 ©VCAA 2010 1MQ8 $m$ and $n$ are positive whole numbers. Matrix $P$ is of order $m \times n$. Matrix Q is of order $n \times m$. The matrix products $PQ$ and $QP$ are both defined

A for no values of $m$ and $n$.
B when $rn$ is equal to $n$ only.
C when $m$ is greater than $n$ only.
D when $m$ is less than $n$ only.
E for all values of $m$ and $n$.

15 ©VCAA 2021 1MQ5 A is a 7 x 7 matrix. $B$ is a 10 x 7 matrix. Which one of the following matrix expressions is defined?

A $AB-2B$
B $A(BA)^{-1}$
C $AB^2$
D $A^2-BA$
E $A(B^T)$

## Exam hack

A good strategy is to leave multiple choice questions where you need to work through each option for your 1 mark until the end.

16 ©VCAA 20071MQ9 Matrix $M$ is a 3 x 4 matrix. Matrix $P$ has five rows. $N$ is another matrix. If the matrix product $M(NP) = \begin{bmatrix} 4 & 1 & 7 & 2 \\ 0 & 9 & 7 & 4 \\ 4 & 3 & 3 & 1 \end{bmatrix}$, then the order of matrix $N$ is

A 3x5
B 5x3
C 4x5
D 5x4
E 5x5


9780170448406

17 ©VCAA 20141MQ6 I The order of matrix $X$ is 3 x 2. The element in row $i$ and column $j$ of matrix $X$ is $x_{ij}$ and it is determined by the rule $X_{ij} = i + j$. The matrix $X$ is

A $\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$ B $\begin{bmatrix} 2 & 3 \\ 4 & 5 \\ 6 & 7 \end{bmatrix}$ C $\begin{bmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}$ D $\begin{bmatrix} 1 & 2 \\ 3 & 3 \\ 4 & 4 \end{bmatrix}$ E $\begin{bmatrix} 2 & 3 \\ 3 & 4 \\ 4 & 5 \end{bmatrix}$

18 The inverse for the matrix $\begin{bmatrix} 1 & x \\ -1 & 2 \end{bmatrix}$ is

A $\begin{bmatrix} 2 & -x \\ 1 & 1 \end{bmatrix}$ B $\frac{1}{2 + x}\begin{bmatrix} 2 & -x \\ 1 & 1 \end{bmatrix}$ C $\frac{1}{2 - x}\begin{bmatrix} 2 & -x \\ 1 & 1 \end{bmatrix}$

D $\frac{1}{2 + x}\begin{bmatrix} 2 & x \\ -1 & 1 \end{bmatrix}$ E $\frac{1}{2 - x}\begin{bmatrix} 2 & x \\ 1 & 1 \end{bmatrix}$

19 ©VCAA 2019N1MQ2 I Four teams, blue (B), green (G), orange (O) and pink (P), played each other once in a competition. There were no draws in this competition. The results of the competition are shown in the matrix below.

| | | Loser | | | |
|---|---|---|---|---|---|
| | | *B* | *G* | *O* | *P* |
| Winner | *B* | – | 1 | v | 1 |
| | *G* | 0 | – | 1 | 1 |
| | O | 0 | *w* | – | 0 |
| | *P* | 0 | 0 | *x* | – |

The letters v, *w* and *x* each have a value of‘0’ or T’.

A T in the matrix shows that the team named in that row defeated the team named in that column.

A ‘O’ in the matrix shows that the team named in that row was defeated by the team named in that column.

A dash (-) in the matrix shows that no game was played.

The values of v, *w* and *x* are

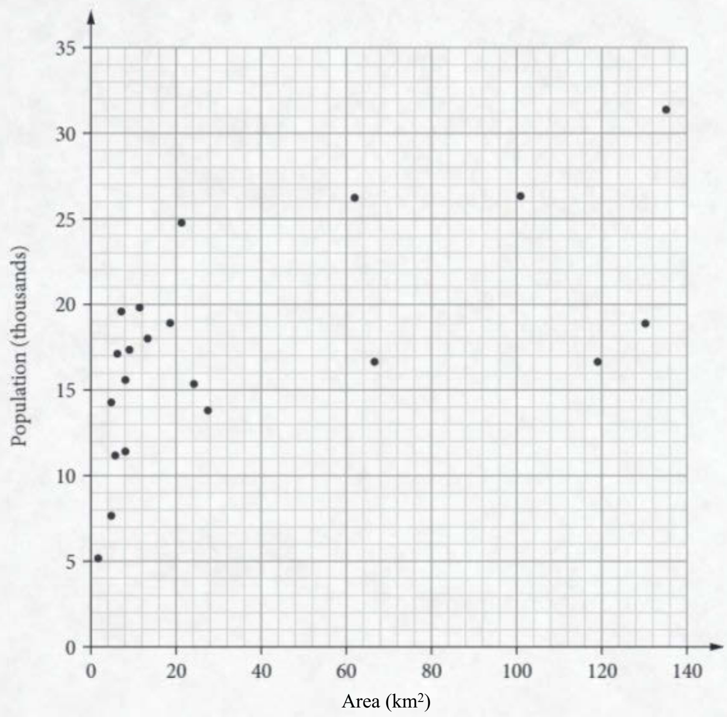
A v = 0, w = 1, x = 0 B v = 0, w = 1, x = 1 C v = 1, w = 0, x = 1

D v = 1, w = 1, x = 0 E v = 1, w = 1, x = 1

# Cumulative examination 2

Total number of marks: 24 Reading time: 7 minutes Writing time: 36 minutes

1 © VCAA 2014 2CQ3 J (2 marks) The scatterplot and table show the *population,* in thousands, and the *area,* in square kilometres, for a sample of 21 outer suburbs of a city.




| Area ($km^2$) | Population (thousands) |
|---|---|
| 1.6 | 5.2 |
| 4.4 | 14.3 |
| 4.6 | 7.5 |
| 5.6 | 11.0 |
| 6.3 | 17.1 |
| 7.0 | 19.4 |
| 7.3 | 15.5 |
| 8.0 | 11.3 |
| 8.8 | 17.1 |
| 11.1 | 19.7 |
| 13.0 | 17.9 |
| 18.5 | 18.7 |
| 21.3 | 24.6 |
| 24.2 | 15.2 |
| 27.0 | 13.6 |
| 62.1 | 26.1 |
| 66.5 | 16.4 |
| 101.4 | 26.2 |
| 119.2 | 16.5 |
| 130.7 | 18.9 |
| 135.4 | 31.3 |

In the outer suburbs, the association between *population* and *area* is non-linear. A log transformation can be applied to the variable *area* to linearise the scatterplot.

a Apply the log transformation to the data and determine the equation of the least squares line of best fit that allows the population of an outer suburb to be predicted from the logarithm of its area. Copy and complete the following, writing the slope and intercept of this line of best fit in the boxes. Write your answers, correct to one decimal place.

*population* = ☐ + ☐ x $\log_{10}$ *(area)* 1 mark

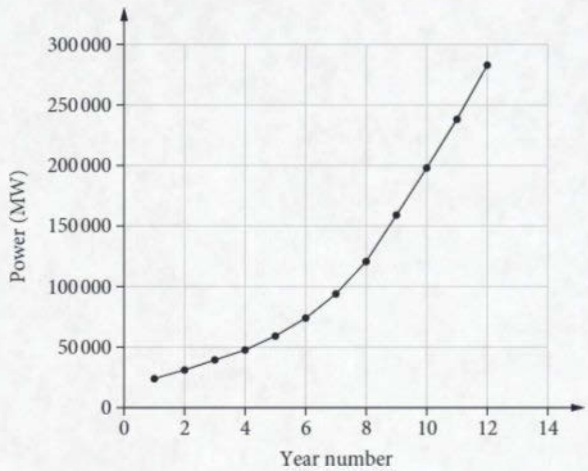
b Use this line of best fit equation to predict the population of an outer suburb with an area of 90 $km^2$. Write your answer, correct to the nearest one thousand people. 1 mark



9780170448406

2 ©VCAA 2018N 2CQ6 J (4 marks) The time series data shows the worldwide growth in electrical *power* generated by wind, in megawatts, for the period 2001-2012. The variable that represents *time,* in years, has been rescaled so that '1' represents 2001, '2' represents 2002, and so on. This new variable is called year *number.* A time series plot for the data is also shown.

| Year number | Power (MW) |
|---|---|
| 1 | 23900 |
| 2 | 31100 |
| 3 | 39431 |
| 4 | 47620 |
| 5 | 59091 |
| 6 | 73957 |
| 7 | 93924 |
| 8 | 120696 |
| 9 | 159052 |
| 10 | 197956 |
| 11 | 238110 |
| 12 | 282850 |




Data: Global Wind Energy Council (GWEC), Global Statistics, 'Global Cumulative Installed Wind Capacity 2001 -2016', <www.gwec.net/>

CUMULATIVE EXAMINATION 2

The relationship between *power and year number* is clearly non-linear. A $\log_{10}$ transformation can be applied to the variable *power* to linearise the data.

a Apply this transformation to the data to determine the equation of the least squares line that can be used to predict $\log_{10}$ *(power)* from *year number.* Copy and complete the following, writing the values of the intercept and slope of this least squares line in the appropriate boxes. Round your answers to four significant figures.

log $i_0$ (power)= ☐ + ☐ x *year number* 2 marks

b Use the equation in part a to predict the electrical *power,* in megawatts, expected to be generated by wind in 2020. Round your answer to the nearest 1000 MW. 2 marks

3 ©VCAA 2005 2BRMQ3 (5 marks) Lena has some money that she wishes to invest for a period of five years. She is considering three investment options,

a Investment Option A

$10000 is deposited into an account with an interest rate of 4.8% per annum compounding monthly for five years. Calculate the value of Investment Option A at the end of five years. Write your answer correct to the nearest cent. 1 mark

b Investment Option B

$4000 is deposited into an account with an interest rate of 4.8% per annum compounding monthly. At the end of each month, for a period of five years, a further $100 is deposited after interest has been paid. Determine the value of Investment Option B at the end of five years (immediately after the $100 has been deposited). Write your answer correct to the nearest cent. 1 mark

c Investment Option C

Investment Option B is followed for two years. After this, the amount deposited at the end of each month changes. With the new monthly deposit, Investment Option C is worth $13 000 at the end of the five years.

i Find the new amount deposited at the end of each month for the remaining three years. Write your answer correct to the nearest cent. 1 mark

ii Determine the total amount of interest earned by Investment Option C over the five-year period. Write your answer correct to the nearest cent. 2 marks

4 ©VCAA 2019N2MQ1 *J* (5 marks) A total of six residents from two towns will be competing at the International Games. Matrix $A$ contains the number of male $(M)$ and the number of female (F) athletes competing from the towns of Gillen (G) and Haldaw (H).

$$A = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} \begin{matrix} G \\ H \end{matrix} \quad (\text{columns } M\ F)$$

a How many of these athletes are residents of Haldaw? 1 mark

Each of the six athletes will compete in one event: table tennis, running or basketball. Matrices $T$ and $R$ contain the number of male and female athletes from each town who will compete in table tennis and running respectively.

Table tennis Running

$$T = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{matrix} G \\ H \end{matrix} \qquad R = \begin{bmatrix} 1 & 1 \\ \text{OOH} \end{bmatrix} \text{G} \quad (\text{columns } M\ F)$$

b Matrix $B$ contains the number of male and female athletes from each town who will compete in basketball. Copy and complete matrix $B$. 1 mark

$$B = \begin{bmatrix} - & - \\ - & - \end{bmatrix} \begin{matrix} G \\ H \end{matrix} \quad (\text{columns } M\ F)$$

Matrix C contains the cost of one uniform, in dollars, for each of the three events: table tennis (T), running $(R)$ and basketball (B).

$$C = \begin{bmatrix} 515 \\ 550 \\ 580 \end{bmatrix} \begin{matrix} T \\ R \\ B \end{matrix}$$

c i For which event will the total cost of uniforms for the athletes be $1030? 1 mark

ii Write a matrix calculation, that includes matrix C, to show that the total cost of uniforms for the event named in part c i is contained in the matrix answer of [1030]. 1 mark

d Matrix $V$ and matrix $Q$ are two new matrices where $V$ = Q x C and:

- matrix Q is a 4 x 3 matrix
- element $v_l$ j = total cost of uniforms for all female athletes from Gillen
- element $v_{21}$ = total cost of uniforms for all female athletes from Haldaw
- element $v_{31}$ = total cost of uniforms for all male athletes from Gillen
- element $v_{41}$ = total cost of uniforms for all male athletes from Haldaw
- $C = \begin{bmatrix} 515 \\ 550 \\ 580 \end{bmatrix} \begin{matrix} T \\ R \\ B \end{matrix}$

Copy and complete matrix Q with the missing values. 1 mark

$$Q = {}^{0}\begin{bmatrix} 1 & - & - \\ & 0 & {}^{1} \\ \text{Oil} & & \\ - & - & 0 \end{bmatrix}$$

5 ©VCAA 2018N 2MQ1 I (4 marks) A region has four districts: North (N), South (S), East (£) and West (*W*). Farmers from each district attended a conference in 2017. Matrix $F_{2017}$ shows the number of farmers from each of these four districts who attended the 2017 conference.

$$F_{2017} = \begin{bmatrix} 36 \\ 20 \\ 28 \\ 16 \end{bmatrix} \begin{matrix} N \\ S \\ E \\ W \end{matrix}$$

a What is the order of matrix $F_{2017}$? 1 mark

b How many of these farmers came from either the North or South district? 1 mark

The table shows the cost per farmer, for each district, to attend the 2017 conference.

| District | Cost per farmer ($) |
|---|---|
| North | 25 |
| South | 20 |
| East | 45 |
| West | 35 |

c Write down a matrix that could be multiplied by matrix $F_{2017}$ to give the total cost for all farmers who attended the 2017 conference. 1 mark

d The number of farmers who attended the 2018 conference increased by 25% for each district from the previous year. Copy and complete the product below with a scalar so that the product gives the number of farmers from each district who attended the 2018 conference.

^2018 =---------------------- X ^2017 1 mark

6 ©VCAA 2012 2MQ1 MODIFIED , (4 marks) Matrix F shows the flight connections for an airline that serves four cities, Anvil (A), Berga (B), Cantor (C) and Dantel (D).

$$F = \text{From} \begin{matrix} A \\ B \\ C \\ D \end{matrix} \overset{\text{To}}{\overset{\begin{matrix} A & B & C & D \end{matrix}}{\begin{bmatrix} 0\ 10\ 0 \\ 10\ 0\ 1 \\ 0\ 10\ 0 \\ 0\ 0\ 10 \end{bmatrix}}}$$

In this matrix, the T' in row C column B, for example, indicates that, using this airline, you can fly directly from Cantor to Berga. The 'O' in row C column D, for example, indicates that you cannot fly directly from Cantor to Dantel.

a Copy and complete the following sentence.

On this airline, you can fly directly from Berga toand 1 mark

b List the route that you must follow to fly from Anvil to Cantor. 1 mark

c Evaluate the matrix product $G = KF$, where $K = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}$. 1 mark

d In the context of the problem, what information does matrix $G$ contain? 1 mark

CHAPTER

# 8 TRANSITION MATRICES

## Study Design coverage

### AREA OF STUDY 2: DISCRETE MATHEMATICS

**Transition matrices**

- use of the matrix recurrence relation: $S_o$ = initial state matrix, $S_{M+1} = TS_n$ *or* $S_{n+1} = LS_n$ where $T$ is a transition matrix, $L$ is a Leslie matrix, and S,, is a column state matrix, to generate a sequence of state matrices (assuming the next state only relies on the current state)
- informal identification of the equilibrium state matrix in the case of regular transition matrices (no noticeable change from one state matrix to the next state matrix)
- use of transition diagrams, their associated transition matrices and state matrices to model the transitions between states in discrete dynamical situations and their application to model and analyse practical situations such as the modelling and analysis of an insect population comprising eggs, juveniles and adults
- use of the matrix recurrence relation $S_o$ = initial state matrix, $S_{,,+1} = TS_n + B$ to extend modelling to populations that include culling and restocking.





**Video playlists (6):**

**8.1** Transition diagrams and matrices
**8.2** Transition matrices and recurrence relations
**8.3** Transition matrices and Long-term trends
**8.4** Transition matrices with restocking and culling
**8.5** Leslie matrices
**VCE question analysis** Transition matrices

To access resources above, visit
cengage.com.au/nelsonmindtap

Nelson MindTap

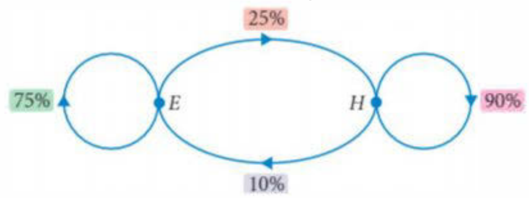
Video playlist
Transition diagrams and matrices

# I@ Transition diagrams and matrices

## Constructing transition diagrams and matrices

A state is a condition at a point in time. A transition matrix is a square matrix that shows a change from one state to another, where the change follows the same rules each time. A transition diagram shows transitions from one state to another using arrows and percentages. The following transition diagram and matching transition matrix show how shoppers at two shopping centres, Eastworld and Highton, change the centre they visit from one week to the next.

Transition diagram




Transition matrix

$$T = \begin{array}{c} \text{This week} \\ \begin{array}{cc} E & H \end{array} \\ \begin{bmatrix} 0.75 & 0.1 \\ 0.25 & 0.9 \end{bmatrix} \begin{array}{c} E \\ H \end{array} \text{ Next week} \end{array}$$

- 75% of shoppers who shop at Eastworld one week will shop at Eastworld the next week
- 25% of shoppers who shop at Eastworld one week will shop at Highton the next week
- 90% of shoppers who shop at Highton one week will shop at Highton the next week
- 10% shoppers who shop at Highton one week will shop at Eastworld the next week

We will only be looking at situations where each of the columns in the transition matrix add up to 1, which means the total number involved in the transition stays the same.

### Transition diagrams and matrices

Transition diagrams

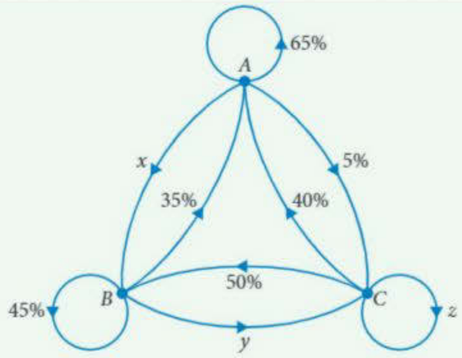
- show transitions using percentages or decimals
- have all the arrow percentages *from* a single point add up to 100%
- do not show transitions that are 0%.

Transition matrices

- are square matrices
- show the transitions as decimals
- have each column adding up to 1
- show transitions that are 0.

## WORKED EXAMPLE 1 Constructing transition matrices from diagrams

Construct transition matrices for each of the following.

| Steps | Working |
|---|---|
| **a** For this transition diagram showing changes from one year to the next<br>**i** find $x$, $y$, $z$<br>**ii** construct the matching transition matrix. |  |
| **i** All the arrow percentages *from* a single point add up to 100%.<br>Solve for the unknowns, using CAS if necessary. | For arrows from $A$:<br>$x + 65 + 5 = 100$<br>$x = 30\%$<br>For arrows from $B$:<br>$45 + y + 35 = 100$<br>$y = 20\%$<br>For arrows from $C$:<br>$50 + 40 + z = 100$<br>$z = 10\%$ |
| **ii** Convert all the percentages to decimals and construct the matrix. | This year<br>$\begin{array}{c} \begin{matrix} A & B & C \end{matrix} \\ \begin{bmatrix} 0.65 & 0.35 & 0.4 \\ 0.3 & 0.45 & 0.5 \\ 0.05 & 0.2 & 0.1 \end{bmatrix} \end{array} \begin{matrix} A \\ B \\ C \end{matrix}$ Next year |
| **b** If a train arrives at a certain station on time, the next train has an 85% chance of being on time. If a train arrives late, the next train has a 30% chance of being late. | |
| **1** Set up a 2 × 2 matrix using $O$ for 'on time' and $L$ for 'late' with a transition from 'This train' to 'Next train'. Enter the percentages from the question as decimals. | This train<br>$T = \begin{array}{c} \begin{matrix} O & L \end{matrix} \\ \begin{bmatrix} 0.85 & \\ & 0.3 \end{bmatrix} \end{array} \begin{matrix} O \\ L \end{matrix}$ Next train |
| **2** Enter the remaining elements of the matrix by using the fact that the columns of a transition matrix must add up to 1. | This train<br>$T = \begin{array}{c} \begin{matrix} O & L \end{matrix} \\ \begin{bmatrix} 0.85 & 0.7 \\ 0.15 & 0.3 \end{bmatrix} \end{array} \begin{matrix} O \\ L \end{matrix}$ Next train |

WB
p. 153

WB

p. 155

## WORKED EXAMPLE 2 Constructing transition diagrams from matrices

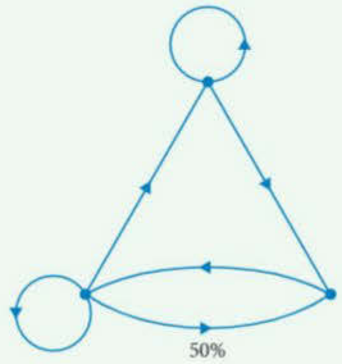
Given the following incomplete transition matrix and matching transition diagram

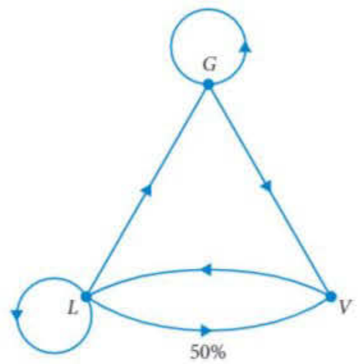
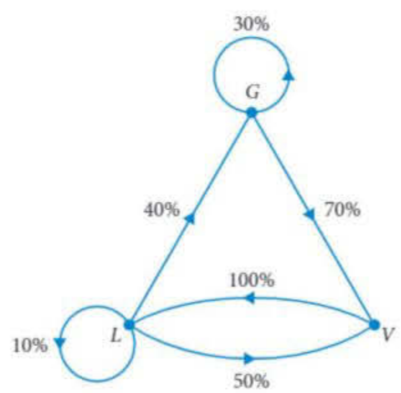
**a** find $x$, $y$, and $z$ to complete the transition matrix

$$\begin{array}{c} \text{This week} \\ \begin{array}{ccc} L & G & V \end{array} \end{array}$$

$$\begin{bmatrix} 0.1 & 0.0 & 1.0 \\ 0.4 & y & 0.0 \\ x & 0.7 & z \end{bmatrix} \begin{matrix} L \\ G \\ V \end{matrix} \text{ Next week}$$

**b** complete the matching transition diagram.




| **Steps** | **Working** |
|---|---|
| **a 1** Each column of a transition matrix must add up to 1. Solve for the unknowns, using CAS if necessary. | $0.1 + 0.4 + x = 1$ <br> $x = 0.5$ <br> $0.0 + y + 0.7 = 1$ <br> $y = 0.3$ <br> $1.0 + 0.0 + z = 1$ <br> $z = 0.0$ |
| **2** Complete the transition matrix. | This week <br> $\begin{array}{ccc} L & G & V \end{array}$ <br> $\begin{bmatrix} 0.1 & 0.0 & 1.0 \\ 0.4 & 0.3 & 0.0 \\ 0.5 & 0.7 & 0.0 \end{bmatrix} \begin{matrix} L \\ G \\ V \end{matrix}$ Next week |
| **b 1** Use the transition matrix to label the points $L$, $G$ and $V$ in the diagram. <br> Which transition in the leading diagonal is 0? <br> Which transition is 0.5? | $V$ to $V$ is 0, so the bottom right point is $V$. <br> The transition from $L$ to $V$ is 0.5, so the bottom left point is $L$. <br> The remaining top point is $G$. <br>  |
| **2** Complete the transition diagram from the transition matrix. |  |







9780170448406

## Interpreting transition matrices

8.1

**WORKED EXAMPLE 3** Interpreting transition matrices

WB p. 157

A colony of butterflies migrates to three different locations A, *B* and C on a peninsula. The butterflies change their location each year according to the transition matrix T.

$$\begin{array}{c} \text{This year} \\ \begin{array}{ccc} A & B & C \end{array} \end{array}$$
$$T = \begin{bmatrix} 0.4 & 0.2 & 0.3 \\ 0.25 & 0.55 & 0 \\ 0.35 & 0.25 & 0.7 \end{bmatrix} \begin{matrix} A \\ B \text{ Next year} \\ C \end{matrix}$$

This year there are 400 butterflies in location A, 660 butterflies in location *B* and 210 butterflies in location C.

a How many butterflies in location *B* this year are expected to be in location C next year?

b How many butterflies in location C this year are expected to remain in location C next year?

c How many butterflies are expected to be in location A next year?

d If 48% of the butterflies were in location *B* one year, what percentage of these butterflies are expected to be in location A the next year?

| Steps | Working |
|---|---|
| a Locate the relevant element in the transition matrix and multiply by the number of butterflies. | 0.25 of butterflies in location *B* one year are expected be in location C the following year.<br>0.25 x 660 = 165 butterflies |
| b Locate the relevant element in the transition matrix and multiply by the number of butterflies. | 0.7 of butterflies in location C one year are expected to remain in location C the following year.<br>0.7 x 210 = 147 butterflies |
| c Locate the relevant elements in the transition matrix, multiply by the number of butterflies in each case, and add. | A to A, B to A, C to A<br>0.4 x 400 + 0.2 x 660 + 0.3 x 210 = 355<br>355 butterflies are expected be in location A next year. |
| d Locate the relevant element in the transition matrix and multiply by the percentage. | 0.2 butterflies in *B* are expected to be in A the next year.<br>0.2 x 48% = 9.6% |

**WORKED EXAMPLE 4** Using permutation matrices fortransitions

WB p. 158

Jeremy completed a multiple-choice test where each question had five possible answers, A, B, C, *D* and *E*. For question number one, Jeremy selected the answer *D*. For the remaining questions, Jeremy selected his answers according to the permutation matrix *P*.

$$\begin{array}{c} \text{This question} \\ \begin{array}{ccccc} A & B & C & D & E \end{array} \end{array}$$
$$P = \begin{bmatrix} 0 & 10 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0\ 0\ 10\ 0 & & & & \end{bmatrix} \begin{matrix} A \\ B \\ C \text{ Next question} \\ D \\ E \end{matrix}$$

a Explain why P is a transition matrix.

b What does the 'I' in column *B* and row A mean?

c What answer did Jeremy select for question number five?

| Steps | Working |
|---|---|
| a Is it a square matrix? Does each column add up to 1? | P is a transition matrix because it is a square matrix and each column adds up to 1. |
| b AT indicates that the transition occurs 100% of the time. | The T in column B and row A means that Jeremy will always select A after he has selected B. |
| c The T s' indicate 100% certainty in choices.<br>Use the transition matrix to list the selections starting from question number one. | Jeremy selected the answer *D* for question one.<br>His first five selections are<br>*D^C^E^B^A*<br>Jeremy selected A for question number five. |

## EXERCISE 8.1 Transition diagrams and matrices

ANSWERS p. 719

### Mastery

1 State whether each matrix could be a transition matrix, giving a reason when it is not one.

a $\begin{bmatrix} 0.6 & 0.7 \\ 0.4 & 0.3 \end{bmatrix}$

b $\begin{bmatrix} 0.6 & 0.7 \\ 0.3 & 0.1 \\ 0.1 & 0.2 \end{bmatrix}$

c $\begin{bmatrix} 0.9 & 0.1 \\ 0.4 & 0.9 \end{bmatrix}$

d $\begin{bmatrix} 0.25 & 0.85 & 0.7 \\ 0.45 & 0.1 & 0.15 \\ 0.3 & 0.05 & 0.15 \end{bmatrix}$

e $\begin{bmatrix} 1 & 0.5 & 0 & 0 \\ 0 & 0.5 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

f $\begin{bmatrix} 0.4 & 0 & 0.5 & 1 \\ 0.6 & 1 & 0.5 & 0 \end{bmatrix}$

g the 6x6 identity matrix

h a 5 x 5 permutation matrix

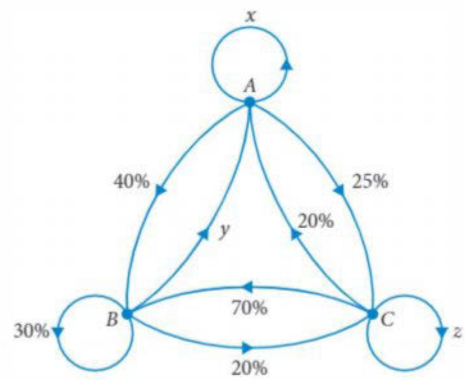
2 H WORKED EXAMPLE 1 Construct transition matrices for each of the following.

a For this transition diagram showing changes from one year to the next

i find x, y, $z$

ii construct the matching transition matrix.

b The Bureau of Meteorology has established that in a particular town, if it rains on any day, there is a 65% chance of it raining the next day, and if it is dry on any day, there is a 90% chance it will be dry the next day.



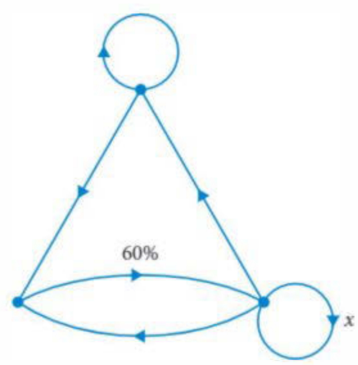

3 H WORKED EXAMPLE 2 Given the following incomplete transition matrix and matching transition diagram

a find x, y, and $z$ to complete the transition matrix

b complete the matching transition diagram.

This week

| S | $H$ | IV | |
|---|---|---|---|
| x | 0.0 | 0.9 | S |
| 0.4 | 0.3 | 0.0 | H Next week |
| $y$ | 0.7 | $z$ | IV |



▶ *S WORKED EXAMPLE 3 J A polling company has established a pattern of voter behaviour in a particular electorate regarding the three major political parties: Labor (L), Coalition (C) and Greens (G). The pattern from one election to the nextis shown in the transition matrix T below.

$$\begin{array}{c} \text{This election} \\ \begin{array}{ccc} L & C & G \end{array} \end{array}$$

$$T = \begin{bmatrix} 0.8 & 0.1 & 0.5 \\ 0 & 0.8 & 0 \\ 0.2 & 0.1 & 0.5 \end{bmatrix} \begin{matrix} L \\ C \text{ Next election} \\ G \end{matrix}$$

In the most recent election, 32425 people voted Labor, 26650 voted Coalition and 10044 voted Greens,

a How many people who voted Labor in this election are expected to vote Greens in the next election?

b How many people who voted Coalition in this election are expected to vote Coalition again in the next election?

c How many people are expected to vote Labor in the next election?

d If 48% of the people voted Coalition one year, what percentage of these Coalition voters are expected to vote Greens in the next election?

5 H WORKED EXAMPLE 4 I Amisha completed a multiple-choice test where each question had five possible answers, A, B, C, *D* and *E*. For question number one, Amisha selected the answer C. For the remaining questions, Amisha selected her answers according to the following the permutation matrix *P*.

$$\begin{array}{c} \text{This question} \\ A\ B\ C\ D\ E \end{array}$$

$$p = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{matrix} A \\ B \\ C \text{ Next question} \\ D \\ E \end{matrix}$$

a Explain why P is a transition matrix.

b What does the T' in column *D* and row *B* mean?

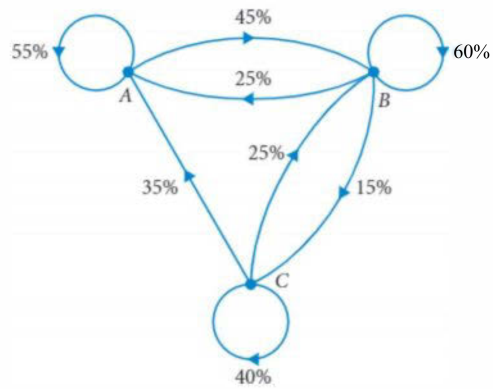
c What answer did Amisha select for question number six?

▶

## ▶ Exam practice

80-100% 60-79% 0-59%

**6** ©VCAA | 20201MQ4 90% In a particular supermarket, the three top-selling magazines are *Angel* (A), *Bella* (B) and *Crystal* (C). The transition diagram shows the way shoppers at this supermarket change their magazine choice from week to week.




A transition matrix that provides the same information as the transition diagram is

**A** This week

$$\begin{array}{c} \quad A \quad\; B \quad\; C \\ \begin{bmatrix} 0.55 & 0.70 & 0.35 \\ 0.70 & 0.60 & 0.40 \\ 0.35 & 0.40 & 0.40 \end{bmatrix} \end{array} \begin{array}{l} A \\ B \text{ Next week} \\ C \end{array}$$

**B** This week

$$\begin{array}{c} ABC \\ \begin{bmatrix} 0.55 & 0.60 & 0.25 \\ 0.45 & 0.15 & 0.35 \\ 0 & 0.25 & 0.40 \end{bmatrix} \end{array} \begin{array}{l} A \\ B \text{ Next week} \\ C \end{array}$$

**C** This week

$$\begin{array}{c} \quad A \quad\; B \quad\; C \\ \begin{bmatrix} 0.55 & 0.25 & 0.35 \\ 0.45 & 0.60 & 0.25 \\ 0 & 0.15 & 0.40 \end{bmatrix} \end{array} \begin{array}{l} A \\ B \text{ Next week} \\ C \end{array}$$

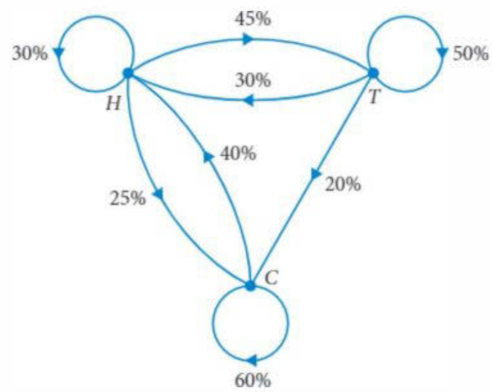
**D** This week

$$\begin{array}{c} ABC \\ \begin{bmatrix} 0.55 & 0.25 & 0.35 \\ 0.45 & 0.60 & 0.25 \\ 0.35 & 0.15 & 0.40 \end{bmatrix} \end{array} \begin{array}{l} A \\ B \text{ Next week} \\ C \end{array}$$

**E** This week

$$\begin{array}{c} \quad A \quad\; B \quad\; C \\ \begin{bmatrix} 0.55 & 0.25 & 0 \\ 0.45 & 0.60 & 0.25 \\ 0 & 0.15 & 0.75 \end{bmatrix} \end{array} \begin{array}{l} A \\ B \text{ Next week} \\ C \end{array}$$

**7** ©VCAA 20161MQ6 86% Families in a country town were asked about their annual holidays. Every year, these families choose between staying at home (H), travelling (T) and camping (C). The transition diagram shows the way families in the town change their holiday preferences from year to year.




A transition matrix that provides the same information as the transition diagram is

**A** From

$$\begin{array}{c} \quad H \quad\; T \quad\; C \\ \begin{bmatrix} 0.30 & 0.75 & 0.65 \\ 0.75 & 0.50 & 0.20 \\ 0.65 & 0.20 & 0.60 \end{bmatrix} \end{array} \begin{array}{l} H \\ T \quad \text{To} \\ C \end{array}$$

**B** From

$$\begin{array}{c} \quad H \quad\; T \quad\; C \\ \begin{bmatrix} 0.30 & 0.30 & 0.40 \\ 0.45 & 0.50 & 0 \\ 0.25 & 0.20 & 0.60 \end{bmatrix} \end{array} \begin{array}{l} H \\ T \quad \text{To} \\ C \end{array}$$

**C** From

$$\begin{array}{c} \quad H \quad\; T \quad\; C \\ \begin{bmatrix} 0.30 & 0.30 & 0.40 \\ 0.45 & 0.50 & 0.20 \\ 0.25 & 0.20 & 0.60 \end{bmatrix} \end{array} \begin{array}{l} H \\ T \text{ To} \\ C \end{array}$$

**D** From

$$\begin{array}{c} \quad H \quad\; T \quad\; C \\ \begin{bmatrix} 0.30 & 0.30 & 0.40 \\ 0.45 & 0.50 & 0.20 \\ 0.25 & 0.20 & 0.40 \end{bmatrix} \end{array} \begin{array}{l} H \\ T \quad \text{To} \\ C \end{array}$$

**E** From

$$\begin{array}{c} \quad H \quad\; T \quad\; C \\ \begin{bmatrix} 0.30 & 0.45 & 0.25 \\ 0.30 & 0.50 & 0.20 \\ 0.40 & 0 & 0.60 \end{bmatrix} \end{array} \begin{array}{l} H \\ T \quad \text{To} \\ C \end{array}$$



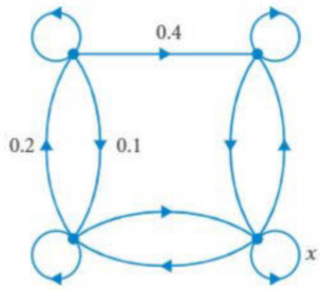
9780170448406

**8** ©VCAA 20181MQ6 81%

A transition matrix, V, is shown.

$$V = \begin{array}{c} \text{This month} \\ \begin{array}{cccc} L & T & F & M \end{array} \\ \left[\begin{array}{cccc} 0.6 & 0.6 & 0.2 & 0.0 \\ 0.1 & 0.2 & 0.0 & 0.1 \\ 0.3 & 0.0 & 0.8 & 0.4 \\ 0.0 & 0.2 & 0.0 & 0.5 \end{array}\right] \end{array} \begin{array}{l} L \\ T \\ F \\ M \end{array} \text{ Next month}$$

The following transition diagram has been constructed from the transition matrix $V$. The labelling in the transition diagram is not yet complete.




The proportion for one of the transitions is labelled $x$. The value of $x$ is

A 0.2 B 0.5 C 0.6

D 0.7 E 0.8

**9** ©VCAA 20151MQ5 72% Wendy buys one type of flower each day. She chooses from tulips (T), roses *(R),* carnations (C), irises (Z) and daisies (D). The type of flower she buys on one day depends on the type of flower she bought the previous day, according to a transition matrix. Today, Wendy bought tulips. The transition matrix that, starting tomorrow, ensures Wendy buys flowers in alphabetical order (C, D, 1, *R, T)* is

A
$$\begin{array}{c} \text{Today} \\ \begin{array}{ccccc} T & R & C & I & D \end{array} \\ \left[\begin{array}{ccccc} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right] \end{array} \begin{array}{l} T \\ R \\ C \\ I \\ D \end{array} \text{ Tomorrow}$$

B
$$\begin{array}{c} \text{Today} \\ \begin{array}{ccccc} T & R & C & I & D \end{array} \\ \left[\begin{array}{ccccc} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right] \end{array} \begin{array}{l} T \\ R \\ C \\ I \\ D \end{array} \text{ Tomorrow}$$

C
$$\begin{array}{c} \text{Today} \\ \begin{array}{ccccc} T & R & C & I & D \end{array} \\ \left[\begin{array}{ccccc} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{array}\right] \end{array} \begin{array}{l} T \\ R \\ C \\ I \\ D \end{array} \text{ Tomorrow}$$

D
$$\begin{array}{c} \text{Today} \\ \begin{array}{ccccc} T & R & C & I & D \end{array} \\ \left[\begin{array}{ccccc} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{array}\right] \end{array} \begin{array}{l} T \\ R \\ C \\ I \\ D \end{array} \text{ Tomorrow}$$

E
$$\begin{array}{c} \text{Today} \\ \begin{array}{ccccc} T & R & C & I & D \end{array} \\ \left[\begin{array}{ccccc} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{array}\right] \end{array} \begin{array}{l} T \\ R \\ C \\ I \\ D \end{array} \text{ Tomorrow}$$

**10** ©VCAA 2019N1MQ5 A population of birds feed at two different locations, $A$ and B, on an island. The change in the percentage of the birds at each location from year to year can be determined from the transition matrix $T$ shown.

$$T = \begin{array}{c} \text{This year} \\ \begin{array}{cc} A & B \end{array} \\ \left[\begin{array}{cc} 0.8 & 0.4 \\ 0.2 & 0.6 \end{array}\right] \end{array} \begin{array}{l} A \\ B \end{array} \text{ Next year}$$

In 2018, 55% of the birds fed at location B. In 2019, the percentage of the birds that are expected to feed at location $A$ is

A 32% B 42% C 48%

D 58% E 62%

▶ **11** ©VCAA | 20141MQ8 60% Wendy will have lunch with one of her friends each day of this week. Her friends are Angela (A), Betty (B), Craig (C), Daniel (D) and Edgar (E). On Monday, Wendy will have lunch with Craig. Wendy will use the transition matrix to choose a friend to have lunch with for the next four days of the week.

$$\begin{array}{cc} & \text{Today} \\ & \begin{array}{ccccc} A & B & C & D & E \end{array} \\ T = & \left[\begin{array}{ccccc} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{array}\right] \begin{array}{l} A \\ B \\ C \text{ Tomorrow} \\ D \\ E \end{array} \end{array}$$

The order in which Wendy has lunch with her friends for the next four days is

A Angela, Betty, Craig, Daniel

B Daniel, Betty, Angela, Craig

C Daniel, Betty, Angela, Edgar

D Edgar, Angela, Daniel, Betty

E Edgar, Daniel, Betty, Angela

**12** ©VCAA 20191MQ4 56% Stella completed a multiple-choice test that had 10 questions. Each question had five possible answers, A, B, C, *D* and *E*. For question number one, Stella chose the answer *E*. Stella chose each of the nine remaining answers, in order, by following the transition matrix, T, below.

$$\begin{array}{cc} & \text{This question} \\ & \begin{array}{ccccc} A & B & C & D & E \end{array} \\ T = & \left[\begin{array}{ccccc} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{array}\right] \begin{array}{l} \text{A} \\ \text{B} \\ \text{C Next question} \\ \text{D} \\ E \end{array} \end{array}$$

What answer did Stella choose for question number six?

A A　　B B　　C C　　D *D*　　E *E*

**13** ©VCAA | 20101MQ9 29% Robbie completed a test of four multiple-choice questions. Each question had four alternatives, A, B, C or *D*. Robbie randomly guessed the answer to the first question. He then determined his answers to the remaining three questions by following the transition matrix shown.

$$\begin{array}{cc} & \text{This question} \\ & \begin{array}{cccc} A & B & C & D \end{array} \\ T = & \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{array}\right] \begin{array}{l} \text{A} \\ \text{B} \\ \text{C Next question} \\ D \end{array} \end{array}$$

Which of the following statements is true?

A It is impossible for Robbie to give the same answer to all four questions.

B Robbie would always give the same answer to the first and fourth questions.

C Robbie would always give the same answer to the second and third questions.

D If Robbie answered A for question one, he would have answered B for question two.

E It is possible that Robbie gave the same answer to exactly three of the four questions.


9780170448406

# @ Transition matrices and recurrence relations

8.2

## The state matrix recurrence relation

The power of transition matrices is that they can be used to make predictions. For example, in a particular industry 1640 people transition between casual (C) and permanent employment (P) from one year to the next according to the following transition matrix $T$.

$$T = \begin{array}{c} \text{This year} \\ \begin{matrix} C & P \end{matrix} \\ \begin{bmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{bmatrix} \end{array} \begin{matrix} C \\ P \end{matrix} \text{ Next year}$$

If we know that one year there were 1200 casual employees and 440 permanent employees, we can write this as a state matrix which shows the numbers at a point in time:

$$S_0 = \begin{bmatrix} 1200 \\ 440 \end{bmatrix} \begin{matrix} C \\ P \end{matrix}$$

Video playlist Transition matrices and recurrence relations

We can then use matrix multiplication to calculate how many casual and permanent employees we expect in the following two years:

$$S_1 = \begin{bmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{bmatrix}\begin{bmatrix} 1200 \\ 440 \end{bmatrix} = \begin{bmatrix} 1004 \\ 636 \end{bmatrix} \begin{matrix} C \\ P \end{matrix} \qquad S_2 = \begin{bmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{bmatrix}\begin{bmatrix} 1004 \\ 636 \end{bmatrix} = \begin{bmatrix} 867 \\ 773 \end{bmatrix} \begin{matrix} C \\ P \end{matrix}$$

The sum of the elements in the state matrices always stays the same:

$S_0$: 1200 + 440 = 1640

$S_1$: 1004 + 636= 1640

$S_2$: 867 + 773 = 1640

The state matrices for the following years can be generated by a matrix recurrence relation:

$$S_0 = \begin{bmatrix} 1200 \\ 440 \end{bmatrix}, \quad S_{n+1} = \begin{bmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{bmatrix} S_n$$

### © Exam hack

Remember the importance of order with matrices. Always multiply the transition matrix on the left.

### The state matrix recurrence relation

The recurrence relation that generates a sequence of state matrices $S_n$ is

$S_0$ = **initial state matrix**, $S_{n+1} = TS_n$

where

$T$ is a transition matrix

$n$ is the number of transitions.

The sum of the elements in $S_n$ is always the same.

WB

p. 159

**WORKED EXAMPLE 5** Using the state matrix recurrence relation

Shoppers at two shopping centres, Eastworld (E) and Highton (H), change the centre they visit from one week to the next according to the following transition matrix:

$$T = \begin{bmatrix} 0.75 & 0.11 \\ 0.25 & 0.89 \end{bmatrix} \begin{matrix} E \\ H \end{matrix}$$

This week: columns $E$, $H$; rows: Next week $E$, $H$

If we know that this week 23000 shoppers go to Eastworld and 12 000 go to Highton

| Steps | Working |
|---|---|
| a find the matrix recurrence relation that generates the sequence of state matrices | |
| Find the recurrence relation given by $S_o$ = initial state matrix, $S_{M+1} = TS_n$ | $S_0 = \begin{bmatrix} 23000 \\ 12000 \end{bmatrix}, S_{n+1} = \begin{bmatrix} 0.75 & 0.11 \\ 0.25 & 0.89 \end{bmatrix} S_n$ |
| b use the recurrence relation to calculate the number of shoppers predicted to go to Eastworld and Highton next week | |
| 1 Use the recurrence relation to find Sj | $S_1 = TS_0$<br>$S_1 = \begin{bmatrix} 0.75 & 0.11 \\ 0.25 & 0.89 \end{bmatrix}\begin{bmatrix} 23000 \\ 12\,000 \end{bmatrix} = \begin{bmatrix} 18570 \\ 16\,430 \end{bmatrix} \begin{matrix} E \\ H \end{matrix}$ |
| 2 Write the answer, rounding to the nearest whole number if necessary. | Next week 18570 shoppers are predicted to go to Eastworld and 16430 shoppers are predicted to go to Highton. |
| c show the calculations for the number of shoppers predicted to go to Eastworld and Highton next week | |
| Write out the steps for the calculations from the matrix multiplication. | Eastworld: 0.75 x 23 000 + 0.11 x 12 000 = 18 570<br>Highton: 0.25 x 23000 + 0.89 x 12000 = 16430 |
| d use the recurrence relation to calculate the number of shoppers expected to go to Eastworld and Highton the week after next. | |
| 1 Use the recurrence relation and your unrounded answer for S, to find $S_2$. | $S_2 = \begin{bmatrix} 0.75 & 0.11 \\ 0.25 & 0.89 \end{bmatrix}\begin{bmatrix} 18570 \\ 16430 \end{bmatrix} = \begin{bmatrix} 15735 \\ 19265 \end{bmatrix} \begin{matrix} E \\ H \end{matrix}$ |
| 2 Write the answer, rounding to the nearest whole number if necessary. | The week after next 15 735 shoppers are expected to go to Eastworld and 19265 shoppers are expected to go to Highton. |

WB

p. 160

**WORKED EXAMPLE 6** Interpreting the state matrix recurrence relation

A species of bird moves between three nesting sites $A$, $B$ and $C$ each year according to the recurrence relation

$$S_0 = \begin{bmatrix} 348 \\ 122 \\ 260 \end{bmatrix} \begin{matrix} A \\ B \\ C \end{matrix}, \ S_{n+1} = TS_n$$

where

$$T = \begin{bmatrix} 0.7 & 0 & 0.3 \\ 0.2 & 1 & 0.4 \\ 0.1 & 0 & 0.3 \end{bmatrix} \begin{matrix} A \\ B \\ C \end{matrix}$$

This year: columns $A$, $B$, $C$; rows: Next year $A$, $B$, $C$

and $S_0$ shows the numbers of birds at each site this year.



9780170448406

| Steps | Working |
|---|---|
| a Explain what the number 1 in the second row and second column of the transition matrix tells us. | |
| A T in the transition matrix indicates this transition is 100% certain to occur. | Once the birds nest at site B, they keep nesting there year after year. |
| b How many birds are expected to nest at the same site next year? | |
| 1 Identify the proportions in the transition matrix that represent no change. | This year $\begin{matrix} A & B & C \end{matrix}$ $T = \begin{bmatrix} 0.7 & 0 & 0.3 \\ 0.2 & 1 & 0.4 \\ 0.1 & 0 & 0.3 \end{bmatrix} \begin{matrix} A \\ B \\ C \end{matrix}$ Next year<br>0.7 of birds at site A this year will be at site *A* next year.<br>All the birds at site *B* this year will be at site *B* next year.<br>0.3 of birds at site C this year will be at site C next year. |
| 2 Use the initial state matrix to calculate the total number of birds that do not change site. | 0.7 x 348 + 1 x 122 + 0.3 x 260 = 443.6 |
| 3 Write the answer, rounding to the nearest whole number if necessary. | We expect 444 birds will nest at the same site next year. |
| c What percentage of birds nest at a different site next year? Round to the nearest percentage. | |
| 1 Subtract the number of birds nesting at the same site next year from the total number of birds. | Total number of birds = 348 + 122 + 260 = 730<br>Number of birds nesting on a different site<br>= 730 - number of birds nesting on the same site<br>= 730 - 444<br>= 286 |
| 2 Calculate the percentage, rounding to the nearest percentage. | The percentage of birds nesting at different site next year<br>$= \frac{286}{730} \times 100\%$<br>= 39% |

**WORKED EXAMPLE 7** Solving the state matrix recurrence relation

WB p. 161

Find the transition matrix Tby solving for the unknowns in the following state matrix recurrence relation.

$$S_0 = \begin{bmatrix} 100 \\ 20 \\ 40 \end{bmatrix}, \$n+l = TS_n$$

where

$$T = \begin{bmatrix} r\,0.1 & 0.2 \\ 5\,0.3 & 0.4 \\ *\,P & <1 \end{bmatrix} \quad S. = \begin{bmatrix} 60 \\ 96 \\ 58 \end{bmatrix}$$

| Steps | Working |
| --- | --- |
| 1 Use the fact that transition matrix columns add to 1. | $0.1 + 0.3 + p = 1$<br>$0.2 + 0.4 + q = 1$<br>$p = 0.6$ $q = 0.4$ |
| 2 Write the recurrence relation for $S_r$ | $S_1 = \begin{bmatrix} r & 0.1 & 0.2 \\ s & 0.3 & 0.4 \\ t & 0.6 & 0.4 \end{bmatrix} \begin{bmatrix} 100 \\ 20 \\ 40 \end{bmatrix} = \begin{bmatrix} 60 \\ 96 \\ 58 \end{bmatrix}$ |
| 3 Multiply the matrix rows with unknowns and solve the equations. | 100r + 0.1 x 20+ 0.2x40 = 60<br>lOOr + 2 + 8 = 60<br>100r=50<br>r = 0.5<br>100s + 0.3 x 20 + 0.4 x 40 = 96<br>100s+ 60+ 16 = 96<br>100s = 20<br>s = 0.2<br>lOOt+ 0.6x20+ 0.4x40 = 58<br>lOOt + 12+ 16 = 58<br>lOOt =30<br>t = 0.3 |
| 4 Write the transition matrix. | $T = \begin{bmatrix} 0.5 & 0.1 & 0.2 \\ 0.2 & 0.3 & 0.4 \\ 0.3 & 0.6 & 0.4 \end{bmatrix}$ |

## The state matrix rule

We can use the state matrix recurrence relation to find a state matrix rule that makes calculations easier, particularly when using CAS.

Using the state matrix recurrence relation:

| | |
| --- | --- |
| initial state: | So |
| after one transition: | $S_l = TxS_Q$ |
| after two transitions: | $S_2 = T$ X $S_1 = T$ X (T X $S_o$) = $T^2$ x $S_o$ |
| after three transitions: | $S_3$=Tx$S_2$=Tx($T^2$x$S_0$)-$r^3$x$S_0$ |
| after four transitions: | $S_4$=Tx$S_3$ = Tx($T^3$x$S_0$) = $T^4$x$S_0$ |
| ... | ... |
| after $n$ transitions: | s„ = rx$s_0$ |

If $T$ is a transition matrix then powers of $T$ are also transition matrices.

**The state matrix rule**

The rule for finding the state matrix $S_n$ after $n$ transitions is

$$S_n = T^n S_0$$

where

$T$ is a transition matrix

$T^n$ is a transition matrix

$S_0$ is the initial state matrix.



9780170448406

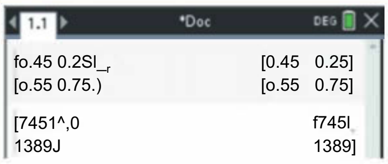
## USING CAS 1 Finding state matrices using the rule

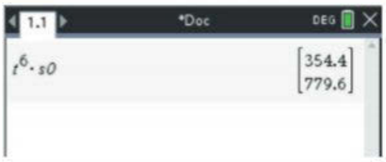
A market research company has analysed the petrol buying patterns of motorists in a small town regarding two petrol stations, GasStop (G) and Oils (O). It has established that the movement between the two occurs according to this transition matrix.

$$T = \begin{matrix} & \text{This week} & \\ & G \quad O & \\ & \begin{bmatrix} 0.45 & 0.25 \\ 0.55 & 0.75 \end{bmatrix} & \begin{matrix} G \\ O \end{matrix} \text{ Next week} \end{matrix}$$

This week, 745 motorists went to GasStop and 389 went to Oils. How many motorists go to each petrol station after six weeks?

**TI-Nspire**



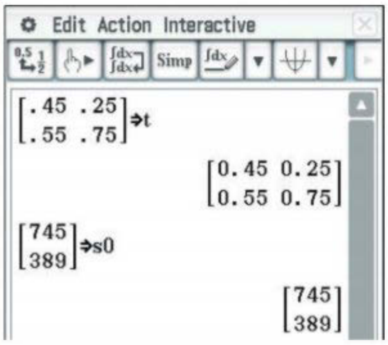

1 In a Calculator page, create a 2x2 transition matrix and store it in *t*.

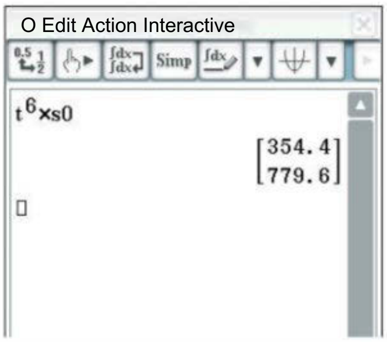
2 Create a 2x1 initial state matrix for the number of motorists and store it in sO.

3 To find the number of motorists after six weeks, calculate $t^6$xs0.

Rounding to the nearest whole number, 354 motorists went to GasStop and 780 went to Oils.

**ClassPad**




1 In the Main application, create a 2x2 transition matrix and store it in *t*.

2 Create a 2x1 initial state matrix for the number of motorists and store it in sO.

3 To find the number of motorists after six weeks, calculate $t^6$xs0.

Rounding to the nearest whole number, 354 motorists went to GasStop and 780 went to Oils.

### Exam hack

Round to the nearest whole number if it makes sense to do so, even if the question doesn't specifically ask you to.

WB

p. 162

**WORKED EXAMPLE 8** Using the state matrix rule

A fleet of trucks starts each day at one of two depots, *A* or *B*. By the end of the day, the trucks end up at either of the two depots according to the transition matrix *T*.

$$\begin{array}{c} \text{This day} \\ \begin{array}{cc} A & B \end{array} \end{array}$$
$$T = \begin{bmatrix} 0.95 & 0.15 \\ 0.05 & 0.85 \end{bmatrix} \begin{matrix} A \\ B \end{matrix} \text{ Next day}$$

At the start of a particular day, there are 40 trucks at depot *A* and 100 trucks at depot *B*.

| Steps | Working |
|---|---|
| **a** How many trucks are at each depot after six days? | |
| **1** Use CAS and the state matrix rule $S_n = T^n S_0$. | $S_0 = \begin{bmatrix} 40 \\ 100 \end{bmatrix}, n = 6$<br>$S_6 = T^6 S_0$<br>$= \begin{bmatrix} 0.95 & 0.15 \\ 0.05 & 0.85 \end{bmatrix}^6 \begin{bmatrix} 40 \\ 100 \end{bmatrix}$<br>$\approx \begin{bmatrix} 88 \\ 52 \end{bmatrix}$ |
| **2** Write the answer. | After six days, 88 trucks will be at depot *A* and 52 trucks will be at depot *B*. |
| **b** How many trucks are at each depot after ten days? | |
| **1** Use CAS and the state matrix rule $S_n = T^n S_0$. | $S_0 = \begin{bmatrix} 40 \\ 100 \end{bmatrix}, n = 10$<br>$S_{10} = T^{10} S_0$<br>$= \begin{bmatrix} 0.95 & 0.15 \\ 0.05 & 0.85 \end{bmatrix}^{10} \begin{bmatrix} 40 \\ 100 \end{bmatrix}$<br>$\approx \begin{bmatrix} 98 \\ 42 \end{bmatrix}$ |
| **2** Write the answer. | After ten days, 98 trucks will be at depot *A* and 42 trucks will be at depot *B*. |
| **c** After ten days, the movement of trucks changes according to the following transition matrix.<br>This day: *A* *B*<br>$R = \begin{bmatrix} 0.55 & 0.3 \\ 0.45 & 0.7 \end{bmatrix} \begin{matrix} A \\ B \end{matrix}$ Next day<br>Calculate the number of trucks expected to be at each depot after fifteen days. | |
| **1** Use CAS and the state matrix rule $S_n = T^n S_0$, making sure you always multiply the transition matrices on the left. | $S_{15} = R^5 T^{10} S_0$<br>$S_{15} = \begin{bmatrix} 0.55 & 0.3 \\ 0.45 & 0.7 \end{bmatrix}^5 \begin{bmatrix} 0.95 & 0.15 \\ 0.05 & 0.85 \end{bmatrix}^{10} \begin{bmatrix} 40 \\ 100 \end{bmatrix}$<br>$= \begin{bmatrix} 0.55 & 0.3 \\ 0.45 & 0.7 \end{bmatrix}^5 \begin{bmatrix} 98 \\ 42 \end{bmatrix}$<br>$= \begin{bmatrix} 56 \\ 84 \end{bmatrix}$ |
| **2** Write the answer. | After fifteen days, 56 trucks are expected to be at depot *A* and 84 trucks are expected to be at depot *B*. |



9780170448406

## EXERCISE 8.2 Transition matrices and recurrence relations

ANSWERS p. 719

### Recap

1 ©VCAA 2009 1MQ8 76% In a country town, people only have the choice of doing their food shopping at a store called Marks (M) or at a newly opened store called Foodies (F). A market researcher predicts that

- of those who do their food shopping at Marks this week, 70% will shop at Marks next week and 30% will shop at Foodies
- of those who do their food shopping at Foodies this week, 90% will shop at Foodies next week and 10% will shop at Marks.

A transition matrix that can be used to represent this situation is

A $T = \begin{matrix} \text{This week} \\ \begin{matrix} M & F \end{matrix} \\ \begin{bmatrix} 0.7 & 0.9 \\ 0.3 & 0.1 \end{bmatrix} \end{matrix} \begin{matrix} M \\ F \end{matrix} \text{ Next week}$

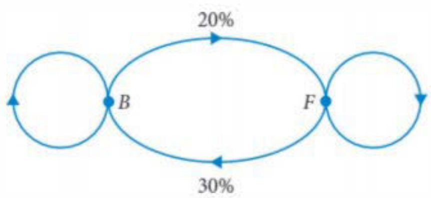
B $T = \begin{matrix} \text{This week} \\ \begin{matrix} M & F \end{matrix} \\ \begin{bmatrix} 0.7 & 0.1 \\ 0.3 & 0.9 \end{bmatrix} \end{matrix} \begin{matrix} M \\ F \end{matrix} \text{ Next week}$

C $T = \begin{matrix} \text{This week} \\ \begin{matrix} M & F \end{matrix} \\ \begin{bmatrix} 0.3 & 0.7 \\ 0.9 & 0.1 \end{bmatrix} \end{matrix} \begin{matrix} M \\ F \end{matrix} \text{ Next week}$

D $T = \begin{matrix} \text{This week} \\ \begin{matrix} M & F \end{matrix} \\ \begin{bmatrix} 0.7 & 0.3 \\ 0.1 & 0.9 \end{bmatrix} \end{matrix} \begin{matrix} M \\ F \end{matrix} \text{ Next week}$

E $T = \begin{matrix} \text{This week} \\ \begin{matrix} M & F \end{matrix} \\ \begin{bmatrix} 0.3 & 0.9 \\ 0.7 & 0.1 \end{bmatrix} \end{matrix} \begin{matrix} M \\ F \end{matrix} \text{ Next week}$

2 ©VCAA 2012 1MQ5 74% There are two fast-food shops in a country town: Big Burgers (B) and Fast Fries (F). Every week, each family in the town will purchase takeaway food from one of these shops. The transition diagram shows the way families in the town change their preferences for fast food from one week to the next.




A transition matrix that provides the same information as the transition diagram is

A $\begin{matrix} \text{From} \\ \begin{matrix} B & F \end{matrix} \\ \begin{bmatrix} 0 & 0.3 \\ 0.2 & 0 \end{bmatrix} \end{matrix} \begin{matrix} B \\ F \end{matrix} \text{ To}$

B $\begin{matrix} \text{From} \\ \begin{matrix} B & F \end{matrix} \\ \begin{bmatrix} 1.2 & -0.3 \\ -0.2 & 1.3 \end{bmatrix} \end{matrix} \begin{matrix} B \\ F \end{matrix} \text{ To}$

C $\begin{matrix} \text{From} \\ \begin{matrix} B & F \end{matrix} \\ \begin{bmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{bmatrix} \end{matrix} \begin{matrix} B \\ F \end{matrix} \text{ To}$

D $\begin{matrix} \text{From} \\ \begin{matrix} B & F \end{matrix} \\ \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix} \end{matrix} \begin{matrix} B \\ F \end{matrix} \text{ To}$

E $\begin{matrix} \text{From} \\ \begin{matrix} B & F \end{matrix} \\ \begin{bmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{bmatrix} \end{matrix} \begin{matrix} B \\ F \end{matrix} \text{ To}$

## Mastery

30 WORKED EXAMPLE 5 Sales of the Breakfast Club chains two most popular items, the bacon roll (B) and the egg wrap (E), have been shown to change from one week to the next according to the following transition matrix

$$T = \begin{array}{c} \text{This week} \\ \begin{array}{cc} B & E \end{array} \\ \begin{bmatrix} 0.4 & 0.35 \\ 0.6 & 0.65 \end{bmatrix} \end{array} \begin{array}{l} B \\ E \end{array} \text{ Next week}$$

If we know that 4243 bacon rolls and 6010 egg wraps were sold this week

a find the matrix recurrence relation that generates the sequence of state matrices

b use the recurrence relation to calculate the number of bacon rolls and egg wraps predicted to be sold next week

c show the calculations that give the number of bacon rolls and egg wraps predicted to be sold next week

d use the recurrence relation to calculate the number of bacon rolls and egg wraps expected to be sold the week after next.

4 Ed WORKED EXAMPLE 6~J A mathematics-inspired theme park has three locations: Algebra World (A), Pythagoras World (P) and Calculus World (C). The number of visitors at each of the three locations is counted every hour. By 10 am on a Sunday, the park reached its capacity and could take no more visitors. For the rest of the day people move between the three locations every hour according to the recurrence relation

$$S_0 = \begin{bmatrix} 3024 \\ 2568 \\ 2006 \end{bmatrix} \begin{array}{l} A \\ p \\ c \end{array} . \ S_{n+1} = TS_n$$

where

$$T = \begin{array}{c} \text{This hour} \\ \begin{array}{ccc} A & P & C \end{array} \\ \begin{bmatrix} 1 & 0.6 & 0.2 \\ 0 & 0.3 & 0.1 \\ 0 & 0.1 & 0.7 \end{bmatrix} \end{array} \begin{array}{l} A \\ P \text{ Next hour} \\ C \end{array}$$

and $S_o$ shows the number of people at each location at 10 am.

a Explain what the number 1 in the first row and first column of the transition matrix tells us.

b How many people are expected to be at the same location at 11 am?

c What percentage of people go to a different location at 11 am? Round to the nearest percentage.

5 Q WORKED EXAMPLE 7 Find the transition matrix $T$ by solving for the unknowns in the following state matrix recurrence relation.

$$S_0 = \begin{bmatrix} 300 \\ 100 \\ 600 \end{bmatrix}, \ S_{n+1} = TS_n$$

where

$$T = \begin{bmatrix} 0.2 & b & 0.6 \\ a & c & e \\ 0.5 & d & 0 \end{bmatrix} \qquad S_1 = \begin{bmatrix} 430 \\ 400 \\ 170 \end{bmatrix}$$

8.2

▶ **60** using CAS 1J A market research company has analysed the fast-food buying patterns of people in a small town regarding two fast-food outlets, FushNChups (F) and BurgerHQ (B). It has established that the movement between the two occurs according to this transition matrix.

$$T = \begin{bmatrix} 0.6 & 0.25 \\ 0.4 & 0.75 \end{bmatrix} \begin{matrix} F \\ B \end{matrix} \text{ Next week}$$

(This week: columns $F$, $B$)

In the current week, 64 people bought from FushNChups and 82 bought from BurgerHQ. How many people go to each fast-food outlet after five weeks?

7 H WORKED EXAMPLE 8 A number of train carriages start each day at one of two depots, $A$ or B. By the end of the day, the train carriages end up at either of the two depots according to the transition matrix T.

$$T = \begin{bmatrix} 0.8 & 0.15 \\ 0.2 & 0.85 \end{bmatrix} \begin{matrix} A \\ B \end{matrix} \text{ Next day}$$

(This day: columns $A$, $B$)

At the start of a particular day, there are 81 train carriages at depot $A$ and 49 train carriages at depot B.

a How many train carriages are at each depot after five days?

b How many train carriages are at each depot after twelve days?

c After twelve days, the movement of train carriages changes according to the transition matrix $R$.

$$R = \begin{bmatrix} 0.35 & 0.5 \\ 0.65 & 0.5 \end{bmatrix} \begin{matrix} A \\ B \end{matrix} \text{ Next day}$$

(This day: columns $A$, $B$)

Calculate the number of train carriages expected to be at each depot after twenty days.

## Exam practice

80-100% 60-79% 0-59%

8 ©VCAA 20091MQ7 I 84% In a country town, people only have the choice of doing their food shopping at a store called Marks (Af) or at a newly opened store called Foodies (F). In the first week that Foodies opened, only 300 of the town's 800 shoppers did their food shopping at Marks. The remainder did their food shopping at Foodies. A state matrix $S_o$ that can be used to represent this situation is

A $S_o = \begin{bmatrix} 300 \\ 800 \end{bmatrix} \begin{matrix} M \\ F \end{matrix}$ B $S_o = \begin{bmatrix} 500 \\ 300 \end{bmatrix} \begin{matrix} M \\ F \end{matrix}$ C $S_o = \begin{bmatrix} 800 \\ 300 \end{bmatrix} \begin{matrix} M \\ F \end{matrix}$

D $SQ - \begin{bmatrix} 300 \\ 500 \end{bmatrix} \begin{matrix} M \\ F \end{matrix}$ E $S_a = \begin{bmatrix} 800 \\ 500 \end{bmatrix} \begin{matrix} M \\ F \end{matrix}$

9 ©VCAA 2020 1MQ10 41% Consider the matrix recurrence relation shown.

$$So \text{“} \begin{bmatrix} 30 \\ 20 \\ 40 \end{bmatrix} > S_{M+1} - TS_n \qquad \text{where } T = \begin{bmatrix} j & 0.3 & I \\ 0.2 & tn & 0.3 \\ 0.4 & 0.2 & n \end{bmatrix}$$

Matrix $T$ is a transition matrix.

Given the information above and that $Sj = \begin{bmatrix} 42 \\ 28 \\ 20 \end{bmatrix}$, which one of the following is true?

A $m > I$ B $j + 1 = 0.7$ C $j = n$ D $j > m$ E $I = nt + n$ ▶

10 ©VCAA 20151MQ9 I 40% A fast-food stand at the football sells pies (P) and chips (C). Each week, 300 customers regularly buy either a pie or chips, but not both, from this stand. For the first five weeks, the customers' choice of pie or chips is expected to change weekly according to the transition matrix $T_{ly}$ where

$$T_1 = \begin{bmatrix} 0.65 & 0.25 \\ 0.35 & 0.75 \end{bmatrix} \begin{matrix} P \\ C \end{matrix}\ \text{Next week}$$

(columns: This week, $P$ $C$)

After the first five weeks, due to expected cold weather, the customers' choice of pie or chips is expected to change weekly according to the transition matrix $T_2$> where

$$T_2 = \begin{bmatrix} 0.85 & 0.25 \\ 0.15 & 0.75 \end{bmatrix} \begin{matrix} P \\ C \end{matrix}\ \text{Next week}$$

(columns: This week, $P$ $C$)

In week 1,150 customers bought a pie and 150 customers bought chips. Let $S_0$ be the initial state matrix. The number of customers expected to buy a pie or chips in week 8 can be found by evaluating

A $T_2^8S_0$ B $r,^8s_0$ C $T_2^3(T,^5S_0)$ D $r,^3(T_2^5s_0)$ E $T_1^5(T_2^3S_0)$

11 ©VCAA 20201MQ7 38% A small shopping centre has two coffee shops: Fatimas (F) and Giorgios (G). The percentage of coffee-buyers at each shop changes from day to day, as shown in the transition matrix $T$.

$$T = \begin{bmatrix} 0.85 & 0.35 \\ 0.15 & 0.65 \end{bmatrix} \begin{matrix} F \\ G \end{matrix}\ \text{Tomorrow}$$

(columns: Today, $F$ $G$)

On a particular Monday, 40% of coffee-buyers bought their coffees at Fatimas. The matrix recursion relation $S_{M+1} = TS_n$ is used to model this situation. The percentage of coffee-buyers who are expected to buy their coffee at Giorgios' on Friday of the same week is closest to

A 31% B 32% C 34% D 45% E 68%

12 ©VCAA | 20171MQ8 I 30% Consider the matrix recurrence relation.

$$S_0 = \begin{bmatrix} 40 \\ 15 \\ 20 \end{bmatrix},\ S_{n+1} = TS_n$$

where

$$T = \begin{bmatrix} 0.3 & 0.2 & V \\ 0.2 & 0.2 & W \\ X & Y & Z \end{bmatrix}$$

Matrix $T$ is a transition matrix. Given this and that $S_1 = \begin{bmatrix} 29 \\ 13 \\ 33 \end{bmatrix}$, which one of the following expressions is not true?

A $W>Z$ B $Y>X$ C $V>Y$

D $V+W+Z=1$ E $X+Y+Z>1$

8.2

13 ©VCAA | 2019 2MQ2 | (4 marks) A theme park has four locations, Air World (A), Food World (F), Ground World (G) and Water World (*W*). The number of visitors at each of the four locations is counted every hour. By 10 am on Saturday the park had reached its capacity of 2000 visitors and could take no more visitors. The park stayed at capacity until the end of the day. The state matrix, $S_0$ , below, shows the number of visitors at each location at 10 am on Saturday.

$$S_0 = \begin{bmatrix} 600 \\ 600 \\ 400 \\ 400 \end{bmatrix} \begin{matrix} A \\ F \\ G \\ W \end{matrix}$$

a | 91% What percentage of the parks visitors were at Water World (W) at 10 am on Saturday? 1 mark

Let $S_n$ be the state matrix that shows the number of visitors expected at each location $n$ hours after 10 am on Saturday. The number of visitors expected at each location $n$ hours after 10 am on Saturday can be determined by the matrix recurrence relation below.

$$S_0 = \begin{bmatrix} 600 \\ 600 \\ 400 \\ 400 \end{bmatrix}, \quad S_{n+1} = T \times S_n$$

where

$$T = \begin{array}{c} \text{This hour} \\ \begin{matrix} \;A & \;\;F & \;\;G & \;\;W \end{matrix} \\ \begin{bmatrix} 0.1 & 0.2 & 0.1 & 0.2 \\ 0.3 & 0.4 & 0.6 & 0.3 \\ 0.1 & 0.2 & 0.2 & 0.1 \\ 0.5 & 0.2 & 0.1 & 0.4 \end{bmatrix} \end{array} \begin{matrix} A \\ F \\ G \\ W \end{matrix} \text{ Next hour}$$

b | 84% Copy and complete the state matrix, $S_1$ to show the number of visitors expected at each location at 11 am on Saturday. 1 mark

$$S_1 = \begin{bmatrix} \_\_\_ \\ \_\_\_ \\ 300 \\ \_\_\_ \end{bmatrix} \begin{matrix} A \\ F \\ G \\ W \end{matrix}$$

c 26% Of the 300 visitors expected at Ground World (G) at 11 am, what percentage was at either Air World (A) or Food World (F) at 10 am? 1 mark

The proportion of visitors moving from one location to another each hour on Sunday is different from Saturday. Matrix V, below, shows the proportion of visitors moving from one location to another each hour after 10 am on Sunday.

$$V = \begin{array}{c} \text{This hour} \\ \begin{matrix} A & F & G & W \end{matrix} \\ \begin{bmatrix} 0.3 & 0.4 & 0.6 & 0.3 \\ 0.1 & 0.2 & 0.1 & 0.2 \\ 0.1 & 0.2 & 0.2 & 0.1 \\ 0.5 & 0.2 & 0.1 & 0.4 \end{bmatrix} \end{array} \begin{matrix} A \\ F \\ G \\ W \end{matrix} \text{ Next hour}$$

Matrix *V* is similar to matrix *T* but has the first two rows of matrix *T* interchanged. The matrix product that will generate matrix *V* from matrix *T* is $V = M \times T$ where matrix *M* is a binary matrix.

d 21% Write down matrix *M*. 1 mark

▶ 14 ©VCAA 2019N 2MQ2 (2 marks) Three television channels, $C_p$ $C_2$ and $C_3$, will broadcast the International Games in the town of Gillen. Gillens 2000 residents are expected to change television channels from hour to hour as shown in the transition matrix $T$ below. The option for residents not to watch television *(NoTV)* at that time is also indicated in the transition matrix.

This hour

| | Q | $c_2$ | Q | *NoTV* | | |
|---|---|---|---|---|---|---|
| $T$ = | 0.50 | 0.05 | 0.10 | 0.20 | Q | Next hour |
| | 0.10 | 0.60 | 0.20 | 0.20 | Q | |
| | 0.25 | 0.10 | 0.50 | 0.10 | $c_3$ | |
| | 0.15 | 0.25 | 0.20 | 0.50 | *NoTV* | |

The state matrix $G_o$ below lists the number of Gillen residents who are expected to watch the games on each of the channels at the start of a particular day (9.00 am). Also shown is the number of Gillen residents who are not expected to watch television at that time.

$$G_0 = \begin{bmatrix} 100 \\ 400 \\ 100 \\ 1400 \end{bmatrix} \begin{matrix} q \\ Q \\ C_3 \\ NoTV \end{matrix}$$

a Copy and complete the calculation below to show that 835 Gillen residents are not expected to watch television *(NoTV)* at 10.00 am that day.

☐ x 100 + ☐ x 400 + ☐ x 100 + ☐ x 1400 = 835 1 mark

b Determine the number of residents expected to watch the games on $C_3$ at 11.00 am that day. 1 mark

15 ©VCAA | 2015 2MQ2^ (3 marks) The ability level of students in a music school is assessed regularly and classified as beginner (B), intermediate (I) or advanced *(A).* After each assessment, students either stay at their current level or progress to a higher level. Students cannot be assessed at a level that is lower than their current level. The expected number of students at each level after each assessment can be determined using the transition matrix, $T_v$

Before assessment

| | C! | Q | $C_3$ | | |
|---|---|---|---|---|---|
| r,- | 0.50 | 0 | 0 | *B* | |
| | 0.48 | 0.80 | 0 | *I* | After assessment |
| | 0.02 | 0.20 | 1 | *A* | |

a The element in the third row and third column of matrix $T_{\ell}$ is the number 1.
Explain what this tells you about the advanced-level students. 1 mark

Let matrix $S_n$ be a state matrix that lists the number of students at beginner, intermediate and advanced levels after $n$ assessments. The number of students in the school, immediately before the first assessment of the year, is shown in matrix $S_o$.

$$S_0 = \begin{bmatrix} 20 \\ 60 \\ 40 \end{bmatrix} \begin{matrix} B \\ I \\ A \end{matrix}$$

b i Write down the matrix Sj that contains the expected number of students at each level after one assessment. Write the elements of this matrix correct to the nearest whole number. 1 mark

ii How many intermediate-level students have become advanced-level students after one assessment? 1 mark ▶

16 ©VCAA 2008 2MQ4 (2 marks) By the end of each academic year, students at the university will have either passed, failed or deferred the year. Experience has shown that

- 88% of students who pass this year will also pass next year
- 10% of students who pass this year will fail next year
- 2% of students who pass this year will defer next year
- 52% of students who fail this year will pass next year
- 44% of students who fail this year will fail next year
- 4% of students who fail this year will defer next year
- 65% of students who defer this year will pass next year
- 10% of students who defer this year will fail next year
- 25% of students who defer this year will defer next year.

Twelve hundred and thirty students began a business degree in 2007. By the end of the 2007 academic year, 880 students had passed, 230 had failed, while 120 had deferred the year. No students have dropped out of the business degree permanently. Use this information to predict the number of business students who will defer the 2009 academic year.

**Exam hack**

For difficult two-mark questions, the first mark can sometimes be reasonably straightforward. In this question, one mark was awarded for finding the transition matrix.

# Transition matrices and long-term trends

Video playlist Transition matrices and Long-term trends

## Regular transition matrices

A regular transition matrix is a transition matrix that either has no zeros itself or any one of its powers has no zeros. For example:

$$T = \begin{bmatrix} 0.2 & 0.5 & 0.8 \\ 0 & 0.2 & 0.1 \\ 0.8 & 0.3 & 0.1 \end{bmatrix} \text{ is a regular transition matrix because } T^2 = \begin{bmatrix} 0.68 & 0.44 & 0.29 \\ 0.08 & 0.07 & 0.03 \\ 0.24 & 0.49 & 0.68 \end{bmatrix} \text{ has no zeros.}$$

When we raise a regular transition matrix to high powers (e.g. 39 or 40), the elements stop changing. For example:

$$T^{39} = \begin{bmatrix} 0.4894 & 0.4894 & 0.4894 \\ 0.05674 & 0.05674 & 0.05674 \\ 0.4539 & 0.4539 & 0.4539 \end{bmatrix} \qquad T^{40} = \begin{bmatrix} 0.4894 & 0.4894 & 0.4894 \\ 0.05674 & 0.05674 & 0.05674 \\ 0.4539 & 0.4539 & 0.4539 \end{bmatrix}$$

**Regular transition matrices**

To check to see if a transition matrix $T$ is regular:

1. Check to see if T has no zeros
2. If $T$ has zeros, then check to see if $T^2$ has no zeros
3. If $T^2$ has zeros, then check to see if $T^3$ has no zeros

and so on until you find a power of $T$ with no zeros.

## The equilibrium state matrix

Although transition matrices involve constant change, for regular transition matrices in the long term the state matrix gets to a point where it stops changing from one transition to the next. This is called the equilibrium state matrix (or steady state matrix). The transition changes continue to occur, but they reach a point where they are cancelling themselves out, so the result ends up the same as the previous one.

To find the equilibrium state matrix we need to use the state matrix rule to show that the state matrix is the same for two consecutive high powers (e.g. 39 and 40).

For example, for the regular transition matrix $T = \begin{bmatrix} 0.2 & 0.5 & 0.8 \\ 0 & 0.2 & 0.1 \\ 0.8 & 0.3 & 0.1 \end{bmatrix}$ and $S_0 = \begin{bmatrix} 3400 \\ 2100 \\ 800 \end{bmatrix}$

$$S_{39} = \begin{bmatrix} 0.2 & 0.5 & 0.8 \\ 0 & 0.2 & 0.1 \\ 0.8 & 0.3 & 0.1 \end{bmatrix}^{39} \begin{bmatrix} 3400 \\ 2100 \\ 800 \end{bmatrix}$$

$$= \begin{bmatrix} 0.4894 & 0.4894 & 0.4894 \\ 0.05674 & 0.05674 & 0.05674 \\ 0.4539 & 0.4539 & 0.4539 \end{bmatrix} \begin{bmatrix} 3400 \\ 2100 \\ 800 \end{bmatrix}$$

$$= \begin{bmatrix} 3083 \\ 357 \\ 2860 \end{bmatrix}$$

$$S_{40} = \begin{bmatrix} 0.2 & 0.5 & 0.8 \\ 0 & 0.2 & 0.1 \\ 0.8 & 0.3 & 0.1 \end{bmatrix}^{40} \begin{bmatrix} 3400 \\ 2100 \\ 800 \end{bmatrix}$$

$$= \begin{bmatrix} 0.4894 & 0.4894 & 0.4894 \\ 0.05674 & 0.05674 & 0.05674 \\ 0.4539 & 0.4539 & 0.4539 \end{bmatrix} \begin{bmatrix} 3400 \\ 2100 \\ 800 \end{bmatrix}$$

$$= \begin{bmatrix} 3083 \\ 357 \\ 2860 \end{bmatrix}$$

So the equilibrium state matrix is $\begin{bmatrix} 3083 \\ 357 \\ 2860 \end{bmatrix}$.

**Exam hack**

Look out for questions that use the words ‘long term’ or describe a situation that is long term. They will be asking for the equilibrium state matrix or for high powers of the transition matrix.

**The equilibrium state matrix**

A regular transition matrix $T$ is a transition matrix that either has no zeros itself or any one of its powers has no zeros.

For large values of $n$

- $T^n$ stops changing and has equal elements in each row
- the state matrix $S_n$ stops changing and becomes the equilibrium state matrix
- both $T^n$ and $S_M$ give information about long-term trends.

To find the equilibrium state matrix, check that two large consecutive values of $n$ give the same state matrix.

**Exam hack**

If a transition matrix doesn't have any zeros, then there will be an equilibrium state matrix.

**WORKED EXAMPLE 9** Finding the equilibrium state matrix

A fleet of trucks starts each day at one of two depots, $A$ or $B$. By the end of the day, the trucks end up at either of the two depots according to the following transition matrix.

$$\begin{array}{c} \text{This day} \\ \begin{array}{cc} A & B \end{array} \end{array}$$

$$T = \begin{bmatrix} 0.95 & 0.15 \\ 0.05 & 0.85 \end{bmatrix} \begin{array}{c} A \\ B \end{array} \text{ Next day}$$

At the start of a particular day, there are 40 trucks at depot $A$ and 100 trucks at depot $B$.

| Steps | Working |
|---|---|
| **a** Explain why there will be an equilibrium state matrix. | |
| Does $T$ or a power $T$ have any zero elements? Check $T$ first, then $T^2$, if necessary. | $T$ has no zero elements, so there will be an equilibrium state matrix. |
| **b** Find the equilibrium state matrix. | |
| **1** Use CAS and the rule $S_n = T^n S_0$ for two large consecutive values of $n$. | $S_0 = \begin{bmatrix} 40 \\ 100 \end{bmatrix} \begin{array}{c} A \\ B \end{array}$, choose $n = 39$ and 40<br>$S_{39} = T^{39} S_0 = \begin{bmatrix} 0.95 & 0.15 \\ 0.05 & 0.85 \end{bmatrix}^{39} \begin{bmatrix} 40 \\ 100 \end{bmatrix} \approx \begin{bmatrix} 105 \\ 35 \end{bmatrix}$<br>$S_{40} = T^{40} S_0 = \begin{bmatrix} 0.95 & 0.15 \\ 0.05 & 0.85 \end{bmatrix}^{40} \begin{bmatrix} 40 \\ 100 \end{bmatrix} \approx \begin{bmatrix} 105 \\ 35 \end{bmatrix}$ |
| **2** Are the two state matrices the same? | $S_{39} = S_{40}$, so $\begin{bmatrix} 105 \\ 35 \end{bmatrix} \begin{array}{c} A \\ B \end{array}$ is the equilibrium state matrix. |
| **c** How many trucks will be at each depot in the long term? | |
| Read from the equilibrium state matrix. | In the long term, 105 trucks will be at depot $A$ and 35 trucks will be at depot $B$. |
| **d** What percentage of trucks are at depot $B$ in the long term? Round your answer to the nearest percentage. | |
| Calculate from the equilibrium state matrix. | In the long term, $\frac{35}{140} \times 100\% = 25\%$ will be at depot $B$. |

WB p. 164

## Long-term trends

We can find long-term trends from a transition matrix alone even if no initial state matrix is given. Also, if a total number is given, we can find the equilibrium state matrix by creating *any* state matrix whose elements add up to the total.

WB

p. 165

## WORKED EXAMPLE 10 Finding long-term trends with transition matrices

A school runs a weekly 'Coping with Year 12' program throughout the year consisting of three electives: Study Skills (S), Managing Stress (M) and Organisational Skills (O). Each week, the Year 12 students choose one of the electives. The transition matrix shows how the students' choices change from week to week.

$$\begin{array}{c} \text{This week} \\ S \quad M \quad O \end{array}$$

$$T = \begin{bmatrix} 0.7 & 0.2 & 0.3 \\ 0.1 & 0.7 & 0 \\ 0.2 & 0.1 & 0.7 \end{bmatrix} \begin{matrix} S \\ M \\ O \end{matrix} \text{ Next week}$$

a Show that there will be an equilibrium state matrix.

b What percentage of students are expected to select Managing Stress each week in the long term? Round your answer to the nearest percentage.

c If there are 270 Year 12 students at the school, find the equilibrium state matrix and hence find how many students are expected to select Study Skills in the long term.

d The school decides to run the same program the following year, and it assumes the same transition matrix applies. If in the long term 108 students are expected to take Organisational Skills, what is the expected number of students who will select Study Skills?

e The school realises the following year that a different transition matrix $R$ applies. Describe the long-term trends shown by

$$\begin{array}{c} \text{This week} \\ S \quad M \quad O \end{array}$$

$$R = \begin{bmatrix} 0.1 & 0.2 & 0 \\ 0.8 & 0.3 & 0 \\ 0.1 & 0.5 & 1 \end{bmatrix} \begin{matrix} S \\ M \\ O \end{matrix} \text{ Next week}$$

| Steps | Working |
|---|---|
| a Does $T$ or a power of $T$ have any zero elements?<br>Check $T$ first, then $T^2$, if necessary. | $T$ has a zero element, so find $T^2$.<br>$T^2 = \begin{bmatrix} 0.7 & 0.2 & 0.3 \\ 0.1 & 0.7 & 0 \\ 0.2 & 0.1 & 0.7 \end{bmatrix}^2 = \begin{bmatrix} 0.57 & 0.31 & 0.42 \\ 0.14 & 0.51 & 0.03 \\ 0.29 & 0.18 & 0.55 \end{bmatrix}$<br>$T^2$ has no zero elements, so there will be an equilibrium state matrix. |
| b 1 Find $T^n$ for two large consecutive values of $n$. | Choose $n = 39$ and 40.<br>$T^{39} = \begin{bmatrix} 0.7 & 0.2 & 0.3 \\ 0.1 & 0.7 & 0 \\ 0.2 & 0.1 & 0.7 \end{bmatrix}^{39} = \begin{bmatrix} 0.4737 & 0.4737 & 0.4737 \\ 0.1579 & 0.1579 & 0.1579 \\ 0.3684 & 0.3684 & 0.3684 \end{bmatrix} \begin{matrix} S \\ M \\ O \end{matrix}$<br>$T^{40} = \begin{bmatrix} 0.7 & 0.2 & 0.3 \\ 0.1 & 0.7 & 0 \\ 0.2 & 0.1 & 0.7 \end{bmatrix}^{40} = \begin{bmatrix} 0.4737 & 0.4737 & 0.4737 \\ 0.1579 & 0.1579 & 0.1579 \\ 0.3684 & 0.3684 & 0.3684 \end{bmatrix} \begin{matrix} S \\ M \\ O \end{matrix}$ |
| 2 Convert the required decimal to a rounded percentage. | The percentage of students expected to select Managing Stress in the long term is $0.1579 \times 100\% = 16\%$. |

| | |
|---|---|
| c 1 Create an initial state matrix from the total given. | Choose an initial state matrix whose total is 270. The elements can be of any value as long as they add up to 270.<br>Let $S_o = \begin{bmatrix} 100 \\ 100 \\ 70 \end{bmatrix} \begin{matrix} S \\ M \\ O \end{matrix}$ |
| 2 Use CAS and the state matrix rule $S_n = T^n S_0$ for two large consecutive values of $n$. | $S_{39}=T^{39}S_0$<br>$= \begin{bmatrix} 0.7 & 0.2 & 0.3 \\ 0.1 & 0.7 & 0 \\ 0.2 & 0.1 & 0.7 \end{bmatrix}^{39} \begin{bmatrix} 100 \\ 100 \\ 70 \end{bmatrix}$<br>$\approx \begin{bmatrix} 128 \\ 43 \\ 99 \end{bmatrix}$<br>$S_{40} = T^{40}S_0$<br>$= \begin{bmatrix} 0.7 & 0.2 & 0.3 \\ 0.1 & 0.7 & 0 \\ 0.2 & 0.1 & 0.7 \end{bmatrix}^{40} \begin{bmatrix} 100 \\ 100 \\ 70 \end{bmatrix}$<br>$\approx \begin{bmatrix} 128 \\ 43 \\ 99 \end{bmatrix}$ |
| 3 Are the two state matrices the same? | $S_{39} = S_{40}$,<br>so $\begin{bmatrix} 128 \\ 43 \\ 99 \end{bmatrix} \begin{matrix} S \\ M \\ O \end{matrix}$ is the equilibrium state matrix. |
| 4 Read from the equilibrium state matrix. | In the long term, 128 will select Study skills each week. |
| d 1 Use $T^n$ for a large value of $n$ and the expected Organisation Skills student number given to find the total number of students. | $T^{40} = \begin{bmatrix} 0.4737 & 0.4737 & 0.4737 \\ 0.1579 & 0.1579 & 0.1579 \\ 0.3684 & 0.3684 & 0.3684 \end{bmatrix} \begin{matrix} S \\ M \\ O \end{matrix}$<br>Expected students for Organisation Skills = 108<br>Total number of students x 0.3684 = 108<br>Total number of students $= \dfrac{108}{0.3684} = 293$ |
| 2 Use the total student number found to calculate the expected Study Skills student number. | Expected students for Study Skills = 293 x 0.4737<br>= 139 |
| e 1 Find $R^n$ for a large value of $n$. | $R^{40} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{matrix} S \\ M \\ O \end{matrix}$ |
| 2 Read from the matrix. | In the long term, all students end up doing Organisational Skills. |

## EXERCISE 8.3 Transition matrices and long-term trends

ANSWERS p. 719

### Recap

1 OVCAA 2009 1MQ9 65% $T = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix}$ is a transition matrix and $S_3 = \begin{bmatrix} 1150 \\ 850 \end{bmatrix}$ is a state matrix.

If $S_3 = TS_2$, then $S_2$ equals

A $\begin{bmatrix} 1000 \\ 1000 \end{bmatrix}$ B $\begin{bmatrix} 1090 \\ 940 \end{bmatrix}$ C $\begin{bmatrix} 1100 \\ 900 \end{bmatrix}$ D $\begin{bmatrix} 1150 \\ 850 \end{bmatrix}$ E $\begin{bmatrix} 1175 \\ 825 \end{bmatrix}$

2 Students at a school choose one mathematics elective activity in each of three terms: Famous Problems in Mathematics (F), Great Mathematicians (G), and History of Mathematics (H). The transition matrix $T$ shows the way in which the students are expected to change their choice of elective activity from term to term. $S_n$ is the state matrix for the number of students expected to choose each elective in Term $n$.

This term

$$T = \begin{array}{c} \begin{array}{ccc} F & G & H \end{array} \\ \begin{bmatrix} 0.1 & 0.2 & 0.4 \\ 0.7 & 0.6 & 0.1 \\ 0.2 & 0.2 & 0.5 \end{bmatrix} \end{array} \begin{array}{l} F \\ G \\ H \end{array} \text{Next term, } S_2 = \begin{bmatrix} 100 \\ 160 \\ 140 \end{bmatrix} \begin{array}{l} F \\ G \\ H \end{array}$$

Which of the following calculations give us the number of students who will not change their elective from Term 2 to Term 3?

A (0.1 x100)+ (0.2 x160)+ (0.4 x140)

B (0.1 x 100)+ (0.6 x 160)+ (0.5 x 140)

C (0.1 x100)+ (0.7 x160)+ (0.2 x140)

D (0.2 x 100) + (0.2 x 160) + (0.5 x 140)

E (0.7 x100)+ (0.6 x160)+ (0.1 x140)

### Mastery

**3** S WORKED EXAMPLE 9 A fleet of buses starts each day at one of two depots, $A$ or $B$. By the end of the day, the buses end up at either of the two depots according to the following transition matrix.

This day

$$T = \begin{array}{c} \begin{array}{cc} A & B \end{array} \\ \begin{bmatrix} 0.75 & 0.45 \\ 0.25 & 0.55 \end{bmatrix} \end{array} \begin{array}{l} A \\ B \end{array} \text{Next day}$$

At the start of a particular day, there are 100 buses at depot $A$ and 120 buses at depot $B$.

a Explain why there will be an equilibrium state matrix.

b Find the equilibrium state matrix.

c How many buses will be at each depot in the long term?

d What percentage of buses are at depot $A$ in the long term? Round your answer to the nearest percentage.

**4B** WORKED EXAMPLE 10 Students at a school either buy their lunch at the canteen (C), bring it from home (H), or do not eat any lunch (N). The transition matrix shows how this changes from day to day.

This day

$$T = \begin{array}{c} \begin{array}{ccc} C & H & N \end{array} \\ \begin{bmatrix} 0.8 & 0.1 & 0.3 \\ 0.2 & 0.7 & 0.1 \\ 0 & 0.2 & 0.6 \end{bmatrix} \end{array} \begin{array}{l} C \\ H \\ N \end{array} \text{Next day}$$

a Show that there will be an equilibrium state matrix.

b What percentage of students are expected to not eat any lunch in the long term? Round your answer to the nearest percentage.

c If there are 1246 students at the school, find the equilibrium state matrix and hence find how many students are expected to bring their lunch from home in the long term.

d The school assumes the same transition matrix applies the following year. If in the long term 423 students are expected to bring their lunch from home, what would the expected number of students that will buy their lunch at the canteen?

e The school realises the following year that a different transition matrix $R$ applies. Describe the long-term trends shown by

$$R = \begin{array}{c} \text{This day} \\ \begin{array}{ccc} C & H & N \end{array} \\ \begin{bmatrix} 1 & 0.2 & 0.2 \\ 0 & 0.4 & 0.7 \\ 0 & 0.4 & 0.1 \end{bmatrix} \end{array} \begin{array}{l} C \\ H \\ N \end{array} \text{Next day}$$

## Exam practice

80-100% | 60-79% | 0-59%

5 ©VCAA | 20131MQ3 63% A coffee shop sells three types of coffee, Brazilian (B), Italian (I) and Kenyan (K). The regular customers buy one cup of coffee each per day and choose the type of coffee they buy according to the following transition matrix, T.

$$T = \begin{array}{c} \text{Choose today} \\ \begin{array}{ccc} B & I & K \end{array} \\ \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0 & 0.8 & 0.1 \\ 0.2 & 0.1 & 0.8 \end{bmatrix} \end{array} \begin{array}{l} B \\ I \\ K \end{array} \text{Choose tomorrow}$$

On a particular day, 84 customers bought Brazilian coffee, 96 bought Italian coffee and 81 bought Kenyan coffee. If these same customers continue to buy one cup of coffee each per day, the number of these customers who are expected to buy each of the three types of coffee in the long term is

| | | | | | |
|---|---|---|---|---|---|
| A Brazilian | 85 | B Brazilian | 87 | C Brazilian | 88 |
| Italian | 85 | Italian | 58 | Italian | 86 |
| Kenyan | 91 | Kenyan | 116 | Kenyan | 87 |
| D Brazilian | 89 | E Brazilian | 116 | | |
| Italian | 89 | Italian | 89 | | |
| Kenyan | 83 | Kenyan | 58 | | |

6 ©VCAA | 20101MQ7 I 60% A new colony of several hundred birds is established on a remote island. The birds can feed at two locations, $A$ and B. The birds are expected to change feeding locations each day according to the transition matrix

$$T = \begin{array}{c} \text{This day} \\ \begin{array}{cc} A & B \end{array} \\ \begin{bmatrix} 0.4 & 0.3 \\ 0.6 & 0.7 \end{bmatrix} \end{array} \begin{array}{l} A \\ B \end{array} \text{Next day}$$

In the beginning, approximately equal numbers of birds feed at each site each day. Which of the following statements is not true?

A 70% of the birds that feed at B on a given day will feed at B the next day.

B 60% of the birds that feed at A on a given day will feed at B the next day.

C In the long term, more birds will feed at B than at $A$.

D The number of birds that change feeding locations each day will decrease over time to zero.

E In the long term, some birds will always be found feeding at each location.

**7** ©VCAA 2008 1MQ8 MODIFIED 57% A large population of mutton birds migrates each year to a remote island to nest and breed. There are four nesting sites on the island, A, B, C and *D*. Researchers suggest that the following transition matrix can be used to predict the number of mutton birds nesting at each of the four sites in subsequent years.

$$\begin{array}{c} \text{This year} \\ \begin{array}{cccc} A & B & c & D \end{array} \end{array}$$

$$T = \begin{bmatrix} 0.4 & 0 & 0.2 & 0 \\ 0.35 & 1 & 0.15 & 0 \\ 0.15 & 0 & 0.55 & 0 \\ 0.1 & 0 & 0.1 & 1 \end{bmatrix} \begin{matrix} A \\ D \\ C \\ D \end{matrix} \text{ Next year}$$

This transition matrix predicts that, in the long term, the mutton birds will

A nest only at site *A*.

B nest only at site B.

C nest only at sites *A* and C.

D nest only at sites B and *D*.

E continue to nest at all four sites.

8 ©VCAA 20161MQ7 I 52% Each week, the 300 students at a primary school choose art (A), music (M) or sport (S) as an afternoon activity. The transition matrix shows how the students' choices change from week to week.

$$\begin{array}{c} \text{This week} \\ \begin{array}{ccc} A & M & S \end{array} \end{array}$$

$$T = \begin{bmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.4 & 0.4 \\ 0.2 & 0.2 & 0.5 \end{bmatrix} \begin{matrix} A \\ M \\ S \end{matrix} \text{ Next week}$$

Based on the information, it can be concluded that, in the long term

A no student will choose sport.

B all students will choose to stay in the same activity each week.

C all students will have chosen to change their activity at least once.

D more students will choose to do music than sport.

E the number of students choosing to do art and music will be the same.

9 ©VCAA J 20181MQFI 46% A public library organised 500 of its members into five categories according to the number of books each member borrows each month. These categories are

*J* = no books borrowed per month

*K* = one book borrowed per month

*L* = two books borrowed per month

*M* = three books borrowed per month

*N* = four or more books borrowed per month

The transition matrix, T, shows how the number of books borrowed per month by the members is expected to change from month to month.

$$\begin{array}{c} \text{This month} \\ \begin{array}{ccccc} J & K & L & M & N \end{array} \end{array}$$

$$T = \begin{bmatrix} 0.1 & 0.2 & 0.2 & 0 & 0 \\ 0.5 & 0.2 & 0.3 & 0.1 & 0 \\ 0.3 & 0.3 & 0.4 & 0.1 & 0.2 \\ 0.1 & 0.2 & 0.1 & 0.6 & 0.3 \\ 0 & 0.1 & 0 & 0.2 & 0.5 \end{bmatrix} \begin{matrix} / \\ K \\ L \\ M \\ N \end{matrix} \text{ Next month}$$

In the long term, which category is expected to have approximately 96 members each month?

*A J* B *K* *CL* *DM* *E N*

10 ©VCAA 2017N1MQ6 I Students at a school choose an afternoon activity every week. They can choose either sport (S), art (A) or music *(M)*. The table shows the number of students who chose sport, art and music in Week 1 of the school term.

| Sport (S) | Art (A) | Music *(M)* |
|---|---|---|
| 150 | 85 | 35 |

The students are expected to change the activity they choose from week to week as shown in the transition matrix *P*.

$$T = \begin{array}{c} \text{This week} \\ \begin{array}{ccc} S & A & M \end{array} \\ \left[\begin{array}{ccc} 0.80 & 0.20 & 0.05 \\ 0.10 & 0.70 & 0.15 \\ 0.10 & 0.10 & 0.80 \end{array}\right] \end{array} \begin{array}{l} S \\ A \text{ Next week} \\ M \end{array}$$

Which one of the following statements is true for this situation?

A 30% of the students will never choose art.

B The number of students who choose music will decrease every week.

C The number of students who choose sport in one week will always be 20% less than the number of students who chose sport in the previous week.

D In Week 3 of the school term, the number of students who choose music will be less than half the number of students who choose art.

E In the long term, the number of students who choose music will be more than the number of students who choose art.

11 ©VCAA 20191MQ6 39% A water park is open from 9 am until 5 pm. There are three activities, the pool (P), the slide (S) and the water jets (W), at the water park. Children have been found to change their activity at the water park each half hour, as shown in the transition matrix, T, below.

$$T = \begin{array}{c} \text{This half hour} \\ \begin{array}{ccc} P & S & W \end{array} \\ \left[\begin{array}{ccc} 0.80 & 0.20 & 0.40 \\ 0.05 & 0.60 & 0.10 \\ 0.15 & 0.20 & 0.50 \end{array}\right] \end{array} \begin{array}{l} P \\ S \text{ Next half hour} \\ W \end{array}$$

A group of children has come to the water park for the whole day. The percentage of these children who are expected to be at the slide (S) at closing time is closest to

A 14% B 20% C 24% D 25% E 62%

12 ©VCAA 2019NIMQ8^ Sheep on the farm are also free to move between the four water sources. The change in the number of sheep at each water source from week to week is shown in matrix *T* below.

$$T = \begin{array}{c} \text{This week} \\ \begin{array}{cccc} P & Q & R & S \end{array} \\ \left[\begin{array}{cccc} 0.4 & 0.3 & 0.2 & 0.1 \\ 0.2 & 0.1 & 0.5 & 0.3 \\ 0.1 & 0.3 & 0.1 & 0.2 \\ 0.3 & 0.3 & 0.2 & 0.4 \end{array}\right] \end{array} \begin{array}{l} P \\ Q \\ R \\ S \end{array} \text{ Next week}$$

In the long term, 635 sheep are expected to be at S each week. In the long term, the number of sheep expected to be at *Q* each week is closest to

A 371 B 493 C 527 D 607 E 635

**13** ©VCAA 2006 2MQ2 (7 marks) A new shopping centre called Shopper Heaven (S) is about to open. It will compete for customers with East own (E) and Noxland (N). Market research suggests that each shopping centre will have a regular customer base, but attract and lose customers on a weekly basis as follows.

- 80% of Shopper Heaven customers will return to Shopper Heaven next week
- 12% of Shopper Heaven customers will shop at Eastown next week
- 8% of Shopper Heaven customers will shop at Noxland next week
- 76% of Eastown customers will return to Eastown next week
- 9% of Eastown customers will shop at Shopper Heaven next week
- 15% of Eastown customers will shop at Noxland next week
- 85% of Noxland customers will return to Noxland next week
- 10% of Noxland customers will shop at Shopper Heaven next week
- 5% of Noxland customers will shop at Eastown next week

a 78% Use this information to copy and complete this transition matrix $T$ (express percentages as proportions, for example, write 76% as 0.76). 2 marks

This week

$$T = \begin{bmatrix} & S & E & N \\ & & & \\ & & & \\ & & & \end{bmatrix} \begin{matrix} \\ S \\ E \\ N \end{matrix} \text{ Next week}$$

During the week that Shopper Heaven opened, it had 300 000 customers. In the same week, Eastown had 120000 customers and Noxland had 180000 customers.

b 78% Use this information to copy and complete this column matrix, $K_o$. 1 mark

$$K_0 = \begin{bmatrix} \\ \\ \\ \end{bmatrix} \begin{matrix} S \\ E \\ N \end{matrix}$$

c 46% Use $T$ and $K_o$ to write and evaluate a matrix product that determines the number of customers expected at each of the shopping centres during the following week, 2 marks

d 46% Show, by calculating at least two appropriate state matrices, that, in the long term, the number of customers expected at each centre each week is given by the matrix 2 marks

$$K = \begin{bmatrix} 194983 \\ 150513 \\ 254504 \end{bmatrix}$$

8.3

14 ©VCAA 2020 2MQ3 I (4 marks) The three major shopping centres in a large city, Eastmall (E), Grandmall (G) and Westmall (*W*), are owned by the same company. An offer to buy the Westmall shopping centre was made by a competitor. A market research project suggested that if the Westmall shopping centre were sold, each of the three centres (Westmall, Grandmall and Eastmall) would continue to have regular shoppers but would attract and lose shoppers on a weekly basis.

Let $S_n$ be the state matrix that shows the expected number of shoppers at each of the three centres *n* weeks after Westmall is sold. A matrix recurrence relation that generates values of $S_n$ is

$S_{w+}$i = *T* x S,,

where

This week

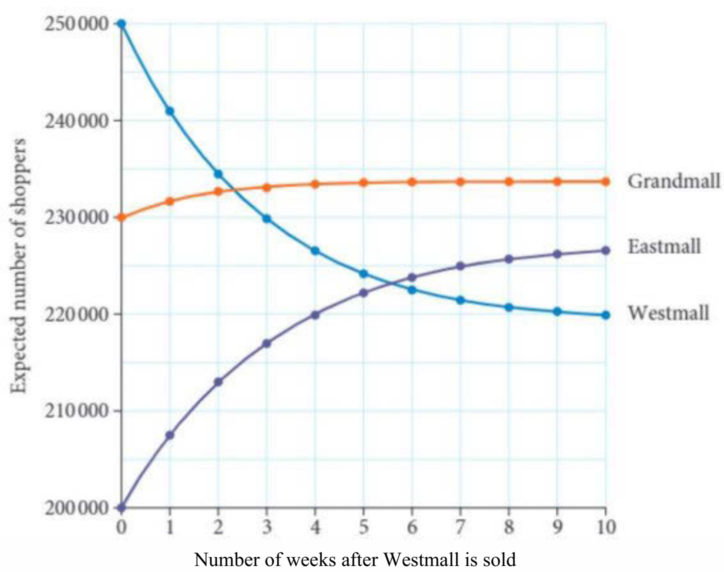
*W G E*

$$T = \begin{bmatrix} 0.80 & 0.09 & 0.10 \\ 0.12 & 0.79 & 0.10 \\ 0.08 & 0.12 & 0.80 \end{bmatrix} \begin{matrix} W \\ G \\ £ \end{matrix} \text{ Next week, } S_o - \begin{bmatrix} 250\,000 \\ 230\,000 \\ 200\,000 \end{bmatrix} \begin{matrix} W \\ G \\ E \end{matrix}$$

a 79% Calculate the state matrix, to show the expected number of shoppers at each of the three centres one week after Westmall is sold. 1 mark

$$S_1 = \begin{bmatrix} \_\_\_\_\_\_ \\ \_\_\_\_\_\_ \\ \_\_\_\_\_\_ \end{bmatrix} \begin{matrix} W \\ G \\ E \end{matrix}$$

Using values from the recurrence relation, the graph below shows the expected number of shoppers at Westmall, Grandmall and Eastmall for each of the 10 weeks after Westmall is sold.




b 64% What is the difference in the expected weekly number of shoppers at Westmall from the time Westmall is sold to 10 weeks after Westmall is sold? Give your answer correct to the nearest thousand. 1 mark

c 27% Grandmall is expected to achieve its maximum number of shoppers sometime between the fourth and the tenth week after Westmall is sold. Write down the week number in which this is expected to occur. 1 mark

d 39% In the long term, what is the expected weekly number of shoppers at Westmall? Round your answer to the nearest whole number. 1 mark

# 11M Transition matrices with restocking and culling

Video playlist Transition matrices with restocking and culling

## Restocking and culling recurrence relations

In the transitions we have dealt with so far, the population involved (shoppers, birds, trucks, etc.) moves between states but the total number in the population remains the same throughout each transition. We are now going to look at transitions where things are added or taken away. Adding something to a transition is called restocking and taking something away is called culling.

For example, the following shows a prediction of the transition of 1640 people between casual (C) and permanent (P) employment in an industry over two years.

$$S_0 = \begin{bmatrix} 1200 \\ 440 \end{bmatrix} \begin{matrix} C \\ P \end{matrix} \qquad Sj = \begin{bmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{bmatrix} \begin{bmatrix} 1200 \\ 440 \end{bmatrix} = \begin{bmatrix} 1004 \\ 636 \end{bmatrix} \begin{matrix} C \\ P \end{matrix}$$

In reality, people enter or leave the industry each year. If we know that each year 24 people enter the industry to work as casuals and 35 permanent employees leave the industry, then we need to add an extra matrix to the calculation:

$$S_0 = \begin{bmatrix} 1200 \\ 440 \end{bmatrix} \begin{matrix} C \\ P \end{matrix} \qquad S_1 = \begin{bmatrix} 0.8 & 0.1 \\ 0.2 & 0.9 \end{bmatrix} \begin{bmatrix} 1200 \\ 440 \end{bmatrix} + \begin{bmatrix} 24 \\ -35 \end{bmatrix} = \begin{bmatrix} 1004 \\ 636 \end{bmatrix} + \begin{bmatrix} 24 \\ -35 \end{bmatrix} = \begin{bmatrix} 1028 \\ 601 \end{bmatrix} \begin{matrix} C \\ P \end{matrix}$$

A positive element means restocking.

A negative element means culling.

With restocking and culling, the sum of the elements in the state matrices changes. The sum of the elements in Sj now equals 1629 while the sum of the elements in $S_0$ is 1640.

**© Exam hack**

Restocking and culling questions sometimes use years in the recurrence relations, so $S_{2025}$ represents the state matrix in the year 2025, $S_{2026}$ represents the state matrix in the year 2026, and so on.

**Restocking and culling recurrence relations**

The restocking and culling recurrence relation that generates a sequence of state matrices $S_n$ is

$S_0$ = initial state matrix, $S_{n+1} = TS_n + B$

where

$T$ is a transition matrix

$n$ is the number of transitions

restocking is adding to the population and culling is removing from the population

$B$ can have both positive elements and negative elements and has the same order as $S_n$.

**Exam hack**

There is no short-cut rule for restocking and culling. The recurrence relation *always* needs to be used to answer these questions.

## WORKED EXAMPLE 11 Restocking and culling recurrence relations

A karate school regularly assesses its students as they progress through three belts. After each assessment, students are classified as white belt (orange belt (O), blue belt (B) or left the school (L). Students cannot be assessed at a belt level lower than their current belt. In the recurrence relation shown

- the initial state matrix $S_o$ shows the number of students at each belt level immediately before the first assessment of the year
- the transition matrix *T* contains the percentages of students who are expected to change their belt or leave the school after each assessment
- the matrix *J* contains the number of students who join the karate school after each assessment.

$$S_0 = \begin{bmatrix} 80 \\ 60 \\ 40 \\ 0 \end{bmatrix} \begin{matrix} W \\ O \\ B \\ L \end{matrix}, \quad S_{n+1} = TS_n + J$$

where

$$T = \overset{\text{Before assessment}}{\begin{matrix} \quad W & \quad O & \quad B & \quad L \end{matrix}} \atop \begin{bmatrix} 0.25 & 0 & 0 & 0 \\ 0.45 & 0.65 & 0 & 0 \\ 0 & 0.2 & 0.7 & 0 \\ 0.3 & 0.15 & 0.3 & 1 \end{bmatrix} \begin{matrix} W \\ O \\ B \\ L \end{matrix} \text{ After assessment} \quad \text{and} \quad 7 = \begin{bmatrix} 16 \\ 3 \\ 4 \\ 0 \end{bmatrix} \begin{matrix} W \\ O \\ B \\ L \end{matrix}$$

**Steps Working**

a Find the number of students with a white belt after the first assessment.

1 Use the recurrence relation

$S_o$ = initial state matrix, $S_{,,+I} = TS_n + J$

to find Sj.

$Si\text{-}TS_{0+} J$

$$S_1 = \begin{bmatrix} 0.25 & 0 & 0 & 0 \\ 0.45 & 0.65 & 0 & 0 \\ 0 & 0.2 & 0.7 & 0 \\ 0.3 & 0.15 & 0.3 & 1 \end{bmatrix} \begin{bmatrix} 80 \\ 60 \\ 40 \\ 0 \end{bmatrix} + \begin{bmatrix} 16 \\ 3 \\ 4 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 20 \\ 75 \\ 40 \\ 45 \end{bmatrix} + \begin{bmatrix} 16 \\ 3 \\ 4 \\ 0 \end{bmatrix} = \begin{bmatrix} 36 \\ 78 \\ 44 \\ 45 \end{bmatrix} \begin{matrix} W \\ O \\ B \\ L \end{matrix}$$

2 Find the answer from $S_r$

There are 36 students with a white belt after the first assessment.

b Find the number of students with a blue belt after the second assessment.

1 Use the recurrence relation to find $S_2$.

$S_2 = TSj + J$

$$S_2 = \begin{bmatrix} 0.25 & 0 & 0 & 0 \\ 0.45 & 0.65 & 0 & 0 \\ 0 & 0.2 & 0.7 & 0 \\ 0.3 & 0.15 & 0.3 & 1 \end{bmatrix} \begin{bmatrix} 36 \\ 78 \\ 44 \\ 45 \end{bmatrix} + \begin{bmatrix} 16 \\ 3 \\ 4 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 9 \\ 67 \\ 46 \\ 81 \end{bmatrix} + \begin{bmatrix} 16 \\ 3 \\ 4 \\ 0 \end{bmatrix} = \begin{bmatrix} 25 \\ 70 \\ 50 \\ 81 \end{bmatrix} \begin{matrix} W \\ O \\ B \\ L \end{matrix}$$

2 Find the answer from $S_2$.

There are 50 students with a blue belt after the second assessment.

### Exam hack

If you find yourself raising the transition matrix to a power in a restocking and culling question, you are making a mistake.

8.4

WB p. 167

WB

p. 168

**WORKED EXAMPLE 12** Restocking and culling with a constant total

A company wishes to start a commercial whiting breeding farm and it starts with 20 000 eggs, 4000 young whiting, and 500 adult whiting. The following could happen in a year:

- eggs ($E$) could die ($D$) or they could live and become young whiting ($Y$)
- young whiting ($Y$) could die ($D$) or they could live and become adult whiting ($A$)
- adult whiting ($A$) could die ($D$) or they could live for a while but will eventually die.

From one year to the next, this situation can be represented by the recurrence relation

$$S_0 = \begin{bmatrix} 20\,000 \\ 4\,000 \\ 500 \\ 0 \end{bmatrix} \begin{matrix} E \\ Y \\ A \\ D \end{matrix}, \quad S_{n+1} = TS_n + W$$

where

$$\begin{matrix} & \text{This year} & & & & \\ & E & Y & A & D & \\ T = & \left[\begin{matrix} 0 \\ 0.3 \\ 0 \\ 0.7 \end{matrix}\right. & \begin{matrix} 0 \\ 0 \\ 0.35 \\ 0.65 \end{matrix} & \begin{matrix} 0 \\ 0 \\ 0.6 \\ 0.4 \end{matrix} & \left.\begin{matrix} 0 \\ 0 \\ 0 \\ 1 \end{matrix}\right] & \begin{matrix} E \\ Y \\ A \\ D \end{matrix} \text{ Next year} \end{matrix}$$

The company wants to buy or sell enough young and adult whiting so that their numbers in the farm remain the same each year. The company also assumes that the female adult whiting will lay 20 000 eggs each year and that all the dead fish and eggs are removed. How many young and adult whiting will the company buy or sell each year?

| Steps | Working |
|---|---|
| **1** The state matrices will remain the same each year. | $S_1 = S_0 = \begin{bmatrix} 20\,000 \\ 4000 \\ 500 \\ 0 \end{bmatrix} \begin{matrix} E \\ Y \\ A \\ D \end{matrix}$ |
| **2** Substitute the known matrices into the recurrence relation. | $S_1 = TS_0 + W$ <br> $\begin{bmatrix} 20\,000 \\ 4000 \\ 500 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0.3 & 0 & 0 & 0 \\ 0 & 0.35 & 0.6 & 0 \\ 0.7 & 0.65 & 0.4 & 1 \end{bmatrix} \begin{bmatrix} 20\,000 \\ 4000 \\ 500 \\ 0 \end{bmatrix} + W$ |
| **3** Solve for $W$. | $\begin{bmatrix} 20\,000 \\ 4000 \\ 500 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 6000 \\ 1700 \\ 16\,800 \end{bmatrix} + W$ <br> $W = \begin{bmatrix} 20\,000 \\ 4000 \\ 500 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 6000 \\ 1700 \\ 16\,800 \end{bmatrix}$ <br> $= \begin{bmatrix} 20\,000 \\ -2000 \\ -1200 \\ -16\,800 \end{bmatrix} \begin{matrix} E \\ Y \\ A \\ D \end{matrix}$ |
| **4** Find the answer from $W$. | Each year the company will sell 2000 young whiting and 1200 adult whiting. |



9780170448406

## WORKED EXAMPLE 13 Solving for the restocking/culling matrix

Pacific Travel analyses the choices their customers make between four main Pacific Island destinations: Fiji (F), Hawaii (H), Tahiti (T) and Vanuatu (V). $S_n$ shows the number of customers who choose each destination in year $n$. $S_{2024}$ and $S_{2025}$ are given below. The matrix $T$ shows how customers are expected to change their choice from year to year.

$$S_{2024} = \begin{bmatrix} 440 \\ 320 \\ 110 \\ 80 \end{bmatrix} \begin{matrix} F \\ H \\ T \\ V \end{matrix} \qquad S_{2025} = \begin{bmatrix} 416 \\ 310 \\ 122 \\ 282 \end{bmatrix} \begin{matrix} F \\ H \\ T \\ V \end{matrix} \qquad T = \overset{\text{This year}}{\overset{\begin{matrix} F & \;\;H & \;\;T & \;\;V \end{matrix}}{\begin{bmatrix} 0.6 & 0.1 & 0.3 & 0.5 \\ 0.25 & 0.45 & 0.1 & 0.1 \\ 0.1 & 0.15 & 0.2 & 0.25 \\ 0.05 & 0.3 & 0.4 & 0.15 \end{bmatrix}}} \begin{matrix} F \\ H \\ T \\ V \end{matrix} \text{ Next year}$$

The destinations chosen can be modelled by $S_{M+1} = TS_n + B$.

a FindB.

b Explain the meaning of the negative element in B.

c Find $S_{2026}$.

**Steps** | **Working**

a 1 Write the recurrence relation in terms of the matrices given in the question.

$S_{n+1} = TS_n + B$

$S_{2025} = TS_{2024} + B$

$$\begin{bmatrix} 416 \\ 310 \\ 122 \\ 282 \end{bmatrix} = \begin{bmatrix} 0.6 & 0.1 & 0.3 & 0.5 \\ 0.25 & 0.45 & 0.1 & 0.1 \\ 0.1 & 0.15 & 0.2 & 0.25 \\ 0.05 & 0.3 & 0.4 & 0.15 \end{bmatrix} \begin{bmatrix} 440 \\ 320 \\ 110 \\ 80 \end{bmatrix} + B$$

2 Solve for $B$.

$$\begin{bmatrix} 416 \\ 310 \\ 122 \\ 282 \end{bmatrix} = \begin{bmatrix} 369 \\ 273 \\ 134 \\ 174 \end{bmatrix} + B$$

$$B = \begin{bmatrix} 416 \\ 310 \\ 122 \\ 282 \end{bmatrix} - \begin{bmatrix} 369 \\ 273 \\ 134 \\ 174 \end{bmatrix}$$

$$= \begin{bmatrix} 47 \\ 37 \\ -12 \\ 108 \end{bmatrix} \begin{matrix} F \\ H \\ T \\ V \end{matrix}$$

b A negative means removing something.

12 customers who travelled to Tahiti will be removed from the analysis.

c Use the restocking/culling recurrence relation. $S_{n+1} = TS_n + B$

$S_{2026} = TS_{2025} + B$

$$= \begin{bmatrix} 0.6 & 0.1 & 0.3 & 0.5 \\ 0.25 & 0.45 & 0.1 & 0.1 \\ 0.1 & 0.15 & 0.2 & 0.25 \\ 0.05 & 0.3 & 0.4 & 0.15 \end{bmatrix} \begin{bmatrix} 416 \\ 310 \\ 122 \\ 282 \end{bmatrix} + \begin{bmatrix} 47 \\ 37 \\ -12 \\ 108 \end{bmatrix}$$

$$= \begin{bmatrix} 458 \\ 284 \\ 183 \\ 205 \end{bmatrix} + \begin{bmatrix} 47 \\ 37 \\ -12 \\ 108 \end{bmatrix}$$

$$= \begin{bmatrix} 505 \\ 321 \\ 171 \\ 313 \end{bmatrix} \begin{matrix} F \\ H \\ T \\ V \end{matrix}$$

8.4

WB

p. 169

## EXERCISE 8.4 Transition matrices with restocking and culling

ANSWERS p. 720

**Recap**

1 ©VCAA | 20141MQ3 68% Regular customers at a hairdressing salon can choose to have their hair cut by Shirley, Jen or Narj. The salon has 600 regular customers who get their hair cut each month. In June, 200 customers chose Shirley (S) to cut their hair, 200 chose Jen (J) to cut their hair and 200 chose Narj *(N)* to cut their hair. The regular customers* choice of hairdresser is expected to change from month to month as shown in the transition matrix, T, below.

$$\begin{array}{c} \text{This month} \\ \begin{array}{ccc} S & 7 & N \end{array} \end{array}$$

$$T = \begin{bmatrix} 0.75 & 0.10 & 0.10 \\ 0.10 & 0.75 & 0.15 \\ 0.15 & 0.15 & 0.75 \end{bmatrix} \begin{matrix} S \\ J \\ N \end{matrix} \text{ Next month}$$

In the long term, the number of regular customers who are expected to choose Shirley is closest to

A 150 B 170 C 185 D 195 E 200

2 ©VCAA 20111MQ7 58% Each night, a large group of mountain goats sleep at one of two locations, *A* or *B*. On the first night, equal numbers of goats are observed to be sleeping at each location. From night to night, goats change their sleeping locations according to a transition matrix *T*. It is expected that, in the long term, more goats will sleep at location *A* than at location *B*. Assuming that the total number of goats remains constant, a transition matrix *T* that would predict this outcome is

A $T- \begin{array}{c}\text{This night} \\ \begin{array}{cc} A & B \end{array} \\ \begin{bmatrix} 0.8 & 0.4 \\ 0.2 & 0.6 \end{bmatrix}\end{array} \begin{matrix} A \\ B \end{matrix} \text{ Next night}$

B $T = \begin{array}{c}\text{This night} \\ \begin{array}{cc} A & B \end{array} \\ \begin{bmatrix} 0.7 & 0.1 \\ 0.3 & 0.9 \end{bmatrix}\end{array} \begin{matrix} A \\ B \end{matrix} \text{ Next night}$

C $T- \begin{array}{c}\text{This night} \\ \begin{array}{cc} A & B \end{array} \\ \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}\end{array} \begin{matrix} A \\ B \end{matrix} \text{ Next night}$

D $T = \begin{array}{c}\text{This night} \\ \begin{array}{cc} A & B \end{array} \\ \begin{bmatrix} 0.6 & 0.2 \\ 0.4 & 0.8 \end{bmatrix}\end{array} \begin{matrix} A \\ B \end{matrix} \text{ Next night}$

E $T = \begin{array}{c}\text{This night} \\ \begin{array}{cc} A & B \end{array} \\ \begin{bmatrix} 0.1 & 0.8 \\ 0.9 & 0.2 \end{bmatrix}\end{array} \begin{matrix} A \\ B \end{matrix} \text{ Next night}$



9780170448406

## Mastery

**30** WORKED EXAMPLE 11 A swim school regularly assesses its students as they progress through three levels. After each assessment, students are classified as Minky *(M),* Flipper (F), Orca (O) or left the school (L). Students cannot be assessed at a level lower than their current level. In the recurrence relation shown

- the initial state matrix $S_0$ shows the number of students at each level immediately before the first assessment of the year
- the transition matrix $T$ contains the percentages of students who are expected to change their level or leave the school after each assessment
- the matrix $J$ contains the number of students who join the swim school after each assessment.

$$S_0 = \begin{bmatrix} 75 \\ 50 \\ 20 \\ 0 \end{bmatrix} \begin{matrix} M \\ F \\ O \\ L \end{matrix}, \quad S_{n+1} = TS_n + J$$

where

$$T = \overset{\text{Before assessment}}{\begin{matrix} & M & F & O & L \\ & \end{matrix}}$$

$$T = \begin{bmatrix} 0.4 & 0 & 0 & 0 \\ 0.35 & 0.6 & 0 & 0 \\ 0 & 0.2 & 0.8 & 0 \\ 0.25 & 0.2 & 0.2 & 1 \end{bmatrix} \begin{matrix} M \\ F \\ O \\ L \end{matrix} \text{ After assessment} \quad \text{and} \quad J = \begin{bmatrix} 12 \\ 7 \\ 6 \\ 0 \end{bmatrix} \begin{matrix} M \\ F \\ O \\ L \end{matrix}$$

a Find the number of students at Minky level after the first assessment,

b Find the number of students at Flipper level after the second assessment.

**40** WORKED EXAMPLE 12 A cooperative wishes to start a commercial salmon breeding farm and it starts with 15000 salmon eggs, 5000 young salmon, and 600 adult salmon. The following could happen in a year:

- eggs (F) could die (D) or they could live and become young salmon (Y)
- young salmon (Y) could die (D) or they could live and become adult salmon (A)
- adult salmon (A) could die (D) or they could live for a while but will eventually die.

From one year to the next, this situation can be represented by the recurrence relation

$$S_0 = \begin{bmatrix} 15000 \\ 5\,000 \\ 600 \\ 0 \end{bmatrix}, \quad S_{n+1} = TS_n + B$$

where

This year

$$T = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0.25 & 0 & 0 & 0 \\ 0 & 0.4 & 0.7 & 0 \\ 0.75 & 0.6 & 0.3 & 1 \end{bmatrix} \begin{matrix} E \\ Y \\ A \\ D \end{matrix} \text{ Next year}$$

(Columns of $T$, this year: $E$, $Y$, $A$, $D$.)

The cooperative wants to buy or sell enough young and adult salmon so that their numbers in the farm remain the same each year. The cooperative also assumes that the female adult salmon will lay 15000 eggs each year and that all the dead fish and eggs are removed. How many young and adult salmon will the cooperative buy or sell each year?

50 WORKED EXAMPLE 13 GoAsia Travel analyses the choices their customers make between four main South-east Asian destinations: Indonesia (f), Singapore (S), Thailand (T) and Vietnam (V). $S_n$ shows the number of customers who choose each destination in year $n$. $S_{2022}$ a^d $2023^{are} given below.
The matrix $T$ shows how customers are expected to change their choice from year to year.

$$S_{2022} = \begin{bmatrix} 530 \\ 460 \\ 320 \\ 90 \end{bmatrix} \begin{matrix} I \\ S \\ T \\ V \end{matrix} \quad S_{2023} = \begin{bmatrix} 590 \\ 386 \\ 294 \\ 142 \end{bmatrix} \begin{matrix} I \\ S \\ T \\ V \end{matrix} \quad T = \overset{\text{This year}}{\overset{I \quad S \quad T \quad V}{\begin{bmatrix} 0.55 & 0.2 & 0.4 & 0.45 \\ 0.2 & 0.4 & 0.1 & 0.1 \\ 0.15 & 0.1 & 0.3 & 0.35 \\ 0.1 & 0.3 & 0.2 & 0.1 \end{bmatrix}}} \begin{matrix} I \\ S \\ T \\ V \end{matrix} \text{ Next year}$$

The destinations chosen can be modelled by

$$S_{n+1} = TS_n + B$$

a Find $B$.

b Explain the meaning of the negative element in $B$.

c Find $S_{2024}$.

## Exam practice

80-100% 60-79% 0-59%

6 ©VCAA 2016S1MQ5 The matrix $S_{n+1}$ is determined from the matrix $S_n$ using the rule

$$S_{n+1} = TS_n - C,$$

where T, $S_0$ and C are defined as follows.

$$T = \begin{bmatrix} 0.5 & 0.6 \\ 0.5 & 0.4 \end{bmatrix}, S_0 = \begin{bmatrix} 100 \\ 250 \end{bmatrix} \text{ and } C = \begin{bmatrix} 20 \\ 20 \end{bmatrix}$$

Given this information, the matrix $S_2$ equals

A $\begin{bmatrix} 100 \\ 250 \end{bmatrix}$ B $\begin{bmatrix} 148 \\ 122 \end{bmatrix}$ C $\begin{bmatrix} 170 \\ 140 \end{bmatrix}$ D $\begin{bmatrix} 180 \\ 130 \end{bmatrix}$ E $\begin{bmatrix} 190 \\ 160 \end{bmatrix}$

7 ©VCAA 20171MQ7 36% At a fish farm:

- young fish (Y) may eventually grow into juveniles (J) or they may die (D)
- juveniles (J) may eventually grow into adults (A) or they may die (D)
- adults (A) eventually die (D).

The initial state of this population, $F_0$, is shown.

$$F_0 = \begin{bmatrix} 50\,000 \\ 10\,000 \\ 7\,000 \\ 0 \end{bmatrix} \begin{matrix} Y \\ J \\ A \\ D \end{matrix}$$

Every month, fish are either sold or bought so that the number of young, juvenile and adult fish in the farm remains constant. The population of fish in the fish farm after $n$ months, $F_n$, can be determined by the recurrence rule

$$F_{n+1} = \begin{bmatrix} 0.65 & 0 & 0 & 0 \\ 0.25 & 0.75 & 0 & 0 \\ 0 & 0.20 & 0.95 & 0 \\ 0.10 & 0.05 & 0.05 & 1 \end{bmatrix} F_n + B$$

where $B$ is a column matrix that shows the number of young, juvenile and adult fish bought or sold each month and the number of dead fish that are removed. Each month, the fish farm will

A sell 1650 adult fish.
B buy 1750 adult fish.
C sell 17 500 young fish.
D buy 50000 young fish.
E buy 10000 juvenile fish.


9780170448406

▶ 8 ©VCAA | 2020 2MQ4a 50% (1 mark) The three major shopping centres in a large city, Eastmall (E), Grandmall (G) and Westmall (W), are owned by the same company. An offer to buy the Westmall shopping centre was made by a competitor. A market research project suggested that if the Westmall shopping centre were sold, each of the three centres (Westmall, Grandmall and Eastmall) would continue to have regular shoppers but would attract and lose shoppers on a weekly basis. Let $R_n$ be the state matrix that shows the expected number of shoppers at each of the three centres $n$ weeks after Westmall is sold.

A matrix recurrence relation that generates values of $R_n$ is

^n+l ~ $TR_n$ + ®

where

$$
\begin{array}{c c}
 & \text{This week} \\
 & W \quad G \quad E \\
T = & \begin{bmatrix} 0.78 & 0.13 & 0.10 \\ 0.12 & 0.82 & 0.10 \\ 0.10 & 0.05 & 0.80 \end{bmatrix}
\end{array}
\begin{array}{l} W \\ G \\ E \end{array} \text{Next week}, \quad B = \begin{bmatrix} -400 \\ 700 \\ 500 \end{bmatrix} \begin{array}{l} W \\ G \\ E \end{array}
$$

The matrix $R_2$ is the state matrix that shows the expected number of shoppers at each of the three centres in the second week after Westmall is sold.

$$
R_2 = \begin{bmatrix} 239060 \\ 250\,840 \\ 192\,900 \end{bmatrix} \begin{array}{l} W \\ G \\ E \end{array}
$$

Determine the expected number of shoppers at Westmall in the third week after it is sold.

9 ©VCAA | 2016 2MQ3abce , (6 marks) A travel company is studying the choice between air (A), land (L), sea (S) or no (N) travel by some of its customers each year. Matrix $T$ shown contains the percentages of customers who are expected to change their choice of travel from year to year.

$$
\begin{array}{c c}
 & \text{This year} \\
 & A \quad L \quad S \quad N \\
T = & \begin{bmatrix} 0.65 & 0.25 & 0.25 & 0.50 \\ 0.15 & 0.60 & 0.20 & 0.15 \\ 0.05 & 0.10 & 0.25 & 0.20 \\ 0.15 & 0.05 & 0.30 & 0.15 \end{bmatrix}
\end{array}
\begin{array}{l} A \\ L \\ S \\ N \end{array} \text{Next year}
$$

Let $S_n$ be the matrix that shows the number of customers who choose each type of travel $n$ years after 2014. Matrix $S_0$ shows the number of customers who chose each type of travel in 2014.

$$
S_0 = \begin{bmatrix} 520 \\ 320 \\ 80 \\ 80 \end{bmatrix} \begin{array}{l} A \\ L \\ S \\ N \end{array}
$$

Matrix Si shows the number of customers who chose each type of travel in 2015.

$$
S_1 = TS_0 = \begin{bmatrix} 478 \\ d \\ e \\ f \end{bmatrix} \begin{array}{l} A \\ L \\ S \\ N \end{array}
$$

a 85% What are the values d, e,/missing from matrix Sj? 1 mark

b 35% Write a calculation that shows that 478 customers were expected to choose air travel in 2015. 1 mark

c 45% Consider the customers who chose sea travel in 2014. How many of these customers were expected to choose sea travel in 2015? 1 mark ▶

In 2016, the number of customers studied was increased to 1360. Matrix $R_2$oi6 shown contains the number of these customers who chose each type of travel in 2016.

$$R_{2016} = \begin{bmatrix} 646 \\ 465 \\ 164 \\ 85 \end{bmatrix} \begin{matrix} A \\ L \\ S \\ N \end{matrix}$$

The company intends to increase the number of customers in the study in 2017 and in 2018. The matrix that contains the number of customers who are expected to choose each type of travel in 2017 ($R_{2017}$) an<^ 2018 ($R_{2018}$) can he determined using the matrix equations shown.

$R_{k2017}$ " ^2016 + $B$

$R_{L2018}$ = ^2017 + $B$

This year

$$T = \begin{matrix} & A & L & S & N & \\ & \begin{bmatrix} 0.65 & 0.25 & 0.25 & 0.50 \\ 0.15 & 0.60 & 0.20 & 0.15 \\ 0.05 & 0.10 & 0.25 & 0.20 \\ 0.15 & 0.05 & 0.30 & 0.15 \end{bmatrix} & \begin{matrix} A \\ L \\ S \\ N \end{matrix} \end{matrix} \text{Next year} \qquad B = \begin{bmatrix} 80 \\ 80 \\ 40 \\ -80 \end{bmatrix} \begin{matrix} A \\ L \\ S \\ N \end{matrix}$$

*d* i 14% The element in the fourth row of matrix *B* is -80. Explain this number in the context of selecting customers for the studies in 2017 and 2018. 1 mark

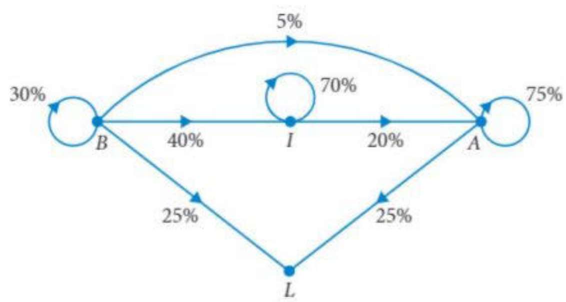
ii 51% Determine the number of customers who are expected to choose sea travel in 2018. Round your answer to the nearest whole number. 2 marks

10 ©VCAA 2015 2MQ3 (7 marks) The ability level of students in a music school is assessed regularly. Students cannot be assessed at a level that is lower than their current level. After each assessment, students are classified as beginner (B), intermediate (Z), advanced (A) or left the school (L). Let matrix $T_2$ be the transition matrix for this model. Matrix $T_2$ contains the percentages of students who are expected to change their ability level or leave the school after each assessment.

Before assessment

$$T_2 = \begin{matrix} & B & I & A & L & \\ & \begin{bmatrix} 0.30 & 0 & 0 & 0 \\ 0.40 & 0.70 & 0 & 0 \\ 0.05 & 0.20 & 0.75 & 0 \\ 0.25 & 0.10 & 0.25 & 1 \end{bmatrix} & \begin{matrix} B \\ I \\ A \\ L \end{matrix} \end{matrix} \text{After assessment}$$

a An incomplete transition diagram for matrix $T_2$ is shown. Copy and complete the transition diagram by adding the missing information. 2 marks



The number of students at each level, immediately before the first assessment of the year, is shown in matrix $R_0$.

$$R_0 = \begin{bmatrix} 20 \\ 60 \\ 40 \\ 0 \end{bmatrix} \begin{matrix} B \\ I \\ A \\ L \end{matrix}$$

b 33% What percentage of students is expected to leave the school after the first assessment? 1 mark

c 33% How many advanced-level students are expected to be in the school after two assessments? Write your answer correct to the nearest whole number, 1 mark

d 33% After how many assessments is the number of students in the school, correct to the nearest whole number, first expected to drop below 50? 1 mark

Another model for the number of students in the school after each assessment takes into account the number of students who are expected to join the school after each assessment. Let $R_n$ be the state matrix that contains the number of students in the school immediately after $n$ assessments. Let $V$ be the matrix that contains the number of students who join the school after each assessment. Matrix $V$ is shown.

$$V = \begin{bmatrix} 4 \\ 2 \\ 3 \\ 0 \end{bmatrix} \begin{matrix} B \\ I \\ A \\ L \end{matrix}$$

The expected number of students in the school after $n$ assessments can be determined using the matrix equation

$$B_{w+1} = T_2 \times B_w + V \text{ where } R_o = \begin{bmatrix} 20 \\ 60 \\ 40 \\ 0 \end{bmatrix} \begin{matrix} B \\ I \\ A \\ L \end{matrix}$$

e 33% Consider the intermediate-level students expected to be in the school after three assessments. How many are expected to become advanced-level students after the next assessment? Write your answer correct to the nearest whole number. 2 marks

**11** ©VCAA 2018N 2MQ4 (2 marks) Areas of farmland in a region are allocated to growing barley (B), corn (C) and wheat (W). This allocation of farmland is to be changed each year, beginning in 2019. The table shows the areas of farmland, in hectares, allocated to each crop in 2018 (n = 0) and 2019 ($n$ = 1).

| Year | 2018 | 2019 |
|---|---|---|
| $n$ | 0 | 1 |
| Barley | 2000 | 2100 |
| Corn | 1000 | 1900 |
| Wheat | 3000 | 2000 |

The planned annual change to the area allocated to each crop can be modelled by

$$H_{n+} j = RH_n + Q \text{ where } \quad R = \overset{\text{This year}}{\begin{matrix} B & C & W \end{matrix}} \begin{bmatrix} 0.7 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.2 \\ 0.2 & 0.1 & 0.7 \end{bmatrix} \begin{matrix} B \\ C \\ W \end{matrix} \text{ Next year}$$

$H_n$ represents the state matrix that shows the area allocated to each crop $n$ years after 2018. Q is a matrix that contains additional fixed changes to the area that is allocated to each crop each year. Copy and complete $W_2$, the state matrix for 2020.

$$H_2 = \begin{bmatrix} \_\_\_ \\ \_\_\_ \\ \_\_\_ \end{bmatrix} \begin{matrix} B \\ C \\ W \end{matrix}$$

▶ 12 ©VCAA 2019 2MQ3 , (3 marks) A theme park has four locations, Air World (A), Food World (F), Ground World (G) and Water World (*W*). The number of visitors at each of the four locations is counted every hour. On Sunday, matrix $V$ is used when calculating the expected number of visitors at each location every hour after 10 am. It is assumed that the park will be at its capacity of 2000 visitors for all of Sunday. Let £$_0$ be the state matrix that shows the number of visitors at each location at 10 am on Sunday. The number of visitors expected at each location at 11 am on Sunday can be determined by the matrix product

$$V \times L_o \text{ where } L_0 = \begin{bmatrix} 500 \\ 600 \\ 500 \\ 400 \end{bmatrix} \begin{matrix} A \\ F \\ G \\ W \end{matrix} \quad \text{and } V = \overset{\text{This hour}}{\overset{\begin{matrix} A & F & G & W \end{matrix}}{\begin{bmatrix} 0.3 & 0.4 & 0.6 & 0.3 \\ 0.1 & 0.2 & 0.1 & 0.2 \\ 0.1 & 0.2 & 0.2 & 0.1 \\ 0.5 & 0.2 & 0.1 & 0.4 \end{bmatrix}}} \begin{matrix} A \\ p \\ G \\ W \end{matrix} \text{ Next hour}$$

a 75% Safety restrictions require that all four locations have a maximum of 600 visitors. Which location is expected to have more than 600 visitors at 11 am on Sunday? 1 mark

b Whenever more than 600 visitors are expected to be at a location on Sunday, the first 600 visitors can stay at that location and all others will be moved directly to Ground World (G). State matrix $R_n$ contains the number of visitors at each location $n$ hours after 10 am on Sunday, after the safety restrictions have been enforced. Matrix $R_{\ell}$ can be determined from the matrix recurrence relation

$$R_0 = \begin{bmatrix} 500 \\ 600 \\ 500 \\ 400 \end{bmatrix} \begin{matrix} A \\ F \\ G \\ W \end{matrix} \quad Q\ y\ R[ = V R\varnothing + \quad B_1$$

where matrix $B_{\ell}$ shows the required movement of visitors at 11 am.

i 25% Determine the matrix $B_{\ell}$. 1 mark

ii 12% State matrix $R_2$ can be determined from the new matrix rule $R_2 = VR_{\ell} + B_2$ where matrix $B_2$ shows the required movement of visitors at 12 noon. Determine the state matrix $R_2$. 1 mark

13 ©VCAA | 2018 2MQ4 (2 marks) Beginning in the year 2021, a company has a contract to maintain a 2700 km highway. Each year sections of highway must be graded (G), resurfaced *(R) or* sealed (S). The remaining highway will need no maintenance *(N)* that year. Let $M_n$ be the state matrix that shows the highway maintenance schedule of this company for the nth year after 2020. The maintenance schedule for 2020 is shown in matrix $M_Q$ below. For these 2700 km of highway, the matrix recurrence relation shown below can be used to determine the number of kilometres of this highway that will require each type of maintenance from year to year.

$$M_{n+}\text{j} = TM_n + B$$

where

$$M_0 = \begin{bmatrix} 500 \\ 400 \\ 300 \\ 1500 \end{bmatrix} \begin{matrix} G \\ R \\ S \\ N \end{matrix}, \quad T = \overset{\text{This year}}{\overset{\begin{matrix} G & R & S & N \end{matrix}}{\begin{bmatrix} 0.2 & 0.1 & 0.0 & 0.2 \\ 0.1 & 0.1 & 0.0 & 0.2 \\ 0.2 & 0.1 & 0.2 & 0.1 \\ 0.5 & 0.7 & 0.8 & 0.5 \end{bmatrix}}} \begin{matrix} G \\ R \\ S \\ N \end{matrix} \text{ Next year, } B = \begin{bmatrix} k \\ 20 \\ 10 \\ -60 \end{bmatrix}$$

a 39% Write down the value of $k$ in matrix $B$. 1 mark

b 23% How many kilometres of highway are to be graded (G) in the year 2022? 1 mark



9780170448406

# @ Leslie matrices

8.5

## Leslie matrices and birth/survival rate tables

We have looked at some examples of transition matrices that deal with animal populations. Leslie matrices give us a better model of how animal populations change by taking into account how birth rates and survival rates change depending on the animal's age. Leslie matrices count only the females in a population as only females give birth.

Leslie matrices are created from tables of information that contain female birth and survival rates of different age groups. The first step to set up a table is to decide on suitable age groups. These depend on the animal's lifespan, which is the maximum time an animal can live. For example:

Video playlist Leslie matrices

| Animal | Lifespan | Age groups |
|---|---|---|
| Golden hamster | 4 years | |
| Labord's chameleon | 6 months | |
| Cockatoo | 50 years | |

| Age (years) | 0-cl | l-<2 | 2-<3 | 3-<4 |
|---|---|---|---|---|

| Age (months) | 0-<l | l-<2 | 2-<3 | 3-<4 | 4-<5 | 5-<6 |
|---|---|---|---|---|---|---|

| Age (years) | 0-<10 | 10-<20 | 20-<30 | 30-<40 | 40-<50 |
|---|---|---|---|---|---|

The following shows the table and corresponding Leslie matrix for a colony of lizards:

Female birth and survival rate table

| Age (years) | 0-<1 | 1-<2 | 2-<3 | 3-<4 |
|---|---|---|---|---|
| Birth rate | 0.2 | 0.6 | 1.4 | 0.9 |
| Survival rate | 0.3 | 0.8 | 0.7 | 0.0 |

Leslie matrix

$$L = \begin{bmatrix} 0.2 & 0.6 & 1.4 & 0.9 \\ 0.3 & 0 & 0 & 0 \\ 0 & 0.8 & 0 & 0 \\ 0 & 0 & 0.7 & 0 \end{bmatrix}$$

On average, female lizards aged

- zero to one give birth to 0.2 female lizards in a year and have a 0.3 (or 30%) chance of surviving to be one
- one give birth to 0.6 female lizards in a year and have a 0.8 (or 80%) chance of surviving to be two
- two give birth to 1.4 female lizards in a year and have a 0.7 (or 70%) chance of surviving to be three
- three give birth to 0.9 female lizards in a year and have a 0.0 (or 0%) chance of surviving to be four.

### © Exam hack

Birth rates can be numbers greater than one, but survival rates can't be greater than one. Unlike transition matrices, the columns of Leslie matrices don't add to 1.

### Leslie matrices and birth/survival rate tables

Leslie matrices are square matrices that calculate female populations where

- birth rates are given for each age group in the first row
- survival rates are given as decimals for each age group in the subdiagonal (the diagonal immediately below the leading diagonal)
- zeros are everywhere else.

Leslie matrices are based on birth/survival rate tables where

- the birth rate is the average number of female babies for a female in an age group
- the survival rate is the chance a female in an age group will live to be the age of the next age group
- the last survival rate in the table, which is always zero, does not appear in the Leslie matrix.

WB
p. 170

## WORKED EXAMPLE 14 Using female birth and survival rate tables

A population of Amazonian capybaras consists of

- 49 female capybaras less than one year old of which 35% will survive to be one
- 62 one-year-old female capybaras of which 43% will survive to be two
- 38 two-year-old female capybaras of which none will survive to be three.

The females in the colony have the following average yearly birth rates:

- Female capybaras less than one year old give birth on average to 4.6 capybaras.
- Female one-year-old capybaras give birth on average to 4.2 capybaras.
- Female two-year-old capybaras give birth on average to 2.8 capybaras.

Assume that half the population is female and half the births are female.

**Steps** | **Working**

a Explain why we need to halve the given capybara birth rates to find the female birth and survival rates.

Check which of the figures given apply to females only and which apply to both males and females.

The birth rates given are for baby capybaras, not just female baby capybaras. So, we need to halve these figures.

b Draw up a female birth and survival rate table that includes the initial number of female capybaras.

Make sure that all the entries apply to females only.

Convert percentage survival rates to decimals.

| Age (years) | 0-<1 | 1-<2 | 2-<3 |
|---|---|---|---|
| Initial number | 49 | 62 | 38 |
| Birth rate | 2.3 | 2.1 | 1.4 |
| Survival rate | 0.35 | 0.43 | 0 |

c Write a calculation from the table that will show how many one-year-old female capybaras are expected to survive to be two.

Multiply the initial number and survival rate from the table. Round to the nearest whole number.

$62 \times 0.43 = 26.66$

So 27 one-year-old female capybaras are expected to survive to be two.

d Write a calculation from the table that will show how many female capybaras are expected to be born after one year.

Multiply the initial numbers of capybaras by each of their birth rates and add them together. Round to the nearest whole number.

$(49 \times 2.3) + (62 \times 2.1) + (38 \times 1.4)$

$= 112.7 + 130.2 + 53.2$

$= 296.1$

So 296 female capybaras are expected to be born after one year.

e Write the Leslie matrix $L$ that represents the information in the table.

Write the birth rates as the first row.

Write the survival rates without the last zero as the subdiagonal.

Write zeros everywhere else.

$$L = \begin{bmatrix} 2.3 & 2.1 & 1.4 \\ 0.35 & 0 & 0 \\ 0 & 0.43 & 0 \end{bmatrix}$$

### i® Exam hack

Assume that half the total population is female and half is male, unless told otherwise in the question.



9780170448406

## The Leslie matrix recurrence relation and rule

The recurrence relation and rule for a Leslie matrix $L$ work the same way as the rules we used for transition matrices. For example, if we know how many female lizards there are in the colony in each age group from the earlier example, we can create an initial state matrix and find the number of female lizards in each age group after one year by multiplying the Leslie matrix on the left.

Initial number of female lizards in each age group

Table

| Age (years) | 0-<l | l-<2 | 2-<3 | 3-<4 |
|---|---|---|---|---|
| Initial number | 48 | 33 | 52 | 29 |

Initial state matrix

$$S_0 = \begin{bmatrix} 48 \\ 33 \\ 52 \\ 29 \end{bmatrix} \begin{matrix} \text{O-cl} \\ \text{1-<2} \\ \text{2-<3} \\ \text{3-<4} \end{matrix}$$

No. of female lizards in each age group after one year

$$S_1 = \begin{bmatrix} 0.2 & 0.6 & 1.4 & 0.9 \\ 0.3 & 0 & 0 & 0 \\ 0 & 0.8 & 0 & 0 \\ 0 & 0 & 0.7 & 0 \end{bmatrix} \begin{bmatrix} 48 \\ 33 \\ 52 \\ 29 \end{bmatrix} = \begin{bmatrix} 128.3 \\ 14.4 \\ 26.4 \\ 36.4 \end{bmatrix} \begin{matrix} \text{0-<l} \\ \text{l-<2} \\ \text{2-<3} \\ \text{3-<4} \end{matrix}$$

No. of female lizards in each age group after two years

$$S_2 = \begin{bmatrix} 0.2 & 0.6 & 1.4 & 0.9 \\ 0.3 & 0 & 0 & 0 \\ 0 & 0.8 & 0 & 0 \\ 0 & 0 & 0.7 & 0 \end{bmatrix} \begin{bmatrix} 128.3 \\ 14.4 \\ 26.4 \\ 36.4 \end{bmatrix} = \begin{bmatrix} 104.2 \\ 38.49 \\ 11.52 \\ 18.48 \end{bmatrix} \begin{matrix} \text{0-cl} \\ \text{1-<2} \\ \text{2-<3} \\ \text{3-<4} \end{matrix}$$

To find the number of female lizards in each age group after 12 years, we use the same rule as for transition matrices: $S_n = L^n S_0$:

$$S_{12} = \begin{bmatrix} 0.2 & 0.6 & 1.4 & 0.9 \\ 0.3 & 0 & 0 & 0 \\ 0 & 0.8 & 0 & 0 \\ 0 & 0 & 0.7 & 0 \end{bmatrix}^{12} \begin{bmatrix} 48 \\ 33 \\ 52 \\ 29 \end{bmatrix} = \begin{bmatrix} 54.038 \\ 17.130 \\ 14.351 \\ 10.741 \end{bmatrix} \begin{matrix} \text{0-<l} \\ \text{1-<2} \\ \text{2-<3} \\ \text{3-<4} \end{matrix}$$

**The Leslie matrix recurrence relation and rule**

The Leslie recurrence relation that generates a sequence of state matrices $S_n$ is

$S_0$ = initial state matrix, $S_{n+1} = LS_n$

The Leslie rule for finding the state matrix $S_n$ is

$S_n = L^n S_0$

where

$L$ is a Leslie matrix

$n$ is the number of time periods.

**Exam hack**

Since Leslie matrices only include the numbers of females, if a question asks to find the total population, multiply your answer by two.

WB
p. 171

## WORKED EXAMPLE 15 Using the Leslie matrix recurrence relation and rule

Data from a study of the four-year lifespan of a fish population has been written as the following initial state matrix $S_0$ and Leslie matrix $L$.

$$S_0 = \begin{bmatrix} 230 \\ 610 \\ 450 \\ 320 \end{bmatrix} \begin{matrix} 0\text{-}<1 \\ 1\text{-}<2 \\ 2\text{-}<3 \\ 3\text{-}<4 \end{matrix} \qquad L = \begin{bmatrix} 0 & 0 & 20.0 & 0.5 \\ 0.3 & 0 & 0 & 0 \\ 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0.1 & 0 \end{bmatrix}$$

| Steps | Working |
|---|---|
| a How do we know from the Leslie matrix that the fish do not breed until they are two years old? | |
| Look at the first two birth rates. | The first two elements in the first row are zero. This means the fish up to two years old have birth rates of zero. |
| b Show a calculation using a recurrence relation to find the number of two-year-old fish after one year. | |
| 1 Find Sj using the Leslie matrix recurrence relation. | $S_0$ = initial state matrix, $S_{n+1} = LS_n$ <br> $S_1 = LS_0 = \begin{bmatrix} 0 & 0 & 20.0 & 0.5 \\ 0.3 & 0 & 0 & 0 \\ 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0.1 & 0 \end{bmatrix} \begin{bmatrix} 230 \\ 610 \\ 450 \\ 320 \end{bmatrix}$ <br> $= \begin{bmatrix} 9160 \\ 69 \\ 122 \\ 45 \end{bmatrix} \begin{matrix} 0\text{-}<1 \\ 1\text{-}<2 \\ 2\text{-}<3 \\ 3\text{-}<4 \end{matrix}$ |
| 2 Find the number from Sj, multiply it by 2, and then round it to the nearest whole number if necessary. | After one year, there are 122 female two-year old fish, so the number of two-year old fish is <br> $122 \times 2 = 244$ |
| c Show a calculation using a recurrence relation to find the total number of fish after two years. | |
| 1 Find $S_2$ using the Leslie matrix recurrence relation. | $S_2 = LS_1 = \begin{bmatrix} 0 & 0 & 20.0 & 0.5 \\ 0.3 & 0 & 0 & 0 \\ 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0.1 & 0 \end{bmatrix} \begin{bmatrix} 9160 \\ 69 \\ 122 \\ 45 \end{bmatrix}$ <br> $= \begin{bmatrix} 2462.5 \\ 2748 \\ 13.8 \\ 12.2 \end{bmatrix} \begin{matrix} 0\text{-}<1 \\ 1\text{-}<2 \\ 2\text{-}<3 \\ 3\text{-}<4 \end{matrix}$ |
| 2 Multiply each of the elements in $S_2$ by 2, rounding to the nearest whole number if necessary, and add them together. | $(2462.5 \times 2) + (2748 \times 2) + (13.8 \times 2) + (12.2 \times 2)$ <br> $= 4925 + 5496 + 27.6 + 24.4$ <br> $= 4925 + 5496 + 28 + 24$ (rounding) <br> $= 10453$ |
| 3 Write the answer. | The total number of fish after two years is 10453. |


9780170448406

8.5

d Find the number of three-year-old fish after ten years.

| Steps | Working |
|---|---|
| Use the Leslie matrix rule. | $S_n = L^n S_0$<br>$S_{10} = L^{10} S_0 = \begin{bmatrix} 0 & 0 & 20.0 & 0.5 \\ 0.3 & 0 & 0 & 0 \\ 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0.1 & 0 \end{bmatrix}^{10} \begin{bmatrix} 230 \\ 610 \\ 450 \\ 320 \end{bmatrix}$<br>$= \begin{bmatrix} 15832 \\ 127.21 \\ 216.72 \\ 79.152 \end{bmatrix} \begin{matrix} 0\text{-}<1 \\ 1\text{-}<2 \\ 2\text{-}<3 \\ 3\text{-}<4 \end{matrix}$ |
| Find the number from $S_{10}$, multiply it by 2, and then round it to the nearest whole number if necessary. | After ten years, there are 79.152 female three-year old fish, so the number of three-year old fish is $79.152 \times 2 = 158.304$, which rounds to 158. |

## Long-term Leslie matrix trends

In the long term, populations modelled by Leslie matrices can increase, decrease or cycle. A cycle involves a repeating pattern of state matrices that start and end with the initial state matrix.

**Long-term Leslie matrix trends**

In the long term the population modelled by an $m \times m$ Leslie matrix $L$ will

- increase if the total population of $S_m >$ the total population of $S_0$
- decrease if the total population of $S_m <$ the total population of $S_0$
- cycle every $m$ time periods if $S_m = S_0$.

When a cycle occurs, $L^m$ equals the identity matrix.

**WORKED EXAMPLE 16** Finding long-term Leslie matrix trends

WB p. 172

For each of the following Leslie matrices $L$ where $S_0 = \begin{bmatrix} 1200 \\ 0 \\ 0 \end{bmatrix}$, show whether in the long term the populations will increase, decrease or cycle.

**a** $L = \begin{bmatrix} 0 & 0 & 200 \\ 0.5 & 0 & 0 \\ 0 & 0.01 & 0 \end{bmatrix}$ **b** $L = \begin{bmatrix} 0 & 0 & 100 \\ 0.5 & 0 & 0 \\ 0 & 0.01 & 0 \end{bmatrix}$ **c** $L = \begin{bmatrix} 0 & 0 & 400 \\ 0.5 & 0 & 0 \\ 0 & 0.01 & 0 \end{bmatrix}$

| Steps | Working |
|---|---|
| **a 1** $L$ is $3 \times 3$, so $S_3$ using the Leslie matrix rule $S_n = L^n S_0$. | $S_3 = L^3 S_0 = \begin{bmatrix} 0 & 0 & 200 \\ 0.5 & 0 & 0 \\ 0 & 0.01 & 0 \end{bmatrix}^3 \begin{bmatrix} 1200 \\ 0 \\ 0 \end{bmatrix}$<br>$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1200 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1200 \\ 0 \\ 0 \end{bmatrix}$ |
| **2** Compare $S_3$ and $S_0$. | $S_0 = S_3$, so the population will cycle every three time periods. |

| | |
|---|---|
| b 1 $L$ is 3 x 3, so find $S_3$ using the Leslie matrix rule $S_n = L”S_0$. | $S_3 = L^3S_O = \begin{bmatrix} 0 & 0 & 100 \\ 0.5 & 0 & 0 \\ 0 & 0.01 & 0 \end{bmatrix}^3 \begin{bmatrix} 1200 \\ 0 \\ 0 \end{bmatrix}$ <br> $= \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.5 \end{bmatrix} \begin{bmatrix} 1200 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 600 \\ 0 \\ 0 \end{bmatrix}$ |
| 2 Compare $S_3$ and $S_o$. | Total population at $S_3 = 600$ <br> Total population at $S_o = 1200$ <br> $600 < 1200$ so in the long term the population will decrease. |
| c 1 $L$ is 3 x 3, so find $S_3$ using the Leslie matrix rule $S_n = L^nS_0$. | $S_3 = L^3S_o$ <br> $= \begin{bmatrix} 0 & 0 & 400 \\ 0.5 & 0 & 0 \\ 0 & 0.01 & 0 \end{bmatrix}^3 \begin{bmatrix} 1200 \\ 0 \\ 0 \end{bmatrix}$ <br> $= \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1200 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 2400 \\ 0 \\ 0 \end{bmatrix}$ |
| 2 Compare $S_3$ and $S_o$. | Total population at $S_o = 1200$ <br> Total population at $S_3 = 2400$ <br> $2400 > 1200$ so in the long term the population will increase. |

## Long-term Leslie matrix growth rates

We have seen that for regular transition matrices in the long term the state matrix gets to a point where it stops changing from one transition to the next. With Leslie matrices in the long term, the numbers in the state matrices generally keep changing, but the proportion of the population in each age group stabilises and we can calculate a long-term growth rate.

For example, if we have calculated state matrices for two large consecutive values of $n$ (e.g. 29 and 30):

$$S_{29} - \begin{bmatrix} 432.2 \\ 585.6 \\ 143.3 \end{bmatrix} \text{ and } SyQ = \begin{bmatrix} 497.1 \\ 673.5 \\ 164.8 \end{bmatrix}$$

We can calculate the growth rates for each age group:

$$\frac{497.1}{432.2} \approx 1.1501 \quad \frac{673.5}{585.6} \approx 1.1501 \quad \frac{164.8}{143.3} \approx 1.1500$$

These values are the same (to three decimal places), so we have found the population has a long-term growth rate of 1.15, which means that after a certain time the population always increases by $(1.15 - 1) \times 100\% = 15\%$ each time period.

### © Exam hack

A long-term growth rate less than 1 gives a percentage *decrease*. For example, for a growth rate of 0.92 we get $(0.92 - 1) \times 100\% = -8\%$, which means the population always *decreases* by 8% each time period.



9780170448406

## Long-term Leslie matrix growth rates

For two state matrices with large consecutive values of n,

if $\dfrac{\text{each of the elements in } S_{n+1}}{\text{matching element in } S_n}$ = the same value $g$, then

- $g$ is the long-term growth rate of the population each time period
- $(g - 1) \times 100\%$ is the percentage increase or decrease per time period.

For $g > 1$, the population is increasing.

For $g < 1$, the population is decreasing.

For $g = 1$, the population is stable.

WB p. 173

**WORKED EXAMPLE 17** Finding long-term Leslie matrix growth rates

For the following initial state matrix $S_0$ and Leslie matrix $L$, find

**a** the long-term growth rate per time period to two decimal places

**b** the percentage increase or decrease per time period to the nearest percentage.

$$S_0 = \begin{bmatrix} 1200 \\ 300 \\ 0 \end{bmatrix} \qquad L = \begin{bmatrix} 0 & 5 & 4 \\ 0.2 & 0 & 0 \\ 0 & 0.55 & 0 \end{bmatrix}$$

| Steps | Working |
|---|---|
| **a 1** Use the rule $S_n = L^n S_0$ for two large consecutive values of $n$. | $S_0 = \begin{bmatrix} 1200 \\ 300 \\ 0 \end{bmatrix}$, choose $n = 29$ and $30$<br>$S_{29} = L^{29} S_0 = \begin{bmatrix} 0 & 5 & 4 \\ 0.2 & 0 & 0 \\ 0 & 0.55 & 0 \end{bmatrix}^{29} \begin{bmatrix} 1200 \\ 300 \\ 0 \end{bmatrix} \approx \begin{bmatrix} 132\,103 \\ 22\,529.8 \\ 10\,566.6 \end{bmatrix}$<br>$S_{30} = L^{30} S_0 = \begin{bmatrix} 0 & 5 & 4 \\ 0.2 & 0 & 0 \\ 0 & 0.55 & 0 \end{bmatrix}^{30} \begin{bmatrix} 1200 \\ 300 \\ 0 \end{bmatrix} \approx \begin{bmatrix} 154\,916 \\ 26\,420.5 \\ 12\,391.4 \end{bmatrix}$ |
| **2** Find the ratios: $\dfrac{\text{each of the elements in } S_{n+1}}{\text{matching element in } S_n}$ | $\dfrac{154\,916}{132\,103} \approx 1.72691$<br>$\dfrac{26\,420.5}{22\,529.8} \approx 1.72691$<br>$\dfrac{12\,391.4}{10\,566.6} \approx 1.72695$ |
| **3** Write the long-term growth rate per time period $g$ to two decimal places. | The long-term growth rate per time period is 1.73. |
| **b 1** Calculate $(g - 1) \times 100\%$. | $(g - 1) \times 100\% = (1.73 - 1) \times 100\%$<br>$= 0.73 \times 100\%$<br>$= 73\%$ |
| **2** Is $g > 1$ (increase) or $g < 1$ (decrease)? Write the answer to the nearest percentage. | The percentage increase per time period is 73%. |

Video
VCE question analysis: Transition matrices

## VCE QUESTION ANALYSIS

©VCAA | 2018 2MQ3 | 2018 Examination 2 Matrices Question 3 (5 marks)

The Hiroads company has a contract to maintain and improve 2700 km of highway. Each year sections of highway must be graded (G), resurfaced *(R)* or sealed (S). The remaining highway will need no maintenance *(N)* that year. Let $S_n$ be the state matrix that shows the highway maintenance schedule for the nth year after 2018. The maintenance schedule for 2018 is shown in matrix $S_0$.

$$S_0 = \begin{bmatrix} 700 \\ 400 \\ 200 \\ 1400 \end{bmatrix} \begin{matrix} G \\ R \\ S \\ N \end{matrix}$$

The type of maintenance in sections of highway varies from year to year, as shown in the transition matrix, *T*.

$$\begin{matrix} & & \text{This year} & & \\ & G & R & S & N \end{matrix}$$

$$T = \begin{bmatrix} 0.2 & 0.1 & 0.0 & 0.2 \\ 0.1 & 0.1 & 0.0 & 0.2 \\ 0.2 & 0.1 & 0.2 & 0.1 \\ 0.5 & 0.7 & 0.8 & 0.5 \end{bmatrix} \begin{matrix} G \\ R \\ S \\ N \end{matrix} \text{ Next year}$$

a Of the length of highway that was graded (G) in 2018, how many kilometres are expected to be resurfaced *(R)* the following year? 1 mark

b Show that the length of highway that is to be graded (G) in 2019 is 460 km by copying the following and writing the appropriate numbers in the boxes.

$$\square \times 700 + \square \times 400 + \square \times 200 + \square \times 1400 = 460$$ 1 mark

The state matrix describing the highway maintenance schedule for the nth year after 2018 is given by

$$S_{n+1} = TS_n$$

c Copy and complete the state matrix, below for the highway maintenance schedule for 2019 (one year after 2018). 1 mark

$$S_1 = \begin{bmatrix} 460 \\ \text{--} \\ \text{------} \\ 1490 \end{bmatrix} \begin{matrix} G \\ R \\ S \\ N \end{matrix}$$

d In 2020, 1536 km of highway is expected to require no maintenance (N). Of these kilometres, what percentage is expected to have had no maintenance (N) in 2019? Round your answer to one decimal place. 1 mark

e In the long term, what percentage of highway each year is expected to have no maintenance (N)? Round your answer to one decimal place. 1 mark

### Reading the question

- Note any totals you are given, particularly those mentioned at the start of a question.
- One of the question parts is a show' question with clear guidance.
- Two question parts require percentage answers and rounding.

### Thinking about the question

- Make sure you know how to read a transition matrix.
- You need a total in order to calculate a percentage.
- Be clear on what sort of techniques you need to use to answer long-term transition questions.

**Worked solution** (/ = 1 mark)

a From the transition matrix

$$T = \begin{array}{c} \text{This year} \\ \begin{array}{cccc} G & R & S & N \end{array} \\ \left[\begin{array}{cccc} 0.2 & 0.1 & 0.0 & 0.2 \\ 0.1 & 0.1 & 0.0 & 0.2 \\ 0.2 & 0.1 & 0.2 & 0.1 \\ 0.5 & 0.7 & 0.8 & 0.5 \end{array}\right] \end{array} \begin{array}{c} G \\ R \\ S \\ N \end{array} \text{Next year}$$

0.1 of highway that was graded (G) in 2018 will be resurfaced *(R)* in 2019.

From $S_0$

$$\begin{bmatrix} 700 \\ 400 \\ 200 \\ 1400 \end{bmatrix} \begin{array}{c} G \\ R \\ S \\ N \end{array}$$

700 km of highway was graded in 2018.

So, 700 x 0.1 = 70 km / of highway that was graded in 2018 is expected to be resurfaced in 2019.

b Multiplying the first row of the transition matrix by $S_0$ gives:

0.2 x 700 + 0.1 x 400 + 0.0 x 200 + 0.2 x 1400 = 460 /

c $S_1 = TS_0 = \begin{bmatrix} 0.2 & 0.1 & 0.0 & 0.2 \\ 0.1 & 0.1 & 0.0 & 0.2 \\ 0.2 & 0.1 & 0.2 & 0.1 \\ 0.5 & 0.7 & 0.8 & 0.5 \end{bmatrix} \begin{bmatrix} 700 \\ 400 \\ 200 \\ 1400 \end{bmatrix} = \begin{bmatrix} 460 \\ 390 \\ 360 \\ 1490 \end{bmatrix} \begin{array}{c} G \\ R \\ S \\ N \end{array}$ /

d From the transition matrix $T$

$$T = \begin{array}{c} \text{This year} \\ \begin{array}{cccc} G & R & S & N \end{array} \\ \left[\begin{array}{cccc} 0.2 & 0.1 & 0.0 & 0.2 \\ 0.1 & 0.1 & 0.0 & 0.2 \\ 0.2 & 0.1 & 0.2 & 0.1 \\ 0.5 & 0.7 & 0.8 & 0.5 \end{array}\right] \end{array} \begin{array}{c} G \\ R \\ S \\ N \end{array} \text{Next year}$$

0.5 of highway requiring no maintenance in one year is expected to require no maintenance the next year.

From part c, the maintenance schedule for 2019 is

$$S_1 = \begin{bmatrix} 460 \\ 390 \\ 360 \\ 1490 \end{bmatrix} \begin{array}{c} G \\ R \\ S \\ N \end{array}$$

so 1490 km of highway is expected to require no maintenance in 2019.

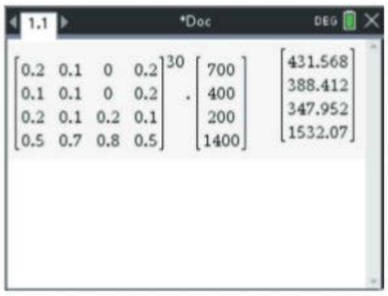
This means that the expected amount of highway not requiring any maintenance in 2020 which did not require any maintenance in 2019 is 0.5 x 1490 = 745 km.

We are told a total of 1536 km of highway will not require any maintenance in 2020, so the percentage of the 1536 km expected to have had no maintenance in 2019 is

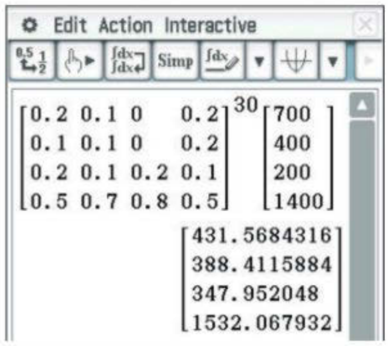
$\frac{745}{1536} \times 100\% = 0.48502... \times 100\% = 48.5\%$ / rounded to one decimal place.

e Method 1

TI-Nspire




ClassPad




Use CAS and the rule $S_n = T^nS_0$ for a large value of M, say $n = 30$. The last value in $S_{30}$ represents the long-term amount of highway not requiring any maintenance, 1532.07 km. Calculating the percentage of the total 2700 km of highway this represents, gives

$$\frac{1532.07}{2700} \times 100\% = 56.7\%$$ /, rounded to one decimal place.

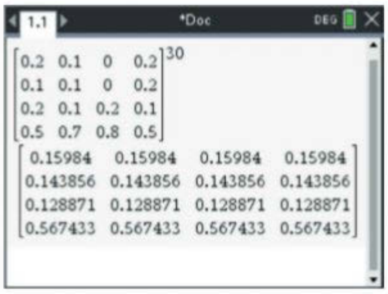
## © Exam hack

For both Method 1 and 2, you would normally need to check to see if $\boldsymbol{n} = 31$ gives the same answer, but in an exam this is only necessary in a show that' question.

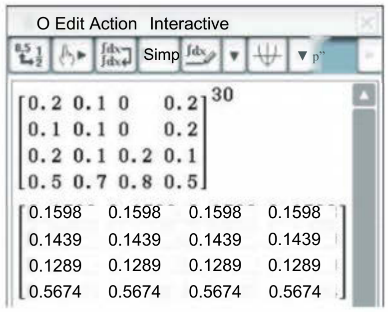
Method 2

This method is quicker when the question asks for a long-term percentage rather than a long-term number.

TI-Nspire




ClassPad




Using CAS to calculate $T^n$ to a high power of M, say $n = 30$, we can see the last row, which represents the long-term proportion of highway expected to require no maintenance, becomes a consistent 0.567 433. Converting this to a percentage gives

$0.567433 \times 100\% = 56.7\%$ /, rounded to one decimal place.

**Student performance** 8o_100% 60‹‹79% o_59%

a 48% Many students gave 390 as their answer, indicating that they had not read the question fully. Some gave 10% or 0.1 rather than finding the length.

b 77% Some students listed the numbers in the *G* column rather than the row.

C 82%

d 17% This question was not answered well, with most students giving 97%, which did not take account of the 0.5 change.

e 39% Some students left the answer as 1532.1 and did not convert to the required percentage.



9780170448406

## EXERCISE 8.5 Leslie matrices

ANSWERS p. 720

### Recap

1 The matrix $S_{,,+1}$ can be calculated from the matrix $S_n$ using the rule $S_{M+1} = TS_n + R$, where T, $S_o$ and $R$ are defined as follows:

$$T = \begin{bmatrix} 0.4 & 0.5 \\ 0.4 & 0.7 \end{bmatrix}, \ S_0 = \begin{bmatrix} 200 \\ 150 \end{bmatrix} \text{ and } R = \begin{bmatrix} 25 \\ 15 \end{bmatrix}$$

Given this information, the matrix $S_2$ equals

A $\begin{bmatrix} 155 \\ 192 \end{bmatrix}$ **B** $\begin{bmatrix} 172 \\ 212 \end{bmatrix}$ C $\begin{bmatrix} 180 \\ 200 \end{bmatrix}$ **D** $\begin{bmatrix} 180 \\ 207 \end{bmatrix}$ E $\begin{bmatrix} 197 \\ 227 \end{bmatrix}$

2 At a frog farm

- tadpoles (T) may eventually grow into juveniles (J) or they may die (D)
- juveniles (J) may eventually grow into adults (A) or they may die (D)
- adults (A) eventually die (D).

The initial state of this population, $F_o$, is shown.

$$F_0 = \begin{bmatrix} 60000 \\ 9000 \\ 5\,000 \\ 0 \end{bmatrix} \begin{matrix} T \\ I \\ A \\ D \end{matrix}$$

Every month, frogs are either sold or bought so that the number of tadpoles, juveniles and adult frogs in the farm remains constant. The population of frogs in the frog farm after $n$ months, $F_n$, can be determined by the recurrence rule

$$F_{n+1} = \begin{bmatrix} 0.5 & 0 & 0 & 0 \\ 0.25 & 0.8 & 0 & 0 \\ 0 & 0.15 & 0.9 & 0 \\ 0.25 & 0.05 & 0.1 & 1 \end{bmatrix} F_n + B$$

where $B$ is a column matrix that shows the number of tadpoles, juveniles and adult frogs bought or sold each month and the number of dead frogs that are removed. Each month, the frog farm will

A buy 850 adult frogs.

B sell 850 adult frogs.

C buy 5850 adult frogs.

D sell 30000 juveniles.

E buy 60 000 tadpoles.

## Mastery

30 WORKED EXAMPLE 14 A colony of rodents consists of

- 93 female rodents less than one year old of which 33% will survive to be one
- 72 one-year-old female rodents of which 46% will survive to be two
- 68 two-year-old female rodents of which none will survive to be three.

The females in the colony have the following average yearly birth rates:

- Female rodents less than one year old give birth on average to 3.9 rodents
- Female one-year-old rodents give birth on average to 4.5 rodents
- Female two-year-old rodents give birth on average to 3.1 rodents.

Assume that half the colony is female and half the births are female.

a Explain why we need to halve the given rodents birth rates to find the female birth and survival rates,

b Draw up a female birth and survival rate table that includes the initial number of female rodents,

c Write a calculation from the table that will show how many female rodents under one are expected to survive to be one.

d Write a calculation from the table that will show how many female rodents are expected to be born after one year.

e Write the Leslie matrix $L$ that represents the information in the table.

4 WORKED EXAMPLE is Data from a study of the four-year lifespan of a population of a particular species of spider has been written as the following initial state matrix $S_0$ and Leslie matrix $L$.

$$S_0 = \begin{bmatrix} 120 \\ 63 \\ 58 \\ 22 \end{bmatrix} \begin{matrix} \text{0-cl} \\ \text{!"<2} \\ \text{2-<3} \\ \text{3-<4} \end{matrix} \qquad L = \begin{bmatrix} 0 & 33 & 12 & 10 \\ 0.1 & 0 & 0 & 0 \\ 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0.05 & 0 \end{bmatrix}$$

a How do we know from the Leslie matrix that the spiders do not breed until they are one year old?

b Show a calculation using a recurrence relation to find the number of three-year-old spiders after one year.

c Show a calculation using a recurrence relation to find the total number of spiders after two years,

d Find the number of two-year-old spiders after twelve years.

5 WORKED EXAMPLE 16 For each of the following Leslie matrices $L$ where

$$S_0 = \begin{bmatrix} 2500 \\ 0 \\ 0 \end{bmatrix},$$

show whether in the long term the populations will increase, decrease or cycle.

a $L = \begin{bmatrix} 0 & 0 & 1000 \\ 0.125 & 0 & 0 \\ 0 & 0.02 & 0 \end{bmatrix}$ b $L = \begin{bmatrix} 0 & 0 & 400 \\ 0.125 & 0 & 0 \\ 0 & 0.02 & 0 \end{bmatrix}$ c $L = \begin{bmatrix} 0 & 0 & 100 \\ 0.125 & 0 & 0 \\ 0 & 0.02 & 0 \end{bmatrix}$

6 H WORKED EXAMPLE 17 J For the initial state matrix $S_o = \begin{bmatrix} 60 \\ 40 \end{bmatrix}$ and each of the following Leslie matrices $L$

a $L$- $\begin{bmatrix} 1 & 4 \\ 0.02 & 0 \end{bmatrix}$

b $L$- $\begin{bmatrix} 1 & 4 \\ 0.01 & 0 \end{bmatrix}$

find

i the long-term growth rate per time period to two decimal places

ii the percentage increase or decrease per time period to the nearest percentage.

**Exam practice**

80-100% 60-79% 0-59%

7 How many of the following could be Leslie matrices?

$\begin{bmatrix} 1.3 & 4.6 & 1.9 \\ 0.2 & 0 & 0 \\ 0.8 & 0 & 0 \end{bmatrix}$ $\begin{bmatrix} 0 & 40 & 35 & 10 \\ 0.3 & 0 & 0 & 0 \\ 0 & 1.2 & 0 & 0 \\ 0 & 0 & 0.7 & 0 \end{bmatrix}$ $\begin{bmatrix} 20 & 40 \\ 0.13 & 0 \\ 0 & 0.5 \end{bmatrix}$ $\begin{bmatrix} 0 & 100 \\ 0.32 & 0 \end{bmatrix}$

A 0 B 1 C 2 D 3 E 4

8 For the initial state matrix $S_o$ and Leslie matrix $L$ shown,

$S_0 = \begin{bmatrix} 0 \\ 38 \\ 55 \end{bmatrix}$ $L$- $\begin{bmatrix} 1.6 & 3.7 & 1.2 \\ 0.42 & 0 & 0 \\ 0 & 0.65 & 0 \end{bmatrix}$

the total population after one time period is closest to

A 230 B 231 C 232 D 462 E 464

*Use the following information to answer the next two questions.*

Data from a study of a female koala population has been written in the following table.

| Age (years) | 0-<2 | 2—<4 | 4-<6 | 6-<8 | 8-<10 |
|---|---|---|---|---|---|
| Initial number | 42 | 38 | 50 | 43 | 29 |
| Birth rate | 0 | 0.8 | 1.1 | 0.7 | 0.6 |
| Survival rate | 0.3 | 0.5 | 0.2 | 0.4 | 0.1 |

9 A calculation from the table that will show how many four and five-year-old female koalas are expected to die before they reach six years old is

A 50x0.2 B 50x0.4 C 50x0.6 D 50x0.8 E 50x1.1

10 A calculation from the table that gives the total number of male and female babies on average that a female koala will give birth to if it lives up to 10 years is

A 0.3+ 0.5+ 0.2+ 0.4+ 0.1

B 0 + 0.8+1.1+0.7+ 0.6

C (42 x 0) + (38 x 0.8) + (50 x 1.1) + (43 x 0.7) + (29 x 0.6)

D (0 + 0.8 + 1.1+0.7 +0.6) x 2

E (0.3 + 0.5 + 0.2+ 0.4 +0.1) x 2

*Use the following information to answer the next three questions.*

The Leslie matrix shown models the population changes of a particular species of toad over five years.

$$\begin{bmatrix} 0 & 2.4 & 2.5 & 0.9 & 0.7 \\ 0.3 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0.2 & 0 \end{bmatrix}$$

11 What does the 0.9 tell us?

A On average 90% of three-year-old toads survive to become four years old.

B On average 90% of four-year-old toads survive to become five years old.

C On average three-year-old toads have 0.9 babies.

D On average four-year-old toads have 0.9 babies.

E On average five-year-old toads have 0.9 babies.

12 What does the 0.4 tell us?

A On average 40% of two-year-old toads survive to become three years old.

B On average 40% of three-year-old toads survive to become four years old.

C On average two-year-old toads have 0.4 babies.

D On average three-year-old toads have 0.4 babies.

E On average four-year-old toads have 0.4 babies.

13 Which one of the following is indicated by the matrix?

A Some toads live to be five years old.

B Three-year-old toads have the highest birth rate.

C On average 20% of toads live to be four years old.

D Two-year-old toads have an 80% chance of dying before they become three years old.

E The toads do not give birth until they are one year old.

14 Which of the following describes the long-term trend of the population modelled by the initial state matrix $S_o$ and Leslie matrix $L$ shown?

$$S_0 = \begin{bmatrix} 2000 \\ 0 \\ 0 \end{bmatrix}. L- \begin{bmatrix} 0 & 0 & 16 \\ 0.125 & 0 & 0 \\ 0 & 0.5 & 0 \end{bmatrix}$$

A long-term decrease

B long-term increase

C cycle every three time periods

D cycle every four time periods

E cycle every five time periods



9780170448406

▶ 15 (9 marks) An investigation into a population of white-banded tanagers in central Brazil wants to establish if this bird species is in long-term decline. The birth and survival rate table for female tanagers is shown.

| Age (years) | 0-<2 | 2-<4 | 4-<6 | 6-<8 |
|---|---|---|---|---|
| Birth rate | 0 | 0.35 | 0.31 | 0.27 |
| Survival rate | 0.42 | 0.71 | 0.65 | 0 |

We also know

- the total female population at the start of the investigation is 34 500
- there are twice as many females as males in this population.

a What is the lifespan of the white-banded tanagers? 1 mark

b What percentage of two-year-old tanagers live to be three? 1 mark

c Find the Leslie matrix $L$ for this population. 1 mark

d If we estimate there are 10000 female tanagers in each of the first three age groups at the start of the investigation, find the initial state matrix. 1 mark

e Find the state matrices after 19 and 20 time periods with all the elements rounded to two decimal places. 2 marks

f Use your answer from part e to find

i the number of white-banded tanagers in the first age group of the population after 20 time periods 1 mark

ii the long-term growth rate $g$ to one decimal place, and comment on whether it supports the theory that the white-banded tanagers are in long-term decline. 2 marks

# (7) Chapter summary

## Transition diagrams and matrices

### Transition diagrams

- show transitions using percentages or decimals
- have all the arrow percentages *from* a single point add up to 100%
- do not show transitions that are 0%.

### Transition matrices

- are square matrices
- show the transitions as decimals
- have each column adding up to 1
- show transitions that are 0.

## The state matrix recurrence relation

The recurrence relation that generates a sequence of state matrices $S_n$ is

$S_o$ = initial state matrix, $S_{n+1} = TS_n$

where

*T* is a transition matrix

*n* is the number of transitions.

The sum of the elements in $S_n$ is always the same.

## The state matrix rule

The rule for finding the state matrix $S_n$ after *n* transitions is

S,, = T"$S_0$

where

*T* is a transition matrix

$T^n$ is a transition matrix

$S_o$ is the initial state matrix.

## The equilibrium state matrix

A regular transition matrix *T* is a transition matrix that either has no zeros itself or any one of its powers has no zeros.

For large *n*

- $T^n$ stops changing and has equal elements in each row
- the state matrix $S_n$ stops changing and becomes the equilibrium state matrix
- both $T^n$ and $S_n$ give information about long-term trends.

To find the equilibrium state matrix check that two large consecutive values of *n* give the same state matrix.

## Restocking and culling recurrence relations

The restocking and culling recurrence relation that generates a sequence of state matrices $S_n$ is

$S_0$ = initial state matrix, $S_{n+1} = TS_n + B$

where

*T* is a transition matrix

*n* is the number of transitions

restocking is adding to the population and culling is removing from the population

*B* can have both positive elements and negative elements and has the same order as $S_n$.


9780170448406

### Leslie matrices and birth/survival rate tables

Leslie matrices are square matrices which calculate female populations where

- birth rates are given for each age group in the first row
- survival rates are given as decimals for each age group in the subdiagonal
- zeros are everywhere else.

Leslie matrices are based on birth/survival rate tables where

- the birth rate is the average number of female babies for a female in an age group
- the survival rate is the chance a female in an age group will live to be the age of the next age group
- the last survival rate in the table, which is always zero, does not appear in the Leslie matrix.

### The Leslie matrix recurrence relation and rule

The Leslie recurrence relation that generates a sequence of state matrices $S_n$ is

$S_o$ = initial state matrix, $S_{w+1} = LS_n$

The Leslie rule for finding the state matrix $S_n$ is

$S_n = L^n S_0$

where

$L$ is a Leslie matrix

$n$ is the number of time periods.

### Long-term Leslie matrix trends

In the long term the population modelled by an $m$ x $m$ Leslie matrix $L$ will

- increase if the total population of $S_{rn}$ > the total population of $S_o$
- decrease if the total population of $S_m$ < the total population of $S_o$
- cycle every $m$ time periods if $S_m = S_o$.

When a cycle occurs, $L^m$ equals the identity matrix.

### Long-term Leslie matrix growth rates

For two state matrices with large consecutive values of n,

if $\dfrac{\text{each of the elements in } S_{n+1}}{\text{matching element in } S_n}$ = the same value, $g$ then

- $g$ is the long term growth rate of the population each time period
- (g - 1) x 100% is the percentage increase or decrease per time period.

Forg > 1, the population is increasing.

Forg < 1, the population is decreasing.

Forg = 1, the population is stable.

# Cumulative examination 1

Total number of marks: 17 Reading time: 7 minutes Writing time: 39 minutes

*Use the following information to answer the next two questions.*

The ordered stem plot shows the percentage of homes connected to broadband internet for 24 countries in 2007.

| Stem | Leaf |
|---|---|
| 1 | |
| 1 | 6 7 |
| 2 | 0 113 4 4 |
| 2 | 5 7 8 9 |
| 3 | 0 0 1 1 1 2 2 3 |
| 3 | 5 7 8 8 |
| 4 | |

Key: 1|6= 16%

1 ©VCAA 2013 1CQ1 The number of these countries with more than 22% of homes connected to broadband internet in 2007 is

A 4 B 5 C 19 D 20 E 22

2 ©VCAA 2013 1CQ2 Which one of the following statements relating to the data in the ordered stem plot is not true?

A The minimum is 16%.

B The median is 30%.

C The first quartile is 23.5%.

D The third quartile is 32%.

E The maximum is 38%.

*Use the following information to answer the next two questions.*

The length (in metres) and wingspan (in metres) of eight commercial aeroplanes are displayed in the table.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Length | 70.7 | 70.7 | 63.7 | 58.4 | 54.9 | 39.4 | 36.4 | 33.4 |
| Wingspan | 64.4 | 59.6 | 60.3 | 60.3 | 47.6 | 35.8 | 28.9 | 28.9 |

3 ©VCAA rnrn Correct to four decimal places, the value of the Pearson correlation coefficient for this data is

A 0.9371 B 0.9583 C 0.9681 D 0.9793 E 0.9839

4 ©VCAA 2005 1CQ9 The equation of the least squares line of best fit for this data is

*wingspan* = -2.99 + 0.96 x *length*

From this equation it can be concluded that, on average, for these aeroplanes, wingspan

A decreases by 2.03 metres with each one metre increase in length.

B increases by 0.96 metres with each one metre increase in length.

C decreases by 0.96 metres with each one metre increase in length.

D increases by 2.99 metres with each one metre increase in length.

E decreases by 2.99 metres with each one metre increase in length.

5 Which one of the following recurrence relations could represent a reducing balance loan?

A $V_o$-5OOO, $V_{,,+1}$- V,, - 525 B $V_0$- 5000, $V_{,,+1}$-0.72V,,

C Vo-5000, $V_{,,+1}$ - 1.06 *V,,* + 1500 D $V_o$ - 5000, $V_{,,+1}$ - 1.003 *V,,* - 2365

E $V_o$ = 5000, $V_{,,+1}$ - 1.062V,,

6 ©VCAA 2005 1BQ7 Gregor invests $10000 and earns interest at a rate of 6% per annum compounding quarterly. Every quarter, after interest has been added, he withdraws $500. At the end of four years, after interest has been added and he has made the $500 withdrawal, the value of the remaining investment will be closest to

A $3720 B $4220 C $5440 D $21660 E $22160

7 ©VCAA 2018N1MQ1 The matrix $\begin{bmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ 8 & 5 & 1 \end{bmatrix}$ is an example of a

A unit matrix. B diagonal matrix. C triangular matrix.

D symmetric matrix. E communication matrix.

8 IfA = $\begin{bmatrix} 8 & 7 \\ 3 & 2 \end{bmatrix}$, which of the following is true?

A det (A) = det ($A^{-1}$) B det (A) = det ($A^2$) C det (A)-det ($A^7$)

D det (A) = det ($A^3$) E det (A) = det($AA^{-1}$)

9 ©VCAA 2012 1MQ6 The table shows the number of classes and the number of students in each class at each year level in a secondary school.

| | Year level | | | |
|---|---|---|---|---|
| | 9 | 10 | 11 | 12 |
| Number of classes | 7 | 5 | 6 | 4 |
| Students per class | 22 | 20 | 18 | 24 |

Let F = [ 1 1 1 1 ], G = $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, A4 = [ 7 5 6 4 ], N = $\begin{bmatrix} 7 \\ 5 \\ 6 \\ 4 \end{bmatrix}$, *P* = $\begin{bmatrix} 22 & 0 & 0 & 0 \\ 0 & 20 & 0 & 0 \\ 0 & 0 & 18 & 0 \\ 0 & 0 & 0 & 24 \end{bmatrix}$

A matrix product that displays the total number of students in Years 9-12 at this school is

A *MxPxF* B *P* x *G* x *M* C Fx *PxN*

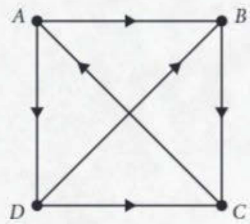
D *PxNxF* E *FxNxP*

10 The matrix $\begin{bmatrix} 0 & 0 & 1 & 1 \\ 0.1 & 0 & 0 & 0 \\ 0 & 0.01 & 0 & 0 \\ 0 & 0 & 0.1 & 0 \end{bmatrix}$ is an example of a

A transition matrix. B dominance matrix. C Leslie matrix.

D communication matrix. E state matrix.

**11** ©VCAA 2013 1NQ5 Four people, Ash ($A$), Binh ($B$), Con ($C$) and Dan ($D$), competed in a table tennis tournament. In this tournament, each competitor played each of the other competitors once. The results of the tournament are summarised in the directed graph. Each arrow shows the winner of a game played in the tournament. For example, the arrow from C to A shows that Con defeated Ash.




In the tournament, each competitor was given a ranking that was determined by calculating the sum of their one-step and two-step dominances. The competitor with the highest sum is ranked number one (1). The competitor with the second-highest sum was ranked number two (2), and so on.

Using this method, the rankings of the competitors in this tournament were

A Dan (1), Ash (2), Con (3), Binh (4)

B Dan (1), Ash (2), Binh (3), Con (4)

C Con (1), Dan (2), Ash (3), Binh (4)

D Ash (1), Dan (2), Binh (3), Con (4)

E Ash (1), Dan (2), Con (3), Binh (4)

12 Which of the following matrices could be a Leslie matrix?

A $\begin{bmatrix} 247 \\ 322 \\ 153 \end{bmatrix}$

B $\begin{bmatrix} 0 & 100 \\ 5.2 & 0 \end{bmatrix}$

C $\begin{bmatrix} 3 & 5 & 4 \\ 0.7 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.2 \end{bmatrix}$

D $\begin{bmatrix} 0 & 0 & 4.2 & 6.1 \\ 0 & 0.3 & 0 & 0 \\ 0 & 0 & 0.9 & 0 \\ 0 & 0 & 0 & 0.4 \end{bmatrix}$

E $\begin{bmatrix} 0 & 0 & 0 & 0.3 & 0.8 \\ 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0.6 & 0 & 0 & 0 \\ 0 & 0 & 0.7 & 0 & 0 \\ 0 & 0 & 0 & 0.1 & 0 \end{bmatrix}$

13 ©VCAA 2018N 1MQ7 A company has selected 200 of its regular customers to rate its performance every month. The rating given by a customer can be poor (P), good (G) or excellent (E). Customers are expected to change their rating from month to month as shown in the transition matrix $T$.

$$\begin{array}{c} \text{This month} \\ \begin{array}{ccc} P & G & E \end{array} \end{array}$$

$$T = \begin{bmatrix} 0.2 & 0.3 & 0.2 \\ 0.7 & 0.3 & 0.5 \\ 0.1 & 0.4 & 0.3 \end{bmatrix} \begin{matrix} P \\ G \\ E \end{matrix} \text{ Next month}$$

The expected number of each rating received after $n$ months can be determined by the recurrence relation

$$S_0 = \begin{bmatrix} 40 \\ 110 \\ 50 \end{bmatrix} \begin{matrix} P \\ G, \\ E \end{matrix} \quad \text{^n+l=^n}$$

where $S_o$ is the state matrix for January.

What percentage of these 200 customers are not expected to change their rating in February?

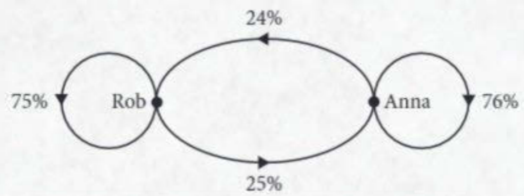
A 28% B 40% C 43% D 56% E 80%

*Use the following information to answer the next two questions.*

Two politicians, Rob and Anna, are the only candidates for a forthcoming election. At the beginning of the election campaign, people were asked for whom they planned to vote. The numbers were as follows.

| Candidate | Number of people who plan to vote for the candidate |
|---|---|
| Rob | 5692 |
| Anna | 3450 |

During the election campaign, it is expected that people may change the candidate that they plan to vote for each week according to the transition diagram.




14 ©VCAA 2011 1MQ5 J The total number of people who are expected to change the candidate that they plan to vote for one week after the election campaign begins is

A 828 B 1423 C 2251 D 4269 E 6891

15 ©VCAA 20111MQ6 J The election campaign will run for ten weeks. If people continue to follow this pattern of changing the candidate they plan to vote for, the expected winner after ten weeks will be

A Rob by about 50 votes. B Rob by about 100 votes. C Rob by fewer than 10 votes.

D Anna by about 100 votes. E Anna by about 200 votes.

16 ©VCAA 2016S1MQ8 A transition matrix, T, and a state matrix, $S_2$, are defined as follows.

$$T- \begin{bmatrix} 0.5 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \end{bmatrix} \quad S_2 = \begin{bmatrix} 300 \\ 200 \\ 100 \end{bmatrix}$$

If $S_2 = TS_t$, the state matrix $S_t$ is

A $\begin{bmatrix} 200 \\ 250 \\ 150 \end{bmatrix}$ B $\begin{bmatrix} 300 \\ 200 \\ 100 \end{bmatrix}$ C $\begin{bmatrix} 300 \\ 0 \\ 300 \end{bmatrix}$ D $\begin{bmatrix} 400 \\ 0 \\ 200 \end{bmatrix}$ E undefined

17 ©VCAA I 2021 1MQ7 The matrix $S_{n+1}$ is determined from the matrix $S_n$ using the recurrence relation $S_{n+1} = TS_n - C$, where

$$T = \begin{bmatrix} 0.6 & 0.1 & 0.3 \\ 0.3 & 0.8 & 0.2 \\ 0.1 & 0.1 & 0.5 \end{bmatrix} \text{» \$0 –} \begin{bmatrix} 21 \\ 51 \\ 31 \end{bmatrix}, \quad S_1 = \begin{bmatrix} 24.0 \\ 54.3 \\ 20.7 \end{bmatrix}$$

and C is a column matrix.

Matrix $S_2$ is equal to

A $\begin{bmatrix} 23.04 \\ 55.78 \\ 16.18 \end{bmatrix}$ B $\begin{bmatrix} 25.34 \\ 56.28 \\ 17.38 \end{bmatrix}$ C $\begin{bmatrix} 26.04 \\ 54.78 \\ 18.18 \end{bmatrix}$ D $\begin{bmatrix} 28.34 \\ 55.28 \\ 19.38 \end{bmatrix}$ E $\begin{bmatrix} 29.04 \\ 53.78 \\ 20.18 \end{bmatrix}$

# Cumulative examination 2

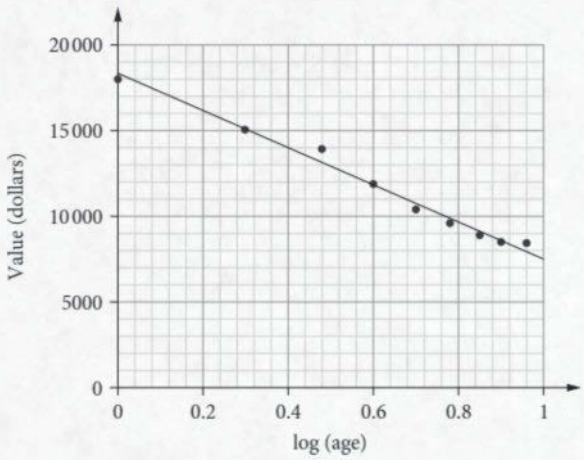
Total number of marks: 33 Reading time: 9 minutes Writing time: 50 minutes

**1** ©VCAA 2005 2CQ1f-h (2 marks) Cars depreciate in value over time. The table gives the average value of a car (of the same brand and model) at different ages. The second row gives the transformed variable, log *(age)*. The table is incomplete.

| Age (years) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| log *(age)* | 0 | 0.30 | 0.48 | 0.60 | 0.70 | 0.78 | 0.85 | 0.90 | |
| Value (dollars) | 18000 | 15050 | 13900 | 11900 | 10400 | 9600 | 8900 | 8500 | 8400 |

a What is the missing value? Write your answer correct to two decimal places. 1 mark

b In the scatterplot, *value* is plotted against log *(age)*. A least squares line of best fit fitted to the transformed data is also drawn.




The equation of this least squares line of best fit is

*value* = 18300 - 10800 x log (age)

Use this equation to predict the value of a car that is three years old. Write your answer correct to the nearest hundred dollars. 1 mark

**2** ©VCAA 2013 2BRMQ4 (2 marks) Hugo took out a reducing balance loan of $25 000 to compete in road races overseas. Interest was charged at a rate of 12% per annum compounding quarterly. His loan is to be repaid fully in four years with equal quarterly payments. After two years, how much of the $25 000 will Hugo have repaid? Write your answer, correct to the nearest dollar.

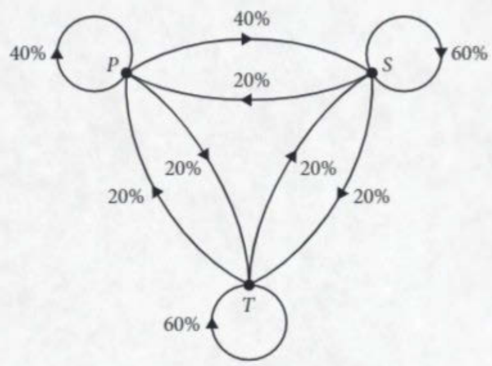
**3** ©VCAA | 2018N2MQF] (2 marks) Five farmers, A, B, C, *D* and E, attended a conference. Pairs of these farmers had previously attended one or more conferences together. The number of conferences previously attended together is shown in matrix *M*. For example, the T' in the bottom row shows that *D* and E had attended one earlier conference together.

$$M = \begin{array}{c} \begin{matrix} A & B & C & D & E \end{matrix} \\ \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 2 & 1 & 3 \\ 1 & 2 & 0 & 1 & 2 \\ 1 & 1 & 1 & 0 & 1 \\ 0 & 3 & 2 & 1 & 0 \end{bmatrix} \end{array} \begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix}$$

a Which two farmers had not previously attended a conference together? 1 mark

b What do the numbers in column *D* indicate? 1 mark

4 © VCAA 2017 2MQ2 J (3 marks) Junior students at a school must choose one elective activity in each of the four terms in 2018. Students can choose from the areas of performance (P), sport (S) and technology (T). The transition diagram shows the way in which junior students are expected to change their choice of elective activity from term to term.




a Of the junior students who choose performance (P) in one term, what percentage are expected to choose sport (S) the next term? 1 mark

Matrix $J_1$ lists the number of junior students who will be in each elective activity in Term 1.

$$J_1 = \begin{bmatrix} 300 \\ 240 \\ 210 \end{bmatrix} \begin{matrix} P \\ S \\ T \end{matrix}$$

b 306 junior students are expected to choose sport (S) in Term 2. Copy and complete the following calculation to show this.

300 x ☐ + 240 x ☐ + 210 x ☐ = 306 1 mark

c In Term 4, how many junior students in total are expected to participate in performance (P) or sport (S) or technology (T)? 1 mark

5 ^^^3~20i9N 2MQ3 J (2 marks) Three television channels, $C_p$ $C_2$ and $C_3$, will broadcast the International Games in the town of Gillen. The basketball finals of the International Games will be televised on channel $C_3$ from 12.00 noon until 4.00 pm. It is expected that 600 Gillen residents will be watching $C_3$ at any time from 12.00 noon until 4.00 pm. The remaining 1400 Gillen residents will not be watching $C_3$ from 12.00 noon until 4.00 pm (represented by *NotC₃*). The transition matrix P shows how the 2000 Gillen residents are expected to change their viewing habits each hour between watching $C_3$ and not watching $C_3$ from 12.00 noon until 4.00 pm.

This hour

$$\begin{matrix} & C_3 & NotC_3 & \\ P = & \begin{bmatrix} v & w \\ 0.35 & x \end{bmatrix} & \begin{matrix} C_3 \\ NotC_3 \end{matrix} & \text{Next hour} \end{matrix}$$

Find the values of v, *w* and x.

6 © VCAA 2021 2MQ1 J (2 marks) Elena imports three brands of olive oil: Carmani (C), Linelli (L) and Ohana (O). The number of 1 litre bottles of these oils sold in January 2021 is shown in matrix *J*.

$$J = \begin{bmatrix} 2800 \\ 1700 \\ 2400 \end{bmatrix} \begin{matrix} C \\ L \\ O \end{matrix}$$

a What is the order of matrix /? 1 mark

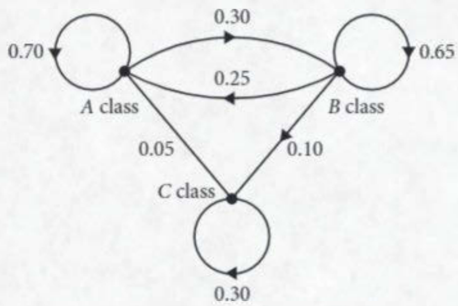
b Elena expected that in February 2021 the sales of all three brands of olive oil would increase by 5%. She multiplied matrix J by a scalar value, k, to determine the expected volume of sales for February. Find *k*. 1 mark

7 ©VCAA 2017N 2MQ2 J (8 marks) The Whiteoak Theatre Club has 200 members who buy tickets for every concert. The members can choose seats from three different classes, A, *B* or C.

For each concert, the choice of seat class for these members can be determined using the transition matrix *T* shown.

$$T = \begin{array}{c} \text{This concert} \\ \begin{array}{ccc} A & B & C \end{array} \\ \left[\begin{array}{ccc} 0.70 & 0.25 & 0.05 \\ 0.30 & 0.65 & 0.65 \\ 0.00 & 0.10 & 0.30 \end{array}\right] \end{array} \begin{array}{l} A \\ B \text{ Next concert} \\ C \end{array}$$

a An incomplete transition diagram for matrix *T* is shown.




Copy and complete the transition diagram by adding all the missing information, 2 marks

b The number of seats in each class chosen by these members for the final concert this year is shown in matrix $S_o$.

$$S_0 = \left[\begin{array}{c} 16 \\ 96 \\ 88 \end{array}\right] \begin{array}{l} A \\ B \\ C \end{array}$$

What percentage of these members chose *A* class seats for the final concert this year? 1 mark

For the first concert next year, some members will choose a different seat class from the seat class that they chose for the final concert this year.

c What percentage of the 200 members are expected to change from *B* class seats at the final concert this year to *A* class seats for the first concert next year? 1 mark

The expected number of these members and their choice of seat class for the nth concert next year can be determined using the recurrence relation

$$S_0 = \left[\begin{array}{c} 16 \\ 96 \\ 88 \end{array}\right], \quad S_{n+1} = TS_n$$

d Write down the state matrix, for the expected number of members and their choice of seat class for the first concert next year. Write your answer correct to one decimal place, 1 mark

e In the long term, how many members would be expected to buy *B* class seats for a concert? 1 mark

f It is expected that, beginning from the third concert next year, the Whiteoak Theatre Club will have more members. Ten new members are expected at every new concert. For their first concert, new members will not be given a seat choice. Matrix $K_2$ contains the expected number of members in each class of seat for the second concert next year. The expected number of members in each class of seat for the third and fourth concerts next year can be determined by

$K, = TK_2 + B$

$K_4\text{-}TK_2 + B$ where?- $\left[\begin{array}{c} 0.70\ 0.25\ 0.05 \\ 0.30\ 0.65\ 0.65 \\ 0.00\ 0.10\ 0.30 \end{array}\right], K_2 = \left[\begin{array}{c} 61 \\ 116 \\ 23 \end{array}\right]$ and $B = \left[\begin{array}{c} 2 \\ 7 \\ 1 \end{array}\right]$

Determine the number of members who are expected to choose *A* class seats for the fourth concert next year. Round your answer to the nearest whole number. 2 marks

8 ©VCAA 2021 2MQ2 MODIFIED] (3 marks) The main computer system in Elenas office has broken down. The five staff members, Alex (A), Brie (B), Chai (C), Dex (D) and Elena (E), are having problems sending information to each other. Matrix $M$ shows the available communication links between the staff members.

$$M = \text{Sender}\ \begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix}\overset{\text{Receiver}}{\overset{\begin{matrix} A & B & C & D & E \end{matrix}}{\begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}}}$$

In this matrix

- the T in row A, column $B$ indicates that Alex can send information to Brie
- the 'O' in row D, column C indicates that Dex cannot send information to Chai.

a Which two staff members can send information directly to each other? 1 mark

b Elena needs to send documents to Chai. What is the sequence of communication links that will successfully get the information from Elena to Chai? 1 mark

c Matrix $M^2$ below is the square of matrix $M$ and shows the number of two-step communication links between each pair of staff members.

$$M^2 = \text{Sender}\ \begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix}\overset{\text{Receiver}}{\overset{\begin{matrix} A & B & C & D & E \end{matrix}}{\begin{bmatrix} 0 & 0 & 1 & 2 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}}}$$

Two pairs of individuals have two different two-step communication links.

List each two-step communication link for each pair. 1 mark

9 (9 marks) A colony of gerbils consists of

- 64 gerbils less than one year old of which 45% will survive to be one
- 96 one-year-old gerbils of which 57% will survive to be two
- 42 two-year-old gerbils of which no gerbils will survive to be three.

The females in the colony have the following average yearly birth rates:

- Female gerbils less than one year old give birth on average to 3.8 gerbils.
- Female one-year-old gerbils give birth on average to 4.8 gerbils.
- Female two-year-old gerbils give birth on average to 2.6 gerbils.

Assume

- half the colony are female and half the births are female
- the survival rates for males and females are the same.

a Explain whywe need to halve the given gerbils birth rates to find the female birth and survival rates. 3 marks

b Copy and complete the female birth and survival rate table.

| Age (years) | 0-<1 | 1-<2 | 2-<3 |
|---|---|---|---|
| Initial number | | | |
| Birth rate | | | |
| Survival rate | | | |

3 marks

CUMULATIVE EXAMINATION 2

c Copy and complete the following calculation from the table that will show the number of female gerbils expected to be born after one year. Round the answer to the nearest whole number.

(---------- x----------- ) + (-----------x----------- ) + (-----------x------------) =----------- 1 mark

d Write the Leslie matrix $L$ that represents the information in the table. 1 mark

e How many gerbils are there in the population after five years? 1 mark



9780170448406

CHAPTER 9

# UNDIRECTED GRAPHS

## Study Design coverage

### AREA OF STUDY 2: DISCRETE MATHEMATICS

#### Graphs and networks

- the concepts, conventions and terminology of graphs including planar graphs and Euler's rule, and directed (digraphs) and networks
- use of matrices to represent graphs, digraphs and networks and their application.

#### Exploring and travelling problems

- the concepts, conventions and notations of walks, trails, paths, cycles and circuits
- Eulerian trails and Eulerian circuits: the conditions for a graph to have an Eulerian trail or an Eulerian circuit, properties and applications
- Hamiltonian paths and cycles: properties and applications.

#### Trees and minimum connector problems

- trees and spanning trees
- minimum spanning trees in a weighed connected graph and their determination by inspection or by Prim's algorithm
- use of minimal spanning trees to solve minimal connector problems.

#### Shortest path problems

- determination of the shortest path between two specified vertices in a graph, digraph or network by inspection
- Dijkstra's algorithm and its use to determine the shortest path between a given vertex and each of the other vertices in a weighted graph or network.

Note: Digraphs are covered in Chapter 10: Directed graphs.



**Video playlists (7):**

**9.1** Graphs and networks
**9.2** Types of graphs
**9.3** Exploring and travelling
**9.4** Graphs and matrices
**9.5** Shortest paths
**9.6** Minimum spanning trees
**VCE question analysis** Undirected graphs

**Worksheets (9):**

**9.2** Planar graphs
**9.3** Eulerian graphs • Eulerian circuits • Eulerian trails and circuits • Hamiltonian paths and cycles
**9.4** Adjacency matrices 1 • Adjacency matrices 2
**9.6** Shortest paths and trees • Minimum spanning trees

9

^Nelson MindTap

To access resources above, visit cengage.com.au/nelsonmindtap


9780170448406

# @ Graphs and networks

9.1

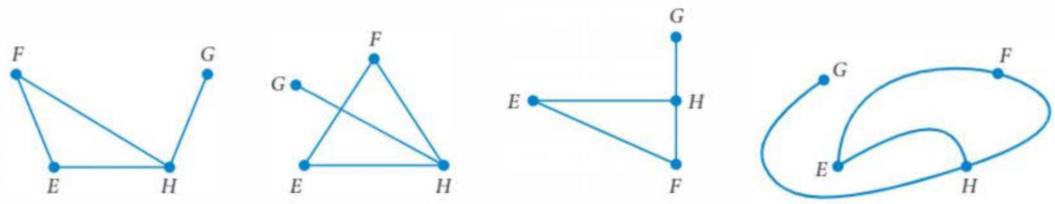
## Graphs, networks and connections

A network is a group of interconnected elements such as people, places or things. A graph shows these connections. In the next two chapters, we will be using the word graph to mean a network diagram. This chapter will only look at undirected graphs, which are graphs where no direction is involved, so if *A* is connected to B, then *B* is also connected to *A*.

## Vertices, edges and isomorphic graphs

A graph consists of points called vertices that are connected by lines called edges. Two vertices that are connected by one or more edges are called adjacent vertices. A vertex is usually labelled by a single capital letter, but it can also appear without a label. Edges are indicated by the two vertices they connect. A graph with vertices E, E, H, *G* and edges EF, EH, *FH* and *GH* is shown below drawn in four different ways.

Video playlist
Graphs and networks




Although these four graphs look different, they show exactly the same network of connections, so they are considered the same graph. In all four of these graphs:

- E and E are connected
- E and *H* are connected
- E and *H* are connected
- *H* and *G* are connected.

**Exam hack**

When deciding whether graphs are isomorphic, look closely at the labels of the vertices and where the edges cross.

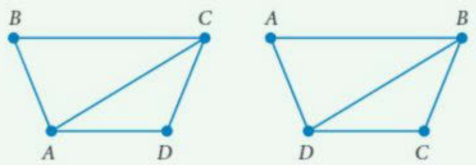
Graphs that show exactly the same connections are called isomorphic graphs.

**WORKED EXAMPLE 1** Identifying isomorphic graphs

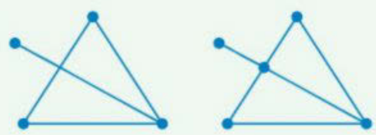
WB
p. 174

For each of the following pairs of graphs, state whether or not they are isomorphic and give a reason for your answer.

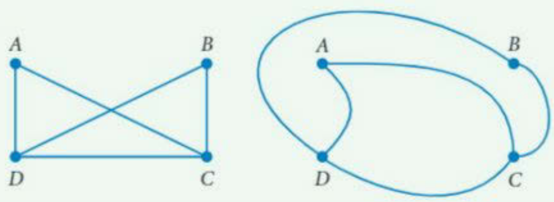
**a**




**b**



**c**

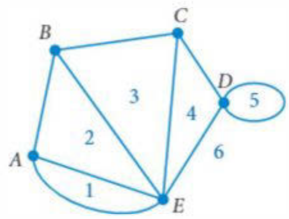


| Steps | Working |
|---|---|
| Do the graphs have the same number of vertices?<br>Do the graphs have the same number of edges?<br>Do the graphs show *exactly* the same connections? | a These two graphs are *not* isomorphic because, although they have the same number of vertices and edges, not all of the connections are the same. In the first graph, *A* and C are connected, but in the second graph, they are not. |
| | b These two graphs are *not* isomorphic because they have a different number of vertices and edges. |
| | c These two graphs are isomorphic because they show exactly the same connections. |

## Features of a graph

Faces are the enclosed regions of a graph. Before you count faces, check that no edges are crossing. If they are, redraw graph with no crossings where possible.

This graph has six faces. Note that




- the loop at D, where an edge starts and ends at the same vertex, creates a face
- the two multiple edges between A and *E* create a face
- the region outside the graph counts as a face.

The degree of a vertex is the number of edges connected to that vertex. For the above graph: *B* has degree 3, *D* has degree 4, and *E* has degree 5. A loop adds 2 to the degree of a vertex because it is connected to the vertex twice.

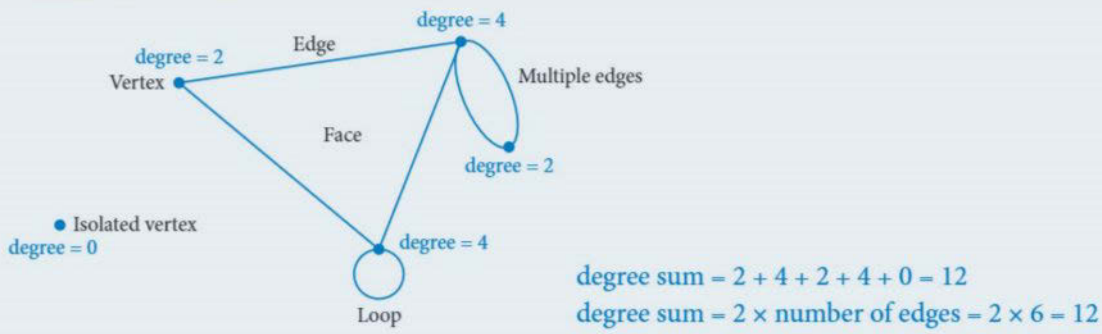
The degree sum of a graph is the sum of the degrees of all the vertices. For any graph,

degree sum = 2 x number of edges

This means the degree sum is always even.

An isolated vertex, a vertex with no edges connected to it, has degree 0.

### Features of a graph



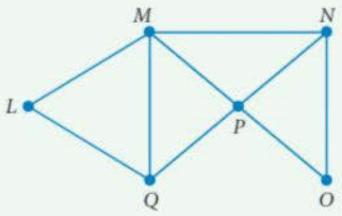
## WORKED EXAMPLE 2 Identifying the features of graphs

For each of the following graphs

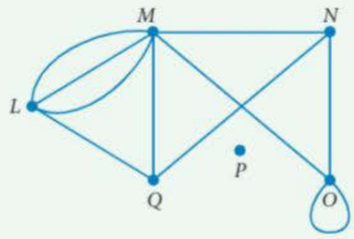
**i** count and list the vertices, edges and faces

**ii** show that the degree sum is twice the number of edges.

**a**




**b**



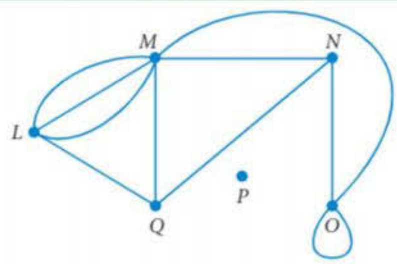

WB p. 175

| Steps | Working |
|---|---|
| **a i** Count and list the number of vertices and edges. | 6 vertices: $L, M, N, O, P, Q$ |
| | 9 edges: $LM, LQ, MQ, MP, MN, NP, NO, PO, PQ$ |
| Count and list the number of enclosed regions plus the region outside the graph. | 5 faces: 4 enclosed and 1 outside the graph |
| **ii** Find the degree of each vertex and then add them. Show that multiplying the number of edges by 2 gives the same result. | see table below |

| Vertex | $L$ | $M$ | $N$ | $O$ | $P$ | $Q$ | Sum |
|---|---|---|---|---|---|---|---|
| Degree | 2 | 4 | 3 | 2 | 4 | 3 | **18** |

degree sum = 2 × number of edges
= 2 × 9
= 18

**b i 1** Redraw the graph to uncross the intersecting edges that have no vertex at the point of intersection.




**2** Count and list the number of vertices and edges.

6 vertices: $L, M, N, O, P, Q$

10 edges: $LM$ × 3, $LQ, MQ, MN, MO, NO, NQ, OO$

Count and list the number of enclosed regions plus the region outside the graph.

7 faces: 6 enclosed and 1 outside the graph.

**ii** Find the degree of each vertex and then add them.

Show that multiplying the number of edges by 2 gives the same result.

| Vertex | $L$ | $M$ | $N$ | $O$ | $P$ | $Q$ | Sum |
|---|---|---|---|---|---|---|---|
| Degree | 4 | 6 | 3 | 4 | 0 | 3 | **20** |

degree sum = 2 × number of edges
= 2 × 10
= 20

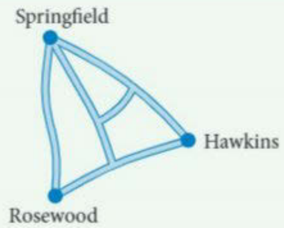
## Applying graphs

Graphs can be used to represent real-life situations such as road systems, maps and friendship networks.
To represent a road system as a network graph we need to identify all the routes between towns. Each route is shown by an edge and each town is shown as a labelled vertex. To represent a road system as a network graph we need to identify all the routes between towns. Each route is shown by an edge and each town is shown as a labelled vertex.

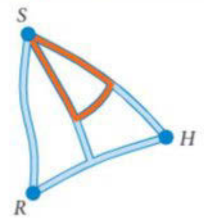
**WORKED EXAMPLE 3** Representing road systems as graphs

WB
p. 176

The road system shows how roads connect the three towns Springfield ($S$), Hawkins ($H$) and Rosewood ($R$). Find all of the routes between the towns that do not go through one of the other towns, and hence, draw a graph representing the road connections.

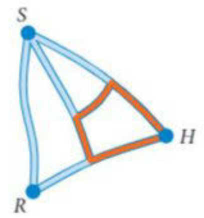



| Steps | Working |
|---|---|
| **1** Find the loops by identifying a route from a town back to the same town that does not go through another town. | There is one route from $S$ to $S$: |




So the graph has a loop at $S$.


There is one route from $H$ to $H$ without going through another town:



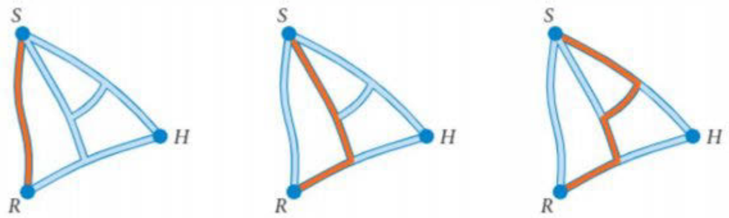

So the graph has a loop at $H$.



9780170448406

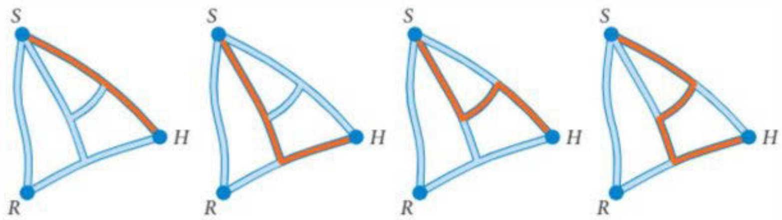
2 Find the edges by identifying routes between two towns that do not go through another town.

There are 3 routes between S and $R$ that do not go through one of the other towns.

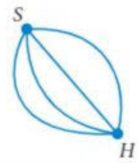



So the graph has 3 multiple edges joining $S$ and $R$.

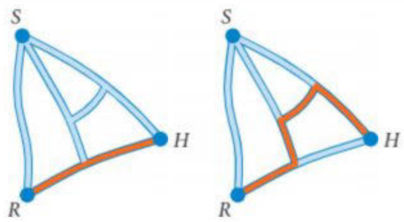

There are 4 routes between S and $H$ that do not go through one of the other towns.




So the graph has 4 multiple edges joining S and $H$.

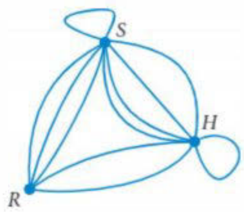



There are 2 routes between $R$ and $H$ that do not go through one of the other towns.




So the graph has 2 multiple edges joining $R$ and $H$.


3 Draw the graph.

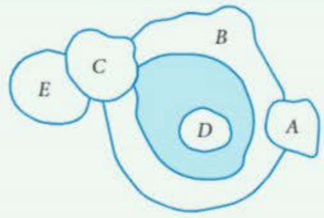


WB

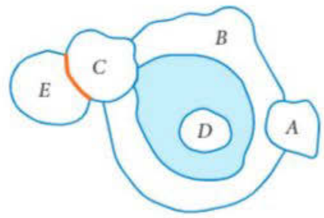
p. 178

## WORKED EXAMPLE 4 Representing maps as graphs

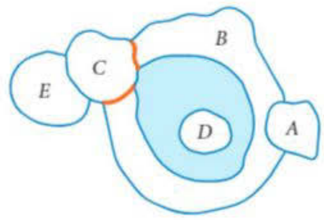
The city of Lakeside is divided into five suburbs labelled as *A* to *E* on the map. A lake in the middle of the city is shown on the map. Draw a graph showing the land connections between the five suburbs.



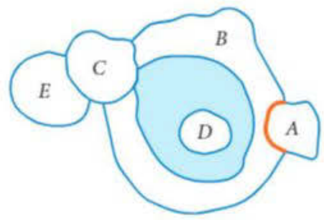

| Steps | Working |
|---|---|
| **1** List all the land connections. | *E* and *C* have 1 land connection: |




*C* and *B* have 2 land connections:

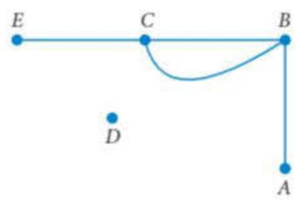



*A* and *B* have 1 land connection:




*D* has no land connections.

**2** Let *A*, *B*, *C*, *D*, *E* be the vertices of the graph. The land connections are the edges. Draw the vertices and connect them with the number of edges.






9780170448406

## EXERCISE 9.1 Graphs and networks

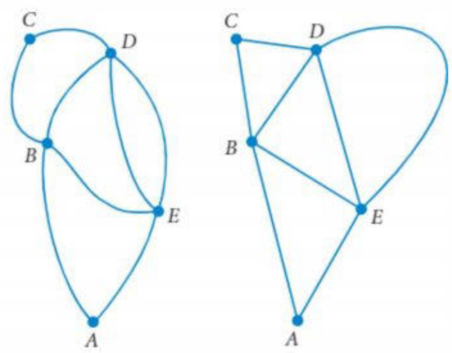
ANSWERS p. 722



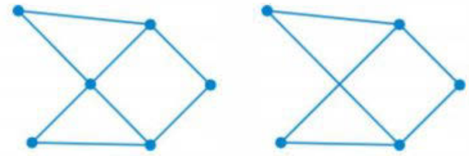
### Mastery

1 EJ WORKED EXAMPLE 1 I For each of the following pairs of graphs, state whether or not they are isomorphic and give a reason for your answer.

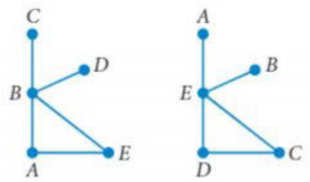
**a**



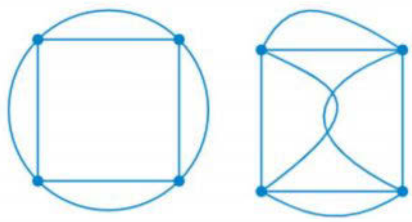

**b**



**c**




**d**

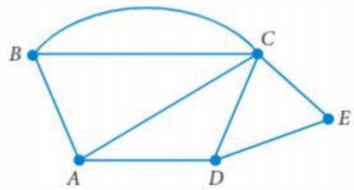


2 H WORKED EXAMPLE 2 I For each of the following graphs

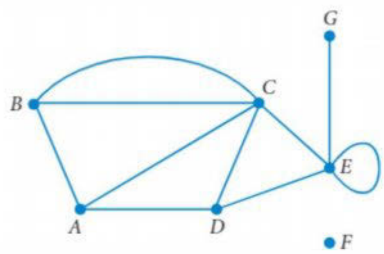
i count and list the vertices, edges and faces

ii show that the degree sum is twice the number of edges.

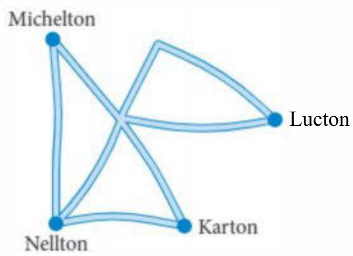
**a**



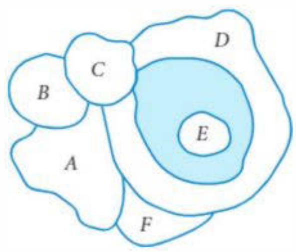

**b**




3 EJ WORKED EXAMPLE 3 I The road system shows how roads connect the four towns Karton (K)» Lucton (L), Michelton *(M)* and Neilton (N). Find all the routes between the towns that do not go through one of the other towns, and hence, draw a graph representing the road connections.

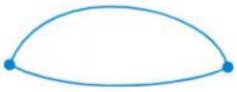
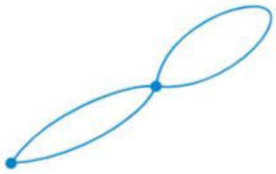


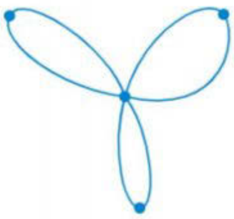
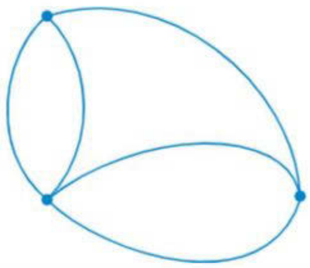

4 E2 WORKED EXAMPLE 4 I The city of Freshwater is divided into six suburbs, labelled as *A* to *F* on the map. A lake in the middle of the city is shown on the map. Draw a network diagram showing the land connections between the six suburbs.



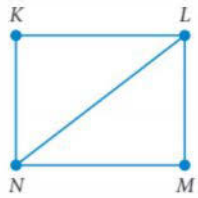
80-100% 60-79% 0-59%

5 ©VCAA 2017 1NQ1 92% Which one of the following graphs contains a loop?

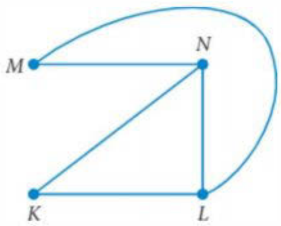
A

B

C

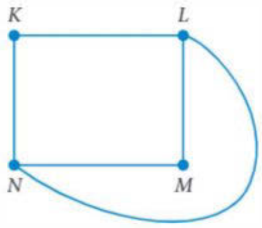
D

E


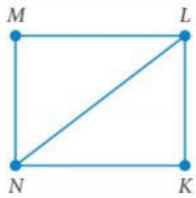
6 ©VCAA | 20151NQ5 92% The graph represents a friendship network. The vertices represent the four people in the friendship network: Kwan (K), Louise (L), Milly *(M)* and Narelle (N). An edge represents the presence of a friendship between a pair of these people. For example, the edge connecting *K* and *L* shows that Kwan and Louise are friends.

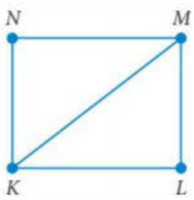


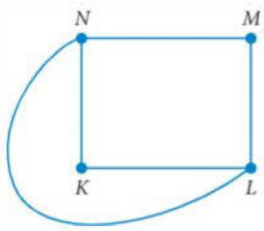

Which one of the following graphs does not contain the same information?

A


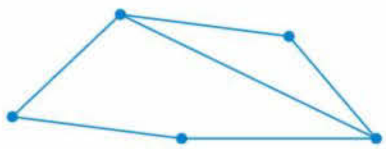
B


C


D


E


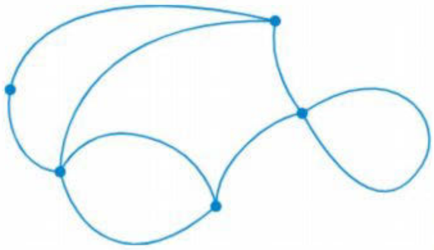

7 ©VCAA 85% In the graph shown the sum of the degrees of the vertices is



A 5 B 6 C 10

D 11 E 12

8 ©VCAA 2019N 1NQ1 The graph has five vertices and eight edges. How many of the vertices in this graph have an even degree?



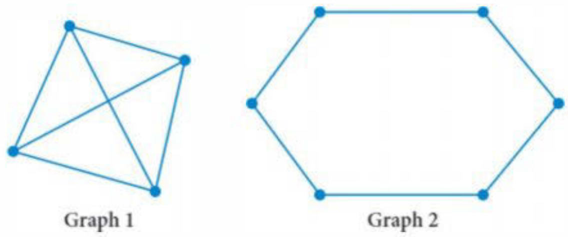
A 0 B 1 C 2

D 3 E 4

 9780170448406

9.1

**9** ©VCAA 20171NQ2 | 83% Two graphs, labelled Graph 1 and Graph 2, are shown.




The sum of the degrees of the vertices of Graph 1 is

A two less than the sum of the degrees of the vertices of Graph 2.

B one less than the sum of the degrees of the vertices of Graph 2.

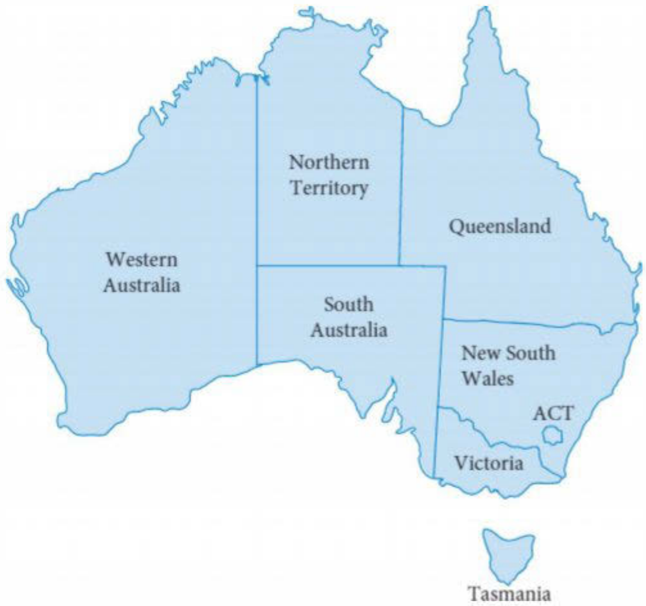
C equal to the sum of the degrees of the vertices of Graph 2.

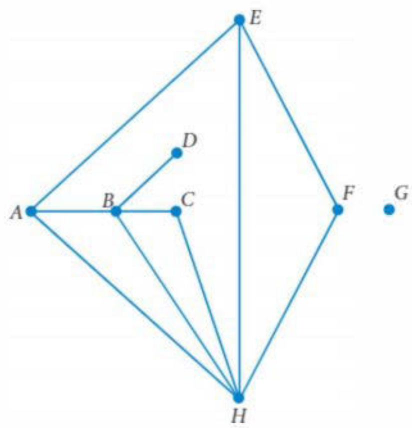
D one more than the sum of the degrees of the vertices of Graph 2.

E two more than the sum of the degrees of the vertices of Graph 2.

*Use the following information to answer the next two questions.*

The map of Australia shows the six states, the Northern Territory and the Australian Capital Territory (ACT). In the network diagram, each of the vertices *A* to *H* represents one of the states or territories shown on the map of Australia. The edges represent a border shared between two states or between a state and territory.



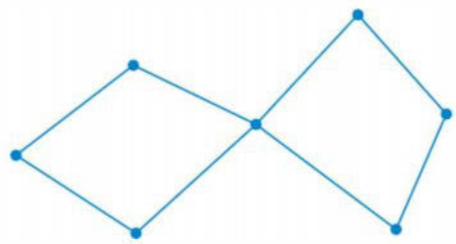

**10** ©VCAA 2011 1NQ3 | 63% In the network diagram, the degree of the vertex that represents the Australian Capital Territory (ACT) is

A 0 B l C 2 D 3 E 4

**11** ©VCAA 2011 1NQ4 | 69% In the network diagram, Queensland is represented by

A vertex *A* B vertex *B* C vertex C D vertex *D* E vertex *E*

▶ 12 ©VCAA 2018NINQ1 , Consider the graph.



Which one of the following statements is not true for this graph?

A This graph has seven vertices.

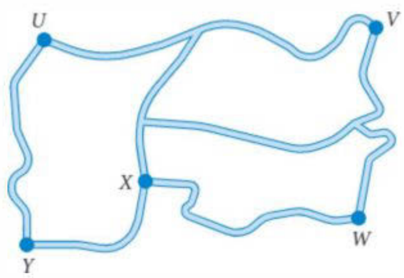
B There are no isolated vertices.

C All vertices have an even degree.

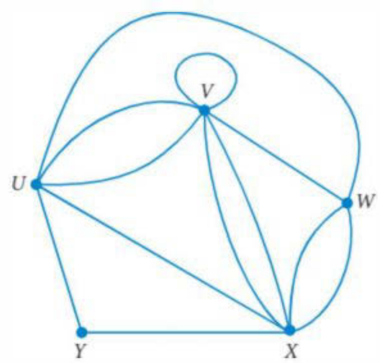
D Six of the vertices have the same degree.

E The sum of the degrees of the vertices is 14.

13 ©VCAA 20151NQ6 59% The map shows all road connections between five towns, *U,* V, W, *X* and *Y.*




A graph, shown here, was constructed to represent this map.




A mistake has been made in constructing this graph. This mistake can be corrected by

A drawing another edge between *V* and W.

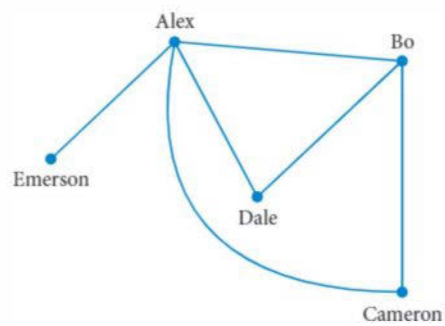
B drawing a loop at *W.*

C removing the loop at *V.*

D removing one edge between *U* and *V.*

E removing one edge between *X* and *V.*

14 ©VCAA 2020 2NQ1 (3 marks) The Sunny Coast Cricket Club has five new players join its team: Alex, Bo, Cameron, Dale and Emerson. The graph shows the players who have played cricket together before joining the team. For example, the edge between Alex and Bo shows that they have previously played cricket together.



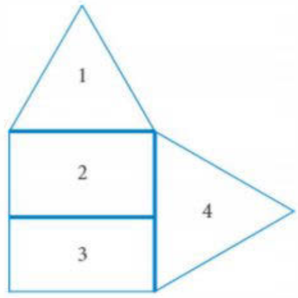

a 95% How many of these players had Emerson played cricket with before joining the team? 1 mark

b 58% Who had played cricket with both Alex and Bo before joining the team? 1 mark

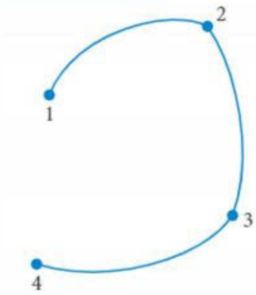
c 91% During the season, another new player, Finn, joined the team.

Finn had not played cricket with any of these players before. Copy the graph and represent this information. 1 mark

**15** ©VCAA 2018N2NQ1 (2 marks) A farmers property is divided into four areas labelled 1 to 4 on the diagram. The bold lines represent the boundary fences between two areas.




In the graph shown, the four areas of the property are represented as vertices. The edges of the graph represent the boundary fences between areas.

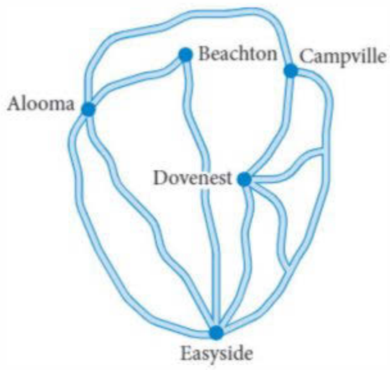



One of the edges is missing from this graph,

a Copy the graph and draw in the missing edge, 1 mark

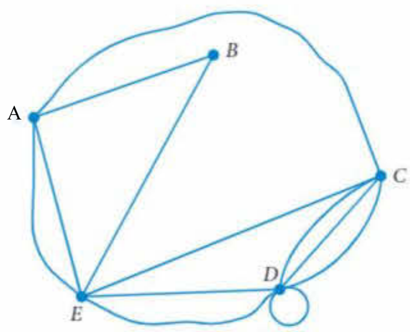
b With this edge drawn in, what is the sum of the degrees of the vertices of the graph? 1 mark

**16** ©VCAA 2016 2NQ1 (3 marks) A map of the roads connecting five suburbs of a city, Alooma (A), Beachton (B), Campville (C), Dovenest (D) and Easyside (E), is shown.




a 83% Starting at Beachton, which two suburbs can be driven to using only one road? 1 mark

b A graph that represents the map of the roads is shown. One of the edges that connects to vertex *E* is missing from the graph.




j 53% Copy the graph and add the missing edge, 1 mark

ii 30% Explain what the loop at *D* represents in terms of a driver who is departing from Dovenest. 1 mark

# @ types of graphs

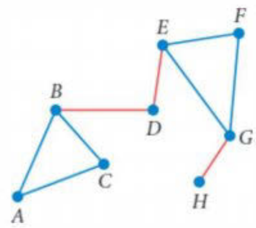
Video playlist Types of graphs

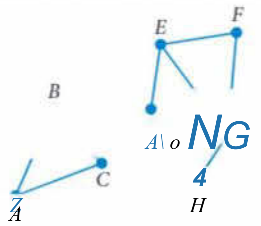
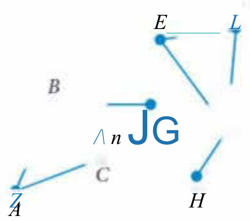
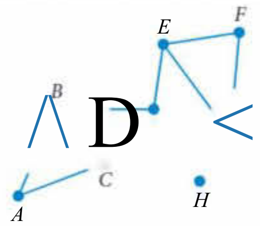
Worksheet Planar graphs

## Connected graphs and bridges

A connected graph is a graph where there is a path from any vertex to any other vertex. A bridge is any edge that keeps a graph connected. So, if you delete a bridge, the graph becomes disconnected.

This is a connected graph that has three bridges BD, *DE* and *GH* shown in red:



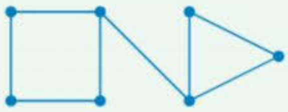

| Removing edge: | *BD* | *DE* | *GH* |
|---|---|---|---|
| Results in a disconnected graph |  |  |  |

WB p. 179

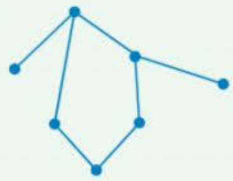
**WORKED EXAMPLE 5** Identifying bridges

How many bridges are in each of these connected graphs? Copy the graphs and indicate the bridges by drawing them in red.

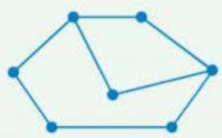
**a**



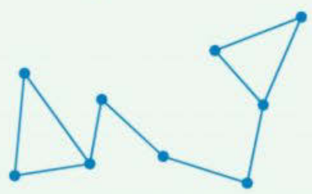
**b**

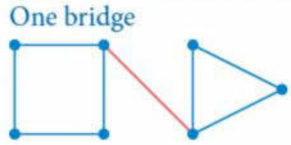
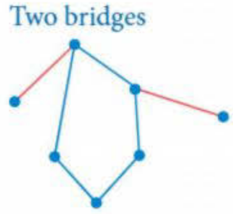
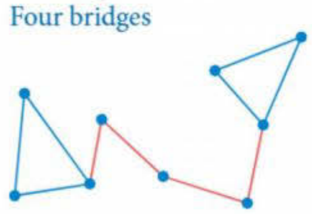


**c**



**d**



| Steps | Working |
|---|---|
| Decide which edge(s) will make the graph disconnected if deleted. | **a** One bridge  **b** Two bridges  **c** No bridges **d** Four bridges  |

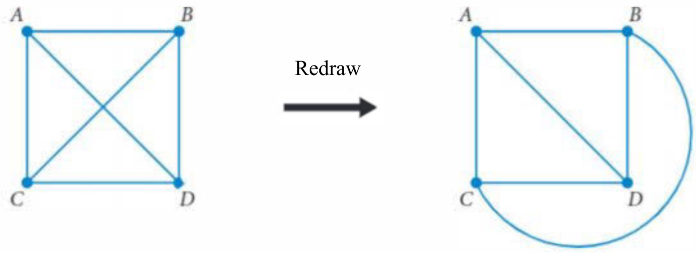


9780170448406

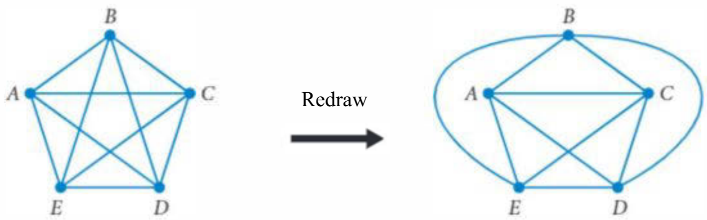
## Planar graphs

Planar graphs are connected graphs that can be drawn so that they don’t have any edges crossing. It doesn’t matter whether it’s actually drawn with crossed edges. What’s important is how it *can* be drawn.

This is a planar graph. Although two edges are crossing, it can be redrawn so that no edges are crossing:




This is *not* a planar graph. It’s possible to redraw it so that *some* of the edges aren’t crossing, but it’s impossible to redraw it with *no* edges crossing:




Trying to uncross the last two edges always results in a different cross.

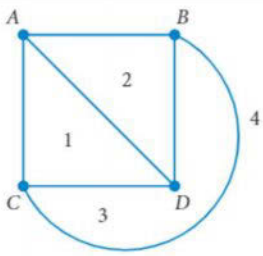
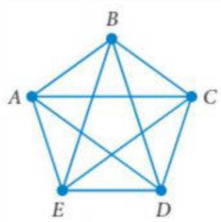
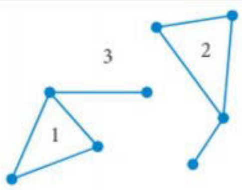
## Euler’s formula

Euler’s (pronounced oilers’) formula applies to graphs that are *both* planar and connected:

number of vertices + number of faces - number of edges = 2

or

$v+f-e=2$

| Graph | Connected/Planar | Euler’s formula |
|---|---|---|
|   | Connected ✓<br>Planar ✓ | $v=4, f=4, e=6$<br>$v+f-e$<br>$=4+4-6$<br>$=2$<br>Euler’s formula works for this graph. |
|   | Connected ✓<br>Planar ×<br>It’s not possible to redraw the graph with all edges uncrossed so it’s not planar. | If we can’t uncross all the edges, we can’t identify all the faces, so Euler’s formula doesn’t work for this graph. |
|   | Connected ×<br>Planar ✓ | $v=8, f=3, e=8$<br>$v+f-e$<br>$=8+3-8$<br>$=3$<br>Euler’s formula doesn’t work for this graph. |

## Euler's formula

For connected planar graphs

$v+f-e=2$

where

$v$ = the number of vertices

$f$ = the number of faces

$e$ = the number of edges.

Note that the formula sheet gives a rearranged version of Eulers formula: $v+f=e+2$.

WB

p. 180

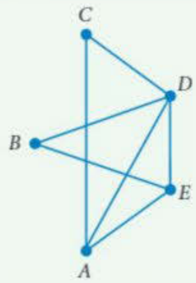
**WORKED EXAMPLE 6** Verifying Euler's formula

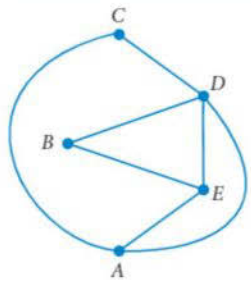
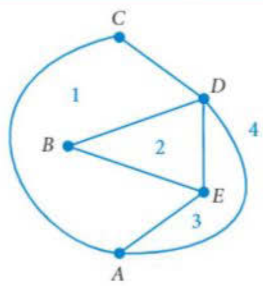
For the graph shown

**a** redraw it to show it is a planar graph

**b** state whether or not it is a connected graph, giving a reason

**c** verify that Euler's formula works or show that it doesn't.




| Steps | Working |
|---|---|
| **a** To uncross the edges, move edges $AC$ and $AD$ around the outside of the graph. |  The graph can be redrawn without any edges crossing, so it is a planar graph. |
| **b** Use the definition of a connected graph. | There is a path from each vertex to every other vertex, so it is a connected graph. |
| **c** Count the number of vertices, faces and edges, and substitute into Euler's formula to see if the result is 2. |  $v=5, f=4, e=7$<br>$v+f-e=5+4-7$<br>$=2$<br>Euler's formula works for this graph. |



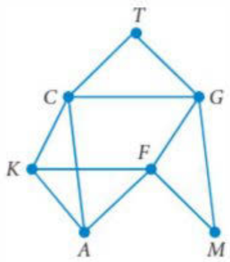
9780170448406

| WORKED EXAMPLE 7 | Using Euler's formula |
|---|---|
| **a** A connected planar graph has 14 edges and 5 faces. How many vertices does it have? | |
| **b** A connected planar graph has 10 vertices and 4 faces. How many edges does this graph have? | |
| **Steps** | **Working** |
| **a** Substitute the known values into Euler's formula and solve to find the number of vertices, $v$. | Substitute $e = 14$ and $f = 5$ into $v + f - e = 2$:<br>$v + 5 - 14 = 2$<br>$v = 2 - 5 + 14$<br>$v = 11$<br>The number of vertices is 11. |
| **b** Substitute the known values into Euler's formula and solve to find the number of edges, $e$. | Substitute $v = 10$ and $f = 4$ into $v + f - e = 2$:<br>$10 + 4 - e = 2$<br>$e = 10 + 4 - 2$<br>$e = 12$<br>The number of edges is 12. |

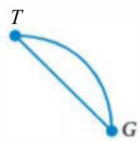
WB p. 181

## Simple and complete graphs

A simple graph is a graph without any loops or multiple edges between adjacent vertices.

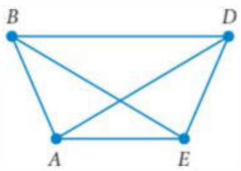



simple graph

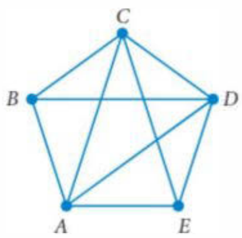



*not* a simple graph because it has two edges between vertices T and G.

A complete graph is a simple graph where every vertex is connected to every other vertex.



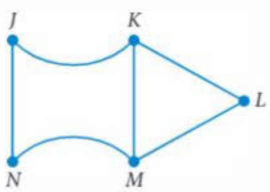

complete graph



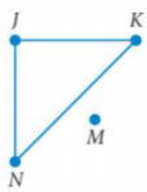

*not* a complete graph because *B* and *E* are not connected

## Subgraphs

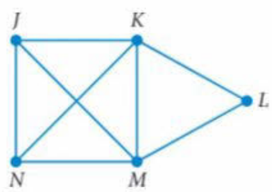
A subgraph is part of a larger graph. It can only have vertices and edges that are in the larger graph.




and




are both subgraphs of




As with all graphs, it doesn't matter whether the edges are drawn curved or straight.

## Simple graphs, complete graphs and subgraphs

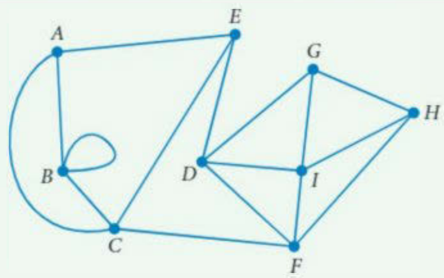
- A simple graph is a graph without any loops or multiple edges.
- A complete graph is a simple graph where every vertex is connected to every other vertex. Every vertex in a complete graph has degree = number of vertices - 1.
- A subgraph is part of a larger graph.

WB
p. 182

### WORKED EXAMPLE 8 Identifying simple graphs, complete graphs and subgraphs

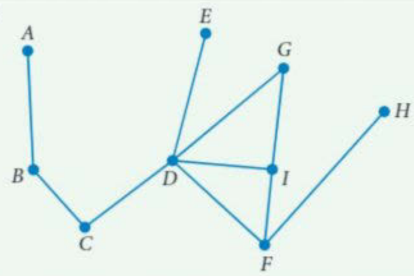
For each of the following graphs state, giving reasons for your answers, whether it is a

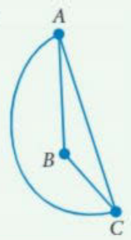
i simple graph
ii complete graph
iii subgraph of

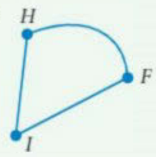



a



b



c



d



| Steps | Working |
|---|---|
| i Does the graph have any loops or multiple edges? | a i It is not a simple graph because it has a loop. |
| ii Is it a simple graph where every vertex is connected by one edge to every other vertex? | ii It is not a complete graph because it is not a simple graph. |
| iii Does the graph contain *only* vertices and edges from the original graph? | iii It is a subgraph because it only has vertices and edges from the larger graph. |
| | b i It is a simple graph because it has no loops or multiple edges. |
| | ii It is not a complete graph because not every vertex is connected by an edge to every other vertex. |
| | iii It is a subgraph because it only has vertices and edges from the larger graph. |
| | c i It is not a simple graph because it has multiple edges connecting A and C. |
| | ii It is not a complete graph because it is not a simple graph. |
| | iii It is not a subgraph because the larger graph has only one edge connecting *A* and C. |
| | d i It is a simple graph because it has no loops or multiple edges. |
| | ii It is a complete graph because every vertex is connected by an edge to every other vertex. |
| | iii It is a subgraph because it only has vertices and edges from the larger graph. |

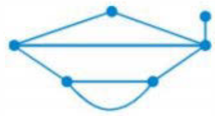

9780170448406

## EXERCISE 9.2 types of graphs

ANSWERS p. 722

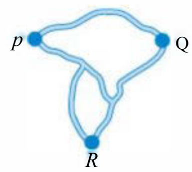
### Recap

1 ©VCAA 2003 1NQ4 62% The sum of the degrees of all the vertices in the network diagram is

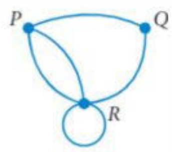


A 6 B 7 **C** 8 D 15 E 16

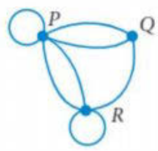
2 ©VCAA | 20131NQ6 I 40% The map shows the road connections between three towns, P, Q and R. The graph that could be used to model these road connections is



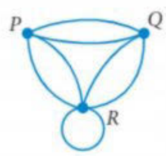

**A**



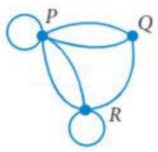

B



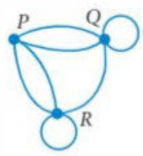

**C**




**D**



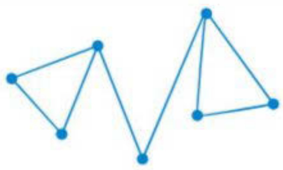

E



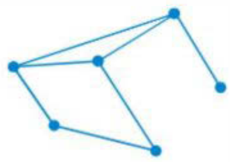

### Mastery

**30** WORKED EXAMPLE 5 How many bridges are in each of these connected graphs? Copy the graphs and indicate the bridges by drawing them in red.

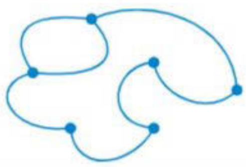
**a**



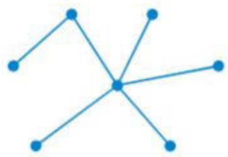
**b**



**c**



**d**



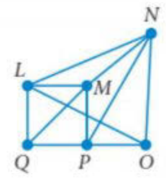
4S WORKED EXAMPLE 6 For each of the following graphs

i redraw it to show it is a planar graph

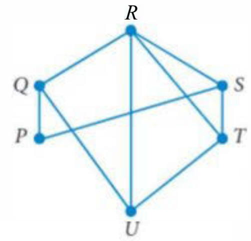
ii state whether or not it is a connected graph, giving a reason

iii verify that Eulers formula works or show that it doesn’t.

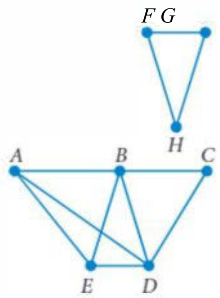
a



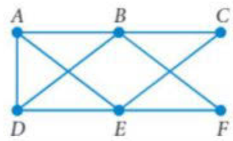

b




c




d




5& WORKED EXAMPLE 7~]

a A connected planar graph has 9 edges and 6 faces. How many vertices does this graph have?

b A connected planar graph has 9 vertices and 7 faces. How many edges does this graph have?

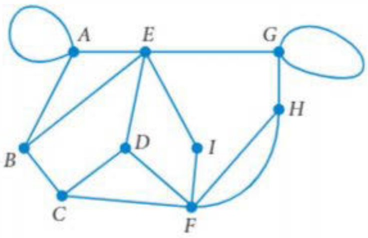
c A connected planar graph has 5 vertices and 7 edges. How many faces does this graph have?

6 S WORKED EXAMPLE 8 For each of the following graphs state, giving reasons for your answers, whether it is a

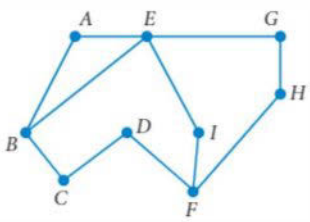
i simple graph

ii complete graph

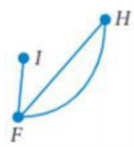
iii subgraph of



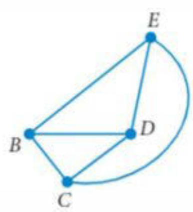

a



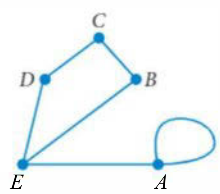

b




c




d






9780170448406

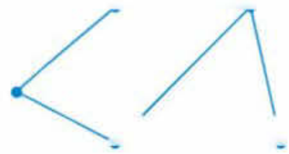
## Exam practice

80-100% 60-79% 0-59%

7 ^^^j20201 NQI I 89% A connected planar graph has seven vertices and nine edges. The number of faces that this graph will have is

A 1 B 2 C 3 D 4 E 5

8 ©VCAA 2010N 1NQ2 Consider the graph.

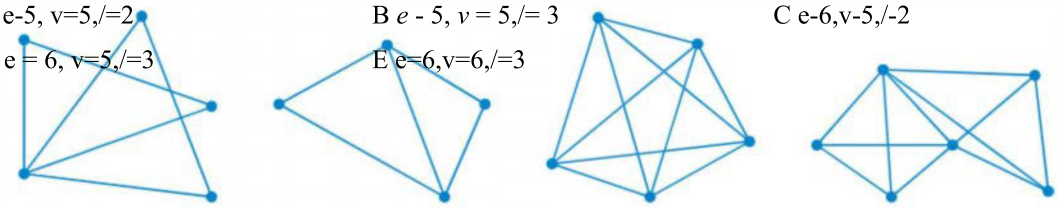


Euler's formula will be verified for this graph. What values of $e$, $v$ and $f$ will be used in this verification?

A $e = 5$, $v = 5$, $f = 2$ B $e = 5$, $v = 5$, $f = 3$ C $e = 6$, $v = 5$, $f = 2$

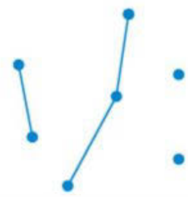
D $e = 6$, $v = 5$, $f = 3$ E $e = 6$, $v = 6$, $f = 3$

9 ©VCAA 2019N 1NQ6 Four graphs are shown below. How many of these graphs are planar?



A 0 B 1 C 2 D 3 E 4

10 ©VCAA 2009 1NQ1 74% Consider the graph. The smallest number of edges that need to be added to make this a connected graph is

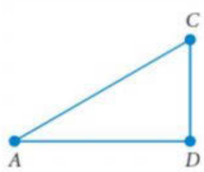


A 1 B 2

C 3 D 4

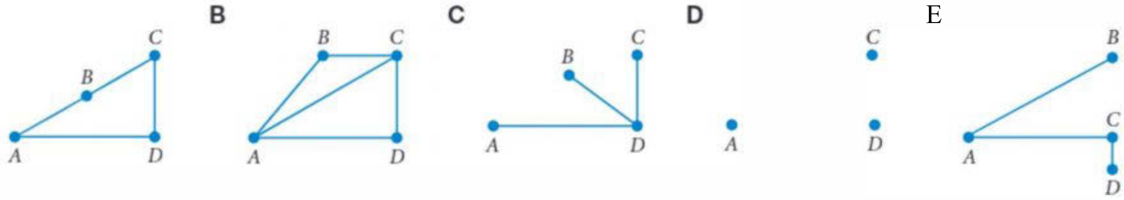
E 5

11 ©VCAA 2008 1NQ2 67%




The graph is a subgraph of which of the following graphs?

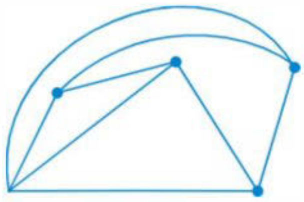
A B C D E




12 ©VCAA 2013 1NQ2 58% The number of edges needed to make a complete graph with four vertices is

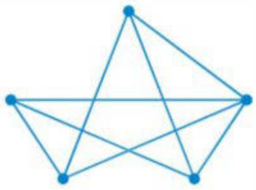
A 2 B 3 C 4 D 5 E 6

13 ©VCAA 20161NQ5 58% Consider the planar graph shown.

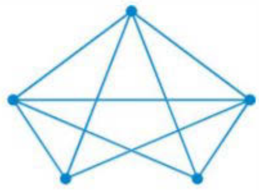


Which one of the following graphs can be redrawn as the planar graph above?

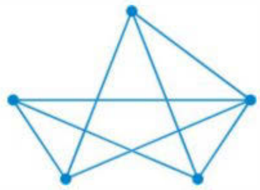
A



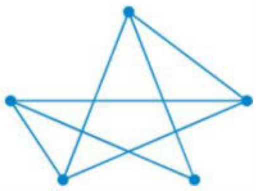
B



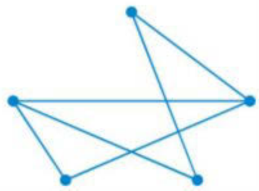
C



D

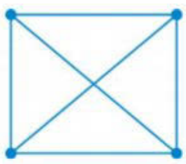


E

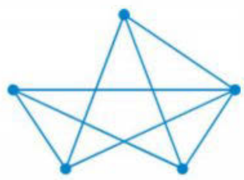


14 ©VCAA 2018 1NQ6 41% Which one of the following graphs is not a planar graph?

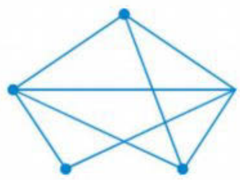
A



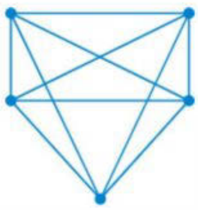
B



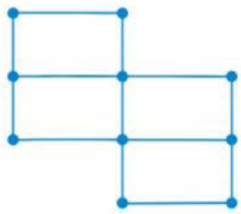
C



D

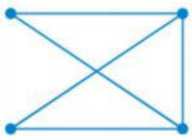


E

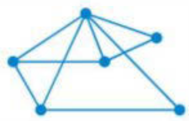


15 ©VCAA 2006 1NQ8 28% Euler's formula, relating vertices, faces and edges, does not apply to which one of the following graphs?

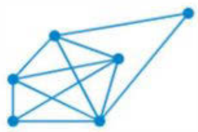
A



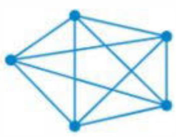
B



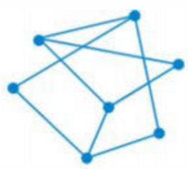
C



D



E



**© Exam hack**

For some multiple-choice questions, go through each option one at a time.



9780170448406

# @ Exploring and travelling

9.3

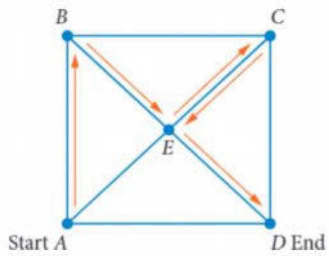
## Types of walks

Graphs can be used to solve travelling and exploring problems where we want to find the shortest route between different locations. These problems involve moving from one vertex to another via an edge and are described by listing the vertices that are visited in order. We need to know the following definitions when dealing with travelling and exploring problems.

A walk is a sequence of connected vertices.

*A-B-E-C-E-D* is an example of a walk in this graph.

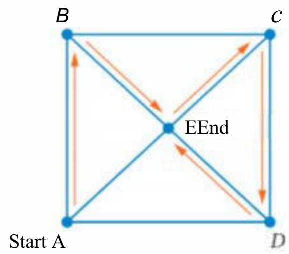
In a walk, edges and vertices can be repeated.
In this example, the edge *EC* is travelled along twice and the vertex *E* is visited twice.




A trail is a walk with no repeated edges.

*A-B-E-C-D-E* is an example of a trail in this graph.

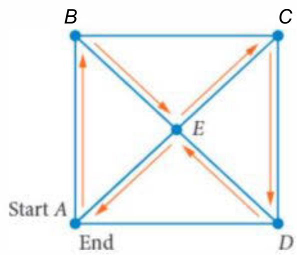
In a trail, vertices can be repeated. In this example, the vertex *E* is visited twice.




A circuit is a walk with no repeated edges that starts and finishes at the same vertex.

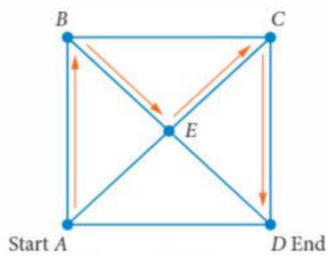
*A-B-E-C-D-E-A* is an example of a circuit in this graph.

In a circuit, vertices can be repeated. In this example, the vertices *A* and *E* are visited twice.



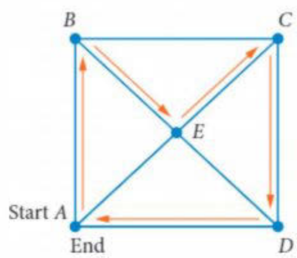

A path is a walk with no repeated vertices.

*A-B-E-C-D* is an example of a path in this graph.




A cycle is a walk with no repeated vertices that starts and finishes at the same vertex.

*A-B-E-C-D-A* is an example of a cycle in this graph.

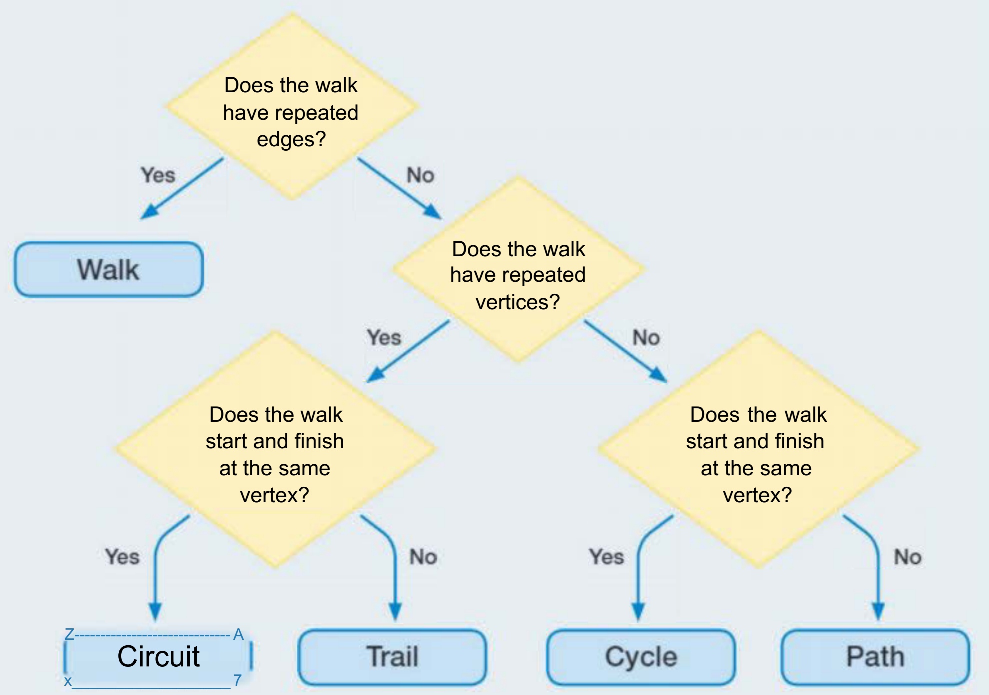



Video playlist
Exploring and travelling

## types of walks

| Walks that: | Trail | Circuit | Path | Cycle |
| --- | --- | --- | --- | --- |
| have no repeated edges | Z | Z | Z | Z |
| have no repeated vertices | | | Z | Z |
| start and finish at the same vertex | | Z | | Z |

## Walk classification chart



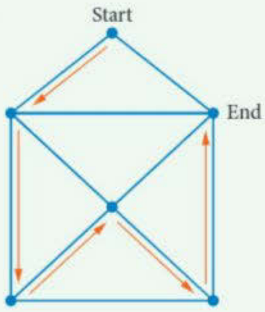



9780170448406

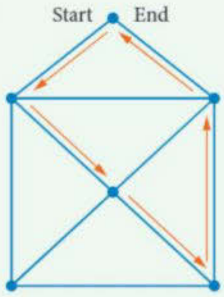
## WORKED EXAMPLE 9 Classifying walks shown on a graph

For each of the following walks, state whether it is a trail, path, circuit, cycle, or walk only and give a reason for your answer.

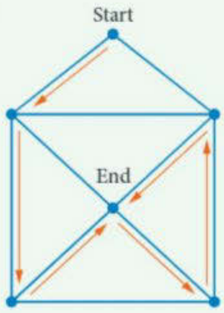
**a**



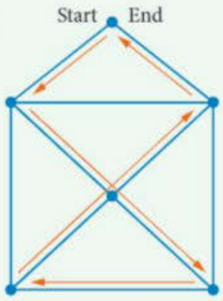

**b**



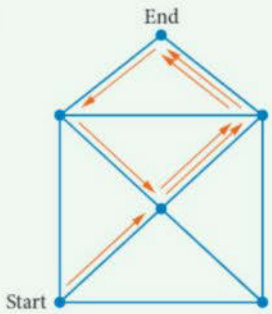

**c**




**d**




**e**




| Steps | Working |
|---|---|
| Use the Walk classification diagram on page 578 and ask the three questions, in this order, for each one:<br>Does the walk have repeated edges?<br>Does the walk have repeated vertices?<br>Does the walk start and finish at the same vertex? | **a** This walk has no repeated edges, no repeated vertices, and does not start and finish at the same vertex, so it's a path.<br>**b** This walk has no repeated edges, no repeated vertices (except the first and last vertex), and starts and finishes at the same vertex, so it's a cycle.<br>**c** This walk has no repeated edges, a repeated vertex, and does not start and finish at the same vertex, so it's a trail.<br>**d** This walk has no repeated edges, a repeated vertex, and starts and finishes at the same vertex, so it's a circuit.<br>**e** This walk has two repeated edges, so it's a walk only. |

9.3

WB p. 183

WB
p. 184

## WORKED EXAMPLE 10 Classifying walks from a list of vertices

Anieka is hiking along tracks in the High Plains National Park. For each of the following walks, state whether it is a trail, path, circuit, cycle, or walk only and give a reason for your answer.


a *C-E-B-A-F-D-C* b *A-F-G-K-H* c *G-K-H-I-J-K-H-G*

d *D-E-C-D-F-G* e *G-K-J-I-H-K-G* f *K-H-I-J-H-G-K*

| Steps | Working |
|---|---|
| Use the Walk classification diagram on page 578 and ask the three questions, in this order, for each one:<br>Does the walk have repeated edges?<br>Does the walk have repeated vertices?<br>Does the walk start and finish at the same vertex? | a This walk has no repeated edges, no repeated vertices (except the first and last vertex), and starts and finishes at the same vertex, so it's a cycle.<br>b This walk has no repeated edges, no repeated vertices, and does not start and finish at the same vertex, so it's a path.<br>c This walk has a repeated edge *KH*, so it's a walk only.<br>d This walk has no repeated edges, a repeated vertex D, and does not start and finish at the same vertex, so it's a trail.<br>e This walk has a repeated edge *(GK* is the same edge as *KG)*, so it's a walk only.<br>f This walk has no repeated edges, a repeated vertex *H*, and starts and finishes at the same vertex, so it's a circuit. |

## Eulerian trails and circuits

Worksheets Eulerian graphs

Eulerian circuits

Eulerian trails and circuits

We often need to consider real-life problems where *every edge* in a graph is included only once.

### Eulerian trail

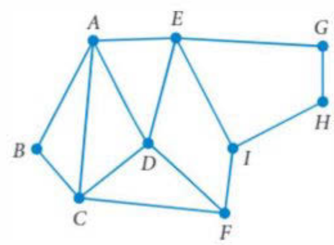
- An Eulerian trail is a walk with no repeated edges that includes every *edge* in a graph.
- An Eulerian trail will only exist if the graph has exactly two vertices of odd degree.
- To find the Eulerian trail, we must start at one of the two vertices of odd degree and finish at the other vertex of odd degree.

### Eulerian circuit

- An Eulerian circuit is a walk with no repeated edges that includes every *edge* in a graph *and* starts and finishes at the same vertex.
- An Eulerian circuit will only exist if the graph has all vertices of even degree.

9.3

For example, this graph has exactly two vertices with odd degrees (F and Z), so an Eulerian trail exists and it will start at *F* and finish at Z, or start at *I* and finish at F. An Eulerian circuit doesn't exist because not all the vertices are of even degree.




### © Exam hack

When a question asks whether an Eulerian trail or Eulerian circuit exists, always find the degree of every vertex to quickly identify whether the correct conditions apply.

## Hamiltonian paths and cycles

We also need to consider real-life problems where *every vertex* in a graph is included only once.

Worksheet Hamiltonian paths and cycles

### Hamiltonian paths and cycles

- A Hamiltonian path is a walk with no repeated vertices that includes every *vertex* in a graph.
- A Hamiltonian cycle is a walk with no repeated vertices that includes every *vertex* in a graph *and* starts and finishes at the same vertex.
- Unlike Eulerian trails and circuits, to find whether a Hamiltonian path or cycle exists, we need to use trial and error.

### Exam hack

To remember the difference between Eulerian and Hamiltonian, remember Eulerian trails and circuits must travel on every Edge, and Hamiltonian paths and cycles are like visiting every House (vertex).

### Walk summary

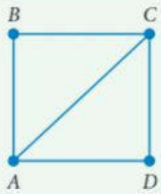
| Walks that: | Trail | Eulerian trail | Circuit | Eulerian circuit | Path | Hamiltonian path | Cycle | Hamiltonian cycle |
|---|---|---|---|---|---|---|---|---|
| have no repeated edges | Z | Z | Z | Z | Z | Z | Z | Z |
| have no repeated vertices | | | | | Z | Z | Z | Z |
| start and finish at the same vertex | | | Z | Z | | | Z | Z |
| include every edge | | Z | | Z | | | | |
| include every vertex | | Z | | Z | | Z | | Z |

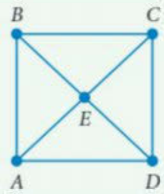
WB

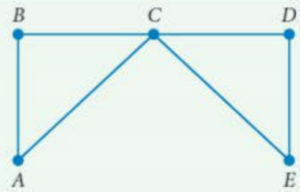
p. 185

## WORKED EXAMPLE 11 Classifying Eulerian and Hamiltonian walks

a For each of these graphs, give a reason why an Eulerian trail, an Eulerian circuit or neither exists and describe three walks for each one that exists.

i



ii



iii


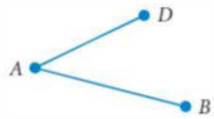

b For each of the graphs in part a, describe a walk that is a Hamiltonian path and a walk that is a Hamiltonian cycle for each one that exists.

| Steps | Working |
|---|---|
| a i 1 Count how many vertices have an odd degree. | Vertex *A* and vertex C are the only two vertices of odd degree. *Exactly two* vertices are of odd degree, so an Eulerian trail exists. |
| 2 An Eulerian trail includes each edge once only and starts and finishes at vertices of odd degree. | Three Eulerian trails are<br>*A-B-C-D-A-C*<br>*C-A-D-C-B-A*<br>*A-C-D-A-B-C*<br>Other answers are possible. |
| ii Count how many vertices have an odd degree. | There are four vertices with odd degrees: A, B, C, *D*.<br>An Eulerian trail exists only if there are exactly two odd vertices and an Eulerian circuit only exists if all the vertices are even, so neither exists for this graph. |
| iii 1 Count how many vertices have an odd degree. | All the vertices have even degree, so an Eulerian circuit exists. |
| 2 An Eulerian circuit includes each edge once only and starts and finishes at the same vertex. | Three Eulerian circuits are<br>*A-B-C-D-E-C-A*<br>*A-C-E-D-C-B-A*<br>*A-B-C-E-D-C-A*<br>Other answers are possible. |
| b A Hamiltonian path includes every vertex in a graph once only. A Hamiltonian cycle also starts and finishes at the same vertex. Trial and error is the only method that can be used. | i Hamiltonian path: *A-B-C-D*<br>Other answers are possible.<br>Hamiltonian cycle: *A-B-C-D-A*<br>Other answers are possible.<br>ii Hamiltonian path: *A-B-C-E-D*<br>Other answers are possible.<br>Hamiltonian cycle: *A-B-C-E-D-A*<br>Other answers are possible.<br>iii Hamiltonian path: *A-B-C-D-E*<br>Other answers are possible.<br>Hamiltonian cycle: Does not exist. |

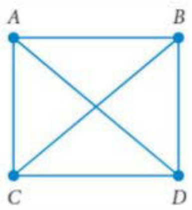
 9780170448406

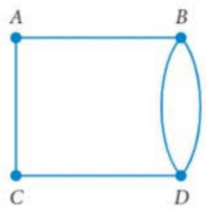
## Recap

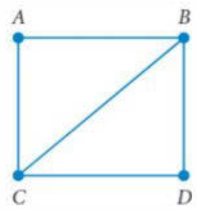
1 ©VCAA 2010 1NQ1 MODIFIED 88% The graph

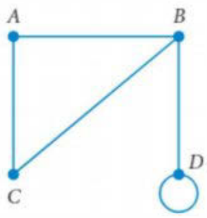


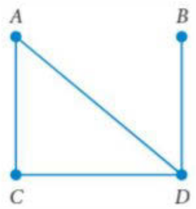

is a subgraph of which one of the following graphs?

a



b



c



d


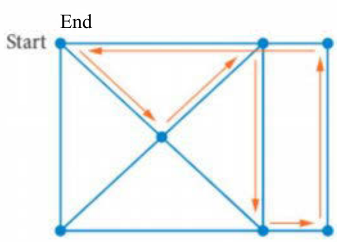

e


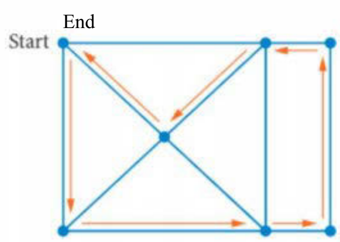

2 ©VCAA 2007 1NQ2 76% A connected planar graph has 12 edges. This graph could have

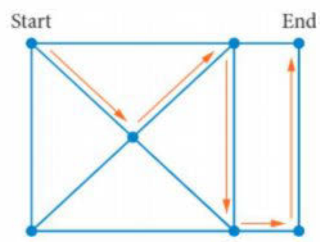
A 5 vertices and 6 faces.
B 5 vertices and 8 faces.
C 6 vertices and 8 faces.
D 6 vertices and 9 faces.
E 7 vertices and 9 faces.

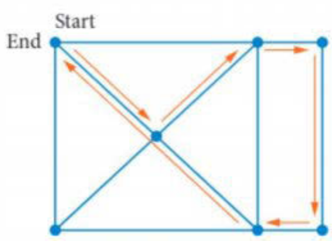
## Mastery

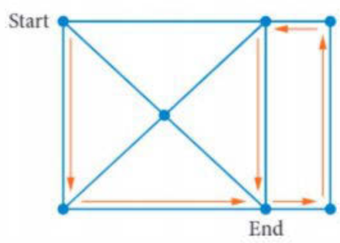
30 WORKED EXAMPLE 9 I For each of the following walks, state whether it is a trail, path, circuit, cycle, or walk only and give a reason for your answer.

a



b


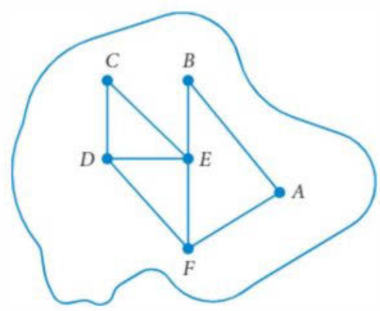

c



d



e


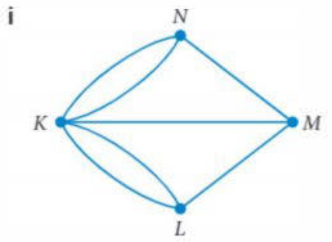

4 S WORKED EXAMPLE 10 I Ahn is trekking along paths on Tropicana Island. For each of the following walks, state whether it is a trail, path, circuit, cycle, or walk only and give a reason for your answer.

a *D-C-E-B-A-F*
b *D-C-E-B-A-F-D*
c *D-C-E-B-A-F-E-D*
d *C-D-E-F-A-B-E-D*
e *F-E-B-A-F-E*
f *C-E-B-A-F-E-D*

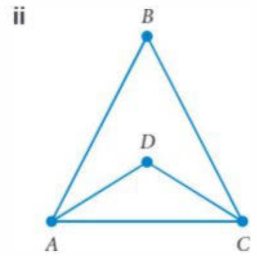


▶ 5 a H WORKED EXAMPLE 11 I For each of these graphs, give a reason why an Eulerian trail, an Eulerian circuit or neither exists and describe three walks for each one that exists.

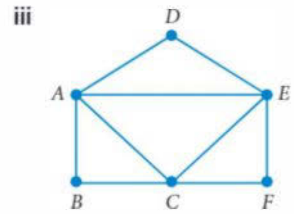
i




ii




iii




b For each of the graphs in part a, describe a walk that is a Hamiltonian path and a walk that is a Hamiltonian cycle for each one that exists.

**Exam practice**

80-100% 60-79% 0-59%

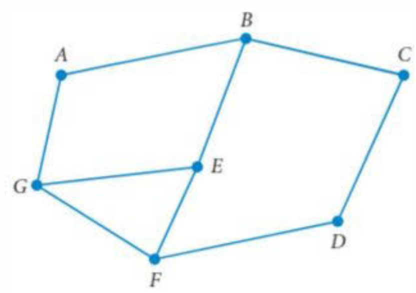
6 ©VCAA 2020 1NQ2 93% Consider the graph.

Which one of the following is not a Hamiltonian cycle for this graph?

A *ABCDFEGA* B *BAGEFDCB*

C *CDFEGABC* D *DCBAGFED*

E *EGABCDFE*

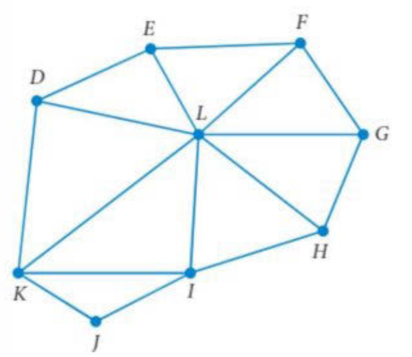



7 ©VCAA | 2008 1NQ3 89% A Hamiltonian cycle for the graph is

A *KJIHGLFEDK* B *DKLIJHGFED*

C *DEFGHIJKD* D *JIKDLHGFE*

E *GHILKJILDEFG*



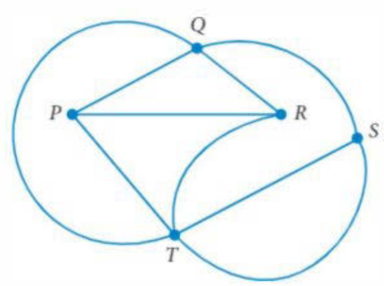

8 ©VCAA 2018 1NQ4 J 88% Consider the graph.

Which one of the following is not a path for this graph?

A *PRQTS* B *PQRTS*

C *PRTSQ* D *PTQSR*

E *PTRQS*




9 © VCAA 2005 1NQ2 65% For the network diagram, an Eulerian trail can be found

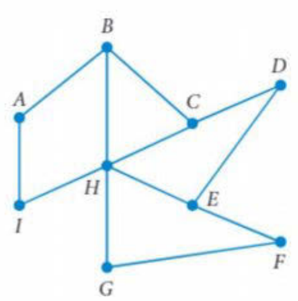
A without altering the network diagram.

B by adding an edge that joins *A* to *H*.

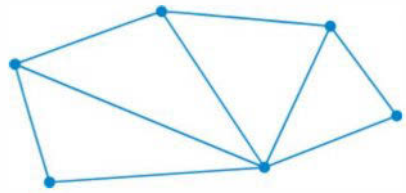
C by adding an edge that joins C to *F*.

D by removing the edge that joins *B* to C.

E by removing the edge that joins *D* to *E*.



10 ©VCAA 20191NQ2 J 64% Consider the graph.

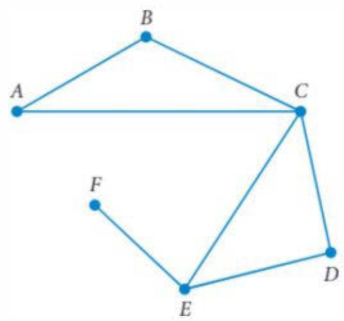


The minimum number of extra edges that are required so that an Eulerian circuit is possible in this graph is

AO Bl C 2

D 3 E 4

11 ©VCAA 2018N1NQ3 I A graph has six vertices and seven edges, as shown in the diagram.




For the graph shown, it is possible to find

A an Eulerian trail and an Eulerian circuit.

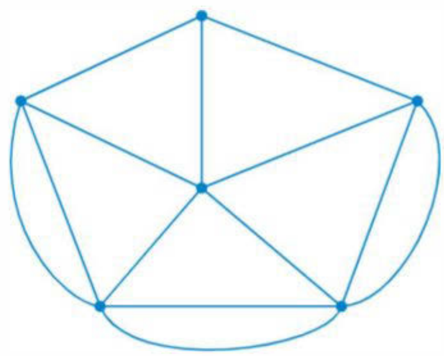
B an Eulerian trail and a Hamiltonian path.

C an Eulerian trail and a Hamiltonian cycle.

D a Hamiltonian path and an Eulerian circuit.

E a Hamiltonian path and a Hamiltonian cycle.

12 ©VCAA 20171NQ6 J 29% An Eulerian trail for the graph shown will be possible if only one edge is removed. In how many different ways could this be done?



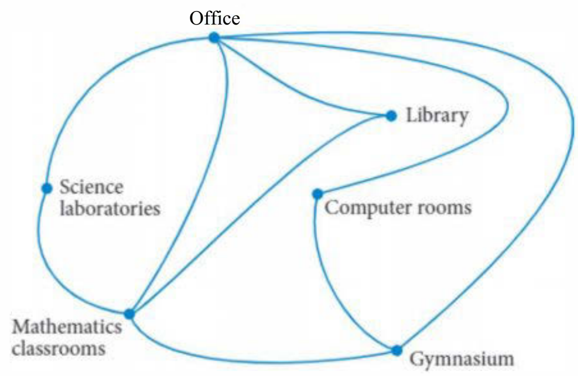
A 1 B 2

C 3 D 4

E 5

13 ©VCAA [2019 2NQ1 J (3 marks)

Fencedale High School has six buildings. The network shows these buildings represented by vertices. The edges of the network represent the paths between the buildings.

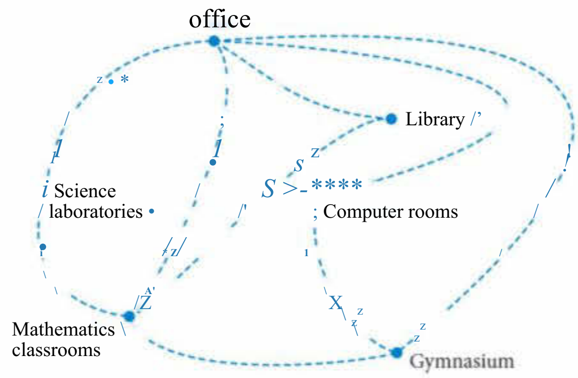



a 96% Which building in the school can be reached directly from all other buildings? 1 mark

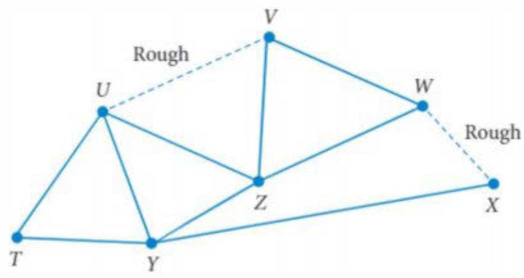
b A school tour is to start and finish at the office, visiting each building only once.

i 70% What is the mathematical term for this route? 1 mark

jj 93% Copy the diagram and draw in a possible route for this school tour. 1 mark



▶ 14 ©VCAA 2016 2NQ2 , (3 marks) The suburb of Alooma has a skateboard park with seven ramps. The ramps are shown as vertices T, *U,* V, W, *X*> *Y* and Z on the graph.




The tracks between ramps *U* and *V* and between ramps W and *X* are rough, as shown on the graph.

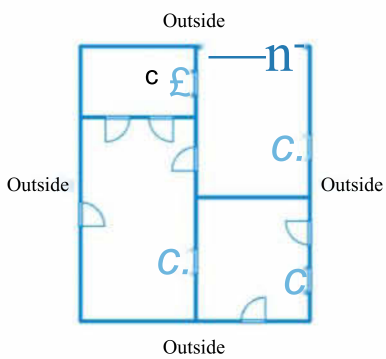
a 72% Nathan begins skating at ramp W and follows an Eulerian trail. At which ramp does Nathan finish? 1 mark

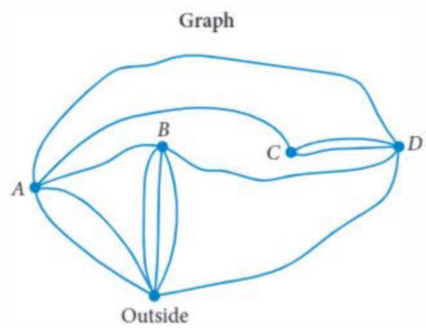
b 77% Zoe begins skating at ramp *X* and follows a Hamiltonian path. The path she chooses does not include the two rough tracks. Write down a path that Zoe could take from start to finish. 1 mark

c 31% Birra can skate over any of the tracks, including the rough tracks. He begins skating at ramp *X* and will complete a Hamiltonian cycle. In how many ways could he do this? 1 mark

15 ©VCAA | 2017N2NQ2 I (4 marks) Simons holiday home has four rooms, A, B, C and *D.* The floor plan shows these rooms and the outside area. There are 12 doors, as shown on the floor plan. Only room C and the outside are labelled. A graph of the floor plan is also shown. On this graph, vertices represent the rooms and the outside area, and edges represent the doors.

Floor plan




a Room C has already been labelled on the floor plan. Use the letters A, *B* and *D* to label the other three rooms on the floor plan. 1 mark

b Simon is in room C and his daughter Zofia is outside. Simon calls Zofia to see him in room C. Zofia visits every other room once on her way to room C. Give the mathematical term that describes Zofias journey. 1 mark

c Simon tries to find a route that passes through every door once only and finishes back at the starting point.

i Explain why this is not possible. Refer to the graph in your answer. 1 mark

ii If two of the doors are locked and only the other doors are considered, then Simons route will be possible. Simon locks the door between room A and room C. Write down the two rooms that are joined by the other door that must be locked. 1 mark

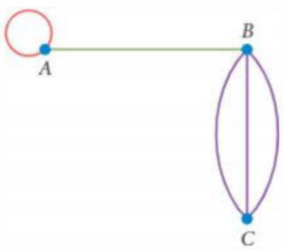
# @ Graphs and matrices

9.4

## Adjacency matrices

An adjacency matrix shows the number of edges between vertices on a graph. For example:

**Graph**




Number of vertices = 3

Number of edges = 5

Adjacency matrix

$$\begin{array}{c} \\ A \\ B \\ C \end{array}\begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 3 \\ 0 & 3 & 0 \end{bmatrix} \end{array}$$

Number of rows = 3

Description

- one loop at A to A
- no loops at *B* to *B* and C to C
- one edge between A and *B*
- three edges between *B* and C
- no edges between A and C

Video playlist Graphs and matrices

Worksheets Adjacency matrices 1

Adjacency matrices 2

Comparing the graph and the adjacency matrix:

Number of vertices in the graph
= number of rows in the matrix
= 3

Number of edges in the graph
= sum of elements on and below the leading diagonal in the matrix
= 1 + 1 + 3
= 5

$$\begin{array}{c} \\ A \\ B \\ C \end{array}\begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 3 \\ 0 & 3 & 0 \end{bmatrix} \end{array}$$

Degree of a vertex
= sum of the elements of the row (+ 1 for each loop)

The degree of vertex A is 3.

The degree of vertex *B* is 4.

The degree of vertex C is 3.

| | *A* | *B* | *C* | Degree |
|---|---|---|---|---|
| A | 1 | 1 | 0 | 1 + 1+1 = 3 |
| *B* | 1 | 0 | 3 | 1+3 = 4 |
| *C* | 0 | 3 | 0 | 3 |

**Adjacency matrices**

An adjacency matrix is a square matrix representing a graph where

- each row and column is labelled as a vertex
- the elements show the number of edges between vertices
- the elements are symmetric about its leading diagonal
- the leading diagonal shows the number of loops.

Number of vertices = number of rows

Number of edges = sum of the elements on and below the leading diagonal

Degree of a vertex = sum of the elements of the row (+ 1 for each loop)

WB

p. 186

## WORKED EXAMPLE 12 Representing graphs using adjacency matrices

Represent the graph using an adjacency matrix.

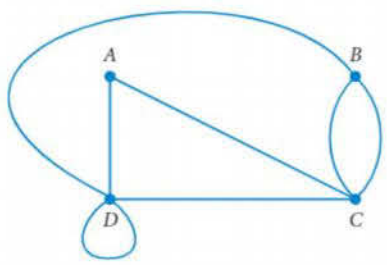

| Steps | Working |
|---|---|
| 1 Label the rows and columns of the matrix to match the graph. | *P Q R S*<br>*p*<br>*Q*<br>*R*<br>*S* [ ] |
| 2 List the connections in terms of the number of edges between vertices. | Two edges between *P* and Q<br>One edge between *P* and *R*<br>One edge between *P* and S<br>One edge between *Q* and *R*<br>One loop from *R* to R |
| 3 Fill in the matrix based on the number of edges.<br>An edge from *A* to *B* is also an edge from *B* to $A_y$ so the matrix is symmetrical around the leading diagonal. | *P Q R S*<br>*P* 2 1 1<br>Q 2 1<br>*R* 1 1 1<br>S 1 |
| 4 Complete the matrix by writing 0 for all the remaining elements. | *P Q R S*<br>*p* 0 2 11<br>*Q* 2 0 10<br>*R* 1110<br>*S* 10 0 0 |


9780170448406

**WORKED EXAMPLE 13** Using adjacency matrices

For the adjacency matrix representing a planar graph:

$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array}\begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 2 & 1 \\ 1 & 2 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \end{array}$$

| Steps | Working |
|---|---|
| **a** how do we know it represents a connected graph? | |
| Check that every vertex is connected to at least one other vertex. | $D$ is connected to $A$, $B$ and $C$, so every vertex is connected to at least one other vertex. |
| **b** find the degree of each vertex and hence, find the degree sum of the graph. | |
| **1** Sum the rows to find the degree of each vertex. Add an extra degree when there's a loop at the vertex. | $\begin{array}{c} \\ A \\ B \\ C \\ D \end{array}\begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 2 & 1 \\ 1 & 2 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ 2 \\ 3 \\ 4 \\ 4+1=5 \end{array}$ (Add 1 because of the loop at $D$.) |
| **2** Find the sum of the degrees of each vertex. | degree sum $= 2+3+4+5$ $= 14$ |
| **c** does the graph have an Eulerian trail? Give a reason for your answer. | |
| A graph has an Eulerian trail if it has exactly 2 vertices of odd degree. | $A$ and $C$ have even degrees. $B$ and $D$ have odd degrees. The graph has an Eulerian trail because it has exactly 2 vertices with odd degrees. |
| **d** how many faces does the graph have? | |
| **1** Find the number of vertices and edges. Number of vertices = number of rows. Number of edges = sum of elements on and below the leading diagonal. | There are 4 vertices: $A$, $B$, $C$, $D$. There are 7 edges. $\begin{array}{c} \\ A \\ B \\ C \\ D \end{array}\begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 2 & 1 \\ 1 & 2 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \end{array}$ |
| **2** This is a connected planar graph, so we can use Euler's formula to find the number of faces. | $v=4, e=7, f=?$ $v+f-e=2$ $4+f-7=2$ $f=5$ The graph has 5 faces. |
| **e** draw the graph without any crossed edges. | |
| Place the four vertices. Use the elements on and below the leading diagonal to draw the edges of the graph. | The edges are $CA$, $CB \times 2$, $DA$, $DB$, $DC$, $DD$  Other versions are possible. |



9.4

WB

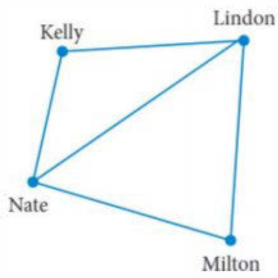
p. 187

## EXERCISE 9.4 Graphs and matrices

ANSWERS p. 724

### Recap

1 ©VCAA 2016S1NQ1 The graph shows the roads connecting four towns: Kelly, Lindon, Milton and Nate.



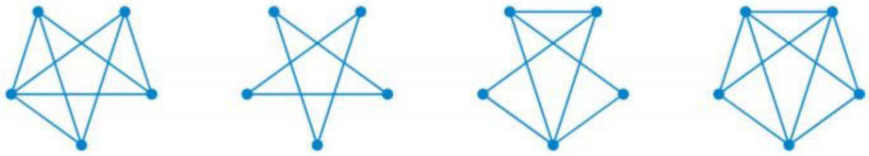

A bus starts at Kelly, travels through Nate and Lindon, then stops when it reaches Milton. The mathematical term for this route is

A a loop. B an Eulerian trail.

C an Eulerian circuit. D a Hamiltonian path.

E a Hamiltonian cycle.

2 ©VCAA 2016S 1NQ5 MODIFIED How many of the four graphs shown have an Eulerian trail?

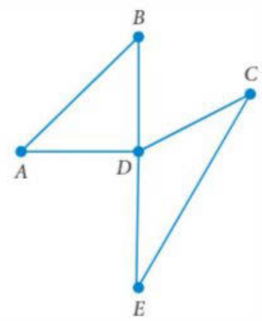


A 0 **B** 1 **C** 2 **D** 3 **E** 4

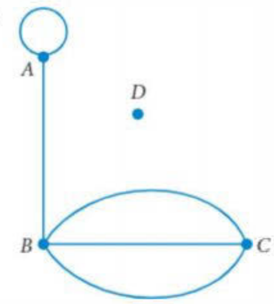
### Mastery

3 WORKED EXAMPLE 12 Represent each graph using an adjacency matrix.

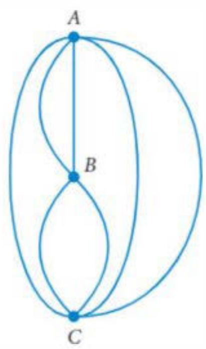
**a**




**b**




**c**




4 WORKED EXAMPLE 13 For each of the following adjacency matrices representing planar graphs:

**a**

$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \end{array}$$

b

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array} \begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 0 & 0 & 1 & 0 & 3 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 3 & 0 & 0 & 1 & 0 \end{bmatrix} \end{array}$$

i how do we know it represents a connected graph?

ii find the degree of each vertex and hence, find the degree sum of the graph.

iii does the graph have an Eulerian trail? Give a reason for your answer.

iv how many faces does the graph have?

v draw the graph without any crossed edges.

## Exam practice

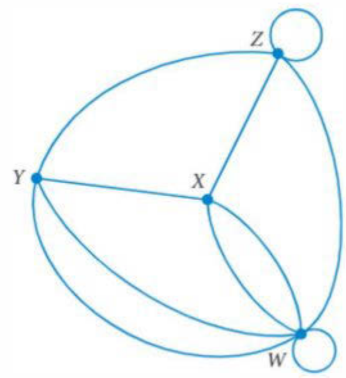
80-100% 60-79% 0-59%

9.4

5 ©VCAA 20171NQ3 I 92% Consider the graph.

The adjacency matrix for this graph, with some elements missing, is shown.

| | W | X | Y | Z |
|---|---|---|---|---|
| W | 1 | _ | _ | _ |
| X | _ | 0 | _ | _ |
| Y | _ | _ | 0 | _ |
| Z | _ | _ | _ | 1 |




This adjacency matrix contains 16 elements when complete. Of the 12 missing elements

A eight are T and four are‘2’.

B four are T and eight are‘2’.

C six are T ’ and six are ‘2’.

D two are ‘O’, six are T ’ and four are ‘2’.

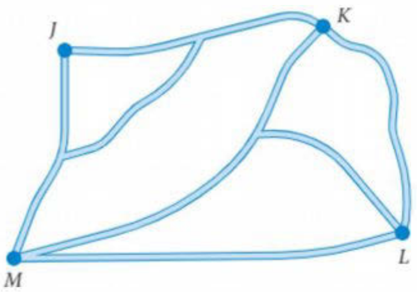
E four are ‘O’, four are T’ and four are ‘2’.

6 © VCAA 20201 NQS 75% The adjacency matrix below shows the number of pathway connections between four landmarks: /, K, *L* and *M*.

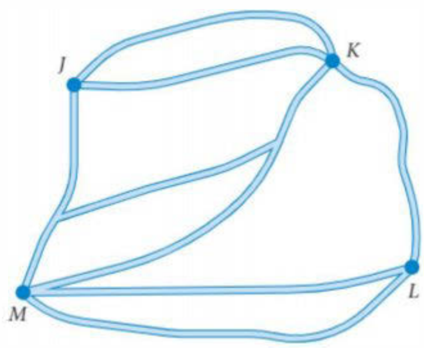
| | J K L M | | |
|---|---|---|---|
| I | 13 | 0 | 2 |
| K | 3 | 0 | 12 |
| L | 0 | 10 | 2 |
| M | 2 2 2 0 | | |

A network of pathways that could be represented by the adjacency matrix is

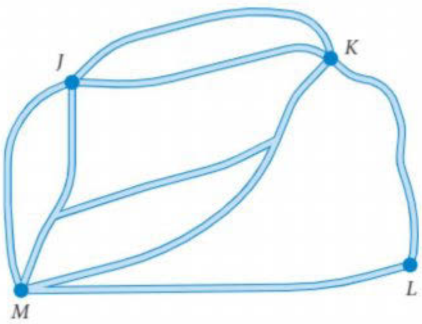
**A**



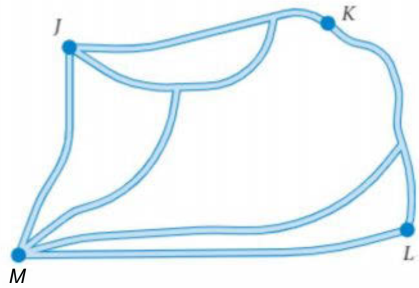

**B**



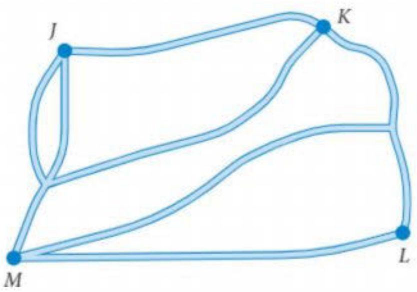

**C**



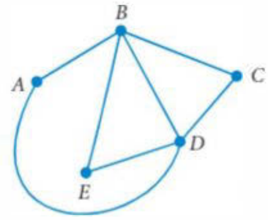

**D**




**E**



▶ 7 ©VCAA 2018N1NQ5 The friendships between five children are summarised in the graph.




The vertices A, B, C, *D* and *E* in the graph represent these children. Each edge between two vertices indicates that the two children are friends. For example, the edge between vertex *B* and vertex C shows that child *B* and child C are friends. An adjacency matrix that summarises the friendships can also be constructed. In this matrix, a ‘0’ indicates no friendship and a‘l’ indicates a friendship. How many zero elements will this adjacency matrix have?

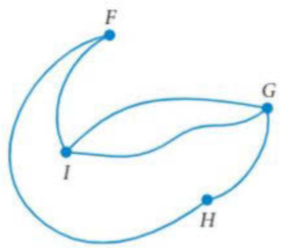
A 7 B 8 C 9 DIO E 11

8 ©VCAA 2017N 1NQ4 The matrix shows the number of road connections between towns F, G, *H* and *I.*

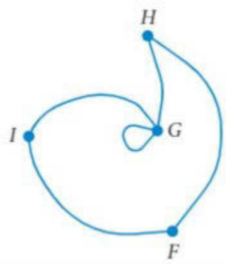
| | *F* | *G* | *H* | *I* |
|---|---|---|---|---|
| *F* | 0 | 0 | 11 | |
| *G* | 0 112 | | | |
| *H* | 110 0 | | | |
| *I* | 12 0 0 | | | |

Which one of the following graphs shows all of these road connections?

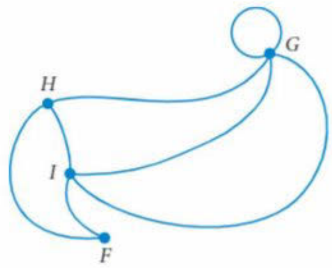
**A**



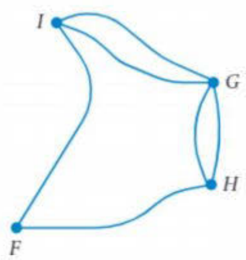

**B**



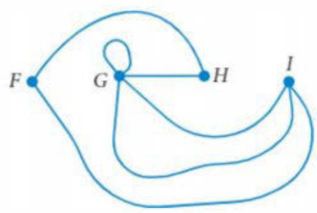

**C**




**D**




**E**




9 © VCAA 20121NQ4 63% The adjacency matrix represents a planar graph with four vertices.

| | *P* | *Q* | *R* | *S* |
|---|---|---|---|---|
| *P* | *0* | *0* | *2* | *1* |
| *Q* | 0 0 11 | | | |
| *R* | 2 10 1 | | | |
| *S* | 1110 | | | |

The number of faces (regions) on the planar graph is

Al B 2 C 3

D 4 E 5

10 ©VCAA 2003 1NQ9 MODIFIED 60% Five graphs are each represented by an adjacency matrix as shown below.

*M* =

| | *A B C D* |
|---|---|
| *A* | *0 2 10* |
| *B* | *2 0 2 0* |
| *C* | 12 0 0 |
| *D* | 0 0 0 2 |

*N* = 

| | *A B C D* |
|---|---|
| *A* | *0 0 2 0* |
| I | 0 0 12 |
| C | 2 10 0 |
| *D* | 0 2 0 0 |

*0* =

| | *A B C D* |
|---|---|
| *A* | *0 0 2 0* |
| *B* | *0 0 0 2* |
| *c* | *2 0 0 0* |
| *D* | *0 2 0 2* |

*P* =

| | *A* | *B* | C | *D* |
|---|---|---|---|---|
| *A* | 0 | 0 | 2 | *0* |
| *B* | 0 | 0 | 1 | *0* |
| *C* | 2 | 1 | 0 | 0 |
| *D* | 0 | 0 | 0 | 2 |

Q =

| | *A B C D* |
|---|---|
| *A* | *0 2 0 0* |
| * | *2 2 0 0* |
| *C* | 0 0 2 1 |
| *D* | 0 0 10 |

Which adjacency matrix represents a *connected* graph?

AM *BN* *CO* *DP* *E Q* ▶

11 ©VCAA 2019N 1NQ7 I A graph has five vertices, A, B, C, *D* and *E*.

The adjacency matrix for this graph is shown.

|   | A | B | C | D | E |
|---|---|---|---|---|---|
| A | 0 | 10 | | 12 | |
| B | 10 10 1 | | | | |
| C | 0 | 110 | | 1 | |
| D | 1 0 0 0 1 | | | | |
| E | 2 1110 | | | | |

Which one of the following statements about this graph is not true?

A The graph is connected.

B The graph contains an Eulerian trail.

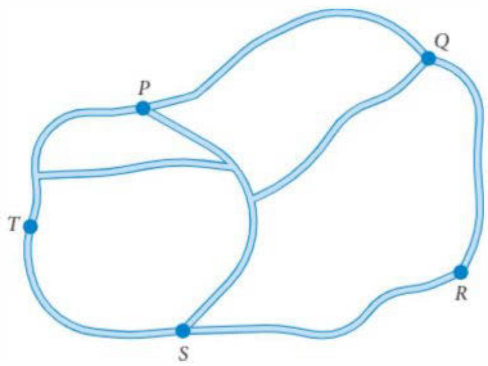
C The graph contains an Eulerian circuit.

D The graph contains a Hamiltonian cycle.

E The graph contains a loop and multiple edges.

12 ©VCAA 20191 NQ6 39% The map shows all the road connections between five towns, P, Q, P, S and *T*.

The road connections could be represented by the adjacency matrix




A

|   | P Q R S T |
|---|---|
| p | 1 3 0 2 2 |
| Q | 3 0 111 |
| R | 0 10 10 |
| S | 2 110 2 |
| T | 2 10 2 0 |

B

|   | P Q R S T |
|---|---|
| P | 1 2 0 2 2 |
| Q | 2 0 111 |
| R | 0 10 10 |
| S | 2 110 2 |
| T | 2 10 2 0 |

C

|   | P Q R S T |
|---|---|
| P | 0 3 0 2 2 |
| Q | 3 0 111 |
| R | 0 10 10 |
| S | 2 110 2 |
| T | 2 10 2 0 |

D

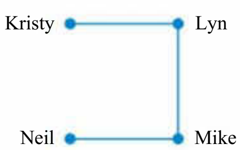
|   | P Q R S T |
|---|---|
| P | 0 2 0 2 2 |
| Q | 2 0 111 |
| R | 0 10 10 |
| S | 2 1112 |
| T | 2 10 2 0 |

E

|   | P Q R S T |
|---|---|
| P | 1 2 0 2 2 |
| Q | 2 0 111 |
| R | 0 10 10 |
| S | 2 1111 |
| T | 2 10 10 |

13 ©VCAA [ 2010 2NQ1 MODIFIED I 72% (2 marks) In a competition, members of a team work together to complete a series of challenges. The members of one team are Kristy (K)» Lyn (L), Mike (M) and Neil (N).

In one of the challenges, these four team members are only allowed to communicate directly with each other as indicated by the edges of the following network diagram.




|   | K L M N |
|---|---|
| K | 0 10 0 |
| L | 10 10 |
| M | 0/01 |
| N | 0^10 |

The adjacency matrix also shows the allowed lines of communication,

a Explain the meaning of a zero in the adjacency matrix, 1 mark

b Write down the values of/and *g* in the adjacency matrix. 1 mark

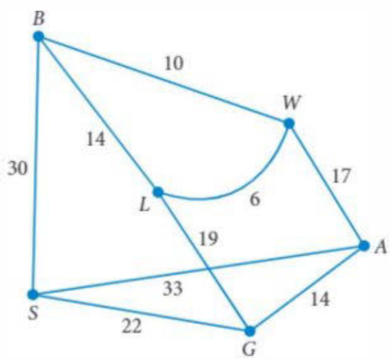
Video playlist
Shortest paths

# @ Shortest paths

## Weighted graphs and shortest paths

A weighted graph (sometimes just called a network) is a graph where extra information, such as distances, times or costs, is labelled on the edges.

The weighted graph on the right shows the distances in kilometres by road between several towns. Note that the lengths of the edges don't need to match the weight. For example, the edge showing a distance of 30 km doesn't have to be drawn three times as long as the edge showing a distance of 10 km. As with all the graphs weve been dealing with, it doesn't matter what length the edges are drawn or whether they are curved or straight.

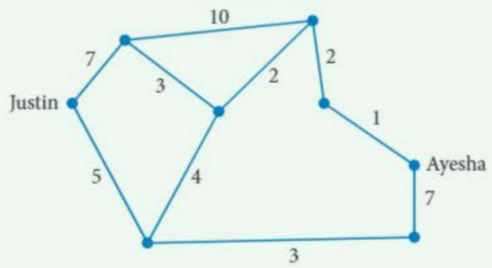


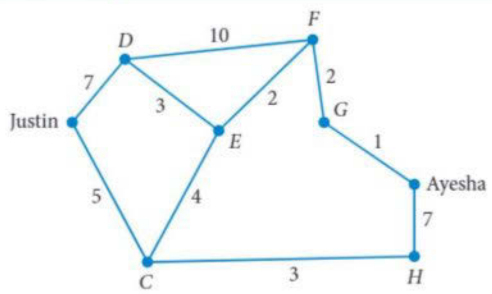

We use weighted graphs to solve shortest path problems. These are problems involving finding the shortest distance, shortest time or least cost from a starting point to an end point. For example, in the graph we might want to find the route from S to W with the shortest distance.

WB
p. 189

### WORKED EXAMPLE 14 Finding the shortest path by inspection

The network shows the travel times, in minutes, along a series of roads. Find the shortest time, in minutes, that it takes Justin to travel from his house to Ayesha's house by listing all the options.




| Steps | Working |
|---|---|
| **1** Add labels to the vertices. |  |
| **2** List the path options and calculate the total time of each option. | Justin–*D*–*F*–*G*–Ayesha takes 7 + 10 + 2 + 1 = 20 min<br>Justin–*C*–*H*–Ayesha takes 5 + 3 + 7 = 15 min<br>Justin–*D*–*E*–*F*–*G*–Ayesha takes 7 + 3 + 2 + 2 + 1 = 15 min<br>Justin–*C*–*E*–*F*–*G*–Ayesha takes 5 + 4 + 2 + 2 + 1 = 14 min<br>Justin–*D*–*E*–*C*–*H*–Ayesha takes 7 + 3 + 4 + 3 + 7 = 24 min<br>Justin–*C*–*E*–*D*–*F*–*G*–Ayesha takes 5 + 4 + 3 +10 + 2 + 1 = 25 min |
| **3** Write the answer. | The shortest time needed for Justin to travel to Ayesha's house is 14 minutes. |




9780170448406

## Dijkstra's algorithm

If the network is complicated, we use Dijkstra's (pronounced 'Dike-stras') algorithm to solve shortest path problems. Dijkstra's algorithm is a systematic series of repeated steps that finds the shortest path between any two vertices.



### Exam hack

You will usually be able to choose whether you use inspection or Dijkstra's algorithm to solve a shortest path problem.

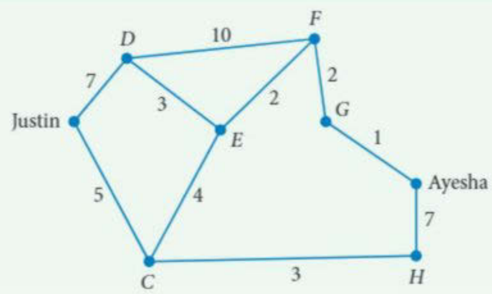
### Dijkstra's algorithm for finding the shortest path

1. Draw a box around the starting vertex label and add a value of 0.
2. For all the vertices connected to the starting vertex, add the value of each edge.
3. Draw a box around the smallest of the unboxed vertices.
4. For all the unboxed vertices connected to the newly boxed vertex, add the value of each edge. If an unboxed vertex already has a value, replace the existing value with the new value if the new value is smaller.
5. Draw a box around the smallest of the unboxed vertices. Repeat the steps for the newly boxed value until the end vertex is boxed. The value in the end box is the shortest amount.
6. To find the path, start from the end vertex and choose the boxed value that gives the correct value when you subtract the edge. Continue backtracking like this until you get back to the start.

WB p. 190

### WORKED EXAMPLE 15 Finding the shortest path using Dijkstra's algorithm

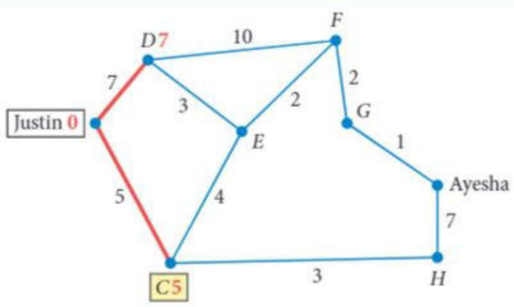
The network shows the travel times, in minutes, along a series of roads.




| Steps | Working |
|---|---|

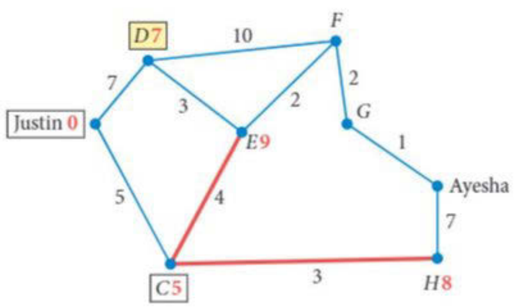
**a** Find the shortest time, in minutes, that it takes Justin to travel from his house to Ayesha's house using Dijkstra's algorithm.

1. Draw a box around the starting vertex label and add a value of 0.
2. For all the vertices connected to the starting vertex, add the value of each edge.
3. Draw a box around the smallest of the unboxed vertices.



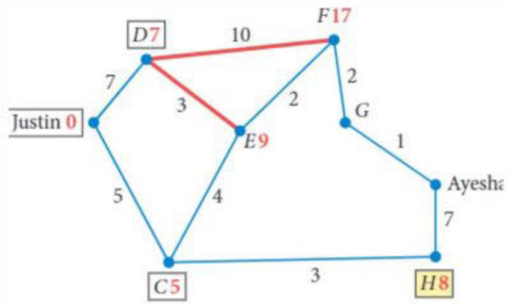

4. For all the unboxed vertices connected to the newly boxed vertex, add the value of each edge.

   If an unboxed vertex already has a value, replace the existing value with the new value if the new value is smaller.
5. Draw a box around the smallest of the unboxed vertices.

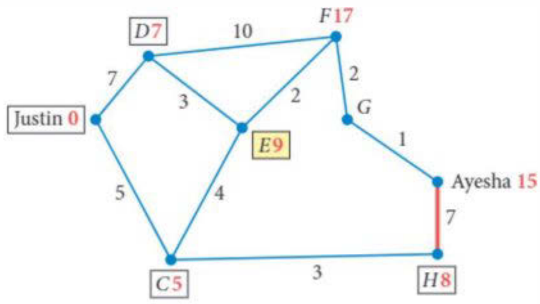


Repeat the steps for the newly boxed value until the end vertex is boxed.

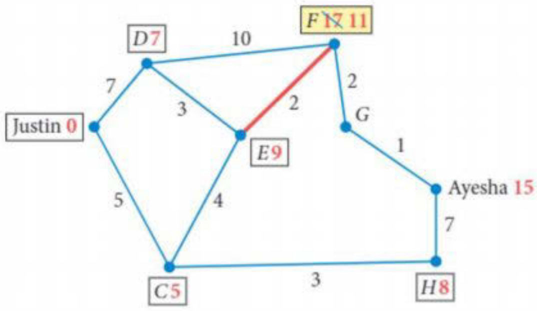
6



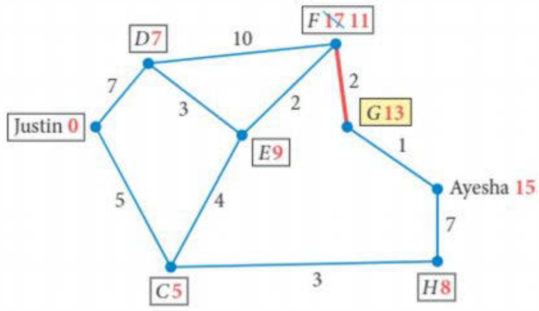

7



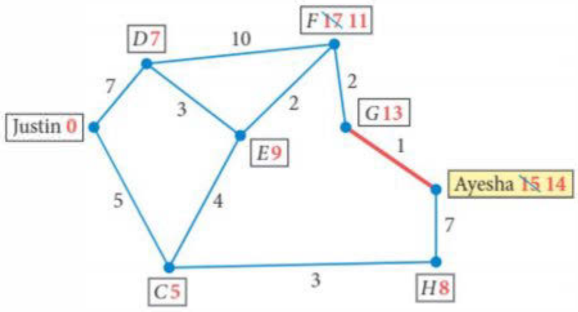

8




9




10




11 The value in the end box is the shortest time. The shortest time needed for Justin to travel to Ayesha's house is 14 minutes.

b Use backtracking to find the shortest time path and draw it on the map.

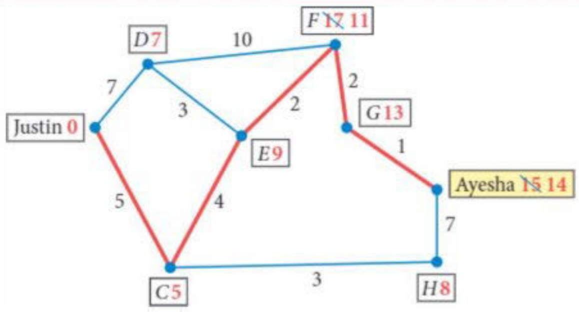
Start from the end vertex and choose the boxed value that gives the correct value when the edge is subtracted.

14 - 7 * 8 so don't move to *H*.

14 - 1 = 13 so move to *G*.

Continue this until you get back to the start.

Draw the path taken.




The shortest path is Justin-C-E-F-G-Ayesha.

## Exam hack

- The smallest unboxed value can be anywhere on the diagram. Don't just look at the most recent values that you've entered.
- Remember that the algorithm is only complete when you have *boxed* the destination value, not when you give the destination vertex a value.
- Sometimes there are two or more equal shortest paths. The backtracking process gives you all of these.


9780170448406

## EXERCISE 9.5 Shortest paths

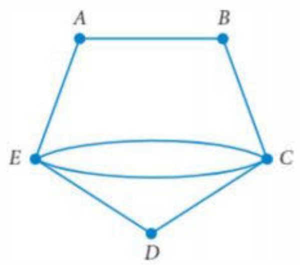
ANSWERS p. 724

### Recap

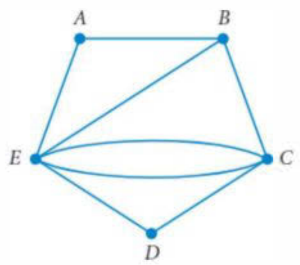
**1** ©VCAA 2010 1NQ3 MODIFIED | 95% A graph that can be drawn from the adjacency matrix is

| | A B C D E |
|---|---|
| A | 0 10 0 1 |
| B | 10 10 1 |
| C | 0 10 12 |
| D | 0 0 10 1 |
| E | 112 10 |

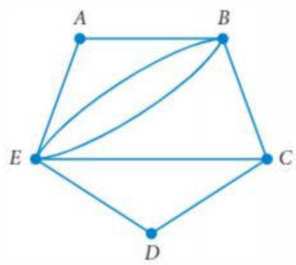
**A**



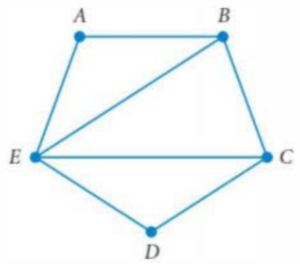

**B**



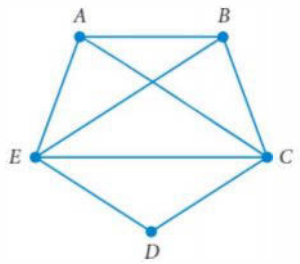

**C**



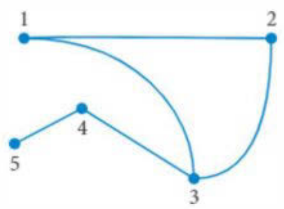

**D**




**E**




2 © VCAA 2002 1NQ3 MODIFIED | 88% Which of the following adjacency matrices could be used to represent the graph?




A

| | A B C D E |
|---|---|
| A | 0 110 0 |
| B | 10 10 0 |
| C | 110 10 |
| D | 0 0 10 1 |
| E | 0 0 0 1 0 |

B

| | A B C D E |
|---|---|
| A | 1 0 0 0 0 |
| B | 0 2 0 0 0 |
| C | 0 0 3 0 0 |
| D | 0 0 0 4 0 |
| E | 0 0 0 0 5 |

C

| | A B C D E |
|---|---|
| A | 0 110 0 |
| B | 0 0 10 0 |
| C | 0 0 0 1 0 |
| D | 0 0 0 0 1 |
| E | 0 0 0 0 0 |

D

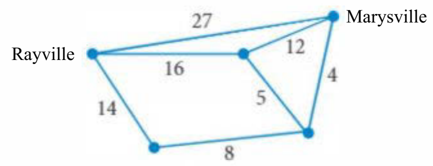
| | A B C D E |
|---|---|
| A | 0 0 0 0 0 |
| B | 1 0 0 0 0 |
| C | 110 0 0 |
| D | 0 0 10 0 |
| E | 0 0 0 1 0 |

E

| | A B C D E |
|---|---|
| A | 1110 0 |
| B | 1110 0 |
| C | 11110 |
| D | 0 0 111 |
| E | 0 0 0 1 1 |

## Mastery

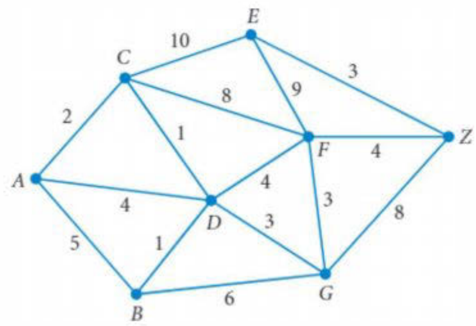
3 H WORKED EXAMPLE 14~| The network shows the travel times, in minutes, along a series of roads. Find the shortest time, in minutes, that it takes to travel from Rayville to Marysville by listing all the options.




4 Pj WORKED EXAMPLE i5~| The network represents the paths that connect different rides in a theme park. After closing time, the gatekeeper wants to lock the front gate at *A* and travel to the exit gate Z in the shortest time. The travel times, in minutes, are given on each edge.

a Find the shortest time, in minutes, that it takes the gatekeeper to travel from A to Z using Dijkstras' algorithm.

b Use backtracking to find the shortest time path. Copy the map and draw the shortest path on it.




## Exam practice

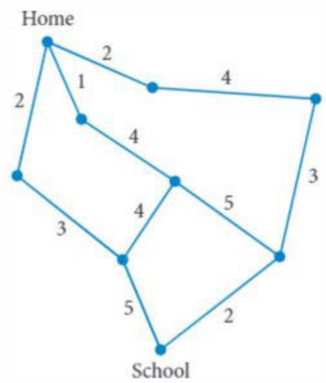
80-100% 60-79% «9%

5 ©VCAA | 2017N INQI Hunter rides his bike to school each day. The edges of the network shown represent the roads that Hunter can use to ride to school. The numbers on the edges give the distance, in kilometres, along each road.

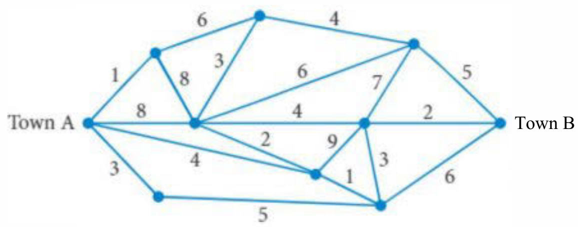
What is the shortest distance that Hunter can ride between home and school?

A 10km B 11 km C 12km

D 14 km E 23 km




6 ©VCAA | 2007 1NQ4 66% The network shows the distances, in kilometres, along a series of roads that connect Town A to Town B.



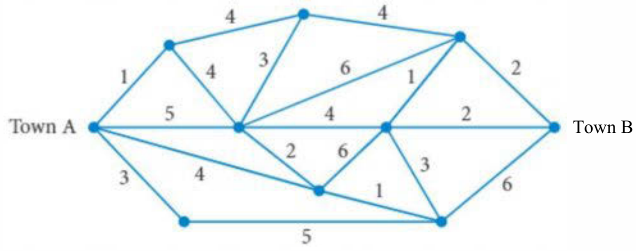

The shortest distance, in kilometres, to travel from Town A to Town B is

A 9 B 10 C 11 D 12 E 13



9780170448406

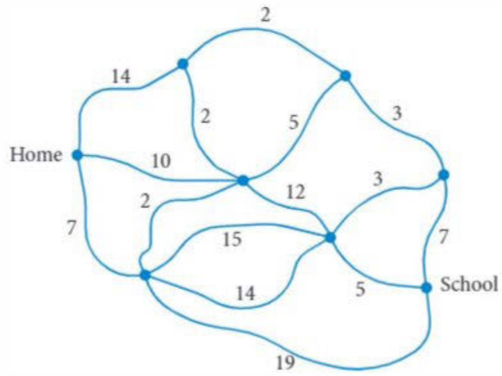
▶ 7 ©VCAA 2016S1NQ3 The following network shows the distances, in kilometres, along a series of roads that connect Town A to Town B.




Using Dijkstras algorithm, or otherwise, the shortest distance, in kilometres, from Town A to Town B is

A 9 B 10 C 11 D 12 E 13

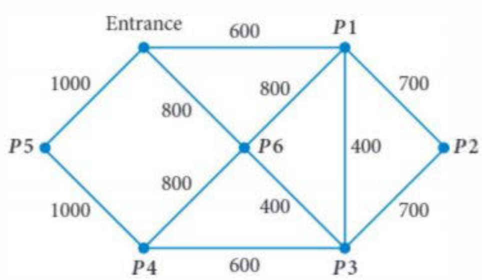
8 ©VCAA 20031NQ8 I 28% The network shows the travel times, in minutes, along a series of roads that connect a students home to school.




The shortest time, in minutes, for this student to travel from home to school is

A 22 B 23 C 24 D 25 E 26

9 ©VCAA 2013 2NQ1 I 90% (5 marks) The vertices in the network diagram show the entrance to a wildlife park and six picnic areas in the park: Pl, P2, P3, P4, P5 and P6. The numbers on the edges represent the lengths, in metres, of the roads joining these locations.




a In this graph, what is the degree of the vertex at the entrance to the wildlife park? 1 mark

b What is the shortest distance, in metres, from the entrance to picnic area P3? 1 mark

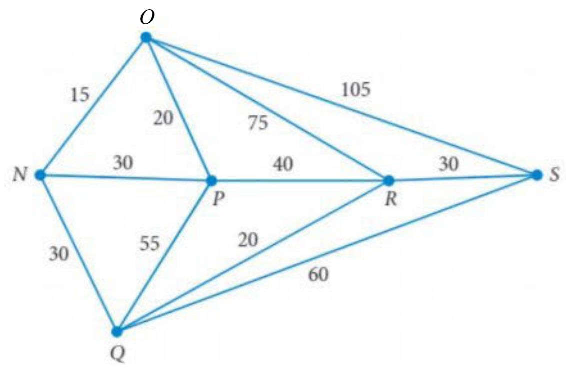
c A park ranger starts at the entrance and drives along every road in the park once.

i At which picnic area will the park ranger finish? 1 mark

ii What mathematical term is used to describe the route the park ranger takes? 1 mark

d A park cleaner follows a route that starts at the entrance and passes through each picnic area once, ending at picnic area Pl. Write down the order in which the park cleaner will visit the six picnic areas. 1 mark

10 IAWJJ 2017 2NQ1 J (3 marks) Bus routes connect six towns. The towns are Northend (N), Opera (O), Palmer (P), Quigley (Q), Rosebush *(R)* and Seatown (S). The graph gives the cost, in dollars, of bus travel along these routes. Bai lives in Northend *(N)* and he will travel by bus to take a holiday in Seatown (S).



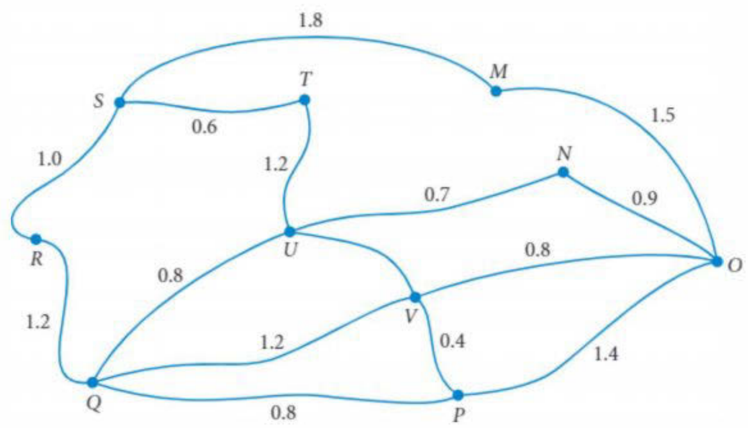

a 94% Bai considers travelling by bus along the route Northend *(N)* - Opera (O) - Seatown (S). How much would Bai have to pay? 1 mark

b 79% If Bai takes the cheapest route from Northend (N) to Seatown (S), which other town(s) will he pass through? 1 mark

c 83% Eulers formula, $v + f = e + 2$, holds for this graph. Copy and complete the formula by writing the appropriate numbers in the boxes.

$\square_v + \square_f = \square_e + 2$ 1 mark

11 ©VCAA | 2020 2NQ3^ (4 marks) A local fitness park has 10 exercise stations: *M* to *V*. The edges on the graph below represent the tracks between the exercise stations. The number on each edge represents the length, in kilometres, of each track.




The Sunny Coast cricket coach designs three different training programs, all starting at exercise station S.

| Training program number | Training details |
|---|---|
| 1 | The team must run to exercise station *O*. |
| 2 | The team must run along all tracks just once. |
| 3 | The team must visit each exercise station and return to exercise station S. |

a 58% What is the shortest distance, in kilometres, covered in training program 1? 1 mark

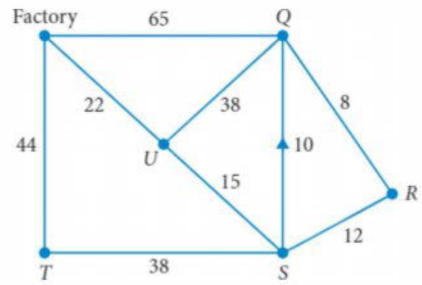
b i 66% What mathematical term is used to describe training program 2? 1 mark

ii 60% At which exercise station would training program 2 finish? 1 mark

c 25% To complete training program 3 in the minimum distance, one track will need to be repeated. Copy and complete the following sentence by filling in the boxes.

This track is between exercise station ☐ and exercise station ☐. 1 mark

▶ 12 ©VCAA 2015 2NQ2 ] (3 marks) A factory supplies groceries to stores in five towns, Q, *R, S, T* and *U*, represented by vertices on the graph.




The edges of the graph represent roads that connect the towns and the factory. The numbers on the edges indicate the distance, in kilometres, along the roads. Vehicles may only travel along the road between towns S and Q in the direction of the arrow due to temporary roadworks. Each day, a van must deliver groceries from the factory to the five towns. The first delivery must be to town T, after which the van will continue on to the other four towns before returning to the factory.

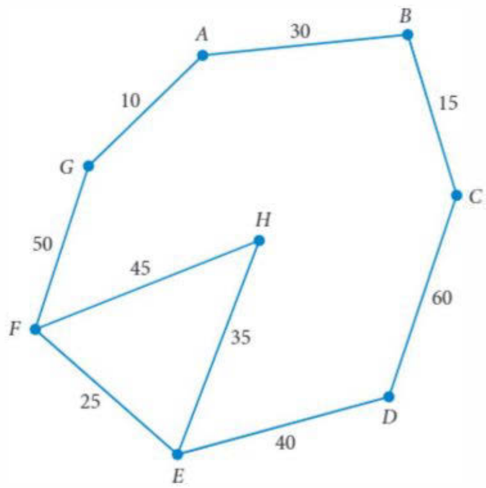
a i The shortest possible circuit from the factory for this delivery run, starting with town T, is not Hamiltonian. Copy and complete the order in which these deliveries would follow this shortest possible circuit:

factory -*T* -- factory 1 mark

ii With reference to the town names in your answer to part a i, explain why this shortest circuit is not a Hamiltonian circuit. 1 mark

b Determine the length, in kilometres, of a delivery run that follows a Hamiltonian circuit from the factory to these stores if the first delivery is to town *T*. 1 mark

13 ©VCAA | 2018N2NQ2 I (3 marks) An area of a property contains eight large bushes that are labelled *A* to H, as shown on the graph. The farmers dog enjoys running around this area, stopping at each bush on the way. The numbers on the edges joining the vertices give the shortest distance, in metres, between bushes.



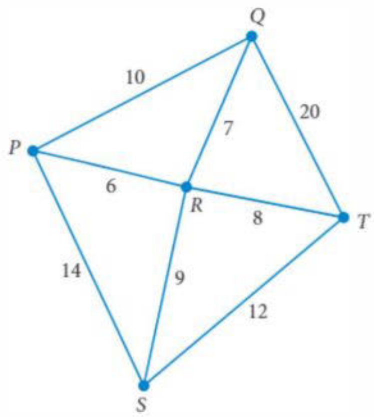

a Explain why the dog could not follow an Eulerian circuit through this network, 1 mark

b If the dog follows the shortest Hamiltonian path, name a bush at which the dog could start and a bush at which the dog could finish. 1 mark

c The sum of all distances shown on the graph is 310 m. The dog starts and finishes at bush *F* and runs along every edge in the network. What is the shortest distance, in metres, that the dog could have run? 1 mark

14 ©VCAA 2018 2NQ4 (2 marks) Parcel deliveries are made between five nearby towns, $P$ to $T$. The roads connecting these five towns are shown on the graph. The distances, in kilometres, are also shown.

A road inspector will leave from town $P$ to check all the roads and return to town $P$ when the inspection is complete. He will travel the minimum distance possible.




a 31% How many roads will the inspector have to travel on more than once? 1 mark

b 14% Determine the minimum distance, in kilometres, that the inspector will travel. 1 mark

Video playlist Minimum spanning trees

Worksheets Shortest paths and trees

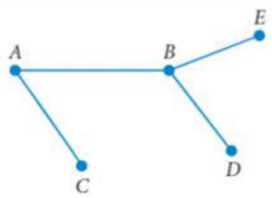
Minimum spanning trees

# @ Minimum spanning trees

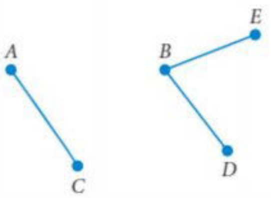
## Trees and spanning trees

A tree is a connected graph with no loops, multiple edges, or cycles.

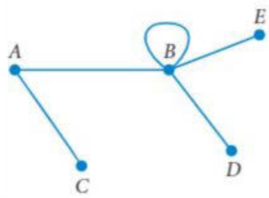
This graph is an example of a tree:



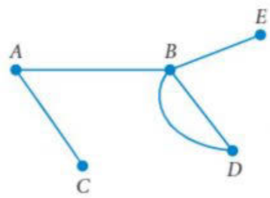

These are *not* trees:



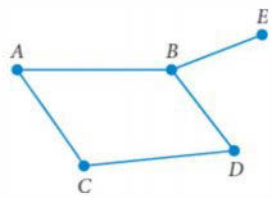

not a connected graph




contains a loop



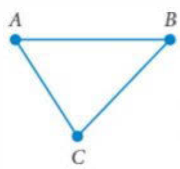

contains multiple edges



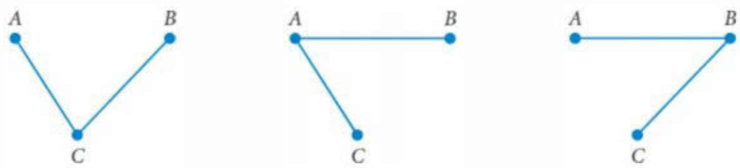

contains a cycle

A spanning tree is a tree subgraph that connects all the vertices of the original graph. Every connected graph has at least one spanning tree.

The three spanning trees for




are shown:





9780170448406

9.6

WB
p. 191

## Trees

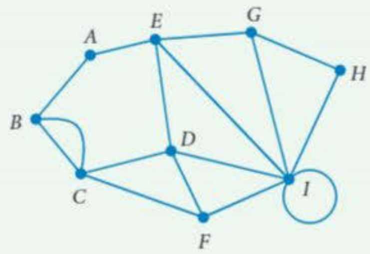
- A tree is a connected graph with no loops, multiple edges or cycles.
- The number of edges in a tree is always one less than the number of vertices.

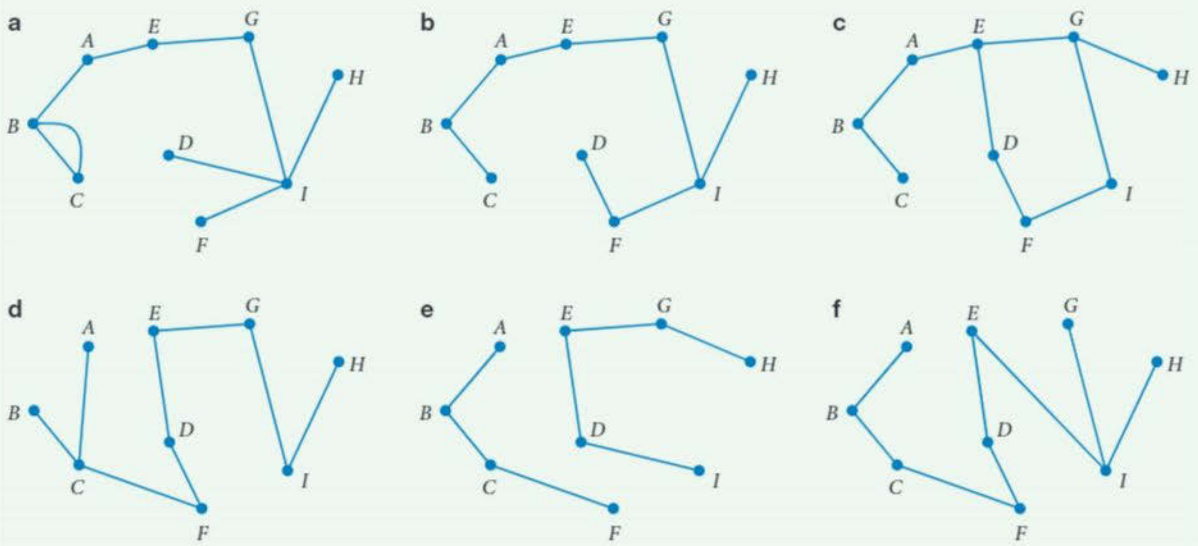
## Spanning trees

- A spanning tree is a tree subgraph that connects all the vertices of the original graph.
- Every connected graph has at least one spanning tree.

### WORKED EXAMPLE 16 Identifying spanning trees

Which of the following graphs are a spanning tree of the graph shown? For those that are spanning trees, verify that the number of edges is one less than the number of vertices. For those that aren't spanning trees, give a reason.




| Steps | Working |
|---|---|
| Is it connected?<br>Does it have all the vertices of the original graph?<br>Does it have no loops?<br>Does it have no multiple edges?<br>Does it have no cycles? | **a** This graph is not a spanning tree.<br>It has a multiple edge between B and C. |
| | **b** This graph is a spanning tree.<br>It has 9 vertices and 8 edges. |
| | **c** This graph is not a spanning tree.<br>It has a cycle. |
| | **d** This graph is not a spanning tree.<br>It has an edge (*AC*) that isn't in the original graph. |
| | **e** This graph is not a spanning tree.<br>It is not connected. |
| | **f** This is a spanning tree.<br>It has 9 vertices and 8 edges. |

## Minimum connector problems

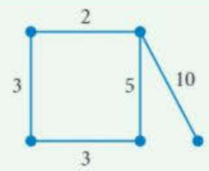
Spanning trees are used to solve minimum connector problems. These problems, like the shortest path problems, involve weights on the edges of the graph that represent quantities like distance and cost. A minimum connector is the minimum weight path connecting *all* the vertices in a graph. This is different to the shortest path, which is the minimum weight path between *two* particular vertices and doesn't have to contain all the vertices in a graph.

To solve a minimum connector problem we need to find the minimum spanning tree, which is the spanning tree with the smallest total weight. Minimum spanning trees are used when connecting broadband networks like the NBN, constructing water pipe systems and in real-time face recognition software.

WB
p. 192

**WORKED EXAMPLE 17** Finding minimum spanning trees by inspection

Find all the spanning trees for the network shown, and hence, find the total weight of the minimum spanning tree.




**Steps**

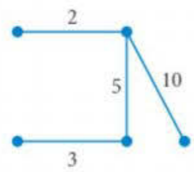
1 Work out how many edges need to be removed from the graph to create a spanning tree. In a spanning tree
- the number of edges is always one less than the number of vertices
- the number of vertices is the same as the original graph.

2 Remove each edge in turn to see which options result in a spanning tree. Calculate the total weight of each spanning tree and find the one with the smallest total weight.

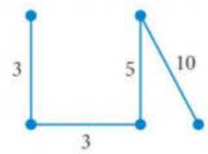
**Working**

A spanning tree has 5 vertices and 4 edges. The original graph has 5 edges, so we need to remove one edge to create a spanning tree.

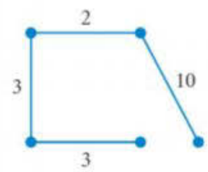
Total weight of spanning tree = 20



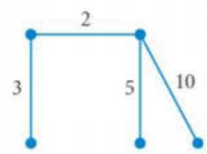

Total weight of spanning tree = 21



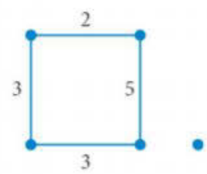

Total weight of spanning tree = 18




Total weight of spanning tree = 20




This isn't a connected graph, so it's not a tree.




The total weight of the minimum spanning tree is 18.



9780170448406

## Prim's algorithm

Prim's algorithm is a series of steps to find a minimum spanning tree for a graph.

> **© Exam hack**
>
> Questions don't always say to use Prim's algorithm, but you can use it for any minimum spanning tree problem.

9.6

**Prim's algorithm for finding a minimum spanning tree**

1 Start at any vertex and choose the edge with the lowest weight connected to this vertex.

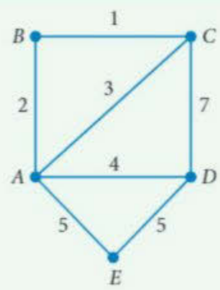
2 Look at *all* the edges connecting to the vertices you've chosen so far *(not just the last vertex connected)* and choose the edge with the lowest weight that doesn't connect to a vertex already in the tree. If there are edges with equal lowest weights, choose one of them.

3 Repeat step 2 until all the vertices in the graph are included in the tree.

WB p. 193

**WORKED EXAMPLE 18** Finding minimum spanning trees using Prim's algorithm

Use Prim's algorithm to find the minimum spanning tree for the weighted graph shown, and hence, find the total weight of the minimum spanning tree.



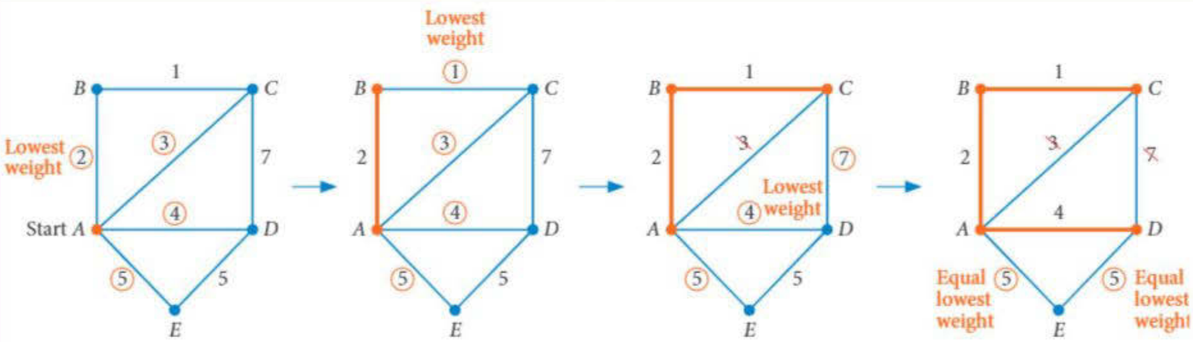

**Steps**

**1** Start at any vertex and choose the edge with the lowest weight connected to this vertex.

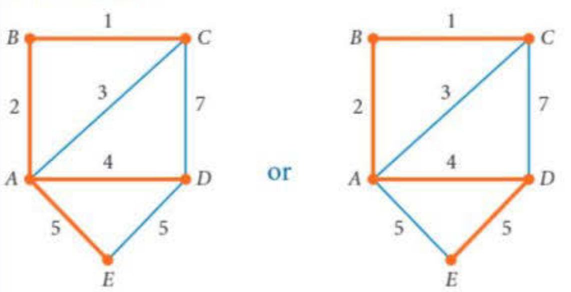
**2** Look at *all* the edges connecting to the vertices you've chosen so far (*not just the last vertex connected*) and choose the edge with the lowest weight that doesn't connect to a vertex already in the tree. If there are edges with equal lowest weights, choose one of them.

**3** Repeat step **2** until all the vertices in the graph are included in the tree.

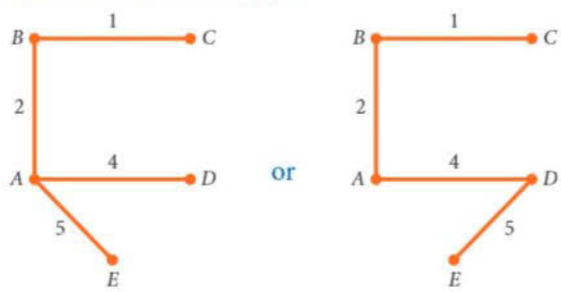
**Working**




With the next step all the vertices have been included:




This gives us two minimum spanning trees with the same total weight:



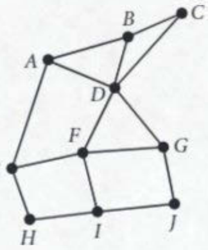

The total weight of the minimum spanning tree is $1 + 2 + 4 + 5 = 12$.

Video VCE question analysis: Undirected graphs

## VCE QUESTION ANALYSIS

©VCAA 2018 2NQ2 2018 Examination 2 Networks and decision mathematics Question 2 (3 marks)

In one area of the town of Zenith, a postal worker delivers mail to 10 houses labelled as vertices A to *J* on the graph,




a Which one of the vertices on the graph has degree 4? 1 mark

For this graph, an Eulerian trail does not currently exist,

b For an Eulerian trail to exist, what is the minimum number of extra edges that the graph would require? 1 mark

c The postal worker has delivered the mail at *F* and will continue her deliveries by following a Hamiltonian path from *F*. Draw a possible Hamiltonian path for the postal worker. 1 mark

### Reading the question

- The key definitions required are the degree of a vertex, an Eulerian trail and a Hamiltonian path.
- You are being asked for a number in part b.
- You are being asked to draw something in part c and you have been given the starting point.

### Thinking about the question

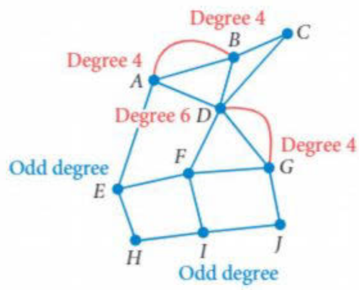
- When does an Eulerian trail exist?
- How do you find a Hamiltonian path?
- The word possible suggests there is more than one answer.

### Worked solution (/ = 1 mark)

a Vertex F / is the only vertex with exactly 4 edges connecting to it, so it has degree 4.

b An Eulerian trail will only exist if the graph has *exactly two* vertices of odd degree. There are six vertices of odd degree, so if we connect two pairs of these with an extra edge, four of the odd degree vertices become even degree vertices, leaving exactly two vertices of odd degree.

For example, adding extra edges *AB* and *DG* means *E* and *I* are the only odd degree vertices:

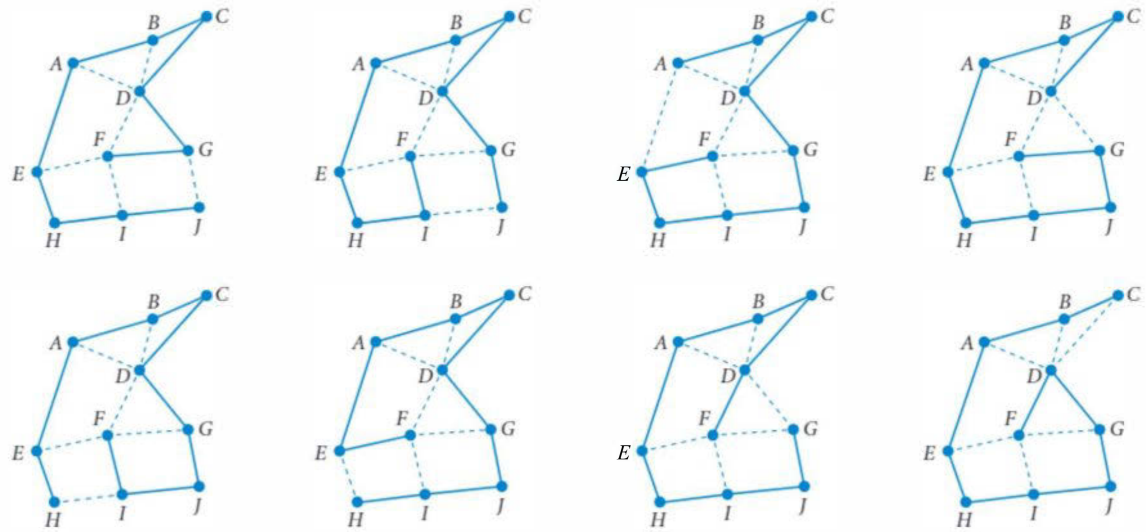



There are different ways of adding extra edges to result in exactly two vertices of odd degree, but it can't be done by adding only one edge. So, the minimum number of extra edges that the graph would require for an Eulerian trail to exist is 2. /



9780170448406

c A Hamiltonian path is a path that includes *every vertex* in a graph once only. We need to find Hamiltonian paths by trial and error. All of them begin at *F* and visit each vertex only once. Some possible answers are below. /



**Student performance** 80-100% 60-79% 0-59%

a 96%

b 54% Some students gave a definition of an Eulerian trail rather than stating how many extra edges were required.

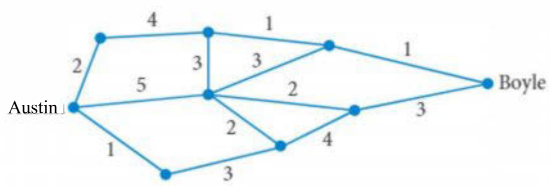
c 62% A reasonable number of students answered this question correctly. Many students drew paths that did not start at *F*. Some drew paths that returned to F, creating a circuit.

## EXERCISE 9.6 Minimum spanning trees

ANSWERS p. 725

### Recap

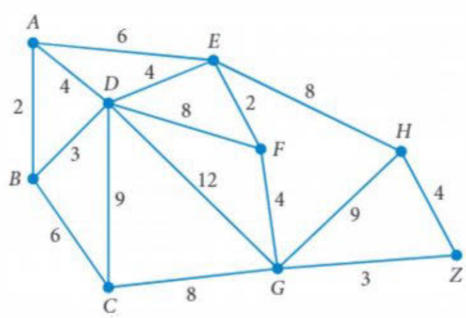
1 ©VCAA 20091NQ2 I 93% The network shows the distances, in kilometres, along roads that connect the cities of Austin and Boyle. The shortest distance, in kilometres, from Austin to Boyle is




A 7 B 8 C 9 D 10 E 11

2 The network represents the paths that connect buildings at a college.

In the morning, Amira has to travel from the entrance at *A* to her class at *Z* in the shortest time. The travel times, in minutes, are given on each edge. Using Dijkstra's algorithm, the shortest travel time from A to Z is



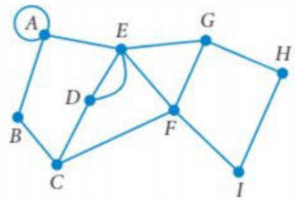

A 14 minutes B 15 minutes

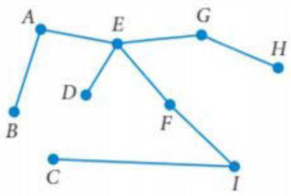
C 16 minutes D 19 minutes

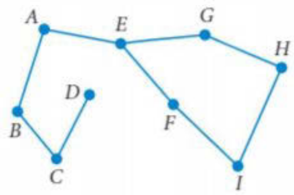
E 20 minutes

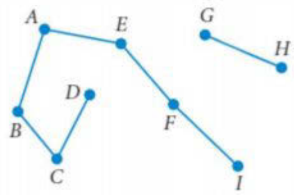
## Mastery

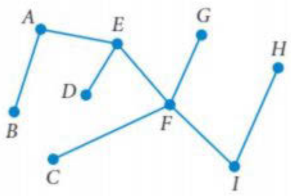
30 WORKED EXAMPLE 16 Which of the following graphs are a spanning tree of the graph shown on the right? For those that are spanning trees, verify that the number of edges is one less than the number of vertices. For those that aren't spanning trees, give a reason.

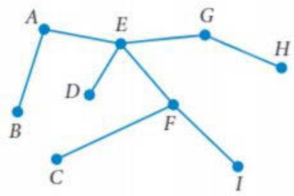


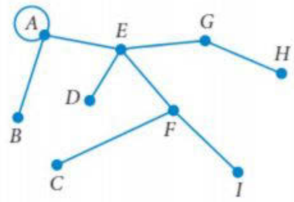

a


b


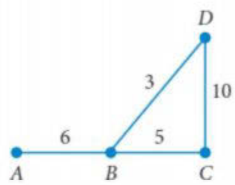
c


d


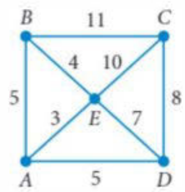
e


f


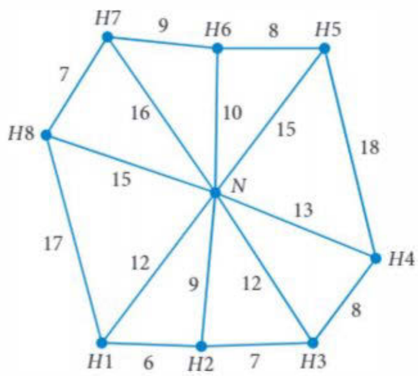
4 WORKED EXAMPLE 17 Find all the spanning trees for the network shown, and hence, find the total weight of the minimum spanning tree.




5 EJ WORKED EXAMPLE 18 Use Prims' algorithm to find the minimum spanning tree for the weighted graph shown, and hence, find the total weight of the minimum spanning tree.




6 WKED EXAMPLEIS A company is laying broadband cable in a neighbourhood. The weighted graph shows the distances, in metres, from the node $N$ to each of the eight houses. Use Prims' algorithm to find the minimum spanning tree and hence, find the minimum length of broadband cable needed to connect to all eight houses.

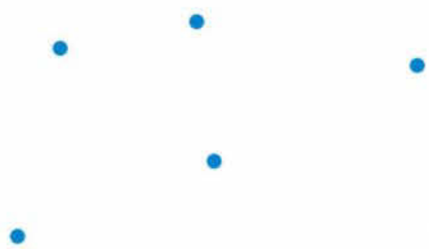



## Exam practice

80-100% 60-79% 0-59%

7 ©VCAA 2018 1NQ1 92% Consider the graph with five isolated vertices shown.

To form a tree, the minimum number of edges that must be added to the graph is



Al B 4 C 5

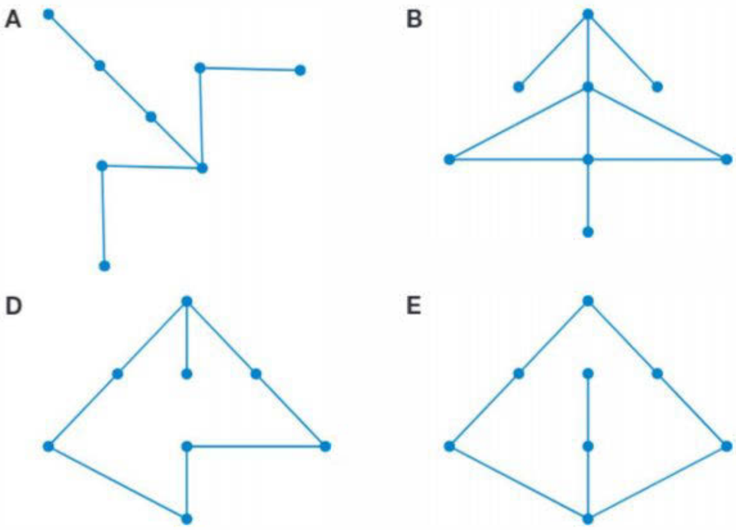
D 6 E 10



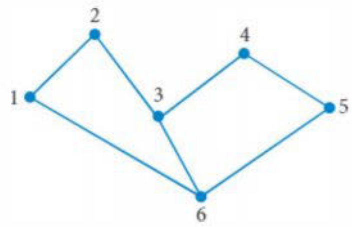
9780170448406

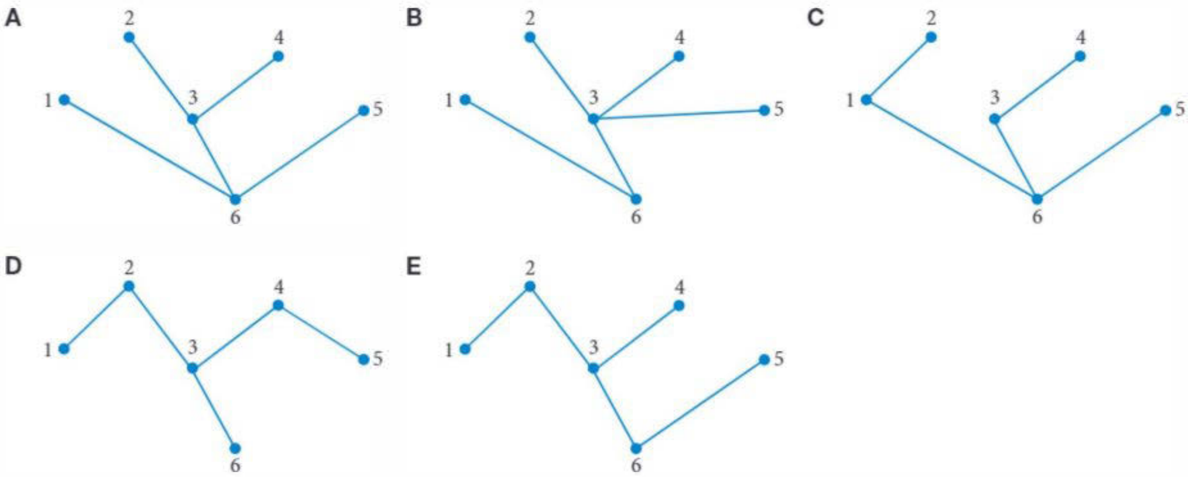
9.6

8 ©VCAA 20131NQ1 I 91% Which of the following graphs is a tree?

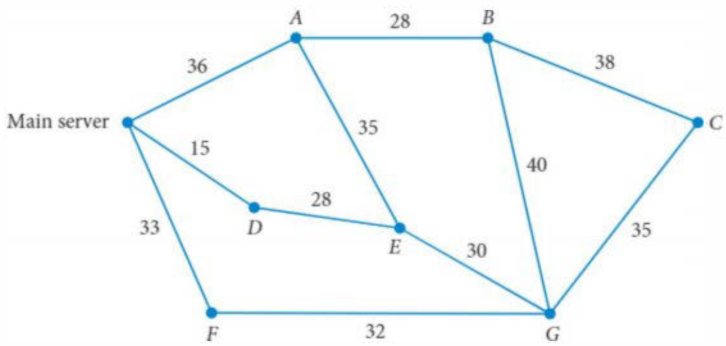



9 ©VCAA 20201NQ3 I 81% Which one of the following is not a spanning tree for the network on the right?



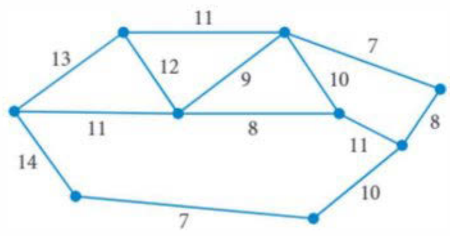

10 ©VCAA 2019 1NQ5 75% The following diagram shows the distances, in metres, along a series of cables connecting a main server to seven points, *A* to G, in a computer network.




The minimum length of cable, in metres, required to ensure that each of the seven points is connected to the main server directly or via another point is

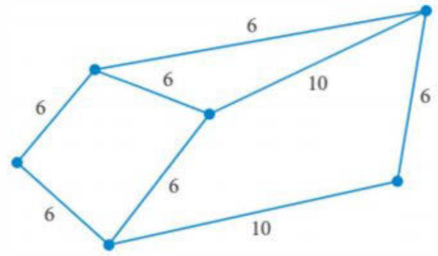
A 175 B 203 C 208 D 221 E 236

11 ©VCAA 2019N1NQ5 J In the graph shown, the vertices represent electricity transformer substations. The numbers on the edges of the graph show the length, in kilometres, of cables that connect these substations. What is the minimum length of cable, in kilometres, that is necessary to make sure that each substation remains connected to the network?




A 65 B 71 C 73 D 74 E 77

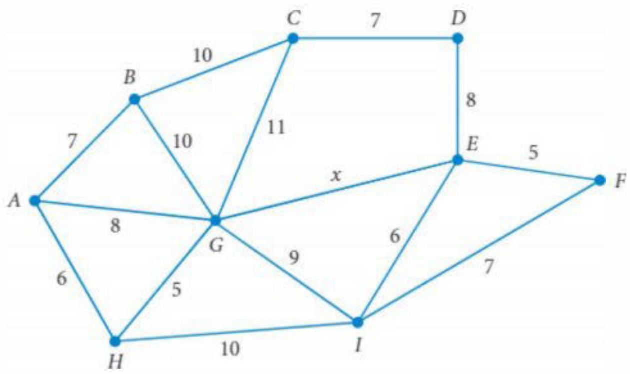
12 ©VCAA 2017N1NQ7 Consider the weighted graph shown.




How many different minimum spanning trees are possible?

A2 B3 C 4 D5 E 6

13 ©VCAA 20201NQ5 56% The network below shows the distances, in metres, between camp sites at a camping ground that has electricity. The vertices A to *I* represent the camp sites.



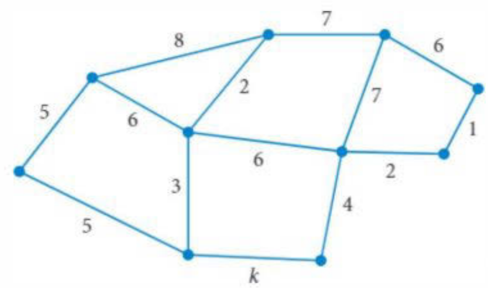

The minimum length of cable required to connect all the camp sites is 53 m. The value of x, in metres, is at least

A 5 B 6 C 8 D 9 E 11

14 ©VCAA 20161NQ4 52% The minimum spanning tree for the network shown includes the edge with weight labelled *k*.

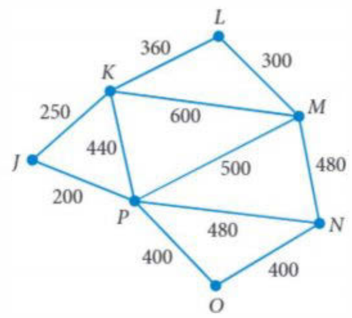
The total weight of all edges for the minimum spanning tree is 33. The value of *k* is

A l B 2 C 3
D 4 E 5



**15** ©VCAA | 2015 2NQ1 I (5 marks) A factory requires seven computer servers to communicate with each other through a connected network of cables. The servers, J, K, £, Af, *N, O* and P, are shown as vertices on the graph.

The edges on the graph represent the cables that could connect adjacent computer servers. The numbers on the edges show the cost, in dollars, of installing each cable.



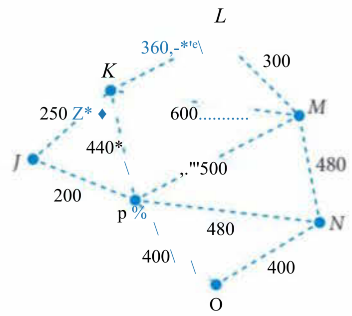

a 90% What is the cost, in dollars, of installing the cable between server *L* and server At? 1 mark

b 90% What is the cheapest cost, in dollars, of installing cables between server *K* and server N? 1 mark

c 90% An inspector checks the cables by walking along the length of each cable in one continuous path. To avoid walking along any of the cables more than once, at which vertex should the inspector start and where would the inspector finish? 1 mark

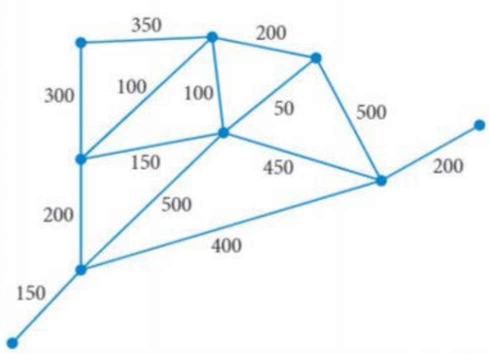
d The computer servers will be able to communicate with all the other servers as long as each server is connected by cable to at least one other server.

i The cheapest installation that will join the seven computer servers by cable in a connected network follows a minimum spanning tree. Copy the plan shown and draw the minimum spanning tree on it. 1 mark



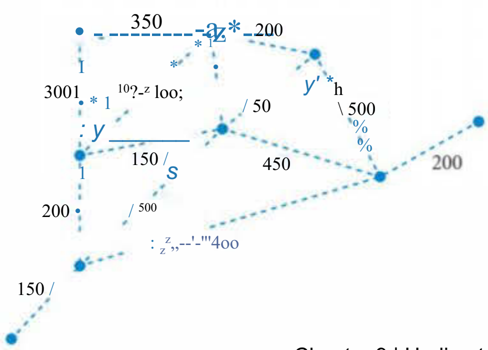

ii The factory's manager has decided that only six connected computer servers will be needed, rather than seven. How much would be saved in installation costs if the factory removed computer server *P* from its minimum spanning tree network? 1 mark

**16** © VCAA 2017 2NQ31 (2 marks) While on holiday, four friends visit a theme park where there are nine rides. On the graph shown, the positions of the rides are indicated by the vertices. The numbers on the edges represent the distances, in metres, between rides.



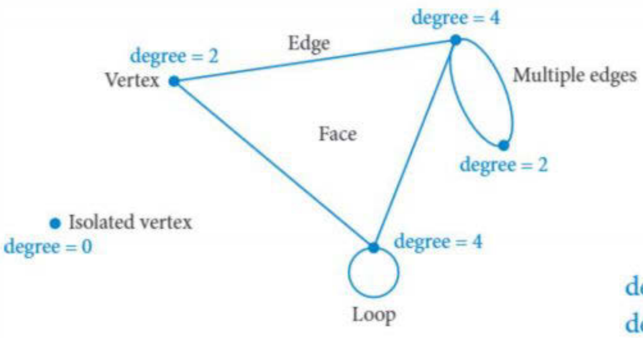

a 53% Electrical cables are required to power the rides. These cables will form a connected graph. The shortest total length of cable will be used. Give a mathematical term to describe a graph that represents these cables. 1 mark

b 53% Copy the following and use it to draw the graph that represents these cables. 1 mark



# (£) Chapter summary

## Features of a graph




degree sum = 2 + 4 + 2 + 4 + 0 = 12
degree sum = 2 × number of edges = 2 × 6 = 12

## Types of graphs

- Isomorphic graphs are graphs that show exactly the same connections.
- A planar graph is a connected graph that can be drawn so that it doesn't have any edges crossing.
- A connected graph is a graph where there is a path from any vertex to any other vertex.
- A bridge is any edge that keeps a graph connected.
- A simple graph is a graph without any loops or multiple edges.
- A complete graph is a simple graph where every vertex is connected to every other vertex.
- Every vertex in a complete graph has degree = number of vertices - 1.
- A subgraph is part of a larger graph.
- A weighted graph is a graph where extra information, such as distances, times or costs, is labelled on the edges.
- A tree is a connected graph with no loops, multiple edges or cycles.
- The number of edges in a tree is always one less than the number of vertices.
- A spanning tree is a tree subgraph that connects all the vertices of the original graph.
- Every connected graph has at least one spanning tree.
- A minimum spanning tree is the spanning tree with the smallest total weight.

## Euler's formula

For connected planar graphs

$$v + f - e = 2$$

where

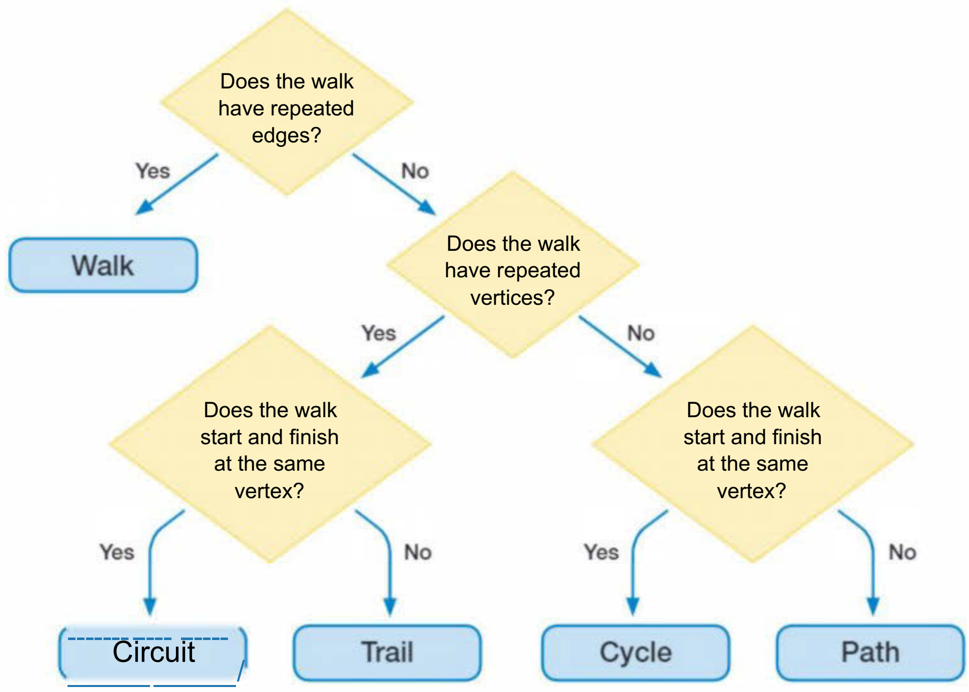
v = the number of vertices

$f$ = the number of faces

$e$ = the number of edges.

## Types of walks

- A walk is a sequence of connected vertices.
- A trail is a walk with no repeated edges.
- A circuit is a walk with no repeated edges that starts and finishes at the same vertex.
- A path is a walk with no repeated vertices.
- A cycle is a walk with no repeated vertices that starts and finishes at the same vertex.




## Eulerian trails and circuits

### Eulerian trail

- An Eulerian trail is a walk with no repeated edges that includes every *edge* in a graph.
- An Eulerian trail will only exist if the graph has *exactly two* vertices of odd degree.
- To find the Eulerian trail, you must start at one of the two vertices of odd degree and finish at the other vertex of odd degree.

### Eulerian circuit

- An Eulerian circuit is a walk with no repeated edges that includes every *edge* in a graph and starts and finishes at the same vertex.
- An Eulerian circuit will only exist if the graph has all vertices of even degree.

## Hamiltonian paths and cycles

- A Hamiltonian path is a walk with no repeated vertices that includes every vertex in a graph.
- A Hamiltonian cycle is a walk with no repeated vertices that includes every vertex in a graph and starts and finishes at the same vertex.
- We need to use trial and error to find whether a Hamiltonian path or cycle exists.

## Walk summary

| Walks that: | Trail | Eulerian trail | Circuit | Eulerian circuit | Path | Hamiltonian path | Cycle | Hamiltonian cycle |
|---|---|---|---|---|---|---|---|---|
| have no repeated edges | Z | Z | Z | Z | Z | Z | Z | Z |
| have no repeated vertices | | | | | Z | Z | Z | Z |
| start and finish at the same vertex | | | Z | Z | | | Z | Z |
| include every edge | | Z | | Z | | | | |
| include every vertex | | Z | | Z | | Z | | Z |

## Adjacency matrices

An adjacency matrix is a square matrix representing a graph where

- each row and column is labelled as a vertex
- the elements show the number of edges between vertices
- the elements are symmetric about its leading diagonal
- the leading diagonal shows the number of loops.

Number of vertices = number of rows

Number of edges = sum of the elements on and below the leading diagonal.

Degree of a vertex = sum of the elements of the row (+ 1 for each loop)

## Finding the shortest path

A weighted graph is a graph where extra information is labelled on the edges.

The shortest path can be found by inspection if the graph isn't complex.

Use Dijkstra's algorithm for complex graphs:

1. Draw a box around the starting vertex label and add a value of 0.
2. For all the vertices connected to the starting vertex, add the value of each edge.
3. Draw a box around the smallest of the unboxed vertices.
4. For all the unboxed vertices connected to the newly boxed vertex, add the value of each edge. If an unboxed vertex already has a value, replace the existing value with the new value if the new value is smaller.
5. Draw a box around the smallest of the unboxed vertices. Repeat the steps for the newly boxed value until the end vertex is boxed. The value in the end box is the shortest amount.
6. To find the path, start from the end vertex and choose the boxed value that gives the correct value when you subtract the edge. Continue backtracking like this until you get back to the start.

## Finding a minimum spanning tree

- A tree is a connected graph with no loops, multiple edges or cycles.
- The number of edges in a tree is always one less than the number of vertices.
- A spanning tree is a tree subgraph that connects all the vertices of the original graph.
- Every connected graph has at least one spanning tree.
- A minimum spanning tree is the spanning tree with the smallest total weight.
- The minimum spanning tree can be found by inspection if the graph isn't complex.

Use Prim's algorithm for complex graphs:

1. Start at any vertex and choose the edge with the lowest weight connected to this vertex.
2. Look at *all* the edges connecting to the vertices you've chosen so far *(not just the last vertex connected)* and choose the edge with the lowest weight that doesn't connect to a vertex already in the tree. If there are edges with equal lowest weights, choose one of them.
3. Repeat step 2 until all the vertices in the graph are included in the tree.

# Cumulative examination 1



Total number of marks: 17 Reading time: 7 minutes Writing time: 39 minutes

*Use the following information to answer the next three questions.*

The following table shows the data collected from a sample of seven drivers who entered a supermarket car park. The variables in the table are:

- *distance* - the distance that each driver travelled to the supermarket from their home
- *sex* - the sex of the driver (female, male)
- *number of children* - the number of children in the car
- *type of car* - the type of car (sedan, wagon, other)
- *postcode* - the postcode of the drivers home

| Distance (km) | Sex (F = female, M = male) | Number of children | Type of car (1 = sedan, 2 = wagon, 3 = other) | Postcode |
|---|---|---|---|---|
| 4.2 | F | 2 | 1 | 8148 |
| 0.8 | M | 3 | 2 | 8147 |
| 3.9 | F | 3 | 2 | 8146 |
| 5.6 | F | 1 | 3 | 8245 |
| 0.9 | M | 1 | 3 | 8148 |
| 1.7 | F | 2 | 2 | 8147 |
| 2.5 | M | 2 | 2 | 8145 |

1 ©VCAA 2014 1CQ3 The mean, $\bar{x}$, and the standard deviation, $s_x$» of the variable *distance* are closest to

A $\bar{x}$ = 2.5, $s_x$=3.3 B $\bar{x}$ = 2.8, $s_x$ = 1.7 C $\bar{x}$ = 2.8, $5_X$ = 1.8

D $\bar{x}$ = 2.9, $s_x$ = 1.7 E $\bar{x}$= 3.3, $s_x$=2.5

2 ©VCAA 2014 1CQ4 The number of categorical variables in this data set is

A 0 **B** 1 **C** 2 **D** 3 E 4

3 © VCAA 2014 1CQ5 The number of female drivers with three children in the car is

A 0 B 1 C 2 D 3 E 4

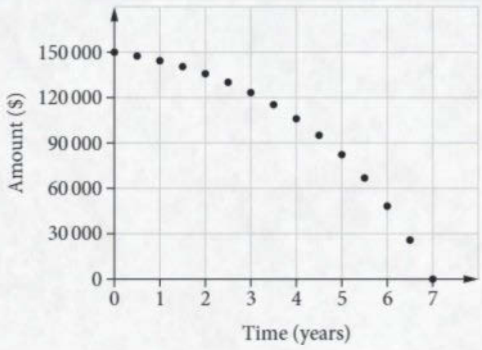
4 ©VCAA | 20181CQ14 , A least squares line is fitted to a set of bivariate data. Another least squares line is fitted with response and explanatory variables reversed. Which one of the following statistics will not change in value?

A the residual values B the predicted values

C the correlation coefficient *r* D the slope of the least squares line

E the intercept of the least squares line

5 Which of the following could this graph model?




A the amount of compound interest earned on a 7-year investment

B the amount in a bank account earning compound interest over 7 years

C the value of an asset over 7 years with flat rate depreciation

D the balance owing on a reducing balance loan over 7 years

E the value of an asset over 7 years with reducing balance depreciation

6 Eric invests $8000 in an account earning 7.2% p.a. interest compounded annually. He deposits an extra $8000 into the account each year after the initial deposit. If $A_n$ is the amount in his account after year n, a recurrence relation for determining $A_n$ is

A $A_o$ = 8000, $A_{n+1}$ = 1.072A,, + 8000

B $A_o$ = 8576, $A_{,,+1}$ = 1.072A,, + 8000

C $A_o$ = 8000, $A_{M+1}$ = 0.072A,, + 8000

D $A_o$ = 8000, $A_{n+1}$ = 1.072(A,, + 8000)

E $A_o$ = 8576, $A_{N+1}$ = 1.072(A,, + 8000)

7 ©VCAA j 2006iMQ<r A large population of birds lives on a remote island. Every night, each bird settles at either location A or location *B*. It was found on the first night that the number of birds at each location was the same. On each subsequent night, a percentage of birds changed the location at which they settled. This movement of birds is described by the transition matrix

$$\begin{array}{c} \\ A \\ B \end{array}\begin{array}{c} A \quad B \\ \begin{bmatrix} 0.8 & 0 \\ 0.2 & 1 \end{bmatrix} \end{array}$$

Assume this pattern of movement continues. In the long term, the number of birds that settle at location A will

A not change.

B gradually decrease to zero.

C eventually settle at around 20% of the islands bird population.

D eventually settle at around 80% of the islands bird population.

E gradually increase.

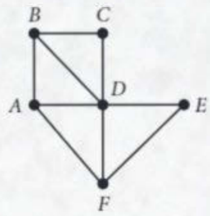
8 ©VCAA 20131MQ8 The matrix $S_{M+1}$ is determined from the matrix $S_n$ using the rule $S_{,,+1} = TS_n - C$ where T, $S_o$ and C are defined as follows.

$$T = \begin{bmatrix} 0.5 & 0.6 \\ 0.5 & 0.4 \end{bmatrix}, S_0 = \begin{bmatrix} 100 \\ 250 \end{bmatrix} \text{ and } C = \begin{bmatrix} 20 \\ 20 \end{bmatrix}$$

Given this information, the matrix $S_2$ equals

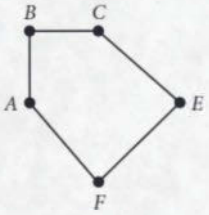
A $\begin{bmatrix} 100 \\ 250 \end{bmatrix}$ B $\begin{bmatrix} 148 \\ 122 \end{bmatrix}$ C $\begin{bmatrix} 170 \\ 140 \end{bmatrix}$ D $\begin{bmatrix} 180 \\ 130 \end{bmatrix}$ E $\begin{bmatrix} 190 \\ 160 \end{bmatrix}$

9 ©VCAA 2004 1NQ1 MODIFIED I Consider the network graph.

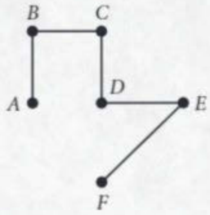



A subgraph of the graph is

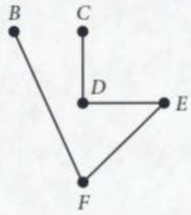
**A**



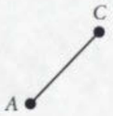

**B**



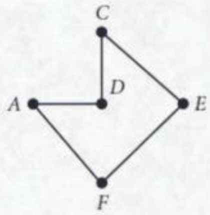

**C**




**D**




**E**




10 ©VCAA 20181NQ3 I A planar graph has five faces. This graph could have

A eight vertices and eight edges.

B six vertices and eight edges.

C eight vertices and five edges.

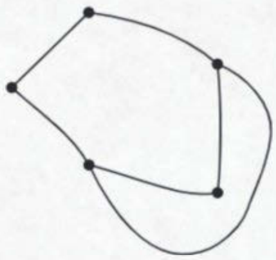
D eight vertices and six edges.

E five vertices and eight edges

11 ©VCAA 2019 1NO4 Two graphs, labelled Graph 1 and Graph 2, are shown below.

Graph 1

Graph 2



Which one of the following statements is not true?

A Graph 1 and Graph 2 are isomorphic.

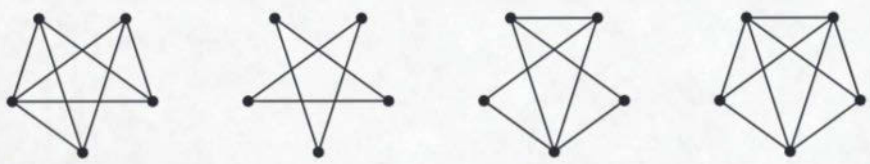
B Graph 1 has five edges and Graph 2 has six edges.

C Both Graph 1 and Graph 2 are connected graphs.

D Both Graph 1 and Graph 2 have three faces each.

E Neither Graph 1 nor Graph 2 are complete graphs.

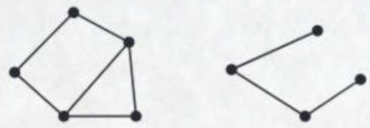
12 evcAA 2014 1NQ6 Consider the following four graphs.



How many of these four graphs have an Eulerian circuit?

A0 Bl C 2 D 3 E 4

13 ©VCAA 2017N1NQ2 Consider the two graphs.



Which one of the following statements is not true?

A Each graph is connected.

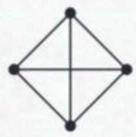
B Each graph contains an Eulerian trail.

C Each graph contains a Hamiltonian cycle.

D Each graph has at least one vertex of degree two.

E The sum of the degrees of the vertices for each graph is even.

14 ©VCAA 2007 1NQ1 A mathematical term that could not be used to describe the graph shown is



A complete B planar C simple D undirected E tree

15 ©VCAA 20081NQ5 A connected planar graph has five vertices, A, *B,C,D* and *E.* The degree of each vertex is given in the table.

| Vertex | Degree |
|---|---|
| *A* | 3 |
| *B* | 4 |
| *C* | 3 |
| *D* | 5 |
| *E* | 3 |

Which of the following statements regarding this planar graph is true?

A The sum of the vertices equals 15.

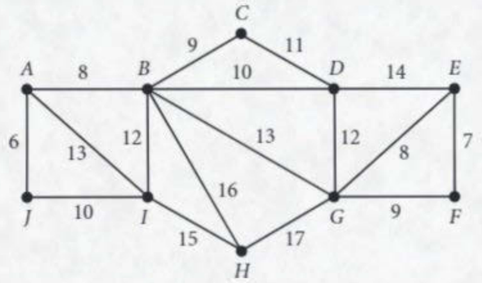
B It contains more than one Eulerian trail.

C It contains an Eulerian circuit.

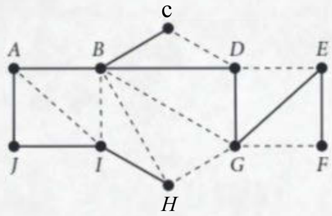
D Eulers formula $v+f=e+2$ could not be used.

E The addition of one further edge could create an Eulerian trail.

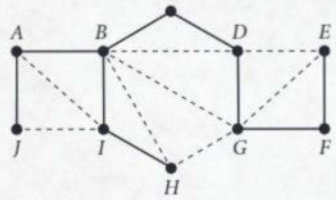
16 ©VCAA 2016S 1NQ4 Which one of the following is the minimal spanning tree for the weighted graph shown?



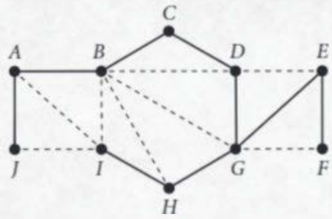

A



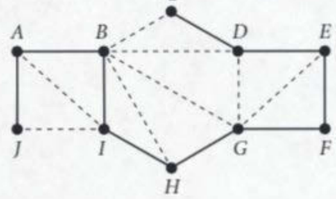

B



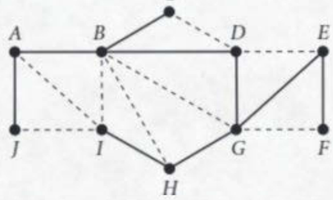

C



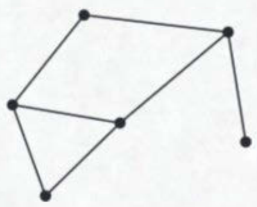

D




E




17 GVCAA 2021 1NQ5 Consider the following five statements about the graph:



- The graph is planar.
- The graph contains a cycle.
- The graph contains a bridge.
- The graph contains an Eulerian trail.
- The graph contains a Hamiltonian path.

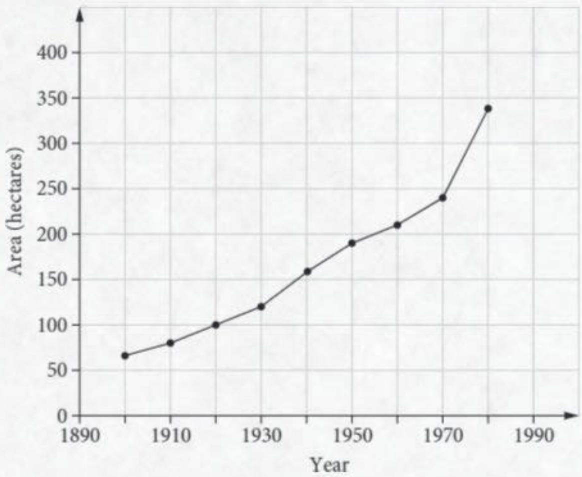
How many of these statements are true?

Al B 2 C 3 D 4 E 5

# Cumulative examination 2

Total number of marks: 32 Reading time: 9 minutes Writing time: 48 minutes

**1** ©VCAA 2017 2CQ4I (5 marks) The following time series plot shows the total *area,* in hectares, of forest eaten by caterpillars in a rural area during the period 1900 to 1980. The data used to generate this plot is also given.




| Year | 1900 | 1910 | 1920 | 1930 | 1940 | 1950 | 1960 | 1970 | 1980 |
|---|---|---|---|---|---|---|---|---|---|
| Area (hectares) | 66 | 80 | 100 | 120 | 160 | 190 | 210 | 240 | 340 |

The association between *area* of forest eaten by the caterpillars and year is non-linear. A $\log_{10}$ transformation can be applied to the variable *area* to linearise the data.

a When the equation of the least squares line that can be used to predict $\log_{10}$ (area) from *year* is determined, the slope of this line is approximately 0.008 538 5. Round this value to three significant figures. 1 mark

b Perform the $\log_{10}$ transformation to the variable *area* and determine the equation of the least squares line that can be used to predict $\log_{10}$ (area) from *year.* Copy and complete the following by writing the values of the intercept and slope of this least squares line in the appropriate boxes. Round your answers to three significant figures. 2 marks

$$\log_{10}(area) = \square + \square \times year$$

c i The least squares line predicts that the $\log_{10}$ (area) of forest eaten by the caterpillars by the year 2020 will be approximately 2.85. Using this value of 2.85, calculate the expected area of forest that will be eaten by the caterpillars by the year 2020. Round your answer to the nearest hectare. 1 mark

ii Give a reason why this prediction may have limited reliability. 1 mark

 9780170448406

2 ©VCAA 2004 2BRM Q2bcdeii MODIFIED (4 marks) Anna borrows $12 000 at 7.5% interest per annum, compounding monthly. The loan is to be fully repaid over four years by equal monthly repayments,

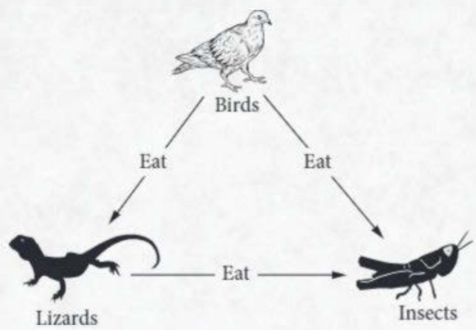
a Determine the monthly repayment for this loan. Write your answer correct to the nearest cent. 1 mark

b Determine the total amount of interest paid on the loan after four years, 1 mark

c After six equal repayments have been made, how much has Anna paid off the loan? Write your answer correct to the nearest dollar. 1 mark

d At the end of six months the interest rate increases to 8.0% per annum. Anna still has to completely pay out the balance of the loan completely within the original period of the loan. Determine the new monthly repayments that now apply. Write your answer correct to the nearest dollar. 1 mark

3 ©VCAA 2011 2MQ1 | (3 marks) The diagram shows the feeding paths for insects (Z), birds (B) and lizards (L). The matrix *E* has been constructed to represent the information in this diagram. In matrix £, a T is read as eat' and a 'O' is read as 'do not eat'.



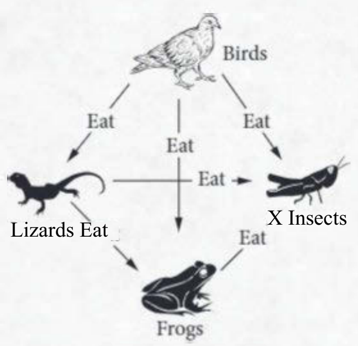

$$Z = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{matrix} I \\ B \\ L \end{matrix} \quad (\text{columns } I\ B\ L)$$

a Referring to insects, birds or lizards

i what does the T' in column B, row L, of matrix *E* indicate? 1 mark

ii what does the row of zeros in matrix *E* indicate? 1 mark

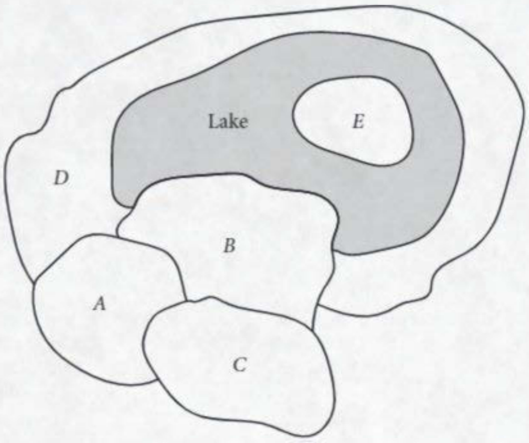
The diagram below shows the feeding paths for insects (Z), birds (B), lizards (L) and frogs ($Z^7$). The matrix Z has been set up to represent the information in this diagram. Matrix *Z* has not been completed.




$$Z = \begin{bmatrix} 0 & 1 & 1 & \\ 0 & 0 & 0 & \_ \\ 0 & 1 & 0 & \_ \\ \_ & \_ & \_ & \_ \end{bmatrix} \begin{matrix} I \\ B \\ L \\ F \end{matrix} \quad (\text{columns } I\ B\ L\ F)$$

b Copy and complete the matrix Z by writing in the seven missing elements. 1 mark

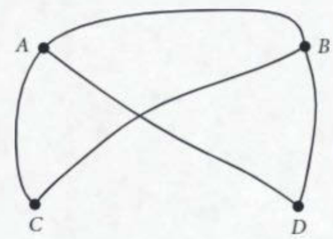
4 ©VCAA 2009 2NQ1 ? (3 marks) The city of Robville is divided into five suburbs labelled as *A* to *E* on the map. A lake that is situated in the city is shaded on the map. An adjacency matrix is constructed to represent the number of land borders between the suburbs.




| | A B C D E |
|---|---|
| *A* | 0 1110 |
| *B* | 10 12 0 |
| *C* | 110 0 0 |
| *D* | 1 2 0 0 0 |
| *E* | 0 0 0 0 0 |

a Explain why all values in the final row and final column are zero. 1 mark

In the network diagram, vertices represent suburbs and edges represent land borders between suburbs. The diagram has been started but is not finished.

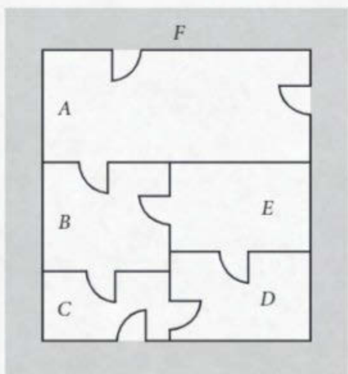


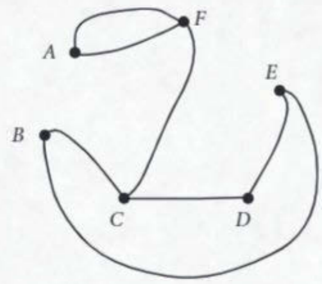

b The network diagram is missing one edge and one vertex. Copy the diagram, and on it

i draw the missing edge 1 mark

ii draw and label the missing vertex. 1 mark

5 ©VCAA | 2021 2NQ1 J (4 marks) Maggies house has five rooms, A, B, C, *D* and F, and eight doors. The floor plan of these rooms and doors is shown below. The outside area, F, is shown shaded on the floor plan. The floor plan is represented by the graph. On this graph, vertices represent the rooms and the outside area. Edges represent direct access to the rooms through the doors. One edge is missing from the graph.




a Copy the graph and draw the missing edge. 1 mark

b What is the degree of vertex E? 1 mark

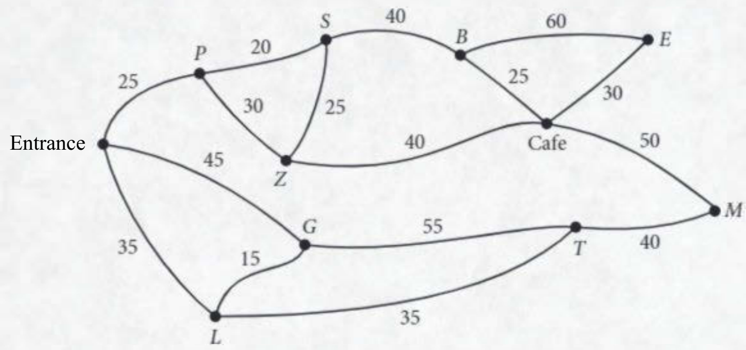
c Maggie hires a cleaner to clean the house. It is possible for the cleaner to enter the house from the outside area, F, and walk through each room only once, cleaning each room as he goes and finishing in the outside area, F.

i Copy and complete the following to show one possible route that the cleaner could take.

*F* – □ – □ – □ – □ – □ – *F* 1 mark

ii What is the mathematical term for such a journey? 1 mark

6 ©VCAA | 2019N2NQ1 | (5 marks) A zoo has an entrance, a cafe and nine animal exhibits: bears (B), elephants (E), giraffes (G), lions (L), monkeys (Af), penguins (P), seals (S), tigers (T) and zebras (Z). The edges on the graph below represent the paths between the entrance, the cafe and the animal exhibits. The numbers on each edge represent the length, in metres, along that path. Visitors to the zoo can use only these paths to travel around the zoo.

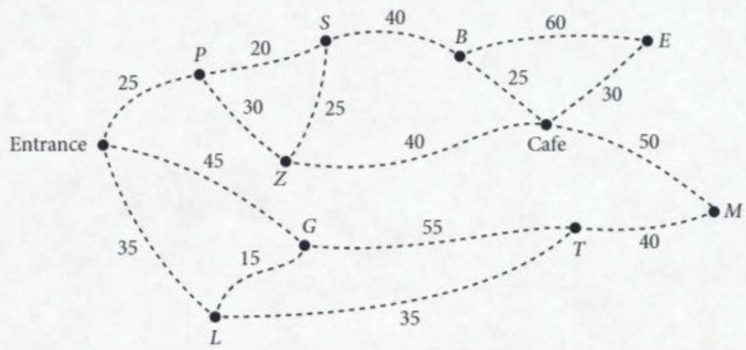



a What is the shortest distance, in metres, between the entrance and the seal exhibit (S)? 1 mark

b Freddy is a visitor to the zoo. He wishes to visit the cafe and each animal exhibit just once, starting and ending at the entrance.

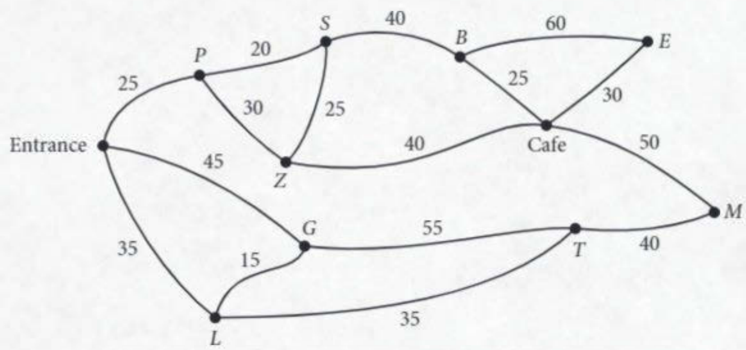
i What is the mathematical term used to describe this route? 1 mark

ii Copy the graph below and draw one possible route that Freddy may take. 1 mark



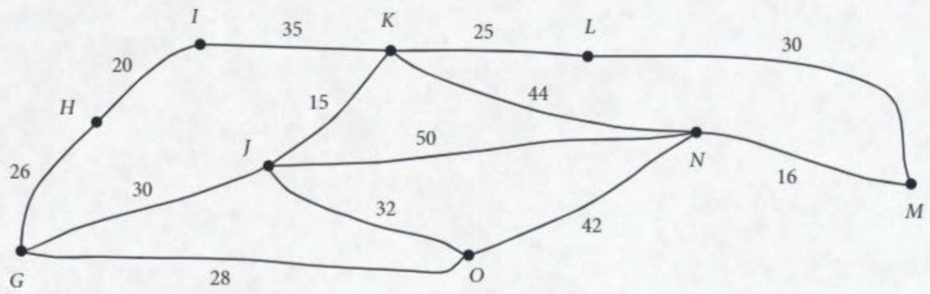

A reptile exhibit *(R)* will be added to the zoo. A new path of length 20 m will be built between the reptile exhibit *(R)* and the giraffe exhibit (G). A second new path, of length 35 m, will be built between the reptile exhibit *(R)* and the cafe,

c Copy the graph below and complete it with the new reptile exhibit and the two new paths added. Label the new vertex *R* and write the distances on the new edges. 1 mark




d The new paths reduce the minimum distance that visitors have to walk between the giraffe exhibit (G) and the cafe. By how many metres will these new paths reduce the minimum distance between the giraffe exhibit (G) and the cafe? 1 mark

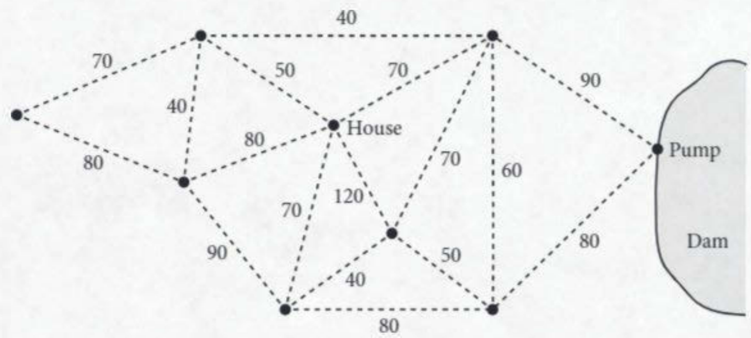
7 ©VCAA 2021 2NQ2 I (2 marks) George lives in Town *G* and Maggie lives in Town *M*.
The diagram shows the network of main roads between Town *G* and Town *M*. The vertices G, $H_y$ *Iy J, Ky Ly My N* and O represent towns. The edges represent the main roads. The numbers on the edges indicate the distances, in kilometres, between adjacent towns.




a What is the shortest distance, in kilometres, between Town *G* and Town Af? 1 mark

b George plans to travel to Maggies house. He will pass through all the towns shown above. George plans to take the shortest route possible. Which town will George pass through twice? 1 mark

8 ©VCAA | 2016S2NQI3 (6 marks) Water will be pumped from a dam to eight locations on a farm. The pump and the eight locations (including the house) are shown as vertices in the network diagram. The numbers on the edges joining the vertices give the shortest distances, in metres, between locations.




a i Determine the shortest distance between the house and the pump. 1 mark

ii How many vertices on the network diagram have an odd degree? 1 mark

iii The total length of all edges in the network is 1180 m. A journey starts and finishes at the house and travels along every edge in the network. Determine the shortest distance travelled. 1 mark

iv A Hamiltonian path, beginning at the house, is determined for this network. How many edges does this path involve? 1 mark

The total length of pipe that supplies water from the pump to the eight locations on the farm is a minimum. This minimum length of pipe is laid along some of the edges in the network,

b i Use the network diagram above to draw the minimum length of pipe that is needed to supply water to all locations on the farm. 1 mark

ii What is the mathematical term that is used to describe this minimum length of pipe in part b i? 1 mark


9780170448406

CHAPTER

# 10

# DIRECTED GRAPHS

## Study Design coverage

### AREA OF STUDY 2: DISCRETE MATHEMATICS

#### Graphs and networks

- the concepts, conventions and terminology of directed graphs (digraphs) and networks
- use of matrices to represent digraphs and networks and their application.

#### Flow problems

- use of networks to model flow problems: capacity, sinks and sources
- solution of small-scale network flow problems by inspection and the use of the 'maximum-flow minimum-cut' theorem to aid the solution of larger scale problems.

#### Shortest path problems

- determination of the shortest path between two specified vertices in a digraph or network by inspection
- Dijkstra's algorithm and its use to determine the shortest path between a given vertex and each of the other vertices in a weighted graph or network.

#### Matching problems

- use of a bipartite graph and its tabular or matrix form to represent a matching problem
- determination of the optimum assignment(s) of people or machines to tasks by inspection or by use of the Hungarian algorithm for larger scale problems.

#### Scheduling problems and critical path analysis

- construction of an activity network from a precedence table (or equivalent) including the use of dummy activities where necessary
- use of forward and backward scanning to determine the earliest starting times (EST) and latest starting times (LST) for each activity
- use of earliest starting times and latest starting times to identify the critical path in the network and determine the float times for non-critical activities
- use of crashing to reduce the completion time of the project or task being modelled.



**Video playlists (6):**

**10.1** The scheduling problem
**10.2** Critical path analysis
**10.3** The assignment problem and bipartite graphs
**10.4** Network flow problems
**10.5** Shortest path problems
**VCE question analysis** Directed graphs

**Worksheets (8):**

**10.1** Project networks • Drawing directed graphs
**10.2** Calculating EST and LST • Critical paths • Critical path analysis • Critical paths and activity float times
**10.3** The Hungarian algorithm
**10.4** Network flow capacity

Nelson MindTap

To access resources above, visit cengage.com.au/nelsonmindtap



9780170448406

# 10.1 The scheduling problem

## Directed graphs

In the previous chapter, we learned that graphs show connections using vertices and edges, and we explored undirected graphs with no arrows on the edges. In this chapter we will learn about directed graphs (or digraphs) where each edge has a direction, shown by an arrowhead. Connections on a directed graph can only go in the direction of the arrows.

Directed graphs can be used to schedule tasks made up of many activities. Each vertex is shown by a circle representing the start and finish of a particular event. An immediate predecessor is any activity that must be completed before the current activity can commence.

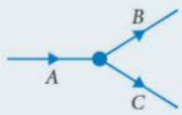
**Different types of connections in a directed graph**

**One preceding event**

Event *B* is preceded by *A*.

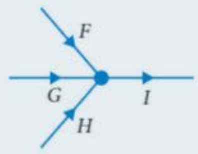

**More than one activity with the same preceding event**

Events *B* and *C* are preceded by *A*.




**An activity with more than one preceding event**

Event *I* is preceded by *F*, *G* and *H*.




## Digraphs with direction and weight

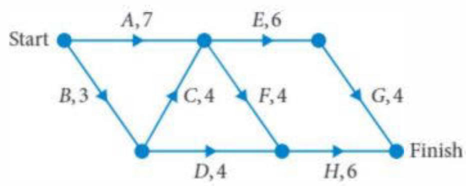
In the previous chapter, we learned about weighted graphs that show extra information on the edges. A weighted directed graph can be used to represent an activity. When a large number of activities are involved in a task, it is often useful to draw a weighted directed graph to show the sequence of activities and the time required for each activity.

Each activity is represented by an edge, shown by a line with an arrow, and is labelled with a letter and a weighting, which represents the time taken to complete the activity. The unit of time will be given in the question in minutes, hours, days etc. In the weighted digraph shown on the right, Activity A takes 5 hours to complete.


## Activity tables

An activity table shows the order and estimated time for each activity. The table below shows eight activities (A to *H*) that must be completed in a task. The directed graph that corresponds to the activity table is shown on the right.

| Activity | Activity time (hours) | Immediate predecessor |
|---|---|---|
| *A* | 7 | – |
| *B* | 3 | – |
| *C* | 4 | *B* |
| *D* | 4 | *B* |
| *E* | 6 | *A*, *C* |
| *F* | 4 | *A*, *C* |
| *G* | 4 | *E* |
| *H* | 6 | *D*, *F* |




10.1

Video playlist The scheduling problem

Worksheets Project networks

Drawing directed graphs

## Guidelines to follow when drawing a network diagram

- Use a vertex to represent the start of the network and label as start’.‘
- Look for activities that do not have any predecessors. These will be your starting activities.
- Multiple predecessors to an activity will all end at the same vertex.
- An activity should not be represented by more than one edge in the network.
- Two vertices can be connected by one edge only.
- A vertex indicating the completion of the project needs to be included in the network and labelled as ‘finish’.

WB

p. 194

### WORKED EXAMPLE 1 Drawing a directed graph from an activity table

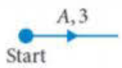
Draw the directed graph for the activity table.

| Activity | Activity time (hours) | Immediate predecessor |
|---|---|---|
| $A$ | 3 | – |
| $B$ | 4 | $A$ |
| $C$ | 3 | $A$ |
| $D$ | 8 | $A$ |
| $E$ | 3 | $B$ |
| $F$ | 5 | $C$ |
| $G$ | 7 | $E, D, F$ |

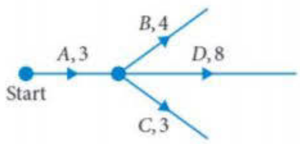
**Steps** | **Working**

**1** Activity $A$ has no predecessors.

Label the first vertex ‘start’ and from this draw a directed edge labelled $A, 3$.



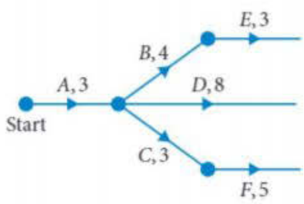

**2** Activities $B$, $C$ and $D$ are all preceded by $A$. From the vertex at the end of edge $A$, draw three directed edges labelled $B, 4$, $C, 3$ and $D, 8$.



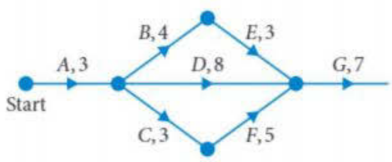

**3** Activity $E$ is preceded by $B$. From the vertex at the end of edge $B$, draw an edge labelled $E, 3$.

Activity $F$ is preceded by $C$.

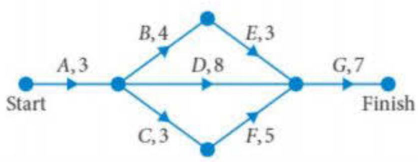
From the vertex at the end of $C$, draw an edge labelled $F, 5$.




**4** Activity $G$ is preceded by activities $E$, $F$ and $D$. Draw a vertex that is connected to the edges $E$, $F$ and $D$. From this vertex, draw an edge labelled $G, 7$.




**5** The final vertex is drawn at the end of edge $G$ and labelled ‘finish’.






9780170448406

## Reachability

**Reachability**

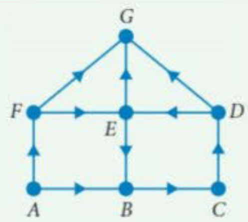
Reachability is the ability to get from one vertex to another vertex in a directed graph.

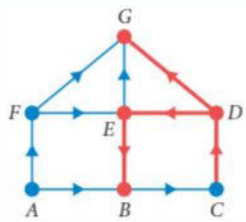
10.1

WB p. 195

**WORKED EXAMPLE 2** Finding the reachability of a vertex

In the directed graph, determine which vertices are not reachable from vertex C.




| Steps | Working |
|---|---|
| 1 Trace the paths from vertex C to every other vertex. The vertices that can be reached are coloured red and the paths have red edges. |  Vertices B, D, *E* and *G* are reachable from vertex C. |
| 2 If a path does not exist from vertex C to a vertex, it is not reachable. | Vertices *A* and *F* are not reachable from vertex C. |

## Dummy activities

A dummy activity is an imaginary or redundant activity that is added to a directed graph to ensure that no two vertices are connected by multiple edges or to maintain precedence structure.

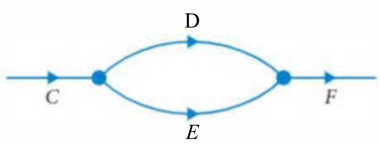
A dummy activity has zero time and is shown as a directed edge with a broken line.

### 1 Multiple edges

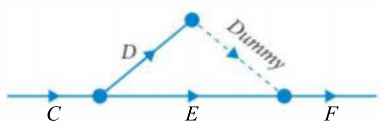
A directed graph for a project cannot be drawn with two activities that have the same beginning and the same end. This causes a loss of identity to the activities and results in errors during network computations. The problem can be solved using a dummy activity.

In the example below, the directed graph cannot be drawn with two vertices both connected by activities *D* and *E*. A dummy activity needs to be included. It can occur before or after activities *D* or *E* to overcome this problem.

| Activity | Immediate predecessor |
|---|---|
| *D* | *C* |
| *E* | *C* |
| *F* | *D*, *E* |



Incorrect

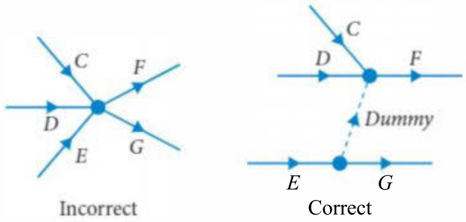


Correct

### 2 Logic difficulties in the precedence structure

The second use of dummy activities is to ensure the logic of the network is maintained. The first network has activity *G* preceded by activities C, *D* and E, which does not follow the logic of the activity table. A second vertex and a dummy activity needs to be included so activities *F* and *G* do not both start from the same vertex and activities C, *D* and *E* do not finish at the same vertex.

| Activity | Immediate predecessor |
|---|---|
| *F* | *C*, *D*, *E* |
| *G* | *E* |




## WORKED EXAMPLE 3 Drawing a directed graph with dummy activities

WB p. 196

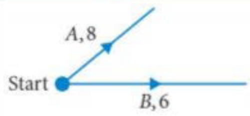
Draw the directed graph for the project shown below, including dummy activities where required.

| Activity | Activity time (hours) | Immediate predecessor |
|---|---|---|
| *A* | 8 | – |
| *B* | 6 | – |
| *C* | 4 | *A*, *B* |
| *D* | 5 | *A*, *B* |
| *E* | 7 | *C*, *D* |

**Steps** | **Working**

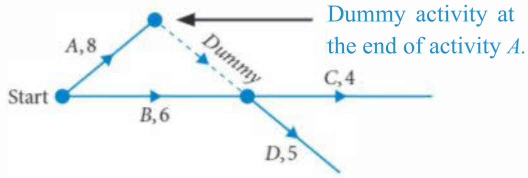
1 Activities *A* and *B* have no predecessors.

Label the first vertex 'start' and from this draw directed edges labelled A, 8 and B, 6.

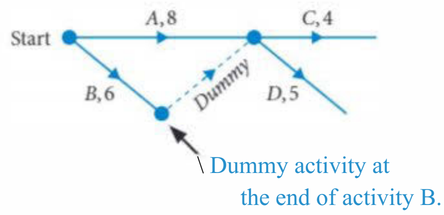



2 Activities C and *D* are both preceded by *A* and B; however, two vertices cannot be connected by multiple edges. A dummy activity must be included after activity *A* or activity B.

If the dummy activity is placed at the end of activity A, connect the end of activity B and the dummy to the same vertex. Draw two directed edges labelled C, 4 and D, 5 from this vertex.

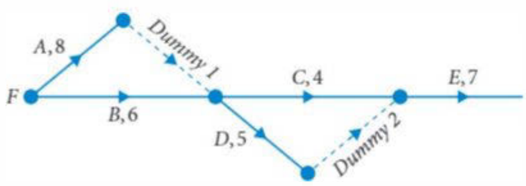



The dummy activity could also be placed at the end of activity B, which would produce




3 Activity *E* is preceded by activities C and *D*. A dummy activity is required so that two vertices are not connected by multiple edges.

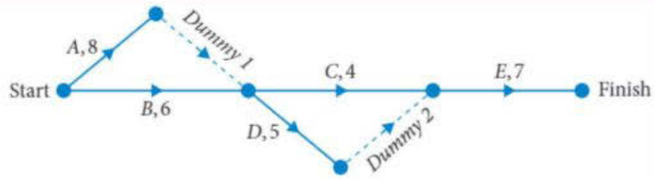
Using the directed graph with dummy activity after A, draw Draw a directed edge labelled E, 7 from the vertex at the end of activity C. As there are two dummy activities, label the first 'dummy T and the second 'dummy 2'.





9780170448406

**4** Draw a vertex labelled 'finish' at the end of activity *E*.




10.1

## Exam hack

Always check your directed graph to ensure that

- it contains all of the connections and times specified in the activity table
- it doesn't have two vertices connected with multiple edges
- it has a start vertex and a finish vertex.

## EXERCISE 10.1 The scheduling problem

ANSWERS p. 726

### Mastery

1 H WORKED EXAMPLE 1 I For a particular project, nine activities must be completed.

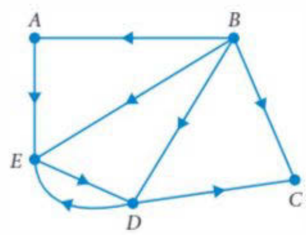
These activities, the activity times and their immediate predecessors are given in the following table.

| Activity | Activity time (days) | Immediate predecessor |
|---|---|---|
| *A* | 1 | – |
| *B* | 4 | – |
| *C* | 6 | *A* |
| *D* | 4 | *A* |
| *E* | 7 | *B*, *C* |
| *F* | 5 | *B*, *C* |
| *G* | 7 | *F* |
| *H* | 6 | *D* |
| *I* | 2 | *G* |

Draw the directed graph for this project.

2 H WORKED EXAMPLE 2 I The directed graph shows a series of rivers that flow between logging towns A, B, C, *D* and *E*. Logs from the towns' lumber yards are transported via the rivers. The direction of flow in each river is indicated on the edge that represents the river.

Which towns can the lumber yard at town *E* transport to?



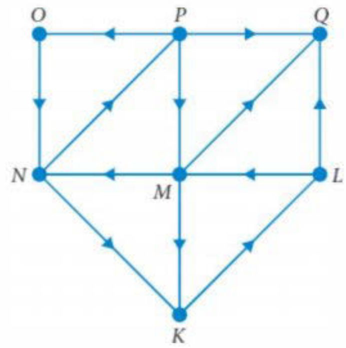

30 WORKED EXAMPLE 3 I Draw the directed graph for the project shown, including dummy activities where required.

| Activity | Activity time (hours) | Immediate predecessor |
|---|---|---|
| *A* | 8 | – |
| *B* | 4 | – |
| *C* | 7 | – |
| *D* | 5 | *A*, *B*, *C* |
| *E* | 7 | *A*, *B*, *C* |

## Exam practice

80-100% 60-79% 0-59%

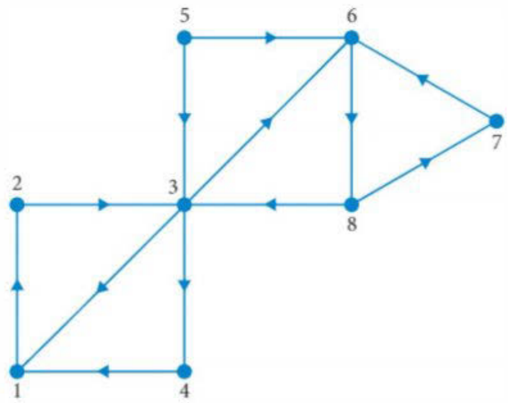
4 CVCAA 2003 1NQ2 89% For the directed graph shown, vertex O can not be reached from vertex




A $L$ B $M$ C N

D $P$ E $Q$

5 CVCAA 20061NQ2 87% The directed graph represents a series of one-way streets with intersections numbered as nodes 1 to 8.




All intersections can be reached from

A intersection 4

B intersection 5

C intersection 6

D intersection 7

E intersection 8

6 ©VCAA 2006 1NQ5 80%

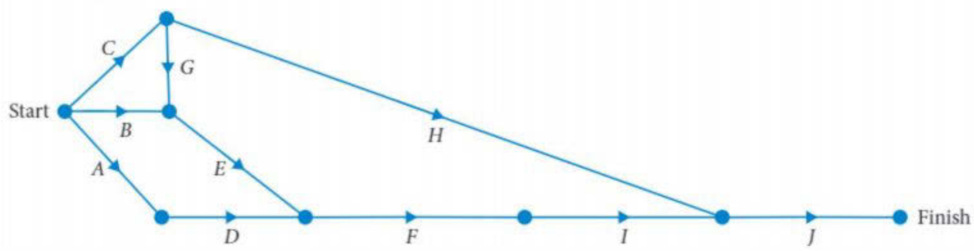
For a particular project there are ten activities that must be completed.

These activities and their immediate predecessors are given in the following table.

| Activity | Immediate predecessors |
|---|---|
| $A$ | – |
| $B$ | – |
| $C$ | – |
| $D$ | $A$ |
| $E$ | $B$ |
| $F$ | $D, E$ |
| $G$ | $C$ |
| $H$ | $C$ |
| $I$ | $F, G$ |
| $J$ | $H, I$ |

A directed graph that could represent this project is

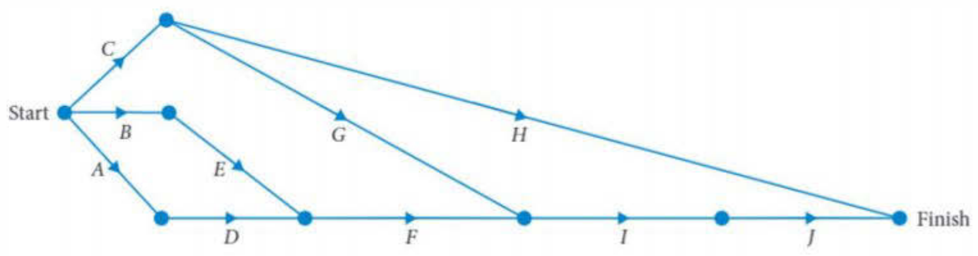
**A**



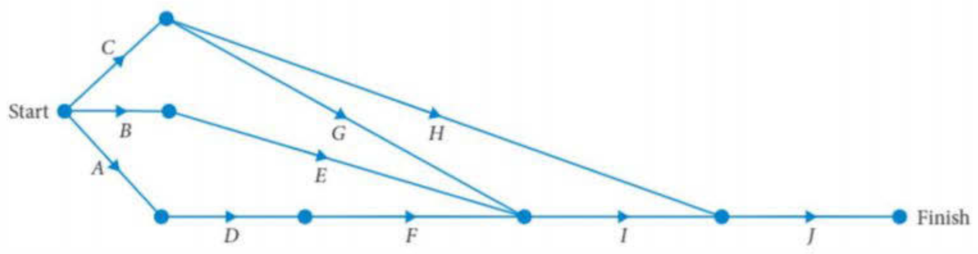



9780170448406

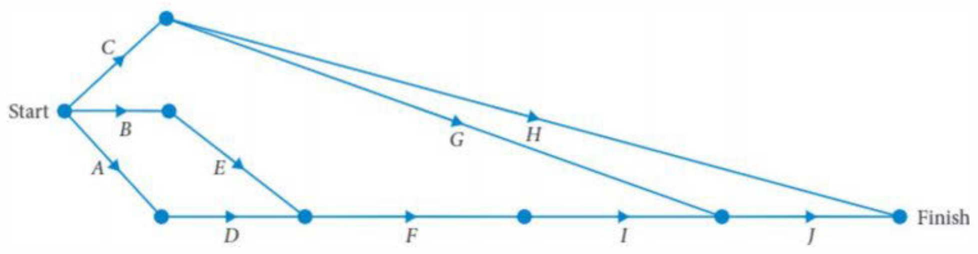
**B**



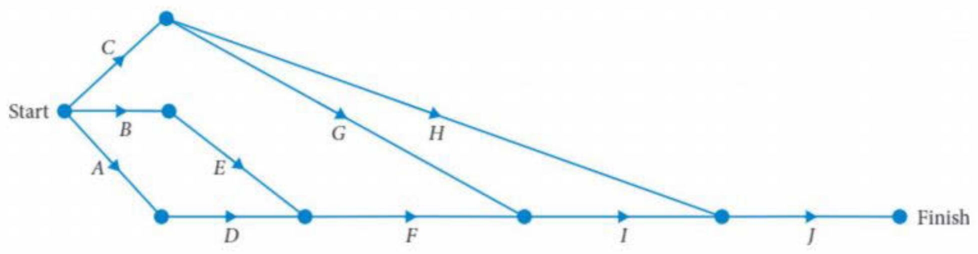

**C**



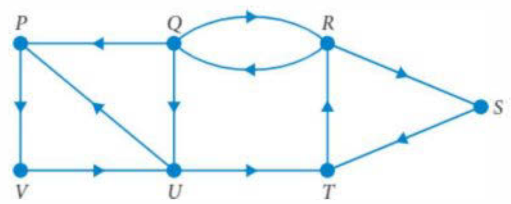

**D**




**E**



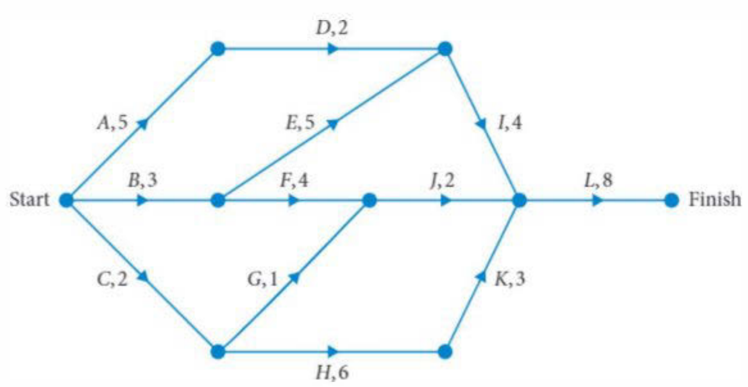

7 ©VCAA | 2O121NQ6~ 74% In the digraph, all vertices are reachable from every other vertex. All vertices would still be reachable from every other vertex if we remove the edge in the direction from




A $Q$ to $U$ B $R$ to S C StoT

D TtoR E Vtol7

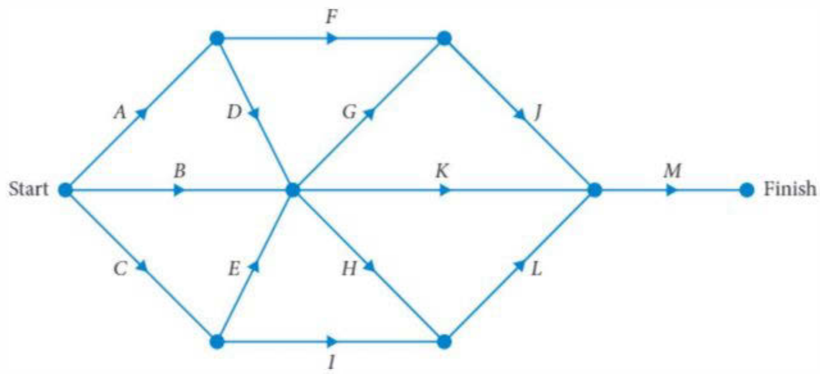
8 ©VCAA 2016 1NQ6 74% The directed graph below shows the sequence of activities required to complete a project. All times are in hours.




The number of activities that have exactly two immediate predecessors is

AO BI C 2 D 3 E 4

9 ©VCAA 2009 1NQ5 45% The network shows the activities that are needed to complete a particular project.




The total number of activities that need to be completed before activity *L* may begin is

A 2 B 4 C 6 D 7 E 8

10 ©VCAA 2019 1NQ7 42% A project involves nine activities, *A* to *I*. The immediate predecessor(s) of each activity is shown in the table.

A directed network for this project will require a dummy activity. The dummy activity will be drawn from the end of

A activity *B* to the start of activity C.

B activity *B* to the start of activity *E*.

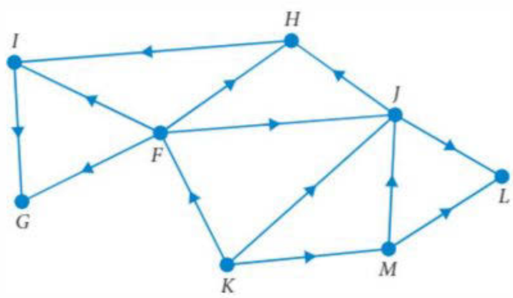
C activity *D* to the start of activity *E*.

D activity *E* to the start of activity *H*.

E activity *E* to the start of activity *F*.

| Activity | Immediate predecessors) |
|---|---|
| *A* | – |
| *B* | *A* |
| *C* | *A* |
| *D* | *B* |
| *E* | *B*, *C* |
| *F* | *D* |
| *G* | *D* |
| *H* | *E*, *F* |
| *I* | *G*, *H* |

11 ©VCAA 2009 2NQ2 92% (2 marks) One of the landmarks in the city is a hedge maze. The maze contains eight statues. The statues are labelled *F* to *M* on the following directed graph. Walkers within the maze are only allowed to move in the directions of the arrows.




a Write down the two statues that a walker could not reach from statue *M* 1 mark

b One way that statue *H* can be reached from statue *K* is along path *KFH*.

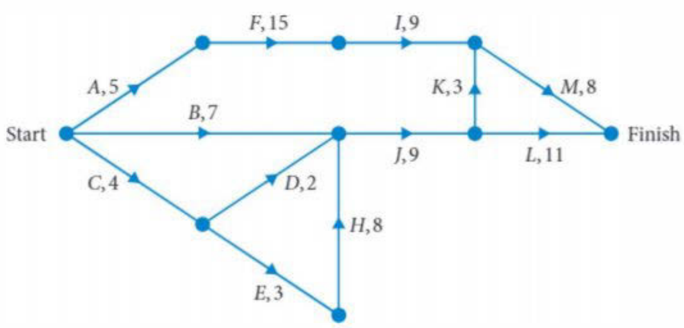
List the three other ways that statue *H* can be reached from statue *K*. 1 mark


9780170448406

▶ 12 ©VCAA J 20i3 2NQ2a J (1 mark) A project will be undertaken in a wildlife park. This project involves the 13 activities shown in the table below. The duration, in hours, and predecessor(s) of each activity are also included in the table.

| Activity | Duration (hours) | Predecessors) |
|---|---|---|
| *A* | 5 | - |
| *B* | 7 | - |
| *C* | 4 | - |
| *D* | 2 | C |
| *E* | 3 | C |
| *F* | 15 | *A* |
| *G* | 4 | *B, D,H* |
| *H* | 8 | *E* |
| *I* | 9 | *F,G* |
| *J* | 9 | *B,D,H* |
| *K* | 3 | *J* |
| *L* | 11 | *J* |
| *M* | 8 | *EK* |

Activity *G* is missing from the network diagram for this project, which is shown below.




Copy and complete the network diagram above by inserting activity *G*.

10.2

## @ Critical path analysis

Video playlist Critical path analysis

Critical path analysis is a step-by-step project management technique that is used to examine every activity in a project and how each activity affects the project completion time.

### Forward scanning to determine EST

Forward scanning through a network enables us to determine the earliest start time (EST) for every activity in the network. The earliest start time (EST) is the earliest time it is possible to start an activity.

## Method for finding the EST

To determine the earliest start times:

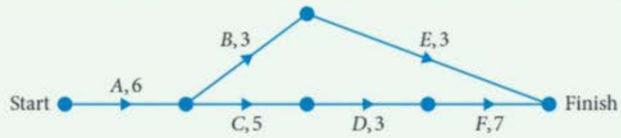
1. Draw a double box at the start vertex for each activity.
2. Begin with the first vertex that has an earliest start time of zero. We will use the convention of writing the earliest starting times in the top box next to each vertex.
3. ESTs are calculated from left to right.
4. Add the activity time to the EST of the previous vertex. If more than one activity leads to the vertex, the highest figure obtained becomes the new EST.
5. Continue until the finish is reached.

WB

p. 197

### WORKED EXAMPLE 4 Finding earliest start times

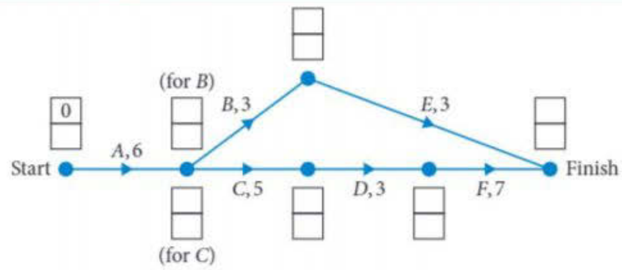
Determine the earliest start times (EST) for each activity in the network shown. Activity times shown are in hours.




**Steps** | **Working**

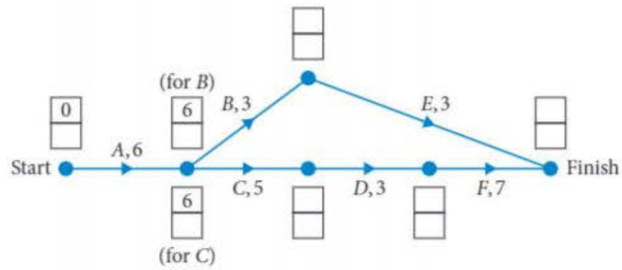
**1** Draw boxes above each vertex and enter 0 as the EST of the 'start' vertex.

Double boxes are required for both activities $B$ and $C$.




**2** To find the EST for the second vertex, add the activity time of 6 to the previous EST.

This gives an earliest start time for activities $B$ and $C$ as 6 hours.

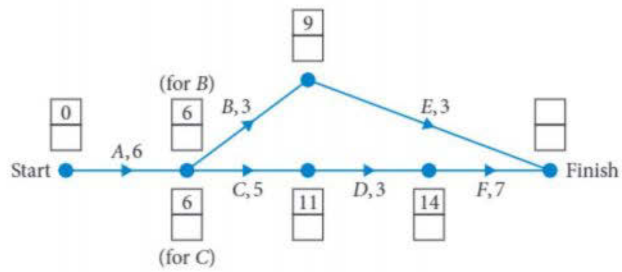



**3** The EST for activity $E$ is $6 + 3 = 9$. This is written in the top box at the start of activity $E$.

This process is repeated:

EST for activity $D = 6 + 5 = 11$

EST for activity $F = 11 + 3 = 14$

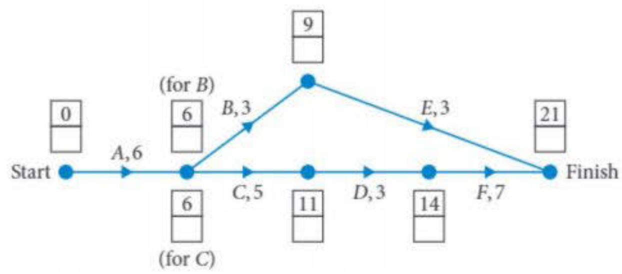



**4** To calculate the EST for the finish, find the total for the paths containing activity $E$ and the path containing activity $F$. The highest figure obtained is the final EST.

Activity $E$: $9 + 3 = 12$

Activity $F$: $14 + 7 = 21$

Therefore, 21 hours is the final EST.



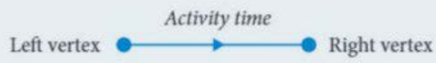
## Backward scanning to determine LST

Backward scanning is the process used to find the latest start times (LST) for an activity. This is the latest time you can start an activity without affecting the project completion time.

10.2

### Latest start times

The LST at the left vertex = the LST at the right vertex − the activity time




### Method for finding the LST

To determine latest start times (LST):

1. Commence LST calculations at the 'finish' vertex. At the 'finish' vertex, LST = EST.
2. Work backwards from right to left. LSTs are written in the bottom box next to each vertex.
3. To find the LST at the left vertex, work backwards subtracting the activity time from the LST at the right vertex. Where there is more than one LST, subtract the activity time from the lowest LST.
4. One of the LST values at the first vertex must be zero.

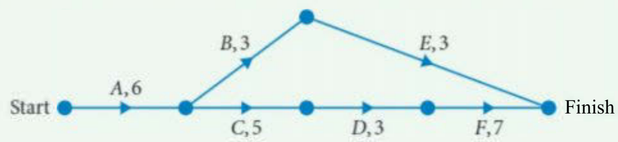
Worksheet Calculating EST and LST

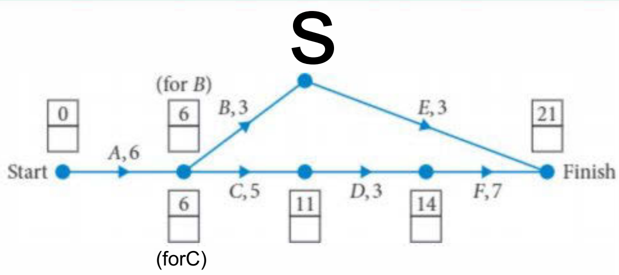
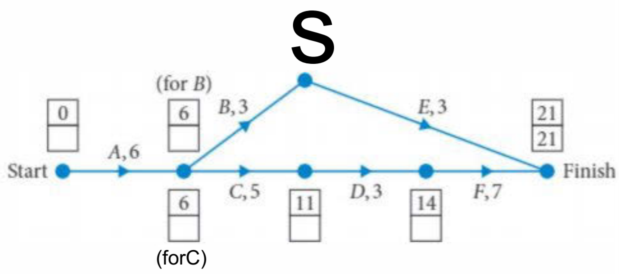
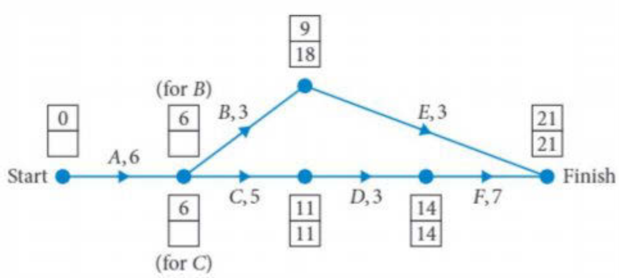
WB

p. 198

**WORKED EXAMPLE 5** Finding latest start times

Determine the latest start times (LST) for each activity in the network. Activity times shown are in hours.




| Steps | Working |
|---|---|
| 1 First calculate ESTs for each activity. These were completed in Worked example 4. |  |
| 2 Work backwards from right to left. The LST for the 'finish' is equal to the EST. |  |
| 3 The LST for activity E is 21 - 3= 18. This is written in the bottom box at the start of activity *E*.<br>The LST for:<br>activity F = 21 - 7 = 14<br>activity D = 14- 3=11 |  |

4 Calculate the LST for activities *B* and C.

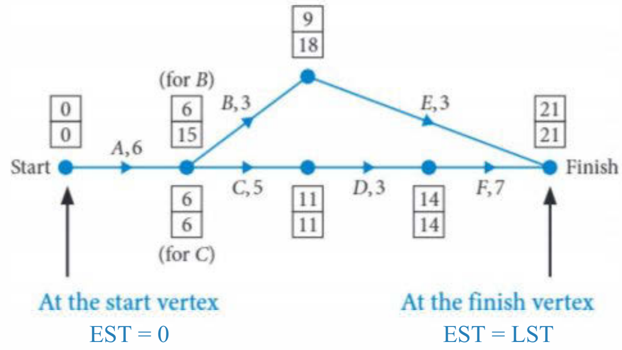
Activity *B*: 18 - 3 = 15

Activity C: 11 - 5 = 6

If there are two LST values at the end of an activity, then the smallest value is used to calculate the LST.

Therefore, use 6 hours to calculate the LST for the start.

The start LST = 6-6 = 0




## Identifying critical paths

Worksheets
Critical paths

Critical path analysis

There are often multiple paths between the start and finish *vertices* of a network. The critical path is the longest time path between the start and finish, and it determines the project completion time.

### The critical path

On the critical path, activities have equal EST and LST values.

Activities on the critical path are called critical activities.

Activities not on the critical path are called non-critical activities.

### Method for finding the critical path

To determine the critical path:

1 Use forward scanning to determine the earliest start times for each activity.

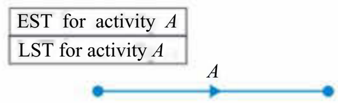
2 Identify the path or paths that produce the final EST.

3 Use backward scanning to determine the latest start times for each activity.

4 Activities for which EST = LST are on the critical path.

### Exam hack

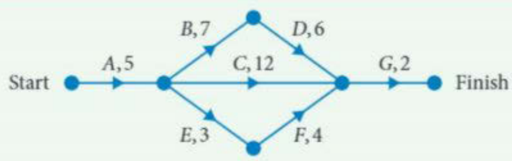
The earliest start time (EST) and latest start time (LST) that are associated with an activity are the values written in the boxes at the start of the activity.

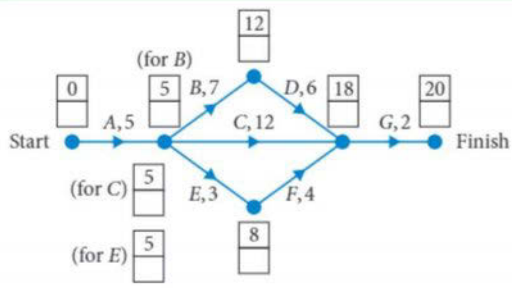



WB
p. 199

### WORKED EXAMPLE 6 Finding the critical path using EST and LST

Determine the earliest start times (EST) and latest start times (LST) for each activity in the network and hence, determine the critical path. Activity times shown are in hours.

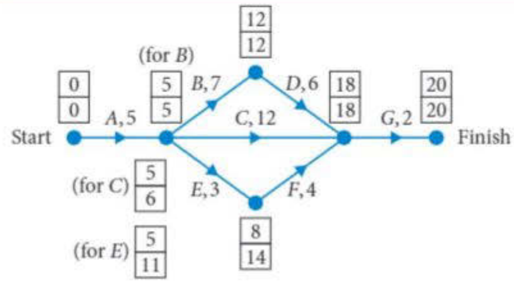



| Steps | Working |
|---|---|
| 1 First calculate ESTs for each activity. |  |



 9780170448406

2 Calculate LSTs for each activity.




3 Activities where EST = LST are critical activities and are on the critical path. | The critical path is *A–B–D–G*.

## Activity float time and project crashing

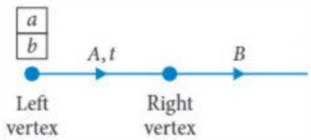
The float time for an activity is the maximum time an activity can be extended or postponed without affecting the project completion time. Activities on the critical path all have float times of zero.

In the diagram shown:

$a$ = the EST of activity $A$

$b$ = the LST for activity $A$

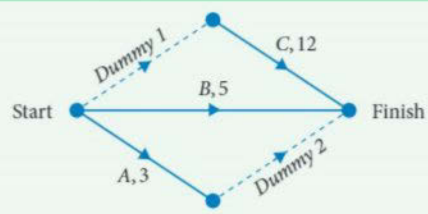
float for activity $A$ = $b$ - $a$

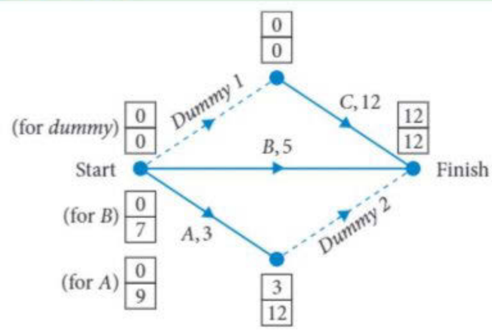



| Float time for an activity | |
|---|---|
| For a non-critical activity<br>float = LST - EST | For critical activities<br>float = 0 |

**WORKED EXAMPLE 7** Finding float times for a directed graph

Determine critical activities and the float times for the non-critical activities for the project shown. Activity times shown are in days.




| Steps | Working |
|---|---|
| 1 First calculate ESTs and LSTs for each activity. |  |
| 2 Identify the critical path.<br>On the critical path EST = LST. | The critical activity is *C*.<br>Dummy activities are NOT listed in the critical path. |
| 3 Calculate float times for the non-critical activities *A* and *B* using the formula<br>float = LST – EST | Float for activity *B* = 7 – 0 = 7 days<br>Float for activity *A* = 9 – 0 = 9 days |



10.2

Worksheet
Critical paths and activity float times

WB
p. 200

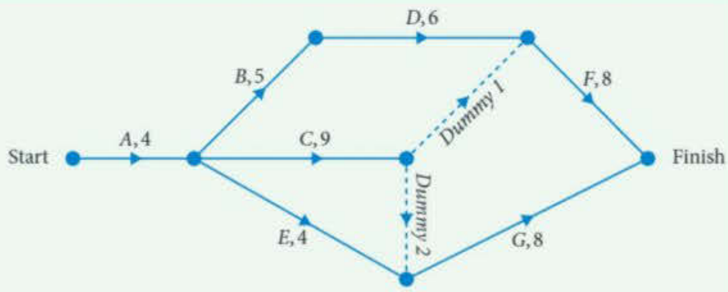
It is possible to speed up a project by employing more resources and this enables some activities to be completed more quickly. When this occurs with critical activities, the process is called crashing.

WB
p. 201

## WORKED EXAMPLE 8 Applying crashing to a directed graph

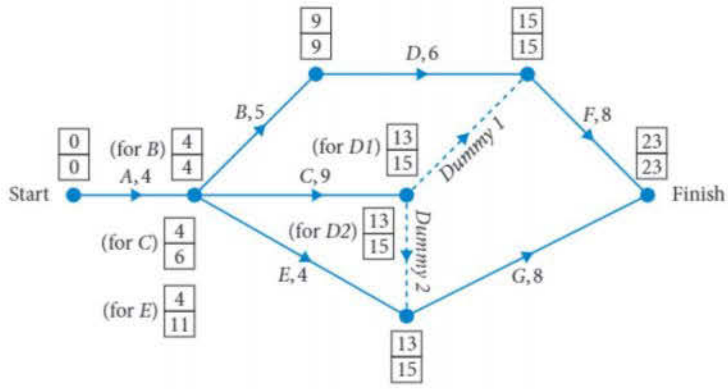
**a** Determine the critical path and the minimum project completion time for the project shown. Activity times shown are in days.

**b** Extra resources are used to speed up the original project resulting in activities *B*, *D* and *F* each being reduced by three days. Find the new critical path and the minimum project completion time.




**Steps** | **Working**

**a 1** First calculate ESTs and LSTs for each activity.




**2** On the critical paths EST = LST. The minimum project completion time is EST or LST for the finish vertex.

Critical path = *A*–*B*–*D*–*F*

Project completion time = 23 days

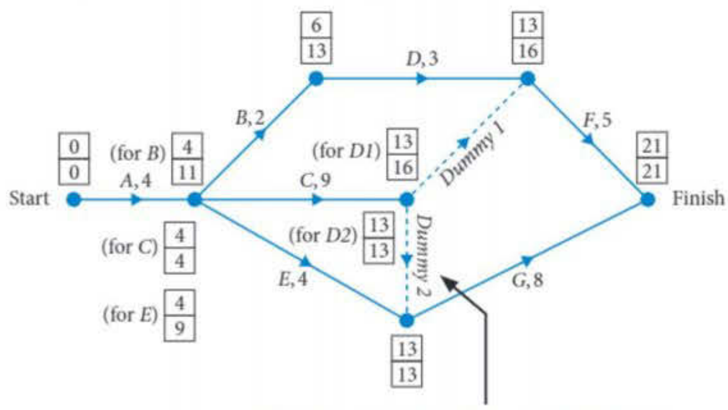
**b 1** Reduce the times for activities *B*, *D* and *F* by three days.

Activity time for *B* = 5 − 3 = 2 days

Activity time for *D* = 6 − 3 = 3 days

Activity time for *F* = 8 − 3 = 5 days

**2** Calculate ESTs and LSTs for each activity with the new network.




The dummy activity (*Dummy 2*) between the end of activity *C* and the start of activity *G* is not included in the critical path.

**3** Identify the activities where EST = LST to find the critical path.

The critical path is *A*–*C*–*G* and the project completion time is 21 days.

## EXERCISE 10.2 Critical path analysis

ANSWERS p. 727

10.2

### Recap

1 In the digraph, which of the statements is true?

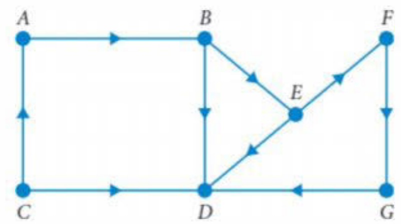
A Vertex A is reachable from vertex *B*.

B Vertex *B* is reachable from vertex *E*.

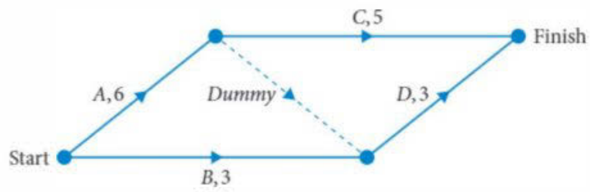
C Vertex *F* is not reachable from vertex *B*.

D Vertex *F* is not reachable from vertex C.

E Vertex *G* is reachable from vertex *A*.




2 The activity table for the directed graph shown is




**A**

| Activity | Time (hr) | Predecessor |
|---|---|---|
| *A* | 6 | – |
| *B* | 3 | – |
| *C* | 5 | *A* |
| *D* | 3 | *B* |

**B**

| Activity | Time (hr) | Predecessor |
|---|---|---|
| *A* | 6 | – |
| *B* | 3 | – |
| *C* | 5 | *A* |
| *D* | 3 | *A*, *B* |

**C**

| Activity | Time (hr) | Predecessor |
|---|---|---|
| *A* | 6 | – |
| *B* | 3 | – |
| *C* | 5 | *B* |
| *D* | 3 | *A* |

**D**

| Activity | Time (hr) | Predecessor |
|---|---|---|
| *A* | 6 | – |
| *B* | 3 | – |
| *C* | 5 | *A*, *B* |
| *D* | 3 | *B* |

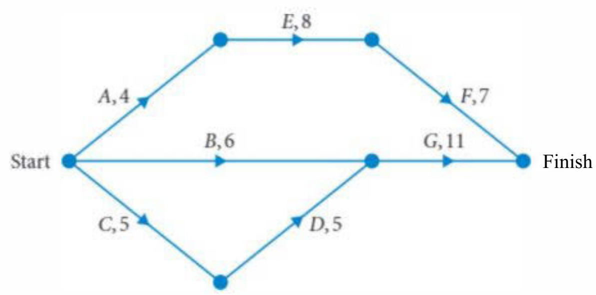
**E**

| Activity | Time (hr) | Predecessor |
|---|---|---|
| *A* | 6 | – |
| *B* | 3 | – |
| *C* | 5 | *B* |
| *D* | 3 | *A*, *B* |

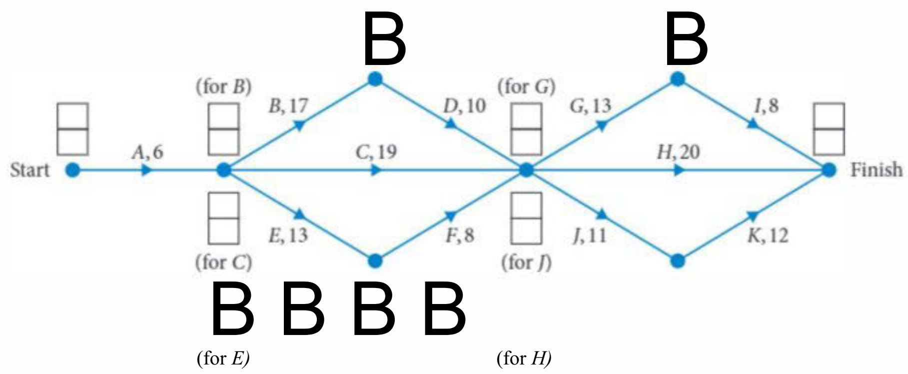
### Mastery

3 E3 WORKED EXAMPLE *ij* Determine the earliest start times (EST) for each activity in the directed graph on the right.

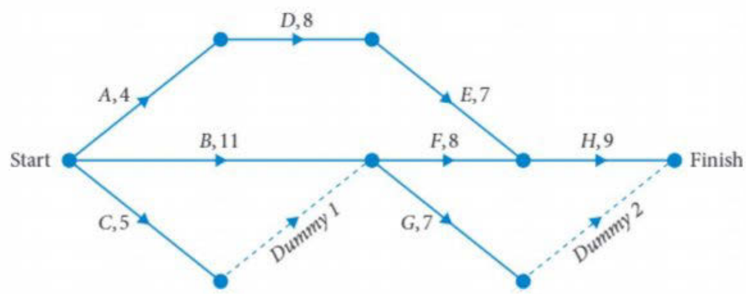
4 H WORKED EXAMPLE 5 J Determine the latest start times (LST) for the directed graph on the right.



5S WORKED EXAMPLE 6 Determine the earliest start times (EST) and latest starting times (LST) for each activity in the network below and hence, determine the critical path. Activity times shown are in hours.



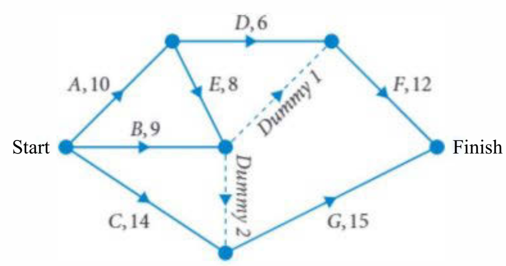

6S WORKED EXAMPLE 7 Determine critical activities and the float times for the non-critical activities for the project shown below. Activity times shown are in days.




7S WORKED EXAMPLE 8 J

a Determine the critical path and the minimum project completion time for the project shown. Activity times shown are in days.

b Extra resources are used to speed up the original project resulting in activities C, *E* and *F* each being reduced by 5 days. Find the new critical path and the minimum project completion time.

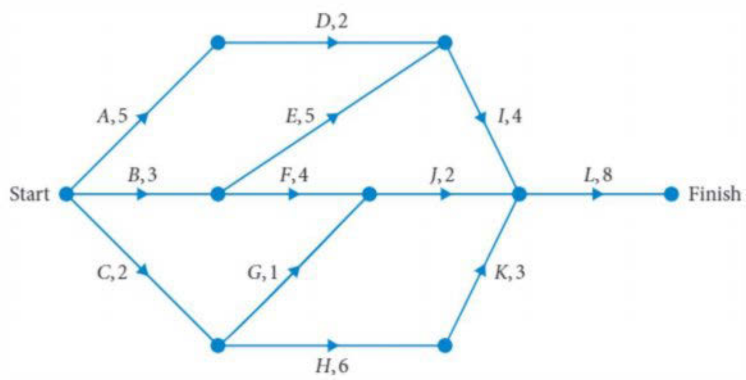



**Exam practice**

80-100% 60-79% 0-59%

*Use the following information to answer the next two questions.*

The directed graph shows the sequence of activities required to complete a project. All times are in hours.




8 ©VCAA | 2016 1NQ6 74% The number of activities that have exactly two immediate predecessors is

A 0 B 1 C 2 D 3 E 4

▶ 9 ©VCAA 20161NQ7 45% There is one critical path for this project.

Three critical paths would exist if the duration of activity

A *I* was reduced by two hours.

B *E* was reduced by one hour.

C *G* was increased by six hours.

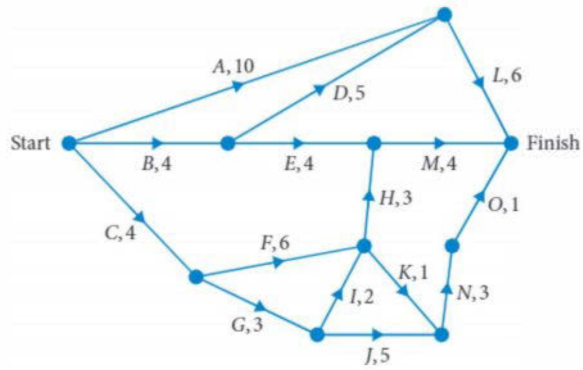
D *K* was increased by two hours.

E *F* was increased by two hours.

10.2

*Use the following information to answer the next two questions.*

The directed graph below shows the sequence of activities required to complete a project. The time to complete each activity, in hours, is also shown.




10 ©VCAA 20171NQ4 65% The earliest starting time, in hours, for activity *N* is

A 3 B 10 C 11 D 12 E 13

11 ©VCAA 2017 1NQ5 51% To complete the project in minimum time, some activities cannot be delayed. The number of activities that cannot be delayed is

A 2 B 3 C 4 D 5 E 6

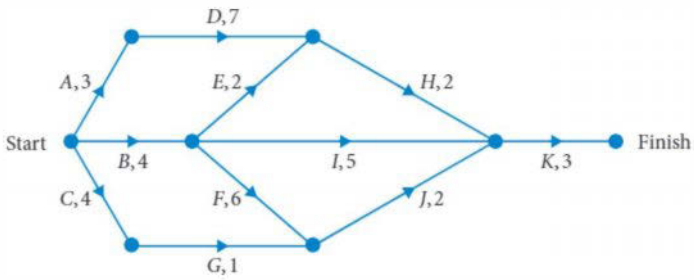
12 ©VCAA 20181NQ7 49% A project requires nine activities *(A-I)* to be completed. The duration, in hours, and the immediate predecessor(s) of each activity are shown in the table below.

| Activity | Duration (hours) | Immediate predecessors) |
|---|---|---|
| *A* | 4 | – |
| *B* | 3 | *A* |
| *C* | 7 | *A* |
| *D* | 2 | *A* |
| *E* | 5 | *B* |
| *F* | 2 | *C* |
| *G* | 4 | *E, F* |
| *H* | 5 | *D* |
| *I* | 3 | *G, H* |

The minimum completion time for this project, in hours, is

A 14 B 19 C 20 D 24 E 35 ▶

13 ©VCAA ^0181NQ5 43% The directed network shows the sequence of 11 activities that are needed to complete a project. The time, in weeks, that it takes to complete each activity is also shown.



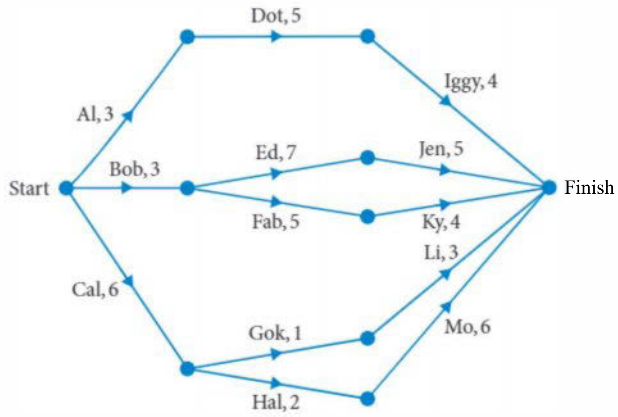

How many of these activities could be delayed without affecting the minimum completion time of the project?

A 3　　B 4　　C 5　　D 6　　E 7

14 ©VCAA 2018N1NQ4 The network shows the sequence of activities required to complete a project.

The name of the person completing each activity and the duration of the activity, in hours, are also shown on the network.

The project is to be completed in the minimum time possible.




The activity that has the latest starting time of 12 hours is completed by

A Iggy.　　B Jen.　　C Ky.　　D Li.　　E Mo.

15 ©VCAA 018N1NQ8 I A project consists of eight activities, *A* to *H*.

The table below shows the immediate predecessor(s) and earliest starting time, in hours, of each activity.

| Activity | Immediate predecessors) | Earliest starting time |
|---|---|---|
| HEBE | – | 0 |
| *B* | – | 0 |
| *C* | *A* | 4 |
| *D* | *B* | 12 |
| *E* | *C* | 15 |
| *F* | *D* | 19 |
| *G* | *E* | 27 |
| *H* | *F*, *G* | 36 |

It is known that activity *G* has a completion time of three hours.

The project can still be completed in minimum time if activity C is delayed. The maximum length of the delay for activity C is

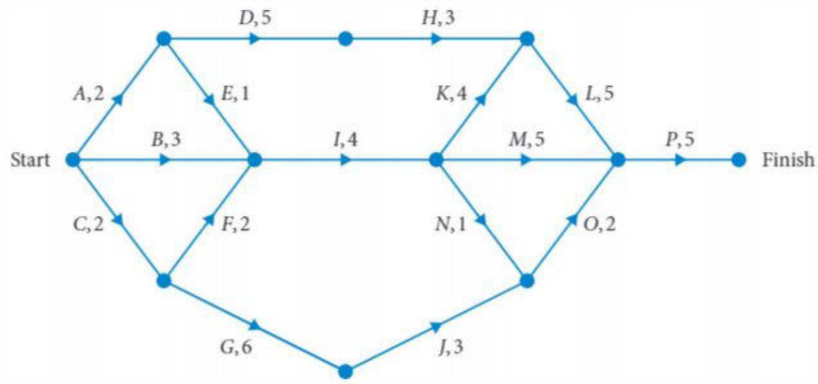
A two hours.　　B four hours.　　C five hours.

D six hours.　　E nine hours.

Use the following information to answer the next two questions.

The activity network shows the sequence of activities required to complete a project.

The number next to each activity in the network is the time it takes to complete that activity, in days.



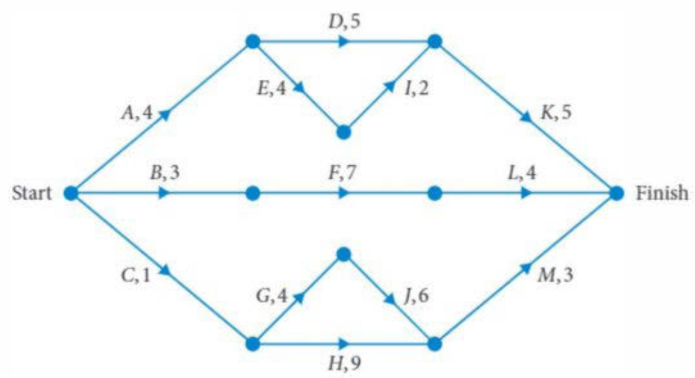

16 ©VCAA 2017N1NQ5 I Beginning with Activity C, the number of paths from start to finish is

A l B 2 C 3 D 4 E 5

17 ©VCAA 2017N1NQ6 What is the latest starting time for Activity I, in days, so that the project is completed in the shortest time possible?

A 3 B 4 C 5 D 6 E 7

18 ©VCAA 2016 2NQ3 J (6 marks) A new skateboard park is to be built in Beachton. This project involves 13 activities, *A* to *M*. The directed network below shows these activities and their completion times in days.




a 61% Determine the earliest start time for activity *M*. 1 mark

b 76% The minimum completion time for the skateboard park is 15 days.

Write down the critical path for this project. 1 mark

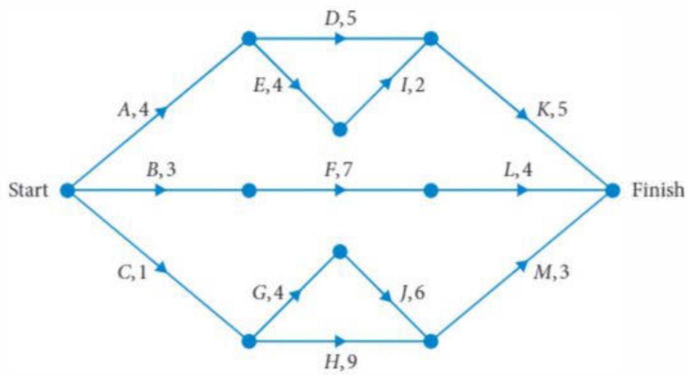
c 37% Which activity has a float time of two days? 1 mark d 21% The completion times for activities *E, F, G, I* and *J* can each be reduced by one day. The cost of reducing the completion time by one day for these activities is shown in the table below.

| Activity | Cost ($) |
|---|---|
| *E* | 3000 |
| *F* | 1000 |
| *G* | 5000 |
| *I* | 2000 |
| *J* | 4000 |

What is the minimum cost to complete the project in the shortest time possible? 1 mark

e The initial skateboard park project in Beachton will be repeated at Campville, but with the addition of one extra activity. The new activity, N, will take six days to complete and has a float time of one day. Activity *N* will finish at the same time as the project,

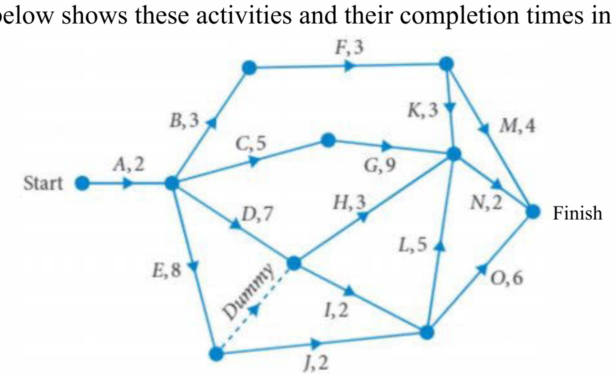
j 21% Copy the network below and add activity *N*. 1 mark




jj 27% What is the latest start time for activity *N?* 1 mark

19 ©VCAA 2017 2NQ4 (5 marks) The rides at the theme park are set up at the beginning of each holiday season. This project involves activities *A* to O.

The directed network below shows these activities and their completion times in days.




a 67% Write down the two immediate predecessors of activity *I.* 1 mark

b The minimum completion time for the project is 19 days.

i 44% There are two critical paths. One of the critical paths is *A-E-J-L-N.* Write down the other critical path. 1 mark

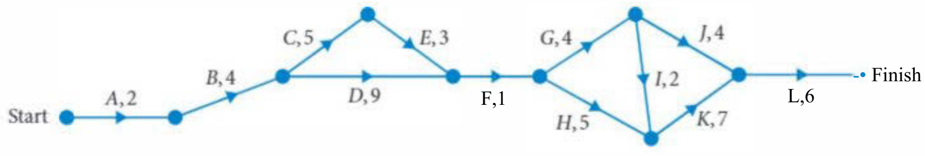
ii 28% Determine the float time, in days, for activity *F.* 1 mark

c The project could finish earlier if some activities were crashed. Six activities, B, D, G, I, *J* and L, can all be reduced by one day. The cost of this crashing is $1000 per activity.

i 35% What is the minimum number of days in which the project could now be completed? 1 mark

jj 15% What is the minimum cost of completing the project in this time? 1 mark

> 20 ©VCAA 2019 2NQ3 | (6 marks) Fencedale High School is planning to renovate its gymnasium. This project involves 12 activities, *A* to *L*. The directed network below shows these activities and their completion times, in weeks.




The minimum completion time for the project is 35 weeks.

a 60% How many activities are on the critical path? 1 mark

b 45% Determine the latest start time of activity *E*. 1 mark

c 45% Which activity has the longest float time? 1 mark

It is possible to reduce the completion time for activities C, D, G, *H* and *K* by employing more workers.

d 24% The completion time for each of these five activities can be reduced by a maximum of two weeks. What is the minimum time, in weeks, that the renovation project could take? 1 mark

e 8% The reduction in completion time for each of these five activities will incur an additional cost to the school. The table below shows the five activities that can have their completion times reduced and the associated weekly cost, in dollars.

| Activity | Weekly cost ($) |
|---|---|
| C | 3000 |
| *D* | 2000 |
| *G* | 2500 |
| *H* | 1000 |
| *K* | 4000 |

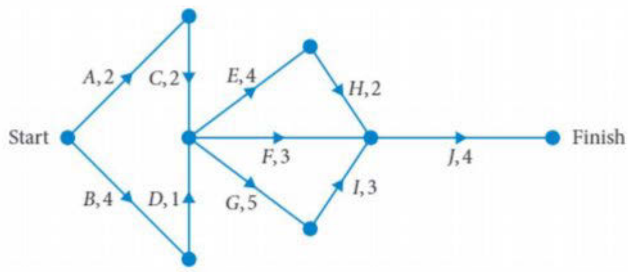
The completion time for each of these five activities can be reduced by a maximum of two weeks. Fencedale High School requires the overall completion time for the renovation project to be reduced by four weeks at minimum cost.

Copy and complete the table below, showing the reductions in individual activity completion times that would achieve this. 2 marks

| Activity | Reduction in completion time (0,1 or 2 weeks) |
|---|---|
| *C* | |
| *D* | |
| *G* | |
| *H* | |
| *K* | |

▶ 21 ©VCAA 2017N 2NQ1 (5 marks) Simon is building a new holiday home for his family.

The directed network below shows the 10 activities required for this project and their completion times, in weeks.




a Write down the two activities that are immediate predecessors of activity *G*. 1 mark

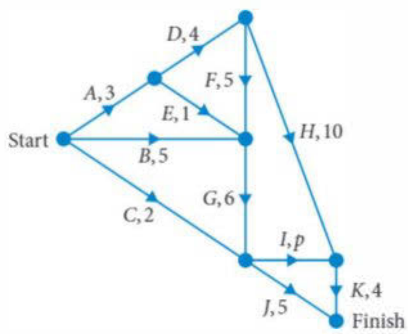
b For activity D, the earliest starting time and the latest starting time are the same.
What does this tell us about activity D? 1 mark

c Determine the minimum completion time, in weeks, for this project, 1 mark

d Determine the latest starting time, in weeks, for activity C. 1 mark

e Which activity could be delayed for the longest time without affecting the minimum completion time of the project? 1 mark

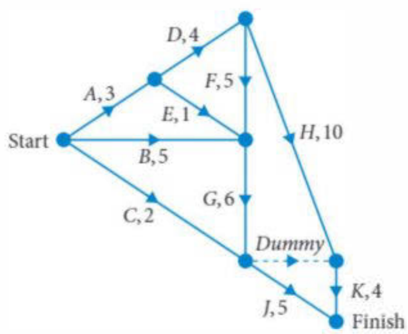
22 ©VCAA 201 SN 2NQ4 , (4 marks) A barn will be built on a property.
This building project will involve 11 activities, *A* to *K*. The directed network shows these activities and their duration in days. The duration of activity *I* is unknown at the start of the project. Let the duration of activity *I* be *p* days.




a Determine the earliest starting time, in days, for activity *I*. 1 mark

b Determine the value of p, in days, that would create more than one critical path, 1 mark

c If the value of *p* is six days, what will be the float time, in days, of activity H? 1 mark




d When a second barn is built later, activity *I* will not be needed. A *dummy* activity is required, as shown on the revised directed network.
Explain what this *dummy* activity indicates on the revised directed network. 1 mark

# 10.3 The assignment problem and bipartite graphs

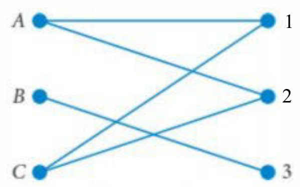
The assignment problem (or allocation problem) involves finding the best way of matching the elements in two sets. For example, a group of workers could be allocated to a set of tasks in order to optimise a stated objective such as minimising cost, distance or time.

## Bipartite graphs

A bipartite graph can be used to display the two sets of elements in an assignment problem. A bipartite graph has its vertices in two distinct sets and the edges join elements in the first set to elements in the second set.

Bipartite graphs can be undirected or directed.

In the bipartite graph below, there are three tasks {1, 2, 3} that can be performed by three people {A, B, C}. The edges show which people are qualified to perform which tasks.




The bipartite graph shows

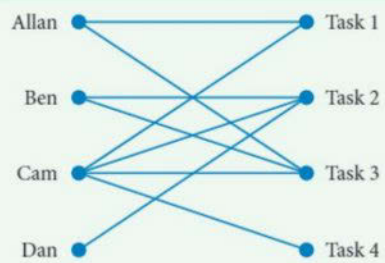
- *A* can perform tasks 1 and 2
- *B* can perform task 3
- C can perform tasks 1 and 2

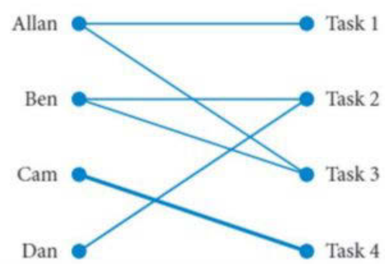
Video playlist The assignment problem and bipartite graphs

**WORKED EXAMPLE 9** Finding the allocation from a bipartite graph

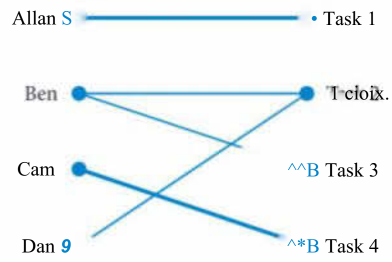
WB
p. 202

The bipartite graph shows the tasks that each of the four people is able to undertake. If each person must complete one task, find a valid allocation.

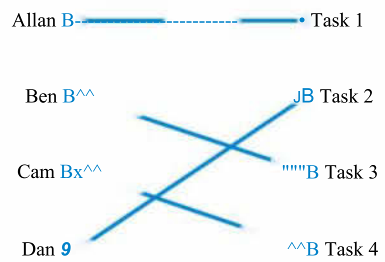



| Steps | Working |
|---|---|
| **1** Identify tasks that have the smallest number of links. | Cam is the only person that can do task 4. |
| **2** Cam is the only person who can do task 4. Allocate Cam to task 4 then eliminate links from Cam to tasks 1, 2 and 3. |  |

3 Task 1 can only be done by Allan so allocate Allan to task 1 then eliminate links from Allan to task 3.




4 Allocate Ben to task 3, which leaves Dan allocated task 2.




Task allocation:

Allan - Task 1 Ben - Task 3

Cam - Task 4 Dan - Task 2

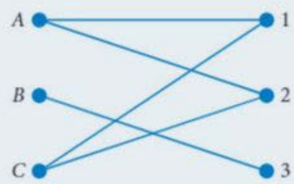
## Adjacency matrices and bipartite graphs

### Adjacency matrix representation for a bipartite graph

An **adjacency matrix** can be used to represent a bipartite graph.

In the matrix, 1 represents a connection and 0 indicates that there is no connection.

The bipartite graph is represented by the matrix

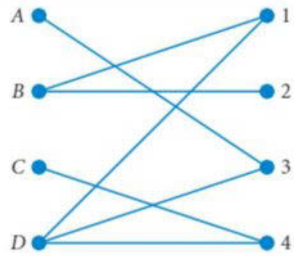
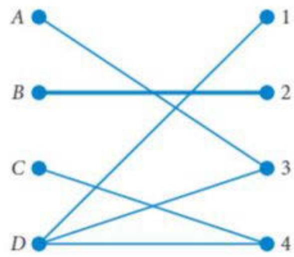
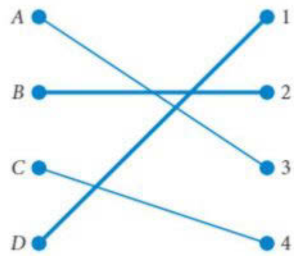



$$\begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \end{array}$$

**WORKED EXAMPLE 10** Drawing a bipartite graph from an adjacency matrix

Four workers $A$, $B$, $C$ and $D$ need to be allocated one task from tasks 1, 2, 3 and 4. The workers are only qualified to perform certain tasks and this is indicated in the following matrix. Draw a bipartite graph from the adjacency matrix and determine a valid allocation.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix} \end{array}$$

| Steps | Working |
|---|---|
| **1** Draw a connection for every 1 in the matrix. $A3$, $B1$, $B2$, $C4$, $D1$, $D3$ and $D4$ |  |
| **2** Start with the tasks with the smallest number of links. Allocate $B$ to task 2 and remove the other link from $B$ to 1. |  |
| **3** Task 1 can only be done by $D$ so allocate $D$ to task 1 and remove any other links from $D$ to tasks 3 and 4. |  The allocation is $A3$, $B2$, $C4$ and $D1$. |

10.3

WB

p. 203

## The allocation problem - Finding an optimum allocation

The aim of the allocation problem is to find a way of assigning the elements in the first set to the elements in the second set that meets an objective like minimising cost or time. The method used to find the optimum allocation is called the Hungarian algorithm and is completed in two stages. Stage one is a row and column reduction. In most cases, this is sufficient to find a solution.

Worksheet
The Hungarian algorithm

**The Hungarian algorithm (stage one) - Row and column reduction**

1. Choose the smallest number in each row and subtract it from every element in the same row.
2. For every column that does not have a zero value, choose the smallest number in the column and subtract it from every element in the same column.
3. Cover all the zeros with the smallest number of lines (horizontal or vertical). The process is complete if the number of lines is equal to the number of rows.

WB
p. 204

## WORKED EXAMPLE 11 Applying stage one of the Hungarian algorithm

Three workers *A*, *B*, and *C* need to be allocated one task from tasks 1, 2 and 3. The time in hours that each worker takes to complete the tasks is shown in the table.

If the tasks must be completed in the minimum time, find how the tasks are assigned.

| | 1 | 2 | 3 |
|---|---|---|---|
| *A* | 8 | 10 | 9 |
| *B* | 9 | 11 | 7 |
| *C* | 7 | 9 | 8 |

| Steps | Working |
|---|---|

**1** Identify the smallest number in each row and subtract it from the other numbers in the same row.

Row *A* subtract 8

Row *B* subtract 7

Row *C* subtract 7

| | 1 | 2 | 3 |
|---|---|---|---|
| *A* | 0 | 2 | 1 |
| *B* | 2 | 4 | 0 |
| *C* | 0 | 2 | 1 |

**2** For every column that does not have a zero value choose the smallest number in the column and subtract it from every element in the same column.

Column 1 has a zero

Column 2 subtract 2

Column 3 has a zero

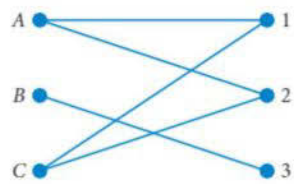
| | 1 | 2 | 3 |
|---|---|---|---|
| *A* | 0 | 0 | 1 |
| *B* | 2 | 2 | 0 |
| *C* | 0 | 0 | 1 |

**3** Cover all the zeros with the smallest number of lines (horizontal or vertical).

| | 1 | 2 | 3 |
|---|---|---|---|
| *A* | 0 | 0 | 1 |
| *B* | 2 | 2 | 0 |
| *C* | 0 | 0 | 1 |

There are three lines and three rows, so a solution has been found.

**4** The zeros indicate the allocations. Draw a bipartite graph connecting *A*1, *A*2, *B*3, *C*1 and *C*2.




*B* needs to complete task 3, but *A* and *C* can complete either task 1 or 2. So there are two possible allocations:

*A*2, *B*3, *C*1 and *A*1, *B*3, *C*2



9780170448406

## The Hungarian algorithm - Stage two

Stage two of the Hungarian algorithm is an extension to row and column reduction when this does not produce an optimum solution.

### The Hungarian algorithm - Stage two

The Hungarian algorithm

1. Perform a row and column reduction and cover all the zeros with the smallest number of horizontal or vertical lines possible. If the number of lines is equal to the number of rows, a solution has been found.
2. If the number of lines does not equal the number of rows, find the smallest uncovered number and add it to every covered number. Numbers that are covered twice have this number added twice.
3. Subtract the smallest number in the matrix or table from all of the numbers in the table.
4. Cover all the zeros with the smallest number of horizontal or vertical lines and if the number of lines is equal to the number of rows, the process is complete. If this is not the case, repeat steps 2, 3 and 4 until a solution is found.
5. Draw a bipartite graph and use it to determine the optimum allocation.

**WORKED EXAMPLE 12** Applying stages 1 and 2 of the Hungarian algorithm

WB p. 205

Four workers A, B, C and *D* all provide quotes for jobs 1, 2, 3 and 4. Their quotes are shown in the matrix, where the rows represent the workers A, B, C and *D* and the columns represent the jobs 1,2, 3 and 4.

a Determine the best allocation of workers that minimises the total quote,

b Find the cost of the four jobs.

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| *A* | 15 | 24 | 40 | 21 |
| *B* | 24 | 18 | 19 | 20 |
| *C* | 28 | 13 | 30 | 22 |
| *D* | 10 | 18 | 17 | 13 |

**Steps** | **Working**

a 1 Perform a row reduction

Row A subtract 15

RowB subtract 18

Row C subtract 13

RowD subtract 10

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| A | 0 | 9 | 25 | 6 |
| B | 6 | 0 | 1 | 2 |
| C | 15 | 0 | 17 | 9 |
| D | 0 | 8 | 7 | 3 |

2 Perform a column reduction

Column 3 subtract 1

Column 4 subtract 2

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| *A* | 0 | 9 | 24 | 4 |
| *B* | 6 | 0 | 0 | 0 |
| *C* | 15 | 0 | 16 | 7 |
| *D* | 0 | 8 | 6 1 | |

3 Cover the zeros

There are three lines and four rows.

As the number of lines does not equal the number of rows, stage 2 of the Hungarian algorithm must be used.

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| A | io | $ | 24 | 4 |
| *B* | 6 | 0 | 0 | 0 |
| *C* | 15 | o | 16 | 7 |
| *D* | iO | & | 6 | 1 |

4 The smallest uncovered number is 1.

Add this to every covered number but add 2 to the numbers covered twice.

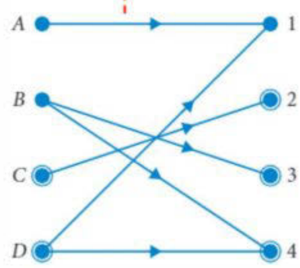
| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| *A* | 1 | 10 | 24 | 4 |
| *B* | 8 | 2 | 1 | 1 |
| *C* | 16 | 1 | 16 | 7 |
| *D* | 1 | 9 | 6 | 1 |

5 The smallest value is 1 so subtract this from every element in the matrix.

Now cover the zeros. There are now four lines and four rows so the allocation is complete.

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| *A* | 0 | 9 | 23 | 3 |
| *B* | 7 | 1 | 0 | 0 |
| *C* | 15 | 0 | 15 | 6 |
| *D* | 0 | 8 | 5 | 0 |

6 Draw the bipartite graph. The zeros in the matrix represent the allocations.




7 Determine the allocation.

Two jobs have a single worker connected.

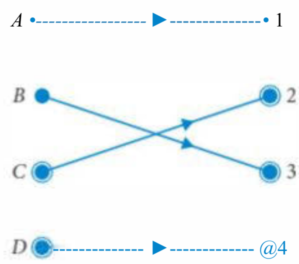
Worker C for job 2

Worker *B* for job 3

This leaves

Worker *D* for job 4

Worker *A* for job 1




Allocation

A1,B3,C2,D4

b Look at the original matrix to determine the cost:

Al = 15, B3 = 19, C2 = 13, D4 = 13

Cost = \$15+ \$19+ \$13+ \$13

= \$60

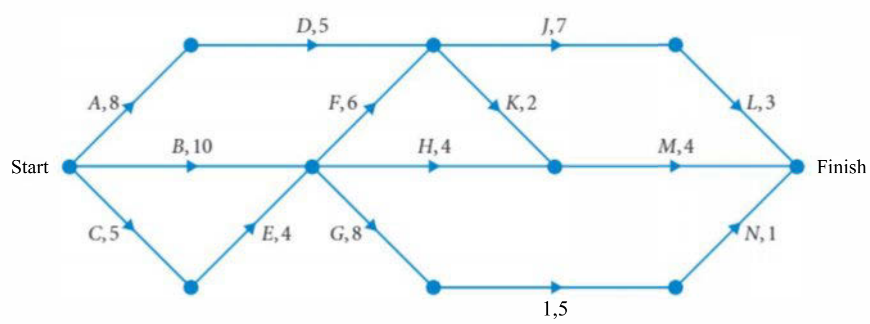
## EXERCISE 10.3 The assignment problem and bipartite graphs

ANSWERS p. 728

### Recap

*Use the following information to answer the next two questions.*

The activities and their completion times (in days) needed to complete a project are shown in the digraph below.




1 For the network shown, the length of the critical path is

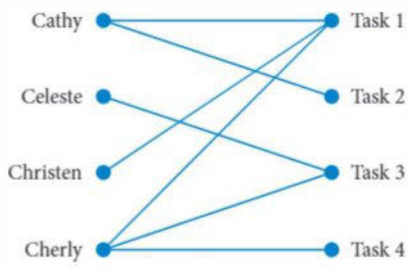
A 22 days. B 23 days. C 25 days. D 26 days. E 28 days.

2 The latest start time for activity *K* is

A 16 days. B 18 days. C 19 days. D 20 days. E 22 days.

## Mastery

30 WORKED EXAMPLE 9 I Determine the allocation of tasks from the following bipartite graph.




**4** S WORKED EXAMPLE io I Workers *Ay B, C* and *D* are each to be assigned one task from tasks 1, 2, 3 and 4. The matrix representing the workers that are qualified to perform the tasks is shown.

Draw a bipartite graph and hence, determine a valid allocation.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \end{array}$$

5 Q WORKED EXAMPLE ijH Perform row and column reductions on the matrix that represents the time, in hours, for workers A, B, C and *D* to complete tasks 1,2, 3 and 4, and hence, determine the optimum allocation and the minimum time to complete all four tasks.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{bmatrix} 25 & 14 & 7 & 15 \\ 6 & 10 & 8 & 9 \\ 13 & 5 & 9 & 5 \\ 8 & 5 & 10 & 11 \end{bmatrix} \end{array}$$

**6** H WORKED EXAMPLE 12 I A taxi company has three taxis A, *B* and C available and there are three customers 1,2 and 3 who require a taxi. The distance (km) that each taxi must travel to get to the customers is shown in the table.

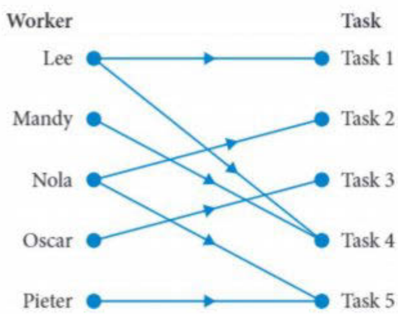
Use the Hungarian algorithm to find the optimum allocation that minimises the distance travelled.

| | 1 | 2 | 3 |
|---|---|---|---|
| A | 11 | 19 | 17 |
| *B* | 21 | 15 | 13 |
| *C* | 15 | 18 | 21 |

## Exam practice

80-100% 60-79% 0-59%

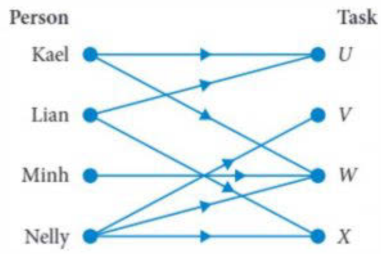
7 ©VCAA 20161 NQI 97% Lee, Mandy, Nola, Oscar and Pieter are each to be allocated one particular task at work. The bipartite graph below shows which task(s), 1-5, each person is able to complete.




Each person completes a different task. Task 4 must be completed by

A Lee. B Mandy. C Nola. D Oscar. E Pieter.

8 ©VCAA 20051NQ3 , (94% | The bipartite graph represents the tasks that four people are able to undertake.




The matrix representation for this task allocation is

**A**

$$\begin{matrix} \\ \text{Kael} \\ \text{Lian} \\ \text{Minh} \\ \text{Nelly} \end{matrix} \begin{matrix} U\ V\ W\ X \\ \begin{bmatrix} 1&1&0&0 \\ 1&0&0&1 \\ 0&0&1&0 \\ 0&1&1&1 \end{bmatrix} \end{matrix}$$

**B**

$$\begin{matrix} \\ \text{Kael} \\ \text{Lian} \\ \text{Minh} \\ \text{Nelly} \end{matrix} \begin{matrix} U\ V\ W\ X \\ \begin{bmatrix} 1&0&1&0 \\ 1&0&0&0 \\ 0&0&1&0 \\ 0&1&1&1 \end{bmatrix} \end{matrix}$$

**C**

$$\begin{matrix} \\ \text{Kael} \\ \text{Lian} \\ \text{Minh} \\ \text{Nelly} \end{matrix} \begin{matrix} U\ V\ W\ X \\ \begin{bmatrix} 1&0&1&0 \\ 1&0&0&0 \\ 0&0&1&0 \\ 0&1&1&1 \end{bmatrix} \end{matrix}$$

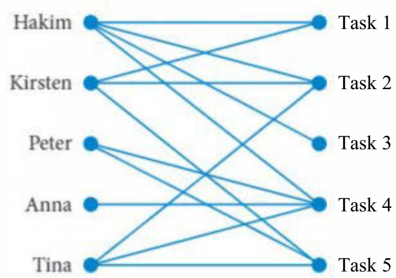
**D**

$$\begin{matrix} \\ \text{Kael} \\ \text{Lian} \\ \text{Minh} \\ \text{Nelly} \end{matrix} \begin{matrix} U\ V\ W\ X \\ \begin{bmatrix} 1&0&1&0 \\ 1&0&0&1 \\ 0&0&1&1 \\ 0&1&1&1 \end{bmatrix} \end{matrix}$$

**E**

$$\begin{matrix} \\ \text{Kael} \\ \text{Lian} \\ \text{Minh} \\ \text{Nelly} \end{matrix} \begin{matrix} U\ V\ W\ X \\ \begin{bmatrix} 1&0&1&0 \\ 1&1&0&1 \\ 0&0&1&0 \\ 0&1&1&1 \end{bmatrix} \end{matrix}$$

9 ©VCAA 2003 1NQ1 92% The bipartite graph below shows the tasks that each of five people are able to undertake.




If each person is to be allocated one task only, then a feasible task allocation is

**A**

| | |
|---|---|
| Hakim | 3 |
| Kirsten | 1 |
| Peter | 5 |
| Anna | 4 |
| Tina | 2 |

**B**

| | |
|---|---|
| Hakim | 3 |
| Kirsten | 2 |
| Peter | 5 |
| Anna | 4 |
| Tina | 1 |

**C**

| | |
|---|---|
| Hakim | 3 |
| Kirsten | 1 |
| Peter | 2 |
| Anna | 4 |
| Tina | 5 |

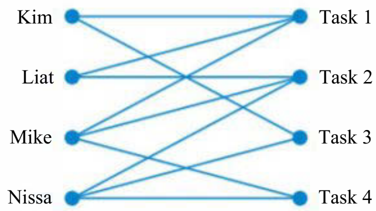
**D**

| | |
|---|---|
| Hakim | 3 |
| Kirsten | 5 |
| Peter | 1 |
| Anna | 4 |
| Tina | 2 |

**E**

| | |
|---|---|
| Hakim | 3 |
| Kirsten | 5 |
| Peter | 1 |
| Anna | 2 |
| Tina | 4 |

**10** ©VCAA 2012 1NQ3 86% The bipartite graph below shows the tasks that each of four people is able to undertake.




All tasks must be allocated and each person can be allocated one task only. A valid task allocation is

**A**

| | |
|---|---|
| Kim | task 1 |
| Liat | task 2 |
| Mike | task 3 |
| Nissa | task 4 |

**B**

| | |
|---|---|
| Kim | task 3 |
| Liat | task 1 |
| Mike | task 2 |
| Nissa | task 3 |

**C**

| | |
|---|---|
| Kim | task 3 |
| Liat | task 2 |
| Mike | task 1 |
| Nissa | task 4 |

**D**

| | |
|---|---|
| Kim | task 1 |
| Liat | task 3 |
| Mike | task 4 |
| Nissa | task 2 |

**E**

| | |
|---|---|
| Kim | task 2 |
| Liat | task 1 |
| Mike | task 4 |
| Nissa | task 3 |

11 ©VCAA 20131NQ4 I 73% Kate, Lexie, Mei and Nasim enter a competition as a team. In this competition, the team must complete four tasks, W, X, *Y* and Z, as quickly as possible.

The table shows the time, in minutes, that each person would take to complete each of the four tasks.

| | Kate | Lexie | Mei | Nasim |
|---|---|---|---|---|
| *W* | 6 | 3 | 4 | 6 |
| *X* | 4 | 3 | 5 | 5 |
| r | 5 | 7 | 9 | 6 |
| Z | 3 | 2 | 3 | 2 |

If each team member is allocated one task only, the minimum time in which this team would complete the four tasks is

A 10 minutes B 12 minutes C 13 minutes
D 14 minutes E 15 minutes

12 ©VCAA 2009 1NQ7 61% Four workers, Anna, Bill, Caitlin and David, are each to be assigned a different task.

The table below gives the time, in minutes, that each worker takes to complete each of the four tasks.

| | task 1 | task 2 | task 3 | task 4 |
|---|---|---|---|---|
| Anna | 7 | 5 | 15 | 9 |
| Bill | 8 | 5 | 18 | 10 |
| Caitlin | 4 | 6 | 22 | 4 |
| David | 7 | 11 | 16 | 10 |

The tasks are allocated so as to minimise the total time taken to complete the four tasks. This total time, in minutes, is

A 21 B 28 C 31 D 34 E 38

▶ 13 ©VCAA 20101NQ9 60% The table below shows the time (in minutes) that each of four people, Aiden, Bing, Callum and Dee, would take to complete each of the tasks *U*, V, W and *X*.

| | Task | | | |
|---|---|---|---|---|
| | *U* | *V* | *W* | *X* |
| Aiden | 3 | 2 | 9 | 9 |
| Bing | 5 | 6 | 12 | 12 |
| Callum | 9 | 6 | 12 | 14 |
| Dee | 8 | 3 | 8 | 12 |

If each person is allocated one task only, the minimum total time for this group of people to complete all four tasks is

A 22 minutes. B 28 minutes. C 29 minutes
D 30 minutes. E 32 minutes

14 ©VCAA 2004 1NQ7 43% Five people are to be each allocated one of five tasks (*A*, *B*, *C*, *D*, *E*). The table shows the time, in hours, that each person takes to complete the tasks.

| | Task | | | | |
|---|---|---|---|---|---|
| | *A* | *B* | C | *D* | *E* |
| Francis | 12 | 15 | 99 | 10 | 14 |
| David | 10 | 9 | 10 | 7 | 12 |
| Herman | 99 | 10 | 11 | 6 | 12 |
| Indira | 8 | 8 | 12 | 9 | 99 |
| Natalie | 8 | 99 | 9 | 8 | 11 |

The tasks must be completed in the least possible total amount of time.

If no person can help another, Francis should be allocated task

A *A* *B B* C C D *D* *EE*

15 ©VCAA 0161NQ8 I 30% Five children, Alan, Brianna, Chamath, Deidre and Ewen, are each to be assigned a different job by their teacher. The table below shows the time, in minutes, that each child would take to complete each of the five jobs.

| | Alan | Brianna | Chamath | Deidre | Ewen |
|---|---|---|---|---|---|
| Job 1 | 5 | 8 | 5 | 8 | 7 |
| Job 2 | 5 | 7 | 6 | 7 | 4 |
| Job 3 | 9 | 5 | 7 | 5 | 9 |
| Job 4 | 7 | 7 | 9 | 8 | 5 |
| Job 5 | 4 | 4 | 4 | 4 | 3 |

The teacher wants to allocate the jobs so as to minimise the total time taken to complete the five jobs. In doing so, she finds that two allocations are possible.

If each child starts their allocated job at the same time, then the first child to finish could be either

A Alan or Brianna.
B Brianna or Deidre.
C Chamath or Deidre.
D Chamath or Ewen.
E Deidre or Ewen.

▶

10.3

16 ©VCAA 20181 NQS 24% Annie, Buddhi, Chuck and Dorothy work in a factory.

Today each worker will complete one of four tasks, 1,2, 3 and 4.

The usual completion times for Annie, Chuck and Dorothy are shown in the table below.

| | Task 1 | Task 2 | Task 3 | Task 4 |
|---|---|---|---|---|
| Annie | 7 | 3 | 8 | 2 |
| Buddhi | $k$ | $k$ | 3 | $k$ |
| Chuck | 5 | 6 | 9 | 2 |
| Dorothy | 4 | 8 | 5 | 3 |

Buddhi takes 3 minutes for Task 3.

He takes $k$ minutes for each other task.

Today the factory supervisor allocates the tasks as follows:

- Task 1 to Dorothy
- Task 2 to Annie
- Task 3 to Buddhi
- Task 4 to Chuck

This allocation will achieve the minimum total completion time if the value of $k$ is at least

AO BI C 2 D3 E4

17 ©VCAA 2017N 1NQ8 Abbey, Barb, Cathal and Dinh are four workers at a business. Each worker will perform one duty.

The time for each worker to complete duties 1, 2, 3 and 4, in minutes, is shown in the table below.

| | Duty 1 | Duty 2 | Duty 3 | Duty 4 |
|---|---|---|---|---|
| Abbey | 6 | 5 | 6 | 8 |
| Barb | 9 | 12 | 10 | 6 |
| Cathal | 8 | 7 | 4 | 8 |
| Dinh | 5 | 3 | 6 | 4 |

The minimum total time for all duties is 19 minutes, with Dinh performing Duty 2.

Before the duties are performed, it is found that Dinh will require 7 minutes for Duty 2 rather than 3 minutes. If the duties are allocated again, the minimum total time for all duties will

A remain the same. B increase by 1 minute. C increase by 2 minutes.

D increase by 3 minutes. E increase by 4 minutes.

18 ©VCAA 2018N1NQ2 I A puzzle has four parts. Four friends, Audrey, Bruce, Christie and Darren, are solving the puzzle together.

The table shows the time it takes each person to complete each part of the puzzle, in minutes.

| | Part 1 | Part 2 | Part 3 | Part 4 |
|---|---|---|---|---|
| Audrey | 5 | 5 | 5 | 5 |
| Bruce | 3 | 4 | 5 | 6 |
| Christie | 6 | 5 | 4 | 3 |
| Darren | 5 | 5 | 5 | 5 |

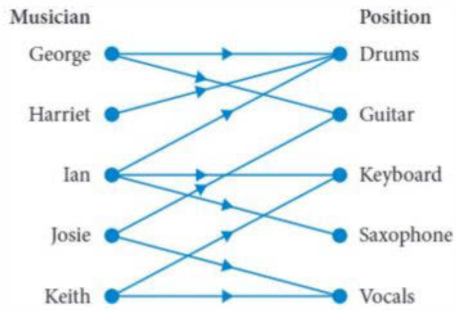
The parts of the puzzle must be solved one after the other, with each friend completing one part. The minimum time taken to complete the entire puzzle, in minutes, is

A 14 B 16 C 18 D 20 E 22

19 ©VCAA 2006 2NQ1 90% (3 marks) George, Harriet, Ian, Josie and Keith are a group of five musicians.

They are forming a band where each musician will fill one position only.

The bipartite graph illustrates the positions that each is able to fill.




a Which musician must play the guitar? 1 mark

b Copy and complete the table showing the positions that the following musicians must fill in the band. 2 marks

| Person | Position |
|---|---|
| Harriet | |
| Ian | |
| Keith | |

20 ©VCAA I 2012 2NQ3 66% (5 marks) Four tasks, W, X, $Y$ and Z, must be completed. Four workers, Julia, Ken, Lana and Max, will each do one task.

Table 1 shows the time, in minutes, that each person would take to complete each of the four tasks.

The tasks will be allocated so that the total time of completing the four tasks is a minimum. The Hungarian algorithm will be used to find the optimal allocation of tasks.

Table 1

| | | Worker | | | |
|---|---|---|---|---|---|
| | | Julia | Ken | Lana | Max |
| Task | $W$ | 26 | 21 | 22 | 25 |
| | $X$ | 31 | 26 | 21 | 38 |
| | $Y$ | 29 | 26 | 20 | 27 |
| | $Z$ | 38 | 26 | 26 | 35 |

Step 1 of the Hungarian algorithm is to subtract the minimum entry in each row from each element in the row.

a Complete step 1 for task X by writing down the number missing from the shaded cell in Table 2. 1 mark

Table 2

| | | Worker | | | |
|---|---|---|---|---|---|
| | | Julia | Ken | Lana | Max |
| Task | $W$ | 5 | 0 | 1 | 4 |
| | $X$ | 10 | 5 | 0 | |
| | $Y$ | 9 | 6 | 0 | 7 |
| | $Z$ | 12 | 0 | 0 | 9 |

The second step of the Hungarian algorithm ensures that all columns have at least one zero. The numbers that result from this step are shown in Table 3.

b Following the Hungarian algorithm, the smallest number of lines that can be drawn to cover the zeros is shown dashed in Table 3. These dashed lines indicate that an optimal allocation cannot be made yet. Give a reason why. 1 mark

Table 3

| | | Worker | | | |
|---|---|---|---|---|---|
| | | Julia | Ken | Lana | Max |
| Task | $W$ | 0 | 0 | 1 | 0 |
| | $X$ | 5 | 5 | 0 | 13 |
| | $Y$ | 4 | 6 | 0 | 3 |
| | $Z$ | 7 | 0 | 0 | 5 |

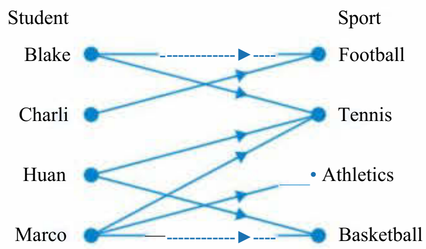
c Complete the steps of the Hungarian algorithm to produce a table from which the optimal allocation of tasks can be made. 1 mark

d Write the name of the task that each person should do for the optimal allocation of tasks. 2 marks


9780170448406

21 ©VCAA | 2019 2NQ2 , (3 marks) Fencedale High School offers students a choice of four sports, football, tennis, athletics and basketball.

The bipartite graph illustrates the sports that each student can play.




Each student will be allocated to only one sport.

a | 93% Copy and complete the table below by allocating the appropriate sport to each student. 1 mark

| Student | Sport |
|---|---|
| Blake | |
| Charli | |
| Huan | |
| Marco | |

b | 63% The school medley relay team consists of four students, Anita, Imani, Jordan and Lola. The medley relay race is a combination of four different sprinting distances: 100 m, 200 m, 300 m and 400 m, run in that order.

The following table shows the best time, in seconds, for each student for each sprinting distance.

| | Best time for each sprinting distance (seconds) | | | |
|---|---|---|---|---|
| Student | 100m | 200 m | 300 m | 400 m |
| Anita | 13.3 | 29.6 | 61.8 | 87.1 |
| Imani | 14.5 | 29.6 | 63.5 | 88.9 |
| Jordan | 13.3 | 29.3 | 63.6 | 89.1 |
| Lola | 15.2 | 29.2 | 61.6 | 87.9 |

The school will allocate each student to one sprinting distance in order to minimise the total time taken to complete the race.

To which distance should each student be allocated? 2 marks

22 ©VCAA | 2008 2NQ3 I 60% (4 marks) Each child is to be driven by his or her parents to one of four different concerts. The following table shows the distance that each car would have to travel, in kilometres, to each of the four concerts.

| | Concert 1 | Concert 2 | Concert 3 | Concert 4 |
|---|---|---|---|---|
| James | 10 | 16 | 18 | 20 |
| Dante | 9 | 14 | 19 | 15 |
| Tahlia | 15 | 13 | 20 | 18 |
| Chanel | 10 | 15 | 21 | 16 |

The concerts will be allocated so as to minimise the total distance that must be travelled to take the children to the concerts. The Hungarian algorithm is to be used to find this minimum value.

a Step 1 of the Hungarian algorithm is to subtract the minimum entry in each row from each element in the row. Copy the table below and complete step 1 for Tahlia by writing the missing values.

| | Concert 1 | Concert 2 | Concert 3 | Concert 4 |
|---|---|---|---|---|
| James | 0 | 6 | 8 | 10 |
| Dante | 0 | 5 | 10 | 6 |
| Tahlia | | | | |
| Chanel | 0 | 5 | 11 | 6 |

1 mark

After further steps of the Hungarian algorithm have been applied, the table is as follows.

| | Concert 1 | Concert 2 | Concert 3 | Concert 4 |
|---|---|---|---|---|
| James | 0 | 5 | 0 | 4 |
| Dante | 0 | 4 | 2 | 0 |
| Tahlia | 3 | 0 | 0 | 0 |
| Chanel | 0 | 4 | 3 | 0 |

It is now possible to allocate each child to a concert.

b Explain why this table shows that Tahlia should attend Concert 2. 1 mark

c Determine the concerts that could be attended by James, Dante and Chanel to minimise the total distance travelled. Copy the table and write in your answers. 1 mark

| | Concert |
|---|---|
| James | |
| Dante | |
| Tahlia | 2 |
| Chanel | |

d Determine the minimum total distance, in kilometres, travelled by the four cars. 1 mark

23 ©VCAA 2017 2NQ2 (2 marks) Bai joins his friends Agatha, Colin and Diane when he arrives for the holiday in Seatown. Each person will plan one tour that the group will take.

Table 1 shows the time, in minutes, it would take each person to plan each of the four tours.

**Table 1**

| | Agatha | Bai | Colin | Diane |
|---|---|---|---|---|
| Tour 1 | 3 | 0 | 3 | 3 |
| Tour 2 | 6 | 4 | 0 | 0 |
| Tour 3 | 0 | 9 | 2 | 0 |
| Tour 4 | 0 | 0 | 1 | 1 |

The aim is to minimise the total time it takes to plan the four tours. Agatha applies the Hungarian algorithm to Table 1 to produce Table 2. Table 2 shows the final result of all her steps of the Hungarian algorithm.

**Table 2**

| | Agatha | Bai | Colin | Diane |
|---|---|---|---|---|
| Tour 1 | 3 | 0 | 3 | 3 |
| Tour 2 | 6 | 4 | 0 | 0 |
| Tour 3 | 0 | 9 | 2 | 0 |
| Tour 4 | 0 | 0 | 1 | 1 |

a 37% In Table 2 there is a zero in the column for Colin.

When all values in the table are considered, what conclusion about minimum total planning time can be made from this zero. 1 mark

b 58% Determine the minimum total planning time, in minutes, for all four tours. 1 mark



9780170448406

# @ Network flow problems

10.4

The applications of flow networks range from transporting water through a network of water pipes to moving people or products by road or a rail network.

## Flow capacity and maximum flow

The first vertex of the network is called the source and the final vertex of the network is called the sink.

### Flow capacity and maximum flow

The inflow capacity is the total flow capacity entering a vertex and the outflow capacity is the total flow capacity leaving the vertex.

The maximum flow through a vertex is the smaller of the total inflow capacity of the vertex and the total outflow capacity of the vertex.

Video playlist Network flow problems

Worksheet Network flow capacity

WB

p. 207

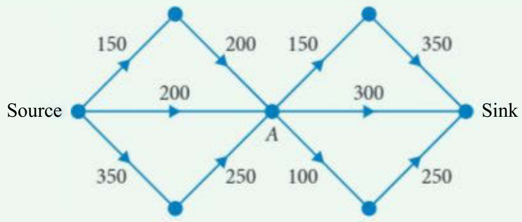
**WORKED EXAMPLE 13** Finding the inflow, outflow and maximum capacity of a vertex

The directed graph shows the flow capacities in litres per minute. Determine

a the inflow capacity of vertex $A$.

b the outflow capacity of vertex $A$.

c the maximum flow out of vertex $A$.

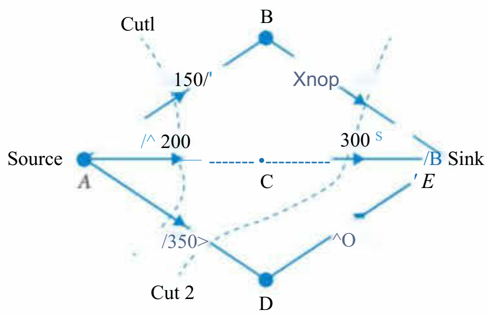



| Steps | Working |
|---|---|
| a The total inflow capacity of vertex $A$ can be found by adding the values on the edges entering vertex $A$. | total inflow capacity of vertex $A$ = 200 + 200 + 250 = 650 litres per minute |
| b The total outflow capacity of vertex $A$ can be found by adding the values on the edges leaving vertex $A$. | total outflow capacity of vertex $A$ = 150 + 300+ 100 = 550 litres per minute |
| c The maximum flow out of vertex $A$ is the smaller of inflow capacity and outflow capacity. | The flow out of vertex $A$ is 550 litres per minute, |

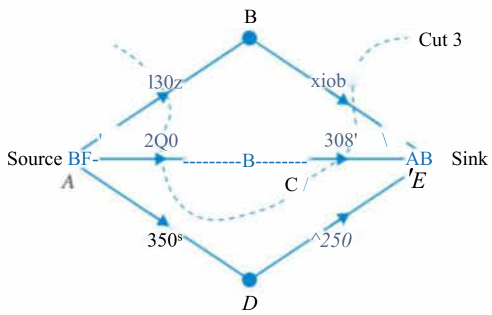
### The capacity of a cut

A cut through a network must stop all flow from the start *(source)* to the end *(sink)*.

One method of finding the maximum flow through a network is to use the minimum capacity of the cuts. In the network shown, Cut 1 and Cut 2 are both valid. The cuts sever the pipes in the network in such a way that there is no flow path from the source *(A)* to the sink (£). Cut 3 is not a valid cut because it does not stop all the flow between the source and the sink. It is still possible for flow to occur from $A$ to $D$ to $E$.




Valid cuts




Invalid cuts

## The capacity of a cut

The capacity of a cut can be determined by adding all the flow capacities that pass across the cut in the direction from source to sink.

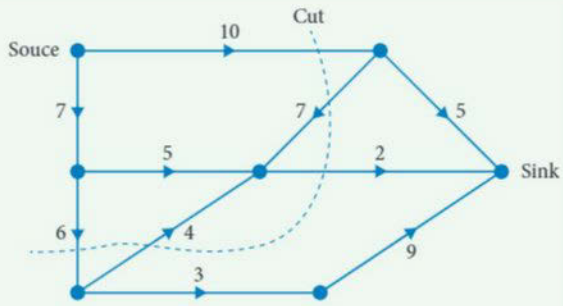
> **Exam hack**
>
> The source is the first vertex (on the left) and the sink is the last vertex (on the right). To determine the capacity of a cut only, add flow capacities that pass across the cut in the direction left (source) to right (sink).

WB

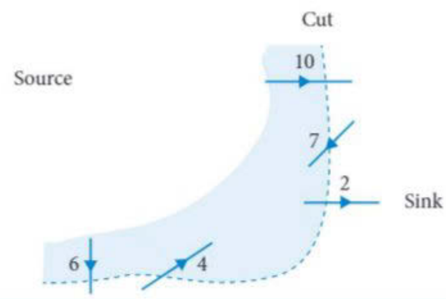
p. 208

**WORKED EXAMPLE 14** Finding the capacity of a cut

In the network, the numbers on the edges show the maximum possible flow between the vertices. The direction of the arrow indicates the direction of the flow. A cut separating the sink from the source is also shown. Determine the capacity of the cut.




| Steps | Working |
|---|---|
| Add the flows where the arrow crosses the cut in the direction from the source side of the cut to the sink side of the cut. | flow of cut = 6 + 2 + 10<br>= 18 |
| Flows 4 and 7 are not included as they flow in the direction sink to source across the cut. | In the diagram below, the source side of the cut has been shaded. The flows 7 and 4 are not included in the cut capacity as they flow toward the shaded source side instead of toward the unshaded sink side of the cut. |




## Maximum flow - minimum cut

The maximum flow from source to sink can be determined by finding the cut that produces the minimum value.

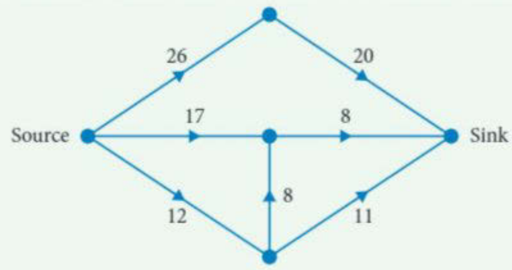
> **Determining the maximum flow for a network**
>
> To determine the maximum flow for a network:
>
> 1 identify cuts through the network
>
> 2 find the capacity of each cut
>
> 3 maximum flow = capacity of the minimum cut.

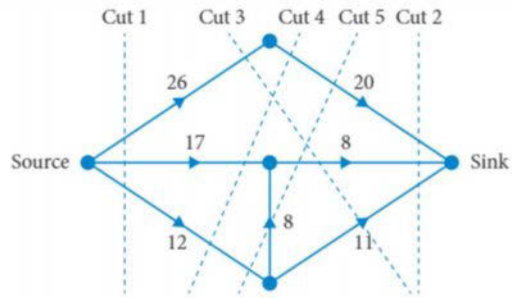


9780170448406

**WORKED EXAMPLE 15** Finding the maximum flow of a network using cut capacities

Find the maximum flow for the network.




| Steps | Working |
|---|---|
| **1** Identify cuts that stop the flow from source to sink. |  |
| **2** Find the capacity of each cut.<br>Note: The vertical edge with a flow capacity of 8 is not included in the cut 5 capacity because its flow direction is from sink to source. | Cut 1 capacity = 26 + 17 + 12 = 55<br>Cut 2 capacity = 20 + 8 + 11 = 39<br>Cut 3 capacity = 26 + 8 + 11 = 45<br>Cut 4 capacity = 12 + 17 + 20 = 49<br>Cut 5 capacity = 12 + 8 + 20 = 40 |
| **3** Maximum flow = minimum cut | The maximum flow from source to sink is 39. |

## EXERCISE 10.4 Network flow problems

ANSWERS p. 728

10.4

WB p. 209

### Recap

*Use the following information to answer the next two questions.*

Four workers, Alli, Brianna, Charles and Don, are each assigned different tasks. The table gives the time, in hours, that each worker takes to complete each of the four tasks. The tasks are allocated to minimise the total time taken to complete the four tasks.

| | Task 1 | Task 2 | Task 3 | Task 4 |
|---|---|---|---|---|
| Alli | 5 | 7 | 6 | 7 |
| Brianna | 6 | 4 | 3 | 6 |
| Charles | 7 | 5 | 4 | 2 |
| Don | 5 | 3 | 8 | 5 |

1 The tasks allocated to Ali and Don are

A Ali - Task 3, Don - Task 2

B Ali - Task 1, Don - Task 3

C Ali - Task 1, Don - Task 2

D Ali - Task 2, Don - Task 4

E Ali - Task 4, Don - Task 3

2 The total minimum time to complete the four tasks is

A 9 hours.

B 10 hours.

C 11 hours.

D 12 hours.

E 13 hours.

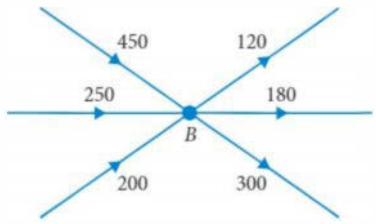
## Mastery

30 WORKED EXAMPLE 13 The traffic capacity, in vehicles per hour, of roads in and out of an intersection, labelled vertex B, is shown. Determine

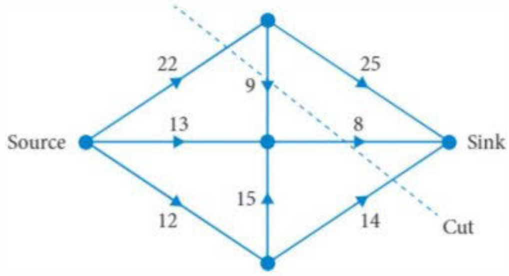
a the inflow capacity of vertex $B$

b the outflow capacity of vertex $B$

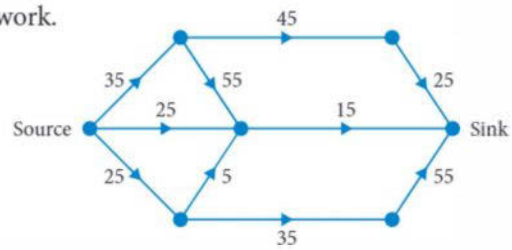
c the maximum flow out of vertex $B$.




4 H WORKED EXAMPLE 14 Determine the capacity of the cut.




**5** WORKED EXAMPLE 15 Find the maximum flow for the network.

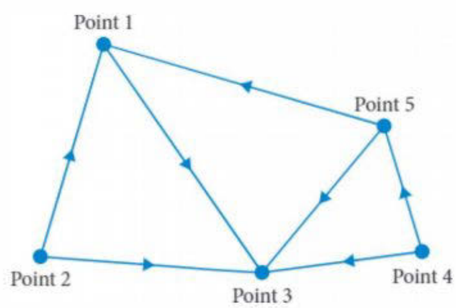



## Exam practice

80-100% 60-79% 0-59%

6 ©VCAA 2008 1NQ1 95% Steel water pipes connect five points underground.

The directed graph shows the directions of the flow of water through these pipes between these points.




The directed graph shows that water can flow from

A point 1 to point 2.

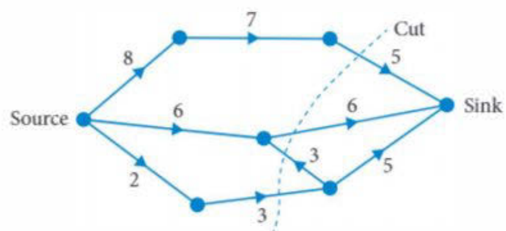
B point 1 to point 4.

C point 4 to point 1.

D point 4 to point 2.

E point 5 to point 2.

7 © VCAA 2019 1NQ3 85% The flow of water through a series of pipes is shown in the network. The numbers on the edges show the maximum flow through each pipe in litres per minute.




The capacity of Cut Q, in litres per minute, is

A 11 B 13 C 14

D 16 E 17



9780170448406

8 ©VCAA 2006 1NQ6 63% In the directed graph below the weight of each edge is non-zero. The capacity of the cut shown is

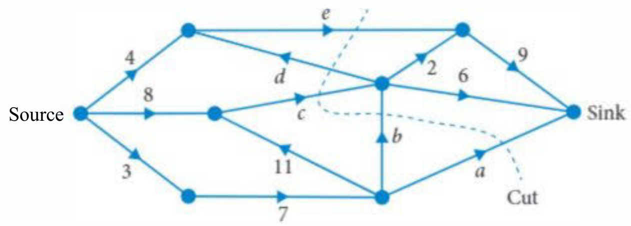
A $a+b+c+d+e$

B $a+c+d+e$

C $a+b+c+e$

D $a+b+c-d+e$

E $a-b+c-d+e$

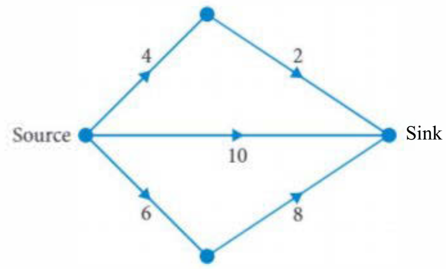



9 ©VCAA 2016 1NQ2 54% The directed graph shows the flow of water, in litres per minute, in a system of pipes connecting the source to the sink.

The maximum flow, in litres per minute, from the source to the sink is

A 10 B 14 C 18

D 20 E 22

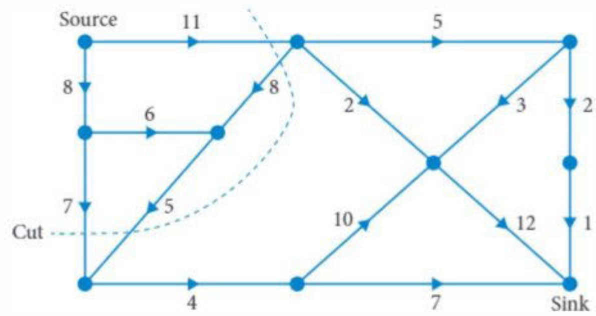



10 ©VCAA 2010 1NQ6 50% In the network, the values on the edges give the maximum flow possible between each pair of vertices. The arrows show the direction of flow. A cut that separates the source from the sink in the network is also shown.

The capacity of this cut is

A 14 B 18 C 23

D 31 E 40




11 ©VCAA 2017 1NQ8 44% The flow of oil through a series of pipelines, in litres per minute, is shown in the network.

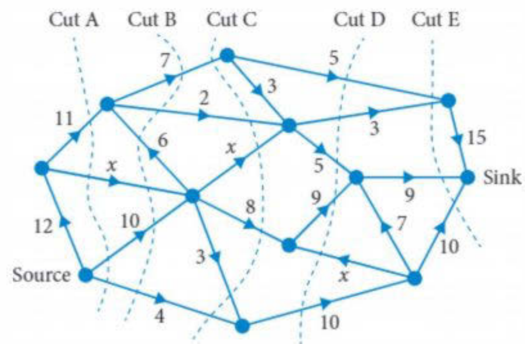
The weightings of three of the edges are labelled $x$. Five cuts labelled $A$-$E$ are shown on the network.

The maximum flow of oil from the source to the sink, in litres per minute, is given by the capacity of

A Cut A if x = 1 B Cut B if x = 2

C Cut C if x = 2 D Cut D if x = 3

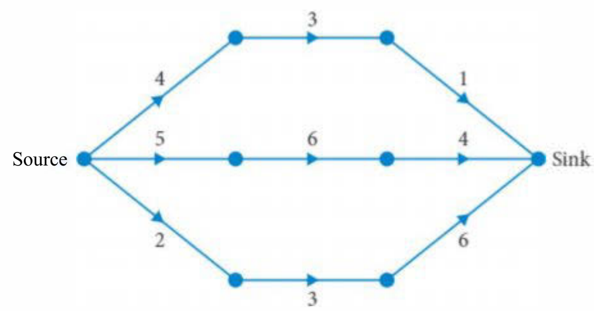
E Cut E if x = 3



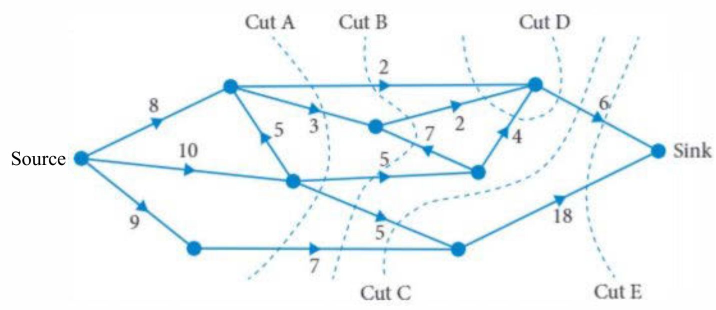

12 ©VCAA 2009 1NQ3 44% The maximum flow from source to sink through the network shown is

A 6 B 7 C 8

D 11 E 16



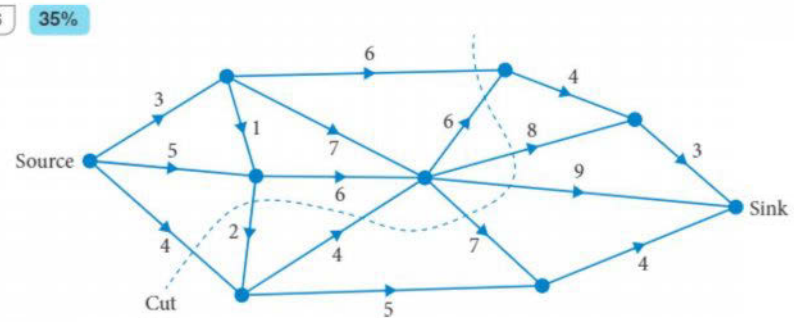
▶ 13 ©VCAA 20051NQ6 39% On the directed graph, the values on the edges give the maximum flow between nodes in the direction of the arrows. Five cuts have been made on the diagram.




Which cut allows you to find the maximum flow from point *X* to point Y?

A cut *A* B cut *B* C cut C D cut *D* E cut *E*

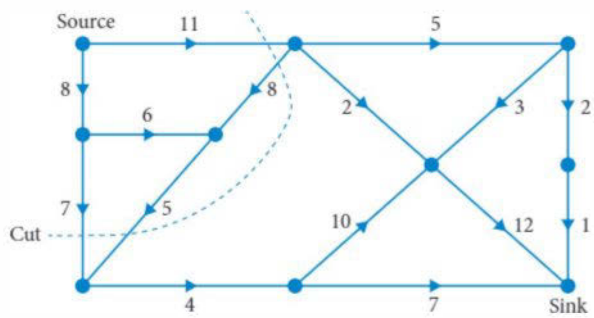
14 ©VCAA 2008 1NQ6 35%




For the graph, the capacity of the cut shown is

A 33 B 36 C 40 D 42 E 46

15 ©VCAA 20101NQ7 24% In the network below, the values on the edges give the maximum flow possible between each pair of vertices. The arrows show the direction of flow. A cut that separates the source from the sink in the network is also shown.



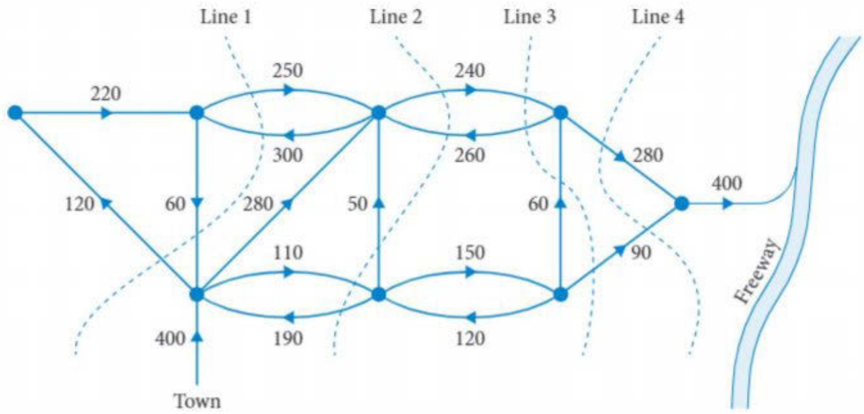

The maximum flow between source and sink through the network is

A 7 B 10 C 11 D 12 E 20

▶ 16 ©VCAA 20121NQ7 I 23% Vehicles from a town can drive onto a freeway along a network of one-way and two-way roads, as shown in the network diagram.

The numbers indicate the maximum number of vehicles per hour that can travel along each road in this network. The arrows represent the permitted direction of travel.

One of the four dotted lines shown on the diagram is the minimum cut for this network.

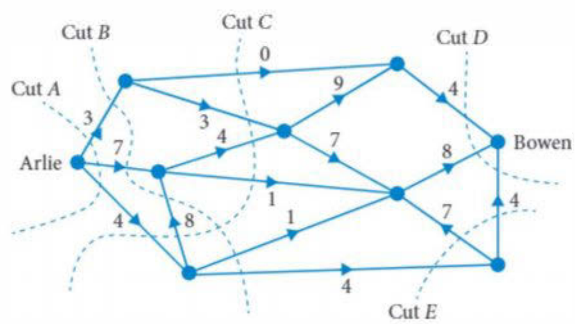



The maximum number of vehicles per hour that can travel through this network from the town onto the freeway is

A 310 B 330 C 350 D 370 E 390

17 ©VCAA 2003 2NQ1 I 51% (5 marks) A train journey consists of a connected sequence of stages formed by edges on the following directed network from Arlie to Bowen.

The number of available seats for each stage is indicated beside the corresponding edge, as shown on the diagram below.




The five cuts, A, $B, C_y D$ and E, shown on the network, are attempts to find the maximum number of available seats that can be booked for a journey from Arlie to Bowen

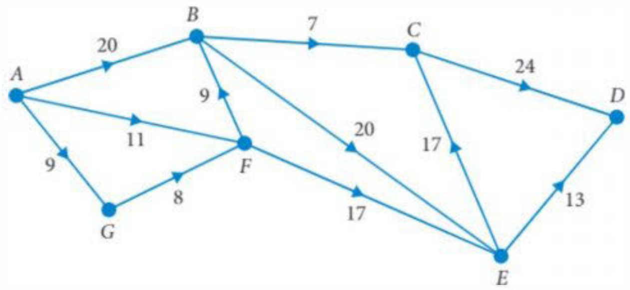
a What is the capacity of Cut A, Cut *B* and Cut C? 3 marks

b Cut *E* is not a valid cut when trying to find the minimum cut between Arlie and Bowen. Why? 1 mark

c Determine the maximum number of available seats for a train journey from Arlie to Bowen. 1 mark

▶ 18 ©VCAA J 2013 2NQ3 41% (4 marks) The rangers at a wildlife park restrict access to the walking tracks through areas where the animals breed.

The edges on the directed network diagram represent one-way tracks through the breeding areas. The direction of travel on each track is shown by an arrow. The numbers on the edges indicate the maximum number of people who are permitted to walk along each track each day.




a Starting at *A*, how many people, in total, are permitted to walk to *D* each day? 1 mark

One day, all the available walking tracks will be used by students on a school excursion. The students will start at *A* and walk in four separate groups to *D*.

Students must remain in the same groups throughout the walk.

b i Group 1 will have 17 students. This is the maximum group size that can walk together from *A* to *D*. Write down the path that group 1 will take. 1 mark

ii Groups 2, 3 and 4 will each take different paths from *A* to *D*.

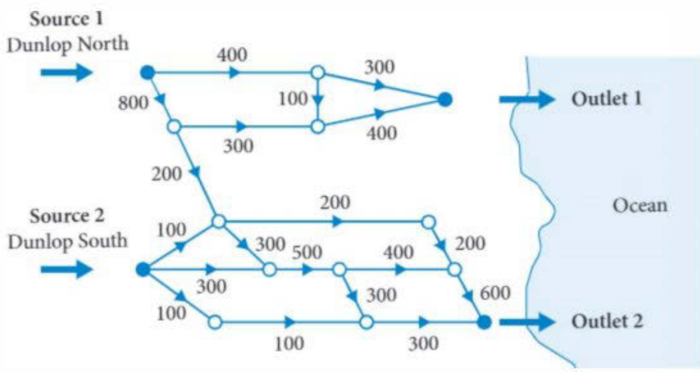
Copy the table below and complete the six missing entries. 2 marks

| Group | Maximum group size | Path taken from *A* to *D* |
|---|---|---|
| 1 | 17 | answered in part **b i** |
| 2 | | |
| 3 | | |
| 4 | | |

19 © VCAA 2011 2NQ4 43% (4 marks) Stormwater enters a network of pipes at either Dunlop North (Source 1) or Dunlop South (Source 2) and flows into the ocean at either Outlet 1 or Outlet 2.

On the network diagram below, the pipes are represented by straight lines with arrows that indicate the direction of the flow of water. Water cannot flow through a pipe in the opposite direction.

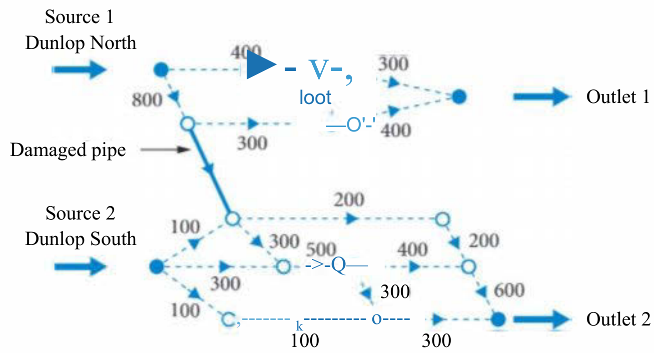
The numbers next to the arrows represent the maximum rate, in kilolitres per minute, at which stormwater can flow through each pipe.




a Copy and complete the following sentence for this network of pipes by writing either the number 1 or 2 in each box.

Stormwater from Source [ ] cannot reach Outlet [ ]. 1 mark b Determine the maximum rate, in kilolitres per minute, that water can flow from these pipes into the ocean at Outlet 1 and Outlet 2. 2 marks

A length of pipe, shown in bold on the network diagram below, has been damaged and will be replaced with a larger pipe

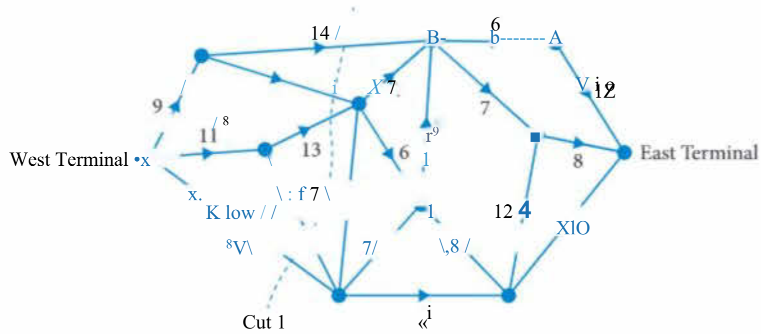



c The new pipe must enable the greatest possible rate of flow of stormwater into the ocean from Outlet 2. What minimum rate of flow through the pipe, in kilolitres per minute, will achieve this? 1 mark

**20** ©VCAA 2007 2NQ3 33% (3 marks) As an attraction for young children, a miniature railway runs throughout a new housing estate.

The trains travel through stations that are represented by nodes on the directed network diagram.

The number of seats available for children, between each pair of stations, is indicated beside the corresponding edge.




Cut 1, through the network, is shown in the diagram above.

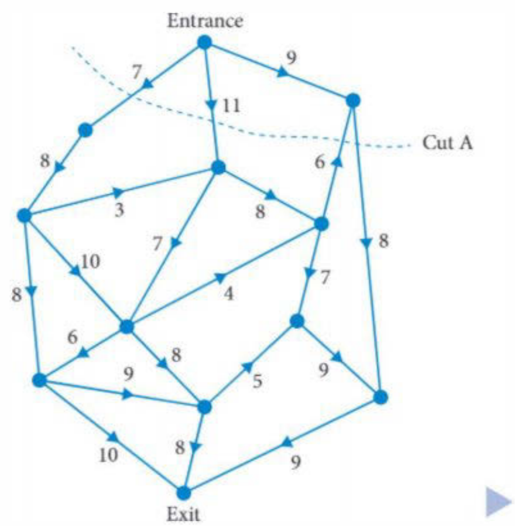
a Determine the capacity of Cut 1. 1 mark

b Determine the maximum number of seats available for children for a journey that begins at the West Terminal and ends at the East Terminal. 1 mark

On one particular train, 10 children set out from the West Terminal. No new passengers board the train on the journey to the East Terminal.

c Determine the maximum number of children who can arrive at the East Terminal on this train. 1 mark

**21** ©VCAA | 2017N2NQ3 (3 marks) Simon built his holiday home on an estate. The estate has one-way streets between the entrance and the exit. There are restrictions on the number of trucks that are allowed to travel along each street per day. On the directed graph, the vertices represent the intersections of the one-way streets. The number on each edge is the maximum number of trucks that are allowed to travel along that street per day.



When considering the possible flow of trucks through this network, many different cuts can be made.

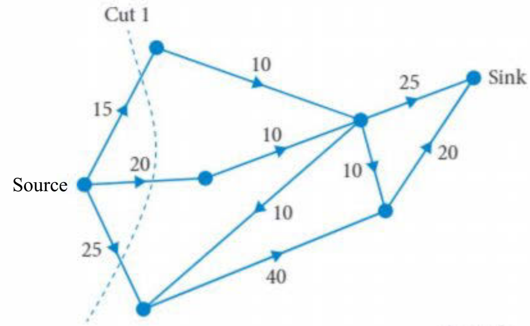
a Determine the capacity of Cut A, shown above. 1 mark

b Find the maximum number of trucks that could travel from the entrance to the exit per day. 1 mark

c A company would like to send one group of trucks from the entrance to the exit. All trucks in this group must follow each other and travel along the same route. The trucks in this group will be the only trucks to use these streets on that day. What is the maximum number of trucks that could be in this group? 1 mark

22 ©VCAA 2018N 2NQ3 (3 marks) All areas of a property require a constant supply of water.

The following directed graph represents the capacity, in litres per minute, of a series of water pipes on a property connecting the source to the sink.




When considering the possible flow through this network, different cuts can be made. Cut 1 is labelled on the graph above.

a What is the capacity of Cut 1 in litres per minute? 1 mark

b Copy the graph above and draw the cut (Cut 2) that has a capacity of 70 litres per minute. Label your answer clearly as Cut 2. 1 mark

c Determine the maximum flow of water, in litres per minute, from the source to the sink. 1 mark

Video playlist Shortest path problems

# 1 @ Shortest path problems

## Dijkstra's shortest path algorithm

Dijkstra's algorithm can also be used for finding the shortest path between two vertices of a directed graph. In these cases, the shortest possible paths are affected by the direction indicated on the edge.

### Dijkstra's algorithm for finding the shortest path

To find the shortest path using Dijkstra's algorithm:

1 Draw a box around the starting vertex label and add a value of 0.

2 For all the vertices connected to the starting vertex, add the value of each edge.

3 Draw a box around the smallest of the unboxed vertices.

4 For all the unboxed vertices connected to the newly boxed vertex, add the value of each edge. If an unboxed vertex already has a value, replace the existing value with the new value if the new value is smaller.

5 Draw a box around the smallest of the unboxed vertices. Repeat the steps for the newly boxed value until the end vertex is boxed. The value in the end box is the shortest amount.

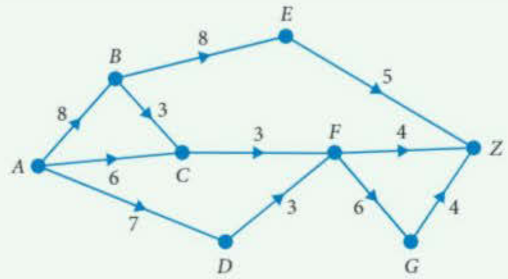
6 To find the path, start from the end vertex and choose the boxed value that gives the correct value when you subtract the edge. Continue backtracking like this until you get back to the start.

**10.5**

WB p. 210

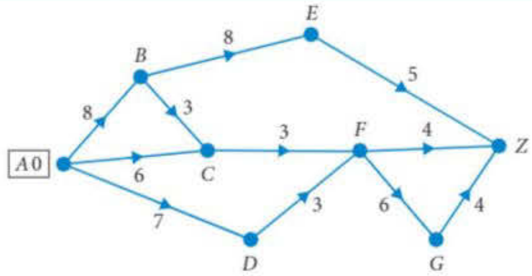
## WORKED EXAMPLE 16 Applying Dijkstra's algorithm to directed graphs

The times, in minutes, to various tracks in an orienteering event are shown in the network. Find the quickest path from $A$ to $Z$ using Dijkstra's algorithm.




| Steps | Working |
|---|---|

**1** Draw a box around the starting vertex $A$ and enter a value of zero.



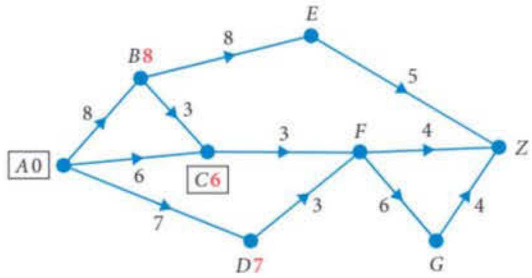

**2** The boxed vertex $A$ is connected to vertices $B$, $C$ and $D$. Add the times on the connecting edges to the boxed value of 0 and write these totals next to the vertices, shown in red on the diagram.

Vertex $B = 0 + 8 = 8$

Vertex $C = 0 + 6 = 6$

Vertex $D = 0 + 7 = 7$

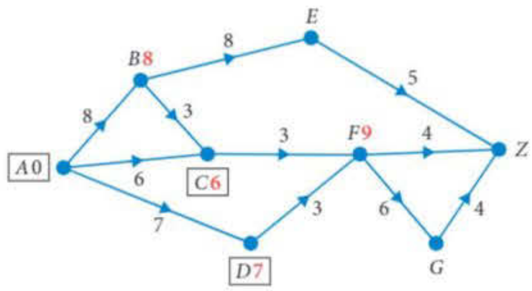
Draw a box around the smallest value, that is, $C6$.




**3** Consider all the unboxed vertices connected to the vertex $C$, that is $F$ only. Add the edge value 3 to the boxed value at $C$.

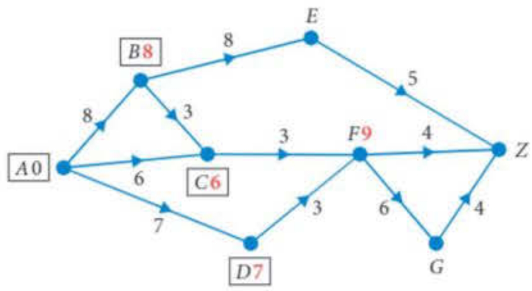
Vertex $F = 6 + 3 = 9$

Draw a box around the smallest value, that is, $D7$.




**4** $D$ is connected to $F$ only: $7 + 3 = 10$, which is larger than the current vertex value, so we leave 9 as the value for $F$.

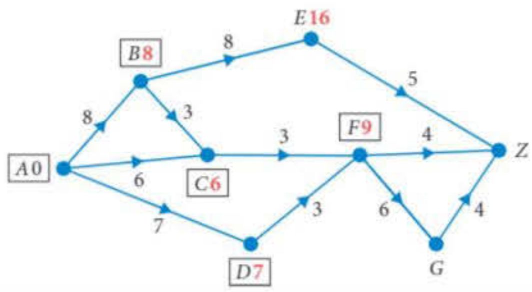
Draw a box around the smallest value, that is, $B8$.




**5** Consider all the unboxed vertices connected to the vertex $B$ which is $E$ only.

Vertex $E = 8 + 8 = 16$

Draw a box around the smallest value, that is, $F9$.

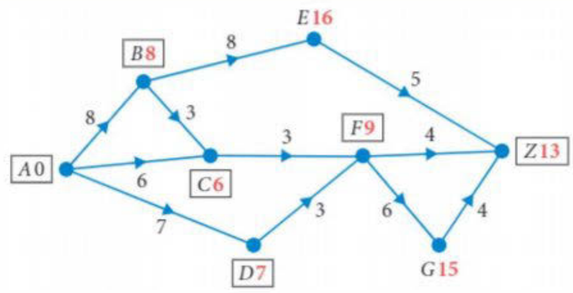


6 Consider all the unboxed vertices connected to the vertex *F*.

Vertex G = 9 + 6 = 15

Vertex Z = 9 + 4 = 13

Draw a box around the smallest value, that is, Z13.



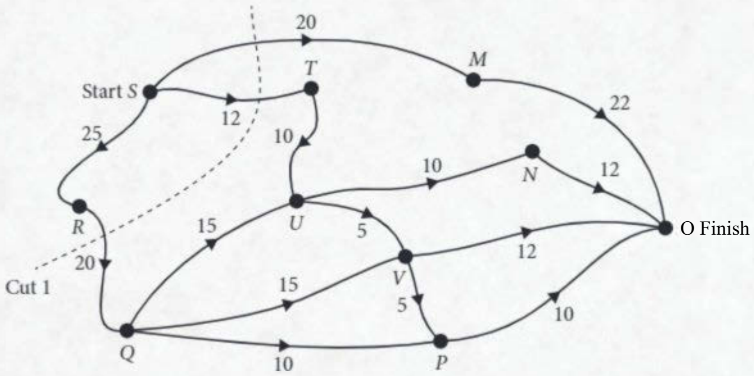

7 Backtrack from vertex Z to vertex *A*. The shortest time path of 13 minutes is *A*-*C*-*F*-*Z*.

## VCE QUESTION ANALYSIS

Video VCE question analysis: Directed graphs

©VCAA 2020 2NQ4 | 2020 Examination 2 Networks and decision mathematics Question 4 (3 marks)

A cricket team has a training program starting from exercise station S and running to exercise station *O*. For safety reasons, the cricket coach has placed a restriction on the maximum number of people who can use the tracks in the fitness park. The directed graph shows the capacity of the tracks, in number of people per minute.




a How many different routes from S to *O* are possible?

When considering the possible flow of people through this network, many different cuts can be made. 1 mark

b Determine the capacity of Cut 1, shown above. 1 mark

c What is the maximum flow from S to O, in number of people per minute? 1 mark

### Reading the question

- Take note of the information in the question about the network flow problem: the source S, the sink O and the numbers on the directed edges representing the maximum number of people per minute.
- Highlight the type of answer required in each question. This may be the number of paths from S to O, the capacity of a cut or the flow capacity.
- Even though you need to identify the number of possible paths, you are not required to list each path.

### Thinking about the question

- The question requires an understanding of network flow.
- You will need to know how to determine the capacity of a cut.
- You will also need to know how to determine the maximum flow capacity of a network.



9780170448406

**Worked solution** (/ = 1 mark)

a The routes from S to *O* are:

*S-M-O*

*S-T-U-N-O*

*S-T-U-V-0*

*S-T-U-V-P-O*

*S-R-Q-U-N-O*

*S-R-Q-U-V-P-O*

*S-R-Q-U-V-0*

*S-R-Q-V-O*

*S-R-Q-V-P-O*

*S-R-Q-P-O*

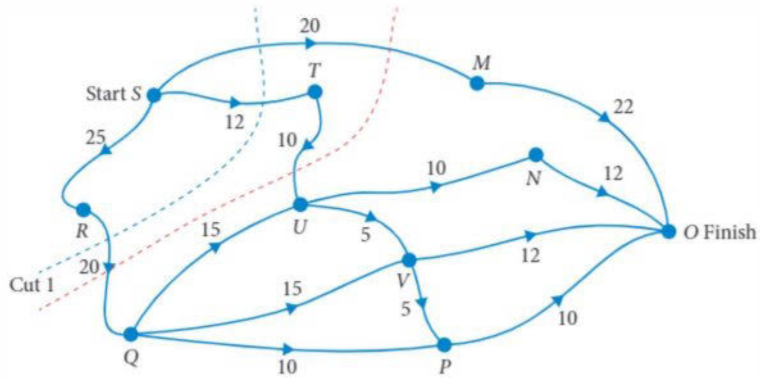
There are 10 / different routes from S to *O*.

b The capacity of this cut is the total capacity of the three edges it passes through: 20 + 12 + 20 = 52 /

c The maximum flow is equal to the capacity of the minimum cut through the network. The minimum cut can be seen in the diagram:

20+ 10 + 20 = 50

Maximum flow capacity is 50 / people per minute.




**Student performance** 80–100% 60–79% 0–59%

a 30% 9 was a common incorrect response,

b 69%

C 32%

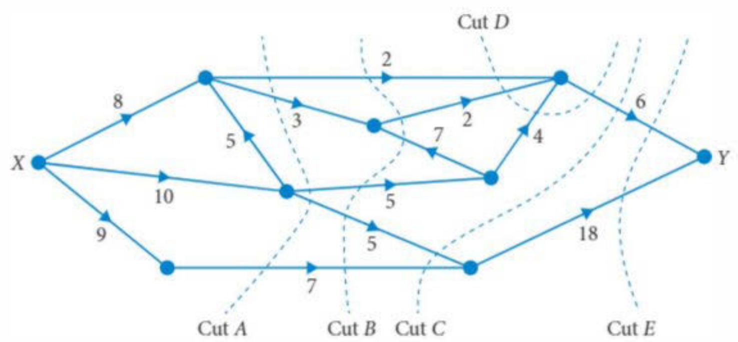
## EXERCISE 10.5 Shortest path problems

ANSWERS p. 729

### Recap

*Use the following information to answer the next two questions.*

In this directed graph, the values on the edges give the maximum flow between vertices in the direction of the arrows.



1 The capacity of cut *B* is

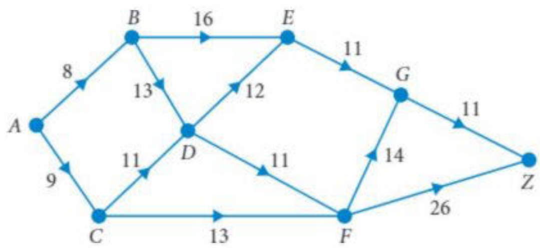
A 18 B 21 C 23 D 24 E 28

2 The invalid cut is

A Cut A B Cut *B* C Cut C D Cut *D* E Cut *E*

## Mastery

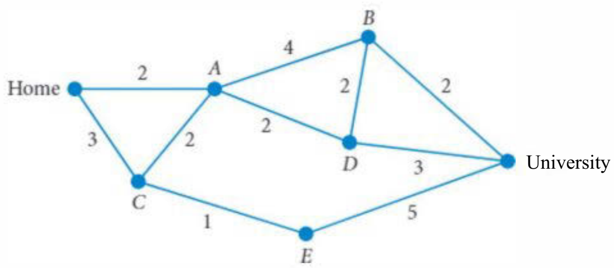
3 H WORKED EXAMPLE 16 Find the shortest path from *A* to Z and give the value of the shortest path.




## Exam practice

80-100% 60-79% 0-59%

4 ©VCAA 2018 1NQ2 68% Niko drives from his home to university. The network below shows the distances, in kilometres, along a series of streets connecting Nikos home to the university. The vertices A, B, C, D and *E* represent the intersection of these streets.

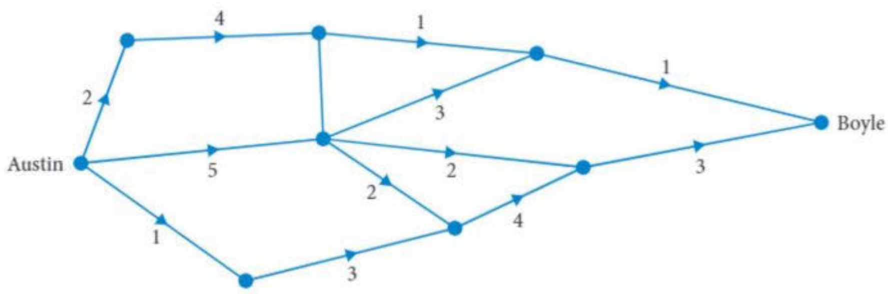



The shortest path for Niko from his home to the university could be found using

A a minimum cut. B Prims algorithm. C Dijkstras algorithm.

D critical path analysis. E the Hungarian algorithm.

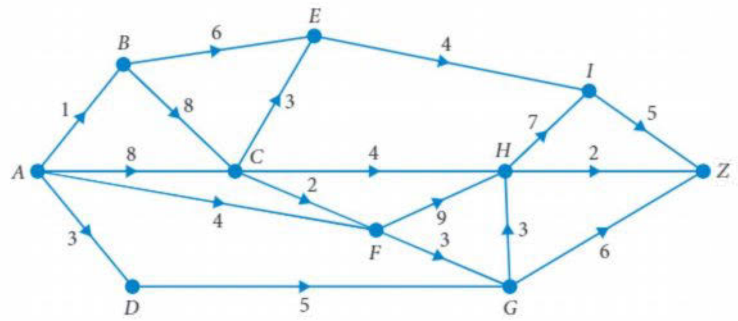
5 The network shows the distances, in kilometres, along a series of one-way roads that connect the cities of Austin and Boyle. The shortest distance between Austin and Boyle is




**A** 8 **B** 9 **C** 10 **D** 11 **E** 14

*Use the following information to answer the next three questions.*

The time, in minutes, taken to traverse a series of one-way tracks is shown in the directed graph below.




6 The shortest time taken to travel from vertex *A* to vertex *H* is

A 4 B 10 C 11 D 12 E 13

7 The shortest path from vertex *A* to vertex Z is

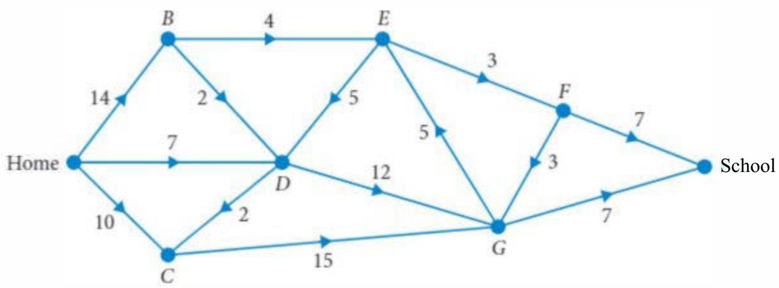
A *A-F-G-H-Z* B *A-F-G-Z* C *A-D-G-Z*
D *A-B-E-I-Z* E *A-B-C-H-Z*

8 The shortest time to travel from vertex *A* to vertex *Z* is

A 11 minutes B 12 minutes C 13 minutes
D 14 minutes E 15 minutes

*Use the following information to answer the next two questions.*

The network below shows the travel times, in minutes, along a series of roads that connect a students home to school. The student may only travel in the direction indicated by the arrows.




9 The shortest time, in minutes, for this student to travel from home to school is

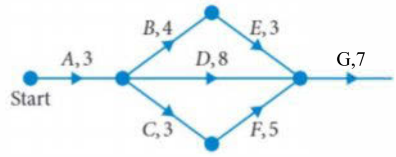
A 22 B 23 C 26 D 28 E 29

10 The shortest path from home to school is

A *home - D - G - school* B *home - D - E - F - school*
C *home - B - E - F - school* D *home - C - G - school*
E *home - B - D - G - school*

# @ Chapter summary

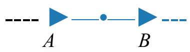
## The scheduling problem




- The edges on a directed graph, or digraph, have a direction. In a weighted digraph, there is a number associated with each edge.
- An activity table shows the order and estimated time for each activity.
- An immediate predecessor is any activity that must be completed before this activity can commence.

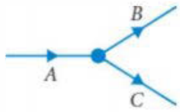
  There are three different types of connections.

  1 One preceding event

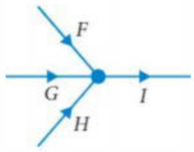
  Event *B* is preceded by *A*.

  2 More than one activity with the same preceding event

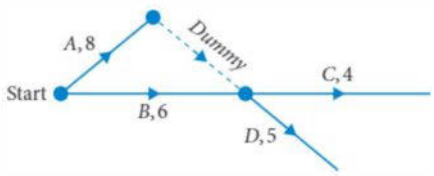
  Events *B* and C are preceded by *A*.

  3 An activity with more than one preceding event

  Event *I* is preceded by F, *G* and *H*.

- When drawing a network
  - use a vertex to represent the start of the network
  - look for any activities that do not have any predecessors - these will be your starting activities
  - multiple predecessors to an activity will all end at the same vertex
  - an activity should not be represented by more than one edge in the network
  - two vertices can be connected by one edge only
  - a vertex indicating the completion of the project needs to be included in the network.
- Reachability is the ability to get from one vertex to another vertex in a directed graph.
- A dummy activity needs to be added to a network to ensure that no two vertices are connected by multiple edges or to maintain precedence structure.
- A dummy activity has zero time and is shown as a directed edge with a broken line.




## Forward scanning to determine EST

- The earliest start time (EST) for an activity is the earliest time it is possible to start the activity.
- Forward scanning through a network enables us to determine the earliest start time (EST) for every activity in the network.

  To determine the earliest start time (EST):

  1 Draw a double box, at the start vertex, for each activity.

  2 Begin with the first vertex, which has an earliest start time of 0. We will use the convention of writing the earliest starting times in the top box next to each vertex.

  3 ESTs are calculated from left to right.

  4 Add the activity time to the EST of the previous vertex. If more than one activity leads to the vertex, the *highest figure* obtained becomes the new EST.

  5 Continue until the finish is reached.


9780170448406

## Backwards scanning to determine LST

- Backward scanning is the process used to find the latest start times (LST) for activities.
- The latest start time (LST) is the latest time you can start an activity without affecting the project completion time.
- The LST at the left vertex = the LST at the right vertex - the activity time


- To determine the latest start time (LST):
  1 Commence LST calculations at the 'finish' vertex. At the 'finish' vertex, LST = EST.
  2 Work backwards from right to left. LSTs are written in the bottom box next to each vertex.
  3 To find the LST at the left vertex, work backwards subtracting the activity time from the LST at the right vertex. If there is more than one LST, subtract the activity time from the lowest LST.
  4 One of the LST values at the first vertex must be zero.

## The critical path

- The critical path is the longest time path in the network
- To determine the critical path
  1 Use forward scanning to determine the earliest start times for each activity.
  2 Identify the path or paths that produce the final EST.
  3 Use backward scanning to determine the latest start times for each activity.
  4 Activities for which EST = LST are on the critical path.

## Activity float time and project crashing

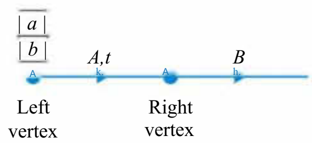
The float time for an activity is the maximum time an activity can be extended or postponed without effecting the project completion time.

In the diagram shown

$a$ = the EST of activity $A$

$b$ = the LST for activity $A$

Float for activity $A = b - a$



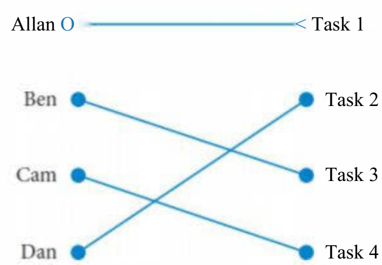

For a

float for non-critical activities = LST - EST

float time for critical activities = 0

## Bipartite graphs

- A bipartite graph has its vertices in two distinct sets and the edges join elements in the first set to elements in the second set.




## The Hungarian algorithm (stage one) - Row and column reduction

1 Choose the smallest number in each row and subtract it from every element in the same row.

2 For every column that does not have a zero value, choose the smallest number in the column and subtract it from every element in the same column.

3 Cover all the zeros with the smallest number of lines (horizontal or vertical). The process is complete if the number of lines is equal to the number of rows.

## The Hungarian algorithm - Stage two

1 Perform a row and column reduction and cover all the zeros with the smallest number of horizontal or vertical lines possible. If the number of lines is equal to the number of rows, a solution has been found.

2 If the number of lines does not equal the number of rows, find the smallest uncovered number and add it to every covered number. Numbers that are covered twice have this number added twice.

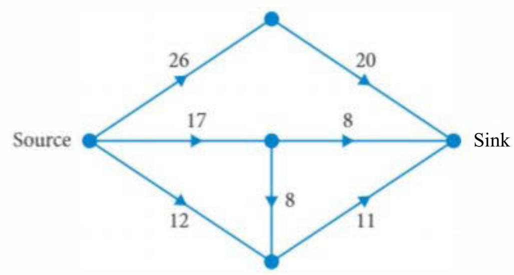
3 Subtract the smallest number in the matrix or table from all the numbers in the table.

4 Cover all the zeros with the smallest number of horizontal or vertical lines and if the number of lines is equal to the number of rows, the process is complete. If this is not the case repeat steps 2,3 and 4 until a solution is found.

5 Draw a bipartite graph and use it to determine the optimum allocation.

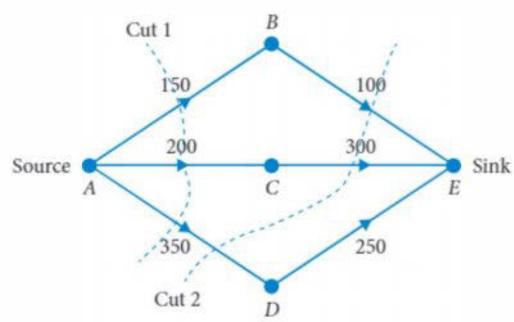
## Flow capacity and maximum flow

- The first vertex of the network is called the source and the final vertex of the network is called the sink.
- The inflow capacity is the total flow capacity entering a vertex and the outflow capacity is the total flow capacity leaving the vertex.
- The maximum flow through a vertex is the smaller of the total inflow capacity of the vertex and the total outflow capacity of the vertex.




## Maximum flow - minimum cut

- The maximum flow from source to sink can be determined by finding the cut that produces the minimum value.
- The maximum flow from source to sink = the capacity of the minimum cut.
- To determine the maximum flow for a network
  1 Identify cuts through the network.
  2 Find the capacity of each cut.
  3 maximum flow = capacity of the minimum cut.




Valid cuts

## Dijkstra's shortest path algorithm

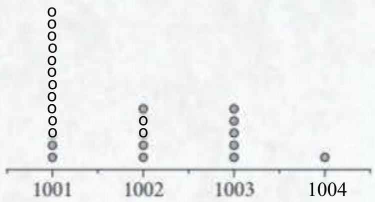
- Dijkstra's algorithm can also be used for finding the shortest path for directed graphs. In these cases, the shortest possible paths are affected by the direction indicated on the edge.
- To find the shortest path using Dijkstra's algorithm:
  1 Box the starting vertex and add a value of 0.
  2 For all the vertices connected to the starting vertex, add the value of each edge.
  3 Box the smallest of the unboxed vertices.
  4 For all the unboxed vertices connected to the newly boxed vertex, add the value of each edge. If an unboxed vertex already has a value, replace the existing value with the new value if the new value is smaller.
  5 Box the smallest of the unboxed vertices. Repeat the steps for the newly boxed value until the end vertex is boxed. The value in the end box is the shortest amount.
  6 To find the path, start from the end vertex and choose the boxed value that gives the correct value when you subtract the edge. Continue backtracking like this until you get back to the start.


9780170448406

# Cumulative examination 1

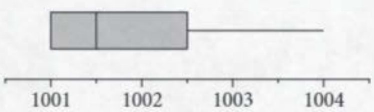
Total number of marks: 18 Reading time: 7 minutes Writing time: 41 minutes

1 OVCAA 2015 1CQ8 A dot plot for a set of data is shown.

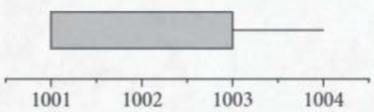



Which one of the following boxplots would best represent the dot plot?

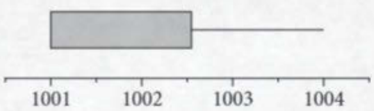
**A**



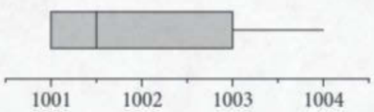

**B**



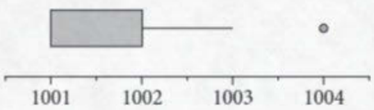

**C**




**D**

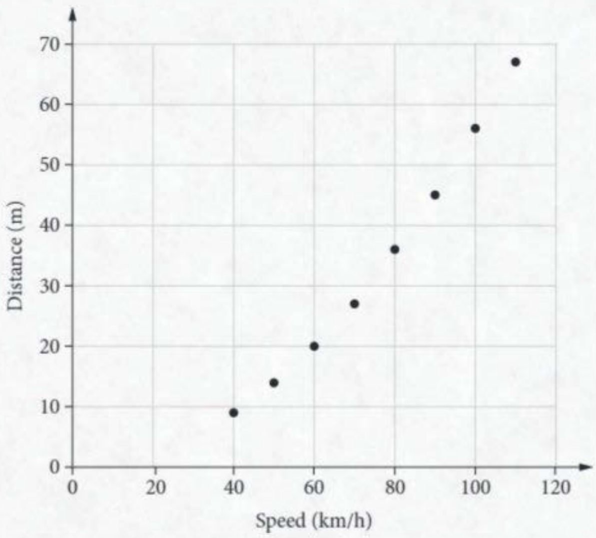



**E**




2 ©VCAA 2017N10012! The table shows the *speedy* in kilometres per hour, and the braking *distance,* in metres, of a car travelling at eight different speeds. A scatterplot has been constructed from this data.

| Speed (km/h) | Distance (m) |
|---|---|
| 40 | 9 |
| 50 | 14 |
| 60 | 20 |
| 70 | 27 |
| 80 | 36 |
| 90 | 45 |
| 100 | 56 |
| 110 | 67 |




Data: ©The State of Queensland (Department of Transport and Main Roads) 2010-2016

The scatterplot shows that the association between *distance* and *speed* is non-linear. A squared transformation is applied to the variable *speed* to linearise the data. A least squares line is then fitted to the transformed data with *distance* as the response variable. The equation of this least squares line is closest to

A *distance* —15.6 + 180 x *speed*$^2$

B *distance* = 0.0056 + 0.092 x *speed*$^2$

C *distance* = 0.092 + 0.0056 x *speed*$^2$

D *speed*$^2$ = 180 - 15.6 x *distance*

E *speed*$^2$ = 0.0056 + 0.092 x *distance*$^2$

*Use the following information to answer the next two questions.*

The seasonal indices (SI) for the daily earnings of a cafe in a tourist town, from Monday to Saturday, are shown in the table. The seasonal index for Sunday is not shown.

| Day | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday | Sunday |
|---|---|---|---|---|---|---|---|
| SI | 0.65 | 0.60 | 0.74 | 0.82 | 1.12 | 1.45 | |

3 ©VCAA 2017N 1CQ15 Last Sunday, a cafe earned $3839. The de-seasonalised earnings for this day were closest to

A $2370 B $2500 C $2650 D $5570 E $6220

4 ©VCAA 2017N 1CQ16 The seasonal index for Wednesday is 0.74. This tells us that Wednesday earnings tend to be

A 26% less than the average daily earnings.

B 26% more than the average daily earnings.

C 35% less than the average daily earnings.

D 35% more than the average daily earnings.

E 74% less than the average daily earnings.

5 ©VCAA I 2016SICQ2Q3 Rohan invests $15000 at an annual interest rate of 9.6% compounding monthly. Let $V_n$ be the value of the investment after $n$ months. A recurrence relation that can be used to model this investment is

A $V_o$= 15000, $V_{n+1}$-0.96V„

B $V_o$= 15000, $V_{„+1}$- 1.008 *V*„

C $V_{(1}$= 15000, $V_{„+1}$ = 1.08V„

D $V_o$= 15000, $V_{B+1}$-1.0096V,

E $V_o$ = 15000, $V_{n+1}$ = 1.096 V„

6 ©VCAA | 2007 IBRMQ9~1 Petra borrowed $250 000 to buy a home. The interest rate is 7% per annum, calculated monthly on the reducing balance over the life of the loan. She will fully repay the loan over 20 years with equal monthly instalments. The total amount of interest she will pay on the loan is closest to

A $215000 B $266000 C $281000 D $350000 E $465000

7 ©VCAA 1 2017N1MQ8 Matrix *A* is an *n* x *n* matrix where *n* > 1. Matrix R is a row matrix. Matrix C is a column matrix. Which one of the matrix products below could result in a 1 x 1 matrix?

A *ACR* B *ARC* C *CAR* D *RAC* E *RCA*

8 ©VCAA 2007 1MQ3 If A = $\begin{bmatrix} 8 & 4 \\ 5 & 3 \end{bmatrix}$ and the product *AX* = $\begin{bmatrix} 5 & 6 \\ 8 & 10 \end{bmatrix}$, then *X* is

A $\begin{bmatrix} 24 & -14 \\ 13 & -7.5 \end{bmatrix}$

B $\begin{bmatrix} -4.25 & -5.5 \\ 9.75 & 12.5 \end{bmatrix}$

C $\begin{bmatrix} -3.75 & 7 \\ -6.5 & 12 \end{bmatrix}$

D $\begin{bmatrix} 25 & 11 \\ -19.5 & -8.5 \end{bmatrix}$

E $\begin{bmatrix} 0.625 & 1.5 \\ 1.6 & 3.333 \end{bmatrix}$

9 The Leslie matrix shown models the changes of an endangered numbat population over 5 years.

$$\begin{bmatrix} 0 & 0 & 1.1 & 0.7 & 0.3 \\ 0.4 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0.6 & 0 & 0 \\ 0 & 0 & 0 & 0.2 & 0 \end{bmatrix}$$

Which one of the following is indicated by the matrix?

A Numbats have 40% chance of dying before they become one year old.

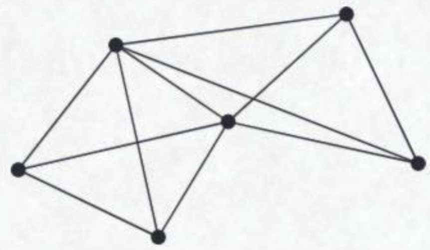
B One-year-old numbats give birth.

C Three-year-old numbats have the highest birth rate.

D Three-year-old numbats have an 80% chance of dying before they become four years old.

E Some numbats live to be five years old.

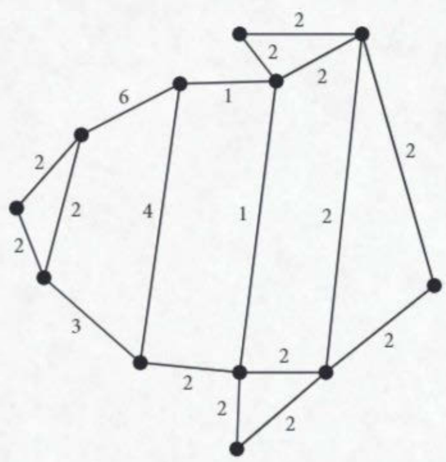
10 ©VCAA 2017N1NQ3 The graph shown is planar.



How many faces does this graph have?

A 5 B 6 C 7 D 8 E 9

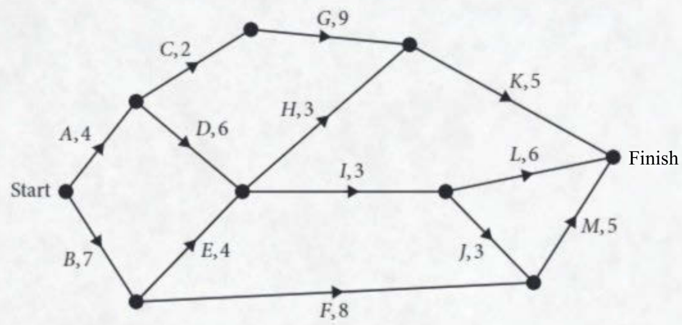
11 ©VCAA 2018N 1NQ6 A minimal spanning tree is to be drawn for the weighted graph.




How many edges with weight 2 will not be included in the minimal spanning tree?

A3 B4 C 5 D 6 E7

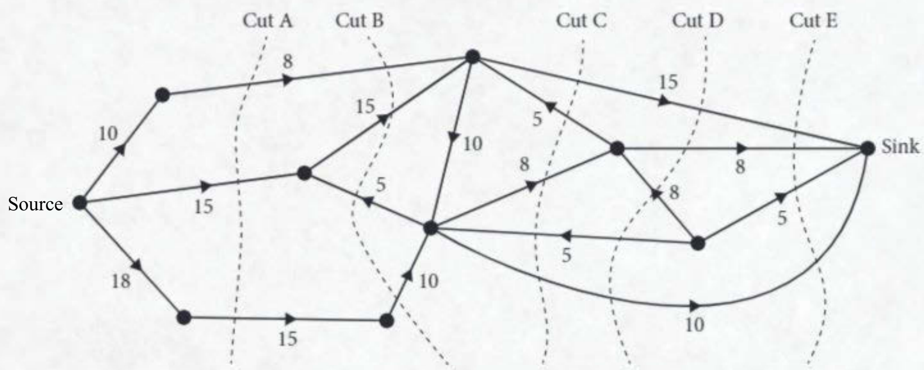
12 ©VCAA 20201NQ6 I The activity network below shows the sequence of activities required to complete a project. The number next to each activity in the network is the time it takes to complete that activity, in days.




The minimum completion time for this project, in days, is

A 18 B 19 C 20 D 21 E 22

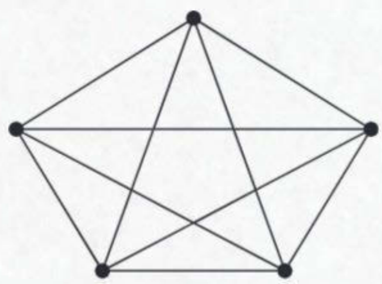
13 ©VCAA 2020 1NQ9 The flow of liquid through a series of pipelines, in litres per minute, is shown in the directed network below.




Five cuts labelled A to E are shown on the network. The number of these cuts with a capacity equal to the maximum flow of liquid from the source to the sink, in litres per minute, is

Al B2 C 3 D 4 E 5

14 ©VCAA 2016 1NQ3 J The following graph with five vertices is a complete graph.

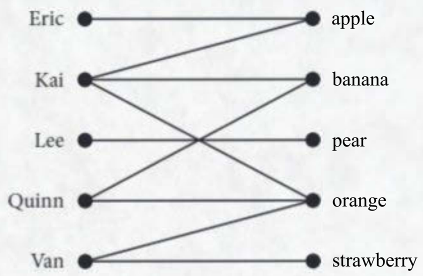


Edges are removed so that the graph will have the minimum number of edges to remain connected. The number of edges that are removed is

A4 B 5 C 6 D9 E10


9780170448406

15 ©VCAA 2021 1NQ2 Five friends ate fruit for morning tea. The bipartite graph shows which types of fruit each friend ate.




Which one of the following statements is not true?

A Only Lee ate pear.

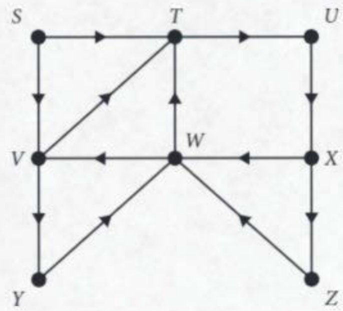
B Eric and Kai each ate apple.

C Van ate only strawberry.

D Quinn and Kai each ate banana.

E Orange was the most eaten type of fruit.

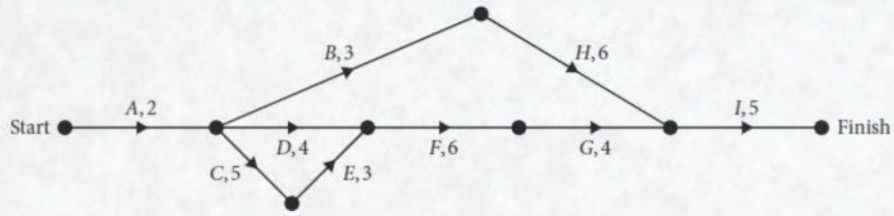
16 ©VCAA 2021 1NQ4 Consider the directed network.




The number of vertices that cannot be reached from $X$ is

A l B 2 C 3 D 4 E5

17 ©VCAA 20201NQ10 The directed network below shows the sequence of activities, $A$ to I, that is required to complete an office renovation. The time taken to complete each activity, in weeks, is also shown.




The project manager would like to complete the office renovation in less time.

The project manager asks all the workers assigned to activity $H$ to also work on activity $F$. This will reduce the completion time of activity $F$ to three weeks.

The workers assigned to activity $H$ cannot work on both activity $H$ and activity $F$ at the same time.

No other activity times will be changed.

This change to the network will result in a change to the completion time of the office renovation.

Which one of the following is correct?

A The completion time will be reduced by one week if activity $F$ is completed before activity $H$ is started.

B The completion time will be reduced by three weeks if activity $F$ is completed before activity $H$ is started.

C The completion time will be reduced by one week if activity $H$ is completed before activity $F$ is started.

D The completion time will be reduced by three weeks if activity $H$ is completed before activity $F$ is started.

E The completion time will be increased by three weeks if activity $H$ is completed before activity $F$ is started.

18 © VCAA 20201NQ7 I Four friends go to an ice-cream shop. Akiro chooses chocolate and strawberry ice cream. Doris chooses chocolate and vanilla ice cream. Gohar chooses vanilla ice cream. Imani chooses vanilla and lemon ice cream.

This information could be presented as a graph. Consider the following four statements:

- The graph would be connected.
- The graph would be bipartite.
- The graph would be planar.
- The graph would be a tree.

How many of these four statements are true?

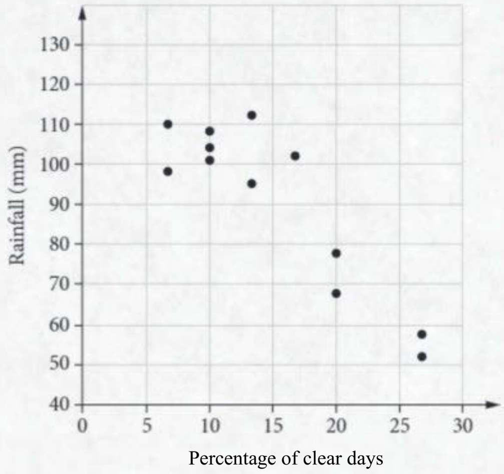
AO Bl C 2 D 3 E 4

# Cumulative examination 2



Total number of marks: 28 Reading time: 8 minutes Writing time: 42 minutes

1 ©VCAA 2009 2CQ3 (5 marks) The scatterplot shows the *rainfall* (in mm) and the *percentage of clear days* for each month of 2008.




An equation of the least squares line of best fit for this data set is

*rainfall* = 131 - 2.68 x *percentage of clear days*

a Copy the scatterplot and draw this line on it. 1 mark

b Use the equation of the least squares line of best fit to predict the rainfall for a month with 35% of clear days. Write your answer in mm correct to one decimal place. 1 mark

**Exam hack**

To ensure accuracy when drawing a straight-line graph from an equation, make sure the two points you choose are not close together.

c The coefficient of determination for this data set is 0.8081.

i Interpret the coefficient of determination in terms of the variables *rainfall* and *percentage of clear days.* 1 mark

**Exam hack**

When asked to 'interpret' an association between two variables, don't use words that imply something *causes* something else.

ii Determine the value of the Pearson correlation coefficient. Write your answer correct to three decimal places. 2 marks

2 ©VCAA 2016S2CQ8J (5 marks) Hugo won \$5000 in a cycling road race. He deposited this money into a savings account. The value of Hugos savings after *n* months, $S_w$, can be modelled by the recurrence relation below.

$S_0 = 5000,\ S_{w+1} = 1.004\ S_n$

*a* What is the annual interest rate (compounding monthly) for Hugo's savings account? 1 mark

b What would be the value of Hugo's savings after 12 months? 1 mark

Using a different investment strategy, Hugo could deposit $3000 into an account earning compound interest at the rate of 4.2% per annum, compounding monthly, and make additional payments of $200 after every month. Let $T_n$ be the value of Hugos investment after $n$ months using this strategy. The monthly interest rate for this account is 0.35%.

c i Write down a recurrence relation, in terms of $T_{n+}$ j and $T_n$, that models the value of Hugos investment using this strategy. 1 mark

ii What is the total interest Hugo would have earned after six months? 2 marks

3 ©VCAA [ 2017N2MQI~] (4 marks) People pay to attend concerts at the Whiteoak Theatre. They can choose their seats for each concert from three classes, A, *B* or C. The table shows the number of seats available in each class and the cost per seat.

| Class | Number of seats available | Cost per seat ($) |
|---|---|---|
| *A* | 100 | 45 |
| *B* | 340 | 35 |
| *C* | 160 | 30 |

a The column matrix *N* contains the number of seats in each class.

$$N = \begin{bmatrix} 100 \\ 340 \\ 160 \end{bmatrix} \begin{matrix} A \\ B \\ C \end{matrix}$$

What is the order of matrix N? 1 mark

b Matrix W contains the cost of each class of seat in the theatre.

*ABC*

W - [ 45 35 30 ]

i Determine the matrix product *WN*. 1 mark

ii Explain what the matrix product *WN* represents. 1 mark

c The number of seats that were sold for the first concert this year is shown in the table.

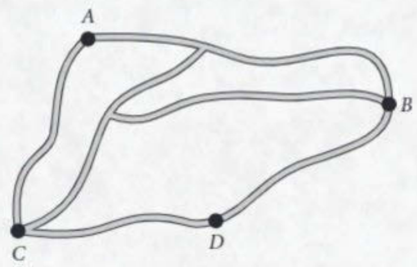
| Class | Number of seats available | Cost per seat ($) |
|---|---|---|
| *A* | 42 | 45 |
| *B* | 179 | 35 |
| *C* | 86 | 30 |

The information in the table is used to construct the matrix P, shown below.

$$\begin{bmatrix} 42 & 0 & 0 \\ 0 & 179 & 0 \\ 0 & 0 & 86 \end{bmatrix} \begin{bmatrix} 45 \\ 35 \\ 30 \end{bmatrix}$$

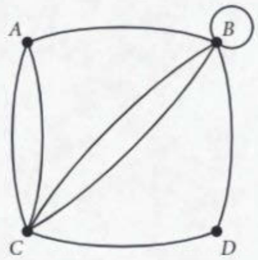
Matrix *P* contains the value of all seats in each class, in dollars, that were sold for the first concert this year. A matrix product *MP* is found where *M* = [0 1 1]. Explain what the matrix product *MP* represents. 1 mark

4 ©VCAA 2008 2NQ2 (3 marks) Four children each live in a different town. The following is a map of the roads that link the four towns, A, B, C and *D*.




a How many different ways may a vehicle travel from town *A* to town *D* without travelling along any road more than once? 1 mark

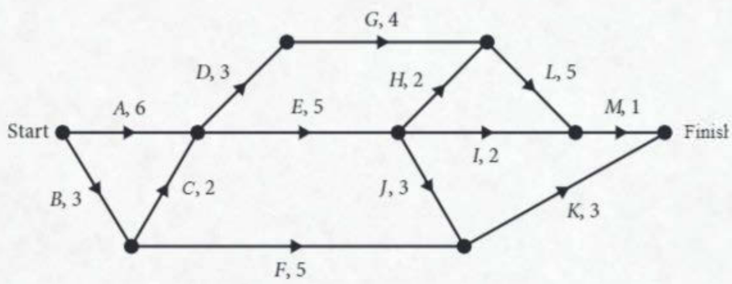
James' father has begun to draw a network diagram that represents all the routes between the four towns on the map. In this network diagram, vertices represent towns and edges represent routes between towns that do not pass through any other town.




b i One more edge needs to be added to complete this network diagram. Copy the diagram and draw in this edge clearly. 1 mark

ii With reference to the network diagram, explain why a motorist at *A* could not drive each of these routes once only and arrive back at *A*. 1 mark

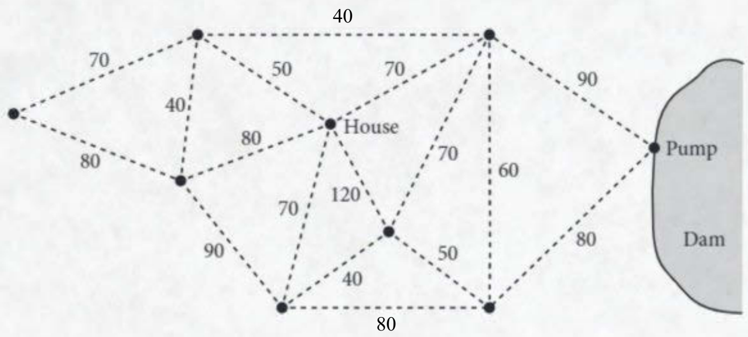
5 © VCAA 2021 2NQ4ab (2 marks) Roadworks planned by the local council require 13 activities to be completed. The network shows these 13 activities and their completion times in weeks.




a What is the earliest start time, in weeks, of activity K? 1 mark

b How many of these activities have zero float time? 1 mark

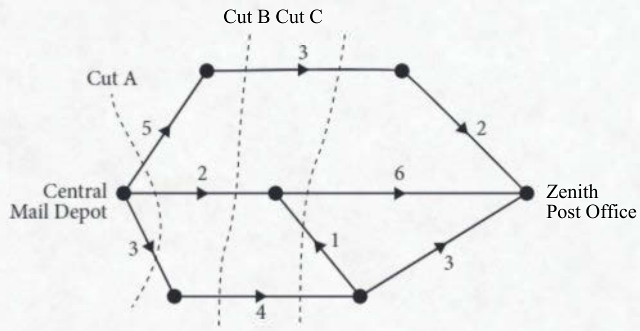
6 ©VCAA 2012 2NQ1b J (2 marks) The total length of pipe that supplies water from a pump to eight locations on a farm (including the house) is a minimum. This minimum length of pipe is laid along some of the edges in the network shown.




a Copy the diagram and draw in the minimum length of pipe that is needed to supply water to all locations on the farm. 1 mark

b What is the mathematical term that is used to describe this minimum length of pipe in part a? 1 mark

7 ©VCAA 2018 2NQ1 (3 marks) The graph below shows the possible number of postal deliveries each day between the Central Mail Depot and the Zenith Post Office. The unmarked vertices represent other depots in the region. The weighting of each edge represents the maximum number of deliveries that can be made each day.




a Cut *Ay* shown on the graph, has a capacity of 10. Two other cuts are labelled as Cut *B* and Cut C.

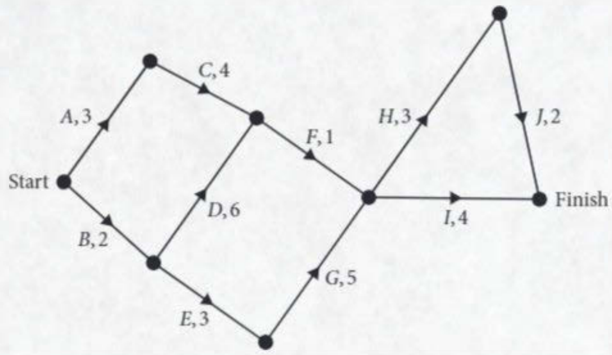
i Write down the capacity of Cut *B*. 1 mark

ii Write down the capacity of Cut C. 1 mark

b Determine the maximum number of deliveries that can be made each day from the Central Mail Depot to the Zenith Post Office. 1 mark


9780170448406

8 ©VCAA I 20182NQ3 J (4 marks) At the Zenith Post Office all computer systems are to be upgraded. This project involves 10 activities, *A* to *J*. The directed network below shows these activities and their completion times, in hours.




a Determine the earliest starting time, in hours, for activity *I*. 1 mark

b The minimum completion time for the project is 15 hours. Write down the critical path, 1 mark

c Two of the activities have a float time of two hours. Write down these two activities. 1 mark

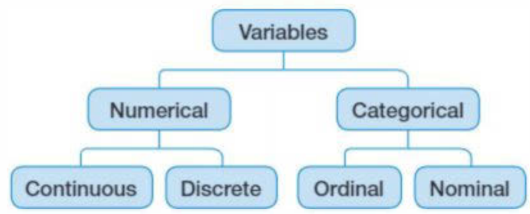
d For the next upgrade, the same project will be repeated but one extra activity will be added. This activity has a duration of one hour, an earliest starting time of five hours and a latest starting time of 12 hours. Copy and complete the following sentence by filling in the boxes.

The extra activity could be represented on the network above by a directed edge from the end of activity ☐ to the start of activity ☐. 1 mark

# Answers

## CHAPTER 1

### EXERCISE 1.1

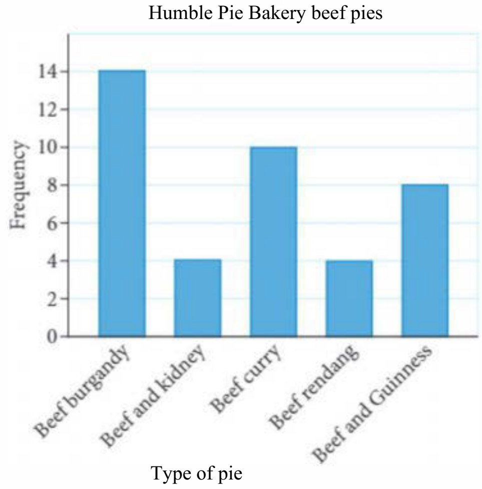
1 



2 a i numerical ii discrete
b i numerical ii continuous
c i categorical ii nominal
d i numerical ii discrete
e i numerical ii continuous
f i categorical ii nominal
9 i categorical ii nominal
h i categorical ii ordinal
i i numerical ii continuous
j i categorical ii ordinal
k i categorical ii ordinal
I i categorical ii ordinal
m i numerical ii discrete

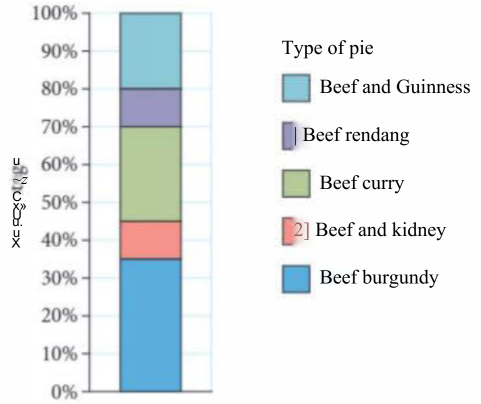
3 a

| Pie | Frequency | Percentage |
|---|---|---|
| Beef burgundy | 14 | 35 |
| Beef and kidney | 4 | 10 |
| Beef curry | 10 | 25 |
| Beef rendang | 4 | 10 |
| Beef and Guinness | 8 | 20 |
| Total | 40 | 100 |

b Humble Pie Bakery beef pies




C Humble Pie Bakery' beef pies




d beef burgundy

4 a i mode = 18 years ii mean = 25 years
iii median = 21 years iv range = 24 years
b i mode = 26 people
ii mean = 41.5 people
iii median = 38.5 people
iv range = 20 people
c i modes = -2.0°C and -1.7°C
ii mean = -0.6°C
iii median =-1.7°C
iv range = 5.8°C

5 C 6 E 7 A 8 D
9 A 10 D 11 E 12 A
13 a 10, 7, 8 b 32%
14 a 31 b 61%

15 a distance
b gender, phone number, language
c 75%
d English
e 1.35 km
f 3
g Phone numbers are nominal variables, and the median can only be calculated for continuous, discrete and ordinal variables.

16 a 6522
b NSW had 45% of the cases, Victoria had 20% of the cases.
c NSW has 32% of the population, but had 45% of the cases, Victoria has 26% of the population and 20% of the cases. Victoria had a lower percentage of cases than its population percentage. NSW had a higher percentage of cases than its population percentage. So, Victoria was more effective in restricting spread of COVID-19 up until 17 April 2020.

## EXERCISE *1.2

1 D

2 E

3 a 9

b positively skewed with no possible outliers

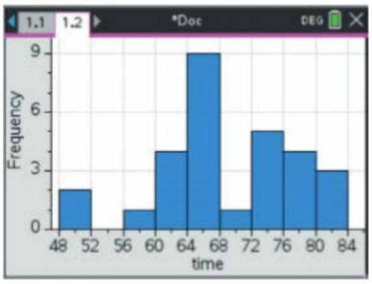
c 0-< 10 years

d 10-<20 years

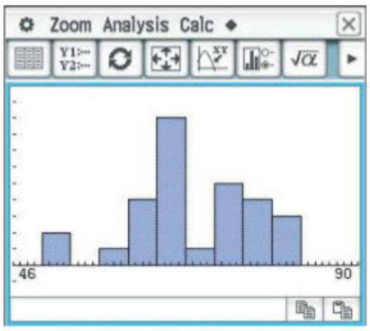
4 a false  b true  c false  d true

e true  f false  g true

5 TI-Nspire




ClassPad




6 B  7 B  8 D  9 D

10 D  11 E  12 C

13 a 14 countries  b 16.4%

14 a positively skewed

b i 26  ii $\frac{20}{103} = 19.4\%$

## EXERCISE 1.3

1 B

2 D

3 a min = 47, Qj = 57.5, median = 68.5, $Q_3$ = 81, max = 92

b 47,51,54, | 61,65, 66, | 71,73,74, | 88, 90,92

Lower quartile = 57.2  Upper quartile = 81

Median = 68.5

i 3 out of a total of 12 data values are less than the lower quartile. = $\frac{3}{12}$ 25%

ii 6 out of a total of 12 data values are less than the median. $\frac{6}{12}$ = 50%

iii 9 out of a total of 12 data values are less than the upper quartile. $\frac{9}{12}$ = 75%

4 a i lower fence:

Qj - 1.5 x IQR = 38.5 - 1.5 x 11 = 22

upper fence:

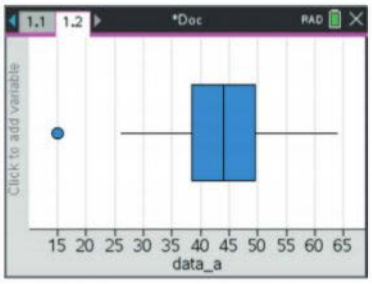
$Q_3$+ 1.5 x IQR = 49.5 + 1.5 x 11 =66

15 is less than 22, so its a possible outlier.

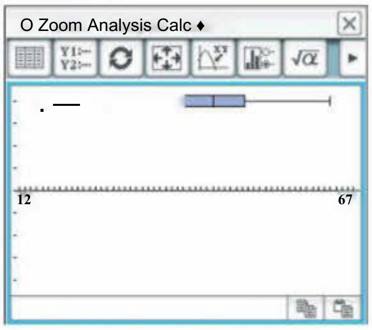
26 isn't less than 22, so its *not* an outlier.

64 isn't greater than 66, so it's *not* an outlier.

ii TI-Nspire




ClassPad




b i lower fence:

Qj - 1.5 x IQR = 88 - 1.5 x 17 = 62.5

upper fence:

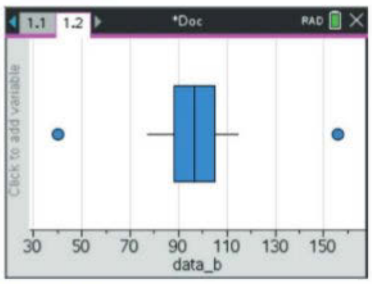
$Q_3$ + 1.5 x IQR = 105 + 1.5 x 17 = 130.5

40 is less than 62.5, so it's a possible outlier.

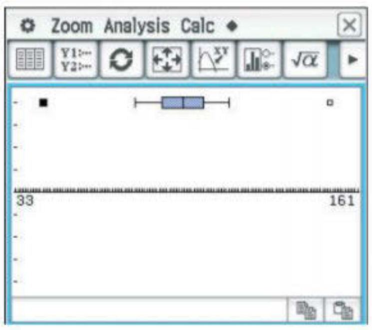
115 isn't greater than 130.5, so it's *not* an outlier.

156 is greater than 130.5, so it's a possible outlier.

ii TI-Nspire




ClassPad



c i lower fence:

Qj - 1.5 x IQR = 21.5 - 1.5 x 5 = 14

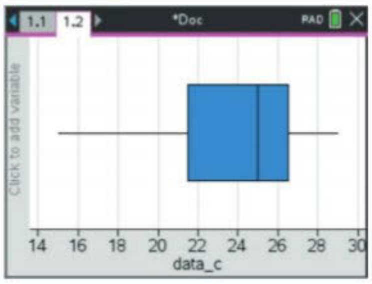
upper fence:

$Q_3$ + 1.5 x IQR = 26.5 + 1.5 x 5 = 34

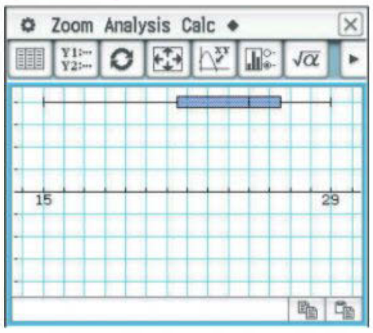
15 isn't less than 14, so it's *not* an outlier.

29 isn't greater than 34, so it's *not* an outlier.

ii TI-Nspire




ClassPad




5 a i min = 1, Qj = 3, median = 5, $Q_3$ = 8, max =11

ii 25% iii 100% iv 75% v 21

vi No scores at the lower end would be considered outliers.

vii scores greater than 15.5

b i min = 7, Q, = 8, median = 11, $Q_3$ = 13, max = 16

ii 75% iii 50% iv 50% v 0

vi scores less than 0.5

vii No scores at the upper end would be considered outliers.

c i min = 1, Q, = 2, median = 3, $Q_3$ = 8, max = 10

ii 25% iii 100% iv 50% v 42

vi No scores at the lower end would be considered outliers.

vii scores greater than 17

6 a negatively skewed; the box and whisker in the negative direction are longer than the box and whisker in the positive direction; two outliers shown by dots

b positively skewed; the box and whisker in the positive direction are longer than the box and whisker in the negative direction; no outliers

c approximately symmetric; median approximately in the middle of box and whiskers about the same length; one outlier shown by dot

7E 8 D 9B 10 D

11 C 12 D 13 D

14 a approximately symmetric with outliers

b 70 c 4000

15 a Qi = 8.0, $Q_3$= 13.5, IQR = 13.5-8.0 = 5.5

Qj - 1.5 x IQR = 8.0 - 1.5 x 5.5 = -0.25.

The lower fence is -0.25 and the minimum value is 2.5, so no value is less than the lower fence.

$Q_3$ + 1.5 x IQR = 13.5 + 1.5 x 5.5 = 21.75.

The upper fence is 21.75 and the maximum value is 20.0, so no value is greater than the upper fence.

Therefore, the boxplot will not include outliers,

b 

–"I—!—I—'—I—'—I—'—I—'—I—'—I—'—I—'—I-—'—I—'—I—!—T"
0 2 4 6 8 10 12 14 16 18 20 22

Sleep time (hours)

## EXERCISE 1.4

1 D 2 B

3 a i 7.42 ii 7.4

b i 12.92 ii 13

c i 0.36 ii 0.36

d i 1800.00 ii 1800

e i 20.67 ii 21

f i 72.00 ii 72

9 i 9.80 ii 9.8

4 a i 1 ii 22 iii 29

b 12.35%

5 a i 100 ii 19.953 iii 10

iv 5.012 v 1 vi 0.1

vii 0.063

b i 2.00 ii 8.00 iii 1.81

iv 2.95 v -0.301 vi 4.30

vii 0.903

6 a 10 100 1000 10000 100000

Annual income ($)

b 0.1 1 10 100 1000

Wingspan (mm)

C 0.0 0.1 0.3 1.0 3.2 10.0 31.6

Meat consumption (kg)

7E 8 D 9 C 10 B

11 D 12 C

13 a positively skewed with 3 (or at least 3) possible outliers

b 71.3% c 37

## EXERCISE 1.5

1 A 2 B

3 a io goals ii 5 goals iii 1 goal

iv 0 goals v 2.5 goals vi 2.5 goals

b positively skewed

4 a i 16% ii 52% iii 23%

iv 19% v 33% vi 14%

b 68 may be an outlier. Check using the upper fence.

$Q_3$ + 1.5 x IQR = 33 + 1.5 x 14 = 54.

68 > 54, so 68 is a possible outlier.

5 a

| Stem | Leaf |
|---|---|
| 4 | 3 5 9 |
| 5 | 0 2 7 8 |
| 6 | 1 2 4 5 7 8 |
| 7 | 0 2 3 9 |
| 8 | 2 4 9 |

Key:4|5-45

(other numbers can be used as the key)

b 20 matches c 89 points

d 25% e *approximately* symmetric

6 D 7 C 8 A 9 C

10 A 11 E 12 C 13 A

14 a day number

b Minimum temperature (°C)

| Stem | Leaf |
|---|---|
| 4 | 1 8 |
| 5 | |
| 6 | 0 7 |
| 7 | 0 5 7 |
| 8 | 0 6 |
| 9 | 0 2 8 |
| 10 | 7 |
| 11 | 8 |
| 12 | 7 |

Key: 4| 1 = 4.1 n«15

c i 12.2 ii 20%

15 a positively skewed b 24.55 c 37.5%

16 a mode = 78; range = 9

b lower fence = Q] - 1.5 x IQR = 70.5.

70 < 70.5, so 70 is a possible outlier.

## EXERCISE 1,6

1 D 2 E

3 $\bar{x}$ « 23.57, $s \approx$ 2.71

4 a x « 26.29, s« 2.42 b $x$ = 7.25, s « 2.40

5 a i mean = 8.6 items

ii median = 9 items

iii The median is an appropriate measure of centre because the distribution is skewed.

iv standard deviation = 1.5 items

v IQR = 2.5 items

vi The IQR is an appropriate measure of spread because the distribution is skewed.

b i mean = 45.2 people

ii median = 45 people

iii The mean and median are both appropriate measures of centre because the distribution is approximately symmetric with no outliers.

iv standard deviation = 13.2 people

v IQR =18 people

vi The standard deviation and IQR are both appropriate measures of spread because the distribution is approximately symmetric with no outliers.

6 a i median ii IQR

b i mean and median

ii standard deviationand IQR

7 D 8E 9 D 10 C

11 A

12 a 0.5

b median = 28, range = 56, IQR = 17

c add 6 to the second line: 1 | 2 4 6

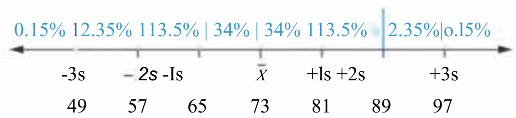
d The distribution is approximately symmetric.

## EXERCISE 1.7

1 E 2 C

3 a yes b no c no d yes

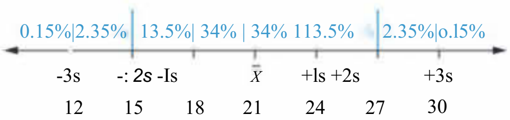
4 a




b 50% C 99.7% d 2.5% e 99.85%

f i 68 ii 290

5 a




b yes c no d yes e yes

f yes g yes h no i yes

j yes k yes

6 D 7 D 8 D 9 B

10 B 11 B 12 D 13 D

14 D

15 a i 20°C ii 23.3%

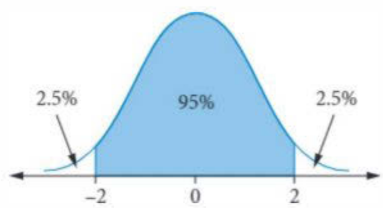
b 97.5%

## EXERCISE 1.8

1 A 2D

3 a -2

b $z$ = —2 means the teacher's height is two standard deviations below the mean.

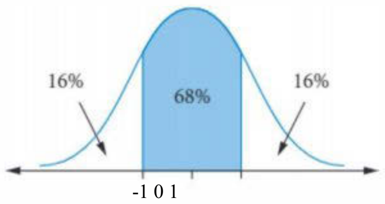



From the diagram, 97.5% of Year 12 teachers are taller than this particular teacher.

c 160.65 cm

4 a The z-scores are 1.2, -0.5,3, -1, -1.2, so the students' best subject was Hospitality,

b Using the following diagram, the student was in the bottom 16% for Psychology and Systems Engineering.




5 C   6 C   7 A   8 A

9 E   10 B   11 D

12 a -0.8

b i 16%   ii 163

## CUMULATIVE EXAMINATION 1

| | | |
|---|---|---|
| 1 B | 2 A | 3 B |
| 4 B 85% | 5 D 71% | 6 E 45% |
| 7 E 63% | 8 E | 9 E 66% |
| 10 C 58% | 11 C 150% | 12 C |
| 13 B | 14 E | |

## CUMULATIVE EXAMINATION 2

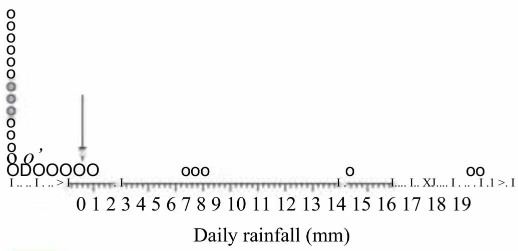
1 a type of mammal

b mean = 9.2, standard deviation = 4.2

c 31.6%   d 5.4 hours

2 a i 17.8 mm 63% ii 0mm 90%

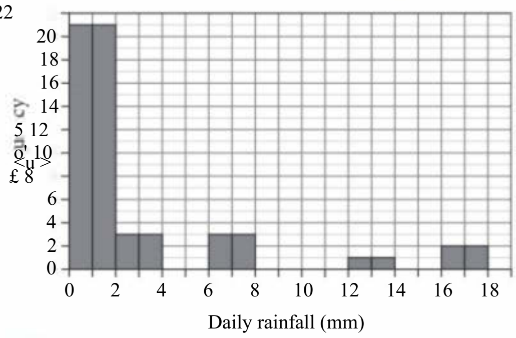
b g




64%

c i 16 days 87%   ii 10% 86%

d 22




56%

3 a i 25.0 years   ii 28.2 years

b 1.1 years

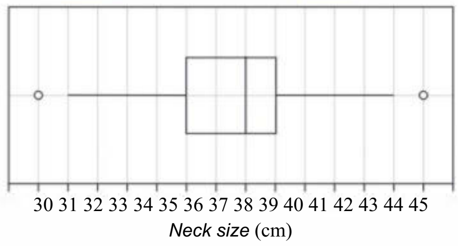
c Qj - 1.5 xIQR = 29.9-1.5x 1.1=28.25

26.0 < 28.25, so the age of 26.0 would be shown on a boxplot as an outlier.

4 a 38 93%

b i 1 40%   ii 1 42%

c




57%

5 a 3 52%   b 84% 60%

c The maximum value is equal to $Q_3$. 47%

d 45.89 metres 30%

# CHAPTER 2

## EXERCISE 2.1

1 a i *average daily screen time* and *ATAR score*

ii Both variables are numerical.

iii The amount of screen time is likely to affect an ATAR score. However, an ATAR score is not likely to affect the amount of screen time. So, *average daily screen time* is the explanatory variable.

b i *stress levels* (scale of 0 to 5 where 0 is none and 5 is extremely high) and *headache levels* (1 = mild, 2 = moderate, 3 = severe)

ii Both variables are categorical.

iii The aim of the research is to see whether headache levels can be predicted from stress levels, so *stress levels* is the explanatory variable.

c i *time taken to complete Fun Run* and *age*

ii Both variables are numerical.

iii A persons' age is likely to affect their time taken to complete a Fun Run. However, the time a person takes to complete a Fun Run is not likely to affect their age. So, *age* is the explanatory variable.

d i *gender* and *hand dominance*

ii Both variables are categorical.

iii A persons' gender is likely to affect their hand dominance. However, a persons' hand dominance is not likely to affect their gender. So, *gender* is the explanatory variable.

e i *driving response times* and *levels of sleep deprivation* (1 = low, 2 = medium, 3 = high)

ii *driving response times* is numerical and *levels of sleep deprivation* (1 = low, 2 = medium, 3 = high) is categorical.

iii The aim of the experiment is to see whether levels of sleep deprivation can explain driving response times, so *levels of sleep deprivation* is the explanatory variable.

f i *English study score* and *number of television sets in the home* ii Both variables are numerical.

iii The number of television sets in the home is likely to affect a students' English study score. However, a students' English study score is not likely to affect the number of television sets in the home. So, *number of television sets in the home* is the explanatory variable.

2D 3 C 4 E

## EXERCISE 2.2

1 E 2 A

3 a no b yes c yes d yes

e no

4 a The explanatory variable is *gender.* The response variable is *number of mobile phones owned.*

b

| Number of mobile phones owned | Gender | | Total |
|---|---|---|---|
| | Male | Female | |
| Only one | 15 | 20 | 35 |
| More than one | 23 | 12 | 35 |
| Total | 38 | 32 | 70 |

5 a

| | Gender | |
|---|---|---|
| Sporting dub membership | Male | Female |
| Sporting club member | 67% | 69% |
| Not a sporting club member | 33% | 31% |
| Total | 100% | 100% |

b The difference between the males and females is only 2%, suggesting that there is no association between gender and sporting club membership.

6 The percentage of flats with only 1 laptop (70%) is considerably higher than the percentage of units with only 1 laptop (55%), and both percentages are considerably higher than the percentage of houses with only 1 laptop (25%). The percentages for 2,3,4, and more than 4 laptops for houses are all considerably greater than for both flats and units. The percentage for 2 laptops for units (35%) is considerably higher than the percentage for 2 laptops for flats (20%). So, this segmented bar chart suggests that there may be an association between type of residence and the number of laptops.

7E 8B 9 B 10 E

11 D 12 D 13 B 14 B

15 a 3 *(city, congestion level, size)*

b 2 *(congestion level, size)*

c Newcastle-Sunderland and Liverpool

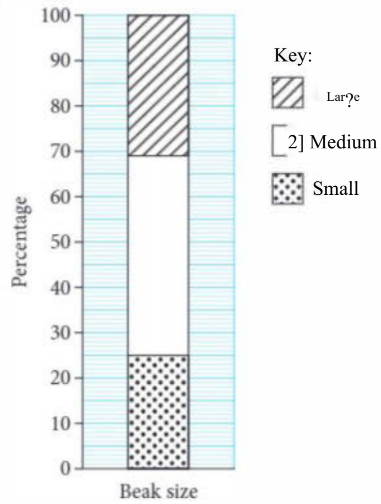
d

| | City size | |
|---|---|---|
| Congestion level | Small | Large |
| high | 4 | 2 |
| medium | 4 | 2 |
| low | 8 | 3 |
| Total | 16 | 7 |

e 25%

16 a 24

b




The three sections can be in any order.

c Yes, the information does provide support for the contention that *beak size* is associated with *sex.* The second mark requires a statement similar to one of the following:

- 50% of males had large beaks, which was higher than females, with 7%.
- 48% of males had medium beaks, which was higher than females, with 38%.
- 2% of males had small beaks, which was lower than females, with 55%.

## EXERCISE 2.3

1 C 2 A

3 a i male range = 52, female range = 43

ii male median = 26, female median = 58

iii male IQR = 22.5, female IQR = 10.5

b i positively skewed

ii approximately symmetrical

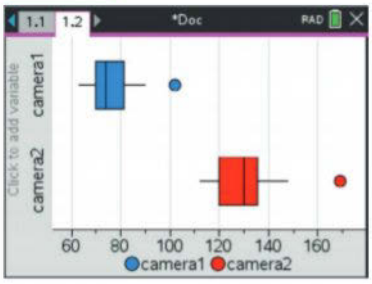
c Yes. The male range (52 years) is considerably more than the female range (43 years). The male median (26 years) is considerably less than the female median (58 years). The male IQR (22.5 years) is considerably more than the female IQR (10.5 years). The male distribution is positively skewed while the female distribution is approximately symmetric.

4 a Year 10 is negatively skewed. Year 11 is positively skewed. Year 12 is neither skewed nor symmetric,

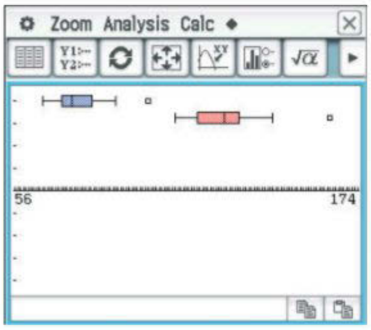
b Year 10: median = 8, range = 7, IQR = 3;
Year 11: median = 6, range = 5, IQR = 3;
Year 12: median = 5, range = 6, IQR = 4.5

c Yes, the dot plots support the contention that number of hours spent binge-watching series during a school week is associated with the school year level. The Year 10 median (8 hours) is noticeably more than the Year 11 median (6 hours) and even more than the Year 12 median (5 hours).

5 TI-NspireJ




ClassPad




6 a The small dog data is negatively skewed. The medium dog data is neither symmetric nor skewed. The large dog data is approximately symmetric,

b small dogs: median =15, range = 17, IQR = 5; medium dogs: median = 10, range =14, IQR = 8; large dogs: median = 5, range = 10, IQR = 2

c Yes, the boxplots support the contention that the lifespans of dogs is associated with their sizes. The shapes of the boxplots are different. The small dog data is negatively skewed, the medium dog data is neither symmetric nor skewed, and the large dog data is symmetric. The small dog range (17 years) is considerably higher than the medium dog range (14 years), which is considerably higher again than the large dog range (10 years).

7E 8 C 9 E 10 E

11 a The shapes of the distributions suggest there is an association because they are all different. The 55-inch screen televisions have a negatively skewed distribution, the 65-inch screen televisions have a distribution that is neither symmetric nor skewed, and the 75-inch screen televisions have a positively skewed distribution.

b The median of the distributions suggest there is an association. The median of the 55-inch screen televisions (8 years) is noticeably more than the median of the 75-inch screen televisions (6 years), which is noticeably more than the median of the 65-inch screen televisions (5 years).

12 a i 5 ii 10 iii 1

b 15

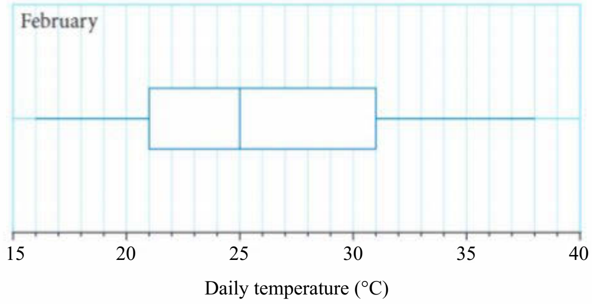
13 a 20% b2.6 c 23

d Yes, because the median BMI changes considerably as *neck size* increases. The median for below average *neck size* was 21.6, which increased to a median of 24.6 for average *neck size,* and increased again to a median of 28.1 for above average *neck size.*

or

Yes, because the IQR changes considerably as *neck size* increases. The IQR for below average *neck size* and average *neck size* was 2.6, which increased to an IQR of 3.7 for above average *neck size.*

14 a i




ii 75%

b i July: positively skewed with an outlier; May: symmetric

ii 15.5°C

iii The medians for the two months differ considerably. In May, the median maximum temperature is about 14.5°C, while in July, the median maximum temperature is about 9°C. (Comparing the two interquartile range (IQR) values is also correct.)

15 a *place of capture* b 20 mm c 16 and 36

d Upper fence = 32 + 1.5 x 12 = 50 and 52 > 50, so 52 mm is an outlier.

e The medians for forest and grassland differ considerably. The forest median is 21 and the grassland median is 30.

16 a 124,148

b The median *arm span* increases with Year level,

c 1.5 x IQR =1.5x10 =15

The lower fence is at Qj - 15 = 160 - 15 = 145.

An actual *arm span* of 140 is lower than this and so is still an outlier.

## EXERCISE 2.4

1 C 2 C

3 a *age* (years) b time (min) c 9 people

d A 50-year-old who was in the cafe for 20 minutes,

e 3 f They were a toddler with his/her parent/caretaker.

4 a i negative, linear and strong

ii The total nap time decreases as a child ages,

b i no association

ii There appears to be no association between the number of bad apples per tonne and the price per kilo.

c i negative, linear and weak

ii There is some indication that the time spent reading to a child under 8 decreases as the child gets older.

d i positive, linear and moderate

ii Weight tends to increase as height increases.

5 a positive association

b no association

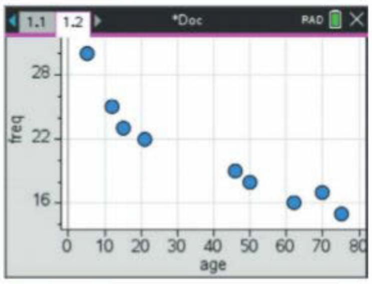
c negative association

d positive association

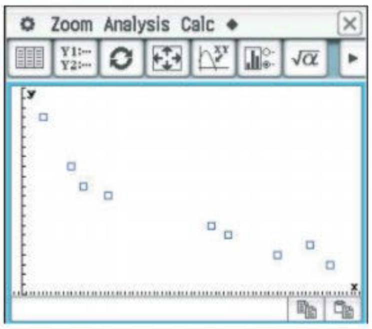
e positive association

f negative association

6 a TI-Nspire

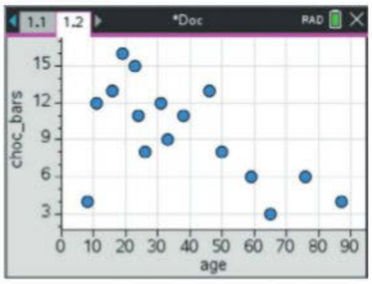



ClassPad

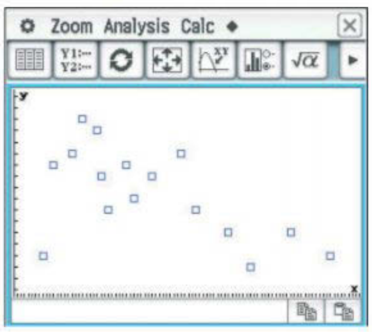



The association between age and frequency can be described as negative, linear and strong.

b TI-Nspire




ClassPad




The association between age and chocolate consumption can be described as negative, linear and moderate.

7 a i scatterplot

ii Both *number of people* and *money spent* are numerical.

b i parallel percentaged segmented bar charts

ii Both *make of car* and *type of pet* are categorical,

c i parallel dot plots and parallel boxplots

ii *Make of car* is categorical with more than two categories and *amount earned* is numerical,

d i back-to-back stem plot, parallel dot plots and parallel boxplots

ii *Gender* (boy, girl) is categorical with two categories and *number of pets* is numerical,

e i parallel boxplots

ii *Types physical activities* is categorical with more than two categories and *amount of time spent watching sport on television* is numerical, and a quick comparison of medians is needed.

8 a i False. A scatterplot can only be used if both variables are numerical.

ii True

iii True

iv False. *Make of car* is categorical and both variables for a scatterplot need to be numerical.

b i False. Scatterplots require two numerical variables and parallel percentaged segmented bar chart require two categorical variables.

ii True

iii False. A back-to-back stem plot can only show two categories.

9 D 10 E 11 D 12 D

13 A

## EXERCISE 2.5

1 E 2 C

3 a The data suggests there is a weak positive linear association between *amount of exercise* and *height.*

b The data suggests there is a moderate positive linear association between *weight* and *time spent sitting down.*

c The data suggests there is a strong negative linear association between *percentage of good peaches* and *time spent in storage.*

d The data suggests there is a moderate positive linear association between *sales* and *temperature.*

e The data suggests there is no association between *number of pets* and *temperature.*

f The data suggests there is a weak negative linear association between *number of home cooked meals* and *income.*

4 a The three assumptions are that both variables are numerical, the association is linear, and there are no outliers.

b An r value of 0.6484 indicates that there is a moderate, positive, linear association between *income* and the *number of cinema movies seen in the past year.*

5 a 0.6 b -0.4 c 0.8 d -0.9

e -0.5 f 0.2

6 a *outdoor temperature* b *population*

c *household income* d *age*

7 A 8 E 9 D 10 D

11 A 12 D 13 C

14 a r= 0.8054 b strong positive

c linear d *cost*

15 a *length of stride* and *running speed*

b Observation. The first researcher is passively observing an existing situation.

c Experimentation. The second researcher is actively manipulating a situation to eliminate other variables *before* observing it.

d The second researchers study has other variables by controlling the factors that can be controlled, randomisation, and repeating the experiment,

e The first researchers' results may only apply specifically to Olympic level runners who, unlike the average person, would have worked hard on stride technique over years. The results could also be influenced by other variables such as training techniques and how wind moves in large stadiums. Other answers are possible. Teacher to check.

16 a Head lice caused people to be healthy.

b observation

c The sickness caused lice to leave.

d The people of the Hebrides had their explanatory and response variables around the wrong way in the association. They believed that the explanatory variable was the *head lice* and the response variable was the *fever,* when in fact the explanatory variable was the *fever* and the response variable was the *head lice.*

e Select a group of people without lice who are sick with fever and randomly introduce lice to the hair of half of them. Monitor the fevers of both groups and record how long it takes for them to recover. It would be unethical to introduce an additional problem to people who are already sick to test such a doubtful proposition. Alternatively, select a group of people with head lice and introduce a sickness that causes fever to half of them. This is also unethical for the same reasons.

## CUMULATIVE EXAMINATION 1

| | | |
|---|---|---|
| 1 D | 2 B 87% | 3 E 93% |
| 4 A 65% | 5 A 71% | 6 A 57% |
| 7 E 95% | 8 B 70% | 9 D 75% |
| 10 B 74% | 11 E 57% | |

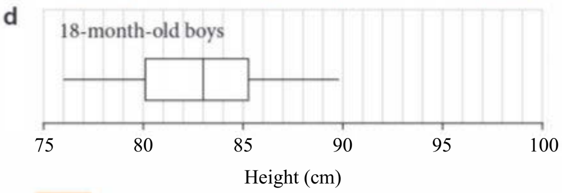
## CUMULATIVE EXAMINATION 2

1 The percentaged segmented bar chart does support the opinion that lunch time activity (walked, sat or stood, ran) is associated with Year level. For example, the percentage who ran changed from around 78% to 40% to 10% from Years 6-8 and 8-10. (The second mark is for quoting relevant percentages.) 77%

2 a 1.81 87% b 0.5 78%

3 a 3.8 62% b −1.4 62% c 97.5% 58%

d




60%

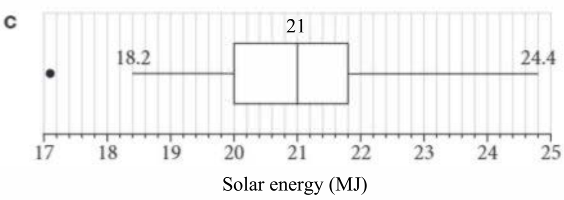
e 89.5 58%

f The median height increases with age. 58%

4 a 3 years

b lower fence = Qj - 1.5 x IQR = 20 - 1.5 x 1.8 = 17.3
17.1 < 17.3, so 17.1 is an outlier.

c




d The median amount of solar energy collected differs from month to month as indicated for April (11 MJ), May (7 MJ) and June (5.8 MJ).

# CHAPTER 3

## EXERCISE 3.1

1 a i explanatory variable: *temperature,*
response variable: *volume of cicadas*

ii $\bar{x}$ =25.3, $s_x$= 5.6, $\bar{y}$ = 83, $s_y$ = 7, r = 0.79

iii *volume of cicadas* = 58 + 0.99 x *temperature*

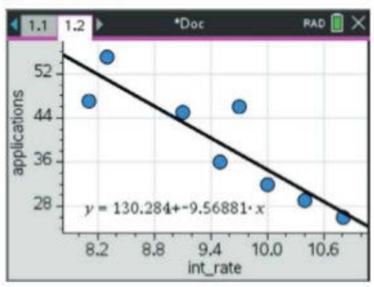
b i explanatory variable: *annual rainfall,*
response variable: *number of bushfires*

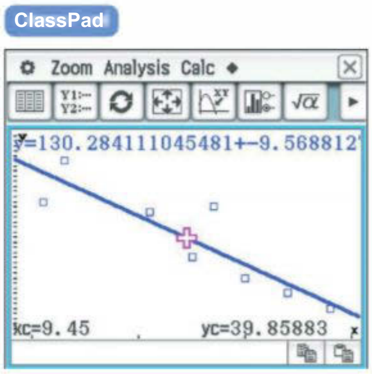
ii $\bar{x}$ = 19.88, $s_x$ = 6.31, $\bar{y}$ = 18.38, $s_y$ = 8.16, r=-0.88

iii *number of bushfires* = 41 - 1.1 x *annual rainfall*

2 a *number of applications* = 130 - 9.6 x *interest rate*

b TI-Nspire






3 a 40 000. The price of the car model was $40 000 when it was released.

b -3890. The price of the car model on average decreases by $3890 for every one-year increase in age.

c *price* = 40 000 - 3890 x *age*

4 a The slope is 0.7. This means the weight on average increases by 0.7 kg for every 1-cm increase in height,

b The y-intercept is -44. This means that weight is -44 kg when height is 0 cm. Negative weight is clearly impossible. This least squares line of best fit only applies from a certain minimum height.

5 A 6 C 7A 8 B

9 B 10 D 11 E

12 a *humidity 9 am*

b *humidity 3pm* = -1.26 + 0.765 x *humidity 9am*

c $r$ = 0.871

13 a strong, positive, linear

b *male* = 9.69 + 0.81 x *female*

14 a *height*

b *arm span* = -16 + 1.1 x *height*

c *Arm span* increases by 1.1 cm for each 1-cm increase in *height.*

## EXERCISE 3.2

1 A 2 A

3 a r2^ 0.736.

b 73.6% of the variation in *shoe size* can be explained by the variation in *height.*

c 26%

d *shoe size* = -20 + 0.17 x *height*

e Yes, this is an appropriate model because of the high $r^2$ value of 73.6%.

4 a r=0.95 b r=-0.79

c $r$ = 0.94 d r=-0.56

5 E 6 A 7 A 8 C

9 B 10 A 11 A

12 a i 24 ii 1.065

b i *weight* ii -0.00112

c 29%

13 a On average, *male* life expectancy increased by 0.88 years for each one-year increase in *female* life expectancy.

b 34.4 years

c 95% of the variation in *male* life expectancy can be explained by the variation in *female* life expectancy.

14 a 0.53

b i 56.9%

ii 56.9% of the variation in *height* is explained by the variation in *age.*

15 a *r*= -0.3755, so the association between *population density* and *area* is weak, negative and linear,

b i z = -0.8 ii 1 suburb iii 2 suburbs

## EXERCISE 3.3

1 A 2 C

3 a 8.95 laps b 5.43 laps c 2.07 laps

4 a 2°C b 8°C c 18°C

5 a 145 grams, interpolation

b 45 grams, extrapolation

c 370 grams, extrapolation

d The prediction for 25 chocolates is the most reliable because it is within the data range of 15 to 50; the other two predictions involve extrapolation,

e 21 chocolates

6 B 7 D 8 C

9 a 3 years (one mark for calculating Australia life expectancy = 87.67... years and UK life expectancy = 84.62... years)

b The least squares lines of best fit equations were used to make predictions outside the available range of data.

10 a *male income* b $350

c i $18250

ii Making the required prediction involved going beyond the data used (extrapolation) to determine the line of best fit equation.

## EXERCISE 3.4

1 C 2 C

3 a i 0% ii 20%

iiii -5% iv -35%

b i 0.5 metres ii -0.7 metres

4 a Non-linear because it appears to be a valley shape,

b Linear because it appears to be randomly scattered,

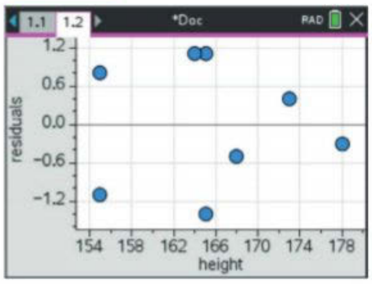
c Non-linear because it appears to be a hill shape,

d Non-linear because it does not appear randomly scattered.

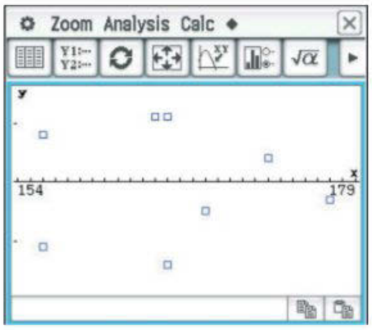
5

| Height (cm) | Femur length (cm) | Residual |
|---|---|---|
| 178 | 50.2 | -0.3 |
| 173 | 48.4 | 0.4 |
| 165 | 45.1 | 1.1 |
| 164 | 44.6 | 1.1 |
| 168 | 45 | -0.5 |
| 165 | 42.6 | -1.4 |
| 155 | 39.9 | 0.8 |
| 155 | 38 | -1.1 |

**TI-Nspire**




**ClassPad**




The data is probably linear because the residual values appear randomly scattered above and below the x-axis.

6 The actual value at $x = 11$ is below the least squares line of best fit, which means the residual must be negative. The residual plot has a positive residual value for $x = 11$.

7 B 8 A 9 B 10 B

11 B

12 a The pressure at 3 pm on average increases by 0.8894hPa for every 1-hPa increase in the pressure at 9 am.

b 1023 hPa c interpolation d 3hPa

e i From the equation for the line of best fit, the

$$\text{slope } b = r\frac{s_y}{s_x}, \text{ so } 0.8894 = r \times \frac{4.1884}{4.5477}.$$

Solving for $r$ gives $r = 0.8894 \div \frac{4.1884}{4.5477} = 0.966$ correct to 3 decimal places.

ii 93.3%

f i The assumption is that there is a linear association between the atmospheric pressure at 3 pm and the atmospheric pressure at 9 am.

ii The residual plot has a clear pattern shows (a valley shape) which suggests that the association is non-linear.

13 a strong, linear, positive

b i *apparent temperature* = -1.7 + 0.94 x *actual temperature*

ii On average, when the actual temperature is 0°C, the apparent temperature is -1.7°C. (Alternative answer: When the actual temperature is 0°C, the predicted apparent temperature is -1.7 °C.)

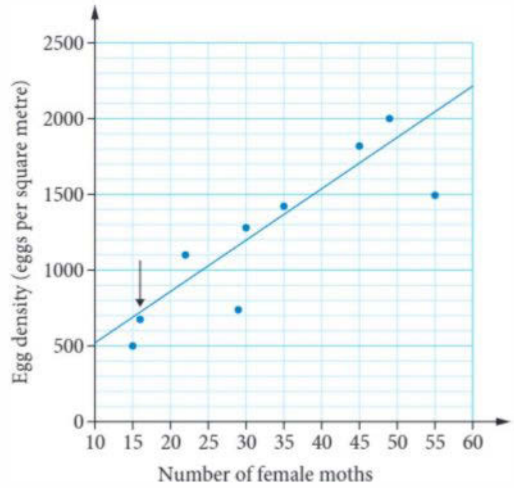
c 97% of the variation in *apparent temperature* can be explained by the variation in *actual temperature*.

d i That the association is linear.

ii Yes, since there is no clear pattern in the residual plot.

14 a egg *density* = -46.8 + 18.9 x *number of male moths*

b i




i i On average, egg density increases by 31.3 eggs/m² for each additional female moth caught.

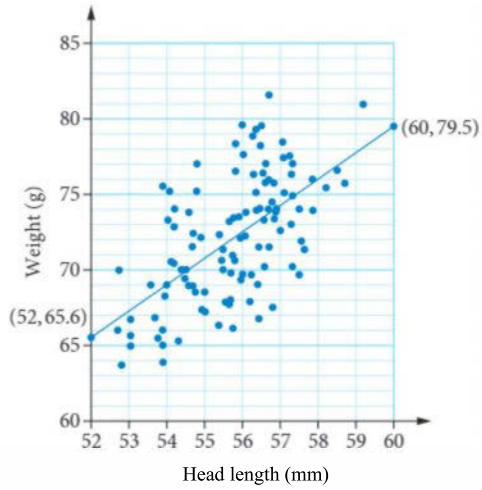
Examples of other correct answers:

- egg density increases by 31.3 as the number of female moths increases by 1
- egg density increases by 31.3 for an increase of 1 in the number of female moths caught
- egg density increases by 31.3 for each additional female moth caught.

iii -412.5

iv 74.3%

15 a




b 60.4 g

c Extrapolation, because 49.0 is outside the data range,

d 80.9 g (one mark for correctly using the predicted value with the residual of 2.78.)

e 64.5%

f There is a linear association.

## EXERCISE 3.5

1 D 2 E

3 a j^ory² b none

c $\frac{1}{x}$, log x, $\frac{1}{y}$ or log y d $x^2$, $\frac{1}{y}$ or log y

e none f $\frac{1}{x}$, log x or $y^2$


9780170448406

4 a *distance = a + bx (time)*$^{1\ 2\ 3}$,

$\frac{1}{distance}$ = *a + bx time,*

log *(distance) = a + b* x *time*

b *weight = a + bx (length)*$^2$,
*(weight)*$^2$ *= a + bx length*

c *score = a + bx* $\frac{1}{study\ time}$,

*score = a +* b x log *(study time),*
*(score)*$^2$ *= a + bx study time*

d *afternoon rainfall = a + bx* $\frac{1}{morning\ rainfall}$,

*afternoon rainfall = a + bx log(morning rainfall),*

$\frac{1}{afternoon\ rainfall}$ *a + bx morning rainfall,*

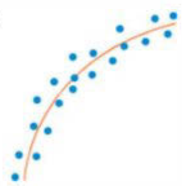
log *(afternoon rainfall) = a + bx morning rainfall*

5 a i 10 seeds for both equations

ii 8 seeds for both equations

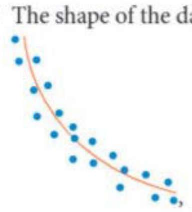
iii 7 seeds for the squared transformation,
6 seeds for the log transformation

b



c log transformation because its coefficient of determination is closer to 1

6 a The shape of the data is



so the options to linearise the data are

$\frac{1}{x}$, logx, $\frac{1}{y}$ or logy.

b For the $\frac{1}{x}$ transformation

| $\frac{1}{x}$ | I | 0.5 | 0.333 | 0.25 | 0.2 |
|---|---|---|---|---|---|
| y | 28 | 18 | 6 | 3 | 2 |

For the logx transformation

| logx | 0 | 0.301 | 0.477 | 0.602 | 0.699 |
|---|---|---|---|---|---|
| y | 28 | 18 | 6 | 3 | 2 |

For the $\frac{1}{y}$ transformation

| y | i | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $\frac{1}{y}$ | 0.036 | 0.056 | 0.167 | 0.333 | 0.5 |

For the logy transformation

| x | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| log7 | 1.447 | 1.255 | 0.778 | 0.477 | 0.301 |

c For the $\frac{1}{x}$ transformation, $r^2$ = 0.93.

For the logx transformation, $r^2$ = 0.97.

For the $\frac{1}{y}$ transformation, $r^2$ = 0.94.

For the logy transformation, $r^2$ = 0.98.

d The coefficient of determination of the logy transformation is closest to 1, so logy is the best choice for linearising the data.

e logy = 1.77 - 0.3 lx

7 C 8 A 9A 10 B

11 E 12 B

13 a *population* = 7700 + 7700 log *(area)*

b 23000

14 a 3.4,6.6 b 13 km/h

15 a 2.39,5.89 b 27.7 years

## CUMULATIVE EXAMINATION 1

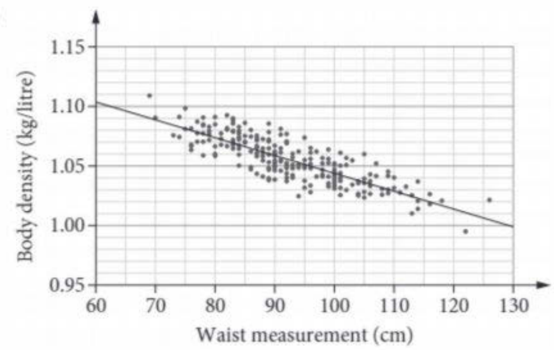
| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | A | 91% | 2 | B | 82% | 3 | B | 46% |
| 4 | D | 68% | 5 | E | 58% | 6 | C | 62% |
| 7 | B | 54% | 8 | B | 43% | 9 | A | 52% |
| 10 | A | 58% | 11 | D | 40% | | | |

## CUMULATIVE EXAMINATION 2

1 a 19% b 29440000

c The percentages of 15-64 year olds in each of the three countries are similar: Australia has 33%, India has 36% and Japan has 36%.

2 a




41%

b 1.10 73%

c extrapolating 38%

d On average, *body density* decreases by 0.001512 kg/litre for each 1-cm increase in *waist measurement.*

44%

e residual = 0.995 - (1.195 - 0.001512 x 122)
= 0.995- 1.010536
= -0.015 536, which rounds to -0.02 51 %

f -0.824 25%

g Yes. Residuals have no clear pattern (or are randomly scattered). 159%

3 a 0.0393,5.28 52% b 38% 31%

4 a 10.4cm 87%

b strong negative 51%

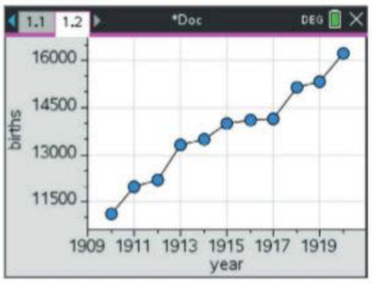
c *mean height* = 171 + -0.169 x *mean age* 23%

d *mean height* = 167.9 - 0.001621 x *{mean age)*$^2$ 30%

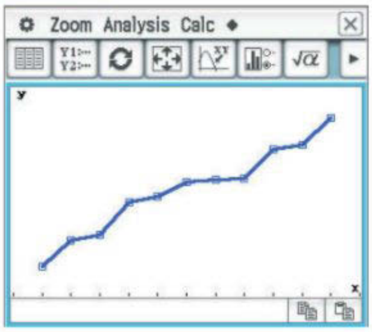
# CHAPTER 4

## EXERCISE 4.1

1 TI-Nspire




ClassPad




2 3200

3 a Likely to show seasonality because more would be bought in winter when there is more rain,

b Likely to show seasonality because more would be bought in summer.

c Likely to show seasonality because more people would have the time to shop for them on weekends rather than weekdays.

d Not likely to show seasonality because milk is consumed regularly every day.

e Likely to show seasonality because the occupancy rates would be higher during the weeks of the holiday periods.

f Not likely to show seasonality because salaries are paid regularly throughout the year.

g Likely to show seasonality because fruit is harvested in certain months.

h Likely to show seasonality because AFL matches are only played during certain months of the year.

4 a decreasing trend and irregular fluctuations

b decreasing trend, seasonality and irregular fluctuations

c increasing trend and irregular fluctuations

d increasing variability and irregular fluctuations

e outlier, structural change and irregular fluctuations

5 E 6 C 7 A

8 D 9 A 10 A

11 The average age at first marriage showed no trend from 1915 to around 1935, and then there was a decreasing trend during the period 1935 to 1970.

## EXERCISE 4.2

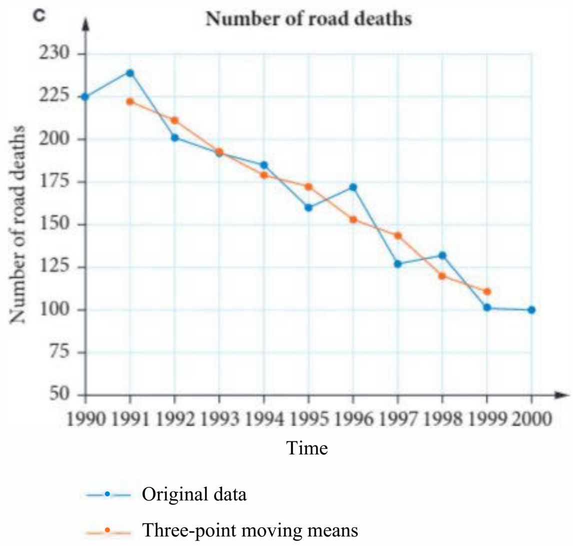
1 C 2 E

3 a

| Year | Number of road deaths | Three-point moving means |
|---|---|---|
| 1990 | 225 | |
| 1991 | 240 | $\frac{225+240+201}{3} = 222$ |
| 1992 | 201 | $\frac{240+201+192}{3} = 211$ |
| 1993 | 192 | $\frac{201+192+185}{3}$ = 192 67 |
| 1994 | 185 | $\frac{192+185+160}{3}$ r9 = |
| 1995 | 160 | $\frac{185+160+172}{3}$ = 172 33 |
| 1996 | 172 | $\frac{160+172+127}{3} =$ 153 |
| 1997 | 127 | $\frac{172+127+132}{3} = 143.67$ |
| 1998 | 132 | $\frac{127+132+101}{3}$ po = |
| 1999 | 101 | $\frac{132+101+100}{3}$ in = |
| 2000 | 100 | |

b The smoothed number of road deaths in 1994 is 179. There was not enough data to calculate the smoothed number of road deaths in 2000.

c




d The graph of the smoothed data indicates a decreasing trend.



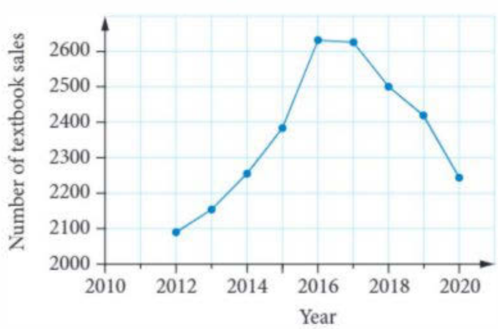
9780170448406

**4 a**

| Year | Number of textbook sales | Four-point moving means | Four-point moving means with centring |
|---|---|---|---|
| 2010 | 2250 | | |
| 2011 | 2230 | | |
| | | $\frac{2250 + 2230 + 2000 + 2010}{4} = 2122.5$ | |
| 2012 | 2000 | | 2090 |
| | | $\frac{2230 + 2000 + 2010 + 1980}{4} = 2057.5$ | |
| 2013 | 2010 | | 2153.75 |
| | | $\frac{2000 + 2010 + 1980 + 3000}{4} =$ 225[illegible] | |
| 2014 | 1990 | | 2255.625 |
| | | $\frac{2010 + 1980 + 3000 + 2045}{4} = 2261.25$ | |
| 2015 | 3000 | | 2383.625 |
| | | $\frac{1980 + 3000 + 2045 + 2989}{4} =$ [illegible] | |
| 2016 | 2045 | | 2632.25 |
| | | $\frac{3000 + 2045 + 2989 + 3000}{4} = 2758.5$ | |
| 2017 | 2989 | | 2627.25 |
| | | $\frac{2045 + 2989 + 3000 + 1950}{4}$ 2496 | |
| 2018 | 3000 | | 2505.375 |
| | | $\frac{2989 + 3000 + 1950 + 2120}{4} = 2514.75$ | |
| 2019 | 1950 | | 2423 |
| | | $\frac{3000 + 1950 + 2120 + 2255}{4} =$ 2331.[illegible] | |
| 2020 | 2120 | | 2243.375 |
| | | $\frac{1950 + 2120 + 2255 + 2297}{4} = 2155.5$ | |
| 2021 | 2255 | | |
| 2022 | 2297 | | |

**b** There is not enough data to calculate the smoothed number of sales for 2011. The smoothed number of sales for 2019 is 2423.

**c** Number of sales (smoothed data)




**d** The graph of the smoothed data indicates an increasing trend until 2016 and a decreasing trend from 2017 to 2022.

**e** The smoothed data value for 2011 is 2177.5 and for 2012 is 2060.

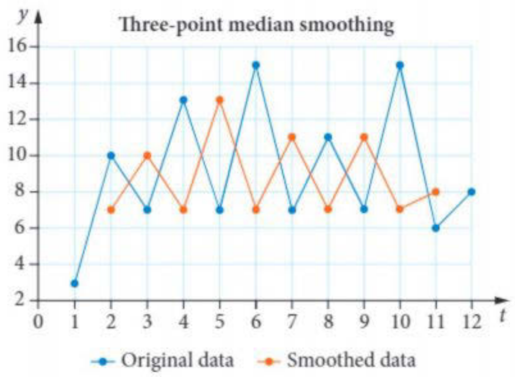
| | | | |
|---|---|---|---|
| 5 B | 6 C | 7 C | 8 C |
| 9 D | 10 D | 11 D | 12 B |
| 13 D | | | |
| 14 a 8 | b 9 | c 9.4 | d 2 |

ANSWERS

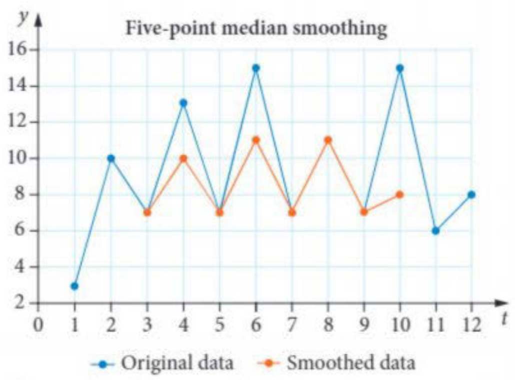
## EXERCISE 4.3

1 C 2D

3 a




b

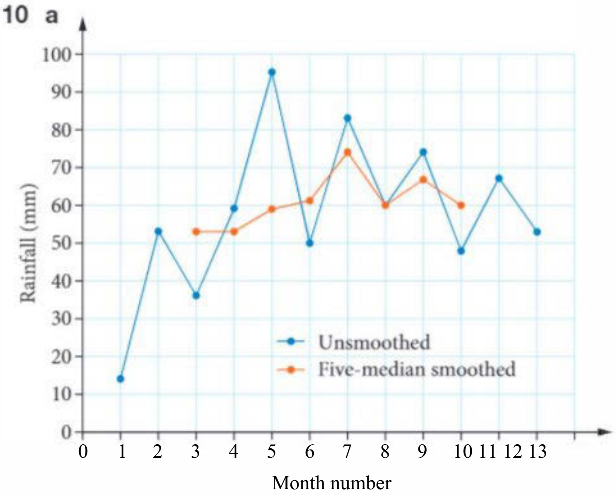



c The graphs of the smoothed data indicate no trend.

4 D 5 E 6 D

7 B 8 B 9 D

10 a




b From the table:

The mean of September and October $= \frac{124 + 140}{2}$

$= 132.$

The mean of October and November = ^4QJL225 / 2

$= 182.5.$

So the two-mean smoothed rainfall centred on

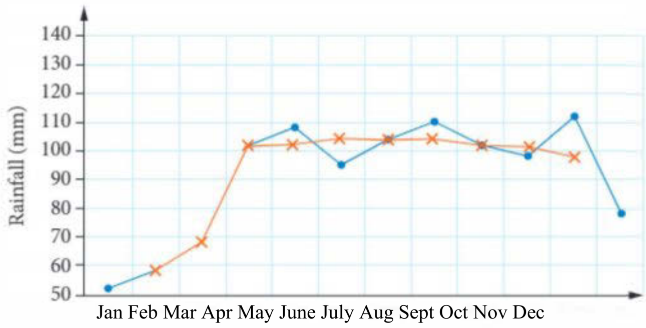
October $= \frac{132 + 182.5}{2} = 157.25$ mm.

11 a 20°C b 20°C

c


12 a November

b




c In the smoothed time series, there are two key trends. Until April, there is an increase in monthly rainfall. It then remains relatively constant for the remainder of the year.

## EXERCISE 4.4

1 E 2 B

3 a 7 months b October c January

d 12 e 1

f

| | |
|---|---|
| Jan | 250% |
| Feb | 201% |
| Mar | 160% |
| Apr | 70% |
| May | 20% |
| Jun | 20% |
| Jul | 10% |
| Aug | 0% |
| Sep | 60% |
| Oct | 90% |
| Nov | 120% |
| Dec | 190% |

9 100% h 150%

4 a

| QI | Q2 | Q3 | Q4 |
|---|---|---|---|
| 1.91 | 1.39 | 0.17 | 0.52 |

b

| Ql | Q2 | Q3 | Q4 |
|---|---|---|---|
| 191% | 139% | 17% | 52% |

c Q2

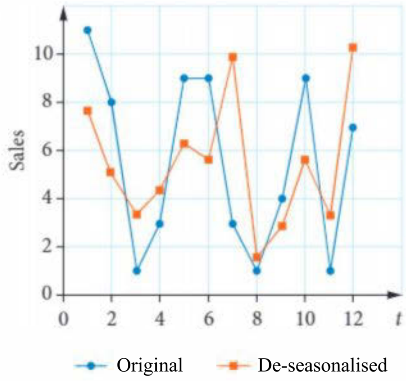
5 a

| Year | Ql | Q2 | Q3 | Q4 |
|---|---|---|---|---|
| Seasonal index | 1.437 | 1.581 | 0.303 | 0.679 |

b

| Year | Q1 | Q2 | Q3 | Q4 |
|---|---|---|---|---|
| 2023 | 7.654 | 5.061 | 3.297 | 4.418 |
| 2024 | 6.263 | 5.694 | 9.892 | 1.473 |
| 2025 | 2.783 | 5.694 | 3.297 | 10.310 |

c




6 a 0.7 is the missing seasonal index.

b 43% c $70000

7 B 8 A 9 E

10 D 11 E 12 C

13 D 14 D 15 A

16 a 1.10 b 241 mm

c The autumn rainfall is 5% above the average for the four seasons of the year.

17 a

| | Summer | Autumn | Winter | Spring 1 |
|---|---|---|---|---|
| Seasonal index | 0.89 | 1.00 | 1.41 | 0.70 |

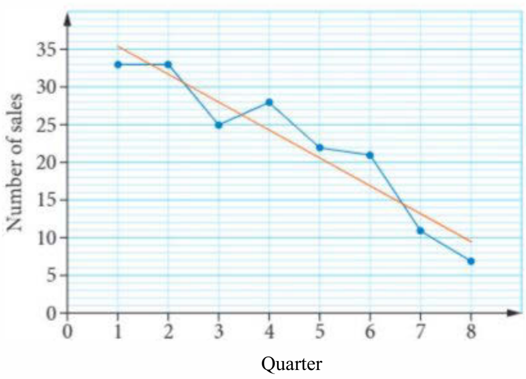
b 186 mm

## EXERCISE 4.5

1 E 2 E

3 a *de-seasonalised number of sales* = 39 - 3.7 x *quarter number*

b De-seasonalised number of sales 2022-2023




During 2022-2023 the sales of the costume decreased on average by 3.7 per quarter.

c 6 costumes

d 5 costumes

4 a *sales* = 211 - 0.127 x *month*

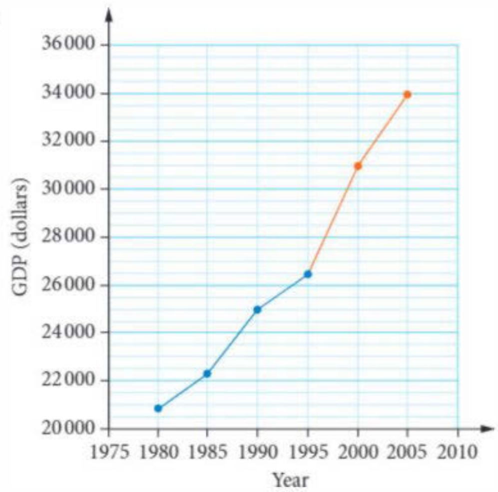
b $207 C$178

5 B 6 E 7 E 8 D

9 C 10 A 11 E 12 C

13 D

14 a




b increasing trend

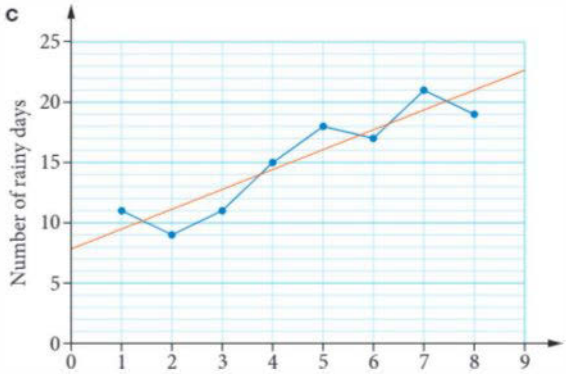
c *GDP* =20000 + 524 x *time*

d $752

15 a increasing trend

b *number of rainy days* = 7.79 + 1.63 x *month number*

c




## CUMULATIVE EXAMINATION 1

1 B 83% 2 B 31% 3 D 71%

4 B 5 E 6 B

7 D 34% 8 C 60% 9 A 62%

10 D 59% 11 D 47%

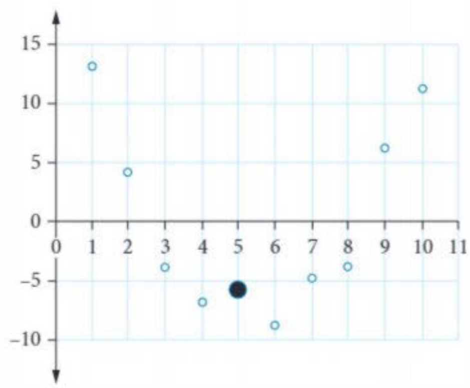
## CUMULATIVE EXAMINATION 2

1 a *year*

b i $10990 ii $185750

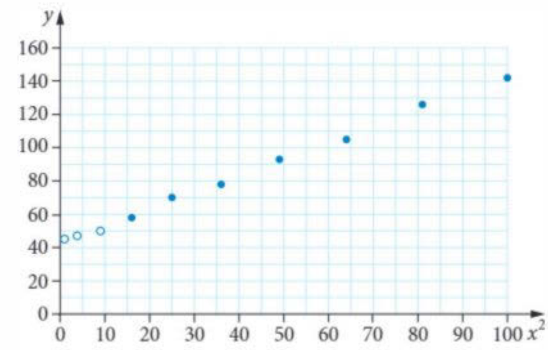
c i −5.85

ii



d i \yeafyx?) | 3 | 4 | 9 | H | 25 | 36 | 49 | 64 | 81 | 100

ii



iii $y = 42.89 + 1.00x^2$

iv $267890

v There is no pattern in this new residual plot, the points appear to be randomly distributed with respect to the horizontal axis.

2 110

## CHAPTER 5

### EXERCISE 5.1

1 a i Start with 3. Multiply each value by 2 and then add 1 to get the next value.

ii 3,7,15

$w_0 = 3$ $u_1 = 2u_0 + 1$ $= 2 \times 3 + 1$ $= 7$ $u_2 = 2U] + 1$ $= 2 \times 7 + 1$ $= 15$

iii «4 = 63

b i Start with 12. Subtract 10 from each value to get the next value.

ii 12, 2, -8

«o - 12 Uj = $u_2$- 10 = 12-10 = 2 $u_2 = u_r$- 10 = 2-10 =-8

iii $«_4$ = -28

2 a 15,13,11,9,7,5

b 2,8,32,128,512,2048

C -8,-1,6,13,20,27

d 64,-32,16,-8,4,-2

e 50,45,40,35,30,25

f 3,7,15,31,63,127

g 1,-2,13,-62,313,-1562

h 40,22,13,8.5,6.25,5.125

3 a $u_0 = 120$, $u_{n+1} = 0.7u_n + 30$

b $u_0$ - 2, $u_{n+1}$ - $5u_n$

c $u_0$ = -6,$u_{n+1}$ = $u_n$+3

d $u_o$-6O,$u_{n+1}$-M,,-4

e $«_0$ = 40, $u_{n+1}$ = 0.5$u_n$

f $w_0$-l,$u_{n+1}$-10«,,+2

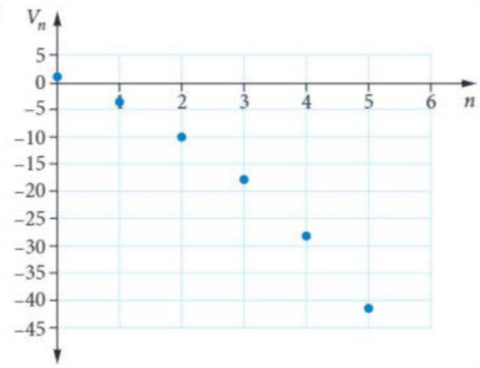
4 a True. Addition is involved. No multiplication is involved. So the graph consists of points in an increasing straight line.

b False. No addition or subtraction is involved. $V_n$ is multiplied by a number between 0 and 1. So the graph consists of points in a decreasing curve, which never reaches zero.

c True. No addition or subtraction is involved. $V_n$ is multiplied by a number greater than 1. So the graph consists of points in an increasing curve,

d False. The graph consists of points in an increasing straight line so it's showing linear growth and not geometric growth.

5



6 D 7 D 8 E 9 A

10 D 11 C 12 E 13 D

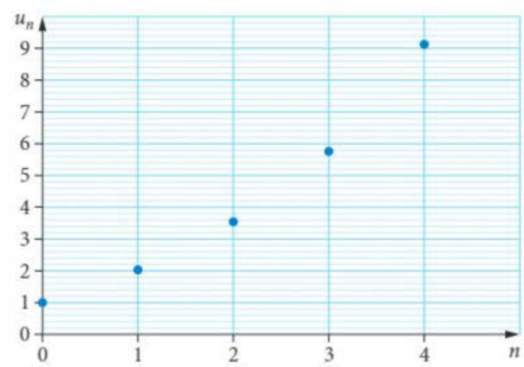
14 C

15 a Start with L Multiply each value by 1.5 and then add 0.5 to get the next value.

b $u_0$=l,$u_{n+1}$ = 1.5$u_n$ + 0.5

c $u_4$ = 9.125

d 1, 2, 3.5, 5.75, 9.125



### EXERCISE 5.2

1 E 2 C

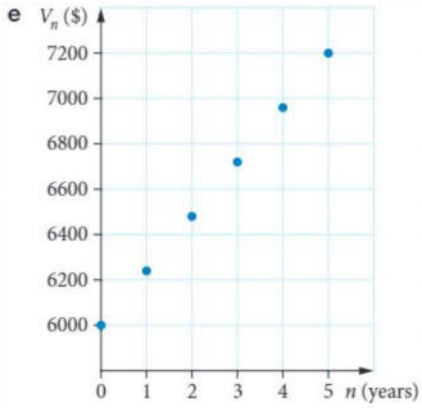
3 a $240

b i $6720 ii 5 years iii $1200

| $n$ | Account balance after $n$ years ($) |
|---|---|
| 0 | 6000 |
| 1 | 6000 + 240 = 6240 |
| 2 | 6240 + 240 = 6480 |
| 3 | 6480 + 240 = 6720 |
| 4 | 6720 + 240 = 6960 |
| 5 | 6960 + 240 = 7200 |

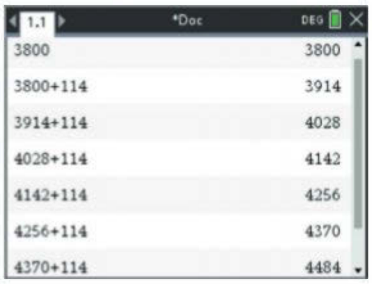
c The recurrence relation is $V_Q = 6000$,
$V_{n+}i = V_{,,+} 240$

d linear growth

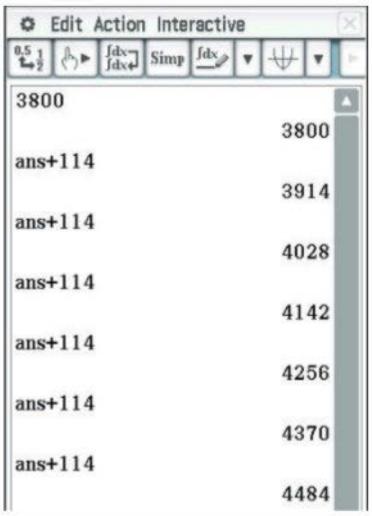
e



4 a $4484

TI-Nspire

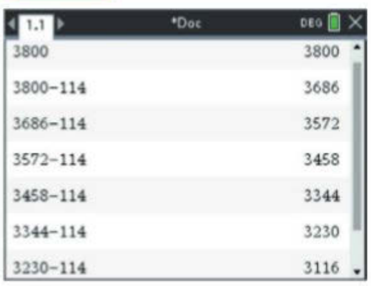



ClassPad

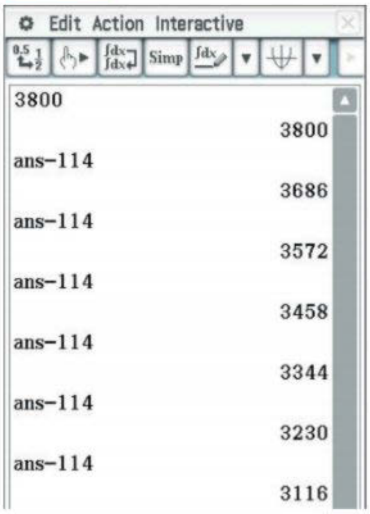



b $3116

TI-Nspire




ClassPad




5 a i d=$400 ii $V_n = 10000+400n$
iiii $14000 iv 25 years
b i <1 - $425 ii $V_n = 8500 + 425n$
iiii $12750 iv 20 years
c i d - $50 ii $V_{,,} = 2000+50n$
iiii $2500 iv 40 years

6 a i $V_o$-8OOO,$V_{,,+1}$- $V_{,,}$ + 1120
ii $V_n = 8000 + 1120n$ iii $16960
b i $V_o = 7500$, $V_{n+1} = V_{,,} + 375$
ii $V_{,,} = 7500 + 375n$ iii $10500
c ■ $V_o$ = 92OO,$V_{,,+1} = V_{,,} + 368$
ii $V_n = 9200 + 368n$ iii $12144
d i $V_o$ = 6OOO,$V_{M+1} = V_n + 720$
ii $V_n = 6000 + 720n$ iii $11760

7 C 8 D 9 B 10 B

11 B

12 a i $2000 ii $2600
b i $3200 ii $3800 iii Year 9

## EXERCISE 5.3

1 E 2D

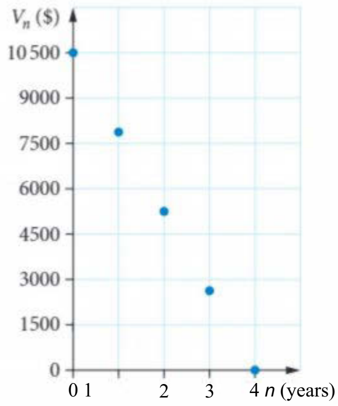
3 a $2625
b i $5250 ii after 3 years iii after 4 years

| $n$ | Value after $n$ years ($) |
|---|---|
| 0 | 10 500 |
| 1 | 10 500 – 2625 = 7875 |
| 2 | 7875 – 2625 = 5250 |
| 3 | 5250 – 2625 = 2625 |
| 4 | 2625 – 2625 = 0 |

c Let $V_n$ = the value of the vacuum cleaner after $n$ years.
$V_o$= 10500, $V_{n+1} = V_{,,}$- 2625

d linear decay

e 10 500, 7875, 5250, 2625, 0




4 a i $112500 ii V,, = 2250000- 112 500K

iii $1350000 iv 20 years

b i $4000 ii V,, = 40 000-4000K

iii $8000 iv 10 years

c i $1640 ii V,, = 20500- 1640K

iii $7380 iv 13 years

5 a $7200 b 12%

6 A 7 C 8 D 9 A

10 C

11 a $V_o$ = 75000

Vj = 75000 - 3375 = 71625

$V_2$ = 71625-3375 = 68250

b i $3375 ii 4.5%

12 a $15000

b Vj = 120000 - 15000 = 105000

$V_2$ = 105000 - 15000 = 90000

c 12.5%

d $V_n$= 120000- 15000K

13 $V_o$ = 60000, $V_{n+1}$ = V,,-4800

14 a $\frac{38000 - 16000}{8}$ = 2750

b $C_o$ = 38000, $C_{n+1}$ = C,, - 2750

15 a $5032 b 10.2%

16 a $180

b $V_0$ = 3OOO

VJ = 3000 - 180 = 2820

So $V_2$ = 2820 - 180 = 2640.

c 6 years

## EXERCISE 5.4

1 D 2 E

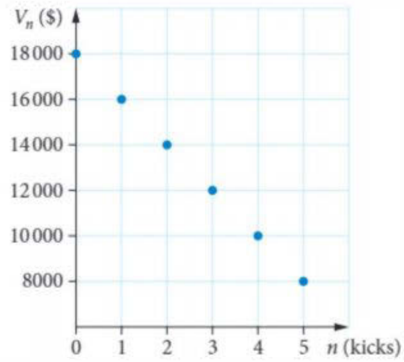
3 a The amount of depreciation is determined by applying a rate per unit of use: $2000 every time the football is kicked.

b i $12000 ii 5 kicks

| n | Value after n units of use ($) |
|---|---|
| 0 | 18000 |
| 1 | 18000-2000= 16000 |
| 2 | 16000-2000= 14000 |
| 3 | 14000-2000= 12000 |
| 4 | 12000-2000= 10000 |
| 5 | 10000 - 2000 = 8000 |

c Let $V_n$ = value of the football after *n* kicks.
*Vo* = 18000, $V_{,,+1}$ = $V_n$- 2000

d linear decay

e



4 a Let $V_n$ = value of vehicle after *n* kilometres.
V,, = 75000 -0.3K

b $33000

c i 200000 km ii 250000 km

5 a $0.03 per page

b $V_n$ = 26000 - 0.03K

6 A 7 E 8 E 9 C

10 E 11 E

12 a $8000

b i 8000-6500= 1500 ii $500

c 12000 km

13 $0.55 per km

14 a $3064 b $\frac{200}{25}$= $8

c G,, = 3264-8XM d 96 concerts


9780170448406

## EXERCISE 5.5

1 A 2 E

3 a

| | Compound | | Simple | |
|---|---|---|---|---|
| $n$ | Interest ($) | Value of investment ($) | Interest ($) | Value of investment ($) |
| 0 | | 4000 | | 4000 |
| 1 | $\frac{10}{100} \times 4000 = 400$ | 4000 + 400 = 4400 | $\frac{10}{100} \times 4000 = 400$ | 4000 + 400 = 4400 |
| 2 | $\frac{10}{100} \times 4400 = 440$ | 4400 + 440 = 4840 | $\frac{10}{100} \times 4000 = 400$ | 4400 + 400 = 4800 |
| 3 | $\frac{10}{100} \times 4840 = 484$ | 4840 + 484 = 5324 | $\frac{10}{100} \times 4000 = 400$ | 4800 + 400 = 5200 |
| 4 | $\frac{10}{100} \times 5324 = 532.40$ | 5324 + 532.40 = 5856.40 | $\frac{10}{100} \times 4000 = 400$ | 5200 + 400 = 5600 |

b $5856.40 c $256.40

4 a i 12 ii 108 iii $\frac{5}{12}$% iv $41.67

b i 52 ii 468 iii $\frac{3}{52}$% iv $20.19

c i 26 ii 234 iii $\frac{4}{13}$% iv $67.69

d i 365 ii 3285 iii $\frac{7}{365}$% iv $7.67

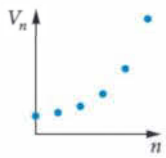
e i 4 ii 36 iii $\frac{3}{2}$% iv $210.00

5 a i $V_o = 13000$, $V_{M+1} = 1.027V_M$

ii $V_o = 18000$, $V_{n+1} = 1.003V_n$

iii $V_o$= liooo, $V_{M+1} = 1.0105V''$

b geometric growth

c 


6 a $6000

b $V_o = \$6000$

V! = 1.028$V_o$ = 1-028 x 6000 = $6168

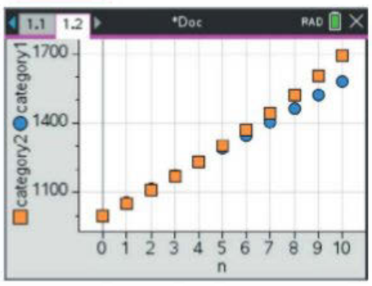
$V_2$ = 1.028^ = 1.028 x 6168 = $6340.70

c 2.8% d after 6 years e $1081.251

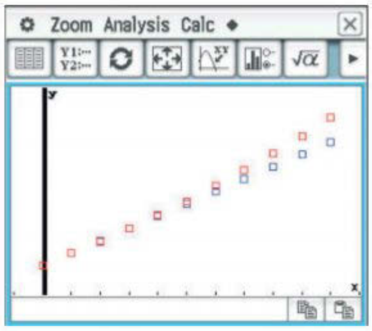
7 a $\frac{9}{12}$% = 0.75% b V= 1.0075" x 55000

C $86112.46 d $85828.01 e monthly

8 a TI-Nspire




ClassPad




b Account 2 is the better option.

9 a 7.1% b 24 months C $12000

10 D 11 B 12 C 13 E

14 a $1584 = 1.056 \times A$, so $A = \frac{1584}{1.056} = \$1500$

b $1865.29 C $Q_{rj}$ = 2080.05 x 1.0046"

15 a $15000

b $V_o$ = 15000, Vi = 1.04 x 15000 = 15600,

so $V_2$ = 1.04 x 15600= 16224

c 4%

d i V"= 1.04" x 15000 ii $22203.66

16 a $12000

b i Vj = 1.0062 x 12000 ii 4 months

c i balance = 12 000 x 1.0062" ii 36

## EXERCISE 5.6

1 B 2 E

3 a Bank 1:7.68%, Bank 2: 7.78%, Bank 3: 7.74%, Bank 4: 7.7%

b Bank 2 earns the most interest (7.78%).

c Bank 1 earns the least interest (7.68%).

d The nominal and effective interest rates for Bank 4 are the same because the rate compounds annually.

4 a 9.38% b 11.61% c 12.36% d 6.18%

5 D 6 D 7 E 8 A

9 C 10 D

11 a 7.8% p.a.

b Aussie bank has the higher effective interest rate because it has the same nominal interest rate as Power Bank but has a higher the number of compounding periods.

c 8% The nominal and effective interest rates for Power Bank are the same because the rate compounds annually.

d Bank of Victoria: 8.03%, and Aussie Bank: 8.16%. Jillian should choose Aussie Bank.

## EXERCISE 5.7

1 E 2 C

3 a Answers can vary slightly depending on when values are rounded.

i $5018 ii $1138 iii the third year

| $n$ | Depreciation after $n$ years ($) | Value after $n$ years ($) |
|---|---|---|
| 0 | | 12000 |
| 1 | $\frac{16}{100}$ x 12000= 1920 | 12000- 1920= 10080 |
| 2 | $\frac{16}{100}$ x 10080= 1613 | 10080- 1613 = 8467 |
| 3 | $\frac{16}{100}$ X 8467 = 1355 | 8467- 1355 = 7112 |
| 4 | $\frac{16}{100}$ x 7112=1138 | 7112- 1138 = 5974 |
| 5 | $\frac{16}{100}$ x 5974 = 956 | 5974 - 956 = 5018 |

b $V_o$ = 12000, $V_{n+1}$= 0.84 $V_n$ C 84%

d geometric decay e about $1000

4 a $V_o$ = 25000

Vi = 0.6 $V_o$ = 0.6 x 2 5000 = 15 000

$V_2$ = 0.6 V, = 0.6 x 15 000 = 9000

b 40% c 5 years

5 a V,, = 0.75" x 200000 b $20023

c 6 years d $8899

6 a $V_o$ = 64000, $n$ = 2, $V_2$ = 40960, r=?

"•■('-TS)"'"1'"

$$V_2 = \left(1 - \frac{r}{100}\right)^2 \times 64000$$

$$40\,960 = \left(1 - \frac{r}{100}\right)^2 \times 64000$$

$$0.64 = \left(1 - \frac{r}{100}\right)^2$$

$$0.8 = 1 - \frac{r}{100}$$

$$\frac{r}{100} = 0.2$$

$$r = 20\%$$

b $7000

7 C 8 D 9 A 10 E

11 A 12 A

13 a Vj = 0.9 x 60000 = 54000

$V_2$ = 0.9x54000 = 48600

b 10% c 11 years

14 5.7%

15 a 15% b $5484.23

16 a 3000 - 2760 = $240 $\frac{240}{3000}$ o.O8 = 8%

b $S_o$ = 3000, $S_{n+1}$ = 0.92S,,

## CUMULATIVE EXAMINATION 1

1 B 51% 2 D 53% 3 D 64%

4 D 81% 5 C 80% 6 B

7 A 83% 8 A 9 D 60%

10 B 39% 11 B 63% 12 A 49%

13 A 55% 14 B

## CUMULATIVE EXAMINATION 2

1 a 65.5%

b The data for each of the four age groups in the table does the support the opinion that age at first marriage is associated with the year of marriage. For example, of all first marriages, the percentage of women aged 25-29 years increased from 23.4% (1986) to 31.7% (1996) to 34.5% (2006).

2 a -0.375 (the association is negative: the smaller the area, the higher the population density)

b -0.8

3 a $120 b $325

4 Flat rate gives a greater depreciation by $28.41. 35%

5 a $5000 b $512.50

c $V_o$ = 5000, $V_{n+1}$ = 1.05 $V_n$

d 9.5% (an answer of 9.6% is also accepted, as this generates over $6000).

6 a Simple Saver annual interest rate is the highest because the first-year increase is larger. 34%

b $920 34%

c i $24\,000 = 8000 \times \left(1 + \frac{r}{100}\right)^{15}$

ii 7.6%

# CHAPTER 6

## EXERCISE 6.1

1 a $62098.75 b 43 months

c i 15 quarters ii 2 months

2 a 7.4% b $135000

3 a N = 156, CpY or C/Y = 52

b N = 10, PpY or P/Y = 2

c N = 48, PV =-15000

d 1% = 3.6, PV = -4000, FV = 5000

e N = 6, PV =-70000, CpY or C/Y = 1

f N = 4,1% =-12.5, FV = 18600

4 a $3942 b $V_4$ = $4846.94

5 E 6 C 7 B 8 C

9 B 10 B 11 C

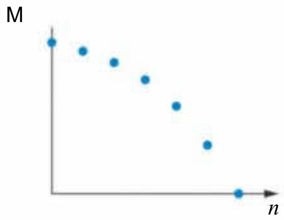
12 a $6089.34 b 14.7%

## EXERCISE 6.2

1 B 2 C

3 a $V_o$ = 30000, $V_{,,+1}$ = 1.0385$V_n$ - 230

b The recurrence relation models a combination of linear decay and geometric growth.

c



4 a $15000 b $465

C $V_o$= 15000

V! = 1.006 $V_Q$ - 465
= 1.006 x 15000-465
= 14625.00

$V_2$ = 1.006 Vj - 465
= 1.006 x 14625.00-465
= 14247.75

$V_3$ = 1.006$V_2$ - 465
= 1.006 x 14247.75 - 465
=13868

d 7.2% e 7 months

5 a $r = \frac{2850.00}{380000.00}$ x 100 = 0.0075 x 100 = 0.75%

$r = \frac{2842.47}{378996.00}$ x 100 = 0.0075 x 100 = 0.75%

b 9% per annum compounding monthly

c

| Payment number | Payment | Interest | Principal reduction | Balance |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 380000.00 |
| 1 | 3854.00 | 2850.00 | 1004.00 | 378996.00 |
| 2 | 3854.00 | 2842.47 | 1011.53 | 377984.47 |
| 3 | 3854.00 | 2834.88 | 1019.12 | 376965.35 |
| 4 | 3854.00 | 2827.24 | 1026.76 | 375938.59 |

6 a $2664.24 b $2643.43 C $106.57

d $7335.76 e $1020.81

f The balance after the last payment is $20.81. It would be $0 if the loan had been paid out.

g $2770.81

7 a Vj = 1.0036 x 80 000 - 288 = 80 000

$V_2$ = 1.0036 x 80000 - 288 = 80000

The value of the loan stays at the principal value $80000 for all compounding periods.

b $200000 C $6825

8 A 9 C 10 D 11 C

12 B 13 B 14 A 15 C

16 B 17 A

18 a i $643.85 ii $317428.45

b $S_o$ = 320000, $S_{n+1}$ = 1.003S,, - 1600

## EXERCISE 6.3

1 D 2 B

3 a 5.2% b 3.8% c 6.6% d 20.9%

e The payments are made to the bank, so the money is moving away from the person.

4 a i $154.66 ii $9279.60 ii i $1279.60

b i $2097.64 ii $755150.40 ii i $455150.40

c i $1514.18 ii $60567.20 ii i $25567.20

d i $442.63 ii $115083.80 ii i $40083.80

5 a i $36157.44 ii 38% ii i 102

iv $195.05 v $1030.05

b i $32815.84 ii 18% ii i 272

iv $221.69 v $471.69

6 a 4.5% b $3517.50 C 3.2% d $9250.00

7 B 8 D 9 C 10 A

11 C 12 C 13 E

14 Yes, Sally will pay it off in the 16th quarter, after 4 years.

15 a $807.23 b 47 months

C i $43234 ii $281.02

16 150

## EXERCISE 6.4

1 B 2 C 3 14 years

4 a $817.14 b $52269.08

5 a $161.50 b $371.48 c $716.93

6 B 7 C 8 C 9 E

10 a $225 b $16801 c $622.75

11 a $107.50 b $250

c $420.40. Using a finance solver:
N = 12,1% = 12.9, PV = 3776.15, PMT = -330,
FV = -90.40065597, P/Y = C/Y = 12

So, the last payment of $330 must be increased by $90.40 to repay the loan.

12 a i 300 months ii $3694.25

b $2265.04

13 a i $65076.22 ii $4676.22

b $28204

## EXERCISE 6.5

1 D 2 B

3 a $V_Q$ = 45000, $V_{M+1}$ = 1.005 $V_n$-250

b 10.4% C $65.39

4 a $380000.00 b $3854.00

c 0.75% per month d 9% p.a.

e $2834.88 f $1004.00

g $376965.35

5 17 years 6 $3355.08

7 D 8 C 9 C 10 C

11 C 12 C 13 C

14 a $a = 200000$, $R = 1.0075$, $d = -2000$

b 186

c i 93 months ii $2299.31

d i 166 months ii $682.57

15 a $3000 b 18 months

## EXERCISE 6.6

1 B 2 C

3 a Vj = 1.006x31000- 186=31000

$V_2 = 1.006 \times 31000 - 186 = 31000$

The value of investment stays at the principal value $31000 for all compounding periods.

b $53571.43 c $13650

4 a 6.3% b $8102 C 8.7% d $45292

5 B 6 E 7 B 8 B

9 E

10 a 3.75% b $20000

11 a $498398.08 b 2.88% c 1.004

12 a 6.3% b $80000 c $35208

## EXERCISE 6.7

1 C 2 D

3 a $V_Q = 27000$, V,,+1 = 1.006V,, + 310

b 7.2% C $839.76

4 a $r = \frac{40.00}{500.00} \times 100 = 0.08 \times 100 = 8\%$

$r = \frac{83.20}{1040.00} \times 100 = 0.08 \times 100 = 8\%$

b 8% p.a.

c

| Payment number | Payment | Interest | Principal addition | Balance |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 500.00 |
| 1 | 500.00 | 40.00 | 540.00 | 1040.00 |
| 2 | 500.00 | 83.20 | 583.20 | 1623.20 |
| 3 | 500.00 | 129.86 | 629.86 | 2253.06 |
| 4 | 500.00 | 180.24 | 680.24 | 2933.30 |

5 a 1% per month b $60.00

c $1272.60 d $1600.00

6 a 8 months b $534.81

7 $1014606.61

8 D 9 C 10 B 11 D

12 E 13 D 14 A

15 a 6.54%

b $11276.52

(1 mark for balance after 3 years $9964.628...)

16 a $1560 b $805.65

17 a $45.33

b $V_o = 2500$, and V,,+1 = 1.0034 $V_M$+150

c 5.87%

## CUMULATIVE EXAMINATION 1

1 E 77% 2 E 73% 3 D 76%

4 A 37% 5 B 79% 6 D 67%

7 B 65% 8 B 9 B 41%

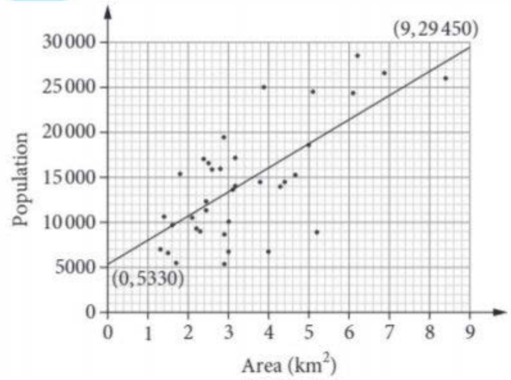
10 A 11 B 12 D

13 A 31%

## CUMULATIVE EXAMINATION 2

1 a Population is the response variable. 86%

b 36%




c The slope of the line is 2680, so the rate of increase in population is approximately 2680 people for every increase of 1 km² in area. 41 %

d i predicted population = 5330 + 2680 x 4 = 16050

residual = 6690 - 16050 = -9360

ii $I^2 = 0.668^2$ ® 0.446

So 44.6% of the variation in the population is explained by the variation in area. 42%

2 a $5060.27 64% b 0.3% 44%

C $610 34%

3 a $1704.03 27% b $45246.67 19%

C $B_o$ = 262 332.33, $B_{n+1}$ = 1.004B,,- 3517.28 35%

4 a $3700 72%

b $1 + \frac{r}{100} = 1.0035$, so r = 0.35%

Annual compound interest rate = 12 x 0.35

= 4.2% 30%

C $92.15 36% d $700 40%

5 a $569377 68%

b (1.001 - l)x 26 x 100% = 2.6% 27%

C $1198.59 15%

# CHAPTER 7

## EXERCISE 7.1

1 a C = $\begin{bmatrix} 100 & 80 & 80 & 75 & 100 \\ 225 & 125 & 150 & 175 & 150 \\ 125 & 100 & 150 & 150 & 175 \end{bmatrix}$

Order is 3 x 5. C has 15 elements,

b [150] Order is 1 x 1.



9780170448406

c $\begin{bmatrix} 225\ 125 & 150\ 175 & 150 \end{bmatrix}$

d $\begin{bmatrix} 100 \\ 80 \\ 80 \\ 75 \\ 100 \end{bmatrix}$

e $\begin{bmatrix} 75\ 175 & 150 \end{bmatrix}$

f $\begin{bmatrix} 435 \\ 825 \\ 700 \end{bmatrix}$

g

| | s | F | B |
|---|---|---|---|
| Chocolate | 100 | 225 | 125 |
| Fruit | 80 | 125 | 100 |
| Tea | 80 | 150 | 150 |
| Banana | 75 | 175 | 150 |
| Butter | 100 | 150 | 175 |

2 a 4x1; column matrix

b 4x3; binary matrix

c 1x4; row matrix, zero matrix

d 3x3; square matrix, binary matrix, permutation matrix

e 4x4; square matrix, binary matrix

f 3x1; column matrix, summing matrix, binary matrix

3 a $A^T = \begin{bmatrix} 7 & 8\ 0\ 3 \\ 2\text{-}194 & \end{bmatrix}$

The order of A is 4 x 2. The order of $A^t$ is 2 x 4.

b $A^T$-[ 6 3 0 ]

The order of $A$ is 3 x 1. The order of $A^t$ is 1 x 3.

c $A^T = \begin{bmatrix} 7\ 12 \\ 1\ 6 \end{bmatrix}$

The order of $A$ is 2 x 2. The order of $A^t$ *is 2x2.*

4 $\begin{bmatrix} -4 & 0 & -10 & 18 \\ 14 & 12 & 9 & 1 \\ 0 & -5 & 13 & -2 \end{bmatrix}$

5 a 5,8,2; upper triangular matrix

b -5, -1,7,1; lower triangular matrix

c 1,1; identity matrix, diagonal matrix, upper triangular matrix, lower triangular matrix, symmetric matrix

d This is not a square matrix so it has no leading diagonal and can not be any of the options,

e 22,25,19; symmetric matrix

6 a 2 x 2; square matrix b 3 x 1; column matrix

c 1x2; row matrix

d 2x2; lower triangular matrix

e 3x3; binary matrix f 4x4; symmetric matrix

7 a false b true c true d false

e true

8 B 9 E 10 D 11 C

12 A 13 C 14 B 15 C

16 a 4x5

b

$$\begin{matrix} & R\ B\ L\ A\ K & \\ & \begin{bmatrix} 110\ 0\ 1 \\ 0\ 10\ 10 \\ 0\ 10\ 0\ 0 \\ 10\ 111 \end{bmatrix} & \begin{matrix} F \\ G \\ S \\ T \end{matrix} \end{matrix}$$

c binary matrix d $\begin{bmatrix} 1\ 0\ 1\ 1\ 1 \end{bmatrix}$

## EXERCISE 7.2

1 C 2 B

3 a Forest SC won the soccer competition 7 times,

b Heights High won the basketball competition 8 times,

c Heights High won 19 competitions.

d The basketball competition was held 10 times.

4 a $\begin{bmatrix} 1 & 1 & 8 \\ 5 & 1 & 7 \end{bmatrix}$ b $\begin{bmatrix} 1 & 1 \\ 6 & 1 \\ 1 & 9 \end{bmatrix}$

c $\begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$ d $\begin{bmatrix} 0 & 1 & 1 \\ 1\ 0\ 1 & & \\ 1\ 1\ 0 & & \end{bmatrix}$

5 B 6 C 7 B 8 C

9 B 10 E 11 C 12 D

## EXERCISE 7.3

1 A 2 D

3 a $\begin{bmatrix} -7\ 0 \\ 0\ 4 \end{bmatrix}$

b Addition is not defined because the matrices have different orders.

c $\begin{bmatrix} 4 & 24 \end{bmatrix}$ d $\begin{bmatrix} 1 & * \\ 7 & 7 \end{bmatrix}$

e $\begin{bmatrix} -5 & -30 \end{bmatrix}$ f $\begin{bmatrix} -9 & 20 \\ 0\ 20 & \end{bmatrix}$

4 a $e = 2J\text{--}3$

b $M = \begin{bmatrix} 10 & -5 \\ 13 & 0 \end{bmatrix}$

c $M+5N = \begin{bmatrix} 12 & 8 & 4 \\ 28 & 24 & 20 \\ 44 & 4(1 & 36 \end{bmatrix}$

5 a 1.5 x $\begin{bmatrix} 50 \\ 120 \\ 840 \end{bmatrix}$

b 1.5 x $\begin{bmatrix} 50 \\ 120 \\ 840 \end{bmatrix} + \begin{bmatrix} 20 \\ 20 \\ 20 \end{bmatrix}$

C 1.5 x $\begin{bmatrix} 50 \\ 120 \\ 840 \end{bmatrix} + \begin{bmatrix} 20 \\ 20 \\ 20 \end{bmatrix} - \begin{bmatrix} 10 \\ 10 \\ 40 \end{bmatrix}$

6 a $\begin{bmatrix} 16 & 14 \\ 16 & 24 \end{bmatrix}$ b $\begin{bmatrix} -1 & -24 \\ 9 & -4 \end{bmatrix}$

c $\begin{bmatrix} 37 & 37 \\ 35 & 56 \end{bmatrix}$ d $\begin{bmatrix} 18 & 62 \\ -2 & 32 \end{bmatrix}$

7 A 8 D 9 C 10 C

11 C 12 D 13 B 14 A

## EXERCISE 7.4

1 D 2D

3 a i $AB$ has order (1 x 3)(3 x 1), number of columns in $A$ = number of rows in B, so $AB$ is defined.

ii lxl iii [60]

b i $BA$ has order (3 x 1)(1 x 3), number of columns in B = number of rows in A, so $BA$ is defined.

ii 3x3 iii $\begin{bmatrix} 6 & 3 & 15 \\ 8 & 4 & 20 \\ 20 & 10 & 50 \end{bmatrix}$

c i BC has order (3 x 1 )(2 x 2), number of columns in B # number of rows in C, so BC is not defined,

d i $BD$ has order (3 x 1 )(3 x 2), number of columns in B # number of rows in D, so BD is not defined,

e i DC has order (3 x 2)(2 x 2), number of columns in D = number of rows in C, so DC is defined.

ii 3 × 2 iii $\begin{bmatrix} 24 & -6 \\ 11 & -1 \\ 16 & -5 \end{bmatrix}$

f i C is a square matrix and powers of square matrices are always defined, so $C^2$ is defined. $C^2$ has order 2x2 and 2C has order 2x2. Matrices must have the same order to be subtracted, so $C^2$ - 2C is defined.

ii 2x2 iii $\begin{bmatrix} -3 & \circ \\ 0 & -3 \end{bmatrix}$

g D has order 3x1. Only powers of square matrices are defined. D is not a square matrix so $D^4$ is not defined.

4 a $\begin{bmatrix} -1 & 2 \\ -2 & 3 \end{bmatrix}\begin{bmatrix} 2 \\ a \end{bmatrix} = \begin{bmatrix} -2+2a \\ -4+3fl \end{bmatrix}$,

$\begin{bmatrix} -2 & 4 \\ 6 & \_1 \end{bmatrix}\begin{bmatrix} b \\ 1 \end{bmatrix} = \begin{bmatrix} -2b+4 \\ 6b-1 \end{bmatrix}$,

If $\begin{bmatrix} -2+2a \\ -4+3a \end{bmatrix} = \begin{bmatrix} -2b+4 \\ 6b-1 \end{bmatrix}$

then $-2 + 2a = -2b + 4$ and $-4 + 3a = 6b - 1$.

Simplifying gives $a = 3 - b$ and $a = 1 + 2b$ so they must be true.

b If $P^2$ is defined then $P$ and $P^2$ are both square matrices.

Let $P^2$ have order *tn* x *tn.*

For $RP^2$ to be defined, $R$ needs to have order 1 x *nt.*

For $P^2C$ to be defined, C needs to have order *tn* x 1.

$RP^2C$ has order (1 x *m)(m* x *tn)(tn* x 1)

= (1 x *m)(tn* x 1)

= (lxl)

Hence, $RP^2C$ is a 1 x 1 matrix.

5 a $\begin{bmatrix} 34.8 & 37.2 \\ 38.1 & 65.7 \end{bmatrix}$ b $\begin{bmatrix} 657 & 700 \\ 900 & 757 \end{bmatrix}$

c $\begin{bmatrix} -552.3 & -576.4 \\ -414.9 & -388.4 \end{bmatrix}$

6 a $\begin{bmatrix} 9 & 20 \\ 3 & 9 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 29 \\ 12 \end{bmatrix} > \frac{1}{2}\begin{bmatrix} 29 \\ 12 \end{bmatrix} = \begin{bmatrix} 14.5 \\ 6 \end{bmatrix}$

b $\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 6 & 4 \\ 1 & 3 \\ 5 & 2 \end{bmatrix} = \begin{bmatrix} 12 & 9 \end{bmatrix}$;

412 9 ] = [4 3 ]
3

7 a $\begin{bmatrix} S \\ T \\ A \\ R \end{bmatrix}$ b P = $\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

8 E 9 D 10 A 11 E

12 D 13 B 14 E 15 E

16 B 17 B

## EXERCISE 7.5

1 D 2 D

3 $\begin{bmatrix} 3 & -4 \\ -2 & q \end{bmatrix}\begin{bmatrix} 3 & 4 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$\begin{bmatrix} 3 & 4 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 3 & -4 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

4 a i det (A) =-2

ii $A^{-*}$ $-\frac{1}{2}\begin{bmatrix} 7 & -8 \\ -2 & 2 \end{bmatrix} = \begin{bmatrix} -\frac{7}{2} & 4 \\ 1 & -1 \end{bmatrix}$

b i det (B) = -5

ii $B^{-1} = -\frac{1}{5}\begin{bmatrix} 1 & -10 \\ -1 & 5 \end{bmatrix} = \begin{bmatrix} -\frac{1}{5} & 2 \\ \frac{1}{5} & -1 \end{bmatrix}$

c i det (C) = 0 ii $C^{-1}$ does not exist,

5 a x = -4 b x = 6

c A + B is defined, so A and B must be of the same order *mxm.*

If A is *tn* x *m,* then $A^{-1}$ is *tn* x *tn* and $A^2$ is *m* x *tn.*

If B is m x m, then $B^3$ is *tn* x *tn.*

The matrix order equation for $A^2A^{-1}B^3$ - B is

*(tti* x *tn)* x *(tn* x *tn)* x *(tti* x *tn)* - *(tn* x *tn)*
= *(tn* x *tn)* - *(tn* x *tti)*

The two matrices being subtracted are of the same order, so $A^2A^{-1}B^3$ - B is defined.

6 a i det (A) = 1

ii A'1 $= \begin{bmatrix} 46 & -51 & 81 \\ -9 & 10 & -16 \\ -26 & 29 & -46 \end{bmatrix}$

b i det (A) = -3

ii A-* $= \begin{bmatrix} \frac{£}{3} & -\frac{1}{3} & \frac{1}{3} \\ -\frac{2}{3} & -\frac{1}{3} & \frac{£}{3} \\ 1 & 1 & 0 \end{bmatrix}$

c i det (A) = 1

ii A'1 $= \begin{bmatrix} 2 & -17 & 9 \\ -1 & 9 & -5 \\ -2 & 19 & -10 \end{bmatrix}$

7 E 8 D 9 D 10 B

11 A 12 B 13 D 14 C

## EXERCISE 7.6

1 B 2 E

3 a $W = \begin{bmatrix} 1 \\ 4 \\ 6 \end{bmatrix}$

b $R = \begin{bmatrix} 26 & 5 & 2 \\ 12 & 2 & 0 \\ 18 & 7 & 2 \\ 9 & 2 & 0 \\ 30 & 3 & 0 \\ 5 & 0 & 0 \\ 3 & 0 & 0 \end{bmatrix}$

c $S = \begin{bmatrix} 26 & 5 & 2 \\ 12 & 2 & 0 \\ 18 & 7 & 2 \\ 9 & 2 & 0 \\ 30 & 3 & 0 \\ 5 & 0 & 0 \\ 3 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \\ 6 \end{bmatrix} = \begin{bmatrix} 58 \\ 20 \\ 58 \\ 17 \\ 42 \\ 5 \\ 3 \end{bmatrix}$

d The award was a tie between Heshan and Ahmat.

e The Little Clunkers made 203, so they lost.

4 a i A = PB-[ 23 37 10 ]

A gives the totals for each type of train spotted by the club on the weekend.

ii fl$_{12}$ tells us that there were a total of 37 Blueghosts sighted on the weekend.

b i $C = \begin{bmatrix} 13 \\ 11 \\ 18 \\ 17 \\ 11 \end{bmatrix}$

C gives the total number of trains spotted by each trainspotter on the weekend.

ii $c_{41}$ tells us that Steve spotted a total of 17 trains on the weekend.

c $\frac{1}{5}PBQ = \frac{1}{5}[70] = [14]$

$\frac{1}{5}PBQ$ tells us that the mean number of trains sighted by club members over the weekend is 14.

5 a $\begin{matrix} \text{Week 1} \\ \text{Week 2} \end{matrix} \begin{bmatrix} 75 & 60 \\ 47 & 82 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 330 \\ 340 \end{bmatrix}$

b i $C = \begin{bmatrix} 330 & 340 \\ 473 & 542 \\ 628 & 745 \\ 263 & 220 \end{bmatrix}$

ii S=1.75C = $\begin{bmatrix} 577.50 & 595.00 \\ 827.75 & 948.50 \\ 1099.00 & 1303.75 \\ 460.25 & 385.00 \end{bmatrix}$

c i profit = S - C

$\begin{bmatrix} 247.50 & 255.00 \\ 354.75 & 406.50 \\ 471.00 & 558.75 \\ 197.25 & 165.00 \end{bmatrix}$

ii 247.50 + 255.00 + ... + 165.00 = $2655.75

6 A 7 D 8 C 9 D

10 B

11 a $2.87 b 3x 1 c $\begin{bmatrix} 1 & 0 & 2 \end{bmatrix}$

12 a 120 students

b i $Q = S_0P = \begin{bmatrix} 5 & 10 & 3 & 2 \\ 15 & 30 & 9 & 6 \\ 10 & 20 & 6 & 4 \end{bmatrix}$

ii 30 intermediate-level students

c i $C \times Q$ or $\begin{bmatrix} 15 & 25 & 40 \end{bmatrix} \times \begin{bmatrix} 5 & 10 & 3 & 2 \\ 15 & 30 & 9 & 6 \\ 10 & 20 & 6 & 4 \end{bmatrix}$

ii $340

13 a 4x 1

b i [6000] (The brackets must be included.)

ii Total booking fees collected for the month

14 a 1x3 b 2800

c i 1296 and 729

ii The number of shoppers (594) in the clothing area of Westmall (at 1.00 pm).

d $\begin{bmatrix} 135 & 143 & 131 \end{bmatrix} \begin{bmatrix} 21.30 \\ 34.00 \\ 14.70 \end{bmatrix} = [9663.20]$

e $\begin{bmatrix} 1.05 & 0 & 0 \\ 0 & 0.85 & 0 \\ 0 & 0 & 0.99 \end{bmatrix}$

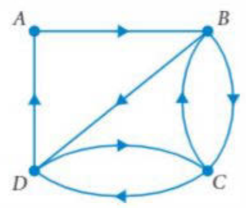
15 a $\begin{bmatrix} 2184 \\ 1782 \\ 1666 \end{bmatrix}$ b $\begin{bmatrix} 1.04 & 0 & 0 \\ 0 & 0.99 & 0 \\ 0 & 0 & 0.98 \end{bmatrix}$

## EXERCISE 7.7

1 C 2 E

3 a i Allie can send direct messages to Ben. Ben can send direct messages to Cassie and Debra. Cassie can send direct messages to Ben and Debra. Debra can send direct messages to Allie and Cassie.

ii The leading diagonal represents links where the sender and receiver are the same. This is not considered communication, so they are redundant links.

iii 


Other versions are possible,

iv Debra Cassie Ben →

b i Sender

| | Receiver | | | | | |
|---|---|---|---|---|---|---|
| | *P* | *Q* | *R* | s | *T* | *u* |
| *P* | 0 | 1 | 0 | 0 | 0 | 0 |
| Q | 1 | 0 | 1 | 0 | 0 | 0 |
| *R* | 0 | 1 | 0 | 1 | 0 | 0 |
| S | 0 | 0 | 1 | 0 | 1 | 0 |
| *T* | 0 | 0 | 0 | 1 | 0 | 1 |
| *U* | 0 | 0 | 0 | 0 | 1 | 0 |

ii The matrix is symmetric because all the communications go both ways.

4 a 2 ways b *A* -> B -> C and *A* -> D -> *C*

C 8 d *D*-* *A* →*D* and D —> C —> D

e 3

5 a i Winner

| | Loser | | | | |
|---|---|---|---|---|---|
| | *S* | *B* | *F* | *A* | *R* |
| S | 0 | 0 | 0 | 0 | 1 |
| B | 1 | 0 | 0 | 0 | 0 |
| *F* | 110 | | | 10 | |
| *A* | 110 | | 0 | | 0 |
| *R* | 0 1110 | | | | |

Total scores

ii S 4
B 2
*F* 7
*A* 4
*R* 9

iii Rhinos (1), Flames (2), Sonics (3), Angels (3), Blazers (5); The overall winner was the Rhinos.

b i Winner

| | Loser |
|---|---|
| | *J K L M* |
| *J* | 0 110 |
| *K* | 0 0 11 |
| *L* | 0 0 0 1 |
| *M* | 10 0 0 |

Total scores

ii
*/* *5*
*K* *4*
*L* *2*
*M* *3*

iii Jackie (1), Kat (2), Maisie (3), Lydia (4); the overall winner was Jackie.

6 a x=l,y = 0, z = 0 b BvsC

7 E 8 D 9 A 10 B
11 A 12 A 13 A 14 A

15 a Ben and Elka b Amara and Dana

16a Arnold and Edgar; Barnaby and Cedric

b Edgar

c Cedric, Barnaby, Arnold, Edgar

d 20

## CUMULATIVE EXAMINATION 1

| | | |
|---|---|---|
| 1 E | 2 A | 3 B |
| 4 A | 5 C | 6 C |
| 7 D | 8 A 92% | 9 E 73% |
| 10 A | 11 D 78% | 12 A 78% |
| 13 D 64% | 14 E 59% | 15 E 45% |
| 16 C 50% | 17 E 42% | 18 B |
| 19 C | | |

## CUMULATIVE EXAMINATION 2

1 a *population* = 7.7 + 7.7 x $\log_{10}$(area) 43%

b 23000 35%

2 a $\log_{10}$ *(power)* = 4.287 + 0.099 02 x *year number*

b 1851000

3 a $12706.41 36%

b $11848.58 36%

C i $129.80 4%

ii $1927.20 (one mark for finding that the total deposits made in the investment for the three year period = $11072.80) 4%

4 a 2

b B = $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ with columns *M F* and rows *G H*

c i table tennis

ii $\begin{bmatrix} 2 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 515 \\ 550 \\ 580 \end{bmatrix}$ =[1030]

d Q = $\begin{bmatrix} 1 & l & 0 \\ o & 0 & 1 \\ o & 1 & 1 \\ l & 0 & 0 \end{bmatrix}$

5 a 4x 1 b 56

c [25 20 45 35 1

d ^2018 = 1-25 X $F_2$017

6 a Anvil and Dantel

b Anvil - Berga - Dantel - Cantor

c G-KF-[ 1 2 1 1 ]

d The matrix *G* lists, for each city, the total number of direct flight connections to that city from another city in the network. 79%



9780170448406

# CHAPTER 8

## EXERCISE 8.1

1 a yes

b no, not a square matrix

c no, columns do not all add to 1

d yes

e no, columns do not all add to 1

f no, not a square matrix

g yes

h yes

2 a i x = 35%, y = 50%, z = 10%

ii This year

$$\begin{array}{c} \begin{matrix} A & B & C \end{matrix} \\ \begin{bmatrix} 0.35 & 0.5 & 0.2 \\ 0.4 & 0.3 & 0.7 \\ 0.25 & 0.2 & 0.1 \end{bmatrix} \end{array} \begin{matrix} A \\ B \\ C \end{matrix} \text{ Next year}$$

b This day

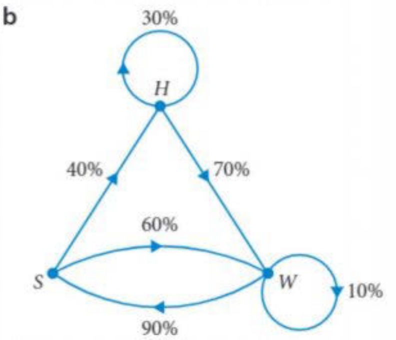
$$T = \begin{array}{c} \begin{matrix} R & D \end{matrix} \\ \begin{bmatrix} 0.65 & 0.1 \\ 0.35 & 0.9 \end{bmatrix} \end{array} \begin{matrix} R \\ D \end{matrix} \text{ Next day}$$

3 a $x = 0.0$, $y = 0.6$, $z = 0.1$

This week

$$\begin{array}{c} \begin{matrix} S & H & W \end{matrix} \\ \begin{bmatrix} 0.0 & 0.0 & 0.9 \\ 0.4 & 0.3 & 0.0 \\ 0.6 & 0.7 & 0.1 \end{bmatrix} \end{array} \begin{matrix} S \\ H \\ W \end{matrix} \text{ Next week}$$

b




4 a 6485 b 21320 c 33627 d 4.8%

5 a P is a transition matrix because it is a square matrix, and each column adds up to 1.

b Amisha will always select *B* after she has selected D.

c C

6 C 7 B 8 C 9 D

10 D 11 E 12 E 13 B

## EXERCISE 8.2

1 B 2 D

3 a $S_o = \begin{bmatrix} 4243 \\ 6010 \end{bmatrix}$ » $S_{n+}j$ - $\begin{bmatrix} 0.4 & 0.35 \\ 0.6 & 0.65 \end{bmatrix}$ S„

b 3801 bacon rolls and 6452 egg wraps

c bacon rolls: 0.4 x 4243 + 0.35 x 6010 = 3801
egg wraps: 0.6 x 4243 + 0.65 x 6010 = 6452

d 3779 bacon rolls and 6474 egg wraps

4 a Once visitors go to Algebra World, they stay there for the rest of the day.

b 5198 c 32%

5 a *a* = 0.3, & - 0.1, c- 0.7, *d* = 0.2, *e* = 0.4

6 56 at FushNChups and 90 at BurgerHQ.

7 a 59 train carriages at depot *A* and 71 at depot *B*.

b 56 train carriages at depot *A* and 74 at depot *B*.

c 57 train carriages at depot *A* and 73 at depot *B*.

8 D 9 E 10 C 11 B

12 A

13 a 20% b $S_1 = \begin{bmatrix} 300 \\ 780 \\ 300 \\ 620 \end{bmatrix} \begin{matrix} A \\ F \\ G \\ W \end{matrix}$

c 60% d $\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

14 a 0.15 x 100 + 0.25 x 400 + 0.20 xlOO + 0.50 x 1400 = 835

b 356

15 a All of the advanced-level students stay as advanced-level students.

b i $S_1 = \begin{bmatrix} 10 \\ 58 \\ 52 \end{bmatrix}$

ii 12 intermediate-level students

16 42 students will defer the 2009 academic year.

$$S_2 = \begin{bmatrix} 0.88 & 0.52 & 0.65 \\ 0.10 & 0.44 & 0.10 \\ 0.02 & 0.04 & 0.25 \end{bmatrix}^2 \begin{bmatrix} 880 \\ 230 \\ 120 \end{bmatrix} = \begin{bmatrix} 996.9 \\ 191.4 \\ 41.7 \end{bmatrix}$$

(One mark for the transition matrix.)

## EXERCISE 8.3

1 C 2 B

3 a *T* has no zero elements, so there will be an equilibrium state matrix.

b $\begin{bmatrix} 141 \\ 79 \end{bmatrix} \begin{matrix} A \\ B \end{matrix}$

c 141 buses will be at depot *A* and 79 buses will be at depot *B*.

d 64%

4 a $T^2 = \begin{bmatrix} 0.66 & 0.21 & 0.43 \\ 0.3 & 0.53 & 0.19 \\ 0.04 & 0.26 & 0.38 \end{bmatrix}$ has no zero elements, so there will be an equilibrium state matrix.

b 18%

c $\begin{bmatrix} 566 \\ 453 \\ 227 \end{bmatrix} \begin{matrix} C \\ H \\ N \end{matrix}$, 453 students are expected to bring their lunch from home.

d 529 students are expected to buy their lunch at the canteen.

e In the long term, all students end up buying their lunch at the canteen.

5 B 6 D 7 D 8 D

9 B 10 E 11 A 12 C

13 a $T = \begin{bmatrix} 0.80 & 0.09 & 0.10 \\ 0.12 & 0.76 & 0.05 \\ 0.08 & 0.15 & 0.85 \end{bmatrix}$

b $K_0 = \begin{bmatrix} 300\,000 \\ 120\,000 \\ 180\,000 \end{bmatrix}$

C $K_X = TK_O = \begin{bmatrix} 0.80 & 0.09 & 0.10 \\ 0.12 & 0.76 & 0.05 \\ 0.08 & 0.15 & 0.85 \end{bmatrix} \begin{bmatrix} 300\,000 \\ 120\,000 \\ 180\,000 \end{bmatrix}$

$= \begin{bmatrix} 268\,800 \\ 136\,200 \\ 195\,000 \end{bmatrix}$

There are 268800 at Shopper Heaven, 136200 at East own and 195000 at Noxland. (One mark for the matrix product and one mark for the statement at the end.)

d Any two products $T^nK_0$ where $n$ 38. For example:

j/' T^S jz K38 ~ 1^0 _ $\begin{bmatrix} 0.80 & 0.09 & 0.10 \\ 0.12 & 0.76 & 0.05 \\ 0.08 & 0.15 & 0.85 \end{bmatrix}^{38} \begin{bmatrix} 300\,000 \\ 120\,000 \\ 180\,000 \end{bmatrix}$

$= \begin{bmatrix} 194\,983 \\ 150\,513 \\ 254\,504 \end{bmatrix}$

and

$K_{39} = T\% = \begin{bmatrix} 0.80 & 0.09 & 0.10 \\ 0.12 & 0.76 & 0.05 \\ 0.08 & 0.15 & 0.85 \end{bmatrix}^{39} \begin{bmatrix} 300\,000 \\ 120\,000 \\ 180\,000 \end{bmatrix}$

$= \begin{bmatrix} 194\,983 \\ 150\,513 \\ 254\,504 \end{bmatrix}$

and this is the same as $X_{38}$.

14 a $\begin{bmatrix} 240\,700 \\ 231\,700 \\ 207\,600 \end{bmatrix}$ b 30000

C 7 d 218884

## EXERCISE 8.4

1 B 2 A

3 a 42 b 60

4 Each year the cooperative will buy 1250 young salmon and sell 1820 adult salmon.

5 a $\begin{bmatrix} 38 \\ 55 \\ 41 \\ -122 \end{bmatrix} \begin{matrix} I \\ S \\ T \\ V \end{matrix}$

b 122 fewer customers travel to Vietnam from one year to the next.

c $\begin{bmatrix} 621 \\ 371 \\ 306 \\ 126 \end{bmatrix} \begin{matrix} I \\ S \\ T \\ V \end{matrix}$

6 B 7 A 8 237966

9 a d = 298, e = 94,/= 130

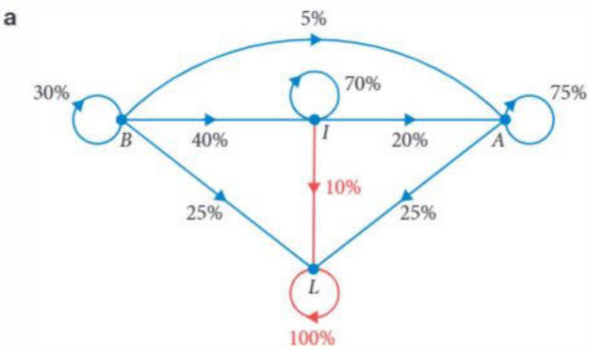
b 0.65 x 520 + 0.25 x 320 + 0.25 x 80 + 0.5 x 80 = 478 customers

c 20 customers

d i 80 customers who have no bookings $I$ no travel in that year will be removed from the study.

ii 190

10 a




One mark for a transition from $I$ to L, including the arrow and the 10% label. Another mark for the loop at $L$, including the 100% label.

b 17.5%

c 43 advanced-level students

d 5 assessments

e 7 intermediate-level students (if the answer was incorrect, one mark for writing

$\wedge = 7\wedge + V = \begin{bmatrix} 10 \\ 52 \\ 46 \\ 21 \end{bmatrix}$,

$R_2 = TR_1 + V = \begin{bmatrix} 7 \\ 42.4 \\ 48.4 \\ 40.2 \end{bmatrix}$ and

$R_3 = TR_2 + V = \begin{bmatrix} 6.1 \\ 34.48 \\ 48.13 \\ 58.29 \end{bmatrix}$ )

11 $\begin{bmatrix} 2160 \\ 2430 \\ 1410 \end{bmatrix}$

12 a Air World

b i $\begin{bmatrix} -210 \\ 0 \\ 210 \\ 0 \end{bmatrix}$ ii $\begin{bmatrix} 600 \\ 288 \\ 512 \\ 600 \end{bmatrix}$

13 a 30 b 457 km

## EXERCISE 8.5

1 E 2 B


9780170448406

3 a The birth rates given are for baby rodents, not just female baby rodents. So, we need to halve these figures.

b

| Age (years) | 0-<1 | 1-<2 | 2-<3 |
|---|---|---|---|
| Initial number | 93 | 72 | 68 |
| Birth rate | 3.9 | 4.5 | 3.1 |
| Survival rate | 0.33 | 0.46 | 0 |

c 93 x 0.33 = 30.69, so 31 one-year-old female rodents are expected to survive to be two.

d (93x3.9)+ (72x4.5)+ (68x3.1)
= 362.7+ 324 + 210.8 = 897.5, so 898 female rodents are expected to be born after one year.

e $L = \begin{bmatrix} 3.9 & 4.5 & 3.1 \\ 0.33 & 0 & 0 \\ 0 & 0.46 & 0 \end{bmatrix}$

4 a The first element in the first row is zero. This means spiders up to one year old have birth rates of zero.

b $S_1 = LS_0 = \begin{bmatrix} 0 & 33 & 12 & 10 \\ 0.1 & 0 & 0 & 0 \\ 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0.05 & 0 \end{bmatrix} \begin{bmatrix} 120 \\ 63 \\ 58 \\ 22 \end{bmatrix}$

$= \begin{bmatrix} 2995 \\ 12 \\ 12.6 \\ 2.9 \end{bmatrix} \begin{matrix} 0\text{-}<1 \\ 1\text{-}<2 \\ 2\text{-}<3 \\ 3\text{-}<4 \end{matrix}$ 2 x 2.9 = 5.8, so there are 6 three-year-old spiders after one year.

c $S_2 = LS_1 = \begin{bmatrix} 0 & 33 & 12 & 10 \\ 0.1 & 0 & 0 & 0 \\ 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0.05 & 0 \end{bmatrix} \begin{bmatrix} 2995 \\ 12 \\ 12.6 \\ 2.9 \end{bmatrix}$

$= \begin{bmatrix} 576.2 \\ 299.5 \\ 2.4 \\ 0.63 \end{bmatrix} \begin{matrix} 0\text{-}<1 \\ 1\text{-}<2 \\ 2\text{-}<3 \\ 3\text{-}<4 \end{matrix}$

(576.2 x 2) + (299.5 x 2) + (2.4 x 2) + (0.63 x 2)
= 1152 + 599 + 5 + 1 = 1757 total spiders after two years

d 6907

5 a increase b cycle c decrease

6 a i 1.07 ii 7% increase

b i 1.04 ii 4% increase

7 B 8 D 9 D 10 D

11 C 12 A 13 E 14 C

15 a 8 years

b 71%

c $L = \begin{bmatrix} 0 & 0.35 & 0.31 & 0.27 \\ 0.42 & 0 & 0 & 0 \\ 0 & 0.71 & 0 & 0 \\ 0 & 0 & 0.65 & 0 \end{bmatrix}$

d $\begin{bmatrix} 10\,000 \\ 10\,000 \\ 10\,000 \\ 4\,500 \end{bmatrix}$

e $L^{19} = \begin{bmatrix} 2.97 \\ 1.93 \\ 2.13 \\ 2.14 \end{bmatrix}$ $L^{20} = \begin{bmatrix} 1.91 \\ 1.25 \\ 1.37 \\ 1.38 \end{bmatrix}$

f i 3 (2 female and 1 male)

ii $g = 0.6$ This means that in the long term, the population of white-banded tanagers is declining by 40% every two years.

## CUMULATIVE EXAMINATION 1

| | | |
|---|---|---|
| 1 C 93% | 2 B 86% | 3 C 71% |
| 4 B 47% | 5 D | 6 A 26% |
| 7 c | 8 c | 9 C 60% |
| 10 c | 11 E 62% | 12 E |
| 13 A | 14 C 68% | 15 E 48% |
| 16 D | 17 A 40% | |

## CUMULATIVE EXAMINATION 2

1 a 0.95 45% b $13100 68%

2 $11029 (One mark for finding the principal remaining after two years is equal to $ 13 971.09.) 126%

3 a *A* and *E*

b Farmer D had attended one earlier conference with all others in this group.

4 a 40% 95%

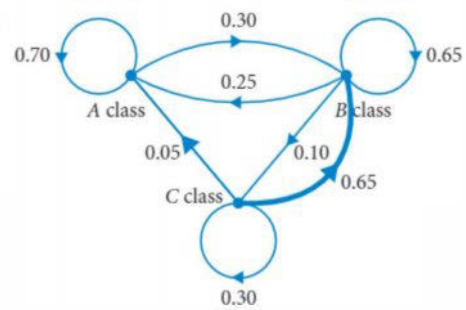
b 300x04 + 240x06 + 210x02 = 306 69%

c 750 30-/o

5 v = 0.65, $w = 0.15$, $x = 0.85$

6 a 3 × 1 93% b 1.05 48%

7 a




(One mark for the arrowhead from C to A, and one mark for the arrowhead from C to *B* and labelled with 0.65.)

b 8% C 12% d $\begin{bmatrix} 39.6 \\ 124.4 \\ 36.0 \end{bmatrix}$

e 100 f 84

8 a Brie and Dex 92%

b Elena - Dex - Brie - Chai 83%

c Alex - Brie - Dex and Alex - Elena - Dex, Chai - Alex - Brie and Chai - Dex - Brie

9 a The gerbil numbers given are total numbers, not just female. So, we need to halve these figures.

The birth rates given are for baby gerbils, not just female baby gerbils. So, we need to halve these figures.

The survival rates for males and females are the same, so we do not halve these.

b

| Age (years) | 0-<1 | 1-<2 | 2-<3 |
|---|---|---|---|
| Initial number | 32 | 48 | 21 |
| Birth rate | 1.9 | 2.4 | 1.3 |
| Survival rate | 0.45 | 0.57 | 0 |

C (32 X 1.9) + (48 X 2.4) + (21 x 1.3) = 203

d $L = \begin{bmatrix} 1.9 & 2.4 & 1.3 \\ 0.45 & 0 & 0 \\ 0 & 0.57 & 0 \end{bmatrix}$

e 15673

# CHAPTER 9

## EXERCISE 9.1

1 a Isomorphic because they show exactly the same connections.

b Not isomorphic because they have different numbers of vertices and edges.

c Not isomorphic because they don't have exactly the same connections. For example, in the first graph *A* connects to B but in the second graph it doesn't,

d Isomorphic because they show exactly the same connections.

2 a i 5 vertices: *A*, B, C, D, *E*

8 edges: *AB, AC, AD, BC* x 2, *CD, CE, DE*

5 faces: 4 enclosed and 1 outside the graph

ii

| Vertex | *A* | B | C | D | *E* | Sum |
|---|---|---|---|---|---|---|
| Degree | 3 | 3 | 5 | 3 | *2* | 16 |

degree sum = 2 x number of edges = 2x8= 16

b i 7 vertices: *A, B, C, D, E, F, G*

10 edges: *AB, AC, AD, BC* x 2, *CD, CE, DE, EE, EG*

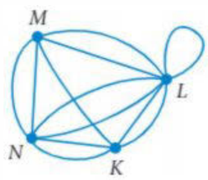
7 faces: 6 enclosed and 1 outside the graph

ii

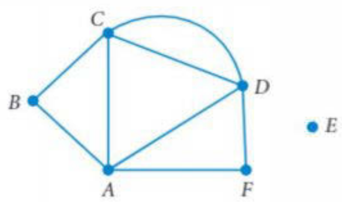
| Vertex | *A* | *B* | C | D | *E* | *F* | *G* | Sum |
|---|---|---|---|---|---|---|---|---|
| Degree | 3 | 3 | 5 | 3 | 5 | 0 | 1 | 20 |

degree sum = 2 x number of edges = 2 x 10 = 20

3 *LL,MNx2, MK,MLx2,NKx2,NLx2, LKx2*




4




5 B 6 D 7 E 8 D

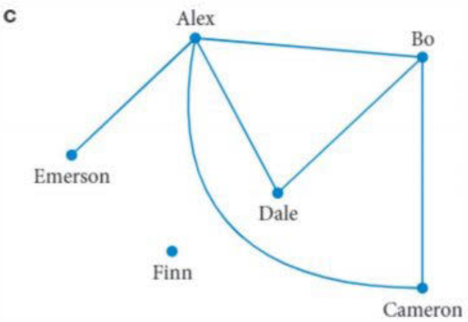
9 C 10 B 11 A 12 E

13 A

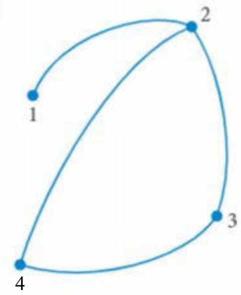
14 a 1 player

b Cameron and Dale

c




15 a




b 8

16 a Alooma and Easyside

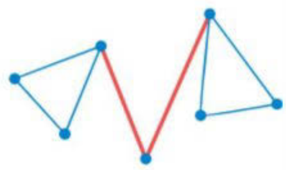
b i Draw a third edge joining *E* and D.

ii The driver can return to Dovenest without going through any other suburb.

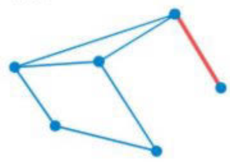
## EXERCISE 9.2

1 E 2 C

3 a two

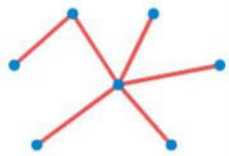


b one

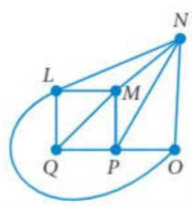


c no bridges

d six



4 a i




Other answers are possible.

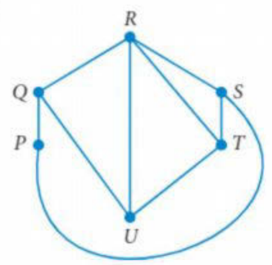
ii It is a connected graph because there is a path from each vertex to every other vertex,

ijj $_v$=6,/=7,e = ll;6 + 7- 11 =2



9780170448406

b i

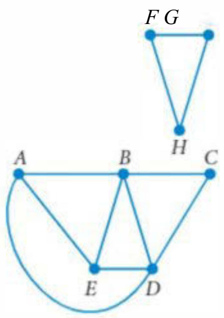



Other answers are possible.

ii It is a connected graph because there is a path from each vertex to every other vertex,

jjj v = 6,/=5,e = 9;; v+/-e = 6 + 5 - 9 = 2

Ci

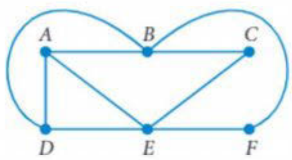



Other answers are possible.

ii It isn't a connected graph because there isn't a path from each vertex to every other vertex (e.g. C to H).

iii Euler's formula doesn't work: v = 8,/= 6,e = 11; 8 + 6 - 11 = 3

d i




Other answers are possible.

ii It is a connected graph because there is a path from each vertex to every other vertex.

iii v = 6,/= 5, e = 9; 6 + 5 - 9 = 2

5 a 5 b 14 c 4

6 a i It is a simple graph because it has no loops or multiple edges.

ii It is not a complete graph because not every vertex is connected by an edge to every other vertex.

iii It is a subgraph because it only has vertices and edges from the larger graph.

b i It is not a simple graph because it has multiple edges.

ii It is not a complete graph because it is not a simple graph.

iii It is a subgraph because it only has vertices and edges from the larger graph.

c i It is a simple graph because it has no loops or multiple edges.

ii It is a complete graph because every vertex is connected by an edge to every other vertex.

iii It is not a subgraph because it has two edges (BD and EC) not in the larger graph,

d i It is not a simple graph because it has a loop.

ii It is not a complete graph because it is not a simple graph.

iii It is a subgraph because it only has vertices and edges from the larger graph.

7 D 8 D 9 D 10 C

11 B 12 E 13 A 14 D

15 D

## EXERCISE 9.3

1 A 2 C

3 a circuit: no repeated edges, a repeated vertex, and starts and finishes at the same vertex,

b cycle: no repeated edges, no repeated vertices (except the first and last vertex), and starts and finishes at the same vertex.

c path: no repeated edges, no repeated vertices, and doesn't start and finish at the same vertex,

d walk only: repeated edge.

e trail: no repeated edges, a repeated vertex, and doesn't start and finish at the same vertex.

4 a path: no repeated edges, no repeated vertices, and doesn't start and finish at the same vertex,

b cycle: no repeated edges, no repeated vertices (except the first and last vertex), and starts and finishes at the same vertex.

c circuit: no repeated edges, a repeated vertex E, and starts and finishes at the same vertex,

d walk only: repeated edge (DE is the same edge as ED),

e walk only: repeated edge EE

f trail: no repeated edges, a repeated vertex E, and doesn't start and finish at the same vertex.

5 a i There are 4 vertices with odd degrees: A, B, C, D. An Eulerian trail exists only if there are exactly two odd vertices and an Eulerian circuit only exists if all the vertices are even, so neither exists for this graph.

ii Vertex *A* and vertex C are the only two vertices of odd degree. Exactly two vertices are of odd degree, so an Eulerian trail exists. Three Eulerian trails are

A-B-C-D-A-C

*C-A-D-C-B-A*

*A-D-C-A-B-C*

Other answers are possible.

iii All the vertices have even degree, so an Eulerian circuit exists. Three Eulerian circuits are

A-B-C-E-E-C-A-E-D-A

*A-D-E-A-C-E-F-C-B-A*

*A-E-C-A-B-C-F-E-D-A*

Other answers are possible.

b i Hamiltonian path: *K-L-M-N*

Other answers are possible.

Hamiltonian cycle: *K-L-M-N-K*

Other answers are possible.

ii Hamiltonian path: *A-B-C-D*

Other answers are possible.

Hamiltonian cycle: *A-B-C-D-A*

Other answers are possible.

iii Hamiltonian path: *B-A-D-E-F-C*

Other answers are possible.

Hamiltonian cycle: *B-A-D-E-F-C-B*

Other answers are possible.

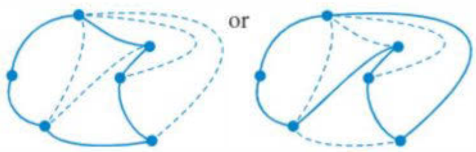
6 D 7A 8 D 9 D

10 C 11 B 12 E

13 a office

b i Hamiltonian cycle

ii




14 a Ramp *V*

b Either: *X-Y-T-U-Z-V-WorX-Y-T-U-Z-W-V.*

*c* Four ways:

. *X-Y-T-U-Z-V-W-X*

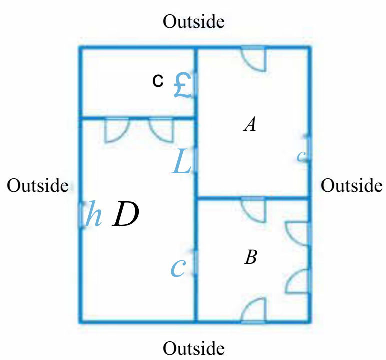
• *X-W-V-Z-U-T-Y-X*

. *X-Y-T-U-V-Z-W-X*

• *X-W-Z-V-U-T-Y-X*

Each Hamiltonian cycle can be reversed.

15 a Floor plan




b Hamiltonian path

c i There is at least one vertex on the graph that is of odd degree.

ii Rooms *B* and D

## EXERCISE 9.4

1 D 2 B

3 a

| | *A* | *B* | c | D | *E* |
|---|---|---|---|---|---|
| *A* | 0 | 1 | 0 | 1 | 0 |
| *B* | 1 | 0 | 0 | 1 | 0 |
| *C* | 0 | 0 | 0 | 1 | 1 |
| *D* | 1 | 1 | 1 | 0 | 1 |
| *E* | 0 | 0 | 1 | 1 | 0 |

b

| | *A B C D* |
|---|---|
| *A* | 110 0 |
| *B* | 10 3 0 |
| *C* | 0 3 0 0 |
| *D* | 0 0 0 0 |

c

| | *ABC* |
|---|---|
| *A* | 0 2 3 |
| *B* | 2 0 2 |
| *C* | 3 2 0 |

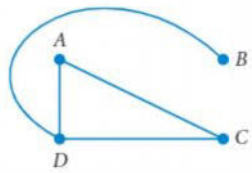
4 a i D is connected to *A, B* and C, so every vertex is connected to at least one other vertex.

ii degree sum =2+1+2+3=8

iii The graph has an Eulerian trail because it has exactly 2 vertices with odd degrees.

iv 2

v




Other versions are possible.

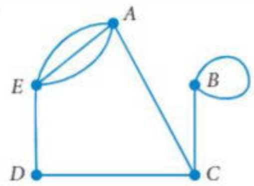
b i C is connected to *A, B* and D, and *A* is connected to $E_y$ so every vertex is connected to at least one other vertex.

ii degree sum = 4 + 3 + 3 + 2 + 4=16

iii The graph has an Eulerian trail because it has exactly 2 vertices with odd degrees.

iv 5

v




Other versions are possible.

5A 6E 7E 8E

9 D 10 B 11 C 12 A

13 a A zero means the two people are not allowed to communicate directly with each other,

b $f$= 1,g = 0

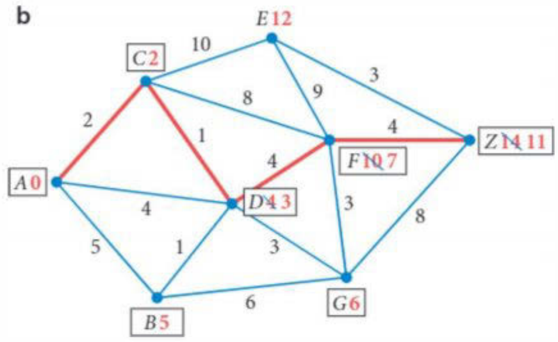
## EXERCISE 9.5

1 B 2 A

3 25 minutes

4 a 11 minutes

b



5 A 6 B 7 B 8 B

9 a 3 b 1000 m

c i *P4* ii Eulerian trail

d *E-P5-P4-P6-P3-P2-P1*

10 a *$120*

b Quigley and Rosebush

c 6 + 7= 11 + 2

11 a 3.2 km

b i Eulerian trail ii *P*

c S and *T*

12 a i factory- *T-S-Q-R-S-* [/-factory

ii Town S is passed through twice,

b 162 km

13 a Not all vertices are of even degree,

b Either one of the following:

- start at C and finish at D
- start at D and finish at C
- start at D and finish at H
- start at H and finish at D
- start at G and finish at H
- start at H and finish at G.

c 335 metres

14 a 2 b 108 km

## EXERCISE 9.6

1 B 2 B

3 a This graph is not a spanning tree. It has an edge *(CI)* that isn't in the original graph.

b This graph is not a spanning tree. It has a cycle,

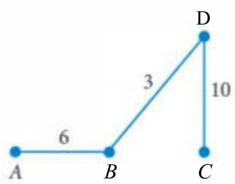
c This graph is not a spanning tree. It is not connected,

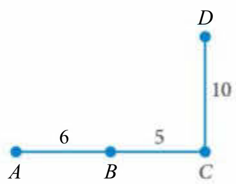
d This graph is a spanning tree. It has 9 vertices and 8 edges.

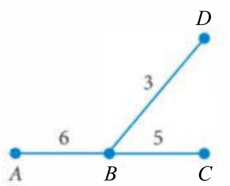
e This graph is a spanning tree. It has 9 vertices and 8 edges.

f This graph is not a spanning tree. It has a loop.

4

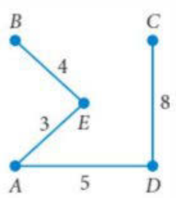


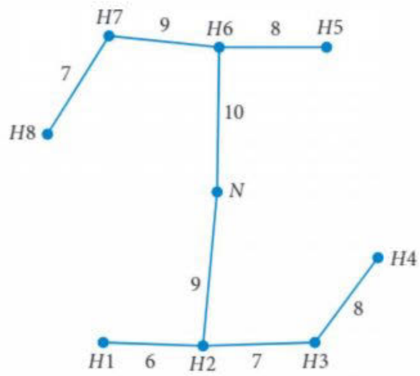

total weight of minimum spanning tree = 14

5




total weight = 20

6 64 m

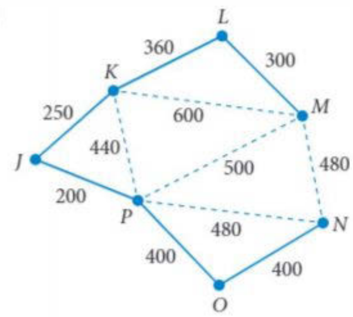



7 B 8 A 9 B 10 B

11 B 12 C 13 D 14 E

15 a $300 b $920 c *N* and *P* (or *P* and *N)*

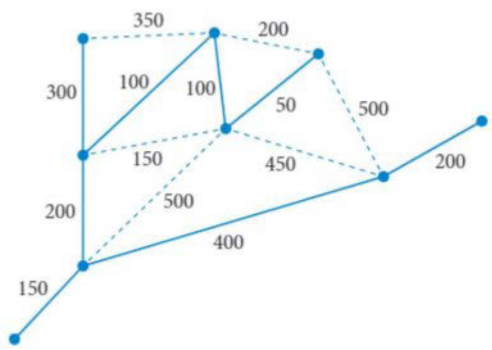
d i




ii $120

16 a minimum spanning tree

b




## CUMULATIVE EXAMINATION 1

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | C | 86% | 2 | D | 29% | 3 | B | 95% |
| 4 | C | 42% | 5 | D | | 6 | B | |
| 7 | B | 40% | 8 | B | 41% | 9 | B | 83% |
| 10 | E | 76% | 11 | A | 71% | 12 | B | 69% |
| 13 | C | | 14 | E | 44% | 15 | E | 41% |
| 16 | A | | 17 | D | 17% | | | |

## CUMULATIVE EXAMINATION 2

1 a 0.00854 44%

b $\log_{10}$ *(area)* = -14.4 + 0.008 54 x *year* 44%

c i 708 hectares 29%

ii This prediction involves extrapolating beyond the data range. 53%

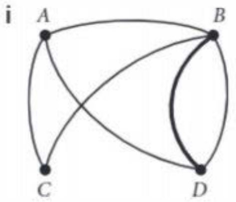
2 a $290.15 b $1927 C $1311 d $293

3 a i birds eat lizards

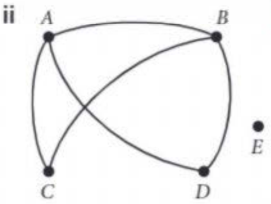
ii no birds, lizards or insects eat birds b Z = $\begin{bmatrix} & I & B & L & F & \\ & 0 & 1 & 1 & 1 & I \\ & 0 & 0 & 0 & 0 & B \\ & 0 & 1 & 0 & 0 & L \\ & 0 & 1 & 1 & 0 & F \end{bmatrix}$

4 a There is no land border between $E$ and any other suburb.

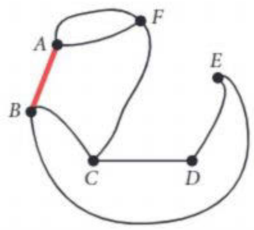
b i




ii 76%




5 a 93%




b 2 88%

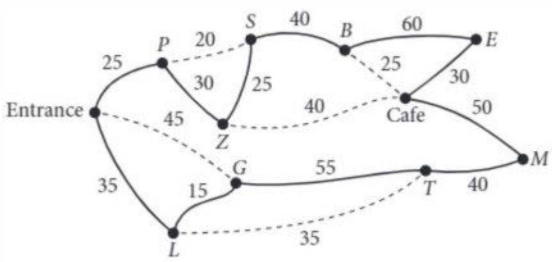
c i A-B-E-D-C or C-D-E-B-A 92%

ii Hamiltonian cycle 60%

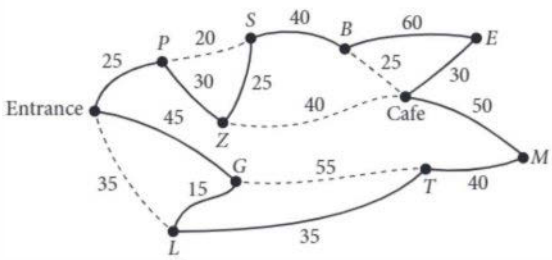
6 a 45 metres

b i Hamiltonian cycle

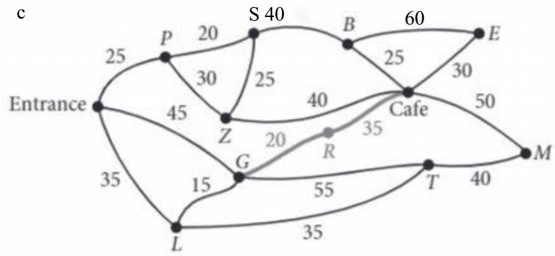
ii *EPZSBECMTGLE* or *EPZSBECMTLGE*




or




c




d 85 metres

7 a 86 km 85%

b $K$ 65%

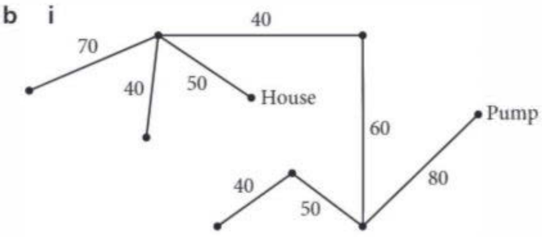
8 a i 160 m

ii 2 vertices

iii 1250 m

iv 8 edges

b i

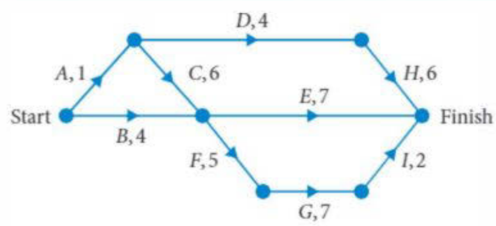



ii minimum spanning tree

## CHAPTER 10

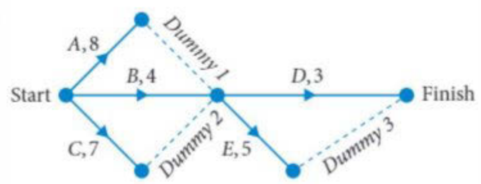
### EXERCISE 10.1

1




2 town $D$ and town $C$

3




4 E

5 B

6 E

7 A

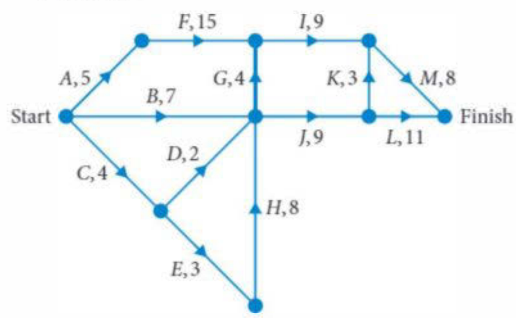
8 C

9 D 10 B

11 a KandF

b $(K)–J–H$, $(K)$-$F$-$J$-$H$ and $(JO$-$M$-$J$-$H$

12



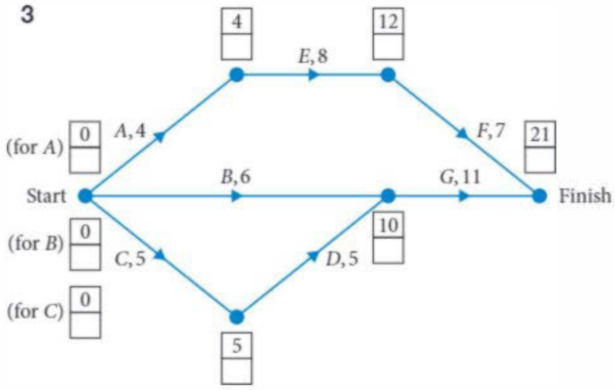



9780170448406

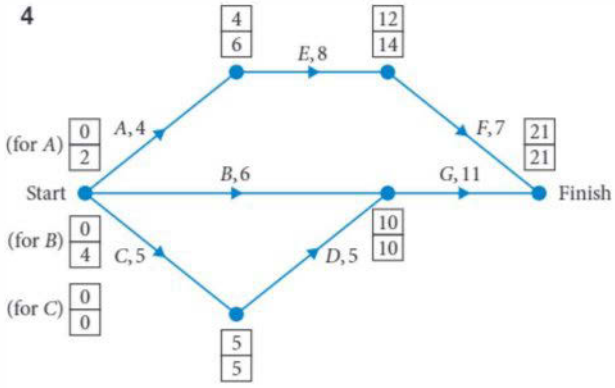
## EXERCISE 10.2

1 E 2 B

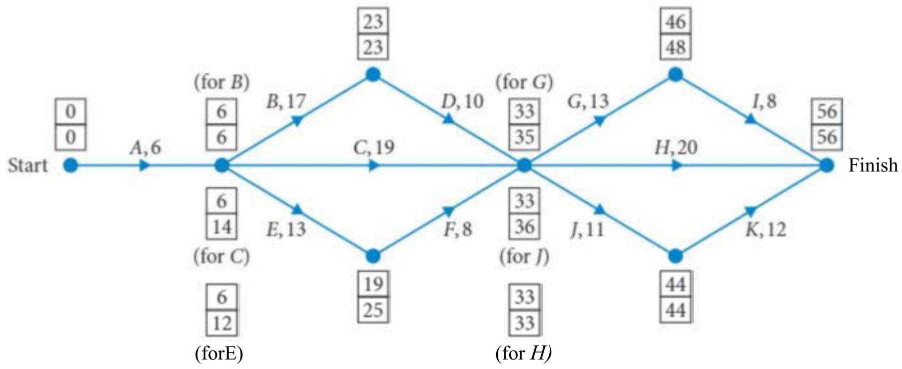
3




4

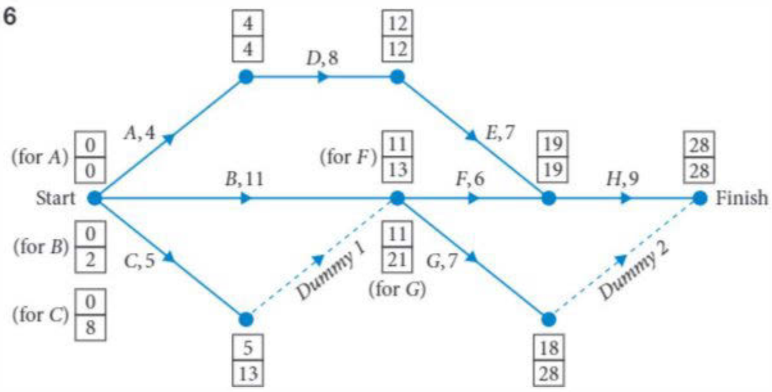



5




Critical path is *A-B-D-J-K*.

6




Critical activities *A, D, E* and *H*

Float for non-critical activities

Float for activity C = 8 - 0 = 8 days

Float for activity *B* = 8 - 0 = 2 days

Float for activity *F* = 13 - 11 = 2 days

Float for activity G = 21-ll = 10 days

7 a critical path *A-E-F,* project time 30 days

b critical path *AEG,* project time 28 days

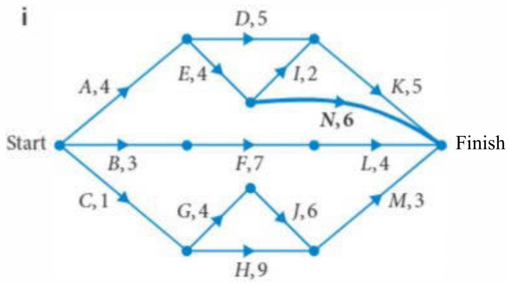
8 C 9 B 10 D 11 C

12 C 13 B 14 D 15 D

16 D 17 B

18a 11 days b *A-E-I-K* c activity *H* d \$2000

e i




ii 9 days

19 a D and *E*

b i *A-E-I-L-N* ii 6 days

c i 17 days ii \$4000

20 a 8 b 12 c / d 29

e *C-0,D- 1,G-2,H*-1,K- 1

21 a activities C and D

b Activity *D* must be on the critical path,

c 17 weeks d 3 weeks e activity *F*

22 a 18 bl c 7

d The *dummy* activity indicates that activity *K* cannot proceed unless activities *H, G* and C are completed.

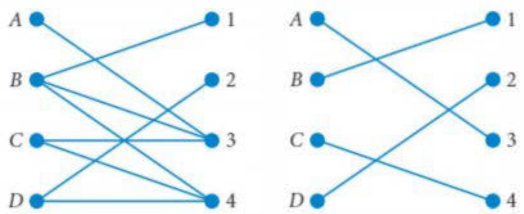
Activities C and *G* are immediate predecessors of *K*.

## EXERCISE 10.3

1 D 2 D

3 task 1 - Christen, task 2 - Cathy, task 3 - Celeste, task 4 - Cheryl

4




Allocation: *A*3, *B*1, *C*4, *D*2

5

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| A | 18 | 7 | 0 | 8 |
| B | 0 | 4 | 2 | 3 |
| C | 8 | 0 | 4 | 0 |
| D | 3 | 0 | 5 | 6 |

Allocation: A3, Bl, C4, D2

Time = 7 + 6 + 5 + 5 = 23 hours

6 *A*1, B3, C2 for a minimum distance of 42 km.

7 B 8 B 9 A 10 C

11 D 12 C 13 B 14 E

15 B 16 C 17 B 18 B

19 a George

b

| Person | Position |
|---|---|
| Harriet | drums |
| Ian | saxophone |
| Keith | keyboard |

20 a 17

b Four lines are required before allocating four tasks to four people and there are only three.

c

| Task | Worker | | | |
|---|---|---|---|---|
| | Julia | Ken | Lana | Max |
| W | 0 | 0 | 4 | 0 |
| *X* | 2 | 2 | 0 | 10 |
| *Y* | 1 | 3 | 0 | 0 |
| *Z* | 7 | 0 | 3 | 5 |

d *W* - Julia, *X* - Lana, *Y* - Max, Z - Ken

21 a

| Student | Sport |
|---|---|
| Blake | tennis |
| Charli | football |
| Huan | basketball |
| Marco | athletics |

b Anita - 400, Imani - 200, Jordan - 100, Lola - 300

22 a 2,0,7, 5

b She is the only child with a zero in the column for Concert 2.

C 3,4,2, lor3,1,2,4 d 56

23 a Colin must plan Tour 2. b 43 minutes

## EXERCISE 10.4

1 C 2 E

3 a 900 vehicles per hour

b 600 vehicles per hour

c 600 vehicles per hour

4 44 5 65 6 C 7 C

8 C 9 C 10 C 11 B

12 B 13 C 14 D 15 B

16 C

17 a CutA = 14, Cut B = 23, Cut C = 12

b Cut *E* does not isolate Bowen from Arlie.

c 12 seats

18 a 37

i *A-B-E-C-D*

ii

| Group | Maximum group size | Path taken from A toD |
|---|---|---|
| 1 | 17 | answered in part b i |
| 2 | 11 | *A-F-E-D* |
| 3 | 7 | *A-G-F-B-C-D* |
| 4 | 2 | *A-B-E-D* |

19 a Stormwater from Source 2 cannot reach Outlet L

b Outlet 1: 700kL/min

Outlet 2: 700kL/min

c 300

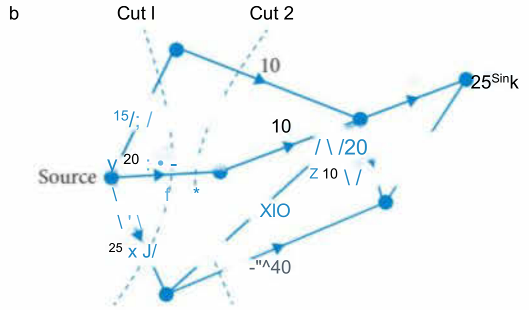
20 a 43 b 22 C 7

21 a 26 trucks b 23 trucks c 8 trucks


9780170448406

22 a 60

b




c 40

## EXERCISE 10.5

1 B 2 D

3 *A-C-F-G-Z* (47)

4 C 5 A 6 D 7 A

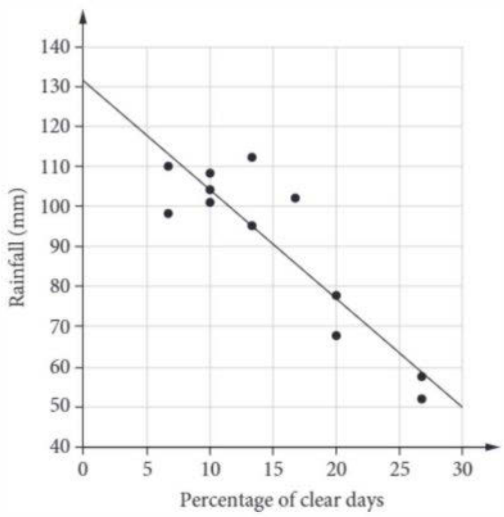
8 B 9 C 10 A

## CUMULATIVE EXAMINATION 1

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | C | 47% | 2 | C | | 3 | A | |
| 4 | A | | 5 | B | | 6 | A | 37% |
| 7 | D | | 8 | B | 65% | 9 | D | |
| 10 | C | | 11 | C | | 12 | E | 61% |
| 13 | C | 60% | 14 | C | 35% | 15 | C | 95% |
| 16 | A | 62% | 17 | A | 24% | 18 | E | 11% |

## CUMULATIVE EXAMINATION 2

1 a




b 37.2 mm

c i 80.81% of the variation in the *rainfall* can be explained by the variation in the *percentage of clear days.*

ii -0.899

2 a 4.8% b $5245.35

c i $T_0 = 3000$, $T_{n+1} = 200 + 1.0035T_n$

ii $74.10

3 a 3x 1

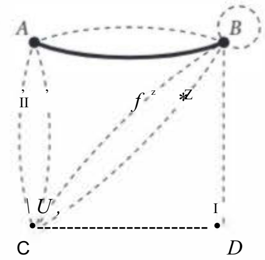
b i [21200] (The brackets must be included)

ii Total cost of all seats in the theatre when every seat is sold.

c Total cost of *B* and C class seats sold.

4 a 7

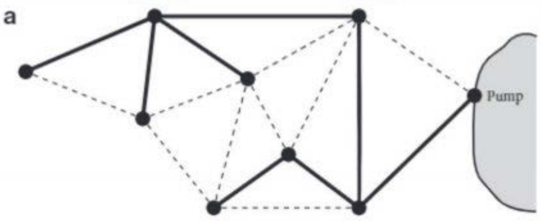
b i




ii For motorist to be able to do this, there needs to be an Eulerian circuit, which can only exist if all vertices have an even degree. In this network, vertices C and *B* have odd degrees.

5 a 14 weeks 46% b 7 22%

6 a




b minimum spanning tree

7 a i 9 90% ii 13 77%

b 7 32%

8 a 10 64%

b *B-E-G-H-J* 76%

c A and C 45%

d end of activity *E* to the start of activity *J* 25%

# Glossary and index

activity (directed graphs) Interconnected steps represented by an edge in a directed graph and shown by a line with an arrow, (p. 627)

activity' table A table showing the order and estimated time for a series of activities, (p. 627)

adjacency matrix A matrix that shows the number of edges between vertices on a graph, (p. 587)

adjacent vertices Two vertices that are connected by one edge. (p. 557)

allocation problem *See* assignment problem,

amortisation table A table that shows the step-by-step calculations of how a loan is reduced, (p. 347)

annuity investment An investment that involves making an initial deposit followed by additional regular payments into an account earning a fixed rate of compound interest, (p. 382)

annuity A type of investment where a sum is invested, interest is compounded at a fixed rate and withdrawals are made at regular intervals, usually until the value of the investment is $0. (p. 369)

appreciation The increase in value of items over time, (p. 284)

asset An item purchased by businesses to help them function, (p. 284)

assignment problem The process of finding the best way to match the elements in two groups, such as a group of workers to a set of tasks, to optimise a stated objective such as minimising cost, distance or time. (p. 649)

back-to-back stem plot A statistical graph used when dealing with two sets of data values for the same variable where the original data values are visible, (p. 98)

balance The value of an investment or loan at any time, (p. 276)

bar chart A graphical display used for categorical data where the frequency of each different category is shown using a vertical column or a horizontal bar. (p. 7)

bell-shaped distribution *See* normal distribution,

bi-modal A distribution with two modes, (p. 9)

binary matrix A matrix where every element is '0' or T. (p. 408)

bipartite graph A graph whose set of vertices can be split into two distinct groups, where vertices cannot connect other vertices from the same group, (p. 649)

birth rate (in a Leslie matrix) The average number of female babies for a female in an age group of an animal population, (p. 529)

bivariate data Data associated with two related variables, (p. 86)

box-and-whisker plot *See* boxplot.

boxplot (box-and-whisker plot) A graphical display of numerical data based on the five-number summary, IQR and outliers, (p. 27)

bridge An edge that keeps a graph connected, (p. 568)

capacity' of a cut The total of all the flow capacities passing across a cut in the direction from source to sink. (p. 664)

categorical variables Variables that represent qualities that cant be counted or measured, (p. 4)

causation A high level of correlation between two variables where one variable is causing a change in the other, (p. 122)

centre of a distribution The single value that best represents the distribution, (p. 9)

centring The extra step of taking two-point moving means of the smoothed values when smoothing with an even number of points, (p. 219)

circuit A walk with no repeated edges that starts and finishes at the same vertex, (p. 577)

coefficient of determination A measure of how useful a line of best fit is as a linear model for a particular set of data. (p. 152)

column matrix A matrix that has just one column, (p. 408)

communication diagram A diagram showing one-way or two-way arrows between points indicating when communication occurs, (p. 455)

communication matrix A square binary matrix where communication is indicated by a T and non-communication is indicated by a 'O', (p. 455)

complete graph A graph where each vertex is connected to every other vertex in the graph, (p. 571)

compound interest Interest that is added to the principal, where the interest for the next time period is calculated using this new balance (p. 298)

compounding period The length of the time period that elapses before interest compounds, (p. 298)

connected graph A graph where there is a path from any vertex to any other vertex, (p. 568)

contingency table *See* two-way frequency table,

continuous numerical variables Numerical variables that can be measured to an increasing level of accuracy, (p. 4)

coordinates A set of values showing an exact position, (p. 228)

correlation coefficient *See* Pearson correlation coefficient,

crashing The process of speeding up a project by employing more resources for critical activities, enabling them to be completed more quickly, (p. 640)

critical path The longest time path between the start and finish, which determines the project completion time, (p. 638)

critical path analysis A step-by-step project management technique that is used to examine every activity in a project and how each affects the project completion time. (p. 635)

culling The process of deleting something from a transition, (p. 518)

cut A line in a network that stops all flow from source to sink (p. 663)

cycle (Leslie matrices) A repeating pattern of state matrices that starts and ends with the initial state matrix, (p. 533)



9780170448406

cycle (networks) A walk with no repeated vertices that starts and finishes at the same vertex, (p. 577)

degree (of a vertex) The number of edges connected to a vertex, (p. 558)

degree sum (of a graph) The sum of the degrees of all the vertices, (p. 558)

depreciation The decrease in value of assets bought by a business over time. (p. 284)

de-seasonalisation The process of using seasonal indices to remove the seasonal component of time series data, using the formula

$$\text{de-seasonalised value} = \frac{\text{actual value}}{\text{seasonal index}} \quad \text{(p. 237)}$$

determinant A number that plays an important role in finding the inverse of a matrix, (p. 441)

diagonal matrix A square matrix where the only non-zero elements are in the leading diagonal, (p. 412)

digraph *See* directed graph.

Dijkstra's algorithm A method used to find the shortest path between two vertices, (p. 595)

directed graph (digraph) A set of vertices connected by edges, where the edges have a direction, (p. 627)

discontinuity (in a time series) A structural change in a time series that is a clear break, (p. 214)

discrete numerical variables Numerical variables that can't be measured to an increasing level of accuracy, (p. 5)

dominance diagram A communication diagram where all the arrows are one-way. (p. 459)

dominance matrix A square binary matrix where dominance is indicated by a T' and other elements are 'O', (p. 459)

dot plot A graphical display where numerical data is represented by dots. (p. 42)

dummy activity An activity of zero time, shown as a directed edge with a broken line, added to a network to ensure that no two vertices are connected by multiple edges or to maintain precedence structure, (p. 629)

earliest start time (EST) The earliest time it is possible to start the activity, (p. 635)

edge The lines connecting the vertices of a graph, (p. 557)

effective interest rate The interest after compounding has been taken into account, which allows us to compare different rates, (p. 311)

element A value in a matrix, (p. 407)

equilibrium state matrix (steady state matrix) A column matrix representing the final state of a transition, (p. 508)

Euler's formula A formula used to test whether a graph is planar:

vertices + faces - edges = 2 (p. 569)

Eulerian circuit A walk with no repeated edges that includes every *edge* in a graph and starts and finishes at the same vertex, (p. 580)

Eulerian trail A walk with no repeated edges that includes every *edge* in a graph, (p. 580)

experimentation A study where the researcher actively manipulates a situation to eliminate other possible variables before observing the results, (p. 122)

explanatory variable A variable that we expect to predict or explain another variable, (p. 86)

extrapolation A prediction made outside the original data range, (p. 163)

face A region bound by edges. The region surrounding a planar graph is also considered a face. (p. 558)

finance solver An application that solves finance problems, (p. 336)

five-number summary Five key points in a data distribution consisting of the minimum value, the lower quartile, the median, the upper quartile and the maximum value, (p. 24)

five-point moving mean smoothing A numerical smoothing technique that involves finding the means of consecutive sets of five data points, (p. 219)

five-point moving median smoothing A graphical smoothing technique that involves finding the medians of consecutive sets of five data points, (p. 229)

flat rate depreciation Depreciation where the future value of an asset is reduced by a fixed amount every year, expressed either in dollars or as a fixed percentage of the purchase price, (p. 284)

float time The maximum time an activity can be extended or postponed without affecting the project completion time. Activities on the critical path all have float times of zero, (p. 639)

four-point moving mean smoothing A numerical smoothing technique that involves finding the means of consecutive sets of four data points, (p. 221)

frequency table A table used to organise large amounts of data with data values in one column and the corresponding frequencies in another, (p. 6)

future value The new reduced value of an asset being depreciated at any point in time or the balance of loans and investments at any point in time. (p. 284)

gradient See slope.

graph (network diagram) A diagram that consists of a set of points, called vertices, that are joined by a set of lines, called edges, (p. 557)

graphical smoothing A smoothing technique that involves working directly from a time series plot. (p. 227)

grouped frequency table A frequency table where numerical data has been grouped into regular intervals, (p. 16)

Hamiltonian cycle A walk with no repeated vertices that includes every *vertex* in a graph and starts and finishes at the same vertex, (p. 581)

Hamiltonian path A walk with no repeated vertices that includes every *vertex* in a graph, (p. 581)

histogram A graphical display of data from a grouped frequency table, (p. 16)

Hungarian algorithm A method used to find the optimum allocation for the assignment problem, (p. 651)

identity matrix A square matrix where all the elements in the leading diagonal are '1' and the other elements are 'O', (p. 412) immediate predecessor An activity that must be completed before another activity can commence, (p. 627)

inflow capacity The total flow capacity entering a vertex, (p. 663)

initial state matrix A column matrix representing the starting state of a transition, (p. 495)

intercept The y-intercept in a line of best fit. (p. 150)

interest The fee for using someone else's money, (p. 276)

interest-only loan A reducing balance loan where the payments are exactly equal to the interest, (p. 350)

interpolation A prediction made within the original data range, (p. 163)

interquartile range (IQR) The measure of the spread of the middle 50% of the data values, (p. 27)

inverse matrix The matrix that when multiplied by another matrix results in the identity matrix, (p. 440)

irregular fluctuation (in time series) Time series data that appears to occur at random with no pattern, (p. 212)

irregular variation *See* irregular fluctuation,

isolated vertex A vertex that has no edges connected to it. (p. 558)

isomorphic graphs Graphs showing exactly the same connections, (p. 557)

latest start time (LST) The latest time it is possible to start an activity without affecting the project completion time, (p. 637)

leading diagonal The diagonal in a matrix running from the upper left to the lower right, (p. 411)

least squares line of best fit (least squares regression line) The method of finding a line of best fit that minimises the sum of the squares of the vertical distances between the line and each data point in a scatterplot. (p. 142)

least squares line *See* least squares line of best fit.

least squares regression line *See* least squares line of best fit.

Leslie matrix A matrix containing female birth and survival rates of different age groups of an animal population, (p. 529)

lifespan The maximum time an animal can live. (p. 529)

line of best fit A straight line that is the best approximation for a set of data. (p. 142)

linear scale A scale used on a plot or graph where the same number is *added* to move from one scale mark to the next, (p. 36)

linearisation The process of putting something into a straight-line form. (p. 181)

log scale A scale used on a plot or graph where the same number is multiplied to move from one scale mark to the next. (p. 36)

logarithmic transformation (or log transformation) A transformation involving finding the log of either the $x$ values or they values, (p. 182)

logarithmic scale *See* log scale.

loop An edge that starts and finishes on the same vertex, (p. 558)

lower fence The value below which a data point may be considered an outlier, (p. 27)

lower quartile The data point that has 25% of the data below it. (p. 24)

lower triangular matrix A square matrix where all the elements above the leading diagonal are 'O', (p. 412)

matrices Plural of matrix.

matrix A rectangular arrangement of numbers organised into rows and columns, usually presented in square brackets, (p. 407)

matrix multiplication The multiplication of a matrix by another matrix, (p. 429)

matrix order equation An equation involving the orders of matrices, (p. 430)

maximum flow' The capacity of the minimum cut. (p. 663)

mean A measure of the centre of a distribution calculated by dividing the sum of all the data values by the total number of values, (p. 9)

median The middle value or the average of the two middle values when data is ordered from smallest to largest, (p. 9)

minimum connector The path connecting all the vertices in a weighted graph with the smallest total weight, (p. 604)

minimum spanning tree The spanning tree with the smallest total weight, (p. 604)

modal category The most frequently occurring category, (p. 9)

modal interval The interval with the highest frequency, (p. 16)

mode The most frequently occurring data value, (p. 9)

moving means A numerical smoothing technique that involves finding a series of means of a fixed number of data points, (p. 218)

moving medians The graphical smoothing technique which involves finding a series of medians of a fixed number of data points, (p. 227)

multiple edges More than one edge connecting two vertices, (p. 558)

negatively skewed distribution Description of a distribution that has a tail at the lower end. (p. 17)

network A group of interconnected people or things, (p. 557)

network diagram *See* graph (network diagram),

nominal interest rate The interest given for an investment or loan which consists of a rate per year and a compounding period, (p. 311)

nominal variables Categorical variables that have no natural order, (p. 5)

normal distribution (bell-shaped distribution) A distribution with a bell shape which is symmetrical about the mean, peaks in the centre and tails off towards zero on both sides, (p. 58)

numerical data Variables represented by quantities or measurements that can be either discrete or continuous, (p. 16)

numerical smoothing A smoothing technique that involves numbers and calculations, (p. 218)

numerical variables Variables that represent quantities that can be counted or measured, (p. 4)


9780170448406

observation A way that data can be gathered by looking at an existing situation, (p. 122)

one-step communication A direct communication between *A* and B. (p. 455)

one-step dominance score The sum of a dominance matrix row. (p. 459)

one-step dominance A direct dominance where A dominates B. (p. 459)

one-way communication A communication where *A* can communicate with B, but B can't communicate with *A*. (p. 456)

order of a matrix The number of rows and columns in a matrix, (p. 407)

ordered stem plots A stem plot where the leaves are ordered from smallest to largest, (p. 50)

ordinal variables Categorical variables that have a natural order, (p. 5)

outflow capacity The total flow capacity leaving the vertex, (p. 663)

outlier An extreme high or low value in the data. (p. 18)

parallel boxplot A statistical graph where two or more boxplots are shown on the same axis. (p. 101)

parallel dot plot A graph where two or more dot plots are shown on the same axes. (p. 100)

parallel percentage segmented bar charts A graph where two or more percentage segmented bar charts are shown on the same axes. (p. 91)

path A walk with no repeated vertices, (p. 577)

Pearson correlation coefficient A number between -1 and 1 that measures the strength and direction of *linear* associations, (p. 119)

per annum (p.a.) per year (p. 276)

percentage segmented bar chart A segmented bar chart where the segments represent percentages rather than frequencies and the length of the bar is 100%. (p. 7)

percentage two-way frequency table A two-way frequency table where the data values have been converted into percentages, (p. 89)

percentaging The process of converting values into percentages, (p. 89)

permutation matrix A square matrix where every row and every column has exactly one T, with zeros everywhere else. (p. 408)

perpetuity A type of annuity where a permanently invested amount of money provides regular payments that continue forever and the balance of the amount invested stays the same forever, (p. 378)

perpetuity investment *See* perpetuity.

planar graph A connected graph that can be drawn without any edges crossing over. (p. 569)

population All items of the group being studied, (p. 12)

positively skewed distribution Description of a distribution that has a tail at the upper end. (p. 17)

present value The current value of an asset, loan or investment, (p. 337)

Prim's algorithm A method used to determine the minimum spanning tree. (p. 605)

principal The amount of money invested or borrowed, (p. 276)

quartiles The three points that divide a set of data into quarters, (p. 24)

range The measure of the spread of a distribution found by subtracting the smallest data value from the largest data value, (p. 10)

ratio The relationship between two numbers indicating how much of one quantity there is compared to the second quantity, (p. 535)

raw data Data that has been collected but not organised, (p. 6)

reachability The ability to get from one vertex to another vertex in a directed graph, (p. 629)

reachable A vertex is reachable if it is connected to another vertex by one or more edges, (p. 629)

reciprocal transformation A transformation involving taking the reciprocal of either the $x$ values or the $y$ values, (p. 183)

recurrence relation A rule that generates a sequence by connecting each value to previous values, (p. 268)

recursion *See* recursive computation.

recursive computation Calculations that continually use the previous answer to find the next answer, (p. 268)

reducing balance depreciation Depreciation where the future value of an asset is reduced every year by a fixed percentage of its value in the preceding year. (p. 315)

reducing balance loan A loan where interest is calculated on the amount still owing after each repayment is made, (p. 344)

redundant links Communication links where the sender and receiver are the same. (p. 455)

regular transition matrix A transition matrix with no zeros or whose powers have no zeros, (p. 507)

re-seasonalisation The process of using seasonal indices to find the original non-seasonalised value using the formula

actual value = de-seasonalised value x seasonal index, (p. 240)

residual The vertical distance between each data point and the least squares line of best fit. (p. 167)

residual plot A plot with the explanatory variable on the x-axis and the residual values on they-axis. (p. 171)

response variable A variable whose changes we expect to be predicted or explained by another variable, (p. 86)

restocking The process of adding something to a transition, (p. 518)

round-robin tournament A competition where every participant plays every other participant once. (p. 459)

row matrix A matrix that has just one row. (p. 408)

scalar A number that is not in a matrix, (p. 422)

scalar multiplication The multiplication of every element in a matrix by the same number, (p. 422) scatterplot A graph used to compare two numerical variables where the explanatory variable is plotted on the x-axis and the response variable on they-axis. (p. 110)

seasonal adjustment An adjustment made to time series data to eliminate seasonality, (p. 236)

seasonal indices Values used to make seasonal adjustments to time series data. (p. 236)

seasonality' Time series data that has regular and predictable changes repeated across a year or less. (p. 212)

segmented bar chart A graphical display involving one bar with several segments, where each segment represents the frequency of a category, (p. 7)

sequence A list of numbers separated by commas, (p. 268)

shape of a distribution A description of data as symmetrical, positively skewed or negatively skewed, (p. 17)

shortest path The path between two vertices of a graph with the smallest weight, (p. 594)

significant figures (rounding to a number of) A method of rounding involving all the non-zero digits of a number plus the zeros that are included between them or that are final zeros and signify accuracy, (p. 34)

simple graph A graph that contains no loops or multiple edges, (p. 571)

simple interest The fixed amount of interest paid at regular time periods calculated as a percentage of the amount of money invested or borrowed, (p. 276)

singular matrix A square matrix that has no inverse, (p. 443)

sink The final vertex of a flow network, (p. 663)

slope (gradient) (of the least squares line of best fit) The coefficient $b = r\frac{s_y}{s_x}$ in the equation for the least squares line of best fit $y = a + bx$. (p. 142)

slope (gradient) The measure of the steepness of a line, (p. 142)

smoothing A technique for levelling out fluctuations in time series data to produce a smoother graph, which allows us to see trends more clearly, (p. 218)

source The first vertex of a flow network, (p. 663)

spanning tree A tree subgraph that includes all the vertices of the original graph, (p. 602)

spread of a distribution A measurement of how much data varies around the centre of a distribution, (p. 10)

square matrix A matrix that has the same number of rows as columns, (p. 408)

squared transformation A transformation involving squaring either the x values or they values, (p. 181)

standard deviation A measurement of the spread of data about the mean. (p. 52)

standardise The process of converting data values from normal distributions to standardised values, (p. 66)

standardised values (z-scores) Values calculated using the formula

$$\text{standardised value} = \frac{\text{data value - mean}}{\text{standard deviation}}$$

that allow us to compare values from different normal distributions, (p. 64)

state matrix A column matrix representing a state at a point in time of a transition, (p. 495)

state A condition or a location at a point in time. (p. 486)

steady state matrix *See* equilibrium state matrix,

stem plot (stem-and-leaf plot) A graphical display where actual data values appear, (p. 43)

stem-and-leaf plot *See* stem plot.

structural change (in statistics) An unexpected shift in the data pattern of a time series, (p. 214)

subdiagonal The diagonal immediately under the leading diagonal, (p. 411)

subgraph A graph that is part of a larger graph, (p. 571)

summing matrix A row and column matrix consisting entirely of 'Is,' which give the sum of the elements when multiplied with other matrices, (p. 408)

survival rate (in a Leslie matrix) The chances a female in an age group of an animal population will live to be the age of the next age group, (p. 529)

symmetric distribution A distribution that is symmetric in shape, (p. 17)

symmetric matrix A square matrix where the elements are symmetric with respect to the leading diagonal, (p. 412)

term (of a sequence) The name for a number in a sequence, (p. 268)

three-point moving mean smoothing A numerical smoothing technique that involves finding means of consecutive sets of three data points (p. 219)

three-point moving median smoothing A graphical smoothing technique that involves finding medians of consecutive sets of three data points, (p. 229)

three-step communication A communication between A and B, where *A* can communicate with D, which can then communicate with C, which can then communicate with B. (p. 457)

time series Bivariate data where the explanatory variable is time measured at equally spaced intervals, (p. 209)

time series plot A scatterplot of time series data where the data points are joined by straight lines, (p. 209)

total dominance score The sum of a dominance matrix row plus the sum of the equivalent row of its square, (p. 459)

trail A walk with no repeated edges, (p. 577)

transformation A way of changing a non-linear association so that the association between the two variables becomes closer to a straight line. (p. 181)

transition diagram A diagram that shows transitions from one state to another using arrows with corresponding percentages, (p. 486)

transition matrix A square matrix that shows a change from one state to another, where the change follows the same rules each time. (p. 486)

transition A change from one state to another, (p. 486)

transpose (of a matrix) A new matrix formed by turning all the rows of the original matrix into columns and vice-versa, (p. 409)

tree A connected graph with no loops, multiple edges or cycles, (p. 602) trend The long-term direction of time series data. (p. 211)

trend line A line of best fit for time series, (p. 245)

trend line forecasting Using a trend line fitted to a time series to make predictions about the future, (p. 245)

two-point moving mean smoothing The process of finding means of consecutive pairs of data points used as an extra step when smoothing with an even number of points, (p. 221)

two-step communication A communication between *A* and B, where *A* can communicate with C, which can then communicate with B. (p. 455)

two-step dominance An indirect dominance of *A* over B, where *A* dominates C, which dominates B. (p. 459)

two-way communication A communication where *A* can communicate with B, and B can communicate with *A*. (p. 456)

two-way frequency table (contingency table) A frequency table used to explore the association between two categorical variables, each with at least two categories, (p. 88)

two-way table A table used to display information when comparing two categorical variables, (p. 90)

undirected graph A graph where there are no directions associated with the edges, (p. 557)

unit cost depreciation Depreciation where the future value of an asset is reduced every year according to the amount of use it has had, not according to its age. (p. 292)

unit matrix *See* identity matrix.

upper fence The value above which a data point may be considered an outlier, (p. 27)

upper quartile The data point that has 75% of the data below it. (p. 24)

upper triangular matrix A square matrix where all the elements below the leading diagonal are ‘O’, (p. 412)

value (of a sequence) *See* term (of a sequence),

variability How much the data points differ from each other, (p. 212)

variable A letter or symbol used to represent a quantity that can have many different values in a particular situation, (p. 4)

vertex The point on a graph or network diagram.

vertices + faces - edges = 2 (p. 557)

vertices Plural of vertex.

walk A sequence of connected vertices, (p. 577)

weight (of a graph) A number attached to an edge on a graph that represents physical quantities such as distance, time and cost. (p. 594)

weighted graph A graph whose edges are labelled with numbers representing physical quantities such as distance, time and cost. (p. 594)

y-intercept The *y* value of a point where the line crosses they-axis. (p. 142)

z-scores *See* standardised values.

zero matrix A matrix where all the elements are ‘O’, (p. 408)