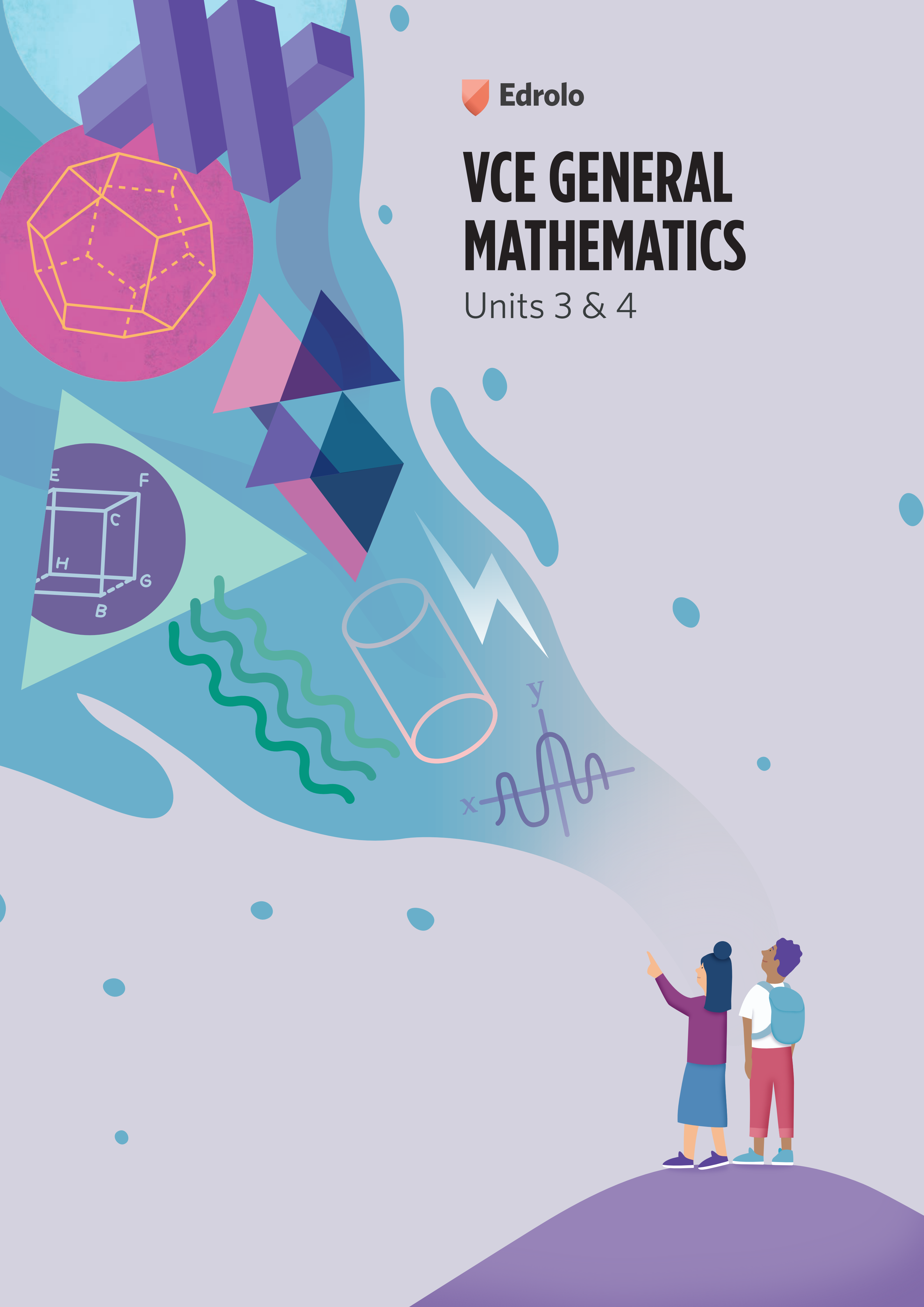
Edrolo

VCE GENERAL MATHEMATICS

Units 3 & 4



Edrolo

VCE GENERAL MATHEMATICS

Units 3 & 4

Robert Borg, Duyen Duong, James Boyce, Zephyr Howson, Sophie Watt,
Clinton Bouphasavanh, Joshua Clements, Victoria Flynn, Nina Miriyagalla,
Angus Plowman, Talia Scott-Hayward, Justin Tan, Ying Qin, Patrick Robertson

Need help?
Email our School Support team at help@edrolo.com.au
Or call 1300 EDROLO | 1300 337 656

Edrolo

At Edrolo, we're transforming the way the students learn and teachers teach.

Our mission is simple: to improve education.

PUBLISHED IN AUSTRALIA BY Edrolo
321 Exhibition Street Melbourne VIC 3000, Australia

Ref: 1.1.1

National Library of Australia Cataloguing-in-Publication data

TITLE: Edrolo VCE General Mathematics Units 3 & 4
CREATOR: Edrolo et al.
ISBN: 978-1-922901-01-9
TARGET AUDIENCE: For secondary school age.
SUBJECTS: General Mathematics--Study and teaching (Secondary)--Victoria
General Mathematics--Victoria--Textbooks.
General Mathematics--Theory, exercises, etc.
OTHER CREATORS/CONTRIBUTORS: Daniel Tram, Hannah Liu, Simon Hamlet, Odette Mawal, Irene Platis, James Vella, James Wallace

LAYOUT DESIGN: Emma Wright and Edrolo
TYPESET BY: Emma Wright, Arslan Khan, Belle Gibson, Esra Yang, Dean Dragonetti
COVER DESIGN BY: Cat MacInnes

Printed in Australia by Ligare Printing Pty Ltd

The paper this book is printed on is in accordance with the standards of the Forest Stewardship Council®. The FSC® promotes environmentally responsible, asocially beneficial and economically viable management of the world's forests.

Proudly Printed In Australia

CONTENTS

Back to contents

FEATURES OF THIS BOOK

Edrolo's VCE General Mathematics Units 3 & 4 product has the following features.

Textbook theory

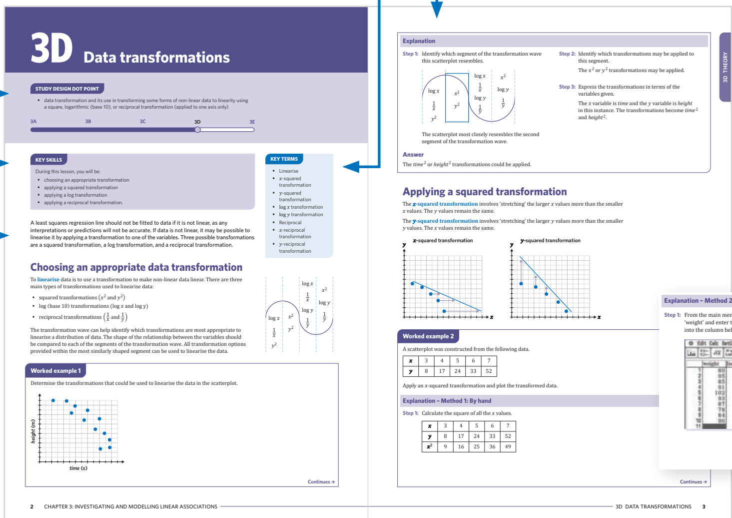
Study design dot points provide explicit links between the content covered in each lesson and the VCAA curriculum.

Key terms identify newly defined mathematical terminology and provide a reference for navigating glossary definitions.

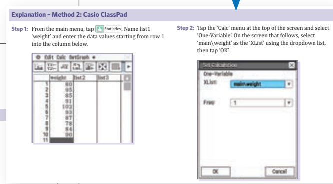
Worked examples provide fully stepped out exemplar solutions.

Key skills break the theory down into smaller chunks that focus on only one skill at a time, with key skill headings replicated throughout the theory, questions, and answers for easy navigation.

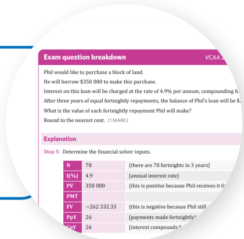
Introductions provide a launchpad for the lesson and serve to give context for the theory.



Calculator methods with screenshots step students through using the 'TI-Nspire' and 'Casio ClassPad' CAS calculators.



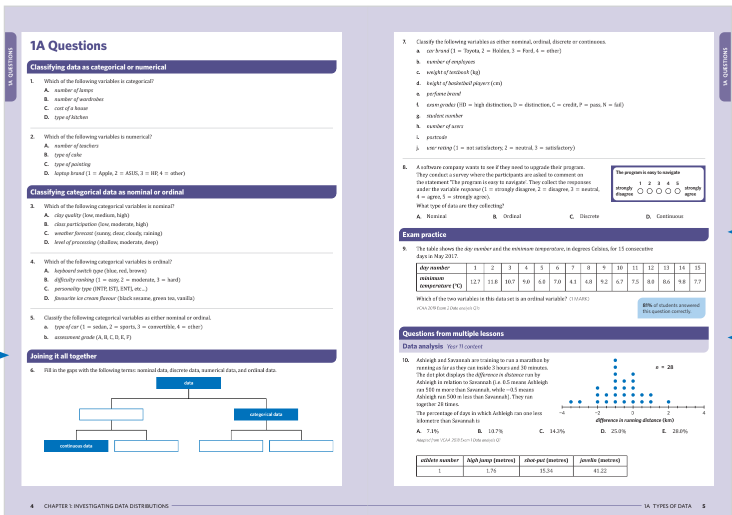
Exam question breakdowns provide an extra level of support by stepping through past exam questions, including the percentage of students who answered the question correctly, as well as common misconceptions and errors made, based on VCAA statistics.



Textbook questions

Key skills questions link to key skills in the theory and ask students to apply only one skill at a time.

Joining it all together questions scaffold students to link multiple skills from the lesson together.

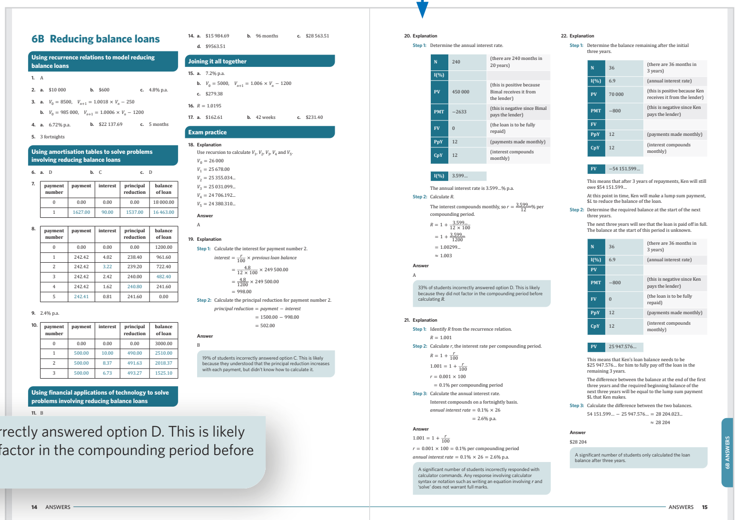


Exam practice questions provide students with past VCAA exam questions to get them ready for exams.

Questions from multiple lessons provide ongoing revision from a range of topics.

Textbook answers

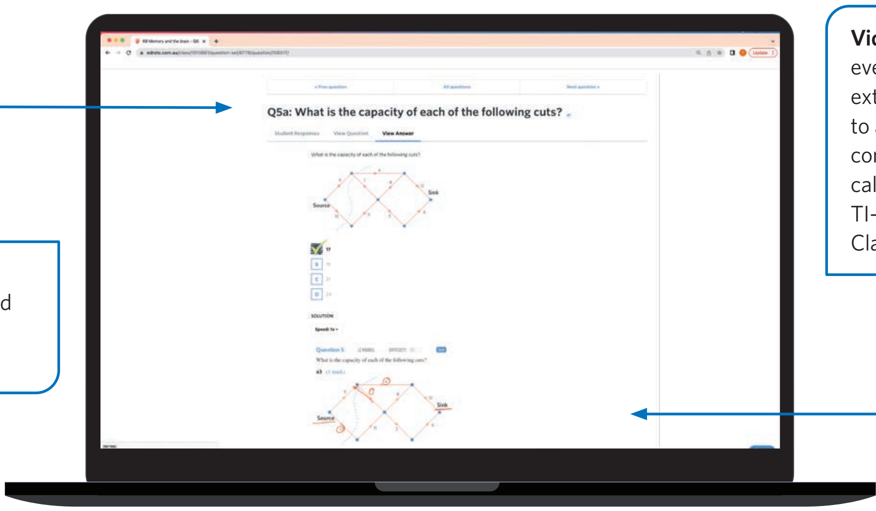
Fully worked solutions are provided for all exam practice questions, complete with commentary on common misconceptions and errors made, based on VCAA statistics (where applicable).



33% of students incorrectly answered option D. This is likely because they did not factor in the compounding period before calculating *R*.

Online – Other resources

Question sets provide the ability to complete all questions online, with instant feedback on student responses.



Video solutions for every question provide extra guidance on how to answer questions, complete with guided calculator solutions for TI-Nspire and Casio ClassPad CAS calculators.

Static solutions provide fully worked solutions for all questions.

Chapter reviews provide visual theory summaries and application questions that scaffold students towards answering questions using multiple skills within the chapter.

Area of study reviews provide teachers with a practice assessment that links concepts from an entire area of study.

Back to contents

AOS 1

Data analysis

CALCULATOR QUICK LOOK-UP GUIDE

Back to contents

1

CHAPTER 1

Investigating data distributions

LESSONS

KEY KNOWLEDGE

- 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 ($Q_1 - 1.5 \times IQR$) and upper fence ($Q_3 + 1.5 \times IQR$) 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.

Image: ChristianChan/Shutterstock.com

Back to contents

1A Types of data

STUDY DESIGN DOT POINT

- types of data

1A 1B 1C 1D 1E 1F 1G 1H 1I

KEY SKILLS

During this lesson, you will be:

- classifying data as categorical or numerical
- classifying categorical data as nominal or ordinal
- classifying numerical data as discrete or continuous.

KEY TERMS

- Data
- Categorical data
- Numerical data
- Nominal data
- Ordinal data
- Discrete data
- Continuous data

In the Information Age, data is becoming increasingly more important to everyday life. Classifying data into data types is necessary before analysis can be performed, or the most appropriate data visualisations can be constructed.

Classifying data as categorical or numerical

Data is a set of values, words or responses, that is collected and ordered by variables.

Data that can be organised into categories or groups is known as **categorical data**. It is also referred to as qualitative data, as it represents a quality or attribute.

Data that can be counted or measured is known as **numerical data**. It is also referred to as quantitative data, as it represents a quantity.

Worked example 1

Classify the following variables as either categorical or numerical.

a. *type of pasta*

Explanation

The variable *type of pasta* is categorised into different pasta types such as gnocchi, fettuccine, spaghetti or lasagne.

Answer

Categorical

b. *number of candles*

Explanation

The variable *number of candles* is counted.

Answer

Numerical

Back to contents

Classifying categorical data as nominal or ordinal

Categorical data can be further classified as either nominal or ordinal.

Categorical data that cannot be sorted into a logical ordered list or hierarchy is called **nominal data**. For example, *type of bread* (white bread, multigrain, sourdough) has no inherent ranking system and is classified as nominal categorical data.

Categorical data that can be ordered into a logical ordered list or hierarchy is called **ordinal data**. For example, *drink size* (small, medium, large) can be ordered such that medium is greater than small, and large is greater than medium. This is an inherent ranking system, so it is classified as ordinal categorical data.

Worked example 2

Classify the following categorical variables as either nominal or ordinal.

a. *type of shoe* (runners, boots, sandals, slides)

Explanation

The categories within the variable *type of shoe* cannot be inherently ordered.

Answer

Nominal

b. *shirt size* (small, medium, large)

Explanation

The categories within the variable *shirt size* can be inherently ordered (small to medium to large).

Answer

Ordinal

Classifying numerical data as discrete or continuous

Numerical variables can be further classified as either discrete or continuous.

Numerical data that can only consist of a set of fixed values within a range is called **discrete data**. Discrete data usually consists of whole numbers and would typically be collected by counting. For example, the *number of steps* taken in a day can only be represented by whole numbers starting from zero, and is classified as discrete numerical data.

Numerical data that can consist of any value within a range is called **continuous data**. Continuous data usually consists of both whole numbers and decimals and would typically be collected by measuring. For example, the *distance* (km) walked in a day is classified as continuous numerical data as it is measured and can consist of any positive value, such as 5.1, 5.01 or even 5.001. Continuous data that has been rounded to the nearest whole number is still considered to be continuous.

Worked example 3

Classify the following numerical variables as either discrete or continuous.

a. *length* (m)

Explanation

The variable *length* (m) can be expressed in decimals and can consist of any value measured on a continuous scale.

Answer

Continuous

b. *number of tennis racquets*

Explanation

The variable *number of tennis racquets* cannot be expressed in decimals and can only be counted.

Answer

Discrete

Exam question breakdown

VCAA 2016 Exam 1 Data analysis Q2

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.

Explanation

Step 1: Classify the variable *blood pressure* (low, normal, high).

The variable *blood pressure* has three categories, low, medium and high. As such, this is a categorical variable.

These categories can be sorted into ascending or descending order. Therefore, *blood pressure* (low, normal, high) can be further classified as an ordinal variable.

Step 2: Classify the variable *age* (under 50 years, 50 years or over).

The variable *age* has two categories; 'under 50 years' and '50 years or over'. As such, this is a categorical variable.

These categories can also be sorted into ascending or descending order. Therefore, *age* (under 50 years, 50 years or over) can be further classified as an ordinal variable.

Answer

B

31% of students answered this question correctly.

45% of students incorrectly chose option D, as they identified the variable *age* (under 50 years, 50 years or over) as a nominal variable. The variable *age* is ordinal since one group of people can be classified as younger than the other group, creating an inherent order between the two categories.

Back to contents

1A Questions

Classifying data as categorical or numerical

1. Which of the following variables is categorical?

A. *number of lamps*

B. *number of wardrobes*

C. *cost of a house*

D. *type of kitchen*

2. Which of the following variables is numerical?

A. *number of teachers*

B. *type of cake*

C. *type of painting*

D. *laptop brand* (1 = Apple, 2 = ASUS, 3 = HP, 4 = other)

3. Classify the following variables as either categorical or numerical.

a. *age*

b. *exam difficulty* (1 = easy, 2 = medium, 3 = hard)

Classifying categorical data as nominal or ordinal

4. Which of the following categorical variables is nominal?

A. *clay quality* (low, medium, high)

B. *class participation* (low, moderate, high)

C. *weather forecast* (sunny, clear, cloudy, raining)

D. *level of processing* (shallow, moderate, deep)

5. Which of the following categorical variables is ordinal?

A. *keyboard switch type* (blue, red, brown)

B. *difficulty ranking* (1 = easy, 2 = moderate, 3 = hard)

C. *personality type* (INTP, ISTJ, ENTJ, etc...)

D. *favourite ice cream flavour* (black sesame, green tea, vanilla)

6. Classify the following categorical variables as either nominal or ordinal.

a. *type of car* (1 = sedan, 2 = sports, 3 = convertible, 4 = other)

b. *assessment grade* (A, B, C, D, E, F)

Classifying numerical data as discrete or continuous

7. Which of the following numerical variables is discrete?

A. *time elapsed*

B. *height*

C. *number of keyboards*

D. *volume of* CO_2 *output*

Back to contents

8. Which of the following numerical variables is continuous?

A. *student enrolments*

B. *tennis tournaments won*

C. *number of dogs*

D. *bone mass*

---

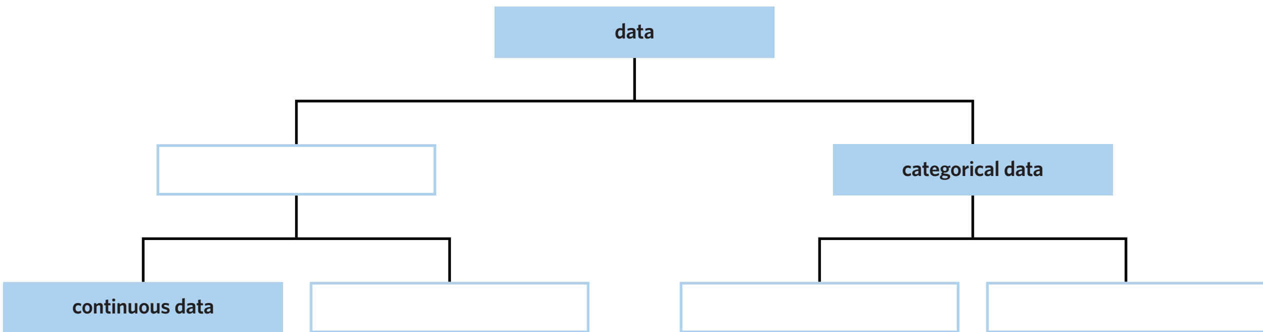
9. Classify the following numerical variables as either discrete or continuous.

a. The *number of parrots* found in different rainforests.

b. The *haemoglobin count* of a group of people, in (g/dl).

Joining it all together

10. Fill in the gaps with the following terms: nominal data, discrete data, numerical data, and ordinal data.



---

11. Classify the following variables as either nominal, ordinal, discrete or continuous.

a. *car brand* (1 = Toyota, 2 = Holden, 3 = Ford, 4 = other)

b. *number of employees*

c. *weight of textbook* (kg)

d. *height of basketball players* (cm)

e. *perfume brand*

f. *exam grades* (HD = high distinction, D = distinction, C = credit, P = pass, N = fail)

g. *student number*

h. *number of users*

i. *postcode*

j. *user rating* (1 = not satisfactory, 2 = neutral, 3 = satisfactory)

---

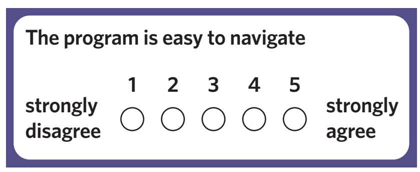
12. A tennis coach collected data on the *number of tennis racquets used* and the *serve speed* (km/h) for several tennis players for the upcoming Australian Open.

a. Which of the two variables is continuous?

b. Which of the two variables is discrete?

---

13. A software company wants to see if they need to upgrade their program. They conduct a survey where the participants are asked to comment on the statement 'The program is easy to navigate'. They collect the responses under the variable *response* (1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree).



What type of data are they collecting?

A. Nominal

B. Ordinal

C. Discrete

D. Continuous

Back to contents

Exam practice

14. The table shows the *day number* and the *minimum temperature*, in degrees Celsius, for 15 consecutive days in May 2017.

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

Which of the two variables in this data set is an ordinal variable? (1 MARK)

VCAA 2019 Exam 2 Data analysis Q1a

81% of students answered this question correctly.

---

15. 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 variables is

A. 1 **B.** 2 **C.** 3
D. 4 **E.** 5

VCAA 2020 Exam 1 Data analysis Q7

69% of students answered this question correctly.

---

16. 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	40.53
5	1.87	13.78	40.62
6	1.73	14.50	45.62
7	1.68	15.08	42.33
8	1.82	13.13	40.88
9	1.83	14.22	39.22
10	1.87	13.62	42.51
11	1.87	12.01	42.75
12	1.80	12.88	38.12
13	1.83	12.68	42.65
14	1.87	12.45	41.32
15	1.78	11.31	42.88

Write down the number of numerical variables in the table. (1 MARK)

VCAA 2021 Exam 2 Data analysis Q1a

52% of students answered this question correctly.

Back to contents

17. The variables *number of moths* (less than 250, 250–500, more than 500) and *trap type* (sugar, scent, light) 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.

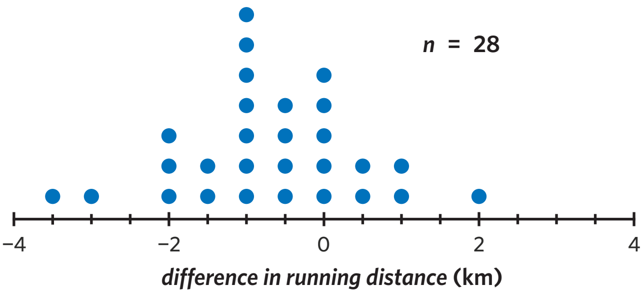
VCAA 2017 Exam 1 Data analysis Q7

46% of students answered this question correctly.

Questions from multiple lessons

Data analysis *Year 11 content*

18. Ashleigh and Savannah are training to run a marathon by running as far as they can inside 3 hours and 30 minutes. The dot plot displays the *difference in distance* run by Ashleigh in relation to Savannah (i.e. 0.5 means Ashleigh ran 500 m more than Savannah, while −0.5 means Ashleigh ran 500 m less than Savannah). They ran together 28 times.



The percentage of days in which Ashleigh ran one less kilometre than Savannah is

A. 7.1% **B.** 10.7% **C.** 14.3% **D.** 25.0% **E.** 28.0%

Adapted from VCAA 2018 Exam 1 Data analysis Q1

Recursion and financial modelling *Year 11 content*

19. Arthur gets \$1000 for his birthday and wants to save his money. He opens a savings account and deposits his \$1000. The account earns interest at a rate of 3% per annum, compounding annually.

Let V_n be the value of Arthur's account n years after he initially deposits his money.

The expected growth of Athur's savings account can be modelled by

A. $V_0 = 1000, \quad V_{n+1} = V_n + 30$

B. $V_1 = 1030, \quad V_{n+1} = V_n + 30$

C. $V_0 = 1000, \quad V_{n+1} = 1.03\,V_n$

D. $V_0 = 1030, \quad V_{n+1} = 1.03\,V_n$

E. $V_1 = 1000, \quad V_{n+1} = 1.3\,V_n$

Adapted from VCAA 2015 Exam 1 Number patterns Q3

Data analysis *Year 11 content*

20. The number of cars that park in a particular car park on each day of one week are counted and recorded in the following table.

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
number of cars	103	84	92	79	93	64	48

From the information given, determine

a. the range. (1 MARK)

b. the percentage of days that had less than 90 cars parked, correct to one decimal place. (1 MARK)

Adapted from VCAA 2017 Exam 2 Data analysis Q1

Back to contents

1B Displaying and describing categorical data

STUDY DESIGN DOT POINTS

- representation, display and description of the distributions of categorical variables: data tables, two-way frequency tables and their associated segmented bar charts
- use of the distribution(s) of one or more categorical or numerical variables to answer statistical questions

1A 1B 1C 1D 1E 1F 1G 1H 1I

KEY SKILLS

During this lesson, you will be:

- constructing frequency tables
- constructing bar charts
- constructing segmented bar charts
- describing the distribution of categorical data.

KEY TERMS

- Frequency table
- Percentage frequency
- Bar chart
- Segmented bar chart
- Percentage segmented bar chart
- Mode

Lists of categorical information can be converted into tables, graphs and charts so that they can be easily read and interpreted. These displays can be used to identify the number, or percentage, of data for each category, as well as the most frequently occurring category.

Constructing frequency tables

A **frequency table** is a table that tallies how often each value in a data set occurs. This is the first step in making a set of data easier to summarise and analyse.

Data can be recorded within a frequency table as either frequency or **percentage frequency**. The percentage frequency is the proportion of times each value or category occurs in relation to the entire data set, represented as a percentage.

$$\textit{percentage frequency} = \frac{\textit{frequency}}{\textit{total frequency}} \times 100$$

Worked example 1

The students in a prep class were asked the question, ‘Would you describe your teacher’s height as short, average or tall?’. Their responses were as follows:

average	short	tall	tall	short	short
average	average	tall	short	average	short
average	average	average	tall	tall	short
average	tall	tall	average	average	tall

Use this data to create a frequency table displaying both frequency and percentage frequency, correct to the nearest decimal place.

Continues →

Back to contents

Explanation

Step 1: Set up a frequency table.

The table should have 3 columns for the variable collected, and the frequency as a number and percentage. There should be an appropriate number of rows to include all the categories. Finally, a row should be included for the total.

teacher's height	**frequency**	
	number	**%**
short		
average		
tall		
total		

Step 2: Fill in the frequency number column by counting from the data set, including the total.

teacher's height	**frequency**	
	number	**%**
short	6	
average	10	
tall	8	
total	24	

Step 3: Calculate the frequency as a percentage for each category, making sure the percentages add up to 100.

Note: When percentages have been rounded, they may not add up to exactly 100. In these situations this is okay, as long as the rounding has been done accurately.

Remember that the question asks for percentages given to the nearest decimal place.

$$\textit{percentage frequency} = \frac{\textit{frequency}}{\textit{total frequency}} \times 100$$

teacher's height	**frequency**	
	number	**%**
short	6	$\frac{6}{24} \times 100 = 25.0$
average	10	$\frac{10}{24} \times 100 \approx 41.7$
tall	8	$\frac{8}{24} \times 100 \approx 33.3$
total	24	100.0

Answer

teacher's height	**frequency**	
	number	**%**
short	6	25.0
average	10	41.7
tall	8	33.3
total	24	100.0

Constructing bar charts

A **bar chart** is a graphical display that is commonly used to display categorical data. The frequency or percentage frequency of each category is represented by columns of varied height. Spaces are included between columns to indicate that the categories are separate.

Worked example 2

24 students in a prep class were asked the question, 'Would you describe your *teacher's height* as short, average or tall?'. Their responses are recorded in the frequency table shown.

Use the frequency table to construct a frequency bar chart.

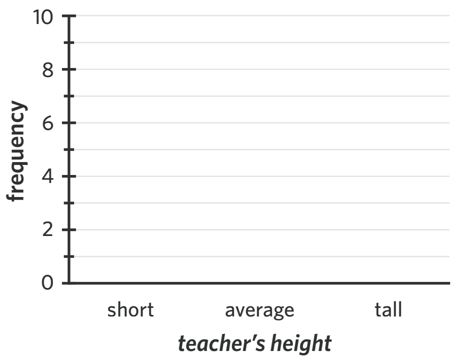
teacher's height	**frequency**	
	number	**%**
short	6	25.0
average	10	41.7
tall	8	33.3
total	24	100.0

Explanation

Step 1: Construct axes with the 'frequency' on the vertical axis and '*teacher's height*' on the horizontal axis.

The vertical axis should at least extend to the maximum value.

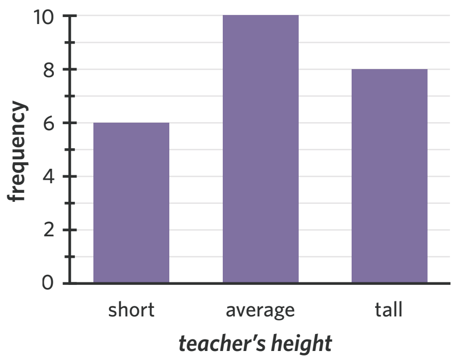
The horizontal axis should include labels for each of the categories.



Step 2: Draw vertical columns for each category according to their value in the frequency table.

Remember that each column should be separated by a gap.

Answer



Constructing segmented bar charts

A **segmented bar chart** is a variation of a bar chart with each category stacked into one column. They are particularly useful for comparing the distribution of categories across different sets of data. This will be explored further later.

Each category within a segmented bar chart has its own segment, with no gaps between segments. The height of each segment indicates the frequency of each category. A legend indicates which segments of the bar relate to which categories. Segmented bar charts can also be constructed for the percentage frequency of a data set. This is called a **percentage segmented bar chart**.

Back to contents

Worked example 3

24 students in a prep class were asked the question, 'Would you describe your *teacher's height* as short, average or tall?'. Their responses are recorded in the frequency table shown.

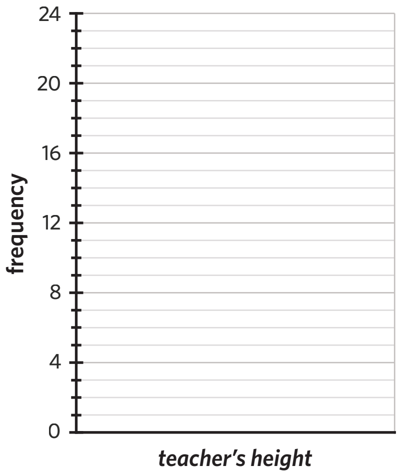
teacher's height	frequency	
	number	**%**
short	6	25.0
average	10	41.7
tall	8	33.3
total	24	100.0

a. Use the frequency table to construct a segmented bar chart.

Explanation

Step 1: Construct axes with the 'frequency' on the vertical axis and '*teacher's height*' on the horizontal axis.

The vertical axis should at least extend to the total frequency.



Step 2: Construct the column by adding the value of each segment.

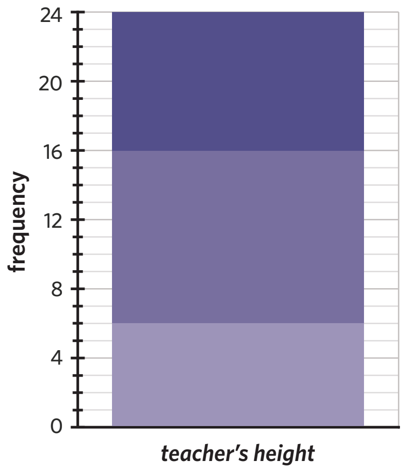
For this segmented bar chart, we will go from short,to average, to tall.

The 'short' segment should end at 6.

The 'average' segment should end at $6 + 10 = 16$.

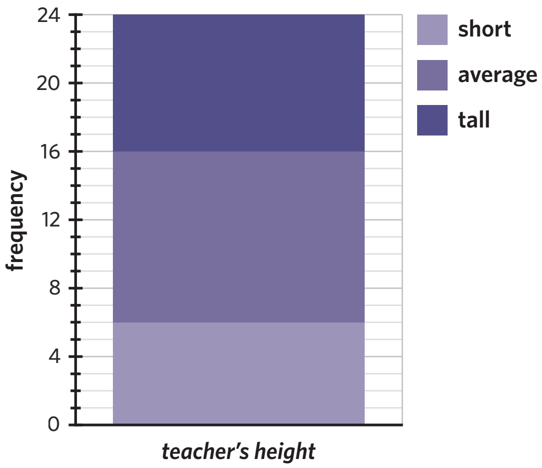
The 'tall' segment should end at $16 + 8 = 24$.

Ensure each segment is clearly defined.



Step 3: Add a legend so the graph can be interpreted correctly.

Answer



Continues →

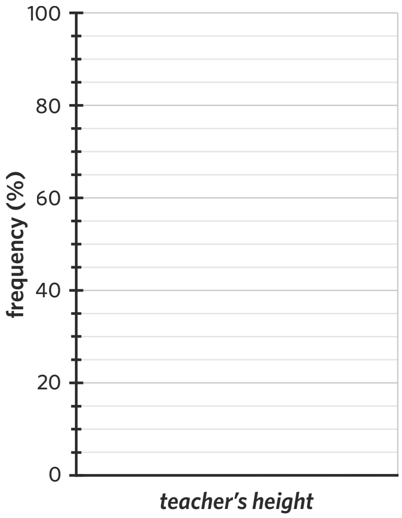
Back to contents

b. Use the frequency table to construct a percentage segmented bar chart.

Explanation

Step 1: Construct axes with the frequency as a percentage on the vertical axis and *'teacher's height'* on the horizontal axis.

The vertical axis should extend to 100%.



Step 2: Construct the column by adding the percentage of each segment.

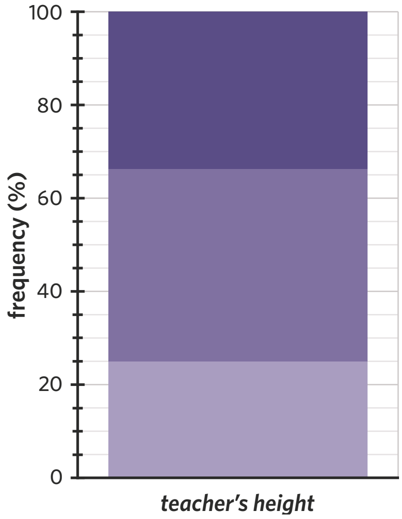
For this percentage segmented bar chart, we will go from short, to average, to tall.

The 'short' segment should end at 25.

The 'average' segment should end at $25 + 41.7 = 66.7$.

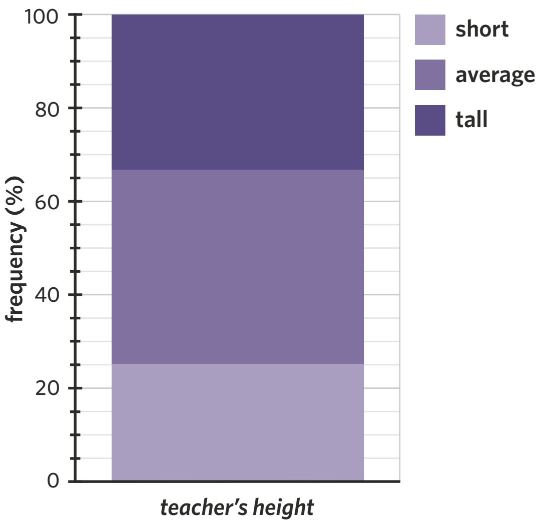
The 'tall' segment should end at $66.7 + 33.3 = 100$.

Ensure each segment is clearly defined.



Step 3: Add a legend so the graph can be interpreted correctly.

Answer



Describing the distribution of categorical data

When describing data, the mean, median and mode are often mentioned as measures of centre. This is the middle, or 'average' value of a distribution. The mode is the only available measure of centre for categorical data as the mean and median only apply to numerical data. The **mode** is the most frequently occurring value in the data set. It can be identified from a bar chart or segmented bar chart by looking at the column or segment with the greatest vertical height.

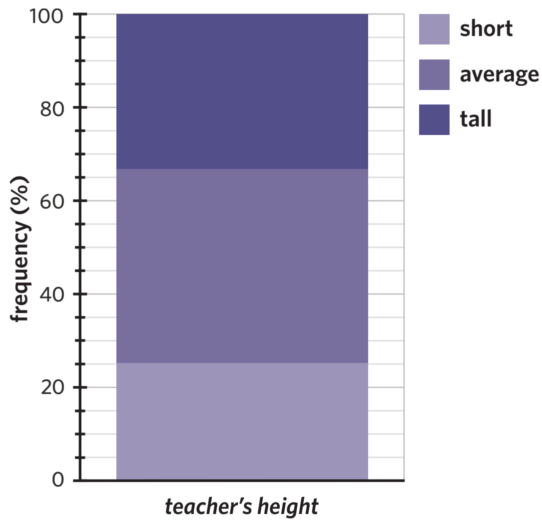
An interpretation of frequency tables, bar charts and segmented bar charts often involves writing a report which can:

- summarise the data type and the number of values represented in the data set
- identify the modal category (if it is obvious)
- compare the percentage frequencies of different categories.

In larger data sets, not all categories need to be mentioned. It might be easier to draw attention to the largest and smallest columns.

Worked example 4

24 students in a prep class were asked the question, 'Would you describe your *teacher's height* as short, average or tall?'. Their responses are shown in the given percentage segmented bar chart.



a. Find the modal category of the data set.

Explanation

Identify the segment with the greatest vertical height.

Answer

Average

b. Describe the distribution of the data set.

Explanation

Consider the elements to be included in the report describing the distribution.

- Number of people surveyed
- Modal category
- Other significant percentages

Answer

24 prep students were surveyed on how tall they thought their teacher was. The most common response was average, accounting for 41.7% of responses, while 25% said their teacher was short, and 33.3% said their teacher was tall.

Back to contents

1B Questions

Constructing frequency tables

1. A group of people were asked whether they preferred coffee, tea, or neither in the morning. Their results are displayed in the following frequency table, with percentages rounded to the nearest whole number.

drink preference	**frequency**	
	number	**%**
coffee	19	59
tea	5	16
neither	8	25
total	32	100

Which of the following statements is true?

A. 19 people were surveyed, and 59% preferred coffee.

B. 32 people were surveyed, and 8% preferred neither.

C. 32 people were surveyed, and 59% preferred coffee.

D. 32 people were surveyed, and 16 of them preferred tea.

2. 20 members of the Italian Club were asked what their *favourite type of pasta* is. Their results were as follows:

penne	penne	spaghetti	fettuccine	penne
fettuccine	macaroni	spaghetti	penne	spaghetti
spaghetti	fettuccine	penne	macaroni	penne
spaghetti	fettuccine	spaghetti	penne	fettuccine

Use these results to construct a frequency table including frequencies and percentages.

Constructing bar charts

3. A class of 312 Year 12 boys were asked their *shoe size*. Their results are recorded in the given bar chart.

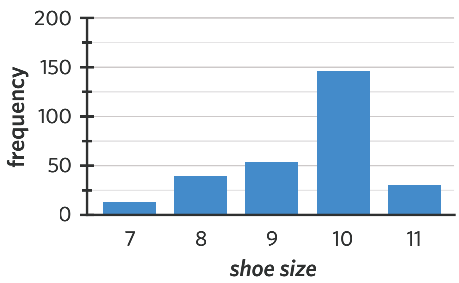
The number of Year 12 boys with a size 9 shoe is closest to

A. 10

B. 30

C. 45

D. 55



4. The *shirt sizes* (extra small, small, medium, large, extra large) of 49 people are displayed in the given frequency table. Percentages are rounded to the nearest decimal place.

a. Use the frequency table to construct a frequency bar chart.

b. Use the frequency table to construct a percentage frequency bar chart.

shirt size	**frequency**	
	number	**%**
extra small	3	6.1
small	14	28.6
medium	17	34.7
large	9	18.4
extra large	6	12.2
total	49	100.0

Back to contents

Constructing segmented bar charts

5. A group of people were asked what their preferred *streaming service* was. Their responses are shown in the frequency segmented bar chart shown

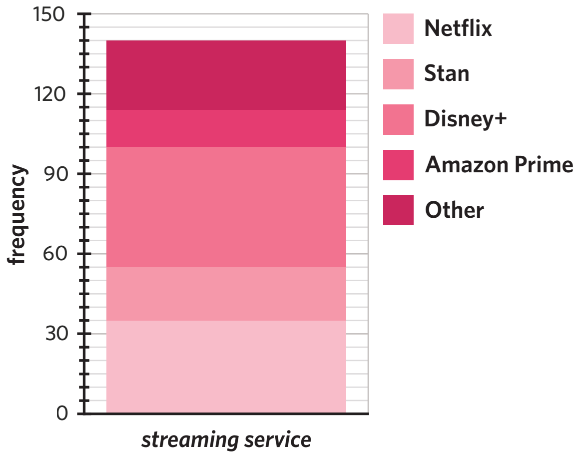
Which of the following statements is false?

A. 20 people prefer Stan.

B. More people prefer Disney+ than Netflix and Stan combined.

C. 140 people were surveyed.

D. 15 people prefer Amazon Prime.



6. 139 people were asked what their *favourite animal* was. The results are shown in the frequency table shown. Percentages have been rounded to the nearest whole number.

a. Use the data from the frequency table to construct a frequency segmented bar chart.

b. Use the data from the frequency table to construct a percentage segmented bar chart.

favourite animal	frequency	
	number	**%**
dog	47	34
cat	52	37
guinea pig	22	16
horse	14	10
snake	4	3
total	139	100

Describing the distribution of categorical data

7. 88 people were asked who their *favourite Avenger* was. The results are shown in the bar chart provided.

The modal superhero is

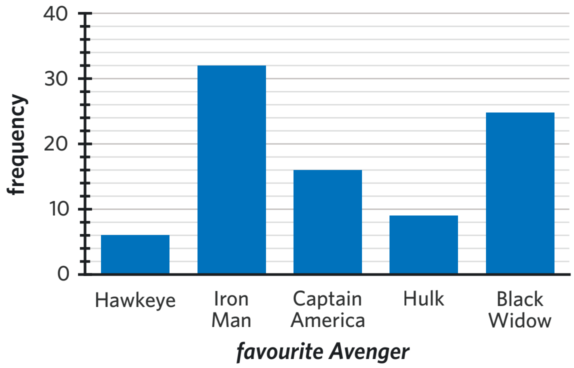
A. Hawkeye

B. Iron Man

C. Captain America

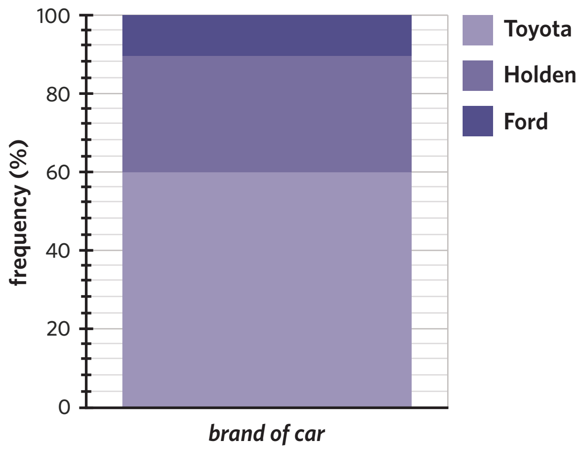
D. Hulk

E. Black Widow



8. The brands of 50 cars entering a carpark were recorded, with the results shown in the percentage segmented bar chart provided. Use this data to fill out the following report template.

The brands of __________ cars were recorded as they entered a car park. All the cars were either 'Holden', 'Ford', or 'Toyota'. The most commonly occurring brand of car was __________, accounting for __________% of all cars. The next most commonly occurring brand was __________, representing __________%. Finally, the last __________% of the cars were __________.



Back to contents

Joining it all together

9. A barista collected information on the type of milk that customers ordered with their coffee. The results are shown in the given percentage segmented bar chart.

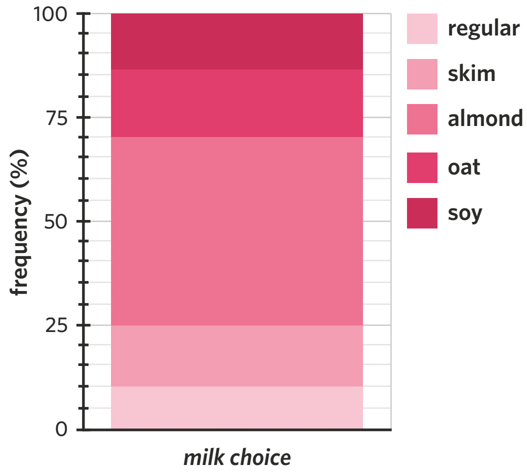
The barista remembered that 36 people ordered almond milk. The number of people that ordered regular milk is closest to

A. 8

B. 10

C. 15

D. 25



10. A group of musicians were asked who their *favourite jazz drummer* was. Their responses were as follows:

Tony Williams	Elvin Jones	Elvin Jones	Brian Blade
Tony Williams	Brian Blade	Buddy Rich	Art Blakey
Elvin Jones	Elvin Jones	Tony Williams	Buddy Rich
Buddy Rich	Art Blakey	Elvin Jones	Buddy Rich
Brian Blade	Elvin Jones	Buddy Rich	Tony Williams

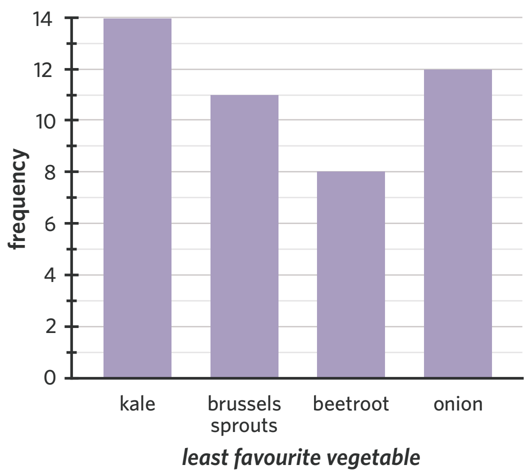
a. Use these results to construct a frequency table.

b. Using the frequency table from part **a**, construct a percentage bar chart to show the results.

c. Using the frequency table from part **a**, construct a frequency segmented bar chart to show the results.

d. Use the data to write a paragraph on the distribution of favourite jazz drummers amongst the musicians.

11. A group of toddlers were asked about their *least favourite vegetable*. The results are represented in the given bar chart.

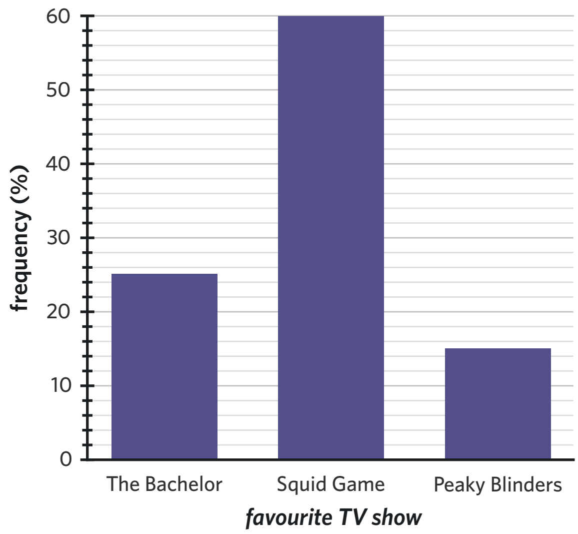


Draw a percentage segmented bar chart to represent this data, correct to the nearest percentage.

Back to contents

12. A group of office workers were asked what their *favourite TV show* was. The results are displayed in the bar chart shown.

6 people said Peaky Blinders was their favourite show. Use this information and the bar chart to construct a frequency table that represents this information.



Exam practice

13. A study was conducted that investigated the *number of moths* caught in a sugar moth trap (less than 250, 250–500, more than 500). The results are summarised in the percentage segmented bar chart shown.

There were 300 sugar traps.

The number of sugar traps that caught less than 250 moths is closest to

A. 30

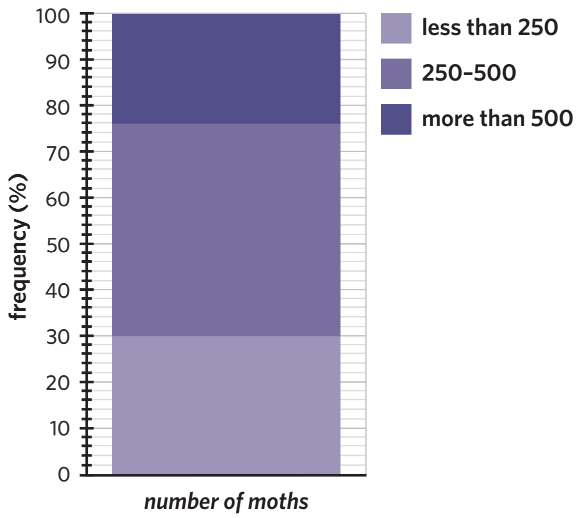
B. 90

C. 250

D. 300

E. 500

Adapted from VCAA 2017 Exam 1 Data analysis Q5



76% of students answered this type of question correctly.

14. The given bar chart shows the distribution of *wind directions* recorded at a weather station at 9:00 am on each of 214 days in 2011.

According to the bar chart, the percentage of the 214 days on which the wind direction was observed to be east or south-east is closest to

A. 10%

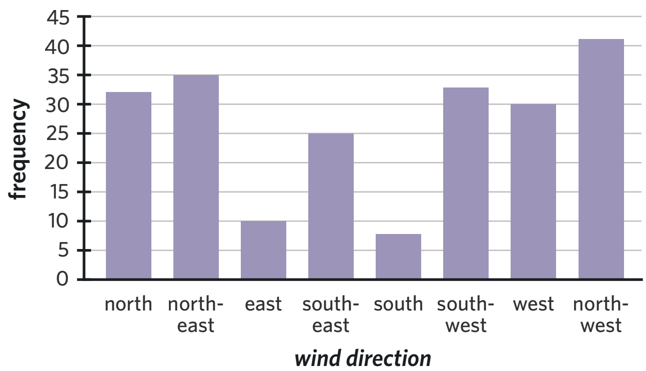
B. 16%

C. 25%

D. 33%

E. 35%

VCAA 2012 Exam 1 Data analysis Q2



68% of students answered this question correctly.

Back to contents

Questions from multiple lessons

Data analysis *Year 11 content*

15. The *clothing size* (small, medium, large), and *age* (under 10 years, 10 years or over) of students at a primary school were collected. In this context, the variables *clothing size* and *age* are

A. ordinal and nominal respectively

B. nominal and ordinal respectively

C. ordinal and continuous respectively

D. both ordinal

E. both nominal

Adapted from VCAA 2016 Exam 1 Data analysis Q2

Recursion and financial modelling *Year 11 content*

16. As part of her new year resolutions, Sarah decides to read every month from January to December for one year. Each month she counts the number of pages that she has read.

In January, she reads 12 pages of a book. In February, she reads 18 pages. In March, she reads 24 pages. In April, she reads 30 pages.

The number of pages she reads each month continues to increase according to this pattern.

The number of pages she reads in September is

A. 48

B. 54

C. 60

D. 66

E. 72

Adapted from VCAA 2014 Exam 1 Number patterns Q1

Recursion and financial modelling *Year 11 content*

17. Alex invested $1000 in a savings account, with interest compounding annually.

M_n is the amount of money in the account after n years.

The following calculations show the amount of money in Alex's account initially, and after one and two years.

$M_0 = 1000$

$M_1 = 1.04 \times 1000 = 1040$

$M_2 = 1.04 \times 1040 = 1081.60$

a. Find a recurrence relation in terms of M_0, M_{n+1}, and M_n that models the amount of money in Alex's savings account after n years. (1 MARK)

b. Alex wants to buy a new laptop for $1250. What is the minimum interest rate per annum that would have been required for Alex to afford this laptop after two years? Give your answer correct to two decimal places. (1 MARK)

Adapted from VCAA 2018NH Exam 2 Recursion and financial modelling Q7c,d

Back to contents

1C Displaying numerical data

STUDY DESIGN DOT POINT

- 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

1A 1B **1C** 1D 1E 1F 1G 1H 1I

KEY SKILLS

During this lesson, you will be:

- displaying data using dot plots
- displaying data using stem plots
- constructing grouped frequency tables
- displaying data using histograms.

KEY TERMS

- Dot plot
- Stem plot
- Grouped frequency table
- Histogram

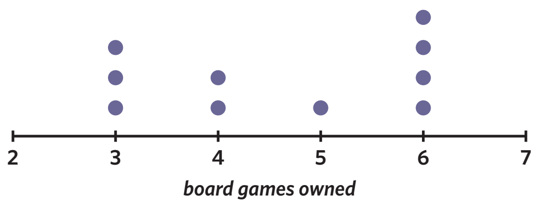
Dot plots, stem plots and histograms are displays that help us visualise the distribution of numerical data. These displays can then be used to identify the number, or percentage, of data within certain ranges of values, as well as the most frequently occurring values.

Displaying data using dot plots

A **dot plot** is a simple way to display discrete numerical data, where each data point is represented by a dot above a single axis.

The number of dots above a value on the axis represents the frequency of the value. The mode of the data set (also known as the modal value) is the value with the most number of dots.

Dot plots are ideal for displaying small/medium-sized data sets with a small range of values.



Worked example 1

Sophie surveyed 12 of the families living on her street.

She asked for the *number of pets* each of them owned and the results were recorded.

3 2 0 1 1 3 0 5 2 1 1 2

a. Construct a dot plot to display this data.

Explanation

Step 1: Rearrange the data set into ascending order and determine the lowest and highest value.

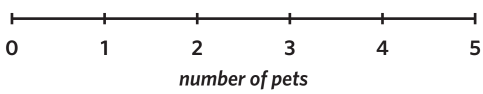
0 0 1 1 1 1 2 2 2 3 3 5

The lowest value is zero.

The highest value is five.

Step 2: Construct a number line with an appropriate scale.

The scale should cover all values between zero and five.



Continues →

Back to contents

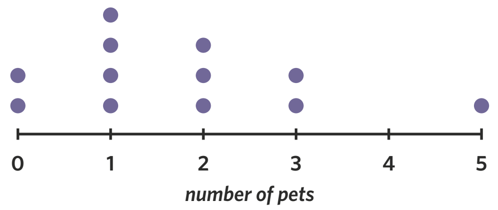
Step 3: Represent each value with a dot.

Mark a dot above the number on the number line each time a value appears in the data set.

If the same data value appears multiple times, illustrate this by placing the corresponding number of dots in a vertical line.

Spacing between each of the vertical dots should be consistent to allow for comparison of frequency across different values.

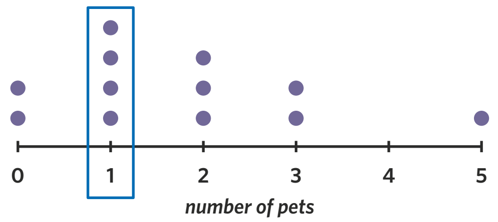
Answer



b. What was the modal *number of pets* owned by families that Sophie surveyed?

Explanation

Find the value with the most dots on the dot plot.



Answer

1 pet

Displaying data using stem plots

See worked example 2

A **stem plot** is a way to display numerical data, where data points are grouped by their leftmost digit(s). Each leaf represents the last digit of an individual data value, and each stem represents the leftmost digit(s) of a group of leaves.

Stems are shown vertically, to the left of a vertical line, ordered from smallest to largest.

leaves are positioned to the right of the vertical line, in line with their corresponding stem. Within each stem, leaves should be ordered from smallest to largest.

Key: 4 | 3 = 43

4	3	5	5	5	7
5	1	3	6	8	
6	7	9			
7	1				
8	0	1	3	9	9
9	1	4			

When constructing a stem plot, always remember to include a key. The key demonstrates the scale of the data. The key allows for data of many forms to be shown in a stem plot. This includes decimals and three (or more) digit numbers.

Back to contents

Decimal:

Key: 1 | 2 = 1.2

1	3 3 4 6 8
2	0 4 9
3	1 1 1 4 5 8
4	2

Three-digit:

Key: 20 | 0 = 200

20	0 0 1 1 2
21	0 1 3 5
22	1 6 9
23	0 4 4 8 8 8
24	5 7

The frequency of a single data value can be found by finding the corresponding stem and counting the number of corresponding leaves within it. The modal value is the value with the most number of identical leaves within a single stem.

In some cases it can be difficult to see the underlying distribution due to having a lot of data within a small range. This problem is solved by 'splitting' the stems. Usually each stem is split into either two or five stems, depending on how close together the data is.

See worked example 3

Key: 1 | 2 = 12

0	1 1 2 3 4
0	5 6 6 8 8 9
1	2 3 3
1	6 7 7 8 9 9

Stem plots are ideal for displaying small/medium-sized data sets with a large range of values.

Worked example 2

Ms Smyth's maths class of 25 students sat their end-of-year exam. Their *results* (%) were recorded.

55 68 76 90 83 89 75 66 59 84 48 62 58

95 80 77 61 92 99 63 84 65 70 81 96

a. Construct a stem plot to display this data.

Explanation

Step 1: Consider the most appropriate scale.

The data values are two-digit numbers.

The stems will refer to 'tens'.

The leaves will refer to 'ones'.

Step 2: Fill in the appropriate stems.

The data values range from 48 to 99.

All values which fall in the 40s, 50s, 60s, 70s, 80s and 90s need to be covered.

The appropriate stems are 4, 5, 6, 7, 8 and 9.

4	
5	
6	
7	
8	
9	

Note: Each stem within the range of the data needs to be included, even if there are no data values within it.

Step 3: Fill in the leaves for each stem.

Start with the smallest stem and fill the corresponding leaves in ascending order.

Repeat this for each stem.

4	8
5	5 8 9
6	1 2 3 5 6 8
7	0 5 6 7
8	0 1 3 4 4 9
9	0 2 5 6 9

Step 4: Construct a key.

A key shows the scale in which the data is represented.

As decided in step 1, the stems refer to 'tens' and the leaves refer to 'ones'.

Demonstrate this scale with an example.

Continues →

Back to contents

Answer

Key: 4 | 8 = 48%

4	8
5	5 8 9
6	1 2 3 5 6 8
7	0 5 6 7
8	0 1 3 4 4 9
9	0 2 5 6 9

b. How many students scored above 70% on the exam?

Explanation

Count the number of leaves that represent a value greater than 70.

This will include any leaves on the '7' stem that are greater than 0 and all leaves on stems greater than 7.

Key: 4 | 8 = 48%

4	8
5	5 8 9
6	1 2 3 5 6 8
7	0 5 6 7
8	0 1 3 4 4 9
9	0 2 5 6 9

Answer

14 students

Worked example 3

Ms Goyle's maths class of 25 students sat their end-of-year exam. Their *results* (%) were recorded.

75 68 76 80 83 69 65 66 79 84 78 62 88

75 80 77 61 62 69 73 84 75 60 81 66

Construct a split stem plot to display this data, with stem intervals of 5%.

Explanation

Step 1: Consider the most appropriate scale.

The data values are two-digit numbers.

The stems will refer to 'tens'.

The leaves will refer to 'ones'.

Step 2: Fill in the appropriate stems.

The data values range from 60 to 88.

All values which fall in the 60s, 70s, and 80s need to be covered.

The question specifies stem intervals of 5%.

The appropriate stems are 6, 6, 7, 7, 8 and 8.

6	
6	
7	
7	
8	
8	

Continues →

Back to contents

Step 3: Fill in the leaves for each stem.

Start with the smallest stem and fill the corresponding leaves in ascending order.

The top stem for each stem value will include leaves from 0–4 and the bottom stem for each value will include leaves from 5–9.

Repeat this for each stem.

6	0 1 2 2
6	5 6 6 8 9 9
7	3
7	5 5 5 6 7 8 9
8	0 0 1 3 4 4
8	8

Step 4: Construct a key.

A key shows the scale in which the data is represented.

As decided in step 1, the stems refer to 'tens' and the leaves refer to 'ones'.

Demonstrate this scale with an example.

Answer

Key: 6 | 0 = 60%

6	0 1 2 2
6	5 6 6 8 9 9
7	3
7	5 5 5 6 7 8 9
8	0 0 1 3 4 4
8	8

Constructing grouped frequency tables

A **grouped frequency table** groups data in regular intervals, and displays the frequency and percentage frequency of each interval. This allows the distribution of the data to be more clearly observed.

The lower bound of each interval is inclusive, whilst the upper bound is not.

For example, this grouped frequency table demonstrates that 46 students scored at least 70% but less than 80% on the test.

test mark	frequency	
	number	**%**
50–<60%	12	10.0
60–<70%	35	29.2
70–<80%	46	38.3
80–<90%	19	15.8
90–<100%	8	6.7
total	120	100.0

Worked example 4

The *height* (cm) of 20 plants in a Year 6 science experiment were recorded.

32.0 40.2 40.5 45.1 47.0 49.1 50.1 53.7 54.2 55.3

56.9 57.2 58.2 67.2 68.9 69.0 72.3 77.6 82.1 88.5

Construct a grouped frequency table with intervals of 10.

Explanation

Step 1: Determine the number of intervals required.

The question specifies intervals of 10.

The data has a minimum of 32.0, so the intervals should start at 30.

The data has a maximum of 88.5, so the intervals should end at 90.

Therefore, six intervals will be required.

Continues →

Back to contents

Step 2: Set up a table.

The table should have 3 columns, for the variable collected and for the frequency (as a number and a percentage). There should be enough rows to include all the intervals, the header and the total.

Each interval includes the lower bound and does not include the upper bound. For example, the first interval is from 30 to less than 40, and is written as '30–<40'.

height (cm)	frequency	
	number	%
30–<40		
40–<50		
50–<60		
60–<70		
70–<80		
80–<90		
total		

Step 3: Fill in the frequency columns.

Count the number of data values within each interval in the data set. This is displayed in the number column. The total of the number column will be the sum of the frequencies of the intervals.

Then find the percentage frequency for each interval. The total percentage frequency will always be 100.

$$\textit{percentage frequency} = \frac{\textit{frequency}}{\textit{total frequency}} \times 100$$

Answer

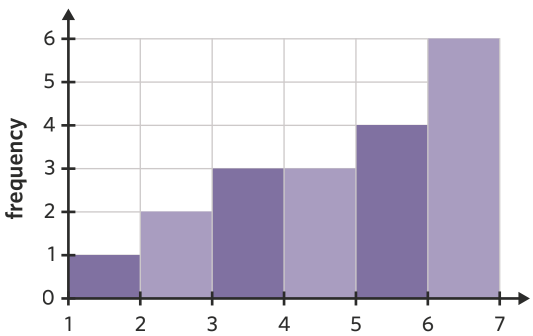
height (cm)	frequency	
	number	%
30–<40	1	5
40–<50	5	25
50–<60	7	35
60–<70	3	15
70–<80	2	10
80–<90	2	10
total	20	100

Displaying data using histograms

A **histogram** is a visual representation of a grouped frequency distribution table. It can display either frequency or percentage frequency on its vertical axis.

See worked example 5

As a histogram depicts intervals of values, there are no spaces between columns.



The height of a column represents the frequency, or percentage frequency, of the interval of values. The modal interval is the interval with the tallest column.

Histograms are best at displaying data sets with a large number of numerical values. A calculator can be helpful in creating a histogram directly from the data.

See worked example 6

Worked example 5

The *height* (cm) of 20 plants in a Year 6 science experiment was recorded in the following grouped frequency table.

***height* (cm)**	**frequency**	
	number	**%**
30–<40	1	5
40–<50	5	25
50–<60	7	35
60–<70	3	15
70–<80	2	10
80–<90	2	10
total	20	100

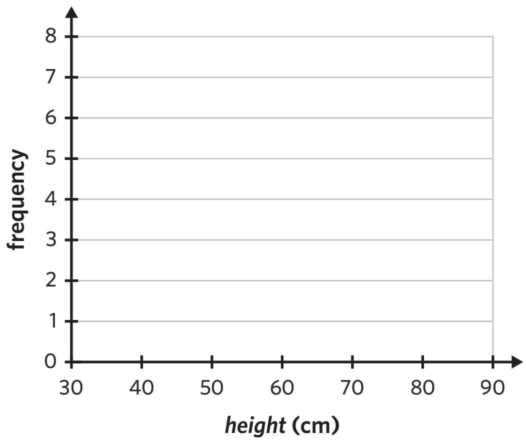
a. Construct a histogram to display this data.

Explanation

Step 1: Construct a set of axes.

Label the horizontal axis '*height* (cm)'. The horizontal axis needs to range from at least 30 to 90, with the boundaries of the intervals labelled.

Label the vertical axis 'frequency'. The vertical axis needs to range from at least 0 to 7.

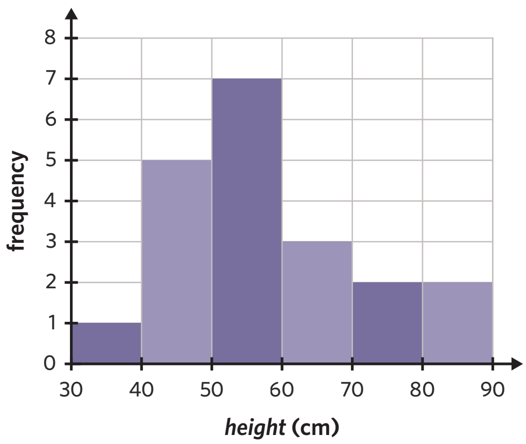


Step 2: Draw a column for each interval.

The height of the column represents the frequency of data values within the corresponding interval.

Histograms do not have spaces between columns.

Answer



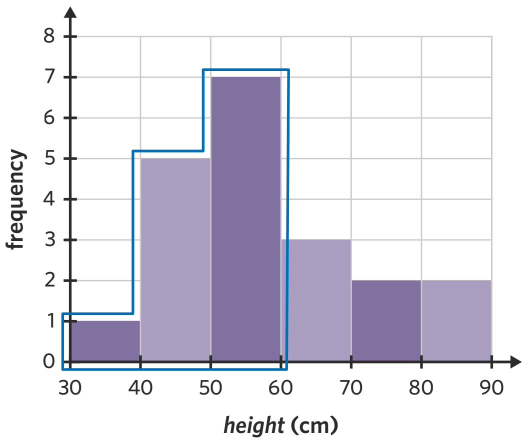
Continues →

Back to contents

b. What percentage of plants were shorter than 60 cm?

Explanation

Step 1: Sum the frequencies of columns with intervals that fall below 60 cm.



$1 + 5 + 7 = 13$

Step 2: Find this as a percentage of the total number of plants.

There are 20 plants in total.

$\frac{13}{20} = 0.65 = 65\%$

Answer

65%

Worked example 6

The *height* (cm) of dogs in a shelter were recorded.

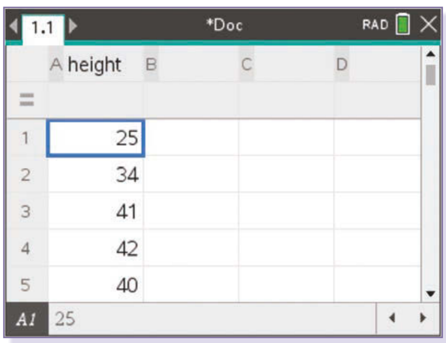
25	34	41	42	40	37	76	47	53	39	72	29	64	61	60
75	39	30	92	45	72	85	50	72	76	42	65	79	64	32

Use the data to construct a histogram.

Explanation – Method 1: TI-Nspire

Step 1: From the home screen, select '1: New' → '4: Add Lists & Spreadsheet'.

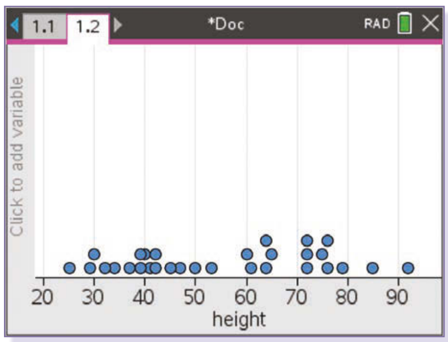
Step 2: Name column A 'height' and enter the data values into column A, starting from row 1.



Step 3: Press ctrl + doc▸ and select '5: Add Data & Statistics'.

Move the cursor to the horizontal axis and select 'Click to add variable'.

Select 'height'.



Step 4: Press menu. Select '1: Plot Type' → '3: Histogram'.

Continues →

Back to contents

Step 5: To adjust the column width and the starting point, press menu and then select ‘2: Plot Properties’ → ‘2: Histogram Properties’ → ‘2: Bin Settings’ → ‘1: Equal Bin Width’.

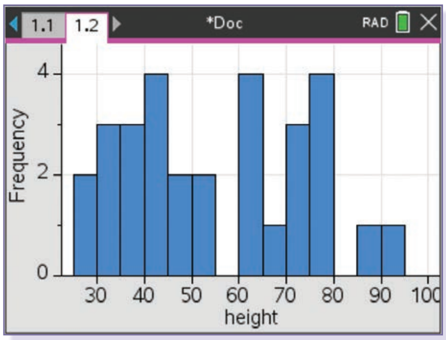
Set the column width to 5 by changing ‘Width’ to ‘5’.

Set the starting point to 25 by changing ‘Alignment’ to ‘25’.

Select ‘OK’.

Note: To change the view of the histogram press menu and use options within ‘5: Window/Zoom’.

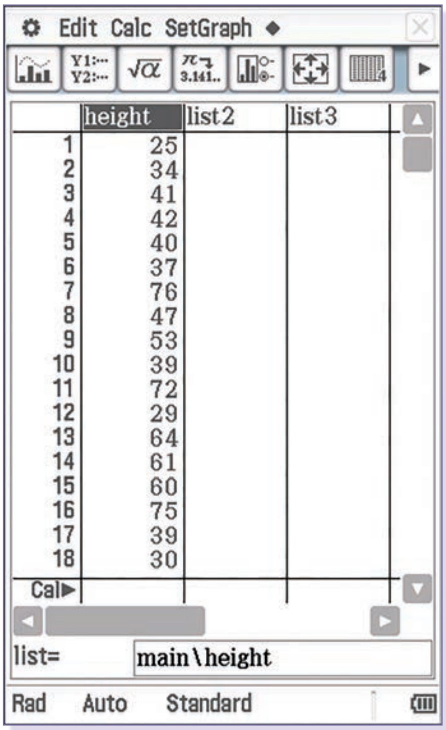
Answer



Explanation – Method 2: Casio ClassPad

Step 1: From the main menu, tap Statistics.

Step 2: Rename list1 to ‘height’ and enter the data values starting from row 1.



Step 3: Configure the settings of the graph by tapping the graph settings icon.

Create a histogram by changing ‘Type:’ to ‘Histogram’.

Specify the data set by changing ‘XList:’ to ‘main\height’.

Tap ‘Set’ to confirm.

Step 4: Tap the graph icon in the icon bar to plot the histogram.

Set the starting point to 25 by changing ‘HStart’ to ‘25’.

Set the column width to 5 by changing ‘HStep’ to ‘5’.

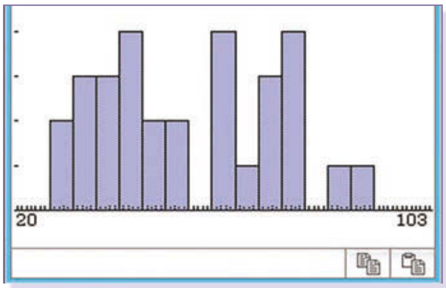
Tap ‘OK’ to confirm.

Note: To change the view of the histogram press the window icon.

Continues →

Back to contents

Answer

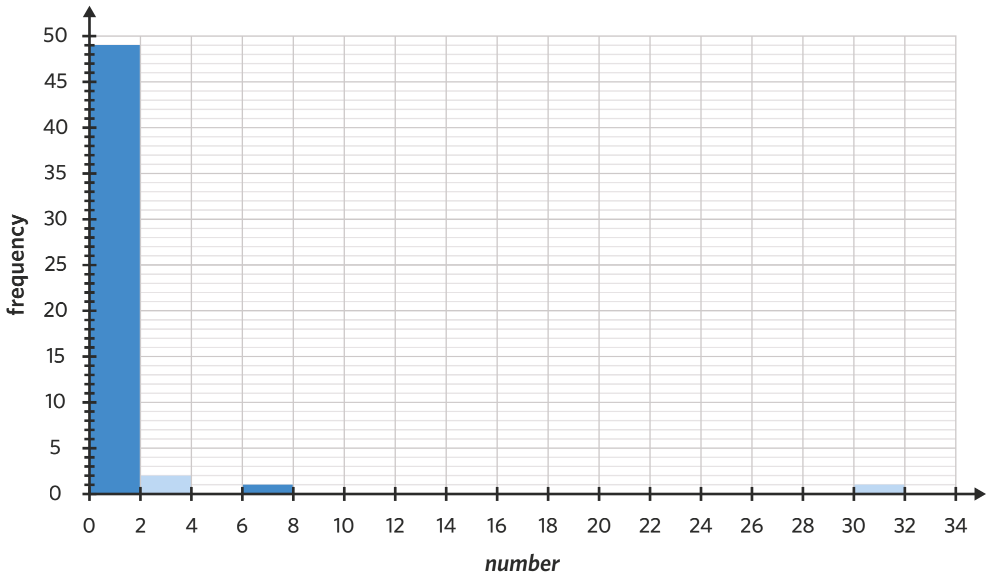


Note: To analyse the histogram, tap in the icon bar to place a (+) marker at the first column. 'xc=25' denotes the starting point of the first column (25) and 'Fc=2' shows the frequency of this interval (2).

Exam question breakdown

VCAA 2016 Exam 1 Data analysis Q6

The following histogram shows the distribution of the *number* of billionaires per million people for 53 countries.



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%

Explanation

Step 1: Sum the frequencies of columns with intervals that fall below 2 billionaires per million people.

The only column that falls below 2 billionaires per million people is the leftmost column.

It has a frequency of 49.

Step 2: Find this as a percentage of the total number of countries.

There are 53 countries in total.

$$\frac{49}{53} = 0.9245...$$

$$\approx 92\%$$

71% of students answered this question correctly.

11% of students incorrectly answered A. This was likely because the number of countries with less than 2 billionaires was 49. However, the question asked for this as a percentage of the total number of countries, which was closest to 92%.

Answer

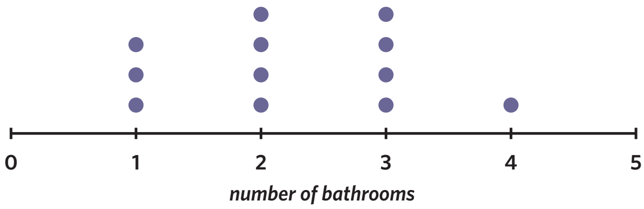
D

Back to contents

1C Questions

Displaying data using dot plots

1. The *number of bathrooms* in 12 houses is shown in the following dot plot.



The data used to construct this dot plot was

A. 1 2 3 1 2 3 3 2 2 3 1 3

B. 1 3 2 4 2 1 3 3 3 2 1 2

C. 3 3 3 4 4 1 1 3 3 2 2 1

D. 3 1 4 2 2 3 1 1 3 2 2 1

---

2. Use the following data to construct a dot plot.

21 28 22 23 21 24 27 28 29 24 28 25 25 24

---

3. Mr Hogan recorded the *amount of time* his students spent over a week studying for history, correct to the nearest hour.

4 3 5 3 2 6 1 2 5 4 3 4 6 4 2 3 2 4 6 3

a. Display Mr Hogan's results in a dot plot.

b. How many students spent two hours studying history over a week?

c. Mr Hogan recommends his students spend 4 hours studying history each week. How many students spent less time than recommended studying history?

d. What percentage of students spent at least five hours studying history over a week?

Displaying data using stem plots

4. The *numbers of fingers* of 11 species of aliens are shown in the following stem plot.

Key: 0 | 5 = 5 fingers

0	1 5 7 8
1	0 3 6
2	2 3 7
3	1

The data used to construct this stem plot was

A. 1 23 7 23 5 16 27 8 10 22 31

B. 27 8 10 7 23 5 16 22 31 10 13

C. 23 5 16 22 1 10 13 23 8 13 7

D. 1 13 8 10 22 31 7 23 5 16 27

---

5. Use the following data to construct a stem plot.

40 89 75 86 89 54 87 82 46 41

74 44 63 71 53 58 91 99 54 47

6. Simon conducted a survey for his statistics class by asking 25 of his classmates how long it took them to get to school on a particular day. He displayed his findings in a stem plot.

Key: 0 | 5 = 5 minutes

0	5 5 6
1	2 3 6 8 9
2	1 4 5 5 6 7 9
3	2 3 4 8 8 8
4	0 1 6
5	2

a. How many students took between 20 and 30 minutes to get to school?

b. How many students took more than 25 minutes to get to school?

c. What is the mode?

---

7. Mrs Jones' physics class did an experiment investigating the *height* that a ball would bounce after being dropped from a ledge. They conducted 20 trials of the experiment and recorded the *heights* (cm) that the ball reached in each of the trials.

178 184 171 180 183 175 189 174 179 184

187 176 170 188 185 190 195 177 181 182

a. Display this data in a split stem plot, with stem intervals of 5 cm.

b. How many times did the ball bounce between 175 cm and 179 cm (inclusive)?

c. How many times did the ball bounce at least 178 cm?

d. What percentage of balls bounced less than 180 cm?

Constructing grouped frequency tables

8. The *weight* of puppies in a litter are recorded in the following grouped frequency table.

weight (kg)	frequency
0.8–<0.9	2
0.9–<1.0	4
1.0–<1.1	7
1.1–<1.2	3
total	16

a. How many puppies weigh less than 1 kg?

A. 2

B. 4

C. 6

D. 14

b. What percentage of puppies weigh 0.9 kg or more?

c. What is the modal weight interval?

---

9. The *maximum flying height* (m) of 16 drones was recorded.

15.3 14.5 15.1 15.2 15.0 16.2 14.8 14.9

15.8 16.2 15.9 16.2 15.3 16.1 15.6 14.9

a. Construct a grouped frequency distribution table with an interval size of 0.5 m to display the data.

b. The company claims that its drones can reach a *maximum flying height* of at least 15 m. What percentage of drones do not meet this criterion?

Displaying data using histograms

10. The number of *games played* by 22 Narrm Demons players is shown in the histogram.

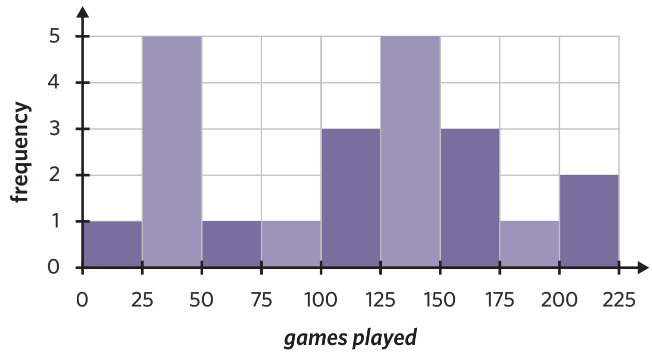
How many players have played less than 100 games?

A. 1

B. 8

C. 11

D. 14



11. A class of 20 students timed themselves to see how long it would take each of them to run around the school oval. Their results are displayed in a histogram.

a. How many students took between

i. 40 and 45 seconds?

ii. 55 and 60 seconds?

iii. 40 and 60 seconds?

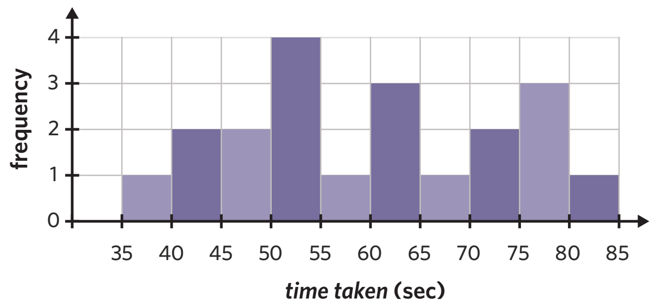
b. How many students ran around the oval in

i. less than 40 seconds?

ii. less than 60 seconds?

iii. more than 70 seconds?

c. What is the modal interval?



12. The *selling prices* of NFTs are shown in the following grouped frequency table.

a. Use the grouped frequency table to construct a histogram.

b. How many NFTs sold for at least \$35 000 but less than \$55 000?

c. Use the grouped frequency table to construct a histogram using percentage frequency.

d. What percentage of the NFTs sold for at least \$40 000, correct to the nearest percent?

selling price **(\$000's)**	**frequency**	
	number	**%**
15–<20	2	6.5
20–<25	1	3.2
25–<30	3	9.7
30–<35	7	22.6
35–<40	3	9.7
40–<45	5	16.1
45–<50	1	3.2
50–<55	4	12.9
55–<60	2	6.5
60–<65	3	9.7
total	31	100.0

13. The *ages* of 20 guests at a wedding are recorded.

19 27 45 25 28 55 22 56 93 12 52 88 39 20 18 43 82 45 76 58

a. Use a calculator to construct a histogram. The first column should start at 10 and it should have a column width of 10.

b. Why was 10 chosen as the start of the first column?

c. What is the starting point for the fifth column?

d. What is the modal interval?

Back to contents

Joining it all together

14. The *maximum temperature* (°C) for days within a fortnight in Bairnsdale are listed.

12 15 10 21 35 42 32 28 31 16 18 8 12 10

a. Which visual display is most appropriate for the data?

b. Construct the most appropriate visual display for the data.

c. On how many days was the maximum temperature 10 °C or colder?

d. On what percentage of the days was the maximum temperature greater than 16 °C?

---

15. The *heights* (m) of the players in a junior level basketball team were recorded.

1.51 1.51 1.63 1.64 1.65 1.67 1.68 1.70 1.75

1.76 1.80 1.85 1.86 1.87 1.89 1.90 1.99

a. Construct a grouped frequency distribution table with an interval size of 0.1. Round percentages to one decimal place.

b. Use the grouped frequency table to construct a histogram.

c. How many members have a height of less than 1.8 m?

d. What is/are the mode(s) of the original data?

e. What is/are the modal interval(s)?

f. Is the mode of the original data a better measure of centre than the modal interval? Explain your answer.

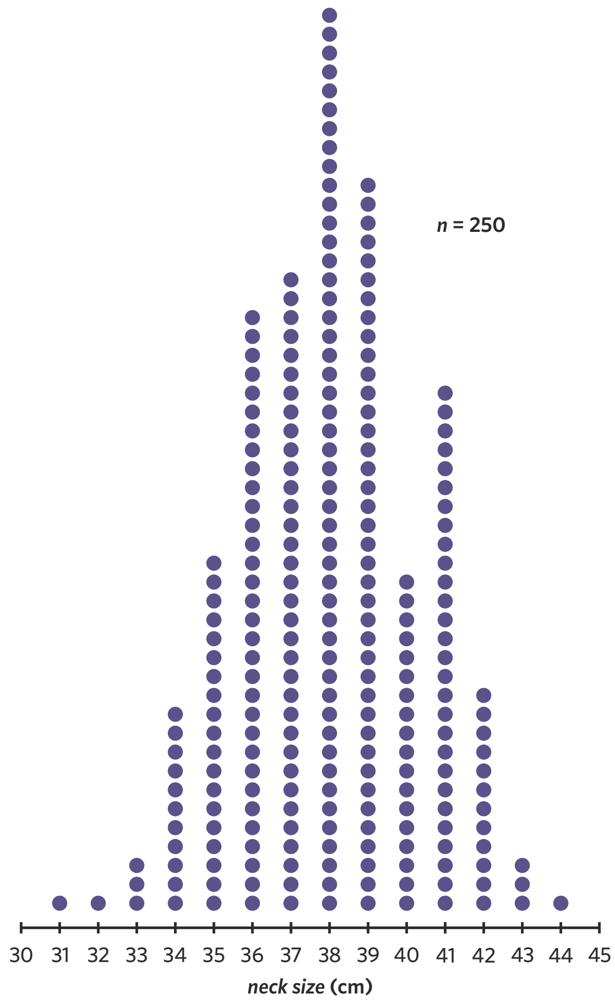
Exam practice

16. The *neck size*, in centimetres, of 250 men was recorded and displayed in the following dot plot.

Write down the modal *neck size*, in centimetres, for these 250 men. (1 MARK)

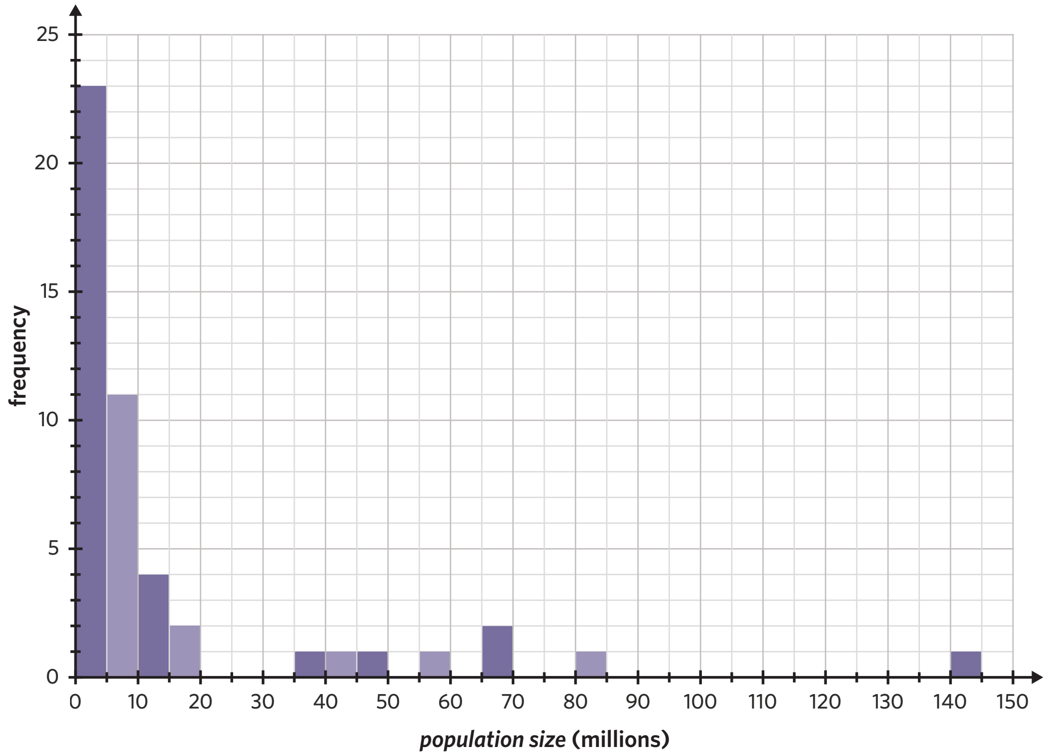
VCAA 2020 Exam 2 Data analysis Q2a

93% of students answered this question correctly.



Back to contents

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



The number of these countries with a *population size* between 5 million and 20 million is

A. 11

B. 17

C. 23

D. 34

E. 35

VCAA 2019 Exam 1 Data analysis Q1

92% of students answered this question correctly.

---

18. The following stem plot displays 30 temperatures recorded at a weather station.

temperature **Key:** 2 | 2 = 2.2 °C

2	2	2	4	4								
2	5	7	8	8	8	8	8	8	9	9	9	9
3	1	2	3	3	4	4	4					
3	5	6	7	7	7	7						
4	1											

The modal *temperature* is

A. 2.8 °C

B. 2.9 °C

C. 3.7 °C

D. 8.0 °C

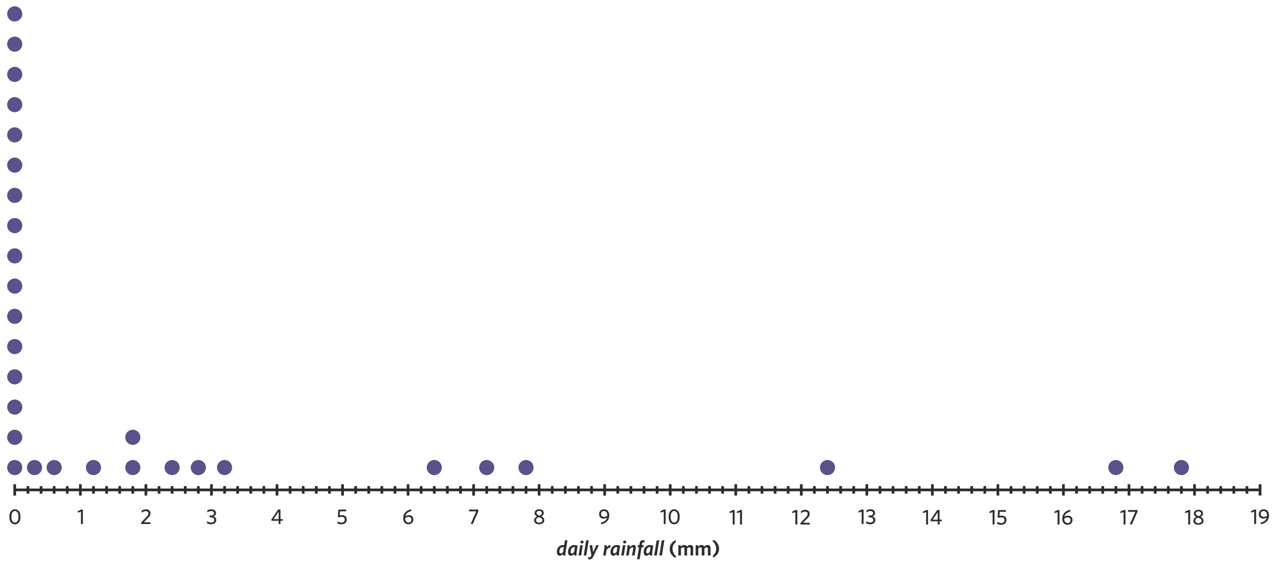
E. 9.0 °C

VCAA 2016 Exam 1 Data analysis Q3

87% of students answered this question correctly.

Back to contents

19. The following dot plot shows the distribution of *daily rainfall*, in millimetres, at a weather station for 30 days in September.



Construct a histogram 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)

Adapted from VCAA 2016 Exam 2 Data analysis Q1d

The average mark on this type of question was **1.1**.

Questions from multiple lessons

Data analysis *Year 11 content*

20. The following two-way frequency table displays the *favourite type of pizza* (margherita, capricciosa, hawaiian) and *sex* (male, female) of 140 people.

What percentage of females chose capricciosa as their favourite type of pizza, correct to the nearest percent?

A. 23%

B. 29%

C. 32%

D. 43%

E. 45%

favourite type of pizza	***sex***	
	male	**female**
margherita	12	28
capricciosa	29	32
hawaiian	24	15
total	65	75

Adapted from VCAA 2016 Exam 1 Data analysis Q1

Data analysis *Year 11 content*

21. The heights, in centimetres, of a sample of eight basketball players were recorded and are displayed in the following table.

height **(cm)**	189	178	190	183	181	194	186	188

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

A. $\bar{x} = 186.1 \quad s_x = 4.88$

B. $\bar{x} = 186.1 \quad s_x = 5.00$

C. $\bar{x} = 186.1 \quad s_x = 5.22$

D. $\bar{x} = 4.88 \quad s_x = 186.1$

E. $\bar{x} = 5.22 \quad s_x = 186.1$

Adapted from VCAA 2017 Exam 1 Data analysis Q3

Back to contents

Data analysis *Year 11 content*

22. Information on 26 different animals is shown in a table.

The four variables in this data set are:

- *species* – name of the animal
- *continent of origin* – continent the animal is originally from
- *size* – size of the animal (small, medium, large)
- *average lifespan* – the average lifespan of the animal (years)

species	***continent of origin***	***size***	***average lifespan* (years)**
aardvark	Africa	medium	30
baboon	Africa	large	45
capybara	South America	medium	10
dodo	Africa	medium	20
echidna	Oceania	small	10
ferret	Europe	small	5
goat	Asia	medium	15
hippopotamus	Africa	large	50
impala	Africa	large	12
jaguar	South America	large	15
kangaroo	Oceania	large	20
lemur	Africa	small	16
meerkat	Africa	small	13
nightingale	Europe	small	2
ocelot	South America	medium	10
penguin	Antarctica	medium	20
quokka	Oceania	small	5
red panda	Asia	small	12
scorpion	Africa	small	4
Tasmanian devil	Oceania	medium	5
uakari	South America	medium	15
vulture	North America	medium	20
wallaby	Oceania	medium	9
x-ray tetra	South America	small	4
yak	Asia	large	20
zebra	Africa	large	25

a. How many variables in this data set are categorical? (1 MARK)

b. How many variables in this data set are ordinal? (1 MARK)

c. List the medium-sized animals from Africa. (1 MARK)

Adapted from VCAA 2018 Exam 2 Data analysis Q1

Back to contents

1D Log scales and graphs

STUDY DESIGN DOT POINTS

- 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

1A 1B 1C **1D** 1E 1F 1G 1H 1I

KEY SKILLS

During this lesson, you will be:

- calculating logarithmic values
- displaying data using a logarithmic scale
- interpreting data displayed using a logarithmic scale.

KEY TERMS

- Logarithms
- Logarithmic scale

There are times in which the range of data collected is very large and is not feasible to plot on a graph. When data involves values from multiple orders of magnitude (i.e. 1, 10, 100, 1000), it can be difficult to plot them on the same set of axes and still be able to see the distribution clearly. Logarithms can convert these values into smaller numbers. For example, the Richter scale measures the amplitude of earthquakes, which are often very large values. Calculating the log returns a magnitude value, which can then be plotted on a graph.

Calculating logarithmic values

Logarithms, commonly referred to as logs, are a mathematical operation. Each logarithm has a base, an argument, and an exponent.

An expression written in logarithmic form can be written in exponent form and vice versa.

If...

$$\log_{\underbrace{b}_{\textbf{base}}}(\overbrace{x}^{\textbf{argument}}) = \underbrace{y}_{\textbf{exponent}}$$

then...

$$b^y = x$$

For example, $\log_{10}(1000) = 3$ since $10^3 = 1000$. In this case, the base is 10, the exponent is 3 and the argument is 1000.

This course only focuses on logarithms with a base of 10. If there is no base specified, then it can be assumed that the base is 10.

If $x > 1$, then $\log(x)$ is positive.

If $x = 1$, then $\log(x)$ is zero.

If $0 < x < 1$, then $\log(x)$ is negative.

If $x \leq 0$, then $\log(x)$ is undefined.

Back to contents

Worked example 1

Use $\log_{10}$ for the following calculations.

a. Calculate the log of 23, correct to two decimal places.

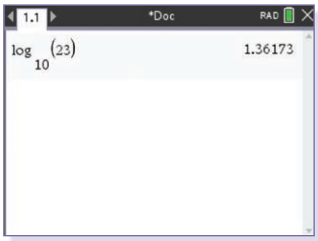
Explanation – Method 1: TI-Nspire

Step 1: From the home screen, select '1: New' → '1: Add Calculator'.

Step 2: Press ctrl + 10^x for the logarithmic function.

Enter in the base, '10', and the argument, '23'.

Press enter.



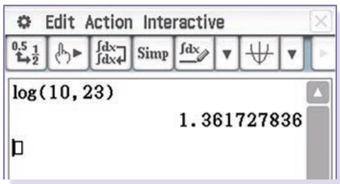
Step 3: Read the value from the screen and round to two decimal places.

Explanation – Method 2: Casio ClassPad

Step 1: From the main menu, tap √α Main.

Step 2: Press keyboard, tap $\log_{\square}\square$ and enter '23'.

Press EXE.



Step 3: Read the value from the screen and round to two decimal places.

Answer – Method 1 and 2

1.36

b. If the log of a number is 5, what is the number?

Explanation

Step 1: Define the unknown number as x.

$\log_{10}(x) = 5$

Step 2: Express this in exponent form and solve for x.

If $\log_{10}(x) = 5$, then $x = 10^5$.

$x = 10^5$

$x = 100\,000$

Answer

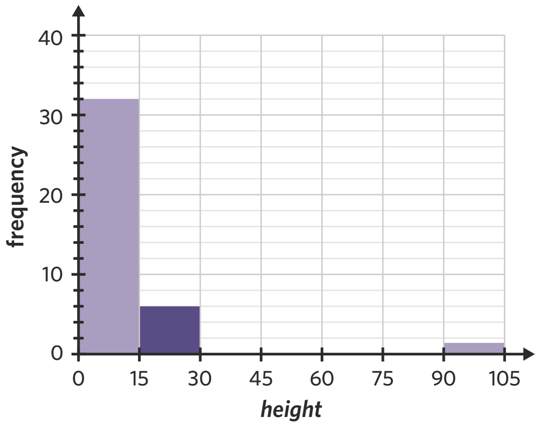
100 000

Back to contents

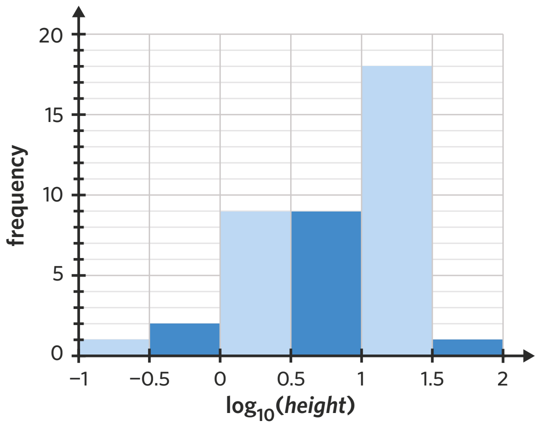
Displaying data using a logarithmic scale

A **logarithmic scale** is a scale with a $\log_{10}$ transformation. It can be used to reveal details about the underlying distribution of data sets with multiple orders of magnitude.

The *height* (cm) of 39 plants are shown on a non-logarithmic scale.



The *height* (cm) of the same plants are shown on a logarithmic scale.



The distribution of values previously hidden within the 0–<15 cm interval is now visible.

Worked example 2

The approximate weights of 20 different vehicles (cars, boats, planes) were recorded in kilograms.

840	910	990	1120	1490	1670	1690	1790	2430	2590
2810	2850	2940	3100	3590	3640	3660	5190	380 000	450 100

Construct a log scale histogram using the data. The histogram should start at 2.9 and have a column width of 0.2.

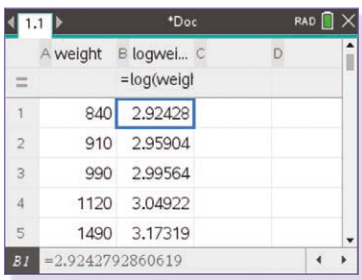
Explanation – Method 1: TI-Nspire

Step 1: From the home screen, select '1: New' → '4: Add Lists & Spreadsheet'.

Step 2: Name column A 'weight' and enter the data values into column A, starting from row 1.

Step 3: Name column B 'logweight'.

Enter '=log(weight)' into the cell below the 'logweight' heading.



Continues →

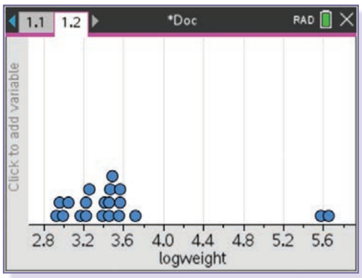
Back to contents

1D THEORY

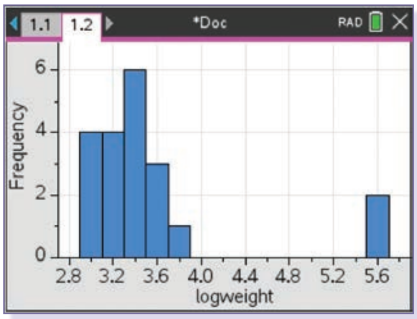
Step 4: Press ctrl + doc▾, and select '5: Add Data & Statistics'.

Move the cursor to the horizontal axis and select 'Click to add variable'.

Select 'logweight'.



Answer



Step 5: Press menu. Select '1: Plot Type' → '3: Histogram'.

Step 6: To adjust the column width and the starting point, press menu and then select '2: Plot Properties' → '2: Histogram Properties' → '2: Bin Settings' → '1: Equal Bin Width'.

Set the column width to 0.2 by changing 'Width' to '0.2'.

Set the starting point to 2.9 by changing 'Alignment' to '2.9'.

Select 'OK'.

Note: To change the view of the histogram, press menu and use options within '5: Window/Zoom'.

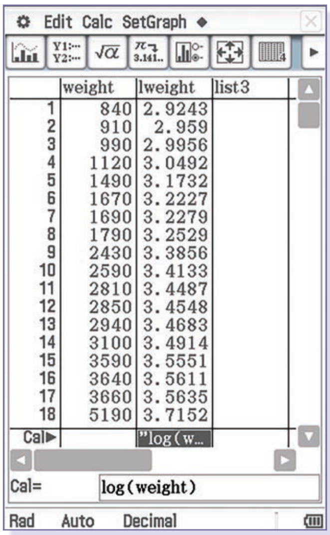
Explanation – Method 2: Casio ClassPad

Step 1: From the main menu, tap Statistics.

Step 2: Rename list1 to 'weight' and enter the data values starting from row 1.

Step 3: Rename list2 'lweight'.

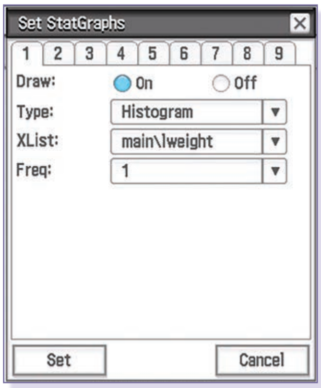
Go down to the calculation cell Cal▶ and enter 'log(weight)'.



Step 4: Configure the settings of the graph by tapping.

Create a histogram by changing 'Type:' to 'Histogram'.

Specify the data set by changing 'XList:' to 'main\lweight'.



Tap 'Set' to confirm.

Step 5: Tap in the icon bar to plot the histogram.

Set the starting point to 2.9 by changing 'HStart' to '2.9'.

Set the column width to 0.2 by changing 'HStep' to '0.2'.

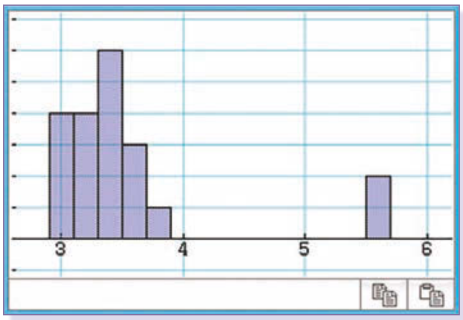
Tap 'OK' to confirm.

Note: To change the view of the histogram, press.

Continues →

Back to contents

Answer



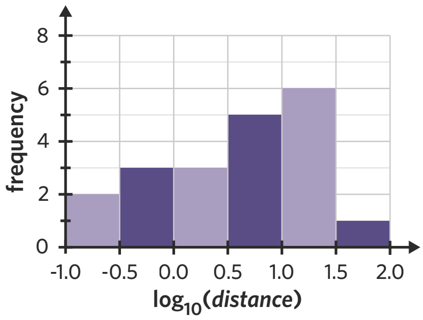
Note: To analyse the histogram, tap in the icon bar to place a (+) marker at the first column. 'xc=2.9' denotes the starting point of the first column (2.9) and 'Fc=4' shows the frequency of this interval (4).

Interpreting data displayed using a logarithmic scale

The height of a column in a histogram displayed using a logarithmic scale continues to represent the frequency of the corresponding interval. However, to interpret the frequency of an interval, it is important to consider the log transformation that has been applied to the horizontal axis.

Worked example 3

A class of twenty students were asked about the *distance*, in kilometres, from their home to the school. The results were summarised in a histogram using a $\log_{10}$ scale.



a. Jane was away the day of the survey. It is known that she lives 9.5 km away from school. Which interval on the histogram reflects this?

Explanation

Step 1: Calculate the appropriate log value.

$\log_{10}(9.5) = 0.9777...$

Step 2: Identify the appropriate interval on the graph.

This would be plotted between 0.5 and 1.0.

Answer

0.5–<1.0

Continues →

Back to contents

b. How many students live more than 10 km away from school?

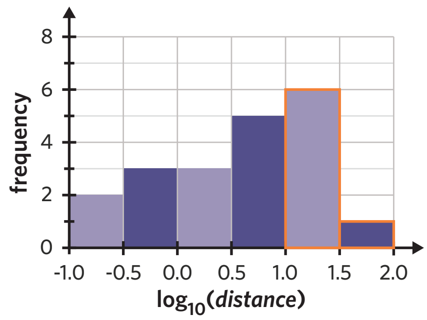
Explanation

Step 1: Calculate the appropriate log value.

$\log_{10}(10) = 1$

This means that on a log scale, 10 is plotted as 1.

Step 2: Sum the columns greater than 1.



$6 + 1 = 7$

Answer

7 students

Exam question breakdown

VCAA 2016 Exam 1 Data analysis Q7

The following histogram shows the distribution of the *number* of billionaires per million people plotted on a $\log_{10}$ scale.

Based on this histogram, the number of countries with one or more billionaires per million people is

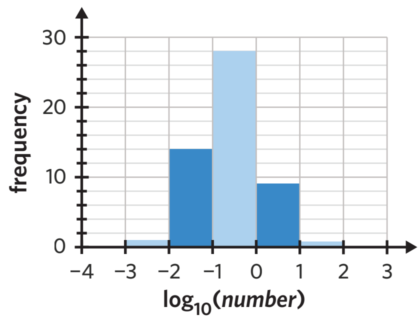
A. 1

B. 3

C. 8

D. 9

E. 10



Data: Gapminder

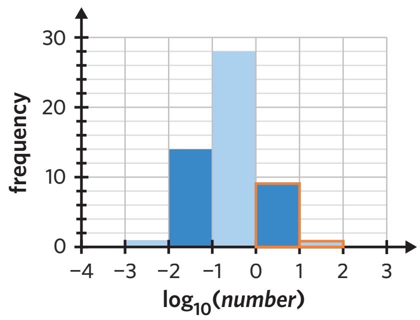
Explanation

Step 1: Calculate the appropriate log value.

$\log_{10}(1) = 0$

This means that on a log scale, 1 is plotted as 0.

Step 2: Sum the columns higher than 0.



$9 + 1 = 10$

Answer

E

45% of students answered this question correctly.

29% of students incorrectly chose option A. These students likely ignored the log scale and counted the column greater than 1, rather than the columns greater than $\log_{10}(1)$.

Back to contents

1D Questions

Calculating logarithmic values

1. What is the value of $\log_{10}(10)$?

A. -1 **B.** 0 **C.** 1 **D.** 10

2. Calculate the log values of the following, correct to two decimal places.

a. 0.05 **b.** 0.5 **c.** 5 **d.** 50
e. 3800 **f.** 380 **g.** 38 **h.** 3.8

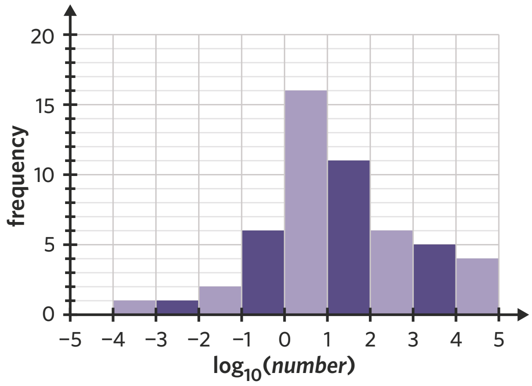
3. Calculate the value of x, correct to two decimal places.

a. $\log_{10}(x) = -0.5$ **b.** $\log_{10}(x) = 0.5$ **c.** $\log_{10}(x) = 1.5$

Displaying data using a logarithmic scale

4. Consider the following histogram with a $\log_{10}$ scale.
If a student wishes to add the *number* 10 358 to the distribution, which interval would need to be adjusted?

A. 1–<2
B. 2–<3
C. 3–<4
D. 4–<5



5. The following table shows the revenues of six businesses.

***revenue* ($)**	100 000	300 000	135 000	120 000	450 000	175 000
$\log_{10}$(*revenue*)	5.00	5.48	5.13	5.08	5.65	5.24

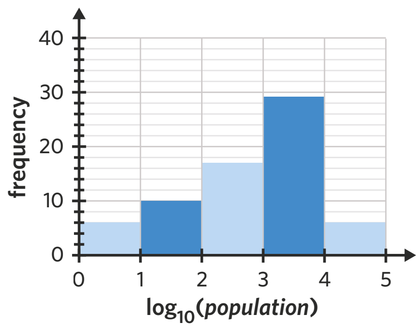
Construct a histogram representing *revenue* with a $\log_{10}$ scale. The histogram should start at 5 and have a column width of 0.1.

6. The *weight* (kg) of animals at the jungle party were recorded. Use a calculator to construct a $\log_{10}$ scale histogram using the data, with an appropriate starting point and a column width of 0.2.

50 56 84 95 87 64 512 204 983 10 200 124 94 43 302 9521

Interpreting data displayed using a logarithmic scale

7. Statisticians have collected data on the *population* of 68 of rural communities. Their data has been displayed as a histogram on a $\log_{10}$ scale.



a. How many communities have a *population* of at least 10 000 people?

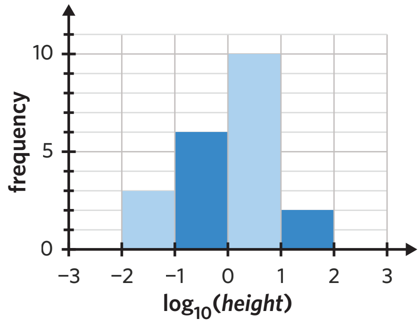
A. 0
B. 3
C. 6
D. 62

b. The percentage of communities with a *population* of less than 10 people is closest to

A. 6% **B.** 8% **C.** 9% **D.** 11%

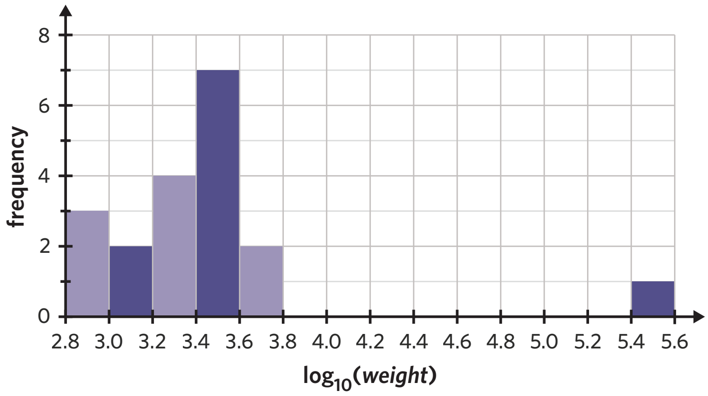
Back to contents

8. The *height* (m) of 21 oak trees have been measured by an environmentalist. The data has been displayed as a histogram with a $\log_{10}$ scale.
   a. How many trees are at least 1 metre tall?
   b. What percentage of trees are shorter than 1 metre? Round to two decimal places.



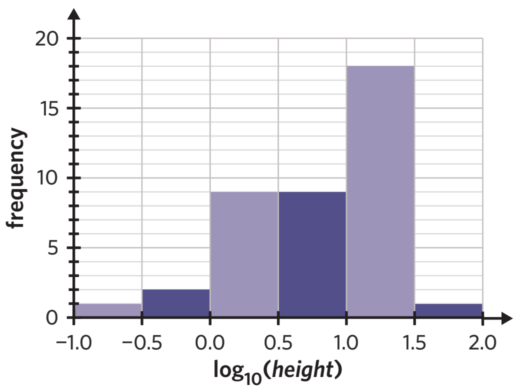
Joining it all together

9. The approximate *weight* of different vehicles (cars, boats, planes) were recorded in kilograms. The data was displayed by a histogram using the $\log_{10}$ scale.



   a. What *weight* is represented by 3.6 on the log scale, correct to the nearest kg?
   b. A truck is known to have a *weight* of 3750 kg. In which interval will this value be plotted?

10. It is the year 3000. Many different alien species have been found across the universe. The $\log_{10}$(*height*), with *height* measured in metres, of a single alien from each species was recorded and displayed in the following histogram.



   a. An alien from planet Zigzagzoon is 1.5 m tall. What is its log value, correct to two decimal places?
   b. How many aliens are at least 10 m tall?
   c. An alien from planet Bazinga is found to have a $\log_{10}$(*height*) value of -0.7, whereas an alien from planet Plazong has a $\log_{10}$(*height*) of 1.6. How much taller (in metres) is the alien from planet Plazong than the alien from planet Bazinga, correct to two decimal places?

Back to contents

11. The weights of 10 vehicles were recorded in kilograms.

a. Using a $\log_{10}$ scale, display the data in the frequency table as a histogram with a column width of 1. Label axes as appropriate.

b. Interpret the value of the modal interval.

vehicle	***weight* (kg)**
bicycle	15
motorcycle	230
truck	6500
van	1500
plane	55 000
hatchback car	1200
sedan car	1300
limousine	2800
scooter	4
tram	32 000

Exam practice

12. The following 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 small animal species with a *weight* of less than 10 000 g is

A. 17%

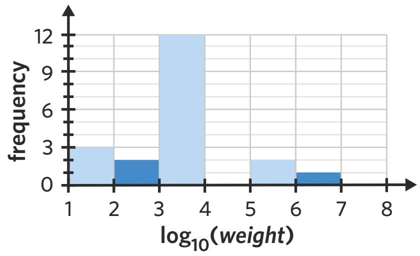
B. 70%

C. 75%

D. 80%

E. 85%

VCAA 2020 Exam 1 Data analysis Q5



78% of students answered this question correctly.

13. The following 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 100 000 people is

A. 1

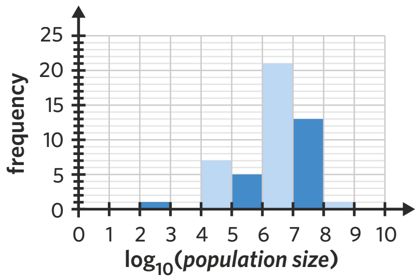
B. 5

C. 7

D. 8

E. 48

VCAA 2019 Exam 1 Data analysis Q3



Data: Worldometers, <www.worldometers.info/>

71% of students answered this question correctly.

14. The following histogram shows the distribution of the $\log_{10}(area)$ with *area* in square kilometres, of 17 islands.

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

A. 2 and 3

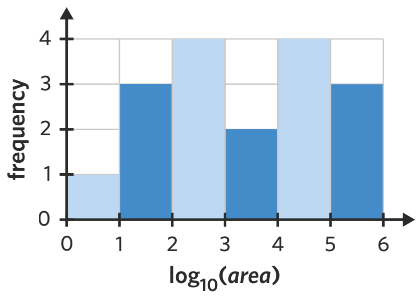
B. 2 and 3, as well as 4 and 5

C. 2 and 5

D. 10 000 and 100 000

E. 100 and 1000, as well as 10 000 and 100 000

Adapted from VCAA 2017 Exam 1 Data analysis Q4



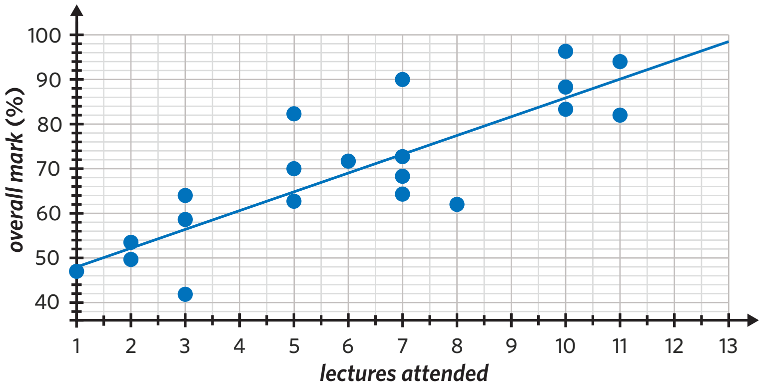
62% of students answered this type of question correctly.

Back to contents

Questions from multiple lessons

Data analysis *Year 11 content*

15. The following scatterplot displays the relationship between *overall mark*, as a percentage, and the number of *lectures attended* by 20 students in a particular unit at university. A line of good fit has been drawn.



The equation of the line of good fit is closest to

A. *overall mark* $=$ 43.49 $-$ 4.23 $\times$ *lectures attended*

B. *overall mark* $=$ 48.24 $+$ 4.23 $\times$ *lectures attended*

C. *overall mark* $=$ 48.24 $-$ 3.98 $\times$ *lectures attended*

D. *overall mark* $=$ 43.49 $+$ 3.98 $\times$ *lectures attended*

E. *overall mark* $=$ 43.49 $+$ 4.23 $\times$ *lectures attended*

Adapted from VCAA 2018 Exam 1 Data analysis Q8

Recursion and financial modelling *Year 11 content*

16. Delores inherits a large sum of money and decides to deposit it in a new savings account which earns interest every year. The following geometric sequence models the value of Delores' investment, in dollars, in each successive year.

850 500, 867 510, 884 860.2, ...

The balance of Delores' savings account after 15 years is closest to

A. \$1 105 650

B. \$1 108 202

C. \$1 110 753

D. \$1 122 217

E. \$1 144 661

Adapted from VCAA 2013 Exam 1 Number patterns Q4

Recursion and financial modelling *Year 11 content*

17. Denton wants to save up some money to buy a new house. He opens a savings account and deposits some money which will earn compound interest every year.

The balance of Denton's account, in dollars, after n years, V_n, can be modelled by the recurrence relation

$V_0 = 87\,500, \quad V_{n+1} = 1.034\,V_n$

a. How many dollars did Denton initially invest? (1 MARK)

b. What is the balance of Denton's account after one year? (1 MARK)

c. How many years will it be until the balance of Denton's account exceeds \$100 000? (1 MARK)

Adapted from VCAA 2018 Exam 2 Recursion and financial modelling Q4

Back to contents

1E The five-number summary and boxplots

STUDY DESIGN DOT POINT

- summary of the distributions of numerical variables; the five-number summary and boxplots (including the use of the lower fence ($Q_1 - 1.5 \times IQR$) and upper fence ($Q_3 + 1.5 \times IQR$) 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

1A 1B 1C 1D **1E** 1F 1G 1H 1I

KEY SKILLS

During this lesson, you will be:

- calculating the five-number summary
- calculating the range and interquartile range
- identifying outliers
- constructing and interpreting boxplots.

KEY TERMS

- Five-number summary
- Minimum
- Maximum
- Median
- Quartiles
- Spread
- Range
- Interquartile range (IQR)
- Outliers
- Fence
- Boxplot

Boxplots show the distribution of a data set based on the five-number summary, instead of displaying the frequency of data values or intervals of data values. The data set is split into quartiles to help visualise the centre and spread of a data set, as well as the existence of any outliers.

Calculating the five-number summary

The **five-number summary** provides key information about a set of data and its distribution including spread and centre. The summary is as follows:

Minimum, Q_1, Median, Q_3, Maximum

The **minimum** is the smallest value in the data set. The **maximum** is the largest value in the data set. It can be helpful to order the data when finding the maximum and minimum values.

The **median** is the middle value in an ordered set of data.

If there are n data values, the median is located at the $\left(\frac{n+1}{2}\right)^{\text{th}}$ position.

When there is an even number of data values, the median will be the average of the two middle data values.

While the median divides a distribution in half, **quartiles** divide a distribution in quarters. The symbols used to refer to the quartiles are Q_1, Q_2 and Q_3.

- Q_1 is the median of the lower half of the data set. The median, Q_2, is excluded if there is an odd number of values.
- Q_2 is the median of the entire data set.
- Q_3 is the median of the upper half of the data set. The median, Q_2, is excluded if there is an odd number of values.

Odd number of values:

3	5	6	6	9	11	15	12	16	17	20
↑		↑			↑			↑		↑
minimum		Q_1			median			Q_3		maximum

Back to contents

Even number of values:

4	5	5	7	8 ⋮	10	11	14	18	19
↑		↑		↑			↑		↑
minimum		Q_1		median			Q_3		maximum

Worked example 1

Construct a five-number summary for the following data.

3 5 1 10 8 9 6 3 8 6

Explanation – Method 1: By hand

Step 1: Arrange the data in ascending order.

1 3 3 5 6 6 8 8 9 10

Step 2: Identify the minimum and maximum values.

$minimum = 1$

$maximum = 10$

Step 3: Determine the median.

Count the number of values in the data set.

$n = 10$

Position of median: $\frac{10 + 1}{2} = 5.5$

The median is the average of the 5^{th} and 6^{th} values.

$median = \frac{6 + 6}{2} = 6$

Step 4: Determine the value of Q_1 and Q_3.

Q_1 is the median of the lower half of the data set.

Lower half: 1 3 3 5 6

$Q_1 = 3$

Q_3 is the median of the upper half of the data set.

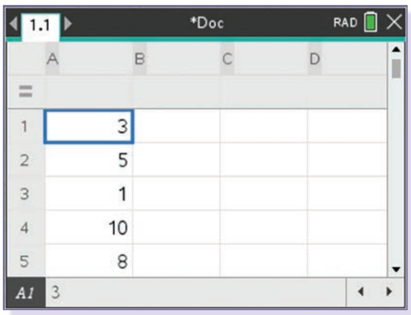
Upper half: 6 8 8 9 10

$Q_3 = 8$

Explanation – Method 2: TI-Nspire

Step 1: From the home screen, select ‘1: New’ → ‘4: Add Lists & Spreadsheet’.

Enter the data values into column A, starting from row 1.



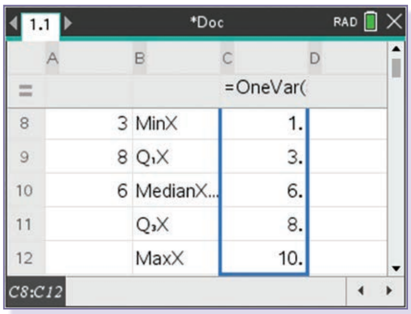
Step 2: Press menu. Select ‘4: Statistics’ → ‘1: Stat Calculations’ → ‘1: One-Variable Statistics’.

Select ‘OK’ to confirm one-variable statistics for one data set only.

Step 3: Specify the data set by entering 'a[]' in ‘X1 List:’.

Select ‘OK’ to exit this window and generate the statistics.

Scroll down to find the five-number summary statistics.



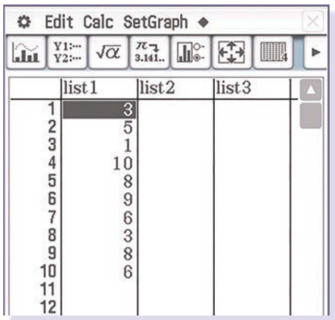
Continues →

Back to contents

Explanation – Method 3: Casio ClassPad

Step 1: From the main menu, tap Statistics.

Enter the data values into list1, starting from row 1.

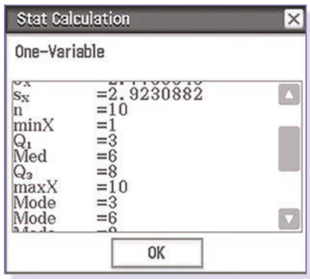


Step 2: Tap 'Calc' → 'One-Variable'.

Specify the data set by keeping 'XList:' as 'list1'.

Tap 'OK' to confirm.

Step 3: Scroll down to find the five-number summary statistics.



Answer – Method 1, 2 and 3

1, 3, 6, 8, 10

Calculating the range and interquartile range

The **spread** of a data set refers to how variable or similar the values are.

The **range** is a measure of spread of an entire data set. It is the difference between the maximum and minimum values, even if they are outliers.

$range = maximum - minimum$

The **interquartile range (IQR)** is a measure of the spread of the middle 50% of a data set. It is sometimes more accurate as a measure of spread than range because it does not include the upper 25% and lower 25% of values. This means that it is rarely affected by outliers.

$IQR = Q_3 - Q_1$

Worked example 2

The following stem plot displays the number of *goals scored* by 16 netball teams over a season.

Key: 2 | 2 = 22 goals

2	2	3	5	7	9		
3	0	1	3	5	7	7	8
4	0	1	2	2			

a. Calculate the range of *goals scored*.

Explanation

Step 1: Identify the minimum and maximum values.

Key: 2 | 2 = 22 goals

2	2	3	5	7	9		
3	0	1	3	5	7	7	8
4	0	1	2	2			

$minimum = 22$

$maximum = 42$

Step 2: Calculate the range.

$$\begin{aligned} range &= maximum - minimum \\ &= 42 - 22 \\ &= 20 \end{aligned}$$

Answer

20 goals

Continues →

Back to contents

b. Calculate the IQR of *goals scored*.

Explanation

Step 1: Determine the lower half and upper half of the data set.

$n = 16$

Position of median: $\frac{16 + 1}{2} = 8.5$

Key: 2 | 2 = 22 goals

2	2	3	5	7	9			
3	0	1	3	5	7	7	8	
4	0	1	2	2				

Lower half: 22 23 25 27 29 30 31 33

Upper half: 35 37 37 38 40 41 42 42

Step 2: Determine the value of Q_1 and Q_3.

Q_1 is the median of the lower half of the data set.

$Q_1 = \frac{27 + 29}{2} = 28$

Q_3 is the median of the upper half of the data set.

$Q_3 = \frac{38 + 40}{2} = 39$

Step 3: Calculate the IQR.

$IQR = Q_3 - Q_1$

$= 39 - 28$

$= 11$

Answer

11 goals

Identifying outliers

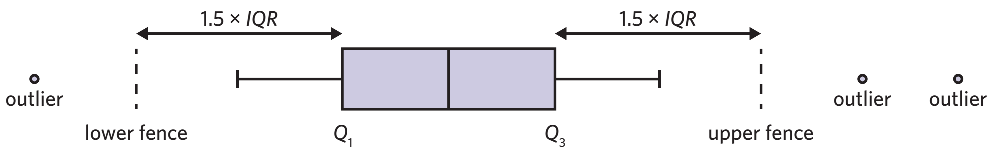
Outliers are values which fall outside of what is normal or reasonable. Outliers can be identified by a series of tests.

A **fence** defines the boundary of what is an outlier, and what is a regular data value.

$lower\ fence = Q_1 - (1.5 \times IQR)$

$upper\ fence = Q_3 + (1.5 \times IQR)$

If a value is less than the lower fence or greater than the upper fence, it is considered to be an outlier. Outliers can still be reported as the maximum or minimum value of a data set.



Note: It is not necessary to mark in the fences when creating a boxplot.

Worked example 3

The *age* of 14 children on a cruise ship was recorded.

12 2 4 10 7 10 8 16 8 7 10 1 15 9

a. Calculate the lower and upper fences of the data set.

Explanation

Step 1: Arrange the data in ascending order.

1 2 4 7 7 8 8 9 10 10 10 12 15 16

Step 2: Determine the lower half and upper half of the data set.

$n = 14$

Position of median: $\frac{14 + 1}{2} = 7.5$

Lower half: 1 2 4 7 7 8 8

Upper half: 9 10 10 10 12 15 16

Continues →

Back to contents

Step 3: Determine the value of Q_1 and Q_3.

Q_1 is the median of the lower half of the data set.

$Q_1 = 7$

Q_3 is the median of the upper half of the data set.

$Q_3 = 10$

Step 4: Calculate the IQR.

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 10 - 7 \\ &= 3 \end{aligned}$$

Step 5: Calculate the lower and upper fences.

$$\begin{aligned} \textit{lower fence} &= Q_1 - (1.5 \times IQR) \\ &= 7 - (1.5 \times 3) \\ &= 2.5 \end{aligned}$$

$$\begin{aligned} \textit{upper fence} &= Q_3 + (1.5 \times IQR) \\ &= 10 + (1.5 \times 3) \\ &= 14.5 \end{aligned}$$

Answer

Lower fence: 2.5 years old

Upper fence: 14.5 years old

b. Identify any outliers.

Explanation

1 and 2 are both less than the lower fence (2.5).

15 and 16 are both greater than the upper fence (14.5).

Answer

1, 2, 15, 16

Constructing and interpreting boxplots

A **boxplot** is a graphical representation of a five-number summary as well as any outliers.

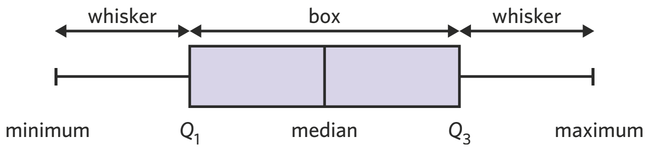
See worked example 4

If there are no outliers, the leftmost and rightmost ends of the whiskers represent the minimum and maximum values. If there is an outlier, the whisker ends at the most extreme value that is not an outlier.

The left and right borders of the box represent Q_1 and Q_3 respectively.

The vertical line in the centre of the box represents the median.

Outliers are indicated by dots which lie outside the range of the box and whiskers.



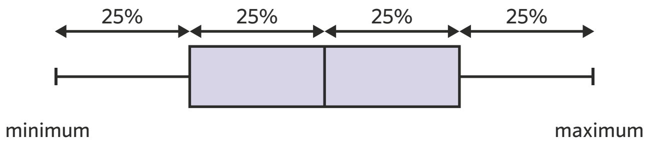
Boxplots can display data sets with a large number of numerical values. A calculator can be helpful in creating a boxplot directly from the data.

See worked example 5

The distance between two consecutive boundaries represents 25% of data. The following intervals all represent 25% of data.

See worked example 6

- Minimum to Q_1
- Q_1 to median
- Median to Q_3
- Q_3 to maximum



Worked example 4

The five-number summary for the following stem plot is:

1, 7, 16.5, 23, 39

Key: 1 | 2 = 12

0	1	2	3	5	5	9
1	0	0	1	5	8	9
2	1	2	3	3	5	
3	3	7	9			

Construct a boxplot using the data.

Explanation

Step 1: Check for any outliers.

Calculate the IQR.

$$IQR = Q_3 - Q_1$$
$$= 23 - 7$$
$$= 16$$

Calculate the lower fence.

$$lower\ fence = Q_1 - (1.5 \times IQR)$$
$$= 7 - (1.5 \times 16)$$
$$= -17$$

Calculate the upper fence.

$$upper\ fence = Q_3 + (1.5 \times IQR)$$
$$= 23 + (1.5 \times 16)$$
$$= 47$$

Locate any outliers.

All data values lie between the lower and upper fences. Therefore, there are no outliers.

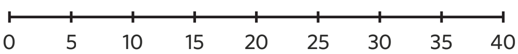
Step 2: Construct an axis with an appropriate scale.

The scale should cover a range such that it shows all data and outliers.

Intervals should be easy to work with.

Data ranges from 1 to 39, so an appropriate scale would run from 0 to 40.

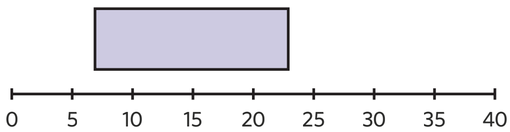
An interval of five makes the larger range easier to work with.



Step 3: Draw the border of the box.

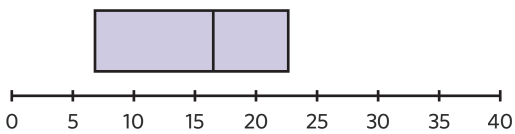
The value of the left border is Q_1 and the value of the right border is Q_3. As given by the five-number summary, Q_1 is 7 and Q_3 is 23.

The height of the box is not important.



Step 4: Mark in the vertical line.

The value of the vertical line in the middle of the box is the median. As given by the five-number summary, the median is 16.5.

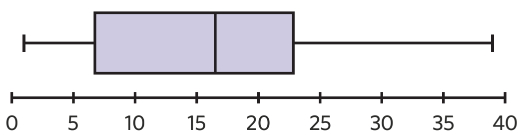


Step 5: Draw the whiskers.

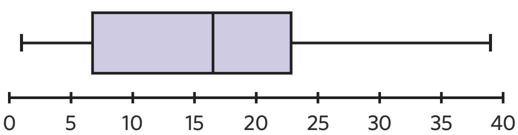
As there is no outlier, the left whisker ends at the minimum value and the right whisker ends at the maximum value.

Draw a whisker ranging from the minimum (1) to the leftmost point of the box.

Draw a whisker ranging from the rightmost point of the box to the maximum (39).



Answer



Back to contents

Worked example 5

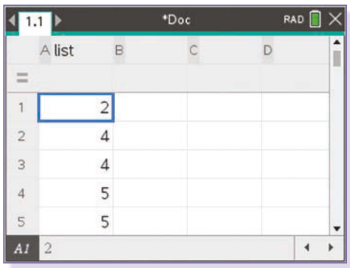
Construct a boxplot for the following data set.

2 4 4 5 5 5 7 6 5 4 6 5 6 7 8 12 3 5

Explanation – Method 1: TI-Nspire

Step 1: From the home screen, select ‘1: New’ → ‘4: Add Lists & Spreadsheet’.

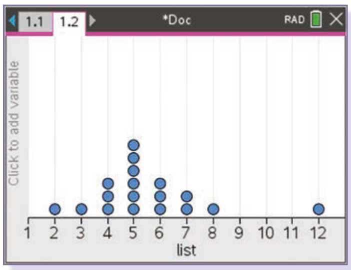
Step 2: Name column A 'list' and enter the data values into column A, starting from row 1.



Step 3: Press ctrl + doc▾ and select ‘5: Add Data & Statistics’.

Move the cursor to the horizontal axis and select ‘Click to add variable’.

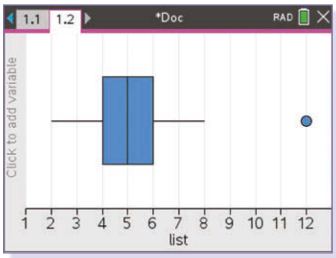
Select ‘list’.



Step 4: Press menu. Select ‘1: Plot Type’ → ‘2: Box Plot’.

Note: To change the view of the histogram press menu and use options within ‘5: Window/Zoom’.

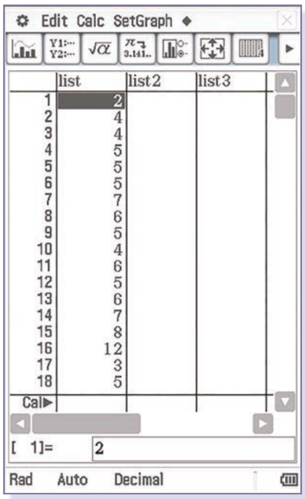
Answer



Explanation – Method 2: Casio ClassPad

Step 1: From the main menu, tap Statistics.

Step 2: Rename list1 to 'list' and input the data values starting from row 1.

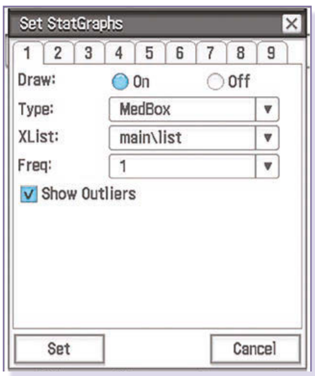


Step 3: Configure the settings of the graph by tapping .

Create a histogram by changing ‘Type:’ to ‘MedBox’.

Specify the data set by changing ‘XList:’ to ‘main\list’.

Tick the ‘Show Outliers’ box.

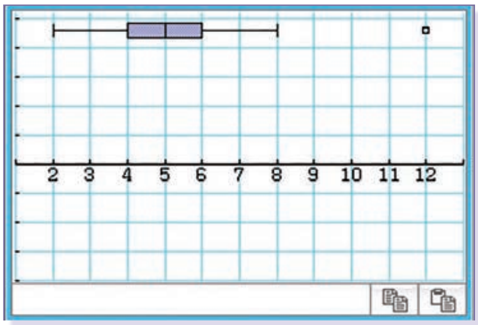


Tap ‘Set’ to confirm.

Step 4: Tap in the icon bar to plot the boxplot.

Continues →

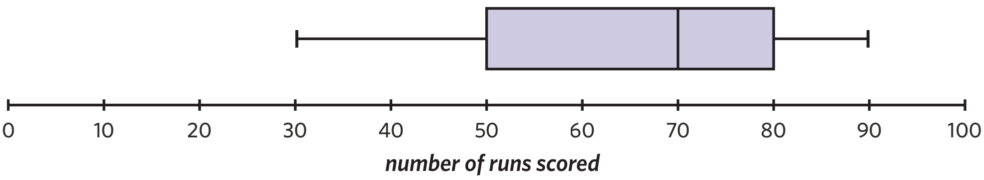
Answer



Worked example 6

The *number of runs scored* by the leading batsman at a local cricket club is represented by the following boxplot.

It is known that data was collected from 20 separate innings.



a. Determine the five-number summary for the data and identify any outliers.

Explanation

Step 1: Locate any potential outliers.

There are no potential outliers indicated on the boxplot.

Step 2: Locate and identify the minimum and maximum values.

As there are no outliers, the leftmost and rightmost ends of the whiskers represent the minimum and maximum values respectively.

$minimum = 30$

$maximum = 90$

Step 3: Locate and identify the value of Q_1 and Q_3.

The left and right borders of the box represent Q_1 and Q_3 respectively.

$Q_1 = 50$

$Q_3 = 80$

Step 4: Locate and identify the median.

The vertical line in the centre of the box represents the median.

$median = 70$

Answer

30, 50, 70, 80, 90

There are no outliers.

b. Between which two values does the middle 50% of data lie?

Explanation

The middle 50% of data lies between Q_1 and Q_3.

$Q_1 = 50$

$Q_3 = 80$

Answer

50 and 80 runs

Continues →

Back to contents

c. In approximately what percentage of innings did they score more than 80 runs?

Explanation

$Q_3 = 80$

25% of data is greater than Q_3.

Answer

25%

d. In approximately how many innings did they score at least 50 runs, but fall short of 100?

Explanation

Step 1: Identify the percentage of data that lies above 50 runs but less than 100 runs.

$Q_1 = 50$

$maximum = 90$

75% of data lies between Q_1 and the maximum.

Step 2: Calculate the number of innings.

There were 20 separate innings.

Calculate 75% of 20:

$0.75 \times 20 = 15$

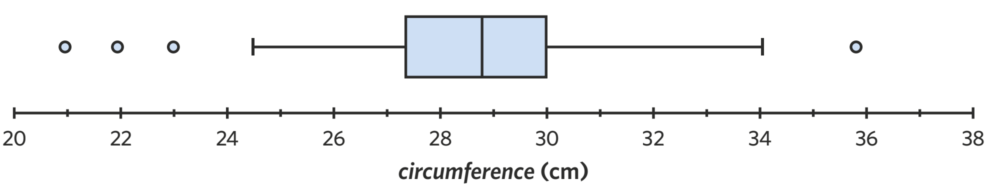
Answer

15 innings

Exam question breakdown

VCAA 2017 Exam 1 Data analysis Q2

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



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

Explanation

Locate and identify the minimum and maximum values.

As there are outliers, the leftmost and rightmost dots represent the minimum and maximum values respectively.

$minimum = 21$

$maximum = 35.9$

The only option with a minimum of 21 and maximum of 35.9 is option B.

Answer

B

58% of students answered this question correctly.

38% of students incorrectly answered option C. Students who answered C ignored the outlier points when determining the maximum and minimum values for the five-number summary. Even though the points are identified as outliers, they are still valid data points within the data set and must be used as maximum and minimum values if appropriate.

Back to contents

1E Questions

Calculating the five-number summary

1. Find the value of Q_3 for the following data set.

7 9 3 8 4 10 2 6 6 4

A. 4 **B.** 6 **C.** 8 **D.** 10

---

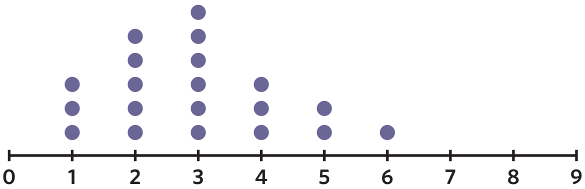
2. A physics class made paper aeroplanes to test aerodynamics. The *distance* (m) that each paper aeroplane flew was recorded.

34 26 27 13 18 16 6 11 25 23 33 31 9 28 20 17 32 35 9 35

Construct a five-number summary for the data.

---

3. Construct a five-number summary for the following dot plot.



---

4. Construct a five-number summary for the following stem plot.

Key: 1 | 7 = 17

0	7 7 8
1	2 3 5 7 9
2	1 1 3 5
3	1 2
4	1

Calculating the range and interquartile range

5. Which of the following statements is true?

A. The IQR is always greater than the range.
B. The IQR is always greater than or equal to the range.
C. The IQR is always less than the range.
D. The IQR is always less than or equal to the range.

---

6. The following data shows the amount of money, correct to the nearest dollar, that Alice spent at her school canteen each day over 2 weeks.

1 4 0 3 6 7 4 0 4 2

a. Find the range of the amount of money Alice spent.

b. Find the IQR.

7. At the annual cheese rolling festival, the *finishing time*, in seconds, of each cheese is recorded in the following stem plot.

Key: 5 | 1 = 5.1 secs

5	1 6 6 9
6	2 2 3 7 9
7	2 8
8	1

What is the IQR?

Identifying outliers

8. 100 Year 12 students were asked how much money they have spent on games in the past year.

The five-number summary for the results is: 0, 50, 90, 230, 395.

The lower fence for the data is

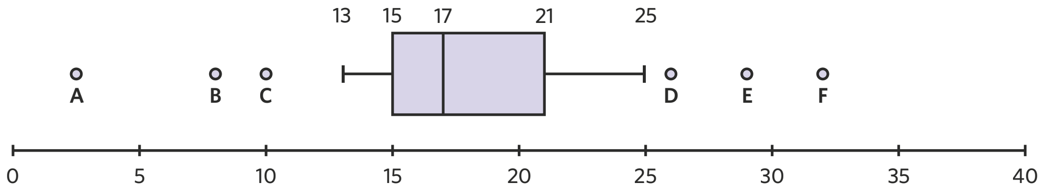
A. −\$220

B. −\$130

C. \$0

D. \$50

9. Identify which data values have been incorrectly marked as outliers on the boxplot.



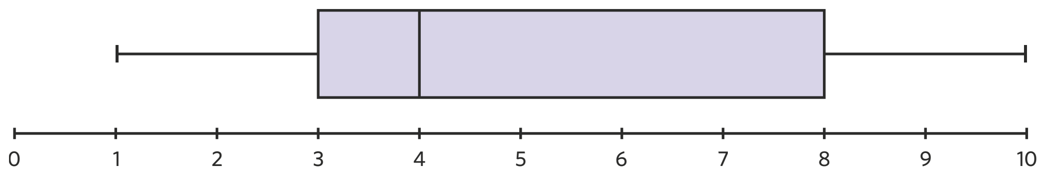
10. Identify any outliers in the following data sets.

a. 3 7 7 8 8 9 9 9 10 14 15

b. 25 24 28 20 30 24 25 24 26 26 19 24

Constructing and interpreting boxplots

11.



The five-number summary for the data shown in the boxplot is

A. 1, 3, 4, 8, 10

B. 3, 3, 4, 8, 8

C. 1, 4, 4, 4, 10

D. 2, 3, 4, 8, 9

Back to contents

12. A group of 20 students, each from different schools, competed in an interschool diving competition. They each performed one dive and received a *score* from 1 to 100.

62 100 39 50 93 25 53 53 44 43 40 28 36 30 84 84 59 100 58 76

a. Use a calculator to construct a boxplot from the data.

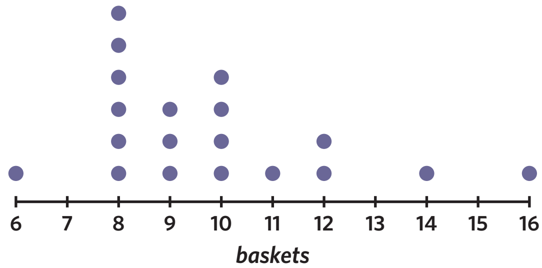
b. The percentage of students that received a *score* of more than 80 is closest to

A. 25%
B. 50%
C. 75%
D. 100%

c. The percentage of students that received a *score* between 39.5 and 80 is closest to

A. 25%
B. 50%
C. 75%
D. 100%

13. A team of basketball players held a mini tournament to see how many *baskets* they can shoot in one minute. The results are shown in the dot plot.



The five-number summary is: 6, 8, 9, 11, 16.

a. Construct a boxplot by hand to represent the data in the dot plot.

b. The percentage of players that scored more than 9 *baskets* is closest to

A. 25%
B. 50%
C. 75%
D. 100%

c. The percentage of players that scored 11 or less *baskets* is closest to

A. 25%
B. 50%
C. 75%
D. 100%

Joining it all together

14. The *goal tally* of the top 20 forwards in the 2000 AFL season is recorded.

51 47 52 70 57 53 56 76 42 53 50 109 49 43 47 44 42 69 40 68

a. Construct a five-number summary for the data.

b. Calculate the IQR.

c. Identify any outliers.

d. Construct a boxplot by hand to represent the data.

Back to contents

15. The owner of a record store documents the number of *records sold* on 17 different days.

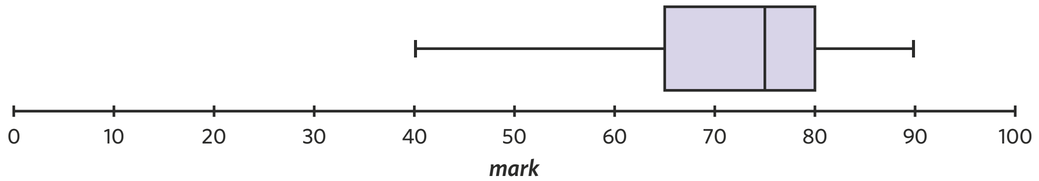
12 11 7 15 8 10 11 13 24 7 16 13 11 15 8 12 7

a. Use a calculator to construct a boxplot.

b. Identify any outliers.

c. Construct a five-number summary for the data.

16. The *mark*, out of 100, for 20 students who sat a university entrance exam was recorded.



a. Use the boxplot to estimate the percentage of students that scored between:

i. 40 and 65

ii. 65 and 80

iii. 75 and 90

iv. 65 and 90

b. The data is correct, but the actual boxplot has been drawn incorrectly. Identify the error and redraw the boxplot given that the second lowest score was 50, the second highest score was 85 and Q_1, the median, and Q_3 are correctly positioned.

Exam practice

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

Key: 4 | 2 = 42 n = 23

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

For this class, the interquartile range (IQR) of *test scores* is

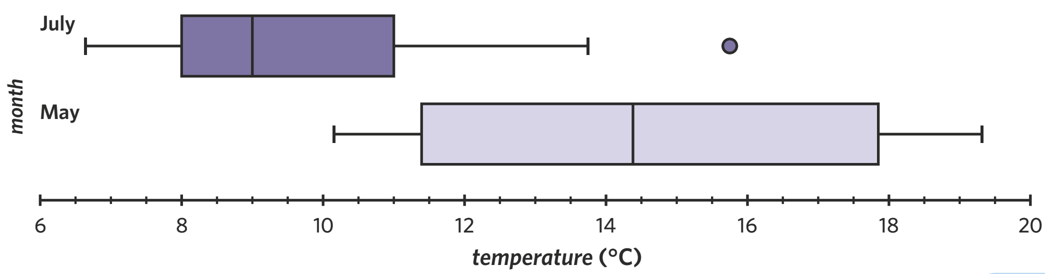
A. 14.5 **B.** 17.5 **C.** 18

D. 24 **E.** 49

VCAA 2019 Exam 1 Data analysis Q5

87% of students answered this question correctly.

18. The following boxplots display the distribution of maximum daily *temperature* for the months of May and July.

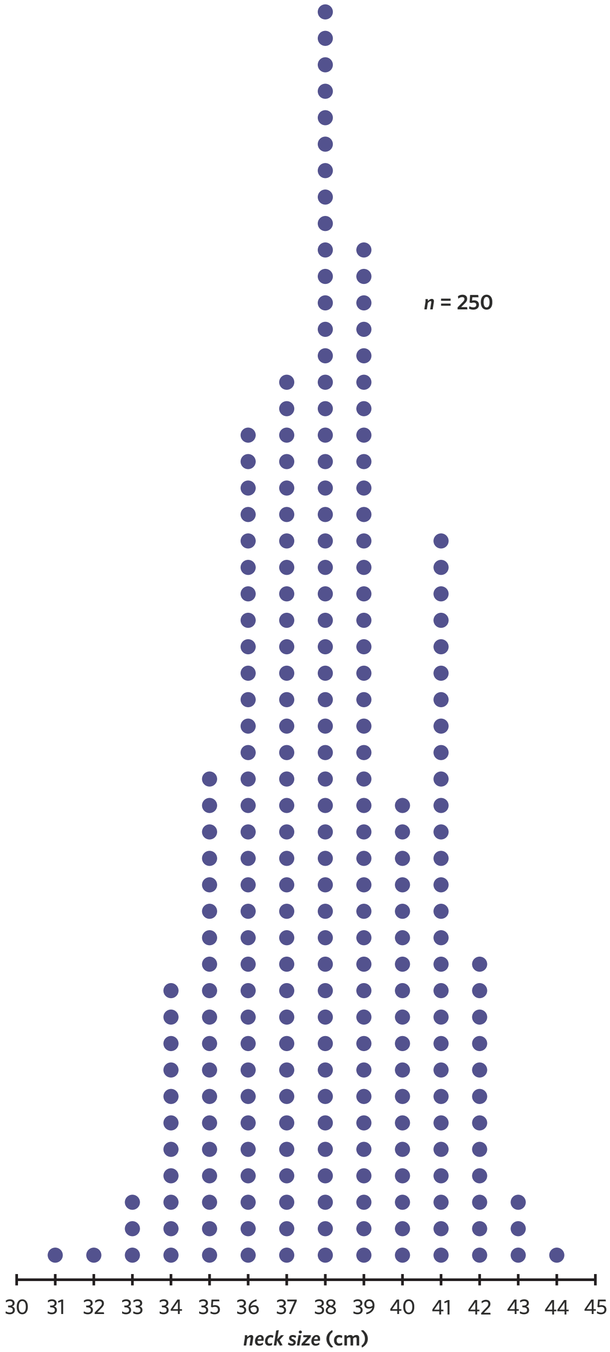


Determine the value of the upper fence for the July boxplot. (1 MARK)

VCAA 2016 Exam 2 Data analysis Q2bii

60% of students answered this question correctly.

19. The *neck size*, in centimetres, of 250 men was recorded and displayed in the following dot plot.



The five-number summary for this sample of neck sizes, in centimetres, is given in the following table.

minimum	first quartile (Q_1)	median	third quartile (Q_3)	maximum
31	36	38	39	44

Use the five-number summary to construct a boxplot, showing any outliers if appropriate. (2 MARKS)

VCAA 2020 Exam 2 Data analysis Q2c

The average mark on this question was **1.1**.

Back to contents

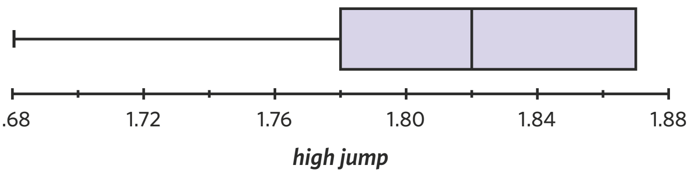
20. In the sport of heptathlon, athletes compete in seven events.

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.

The following boxplot 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)

VCAA 2021 Exam 2 Data analysis Q1e

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	40.53
5	1.87	13.78	40.62
6	1.73	14.50	45.62
7	1.68	15.08	42.33
8	1.82	13.13	40.88
9	1.83	14.22	39.22
10	1.87	13.62	42.51
11	1.87	12.01	42.75
12	1.80	12.88	38.12
13	1.83	12.68	42.65
14	1.87	12.45	41.32
15	1.78	11.31	42.88

47% of students answered this question correctly.

Questions from multiple lessons

Data analysis *Year 11 content*

21. George works for a furniture company that has a number of stores across Victoria. He is asked to investigate the association between *number of couches sold* and *store location* (Shepparton, Fitzroy, Vermont, Bundoora). These variables are

A. a numerical variable and an ordinal variable respectively.

B. a numerical variable and a nominal variable respectively.

C. an ordinal variable and a numerical variable respectively.

D. both numerical variables.

E. both categorical variables.

Adapted from VCAA 2017NH Exam 1 Data analysis Q4

Recursion and financial modelling *Year 11 content*

22. Sheldon invests \$20 000 for a total of 365 days.

His interest compounds daily and he does not touch his investment for its entire duration.

If it reaches a value of \$20 506.28 at the end of the 365 days, the interest rate per annum is closest to

A. 1.0% p.a

B. 1.5% p.a.

C. 2.0% p.a.

D. 2.5% p.a.

E. 3.0% p.a.

Adapted from VCAA 2014 Exam 1 Business-related mathematics Q3

Back to contents

Data analysis *Year 11 content*

23. Students in a biology lab are experimenting to see the relationship between the number of eggs in a termite colony and its size by the end of the trial period.

The line of good fit for this data is:

size of colony $= -4 + 1.05 \times$ *number of eggs*

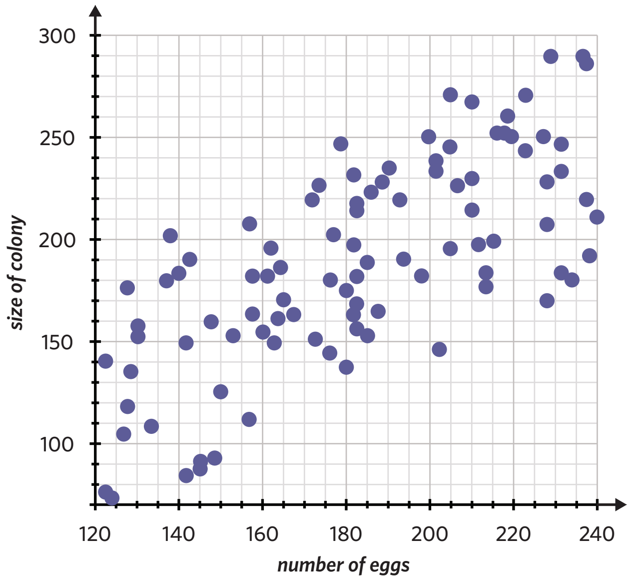
a. Draw the line of good fit on the scatterplot. (1 MARK)

b. Suppose the students had started one group with exactly 100 eggs. Estimate the population projection for this group of experiments.

Round your answer to the nearest whole number. (1 MARK)

c. In reference to the prediction made in part **b**, is this an example of interpolation or extrapolation? (1 MARK)

Adapted from VCAA 2018NH Exam 1 Data analysis Q5



Back to contents

1F Describing numerical data

STUDY DESIGN DOT POINTS

- 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

1A 1B 1C 1D 1E **1F** 1G 1H 1I

KEY SKILLS

During this lesson, you will be:

- describing the distribution of histograms
- describing the distribution of dot plots and stem plots
- identifying the best measure of centre
- describing the distribution of boxplots.

KEY TERMS

- Positively skewed
- Negatively skewed
- Symmetric
- Bimodal
- Centre

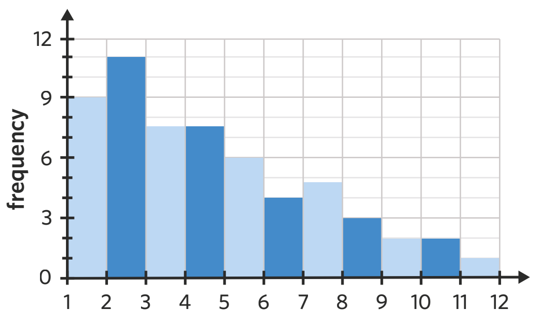
Once numerical data has been collected and graphed, it is important to be able to describe what it displays. This allows people using the data to understand exactly what the data is suggesting, and helps to convey underlying trends.

Describing the distribution of histograms

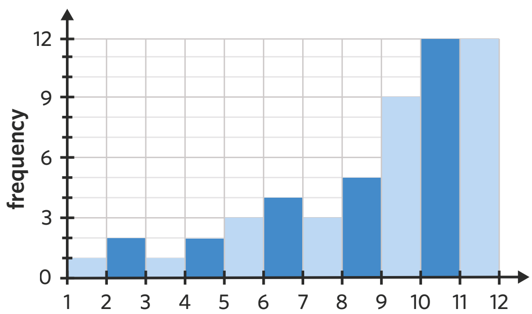
Numerical data distributions are generally described in terms of their shape, centre and spread.

Shape

A **positively skewed** distribution trails off in a positive direction on the horizontal axis.

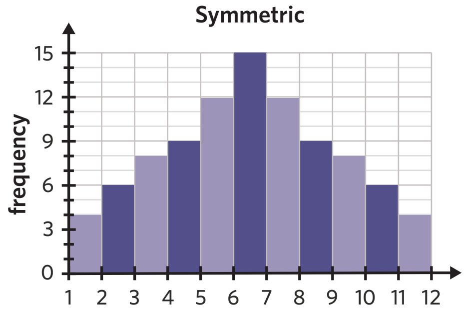


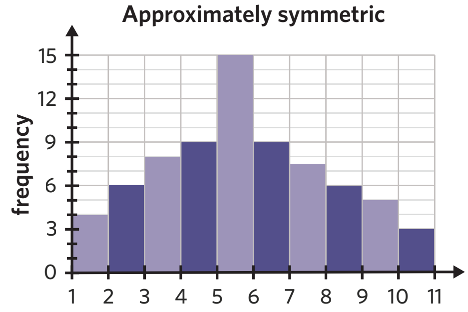
A **negatively skewed** distribution trails off in a negative direction on the horizontal axis.



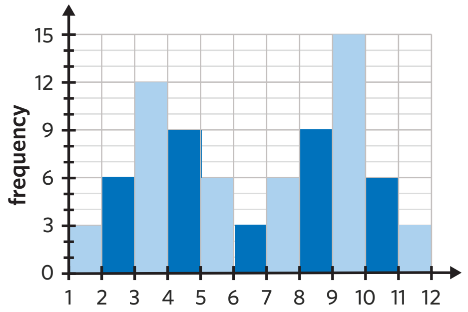
Back to contents

A **symmetric** distribution is the same on both sides of the centre. If the distribution isn't exactly symmetric, it is important to describe the shape as approximately symmetric.



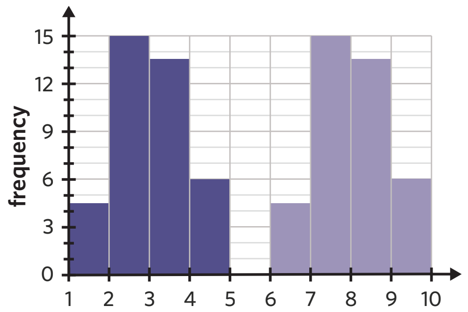


A symmetric or approximately symmetric distribution can also be **bimodal**. This occurs when there are two distinct peaks in the distribution. These peaks do not necessarily have to be equal.



Centre

The **centre** refers to the middle of a distribution. The following two distributions have the exact same shape, but different centres.



Either the mean or median can be used as a measure of centre. This lesson will focus on the median.

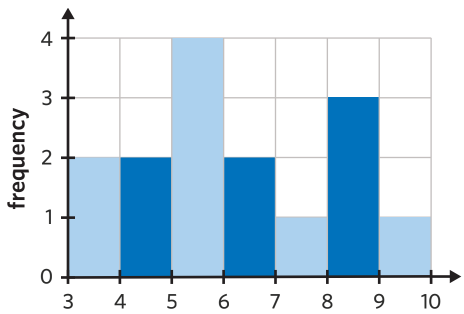
Spread

Recall that the spread refers to the spacing of values within a data set.
For histograms and other numerical distributions, the range can be used as the measure of spread.

range = *maximum value* − *minimum value*

This histogram has a range of 7.

range = 10 − 3 = 7

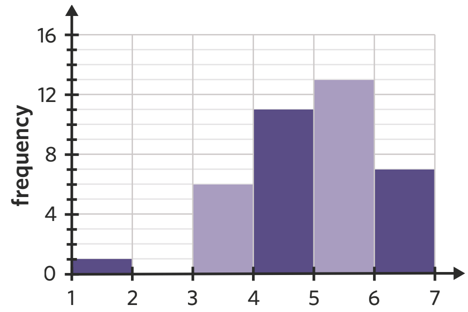


It is also important to identify potential outliers in a histogram. These are values that fall outside of what looks normal or reasonable, and can be identified by eye. This means they need to be referred to as potential outliers.

Back to contents

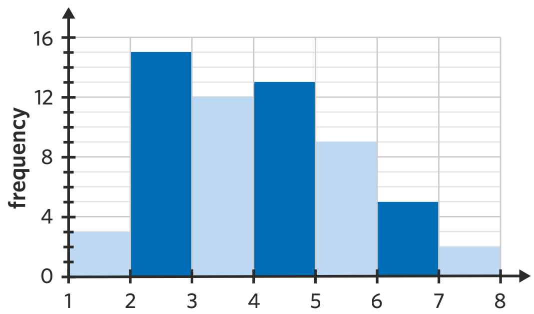
This data set has a potential outlier in the interval 1–<2.

Note: The interquartile range (IQR) is often used as a more reliable measure of spread when outliers are present. This lesson however only focuses on range as the measure of spread.



Worked example 1

Consider the histogram.



a. Estimate the median.

Explanation

Step 1: Calculate the number of data points by summing the frequency of each column.

$3 + 15 + 12 + 13 + 9 + 5 + 2 = 59$

Step 2: Determine the position of the median.

Remember that the median is located in the $\left(\frac{n+1}{2}\right)^{\text{th}}$ position.

In this case, $n = 59$.

$$\frac{59+1}{2} = \frac{60}{2}$$
$$= 30$$

Step 3: Determine in which interval the 30$^{\text{th}}$ data point lies by calculating the cumulative frequency for each interval.

interval	cumulative frequency
1–<2	3
2–<3	3 + 15 = 18
3–<4	18 + 12 = 30

The 30$^{\text{th}}$ data point occurs in the 3–<4 interval.

Step 4: Estimate the median.

$$\frac{3+4}{2} = \frac{7}{2}$$
$$= 3.5$$

Answer

Approximately 3.5

b. Describe the histogram in terms of shape, spread and potential outliers.

Explanation

Step 1: Determine the shape of the histogram.

The majority of the data is towards the left of the distribution.

The distribution trails off in the positive direction.

This means the histogram is positively skewed.

Step 2: Calculate the spread of the histogram.

$$\textit{range} = \textit{maximum value} - \textit{minimum value}$$
$$= 8 - 1$$
$$= 7$$

Continues →

Back to contents

Step 3: Identify any potential outliers.

There are no data values that fall outside of what looks reasonable.

Answer

The histogram is positively skewed with a range of 7 and no potential outliers.

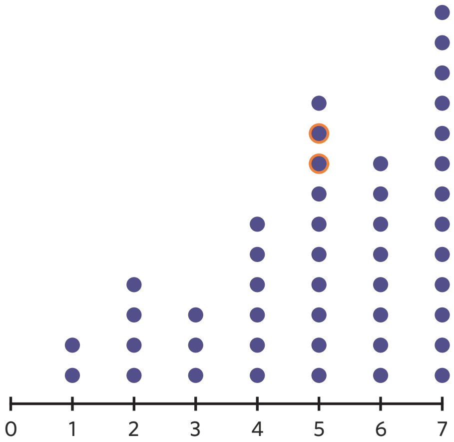
Describing the distribution of dot plots and stem plots

Dot plots and stem plots are also described using shape, centre and spread.

When calculating the median for dot plots and stem plots, the exact location of the median can be identified, unlike in histograms where only an approximate value can be calculated.

For example, consider the following dot plot.

See worked example 2



The median is the average of the 23rd and 24th values. These are highlighted on the dot plot.

The median is $\frac{5 + 5}{2} = 5$.

Stem plots are displayed vertically, not horizontally like histograms and dot plots. The distribution will still trail off in the same direction (in the positive or negative direction) for a positively or negatively skewed distribution.

See worked example 3

Positively skewed

Key: 4 | 0 = 40

4	0 1 4 6 7 7 8
5	3 4 4 5 6
6	0 1 1 3 9
7	1 4 5 8
8	2 6
9	1

Negatively skewed

Key: 4 | 0 = 40

4	0
5	3 4 4
6	0 1 1
7	1 4 5 8
8	2 3 3 5 7 9
9	1 2 4 4 6 7 8

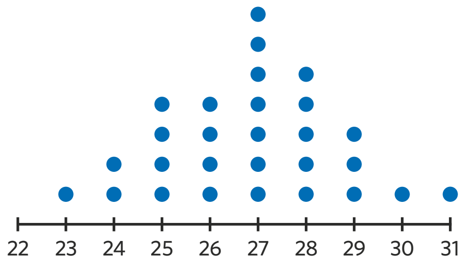
Approximately symmetric

Key: 4 | 0 = 40

4	0
5	3 4 4
6	0 1 1 3 9
7	1 4 5 8 8 9
8	2 6
9	1

Worked example 2

Consider the dot plot.



a. Determine the median.

Explanation

Step 1: Count the number of data points.

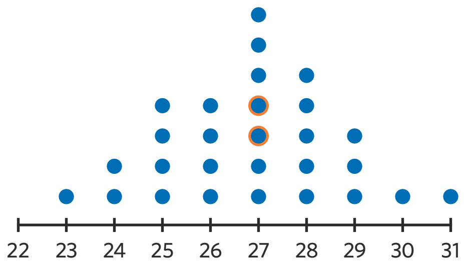
There are 28 data points.

Step 2: Determine the location of the median.

$$\frac{28 + 1}{2} = \frac{29}{2}$$
$$= 14.5$$

The median will be the average of the 14^{th} and 15^{th} data points.

Step 3: Locate the data points and calculate the median.



$$\frac{27 + 27}{2} = 27$$

Answer

27

b. Describe the dot plot in terms of shape, spread and potential outliers.

Explanation

Step 1: Determine the shape of the dot plot.

The majority of the data is in the middle of the distribution.

The distribution trails off in both directions approximately equally.

This means the dot plot is approximately symmetric.

Step 2: Calculate the spread of the dot plot.

$$range = maximum\ value - minimum\ value$$
$$= 31 - 23$$
$$= 8$$

Step 3: Identify any potential outliers.

There are no data values that fall outside of what looks reasonable.

Answer

The dot plot is approximately symmetric with a range of 8 and no potential outliers.

Back to contents

Worked example 3

Consider the stem plot.

Key: 40 | 7 = 407

40	7
41	
42	
43	
44	4 6
45	1 3 3
46	2 6 7
47	4 7 8 9
48	0 0 4 5 6 8
49	1 2 5 5 5 9
50	3 4 4 7 8 8 9 9

a. Calculate the median.

Explanation

Step 1: Count the number of data points.

There are 33 data points.

Step 2: Determine the location of the median.

$$\frac{33 + 1}{2} = \frac{34}{2} = 17$$

The median will be the 17th data point.

Step 3: Locate the data point.

Key: 40 | 7 = 407

40	7
41	
42	
43	
44	4 6
45	1 3 3
46	2 6 7
47	4 7 8 9
48	0 0 4 5 6 8
49	1 2 5 5 5 9
50	3 4 4 7 8 8 9 9

Answer

485

b. Describe the stem plot in terms of shape, spread and potential outliers.

Explanation

Step 1: Determine the shape of the stem plot.

The majority of the data is towards the larger values in the distribution.

The distribution trails off in the negative direction.

This means the stem plot is negatively skewed.

Step 2: Calculate the spread of the stem plot.

$$\begin{aligned} range &= maximum\ value - minimum\ value \\ &= 509 - 407 \\ &= 102 \end{aligned}$$

Step 3: Identify any potential outliers.

There is one data point (407) that appears to lie outside the normal range of data.

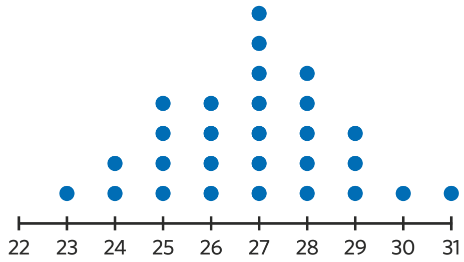
Continues →

Back to contents

Answer

The stem plot is negatively skewed with a range of 102 and one potential outlier.

Identifying the best measure of centre



Either the mean or the median can be used as the best measure of centre for a distribution of data.

The mean is the best measure of centre for distributions that have no outliers and are approximately symmetric.

For skewed distributions or distributions with outliers, the median is the best measure of centre as it refers to the middle value and doesn't include the numerical value of any outliers.

Consider the dot plot. The median is 27, while the mean is 26.86.

The distribution is approximately symmetric with no outliers. Therefore, the mean hasn't been affected by any skewed data, and provides a more precise measure of centre than the median.

Consider the stem plot.

Key: 5 | 1 = 5.1

0	2
1	
2	
3	
4	
5	1
6	6 6
7	2 5
8	1 6 7 8
9	4 5 5 6 8 9
10	2 4 5 8 8 8 9 9

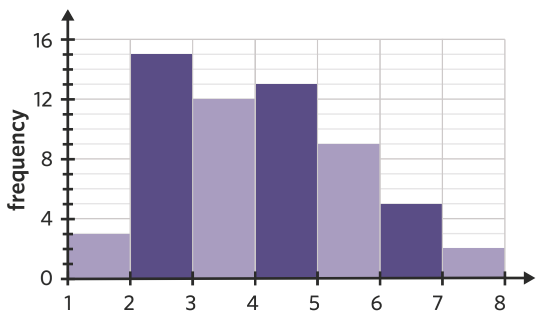
The median is 9.5, while the mean is 8.77. The mean is affected by the value of the outlier (0.2), so the median is the preferred measure of centre.

If the outlier is ignored, the median remains at 9.5 and the mean changes to 9.14. The median is still a better measure of centre because it is weighted more heavily to the bulk of the distribution.

Worked example 4

Identify the best measure of centre for the following distributions.

a.



Continues →

Back to contents

Explanation

Step 1: Identify the shape of the distribution.

The bulk of the data is towards the smaller values in the distribution.

The distribution trails off in the positive direction.

This means the histogram is positively skewed.

Step 2: Determine the best measure of centre.

When data is skewed, the best measure of centre is the median.

Answer

Median

b. **Key:** 4 | 0 = 40

4	0
5	3 4 4
6	0 1 1 3 9
7	1 4 5 8 8 9
8	2 6
9	1

Explanation

Step 1: Identify the shape of the distribution.

The majority of data is in the middle of the distribution.

The distribution trails off in both directions approximately equally.

This means the stem plot is approximately symmetric.

Step 2: Identify any potential outliers.

There are no data values that fall outside of what looks reasonable

Step 3: Determine the best measure of centre.

When data is approximately symmetric with no outliers, the best measure of centre is the mean.

Answer

Mean

Describing the distribution of boxplots

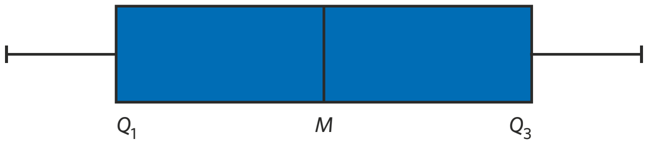
Boxplots can also be described in terms of their shape, centre and spread. The same principles apply as with histograms, dot plots and stem plots however the method of interpreting the shape is slightly different.

A positively skewed distribution trails off in the positive direction. For a boxplot, this means the median is located towards the left of the box. The left whisker will be short whereas the right whisker will be longer.

A negatively skewed distribution trails off in the negative direction. For a boxplot, this means the median is located towards the right of the box. The right whisker will be short whereas the left whisker will be longer.

Back to contents

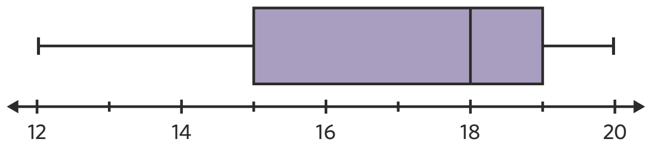
A symmetric distribution is evenly distributed and does not trail off in either direction more than the other. For a boxplot, this means the median is approximately in the centre of the box. The distances from the median to Q_1 and Q_3 are approximately equal, and the whiskers are also approximately equal.



Worked example 5

Consider the boxplot.

Describe the distribution in terms of shape, centre, spread and outliers.



Explanation

Step 1: Determine the shape of the boxplot.

The median is located towards the right of the box.

The distribution trails off in the negative direction.

This means the boxplot is negatively skewed.

Step 2: Identify the median of the boxplot.

The median is represented by the vertical line inside the box.

$median = 18$

Step 3: Calculate the spread of the boxplot.

$$\begin{aligned} range &= maximum\ value - minimum\ value \\ &= 20 - 12 \\ &= 8 \end{aligned}$$

Step 4: Identify any outliers.

There are no outliers in the boxplot.

Answer

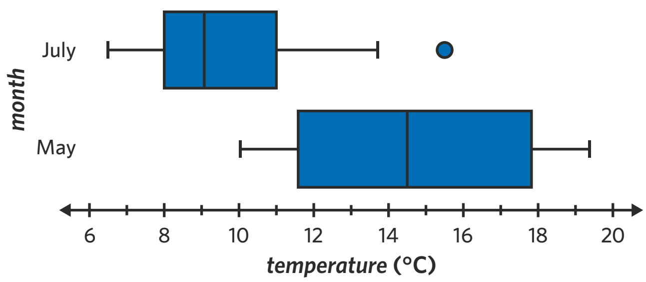
The boxplot is negatively skewed with no outliers, a median of 18 and a range of 8.

Exam question breakdown

VCAA 2016 Exam 2 Data analysis Q2bi

The following boxplots display the distribution of maximum daily *temperature* for the months of May and July.

Describe the shapes of the distributions of daily *temperature* (including outliers) for July and for May. (1 MARK)



Explanation

Step 1: Identify the shape of the distribution for July.

The median is located towards the left of the boxplot.

The distribution trails off in the positive direction.

The boxplot is positively skewed.

Step 2: Identify any outliers.

There is one outlier for July.

Step 3: Identify the shape of the distribution for May.

The median is located approximately in the middle of the boxplot.

Q_1 and Q_3 are located approximately the same distance from the median.

The whiskers are approximately the same length.

Step 4: Identify any outliers.

There are no outliers for May.

Continues →

Back to contents

Answer

July: Positively skewed with one outlier

May: Approximately symmetric with no outliers

56% of students answered this question correctly.

For this question, many students answered either symmetrically skewed, evenly distributed, bell-shaped or normally distributed for July. These are either incorrect descriptions of the shape of a distribution or inaccurate descriptions of the distribution for July.

1F Questions

Describing the distribution of histograms

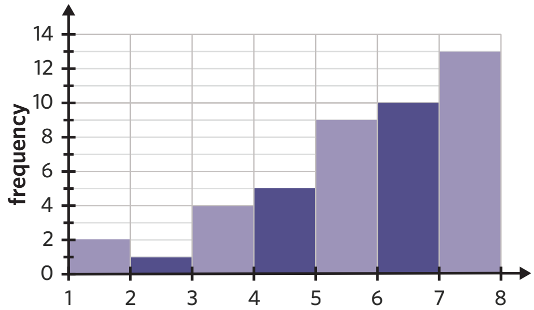
1. Describe the shape of the following histogram.

A. Positively skewed

B. Positively skewed with a potential outlier

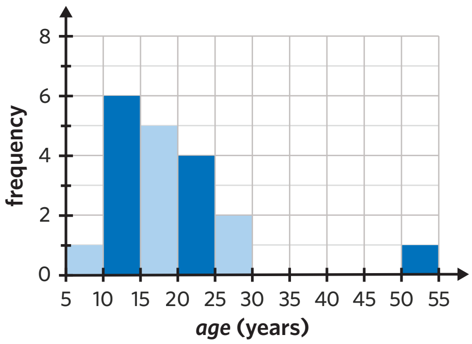
C. Negatively skewed

D. Approximately symmetric

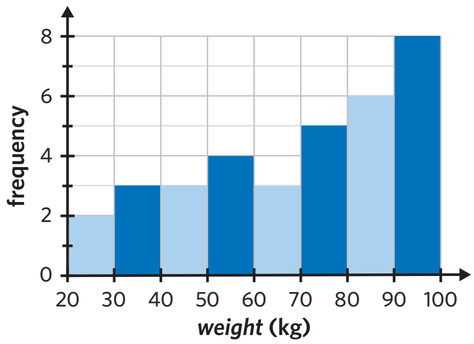


2. Describe the shape of each of the following histograms and identify any potential outliers.

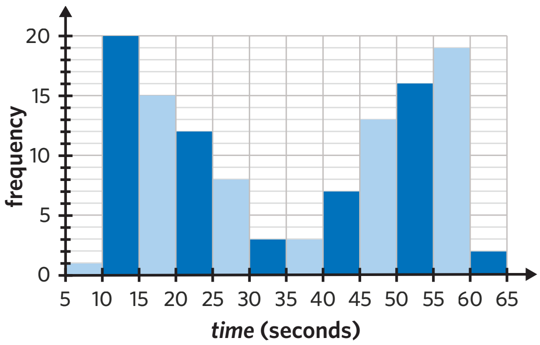
a.



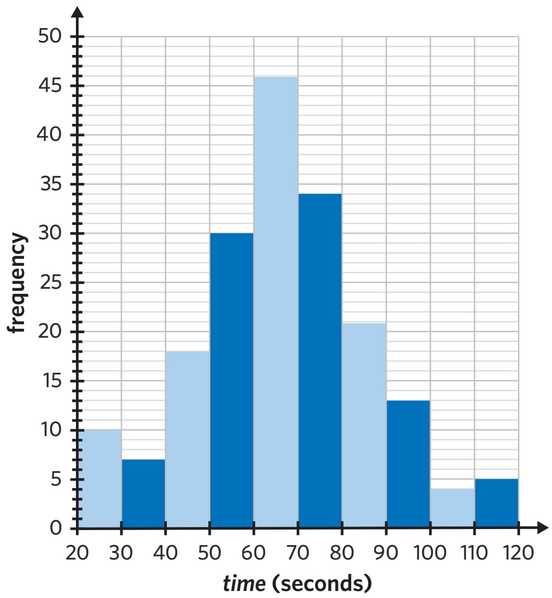
b.



c.



d.

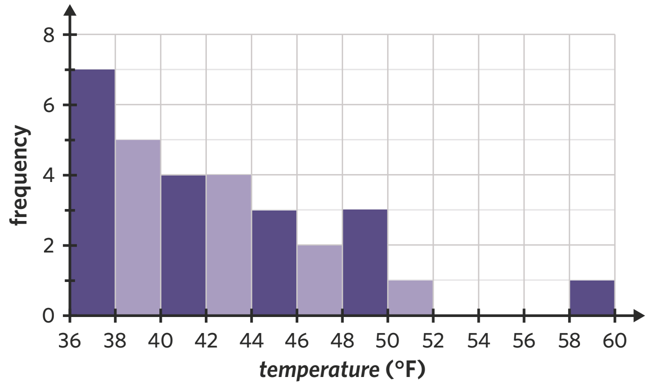


Back to contents

3. The *temperature*, in °F, in New York was measured every day for the 30 days of November. The data has been placed into the following histogram.

a. Estimate the median *temperature*.

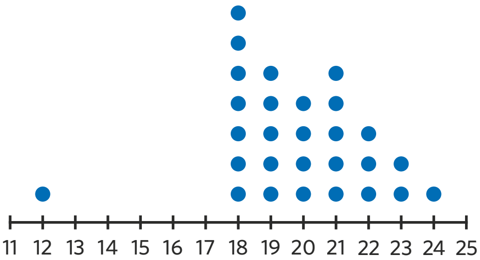
b. Describe the histogram in terms of shape, spread and potential outliers.



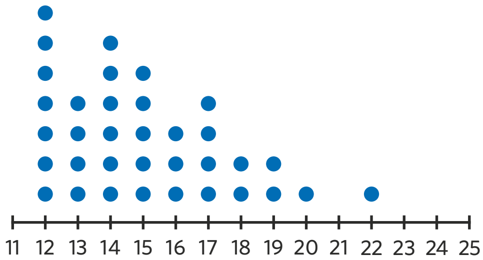
Describing the distribution of dot plots and stem plots

4. Which of the following dot plots is negatively skewed with a potential outlier?

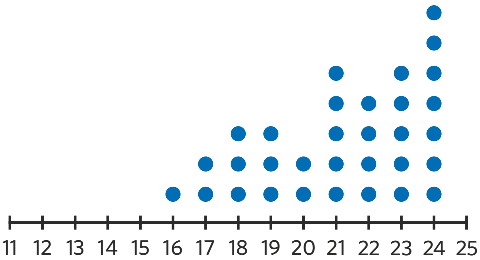
A.



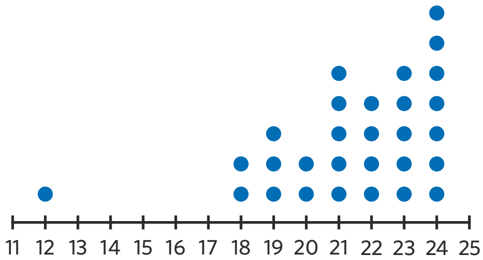
B.



C.

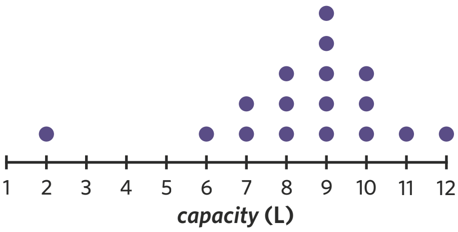


D.

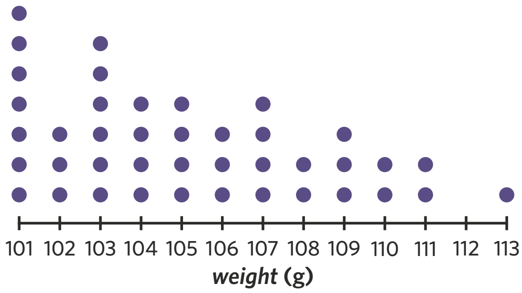


5. Describe the shape of each of the following dot plots and identify any potential outliers.

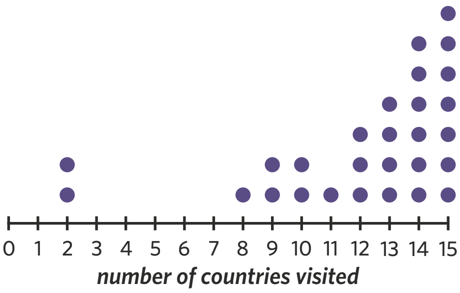
a.



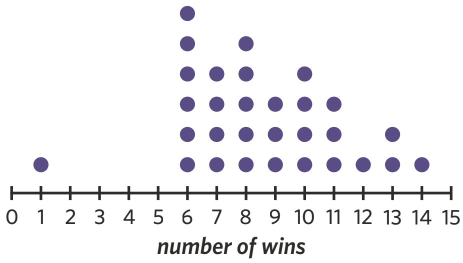
b.



c.



d.



Back to contents

6. Describe the shape of each of the following stem plots and identify any possible outliers.

a. **Key:** 1 | 7 = 17 cm

1	7
2	1 8
3	0 5 5
4	4 7
5	1 3 8
6	2 2 9
7	3 4 4 8
8	1 1 4 5 6 9

b. **Key:** 20 | 1 = 20.1 seconds

20	1
21	2 4 5
22	0 7 7
23	4 5 7 8 8
24	4 6 6 7 9 9
25	1 3 7 8
26	2 6
27	4 9
28	
29	
30	
31	9

c. **Key:** 9 | 0 = 90 days

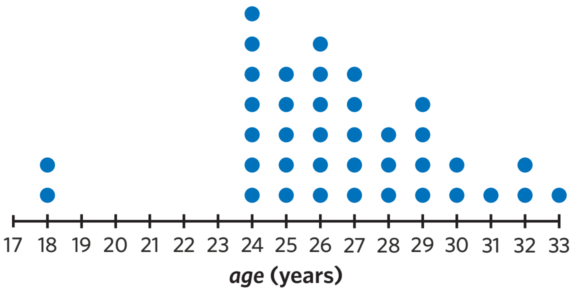
9	0 2 5 5 5 6 7 9
10	2 4 6 7 9 9
11	0 1 5
12	6 6 7 8
13	1 2 2
14	3
15	5
16	
17	
18	
19	4

d. **Key:** 200 | 4 = $2004

200	4
201	2 6
202	4 4 5 7
203	3 6 6 7 8
204	2 2 2 6 7
205	0 0 3 4 8 8 9
206	
207	
208	
209	3 5

---

7. Consider the following dot plot.



a. Calculate the median *age*.

b. Describe the dot plot in terms of shape, spread and potential outliers.

Back to contents

8. Sophie is a newspaper editor and records the *number of edits* she makes for each newspaper that gets printed. The results are shown in the following stem plot.

Key: 15 | 7 = 157

15	7
16	
17	
18	1 2
19	5 6 7 9
20	3 3 4 6 6 7
21	0 0 1 2 5 8 9 9
22	0 3 3 3 4 6 7 7 8 9

a. Calculate the median *number of edits*.

b. Describe the stem plot in terms of shape, spread and potential outliers.

Identifying the best measure of centre

9. Fill in the missing words.

The mean is the best measure of centre when the distribution is __________ with no __________.

A. negatively skewed, outliers

B. outliers, positively skewed

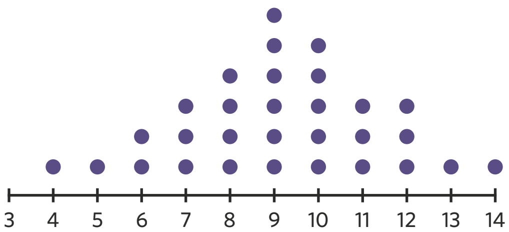
C. approximately symmetric, outliers

D. outliers, approximately symmetric

---

10. Identify whether the mean or median would be the best measure of centre for the following distributions.

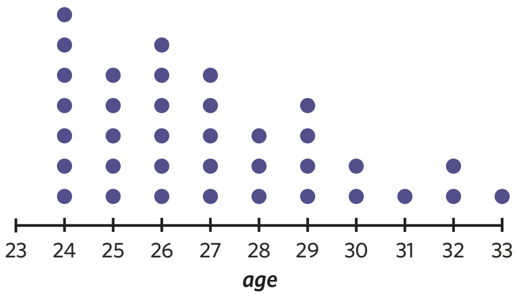
a.



b. **Key:** 0 | 3 = 3

0	3
1	
2	
3	
4	2 7
5	2 2 5
6	1 7 8
7	3 6 9 9
8	1 1 4 6 6 8
9	0 0 1 5 6 9
10	4 6 6 7 8 9 9 9

c.



d. **Key:** 50 | 2 = 50.2

50	2
51	3 4 6
52	3 8 8
53	0 3 6 8 8
54	2 2 6 7 8 9
55	0 1 8 9
56	2 6
57	5 8
58	
59	
60	
61	7

11. Rodney and Sally are discussing their answers to question **10d**. Rodney says the best measure of centre is the mean because the data is approximately symmetric. Sally says the best measure of centre is the median because there is a potential outlier. Who is correct and why?

Describing the distribution of boxplots

12. Fill in the blanks in the following statement.

For a negatively skewed boxplot, the median is towards the __________ of the __________.
The right whisker will be __________ whereas the left whisker will be __________.

A. left, centre, short, longer

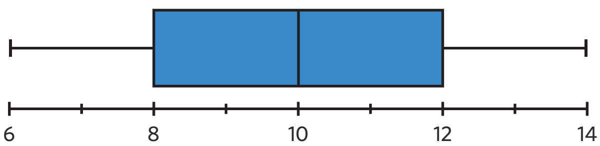
B. right, box, short, longer

C. middle, distribution, long, shorter

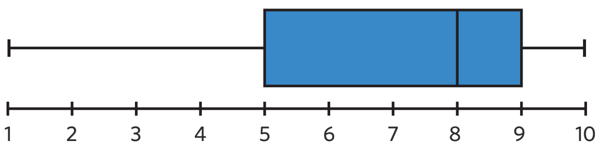
D. right, box, long, shorter

13. Describe the following boxplots in terms of shape, centre, spread and outliers.

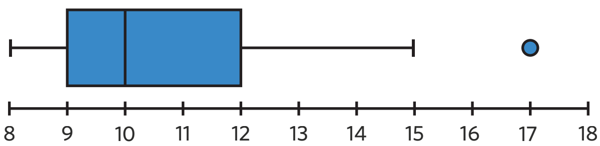
a.



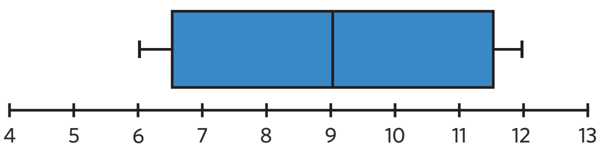
b.



c.



d.



Back to contents

Joining it all together

14. Which of the following statements is false?

A. A bimodal distribution has two clear peaks that are not necessarily equal.

B. The median can only be estimated from a histogram but can be calculated exactly for both dot plots and stem plots.

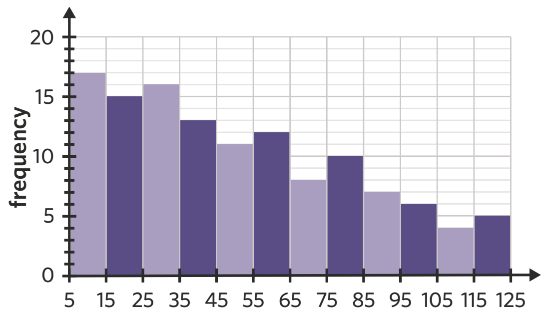
C. The median should always be used as the best measure of centre.

D. The median for a negatively skewed boxplot will be located towards the right side of the box.

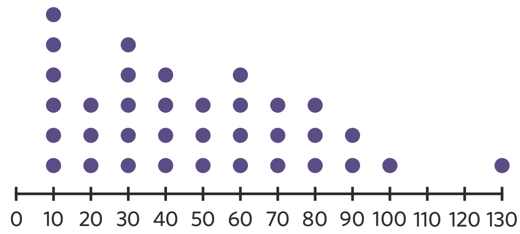
---

15. Sashi constructed a distribution that is positively skewed with an outlier. The distribution has a median of 50 and a range of 120. Which of the following could be the distribution that Sashi constructed?

A.



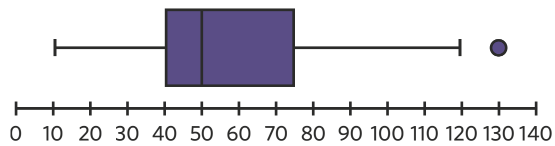
B.



C. **Key:** 1 | 8 = 18

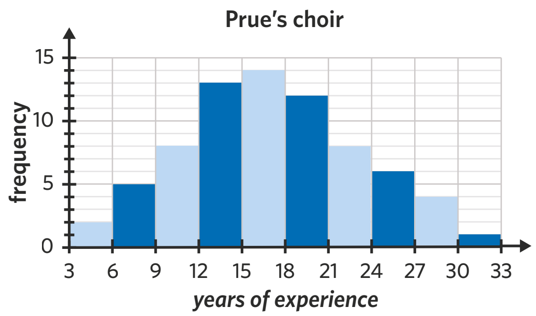
1	8
2	
3	
4	
5	5
6	3 4
7	0 1
8	5 7 8 8
9	1 1 6
10	3 4 8 9
11	0 0 0 5 8
12	2 5 6 6 6
13	1 2 5 7 7 8

D.



---

16. Prue and Gil are both singers, and are members of two different choirs. They each decided to collect data on the amount of singing experience each member of their choir has, in years. They produced the following distributions.



Gil's choir

Key: 0 | 6 = 6 years

0	6 7
1	1 4 4
2	0 0 3 5 7 8 9
3	1 3 7 7
4	5
5	
6	
7	1

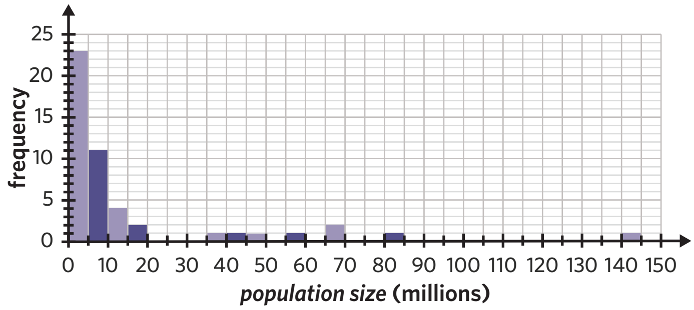
Back to contents

a. Describe both distributions in terms of shape, centre, spread and potential outliers.

b. Identify and explain the best measure of centre for each distribution.

Exam practice

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



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.

VCAA 2019 Exam 1 Data analysis Q2

86% of students answered this question correctly.

---

18. Each dalmatian in a sample of 32 dalmatians had their *weight* recorded. The data is displayed in the following ordered stem plot.

Key: 21 | 6 = 21.6 kg $n = 32$

21	6 9 9
22	1 2 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

a. Describe the shape of the distribution. (1 MARK)

b. Determine the median *weight* for this group of dalmatians. (1 MARK)

Adapted from VCAA 2020 Exam 2 Data analysis Q1

Part **a**: **86%** of students answered this type of question correctly.
Part **b**: **70%** of students answered this type of question correctly.

---

19. The *times* between successive nerve impulses, in milliseconds, were recorded.

The following table shows the mean and the five-number summary calculated using 800 recorded data values.

	time (milliseconds)
mean	220
minimum value	10
first quartile (Q_1)	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

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.

VCAA 2020 Exam 1 Data analysis Q3

59% of students answered this question correctly.

Back to contents

Questions from multiple lessons

Data analysis *Year 11 content*

20. 25 students in a Year 11 class recorded their *heights* and displayed them in the following stem plot.

Key: 16 | 3 = 163 cm

16	3 3 4
16	5 6 7 9 9
17	0 0 0 0 1 1 1 3 4
17	5 7 7 7 9
18	0 1 1

The modal *height* is

A. 17 cm

B. 18 cm

C. 170 cm

D. 171 cm

E. 177 cm

Adapted from VCAA 2016 Exam 1 Data analysis Q3

Data analysis *Year 11 content*

21. A study examined the relationship between *IQ* and *test score* (%) on a logic test.

The following least squares regression equation was derived.

$\textit{test score} = -25.9 + 0.93 \times IQ$

Which of the following conclusions drawn from the regression equation is true?

A. The correlation coefficient is 0.93.

B. A person's *test score*, as a percentage, can be determined by subtracting 25.9 from their *IQ*.

C. A person's *IQ* can be determined by subtracting 25.9 from their *test score*.

D. An increase of 1 point in *IQ* is associated with an increase of 0.93% in *test score*.

E. An increase of 1% in *test score* is associated with an increase of 0.93 points in *IQ*.

Adapted from VCAA 2018 Exam 1 Data analysis Q10

Recursion and financial modelling *Year 11 Content*

22. An electricity company, Edrolicity, charges interest for overdue bills. If a bill is not paid by the due date, interest is charged at a rate of 2.2% per month, compounding monthly.

a. Anna received an Edrolicity bill for $120. If she paid the bill one month after its due date, how much did she pay in total? (1 MARK)

b. Brody received a bill for $246, and failed to pay it by the due date. Write a recurrence relation in terms of T_0, T_{n+1} and T_n that shows the total amount of the bill n months after the due date. (2 MARKS)

c. Brody paid his bill five months after the due date. How much did Brody pay in interest only, correct to the nearest cent? (1 MARK)

Adapted from VCAA 2017 Exam 2 Recursion and financial modelling Q6

Back to contents

1G Introduction to standard deviation

STUDY DESIGN DOT POINT

- summary of the distributions of numerical variables; the five-number summary and boxplots (including the use of the lower fence ($Q_1 - 1.5 \times IQR$) and upper fence ($Q_3 + 1.5 \times IQR$) 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

1A 1B 1C 1D 1E 1F **1G** 1H 1I

KEY SKILLS

During this lesson, you will be:

- calculating the sample mean
- calculating the sample mean and standard deviation using technology.

KEY TERMS

- Population
- Sample
- Mean
- Standard deviation

The mean and standard deviation provide an important way for data analysts to investigate data sets. The mean gives a standard measure of centre, whilst the standard deviation is another way to measure how spread out a data set is. It provides information about how far away each data point is from the mean and sets the foundation for the standardisation of scores across a variety of topics.

Calculating the sample mean

When data is collected, it may be taken from a **population** (the entire group from which a conclusion can be drawn) or a **sample** (a smaller subset of the population). As data collected from a population is not always easily accessible, the scope of this course focuses on data collected from a sample. As such, when calculating the mean of a data set in this course, this is actually a reference to the mean of a sample (or sample mean), which is represented as $\bar{x}$.

The **mean** is a measure of centre that averages out all values into even groupings. It is calculated by adding all data values in a sample and then dividing the sum by the number of values. This can be expressed in the formula:

$\bar{x} = \frac{\Sigma x}{n}$, where Σx is the 'sum of all values', and n is the number of values in the data set.

Worked example 1

The amount of *money spent*, in dollars, of 10 customers at a supermarket is shown. Calculate the mean amount of *money spent*.

11.60 55.50 7.95 42.15 17.10 2.00 82.55 26.85 5.40 21.90

Explanation

Step 1: Calculate Σx and determine n.

$$\begin{aligned}\Sigma x &= \text{sum of all values}\\ &= 11.60 + 55.5 + 7.95 + 42.15 + 17.10 + 2.00\\ &\quad + 82.55 + 26.85 + 5.40 + 21.90\\ &= 273\end{aligned}$$

$$\begin{aligned}n &= \text{number of data values}\\ &= 10\end{aligned}$$

Step 2: Calculate the mean.

$$\begin{aligned}\bar{x} &= \frac{\Sigma x}{n}\\ &= \frac{273}{10}\\ \bar{x} &= 27.3\end{aligned}$$

Continues →

Back to contents

Answer

\$27.30

Calculating the sample mean and standard deviation using technology

Once the mean has been calculated, the standard deviation can be used to give an indication of the spread of a data set. The standard deviation referenced in this course is for a sample and is represented as s_x. The **standard deviation** is a measure of spread that is based on the average deviation (or difference) of each data point compared to the mean. The standard deviation of a sample can be calculated manually with the formula

$$s_x = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}},$$

where $\Sigma(x - \bar{x})^2$ is the sum of the squared differences between each data point and the mean, and n is the number of values in the data set; however, it is more commonly calculated using technology rather than by hand.

Worked example 2

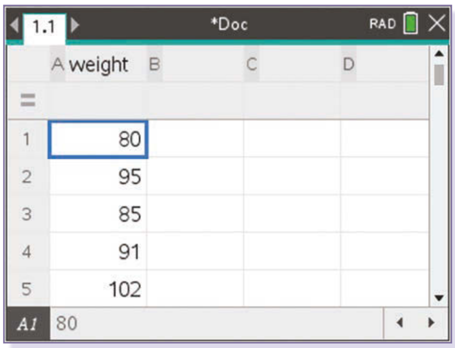
The *weight*, in kilograms, of a sample of 10 rugby players are shown. Determine the mean and standard deviation of the rugby players' *weight*, rounded to two decimal places.

80 95 85 91 102 93 87 78 84 90

Explanation – Method 1: TI-Nspire

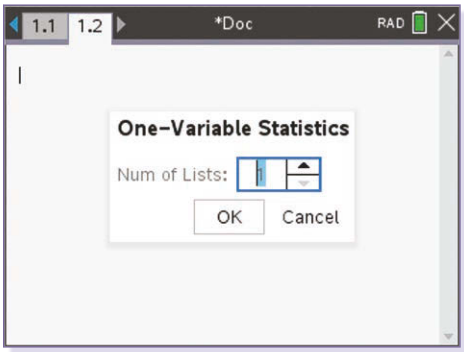
Step 1: From the home screen, select '1: New' → '4: Add Lists & Spreadsheet'.

Step 2: Name column A 'weight' and enter the data values starting from row 1 into the column below.

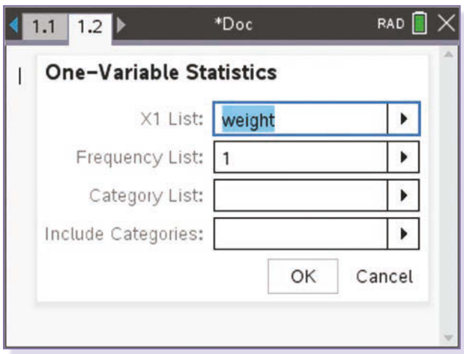


Step 3: Press ctrl + doc▾, and select '1: Add Calculator'.

Step 4: Press menu. Select '6: Statistics' → '1: Stat Calculations' → '1: One-Variable Statistics'. As there is only one data set, on the screen that follows, select 'OK'.

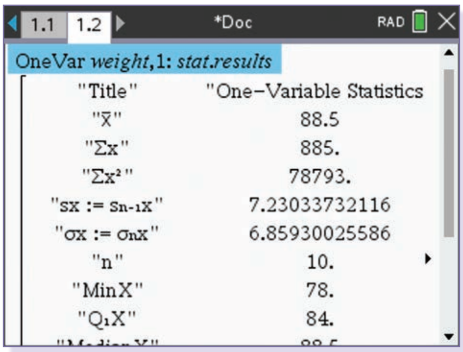


Step 5: On the next screen, select 'weight' as the 'X1 List' using the dropdown list, then select 'OK'.



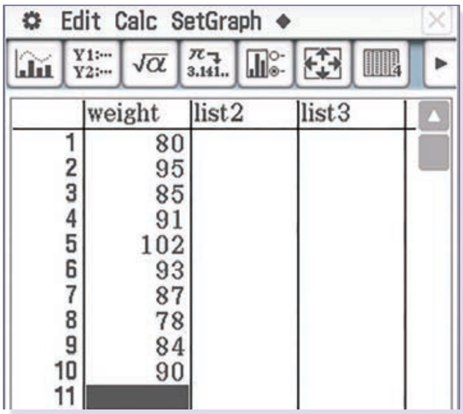
Continues →

Step 6: Identify the sample mean, $\bar{x}$, and standard deviation, s_x, (scrolling up may be required).

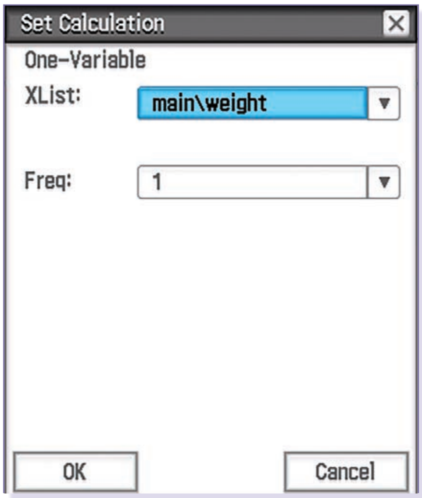


Explanation – Method 2: Casio ClassPad

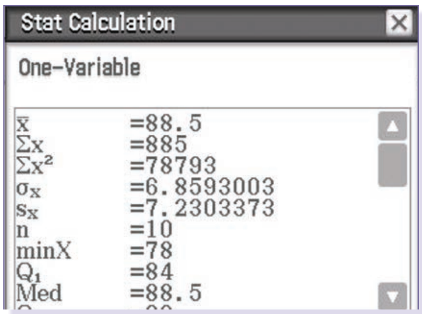
Step 1: From the main menu, tap Statistics. Name list1 'weight' and enter the data values starting from row 1 into the column below.



Step 2: Tap the 'Calc' menu at the top of the screen and select 'One-Variable'. On the screen that follows, select 'main\weight' as the 'XList' using the dropdown list, then tap 'OK'.



Step 3: Identify the sample mean, $\bar{x}$,and standard deviation, s_x.



Answer – Method 1 and 2

Mean: 88.50 kg

Standard deviation: 7.23 kg

Back to contents

Exam question breakdown

VCAA 2017 Exam 1 Data analysis Q3

The following 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. $\bar{x} = 1.58$ $s_x = 27.8$

B. $\bar{x} = 1.66$ $s_x = 27.8$

C. $\bar{x} = 27.8$ $s_x = 1.58$

D. $\bar{x} = 27.8$ $s_x = 1.66$

E. $\bar{x} = 27.8$ $s_x = 2.30$

Explanation – Method 1: TI-Nspire

Step 1: From the home screen, select '1: New' → '4: Add Lists & Spreadsheet'.

Step 2: Name column A 'circumf' and enter the data values starting from row 1 into the column below.

Step 3: Press ctrl + doc ▾, and select '1: Add Calculator'.

Step 4: Press menu. Select '6: Statistics' → '1: Stat Calculations' → '1: One-Variable Statistics'. As there is only one data set, on the screen that follows, select 'OK'.

Step 5: On the next screen, select 'circumf' as the 'X1 List' using the dropdown list, then select 'OK'.

Step 6: Identify the sample mean, $\bar{x}$, and standard deviation, s_x.

$\bar{x} = 27.8$ and $s_x = 1.66$

Explanation – Method 2: Casio ClassPad

Step 1: From the main menu, tap Statistics. Name list1 'circumf' and enter the data values starting from row 1 into the column below.

Step 2: Tap the 'Calc' menu at the top of the screen and select 'One-Variable'. On the screen that follows, select 'main\circumf' as the 'XList' using the dropdown list, then tap 'OK'.

Step 3: Identify the sample mean, $\bar{x}$, and standard deviation, s_x.

$\bar{x} = 27.8$ and $s_x = 1.66$

Answer - Method 1 and 2

D

88% of students answered this question correctly.

The most common incorrect responses were C, where students read the population standard deviation, σ_x, instead of the sample standard deviation, s_x, and B, where students read the mean and standard deviation in the wrong order.

1G Questions

Calculating the sample mean

1. A data set contains the following values.

35 21 59 43 56 30

The mean of the data set is closest to

A. 40.0 **B.** 40.6 **C.** 40.7 **D.** 41.0

2. Calculate the mean of the following data set.

141 159 130 164 150 147

3. Calculate the mean *number of pets* from the following data set, rounded to one decimal place.

number of pets	frequency
0	5
1	8
2	4
3	2
total	19

Calculating the sample mean and standard deviation using technology

4. For the following data set,

91 56 67 34 65 89 76 99 33 21 23 54 34 32 76 67 84

the sample mean and standard deviation are closest to

A. *mean* = 58.9 *standard deviation* = 24.6

B. *mean* = 58.9 *standard deviation* = 25.4

C. *mean* = 25.4 *standard deviation* = 58.9

D. *mean* = 58.8 *standard deviation* = 24.6

5. Calculate the sample mean and standard deviation for the following data set (to the nearest one decimal place).

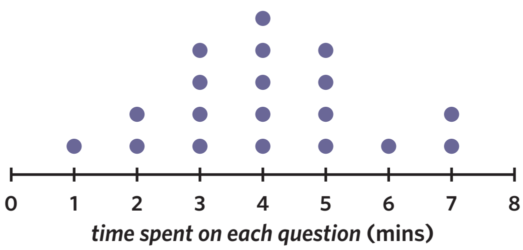
151.4 147.6 134.1 156.5 184.3 165.7 167.0 164.1 155.9 157.2

6. Calculate the sample mean and standard deviation for the following data set, rounded to one decimal place.

number of pens in a pencil case	0	1	2	3	4	5
frequency	3	10	5	4	1	1

Joining it all together

7. While studying for her geography exam, Layla recorded the *time spent on each question* in order to determine her speed. Her results are displayed in the dot plot, correct to the nearest minute.



a. Identify Σx and n for the data set.

b. Hence, calculate the mean amount of *time spent on each question*, rounded to the nearest minute.

c. Determine the standard deviation, rounded to two decimal places.

Back to contents

8. The following stem plot displays the *daily sales* at a crepe shop each day for two weeks.

Key: 3 | 1 = 31 sales

3	1
4	3
5	4 7
6	1 2 4
7	2 3 5 9
8	1 1 3

a. Identify Σx and n for the data set.

b. Hence, calculate the mean amount of *daily sales* at the crepe shop over the two weeks, rounded to the nearest whole number of sales.

c. Determine the standard deviation, rounded to two decimal places.

9. Ahava collects data from her colleagues on the number of *disposable coffee cups* used in a week. She is staggered to find that in total, her colleagues throw away 474 *disposable coffee cups* every week, with an average of 6 per person.

a. How many colleagues did Ahava ask?

b. Ahava wants to reduce the average number of disposable coffee cups that her colleagues use over the week. Although she would ideally like to eliminate their use altogether, she decides to start small and would like to lower the average coffee cups used by each person to 5 over the next week.

How many less coffee cups will need to be disposed of in the coming week to achieve this target?

Exam practice

10. 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.

Table 1

athlete number	1	2	3	4	5	6	7	8
high jump (metres)	1.76	1.79	1.83	1.82	1.87	1.73	1.68	1.82
shot-put (metres)	15.34	16.96	13.87	14.23	13.78	14.50	15.08	13.13
javelin (metres)	41.22	42.41	46.53	40.53	40.62	45.62	42.33	40.88

athlete number	9	10	11	12	13	14	15
high jump (metres)	1.83	1.87	1.87	1.80	1.83	1.87	1.78
shot-put (metres)	14.22	13.62	12.01	12.88	12.68	12.45	11.31
javelin (metres)	39.22	42.51	42.75	38.12	42.65	41.32	42.88

Complete Table 2 by calculating the mean height jumped for the *high jump*, in metres, by the 15 athletes. Write the answer in the space provided in the table. (1 MARK)

Table 2

statistic	*high jump* (metres)	*shot-put* (metres)
mean		13.74
standard deviation	0.06	1.43

VCAA 2021 Exam 2 Data analysis Q1b

87% of students answered this question correctly.

Back to contents

11. The *body density*, in kilograms per litre, and *weight*, in kilograms, of a sample of 12 orangutans are shown in the table.

***body density* (kg/litre)**	1.07	1.07	1.08	1.08	1.03	1.05	1.07	1.06	1.07	1.09	1.02	1.09
***weight* (kg)**	70.1	90.4	73.2	85.0	84.3	95.6	71.7	95.0	80.2	87.4	94.9	65.3

For these 12 orangutans, determine the mean of their *body density*, in kilograms per litre. (1 MARK)

Adapted from VCAA 2020 Exam 2 Data analysis Q4aii

77% of students answered this type of question correctly.

Questions from multiple lessons

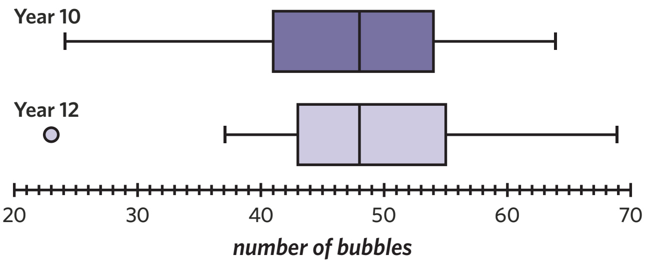
Data analysis *Year 11 content*

12. The parallel boxplots shown display the data for twenty Year 12 students competing against twenty Year 10 students to see who can blow the most bubbles.

The five-number summary of the *number of bubbles* for Year 12 students is equal to:

A. 23, 43, 48.5, 55, 69
B. 26, 41, 48, 54, 64
C. 23, 41, 48.5, 54, 69
D. 37, 43, 48, 54, 64
E. 37, 43, 48.5, 55, 69

Adapted from VCAA 2017 Exam 1 Data analysis Q2



Recursion and financial modelling *Year 11 content*

13. Wally decides that he needs to practise playing the piano. On the first day, he spends a total of two hours practising.

Each following day, he spends five minutes less than the previous day.

Let w_n be the number of minutes that Wally spends practising the piano on day n.

A recurrence relation that can be used to model this behaviour for $1 \le n \le 24$ is

A. $w_{n+1} = w_n + 5, \quad w_1 = 2$
B. $w_{n+1} = w_n \times 1.05, \quad w_1 = 120$
C. $w_{n+1} = w_n \times 0.95, \quad w_1 = 120$
D. $w_{n+1} = w_n - 5, \quad w_1 = 2$
E. $w_{n+1} = w_n - 5, \quad w_1 = 120$

Adapted from VCAA 2014 Exam 1 Number patterns Q4

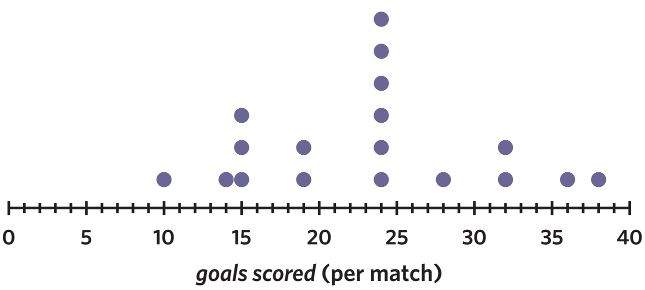
Data analysis *Year 11 content*

14. The following dot plot shows the distribution of *goals scored*, per match, throughout 18 matches in a local football league.

a. Write down the
  i. range (1 MARK)
  ii. median (1 MARK)

b. What is the number of goals scored at Q_3? (1 MARK)

Adapted from VCAA 2016 Exam 1 Data analysis Q1



Back to contents

1H The normal distribution

STUDY DESIGN DOT POINT

- 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

1A 1B 1C 1D 1E 1F 1G **1H** 1I

KEY SKILLS

During this lesson, you will be:

- calculating proportions in normal distributions
- calculating values in normal distributions
- determining the mean and standard deviation of a normal distribution.

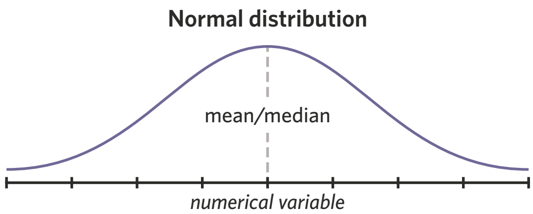
KEY TERMS

- Normal distribution

Data sets can take on a variety of different distributions and shapes. For data sets that exhibit a normal distribution, there are assumptions and generalisations that can be made that assist in making predictions about the data. These predictions make it possible to further analyse a data set by utilising unique properties of the mean and standard deviation of normal distributions.

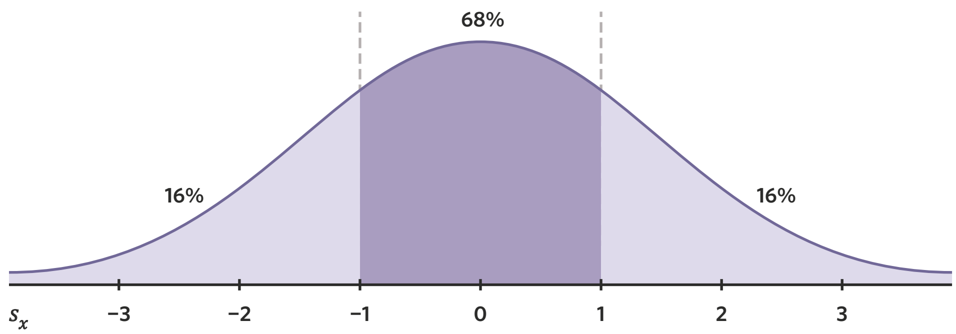
Calculating proportions in normal distributions

A **normal distribution** is a symmetrical (or approximately symmetrical) numerical data set that is centred around the mean, with a width determined by the standard deviation. Normal distributions are commonly known to be 'bell-shaped'. In a normal distribution, the mean and median are equal and are located through the central mirror line of the distribution. Some instances where data may be approximately normally distributed include study scores and IQ.

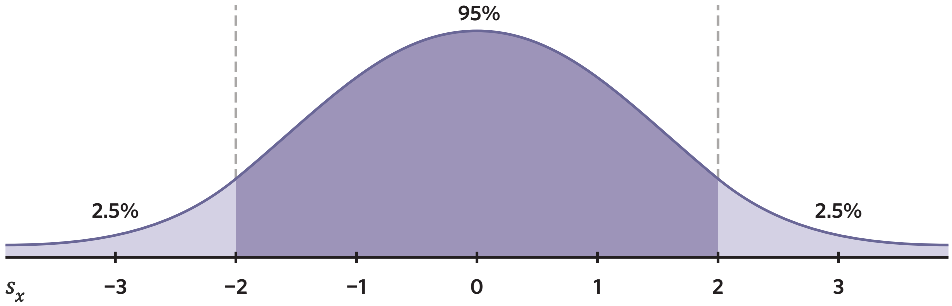


For a data set that is normally distributed, the data is spread around the mean with respect to the standard deviation. The 68–95–99.7% rule states that:

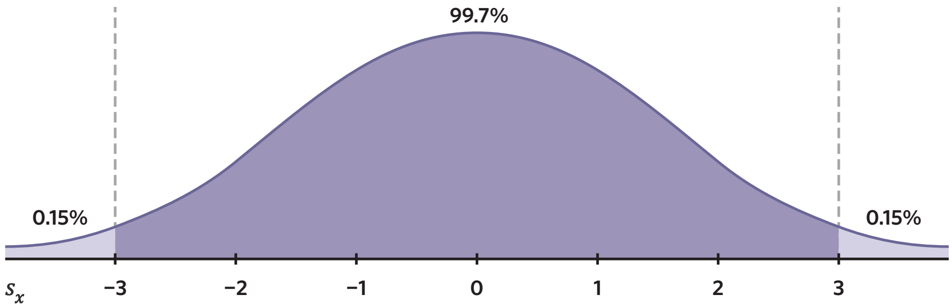
- 68% of data lies within one standard deviation (s_x) on either side of the mean ($\pm 1\ s_x$)



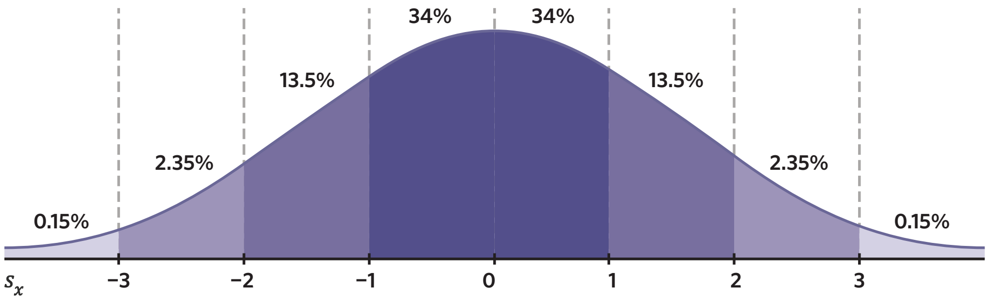
- 95% of data lies within two standard deviations on either side of the mean ($\pm 2\ s_x$)



- 99.7% of data lies within three standard deviations on either side of the mean ($\pm 3\ s_x$)



This allows a percentage breakdown of the data into groups based on the location of a data point with respect to the mean.



This concept can then be applied to calculate the proportion of data that lies between specified boundaries, as well as calculating the expected number within a sample that fulfil a specified criteria.

Worked example 1

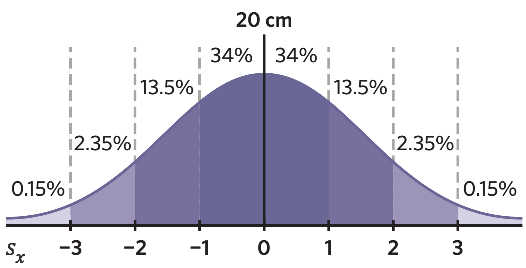
An entire girl's basketball club decided to shave their heads to raise money for cancer. They also donated their hair to be made into wigs for those undergoing treatment. The *length of hair donated* was approximately normally distributed with a mean of 20 cm and a standard deviation of 4 cm.

---

a. What percentage of hair lengths donated were between 16 cm and 32 cm?

Explanation

Step 1: Locate and label the mean on the normal distribution graph.



Continues →

Back to contents

Step 2: Calculate and label the values one, two and three standard deviations on either side of the mean.

1 standard deviation:

$20 - 1 \times 4 = 16$

$20 + 1 \times 4 = 24$

2 standard deviations:

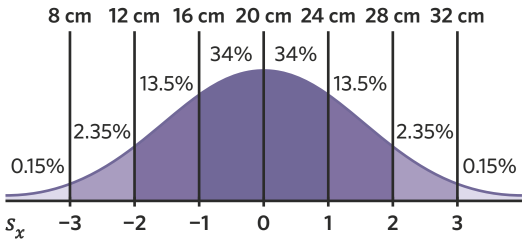
$20 - 2 \times 4 = 12$

$20 + 2 \times 4 = 28$

3 standard deviations:

$20 - 3 \times 4 = 8$

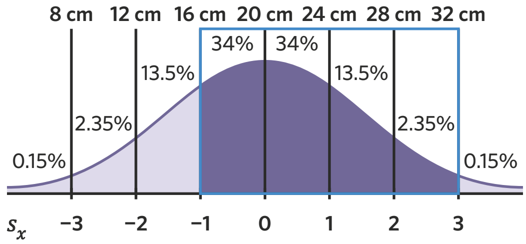
$20 + 3 \times 4 = 32$



Step 3: Identify the boundaries stated in the question.

The lower boundary is 16 cm.

The upper boundary is 32 cm.



Step 4: Calculate the sum of the percentages within the boundaries.

$34 + 34 + 13.5 + 2.35 = 83.85$

Answer

83.85%

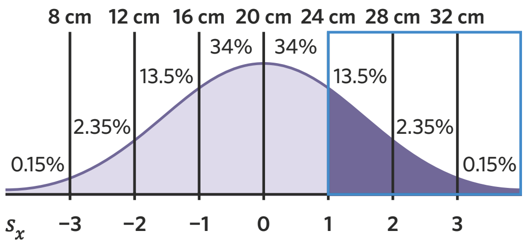
---

b. Donated hair lengths greater than 24 cm are highly sought after as they can be customised for a diverse range of hairstyles. What percentage of the girl's basketball club donated hair greater than 24 cm?

Explanation

Step 1: Determine the boundary, and the equivalent number of standard deviations.

A lower boundary of 24 cm is equivalent to 1 standard deviation greater than the mean.



Step 2: Calculate the percentage of data that fulfils the criteria.

$13.5 + 2.35 + 0.15 = 16$

Answer

16%

---

c. If there are 30 girls at the basketball club, how many of them are expected to have donated hair lengths longer than 24 cm, rounded to the nearest person?

Explanation

From part **b**, 16% of the hair lengths donated were longer than 24 cm.

Therefore, calculate 16% of 30 girls.

$0.16 \times 30 = 4.8$

Answer

5 girls

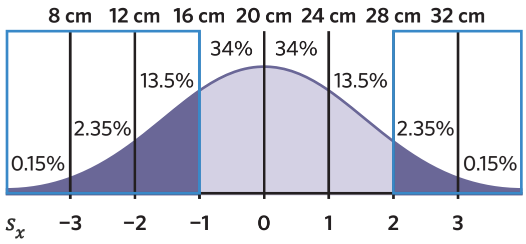
Continues →

d. How many girls are expected to have donated hair lengths that were either less than 16 cm or greater than 28 cm, rounded to the nearest person?

Explanation

Step 1: Using the labelled normal distribution graph, identify the boundaries stated in the question.

There are two separate boundary conditions that do not overlap.



Step 2: Calculate the sum of the percentages within the boundaries.

$0.15 + 2.35 + 13.5 + 2.35 + 0.15 = 18.5\%$

Step 3: Calculate the expected value.

18.5% of 30 girls.

$0.185 \times 30 = 5.55$

Answer

6 girls

Calculating values in normal distributions

Data sets that exhibit a normal distribution can also be used to determine the values within which a certain proportion of the data will lie.

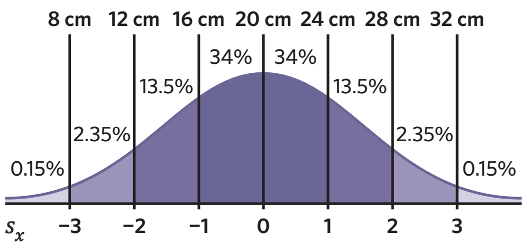
Worked example 2

An entire girl's basketball club decided to shave their heads to raise money for cancer. They also donated their hair to be made into wigs for those undergoing treatment. The *length of hair donated* was approximately normally distributed with a mean of 20 cm and a standard deviation of 4 cm.

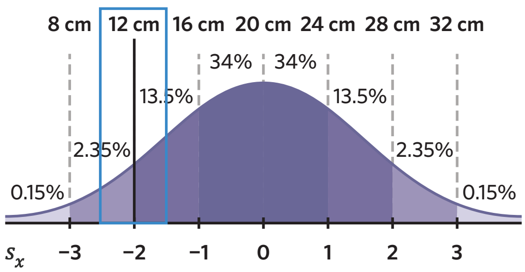
a. 97.35% of the hair lengths will fall within 12 cm and which other value?

Explanation

Step 1: Identify and label the mean, and the values that lie 1, 2, and 3 standard deviations on either side of the mean.



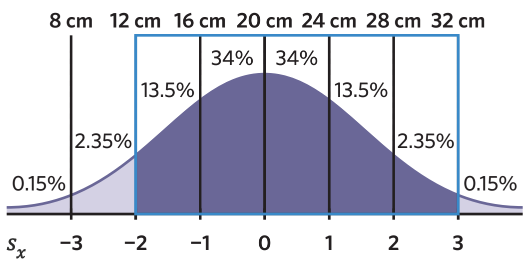
Step 2: Locate the boundary stated in the question.



Step 3: Determine the second boundary.

As 97.35% of the data does not lie below 12 cm, add percentages that are greater than 12 cm until 97.35% is obtained.

$13.5 + 34 + 34 + 13.5 + 2.35 = 97.35$



Continues →

Back to contents

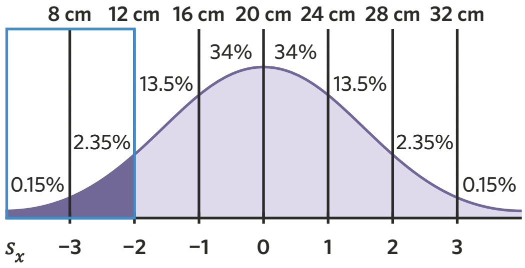
Answer

32 cm

b. 2.5% of hair donated by the team will be disposed of, as it falls below a required minimum length. What is this required minimum length?

Explanation

Using the labelled normal distribution graph, identify the appropriate boundary.



$0.15 + 2.35 = 2.5\%$

Therefore, 2.5% of hair lengths are below 12 cm.

Answer

12 cm

Determining the mean and standard deviation of a normal distribution

The 68–95–99.7% rule can also be used to calculate the mean and standard deviation of a data set if unknown.

Worked example 3

Duncan's light bulb factory produces light bulbs.

The *lifetime* of their light bulbs are approximately normally distributed. It is known that approximately 95% of the light bulbs have a *lifetime* ranging from 800 hours to 1600 hours.

a. Calculate the standard deviation for the *lifetime* of a light bulb.

Explanation

Step 1: Determine the number of standard deviations.

95% of data lies between two standard deviations either side of the mean. This is a total of four standard deviations.

Step 2: Calculate the value of one standard deviation.

Four standard deviations have a value of $1600 - 800 = 800$.

Therefore, one standard deviation is equal to $800 \div 4 = 200$.

Answer

200 hours

Continues →

Back to contents

b. Calculate the mean *lifetime* of a light bulb.

Explanation

The mean will lie two standard deviations greater than the lower boundary or two standard deviations less than the upper boundary.

$$800 + 2 \times 200 = 800 + 400$$
$$= 1200$$

Answer

1200 hours

Exam question breakdown

Adapted from VCAA 2021 Exam 2 Data analysis Q1d

In a qualifying competition, the heights jumped in the high jump are expected to be approximately normally distributed.

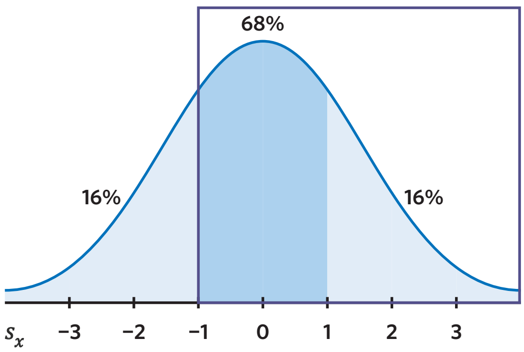
Chara's jump in this competition places her one standard deviation below the mean.

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)

Explanation

Step 1: Identify the boundaries using the normal distribution graph.

The question places Chara's jump one standard deviation below the mean (–1) and asks for athletes that have jumped higher.



Answer

84%

Step 2: Calculate the sum of the percentages within the boundaries.

$68 + 16 = 84$

60% of students answered this type of question correctly.

The most common incorrect answer was 16%. This is most likely caused by students finding the percentage of athletes who were expected to jump lower than Chara, rather than higher.

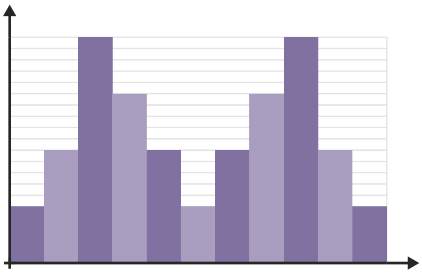
Back to contents

1H Questions

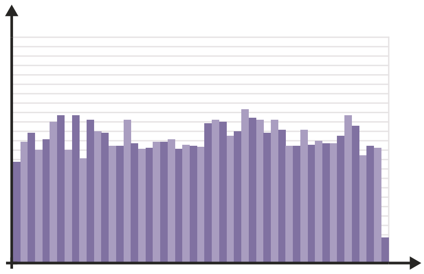
Calculating proportions in normal distributions

1. Which of the following is an example of a normal distribution?

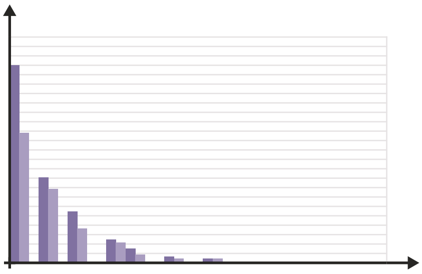
A.



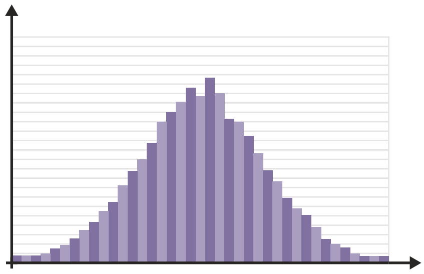
B.



C.



D.



2. 3000 students were surveyed about the *volume of milk* (mL) they add to their breakfast tea each morning. The results were approximately normally distributed with a mean of 30 mL and a standard deviation of 5 mL. What percentage of students add a *volume of milk* that is:

a. less than 25 mL?

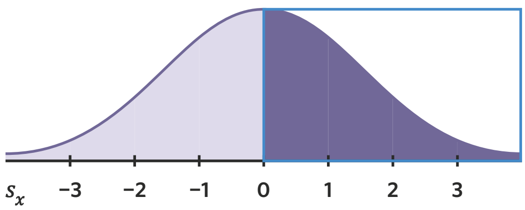
b. either less than 20 mL or greater than 40 mL?

3. Malia goes to a turkey farm before Christmas to have a look at their range of turkeys. She finds that the distribution of *weight* for the male turkeys is approximately normally distributed with a mean of 8 kilograms and a standard deviation of 1 kilogram. Malia has all of her extended family coming for Christmas lunch this year and will need a turkey that weighs more than 11 kilograms. If there are 40 000 turkeys available on the day that she visits the farm, how many turkeys will be suitable for her?

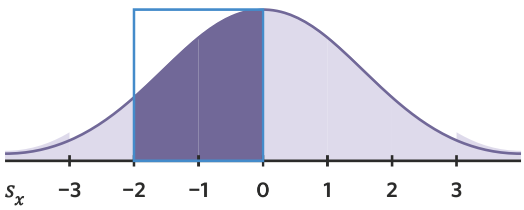
Calculating values in normal distributions

4. Which of the following graphs identifies 49.85% of data?

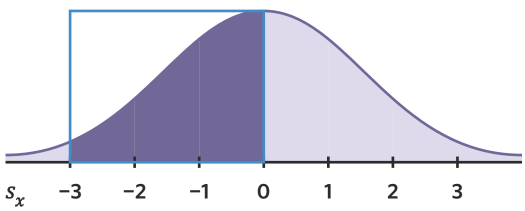
A.



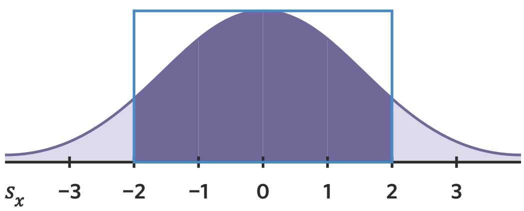
B.



C.



D.



5. The *weight*, in kilograms, of mountain zebras is approximately normally distributed with a mean of 280 kg and a standard deviation of 21 kg.

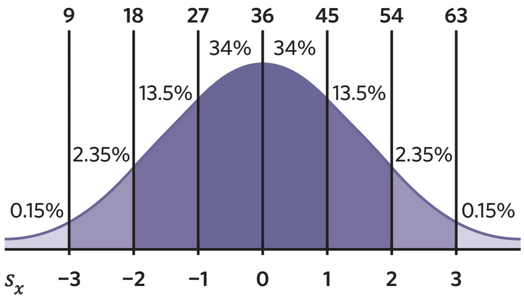
   a. Approximately 16% of mountain zebras will weigh more than what weight?

   b. Approximately 81.5% of mountain zebras will weigh between 259 kg and what other weight?

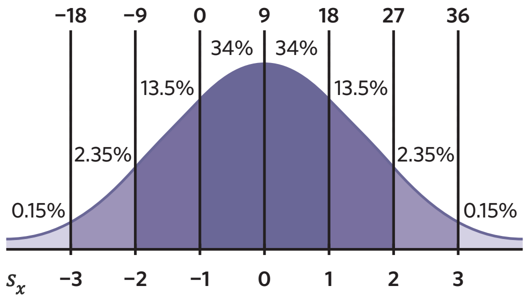
Determining the mean and standard deviation of a normal distribution

6. Which of the following graphs shows a normal distribution with a mean of 63 and a standard deviation of 9?

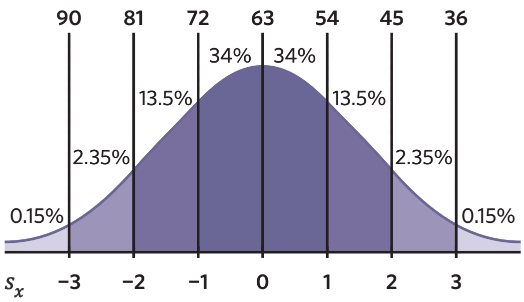
A.



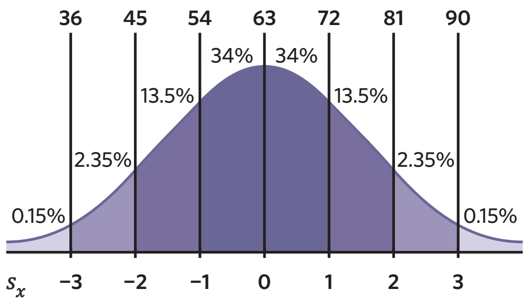
B.



C.



D.



7. A manufacturing company produces plastic bottles from recycled plastics that have been collected from waterways. The *density*, in g/cm^3, of each bottle is approximately normally distributed. The company claims that their manufacturing process is so consistent that 99.7% of the bottles produced have a *density* ranging between 0.92 and 1.16 g/cm^3.

   a. Calculate the standard deviation for the *density* of the plastic bottles produced at the manufacturing company.

   b. Calculate the mean *density* of the plastic bottles.

Joining it all together

8. The heights of 2000 seedlings planted by Michelle in her garden last year were measured and found to be approximately normally distributed. The mean of the heights was found to be 140 cm and the standard deviation was 12 cm.

   a. What percentage of Michelle's plants were between

      i. 128 cm and 140 cm?

      ii. 104 cm and 116 cm?

   b. Michelle is a diligent gardener and knows that she needs to harvest any plants that grow taller than 176 cm. Approximately how many plants will she need to harvest from her garden?

   c. The smallest 2.5% of seedlings will need extra care if they are to grow to their full potential. What height would the tallest of these seedlings be?

9. It is known that travel time from home to school for students at Schoolrolo Secondary College is approximately normally distributed with a mean of 45 minutes and a standard deviation of 10 minutes.

   a. Find the percentage of students who take less than 55 minutes to arrive at school.

Back to contents

b. Given that Schoolrolo Secondary College has 1800 students, find the approximate number of students who take more than 75 minutes to arrive at school. Round to the nearest number of students.

c. Find a time interval in which approximately 95% of the students will take to arrive at school.

---

10. The *mean time before failure* (*MTBF*) of an aeroplane engine is approximately normally distributed with a mean of 25 000 hours. Assume that one airline has a fleet of 200 aircrafts with each plane having two engines.

a. How many engines does the airline have in total?

Approximately 380 of these engines will have an *MTBF* between 24 000 and 26 000 hours.

b. Calculate the standard deviation.

c. Find the number of engines that have an *MTBF* more than 25 500 hours.

d. State a time interval in which approximately 68% of the engines will fail.

Exam practice

11. 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%

VCAA 2019 Exam 1 Data analysis Q6

79% of students answered this question correctly.

---

12. The *temperature difference* between the *minimum daily temperature* and the *maximum daily temperature* in November 2017 at a 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)

VCAA 2019 Exam 2 Data analysis Q2b

58% of students answered this question correctly.

---

13. 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.

Using the 68–95–99.7% rule, determine

a. the percentage of clusters expected to contain more than 140 eggs (1 MARK)

b. the number of clusters expected to have less than 215 eggs in a sample of 1000 clusters. (1 MARK)

VCAA 2017 Exam 2 Data analysis Q1b

Part **a**: **69%** of students answered this question correctly.
Part **b**: **47%** of students answered this question correctly.

Questions from multiple lessons

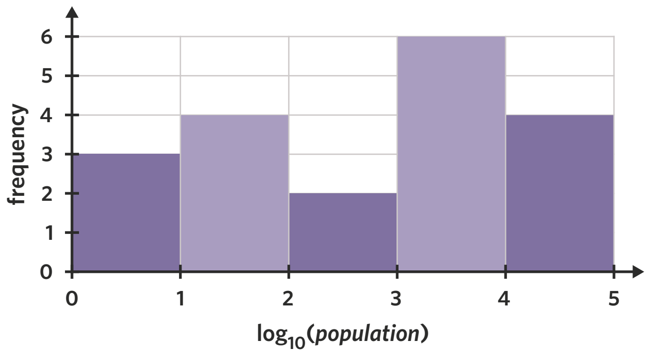
Data analysis

14. The following histogram displays the distribution of the $\log_{10}(population)$ of 19 planets in a faraway galaxy.

The median population is between

A. 10 and 100
B. 100 and 1000
C. 1000 and 10 000
D. 2 and 3
E. 3 and 4

Adapted from VCAA 2017 Exam 1 Data analysis Q4



Recursion and financial modelling *Year 11 content*

15. A dress shop is having a closing-down sale and is selling the 1800 dresses it has in stock.

On the first day of the sale, 100 dresses are sold.

On the second day, 125 dresses are sold.

On the third day, 150 dresses are sold.

This pattern continues until all 1800 dresses have been sold.

How long does it take to sell all of the dresses?

A. 8 days **B.** 9 days **C.** 18 days **D.** 19 days **E.** 67 days

Adapted from VCAA 2013 Exam 1 Number patterns Q6

Data analysis *Year 11 content*

16. A study was conducted investigating the relationship between *sleep*, measured in hours, and *blood pressure*, measured in mmHg. The following least squares regression equation was obtained.

$\textit{blood pressure} = 220 - 10.47 \times \textit{sleep}$

a. Which variable is the response variable? (1 MARK)

b. Interpret the slope of the regression equation in terms of the given variables. (2 MARKS)

Adapted from VCAA 2017NH Exam 2 Data analysis Q2bi,ii

Back to contents

1I z-scores

STUDY DESIGN DOT POINT

- 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

1A 1B 1C 1D 1E 1F 1G 1H 1I

KEY SKILLS

During this lesson, you will be:

- calculating standardised (z) scores
- calculating actual scores from standardised (z) scores
- using standardised (z) scores to interpret data.

KEY TERMS

- Standardised score (z-score)
- Actual score

Comparing several data sets that exhibit a normal distribution can be difficult when the number of data values in the samples differ, especially if by a large amount. Standardised scores (or z-scores) are an important tool that statisticians use to compare normal distributions. Standardised scores extend on the 68–95–99.7% rule and provide a more precise measure of the location of each data value within a sample.

Calculating standardised (z) scores

A **standardised score**, also known as a **z-score**, is a measure of the number of standard deviations between the mean and a data value. When performing standardised score calculations, each data value in a data set is referred to as an '**actual score**'.

Standardised scores can be:

- positive, indicating that the actual score is above the mean
- zero, indicating that the actual score is equal to the mean
- negative, indicating that the actual score is below the mean.

A standardised score is calculated by subtracting the mean from the actual score, and then dividing the result by the standard deviation.

$$z = \frac{x - \bar{x}}{s_x}$$

- z is the standardised score
- x is the actual score
- $\bar{x}$ is the mean
- s_x is the standard deviation

Worked example 1

The weights of puppies within a litter are approximately normally distributed with a mean of 4 kg and a standard deviation of 0.3 kg.

Determine the z-score of a puppy that weighs 4.6 kg.

Continues →

Back to contents

Explanation

Step 1: Identify the mean, standard deviation and actual score.

$\bar{x} = 4$

$s_x = 0.3$

$x = 4.6$

Step 2: Calculate the z-score.

Substitute the values into the formula $z = \frac{x - \bar{x}}{s_x}$.

$$z = \frac{4.6 - 4}{0.3} = \frac{0.6}{0.3} = 2$$

Answer

2

Calculating actual scores from standardised (z) scores

A standardised score can also be used to calculate the actual score using the formula:

$x = \bar{x} + (z \times s_x)$

Worked example 2

The number of black stripes on tigers at the Royal Melbourne Zoo is known to be normally distributed with a mean of 55 and a standard deviation of 2. One of the tigers has a standardised number of stripes of $z = -2.5$. How many stripes does the tiger actually have?

Explanation

Step 1: Identify the mean, standard deviation and standardised score.

$\bar{x} = 55$

$s_x = 2$

$z = -2.5$

Step 2: Calculate the actual score.

Substitute the values into the formula $x = \bar{x} + (z \times s_x)$.

$$x = 55 + (-2.5 \times 2) = 55 - 5 = 50$$

Answer

50 stripes

Using standardised (z) scores to interpret data

Comparing raw data values across multiple data sets may be misleading. Standardised scores help to look at data in a way which allows for comparison between different data sets.

For example, a test score of 70 might be considered low for an easy test in which everyone else did well on, but high for a more difficult test, which people struggled with.

The standardised score takes into account the mean and standard deviation of each data set and, when paired with the 68–95–99.7% rule, allows for a better comparison to be made.

Worked example 3

The results for a Chemistry exam were approximately normally distributed, with a mean of 71 and a standard deviation of 3. The results for a Physics exam were also approximately normally distributed, with a mean of 95 and a standard deviation of 6.

a. A student taking both Chemistry and Physics scored 77 in her final exam for both subjects. Did she do equally well in both exams in relation to her peers? Use standardised scores and the 68–95–99.7% rule to justify your answer.

Continues →

Back to contents

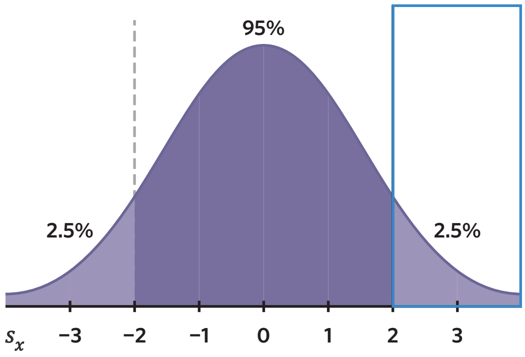
Explanation

Step 1: Calculate the standardised score for Chemistry.

$$z = \frac{77 - 71}{3}$$
$$= 2$$

Step 2: Interpret the standardised score for Chemistry using the 68–95–99.7% rule.

A z-score of 2 indicates that the student is two standard deviations above the mean Chemistry exam score. This means she is in the top 2.5% of the Chemistry class.

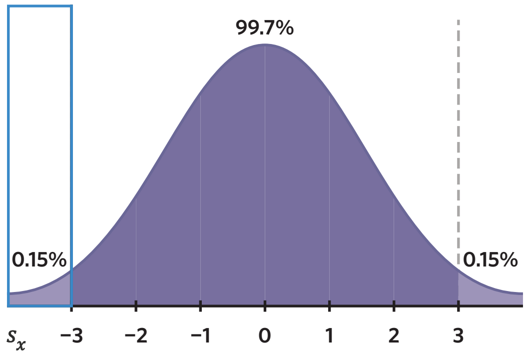


Step 3: Calculate the standardised score for Physics.

$$z = \frac{77 - 95}{6}$$
$$= -3$$

Step 4: Interpret the standardised score for Physics using the 68–95–99.7% rule.

A z-score of -3 indicates that the student is three standard deviations below the mean Physics exam score. This means she is in the bottom 0.15% of the Physics class.



Step 5: Write a worded answer.

Ensure the question is answered directly. Justify your answer using the information from the previous steps.

Answer

In Chemistry, the student's standardised score was $z = 2$ placing her in the top 2.5% of students.
In Physics, the student's standardised score was $z = -3$ placing her in the bottom 0.15% of students.
Therefore, she did not do equally well in both exams when compared to her peers.

---

b. Another student scores 66 on his Chemistry exam. Use standardised scores and the 68–95–99.7% rule to compare this student's score relative to the class.

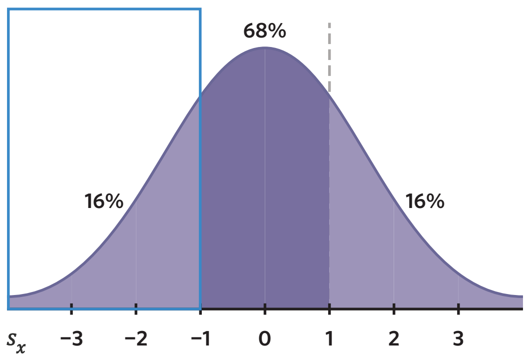
Explanation

Step 1: Calculate the standardised score for Chemistry.

$$z = \frac{66 - 71}{3}$$
$$= -1.666...$$

Step 2: Interpret the standardised score for Chemistry using the 68–95–99.7% rule.

A z-score of -1.67 indicates that the student is between one and two standard deviations below the mean Chemistry exam score. This means he is in the bottom 16% of the Chemistry class.



Continues →

Back to contents

Step 3: Write a worded answer.

Ensure the question is answered directly.
Justify your answer using the information from the previous steps.

Answer

In Chemistry, the student's standardised score was $z = -1.67$ placing him in the bottom 16% of students.

Exam question breakdown

VCAA 2016 Exam 1 Data analysis Q5

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.

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.

Explanation

To solve this question, check whether each option is true or false.

A: This is true. As the z-scores for both Brett and Sanjeeva are negative, this indicates that their weights are both below the mean weight. ✗

B: This is true. As Sanjeeva's z-score is higher than Brett's, this indicates that Sanjeeva weighs more than Brett. ✗

C: This is false.

Sanjeeva's actual weight:

$$x = \bar{x} + (z \times s_x)$$
$$= 78.6 + (-0.26 \times 9.3)$$
$$= 76.182$$

Add 2 kg to Sanjeeva's actual weight.

$76.182 + 2 = 78.182$

78.182 is less than the mean of 78.6. ✓

D: This is true.

Brett's actual weight:

$$x = \bar{x} + (z \times s_x)$$
$$= 78.6 + (-0.96 \times 9.3)$$
$$= 69.672$$ ✗

E: This is true. As Sanjeeva's z-score is negative, this indicates that her weight is below the mean weight, and therefore less than 50% of the competition's weight. ✗

Answer

C

58% of students answered this question correctly.

Students who did not answer this question correctly would have found the question difficult to navigate, and probably did not take a methodical approach. As the possible responses are quite word-heavy, students most likely were overwhelmed with the choices.

Back to contents

1I Questions

Calculating standardised (z) scores

1. A standardised score of $z = -2.4$ indicates an actual score is

A. less than 2 standard deviations greater than the mean.
B. more than 2 standard deviations greater than the mean.
C. equal to the mean.
D. more than 2 standard deviations less than the mean.

2. A school hockey team underwent a cardiovascular fitness test. The heart rates of the students (in beats per minute, bpm) were measured, and found to be approximately normally distributed, with a mean of 65 bpm and a standard deviation of 8 bpm.

Calculate the standardised heart rate, correct to two decimal places, for a student whose actual heart rate was

a. 81 bpm. **b.** 77 bpm. **c.** 63 bpm. **d.** 54 bpm.

3. A PE class played soccer every lesson for one term. Over this period, the number of goals kicked by each student was recorded and was found to be approximately normally distributed, with a mean of 13 and a standard deviation of 3.

Calculate a student's standardised score, correct to three significant figures, given that they kicked

a. 7 goals. **b.** 19 goals. **c.** 15 goals. **d.** 6 goals. **e.** 2 goals.

Calculating actual scores from standardised (z) scores

4. Which of the following actual scores will give a standardised score of zero in a data set with a mean of 52.0 and a standard deviation of 3.4?

A. 0.0 **B.** 3.4 **C.** 52.0 **D.** 62.2

5. The weights of a group of cats are approximately normally distributed with a mean weight of 3.5 kg and a standard deviation of 0.6 kg.

Correct to two decimal places, calculate the actual weight that corresponds to each of the following standardised weights.

a. 1.5 **b.** -0.8 **c.** -2.5

6. After a local survey, the annual wage of workers in a small town was found to be approximately normally distributed. The workers had a mean wage of \$62 377 with a standard deviation of \$7256. Calculate a worker's actual wage, correct to the nearest dollar, for the following standardised scores.

a. $z = 1.00$ **b.** $z = -1.50$ **c.** $z = 2.50$ **d.** $z = 1.74$ **e.** $z = -2.83$

Using standardised (z) scores to interpret data

7. Which of the following scenarios will most likely result in equal z-scores?

A. One student scoring 54% on two separate tests.
B. Two students scoring 54% on the same test in the same class.
C. Two students scoring 54% on the same test in different classes.
D. One student scoring 54% on one test and 45% on another test.

Back to contents

8. Matt's height is measured at 182.2 cm. In Australia, male heights are approximately normally distributed with a mean of 175.6 cm and a standard deviation of 6.6 cm. In the Netherlands, male heights are also approximately normally distributed with a mean of 183.8 cm and a standard deviation of 7.1 cm.

Use standardised scores and the 68–95–99.7% rule to compare Matt's height relative to males in both countries.

---

9. The men's 100 m sprint results are approximately normally distributed. Times continue to drop every year. The mean times and standard deviations, in seconds, in the years 2017, 2018 and 2019 are shown in the table.

year	mean	standard deviation
2017	10.43	0.09
2018	10.17	0.23
2019	9.91	0.10

Kamil has also seen a reduction in his times over the three years. In 2017, he recorded a best time of 10.55 seconds. In 2018, his personal best had dropped to 10.12 seconds. In 2019, it had dropped further to 9.97 seconds.

In which year did Kamil run fastest when compared against other athletes? Use standardised scores and the 68–95–99.7% rule to justify your answer.

Joining it all together

10. Suppose that the height of 18-year-old girls is normally distributed with a mean of 169.0 cm and a standard deviation of 2.6 cm.

a. Camilla is 18 years old and 164.0 cm tall. What is her standardised height, correct to two significant figures?

b. Jing is 18 years old and has a standardised height of $z = -0.65$. What is her actual height, correct to four significant figures?

c. Amara is 18 years old and has a standardised height of $z = 2$.

i. What is Amara's actual height?

ii. What percentage of 18 year old girls are shorter than Amara?

---

11. Two classes sat tests in both Chemistry and Physics. The tests were marked out of 120. Table 1 shows the mean and standard deviation of both classes for these tests. One value is missing.

Table 1

	class A		class B	
	$\bar{x}$	s_x	$\bar{x}$	s_x
Chemistry	79	5	83	
Physics	85	7	65	12

Isaac and Mohit are in class A, whilst Sarika is in class B. Table 2 shows their results.

Table 2

	Isaac	Mohit	Sarika
Chemistry	92	89	76
Physics	92	81	81

a. Use standardised scores to compare Isaac's performance, relative to his class, in both subjects.

b. Use standardised scores to compare Mohit and Sarika's performances in Physics, relative to their respective classes.

c. What percentage of class A did better than Mohit in Chemistry?

Back to contents

d. Sarika has a standardised score of $z = -1$ in Chemistry, relative to her class. What is the standard deviation of class B in Chemistry?

e. What percentage of class B did worse than Sarika in Chemistry?

f. The pass mark for the Chemistry test in class B was such that only 50% of the class passed the test. What percentage of the class did better than Sarika but still failed the test?

Exam practice

12. 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.

A student selected at random from this population has a standardised pulse rate of $z = -2.5$.

This student's 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.

VCAA 2018 Exam 1 Data analysis Q3

86% of students answered this question correctly.

13. 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. In shot-put, athletes throw a heavy spherical ball (a shot) as far as they can.

Fifteen female athletes competed to qualify for the heptathlon at the Olympic Games.

For shot-put, athlete number six, Jamilia, threw the shot 14.50 m.

The table shows the mean and standard deviation of the shot-put results for the 15 athletes.

statistic	*shot-put* (metres)
mean	13.74
standard deviation	1.43

Calculate Jamilia's standardised score (z).

Round your answer to one decimal place. (1 MARK)

VCAA 2021 Exam 2 Data analysis Q1c

78% of students answered this question correctly.

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

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)

VCAA 2017 Exam 2 Data analysis Q1c

67% of students answered this question correctly.

15. The temperatures over a number of days in a city have a mean of 3.5 °C and a standard deviation of 3 °C.

The temperature for one of the data points is found to be −4.3 °C.

The standardised value of this temperature is

A. −4.3

B. −2.6

C. 2.6

D. 3.5

E. 4.3

Adapted from VCAA 2017 Exam 1 Data analysis Q10

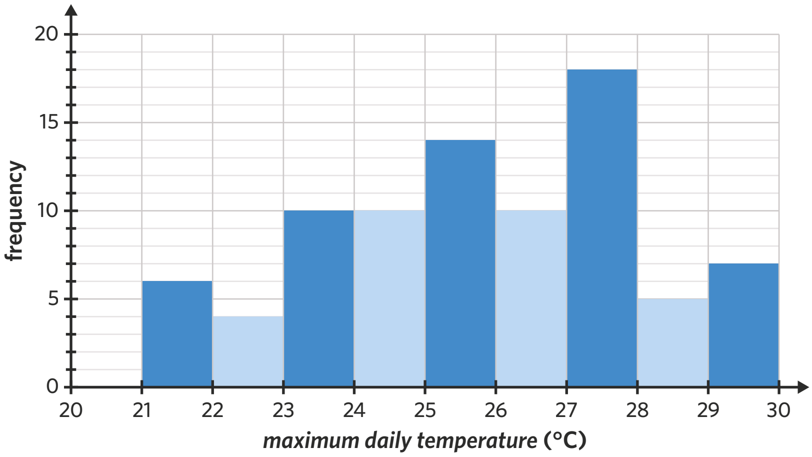
58% of students answered this type of question correctly.

Back to contents

Questions from multiple lessons

Data analysis *Year 11 content*

16. The following histogram displays the *maximum daily temperature* (°C) in a sample of 84 cities on a particular day.



The interquartile range for this distribution is closest to

A. 3 °C **B.** 4 °C **C.** 5 °C **D.** 7 °C **E.** 9 °C

Adapted from VCAA 2018NH Exam 1 Data analysis Q4

Recursion and financial modelling *Year 11 content*

17. Lorenzo bought an exotic pet python with a retail price of $6000.

However, he did not have enough money upfront. He paid a deposit of $2000 and repaid the rest of the balance with 36 monthly repayments of $200.

How much interest was Lorenzo charged?

A. $1200 **B.** $2200 **C.** $3200 **D.** $5200 **E.** $9200

Adapted from VCAA 2015 Exam 1 Business-related mathematics Q5

Data analysis *Year 11 content*

18. Students in a maths class recorded their heights and used the results to construct a five-number summary, as shown in the table. The distribution contains no outliers.

	minimum	Q_1	median	Q_3	maximum
height (cm)	161	168	173	178	182

a. Construct a boxplot displaying the students' heights. (1 MARK)

b. What percentage of students had a height of 178 cm or less? (1 MARK)

Adapted from VCAA 2016 Exam 2 Data analysis Q2ai,ii

Back to contents

2

CHAPTER 2

Investigating associations between two variables

LESSONS

KEY KNOWLEDGE

- 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.

Back to contents

2A Associations between two categorical variables

STUDY DESIGN DOT POINTS

- 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
- answering statistical questions that require a knowledge of the associations between pairs of variables

2A 2B 2C 2D

KEY SKILLS

During this lesson, you will be:

- displaying bivariate data using two-way frequency tables
- displaying bivariate data using grouped bar charts
- displaying bivariate data using percentage segmented bar charts
- describing the association between two categorical variables.

KEY TERMS

- Two-way frequency table
- Two-way percentage frequency table
- Grouped bar chart

Data displays such as two-way frequency tables, grouped bar charts, and percentage segmented bar charts, can help visually compare the distributions of two categorical variables. If the distributions differ between categories, this indicates that there may be an association between the two variables.

Displaying bivariate data using two-way frequency tables

A **two-way frequency table** is a data display used to summarise bivariate data. They allow two sets of categorical data to be compared. The columns of the two-way frequency table are defined by the explanatory variable, and the rows are defined by the response variable. The explanatory variable is used to explain or predict a change in the response variable.

A difference in sample size between the categories of the explanatory variable can cause misleading patterns. As a result, further calculations are needed to accurately compare the distributions between each data set.

Data can be recorded within a frequency table as either a frequency or percentage frequency. In a **two-way percentage frequency table**, the percentages represent the proportion of times each value or category occurs.

Worked example 1

A group of people were surveyed on their opinion towards offering a student discount for public transport. Their *studying status* was recorded as 'studying' or 'not studying' along with their *opinion* 'for' or 'against' the student discount.

Complete the percentage frequency table by filling out the missing columns with the percentage frequency of each category, rounded to two decimal places.

	studying status			
opinion	studying		not studying	
for	20		5	
against	16		15	
total	**36**		**20**	

Continues →

Back to contents

Explanation

Step 1: Calculate the percentage frequency for 'studying'.

The total percentage frequency should add up to 100%.

	studying status			
opinion	studying		not studying	
for	20	$\frac{20}{36} \times 100 \approx 55.56\%$	5	
against	16	$\frac{16}{36} \times 100 \approx 44.44\%$	15	
total	**36**	**100%**	**20**	

Step 2: Calculate the percentage frequency for 'not studying'.

The total percentage frequency should add up to 100%.

	studying status			
opinion	studying		not studying	
for	20	55.56%	5	$\frac{5}{20} \times 100 \approx 25.00\%$
against	16	44.44%	15	$\frac{15}{20} \times 100 \approx 75.00\%$
total	**36**	**100%**	**20**	**100%**

Answer

	studying status			
opinion	studying		not studying	
for	20	55.56%	5	25.00%
against	16	44.44%	15	75.00%
total	**36**	**100%**	**20**	**100%**

Displaying bivariate data using grouped bar charts

A **grouped bar chart** is a type of bar chart that visually displays two categorical variables. It displays the categories of one variable on the horizontal axis, while the categories of the other variable are represented with different bars for each category.

The frequency of each category is represented by the height of the columns. These graphs should follow the same rules as a regular bar chart. The only difference is that there will be multiple columns grouped together, based on the categories. Spaces are included between groups of columns to indicate separate categories.

Data sets for grouped bar charts are usually presented in a frequency table. The chart typically has frequency plotted on the vertical axis, and the explanatory variable on the horizontal axis. The response variable is then represented by each of the columns. Colours or patterns can be used to distinguish between one category and another, and a legend indicates which bars relate to which categories.

Worked example 2

A group of individuals between the ages of 20 and 59 were asked how much coffee they drink a day. Their *caffeine intake* was categorised as 'low', 'medium' and 'high'. Their *age* was also recorded and grouped as '20–29', '30–39', '40–49' and '50–59'. The results are displayed in a two-way frequency table.

Use the frequency table to construct a grouped bar chart.

	age			
caffeine intake	20–29	30–39	40–49	50–59
low	22	16	10	2
medium	5	10	14	17
high	3	4	6	11

Explanation

Step 1: Identify the explanatory variable.

Since *age* is more likely to dictate the level of *caffeine intake*, this would make it the explanatory variable. The variable *age* will be plotted on the horizontal axis.

Step 2: Construct a set of axes with 'frequency' on the vertical axis and *age* on the horizontal axis.

The frequency should at least extend to the maximum value, which is 22.

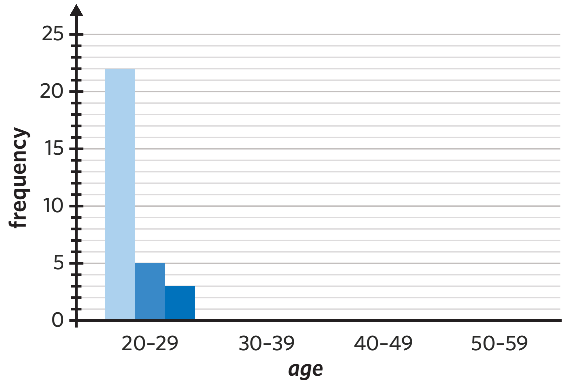
The horizontal axis should include labels for each of the *age* categories.

Continues →

Step 3: Plot the frequencies for the '20–29' *age* group.

Draw vertical columns for each *caffeine intake* category according to their frequency.

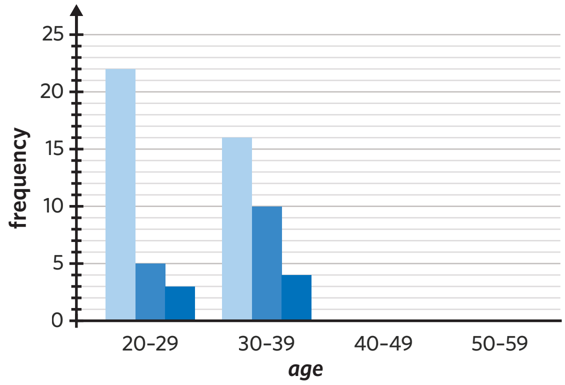
The categories will be graphed starting from 'low' to 'medium' to 'high'.



Step 4: Plot the frequencies for the '30–39' *age* group.

Draw vertical columns for each *caffeine intake* category according to their value in the frequency table.

The categories will be graphed starting from 'low' to 'medium' to 'high'.

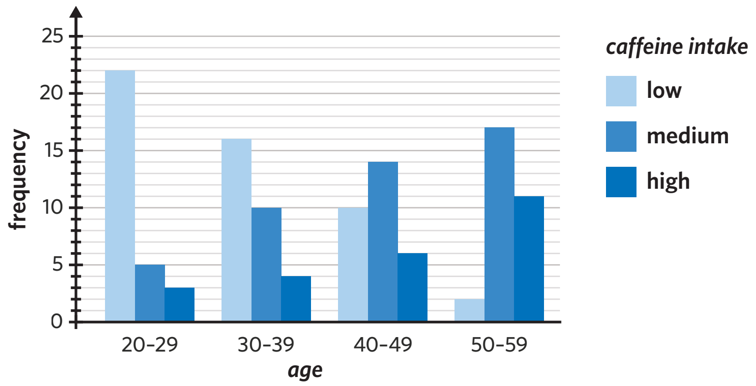


Step 5: Repeat this process for the remaining categories.

The remaining categories that need to be plotted are 40–49 and 50–59.

Step 6: Add a legend.

Answer



Displaying bivariate data using percentage segmented bar charts

Percentage segmented bar charts that display categorical variables follow the same rules as regular percentage segmented bar charts. The only difference is that there are multiple columns side by side. Previously, segmented bar charts were used to display one categorical variable, but they can be used to display data for two categorical variables by adding more columns.

There should be a gap between the columns and each should have a height of 100%. The percentage frequency is plotted on the vertical axis, and the explanatory variable on the horizontal axis. The categories of the response variable are then represented by each of the segments.

Back to contents

Worked example 3

A group of students in Year 11 and Year 12 were asked to describe their *height* as 'short', 'average' or 'tall'. Their responses, as well as their *year level*, were recorded in the following percentage frequency table.

Use the frequency table to construct a percentage segmented bar chart.

	year level	
height	Year 11	Year 12
short	15%	20%
average	20%	30%
tall	65%	50%
total	**100%**	**100%**

Explanation

Step 1: Identify the explanatory variable.

Since *year level* is more likely to dictate the *height*, this would make it the explanatory variable. The variable *year level* will be plotted on the horizontal axis.

Step 2: Construct a set of axes with 'frequency (%)' on the vertical axis and *year level* on the horizontal axis.

The vertical axis should extend to the total percentage frequency, 100%.

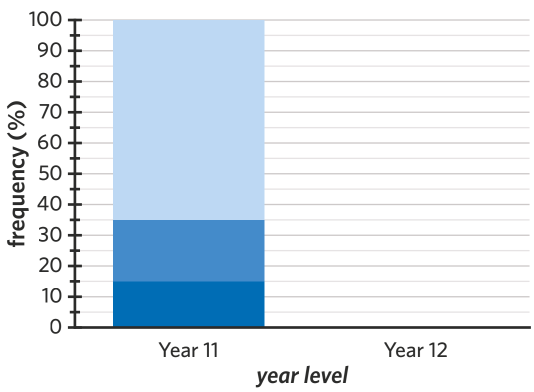
Step 3: Construct the column for 'Year 11' by adding the value of each segment.

The percentages will be graphed starting from 'short' to 'average' to 'tall'.

The 'short' segment should end at 15.

The 'average' segment should end at 15 + 20 = 35.

The 'tall' segment should end at 35 + 65 = 100.



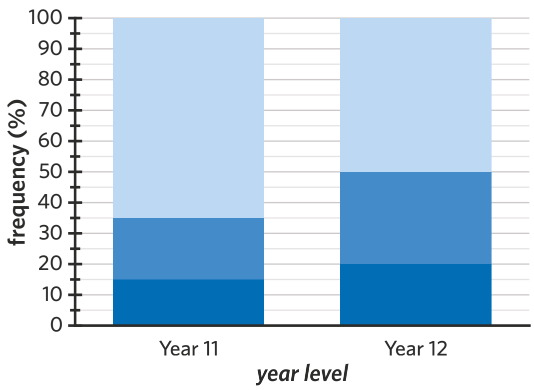
Step 4: Construct the column for 'Year 12' by adding the value of each segment.

The percentages will be graphed from 'short', to 'average', to 'tall'.

The 'short' segment should end at 20.

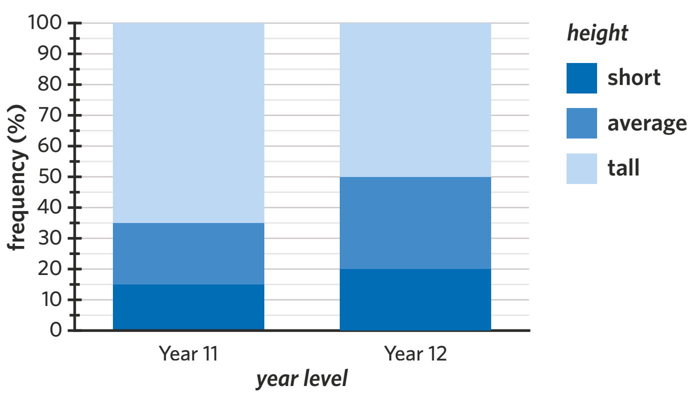
The 'average' segment should end at 20 + 30 = 50.

The 'tall' segment should end at 50 + 50 = 100.



Step 5: Add a legend.

Answer



Back to contents

Describing the association between two categorical variables

See worked example 4

When two categorical variables are compared, they can be analysed for any associations or patterns that may exist. A brief report describing an association, or lack of association, can then be written for a two-way percentage frequency table. The report should always include whether or not an association may exist, and appropriate percentages that support the finding.

See worked example 5

Associations between two sets of categorical data can also be analysed from percentage segmented bar charts.

Worked example 4

A group of people were surveyed on their *preference* towards reading hardcopy or digital books. Their *age* (under 20, 20 and over) was recorded along with their book *preference* (hardcopy, digital).

The results were displayed in a two-way percentage frequency table.

Is there an association between *age* and *preference*? Justify your answer by quoting appropriate percentages.

	age			
preference	under 20		20 and over	
hardcopy	20	56%	15	75%
digital	16	44%	5	25%
total	**36**	**100%**	**20**	**100%**

Explanation

Step 1: Consider whether there is a large percentage difference in *preference* for the different *age* categories.

- 56% of people under the age of 20 prefer hardcopy books.
- 75% of people aged 20 and over prefer hardcopy books.

There is a 19% difference between each group, which is a significant difference.

Step 2: Determine if there is an association.

An association can be determined by comparing one of the *preference* categories for both *age* categories.

Answer

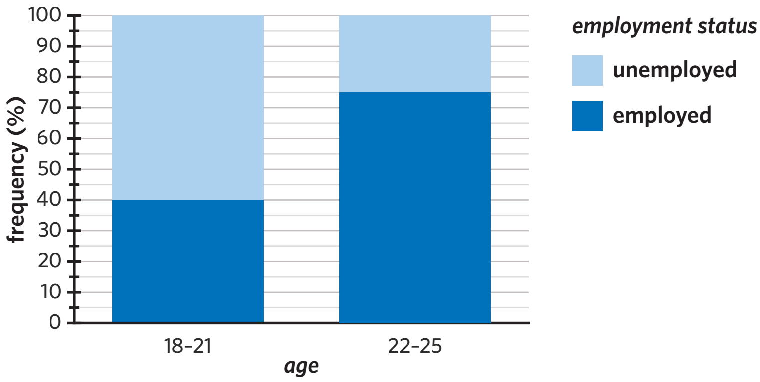
Yes, there is an association between *preference* and *age*. The percentage of people that prefer hardcopy books differs depending on their age, since only 56% of those under 20 prefer it while 75% of those aged 20 years and over do.

Answers may vary.

Worked example 5

A large group of people between the ages of 18 and 25 were asked what their current *employment status* was. They were classified as either ‘employed’ or ‘unemployed’. Their *age* was also recorded and classified as ‘18–21’ or ‘22–25’.

The results have been displayed in a percentage segmented bar chart.



Is there an association between *employment status* and *age*? Justify your answer by quoting appropriate percentages.

Continues →

Back to contents

Explanation

Step 1: Consider whether *employment status* is significantly different between the categories '18–21' and '22–25'.

- 40% of people between the ages of 18–21 are employed.
- 75% of people between the ages of 22–25 are employed.

There is a 35% difference between each age group. This is quite a large gap.

- 60% of people between the ages of 18–21 are unemployed.
- 25% of people between the ages of 22–25 are unemployed.

There is a 35% difference between each age group. This is quite a large gap.

Step 2: Determine if there is an association.

An association can be determined by comparing one of the two categories for *employment status*.

Answer

Yes, there is an association between *employment status* and *age*. The percentage of people that are employed tends to increase with age, with 40% of 18–21 year olds employed and 75% of 22–25 year olds employed.

Answers may vary.

Exam question breakdown

VCAA 2016 Exam 1 Data analysis Q1

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.

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%

	age	
blood pressure	under 50 years	50 years or over
low	15	5
normal	32	24
high	11	23
total	**58**	**52**

Explanation

Step 1: Determine the number of adults under 50 years of age with high blood pressure.

There are 11 adults aged under 50 with high blood pressure.

Step 2: Represent this as a percentage frequency.

$$\textit{percentage frequency} = \frac{\textit{frequency}}{\textit{total frequency}} \times 100$$

$$= \frac{11}{58} \times 100$$

$$= 18.96...\%$$

Answer

B

83% of students answered this question correctly.

12% of students incorrectly answered option A. These students likely did not convert the frequency to a percentage frequency.

Back to contents

2A Questions

Displaying bivariate data using two-way frequency tables

1. A group of people were surveyed on *commute time* from home to school.

Their responses were classified as 'less than an hour' or 'an hour or more'. They were also grouped by the *location* of their home as either 'urban' or 'rural'. The results were displayed in the two-way percentage frequency table.

	location	
commute time	urban	rural
less than an hour	97%	62%
an hour or more	3%	38%
total	**100%**	**100%**

a. What percentage of respondents living in an 'urban' *location* are 'less than an hour' away from their school?

A. 3%
B. 38%
C. 62%
D. 97%

b. What percentage of respondents living in a 'rural' *location* are 'less than an hour' away from their school?

A. 3%
B. 38%
C. 62%
D. 97%

2. A survey was conducted on people's preferences for *reusable cup* brands (frank green, KeepCup, other) and their *coffee consumption* (often, rarely). The results were displayed in a two-way percentage frequency table, but some values are missing.

a. Fill in the table with the missing values.

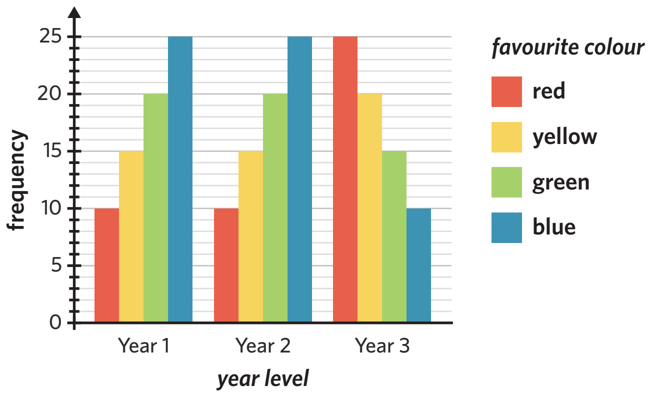
b. What percentage of people who often drink coffee prefer to use a KeepCup?

	coffee consumption	
reusable cup	often	rarely
frank green	41%	17%
KeepCup		11%
other	19%	
total	**100%**	**100%**

Displaying bivariate data using grouped bar charts

3. A study was conducted to test the relationship between a student's *year level* (Year 1, Year 2, Year 3) and their *favourite colour* (red, yellow, green, blue). The results were plotted on a grouped bar chart.

Determine the frequency table that matches the grouped bar chart.



A.

	year level		
favourite colour	Year 1	Year 2	Year 3
red	10	10	20
yellow	10	15	20
green	10	20	10
blue	10	25	20

B.

	year level		
favourite colour	Year 1	Year 2	Year 3
red	10	25	25
yellow	15	20	20
green	20	15	15
blue	25	10	10

Back to contents

C.

	year level		
favourite colour	Year 1	Year 2	Year 3
red	25	10	25
yellow	20	15	20
green	15	20	15
blue	10	25	10

D.

	year level		
favourite colour	Year 1	Year 2	Year 3
red	10	10	25
yellow	15	15	20
green	20	20	15
blue	25	25	10

4. A survey was conducted for a group of students to see if their current *education level* (primary, secondary, tertiary) had a relationship with their *preferred sport* (soccer, football, volleyball). The results are displayed in a frequency table.

	education level		
preferred sport	primary	secondary	tertiary
soccer	5	7	9
football	8	4	1
volleyball	2	4	5
total	**15**	**15**	**15**

Display the data using a grouped bar chart.

Displaying bivariate data using percentage segmented bar charts

5. A scientific experiment was conducted to determine if there is a relationship between *light intensity* (dark, bright) and *respiration rate* (low, medium, high) for a small sample of spinach leaves.

The results were displayed in a percentage segmented bar chart.

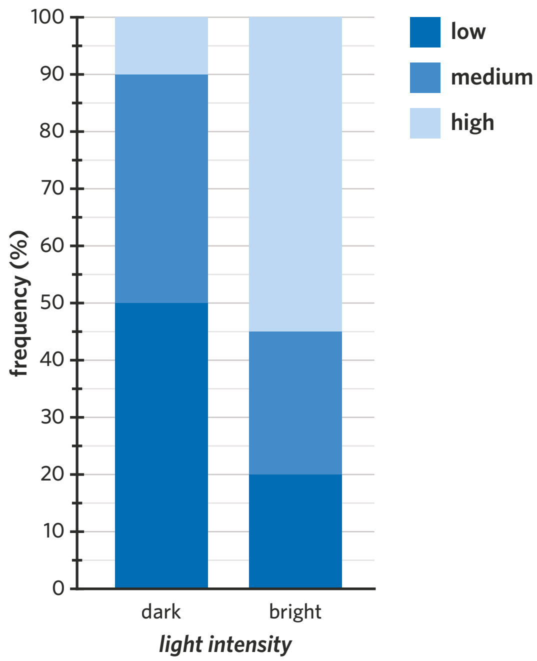
Which of the following statements is correct?

A. 25% of spinach leaves subjected to a bright light intensity had a low respiration rate.

B. 55% of spinach leaves subjected to a dark light intensity had a low respiration rate.

C. 35% of spinach leaves subjected to a dark light intensity had a medium respiration rate.

D. 55% of spinach leaves subjected to a bright light intensity had a high respiration rate.



6. A chef is experimenting with *cooking time* (short, medium, extended) for various pasta types to see if there is a relationship with its *doneness* (undercooked, al dente, overcooked).

The results were recorded in the percentage frequency table. Represent this information using a percentage segmented bar chart.

	cooking time		
doneness	short	medium	extended
undercooked	80%	20%	5%
al dente	15%	70%	20%
overcooked	5%	10%	75%
total	**100%**	**100%**	**100%**

Back to contents

Describing the association between two categorical variables

7. Each morning at the Edrolo office, a team member volunteers to do a coffee run. Everyone is able to choose either 'cappuccino', 'flat white' or 'latte'. Their *coffee preference* has been recorded alongside their *employment duration* (less than three years, three years or more) in a two-way percentage frequency table.

	employment duration	
coffee preference	less than three years	three years or more
cappuccino	13.1%	11.7%
flat white	23.7%	23.6%
latte	63.2%	64.7%
total	**100%**	**100%**

The data displayed supports the contention that there is no obvious association between *coffee preference* and *employment duration* because

A. the coffee preferences between employment duration differ by a significant amount.

B. 23.7% of team members who have been working for less than 3 years prefer a flat white.

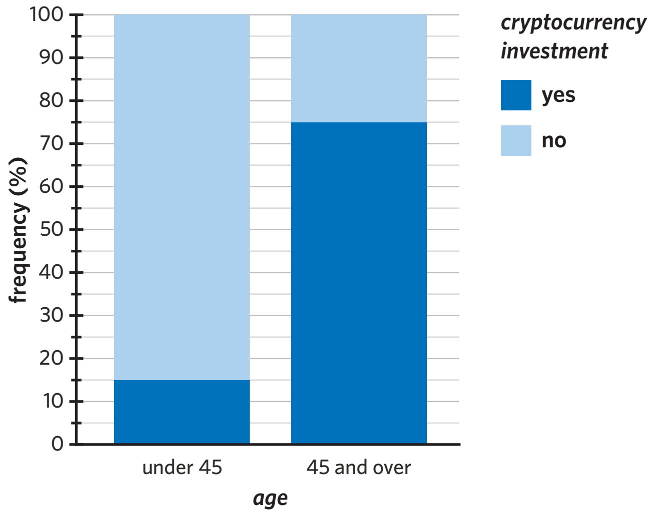
C. 63.2% of team members who have been working for less than 3 years prefer a latte and 64.7% of those who have been working for 3 or more years prefer a latte.

D. 63.2% of team members who have been working for less than 3 years prefer a latte and 11.7% of those who have been working for 3 or more years prefer a cappuccino.

8. A study was conducted to see if individuals had a *cryptocurrency investment* (yes, no). Participants also had their *age* (under 45, 45 and over) recorded.

The results have been displayed in a percentage segmented bar chart.

Is there an association between *age* and *cryptocurrency investment*? Ensure to quote appropriate percentages.



Joining it all together

9. Mark wants to see if there is an association between the number of *hours studied* (0 to 9 hours, 10 to 19 hours, 20 to 29 hours, 30 or more hours) for an assessment and the *test score* (low, medium, high) that his students get.

a. Which variable is most likely to be plotted on the horizontal axis of a percentage segmented bar chart?

b. Mark fills out a table with the data he gathered from the test. However, Mark spilled his coffee on his paper on his drive to work, leaving a couple of stains.
Fill out the rest of the table.

	hours studied			
test score	0 to 9 hours	10 to 19 hours	20 to 29 hours	30 or more hours
low		30%	20%	0%
medium	20%			10%
high	0%	10%	30%	90%
total	**100%**	**100%**	**100%**	**100%**

c. Mark's colleague Julie has also filled out a two-way percentage frequency table for her General Mathematics class. Construct a percentage segmented bar chart from the following information.

Back to contents

	hours studied			
test score	0 to 9 hours	10 to 19 hours	20 to 29 hours	30 or more hours
low	50%	40%	40%	0%
medium	50%	50%	30%	10%
high	0%	10%	30%	90%
total	**100%**	**100%**	**100%**	**100%**

d. Is there an association between the number of *hours studied* and the *test score* in Julie's class? Justify your answer by quoting appropriate percentages.

10. An astronomical observation was conducted to determine if there is a relationship between a star's *size* (small, large) and its *colour* (yellow, blue). The results are as follows:

- 12% of small stars observed were blue, with the rest being yellow.
- 15% of large stars observed were yellow, with the rest being blue.

a. Construct a percentage segmented bar chart to represent this information.

b. Is there an association between a star's *size* and its *colour*? Justify your answer by quoting appropriate percentages.

Exam practice

11. The following data relates to the impact of traffic congestion in 2016 on travel times in 23 cities in the United Kingdom (UK).

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

Data: TomTom International BV, <www.tomtom.com/en_gb/trafficindex>

Back to contents

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).

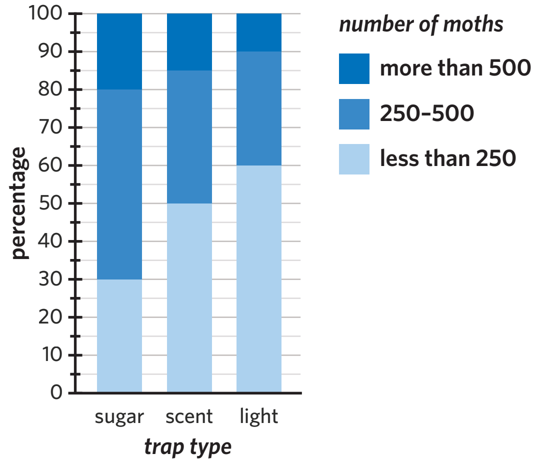
Use the data to complete the following two-way frequency table. (2 MARKS)

	size	
congestion level	small	large
high	4	
medium		
low		
total	**16**	

VCAA 2018 Exam 2 Data analysis Q1d

The average mark on this question was **1.9**.

12. 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.



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.

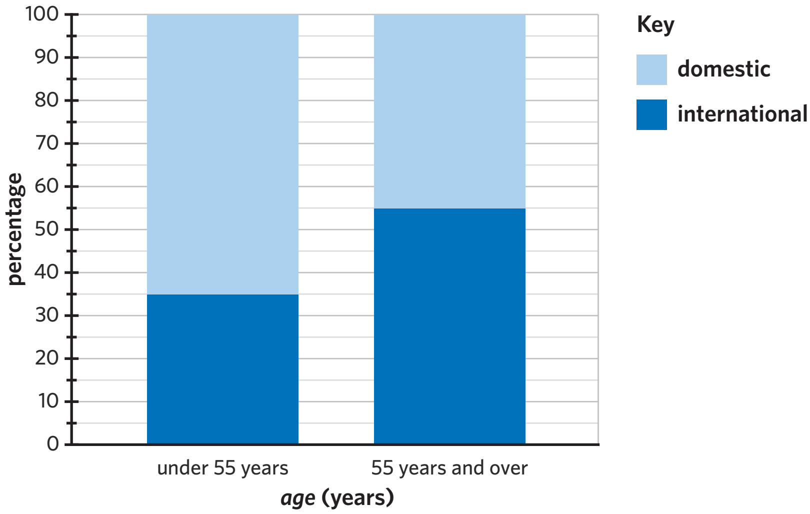
D. 20% of sugar traps contained more than 500 moths while 50% of light traps contained less than 250 moths.

E. 20% of sugar traps contained more than 500 moths while 10% of light traps contained more than 500 moths.

VCAA 2017 Exam 1 Data analysis Q6

71% of students answered this question correctly.

13. The following percentage segmented bar chart shows the *age* (under 55 years, 55 years and over) of visitors at a travel convention, segmented by *preferred travel destination* (domestic, international).



The results could also be summarised in a two-way frequency table.

Which of the following frequency tables could match the percentage segmented bar chart?

A.

	age	
preferred travel destination	under 55 years	55 years and over
domestic	91	90
international	49	110
total	**140**	**200**

B.

	age	
preferred travel destination	under 55 years	55 years and over
domestic	65	35
international	45	55
total	**110**	**90**

C.

	age	
preferred travel destination	under 55 years	55 years and over
domestic	35	55
international	65	45
total	**100**	**100**

D.

	age	
preferred travel destination	under 55 years	55 years and over
domestic	50	70
international	100	50
total	**150**	**120**

E.

	age	
preferred travel destination	under 55 years	55 years and over
domestic	71	39
international	29	61
total	**100**	**100**

VCAA 2021 Exam 1 Data analysis Q3

50% of students answered this question correctly.

Questions from multiple lessons

Data analysis

14. The following data represents the height in centimetres of eighteen year 7 students.

144 162 131 156 165 171 149 148 159

182 171 165 167 166 158 151 158 132

What is the mean, $\bar{x}$, and the standard deviation, s_x, of the heights, in centimetres, of the year 7 class?

A. $\bar{x} = 157.5, \quad s_x = 12.92$

B. $\bar{x} = 156.5, \quad s_x = 12.92$

C. $\bar{x} = 157.5, \quad s_x = 13.29$

D. $\bar{x} = 156.5, \quad s_x = 13.29$

E. $\bar{x} = 158.5, \quad s_x = 12.92$

Adapted from VCAA 2018NH Exam 1 Data analysis Q7

Data analysis

15. Data was collected to investigate the association between the variables *height* (cm) and *weight* (kg).

Which one of the following is most appropriate to display this data?

A. Back-to-back stem plot

B. Bar chart

C. Parallel boxplots

D. The coefficient of determination

E. Scatterplot

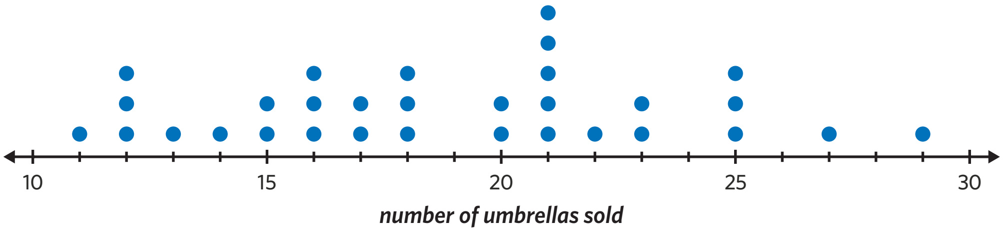
Adapted from VCAA 2018 Exam 1 Data analysis Q6

Back to contents

Data analysis *Year 11 content*

16. The owner of an umbrella shop records the number of umbrellas he sells each day.

His sales for the month of January are shown in the following dot plot.



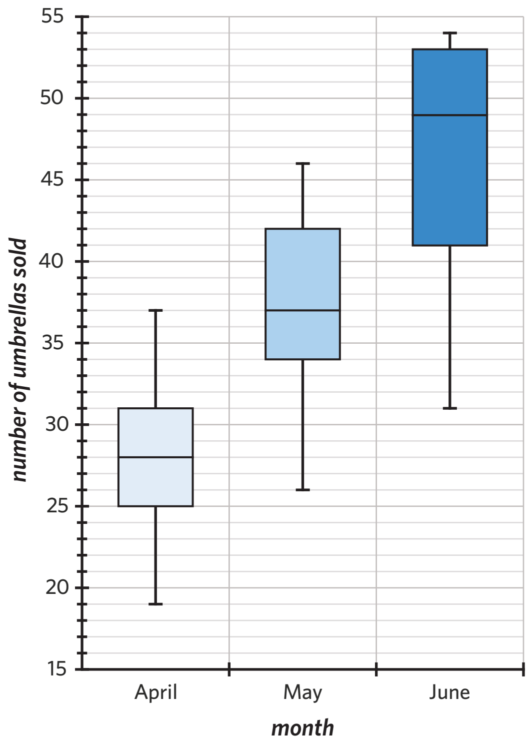
a. Construct a boxplot based on the data in the dot plot. (2 MARKS)

b. The number of umbrellas sold at the shop in April, May, and June are shown in the following parallel boxplots.

The parallel boxplots indicate that the number of umbrellas sold could be associated with the month of the year.

Explain, stating values of an appropriate statistic, why this conclusion could be made. (2 MARKS)

Adapted from VCAA 2017 Exam 2 Data analysis Q1c,d



Back to contents

2B Associations between numerical and categorical variables

STUDY DESIGN DOT POINTS

- 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
- answering statistical questions that require a knowledge of the associations between pairs of variables

2A 2B 2C 2D

KEY SKILLS

During this lesson, you will be:

- displaying data using back-to-back stem plots
- displaying data using parallel dot plots
- displaying data using parallel boxplots
- describing the association between numerical and categorical variables.

KEY TERMS

- Back-to-back stem plot
- Parallel dot plots
- Parallel boxplots

Data displays, such as back-to-back stem plots, parallel dot plots, and parallel boxplots help visually compare the distributions of two or more categories. If the distributions differ, this may signal that there is an association between the numerical and categorical variables.

Displaying data using back-to-back stem plots

A **back-to-back stem plot** is a stem plot that displays and compares the distribution of two categories.

The categories share the same stem, with the leaves of one category on the left and the leaves of the other category on the right. The leaves are ordered such that they get larger as they move away from the stem.

Back-to-back stem plots can be used to identify an association between two categories.

The following back-to-back stem plot shows the *number of points scored* in each game by two U18 basketball teams, the 'Crocodiles' and the 'Zebras'.

Key: 3 | 7 = 37 points

Crocodiles		Zebras
4 0	2	
8 7 5 2	3	7
7 4 4	4	
8 2	5	2 8
3 1	6	0 1 5 9
	7	0 3 5 8
0	8	1 3 4

Back to contents

Worked example 1

The 'Abbotsford Avengers' soccer team beat the 'East Melbourne Eagles' in the final. In order for the 'East Melbourne Eagles' to see what they needed to improve on, they collected the *successful passes* (%) for each member of both teams.

Abbotsford Avengers: 64 68 70 53 57 62 74 77 36 41 66

East Melbourne Eagles: 59 90 83 87 88 76 82 83 91 78 75

a. Construct a back-to-back stem plot to display this data, with the 'Abbotsford Avengers' on the left.

Explanation

Step 1: Consider the most appropriate stem.

The data values are two digit numbers.

The stems will refer to 'tens'.

The leaves will refer to 'ones'.

The data (from both teams) ranges from 36 to 91.

The appropriate stems are 3, 4, 5, 6, 7, 8 and 9.

3
4
5
6
7
8
9

Step 2: Fill in the leaves for the left category.

They should increase as they move further from the stem. Make sure to add the category title.

Abbotsford Avengers	
6	3
1	4
7 3	5
8 6 4 2	6
7 4 0	7
	8
	9

Step 3: Fill in the leaves for the right category.

They should increase as they move further from the stem. Make sure to add the category title.

Abbotsford Avengers		East Melbourne Eagles
6	3	
1	4	
7 3	5	9
8 6 4 2	6	
7 4 0	7	5 6 8
	8	2 3 3 7 8
	9	0 1

Step 4: Construct a key.

As decided in step 1, the stems refer to 'tens' and the leaves refer to 'ones'.

Answer

Key: 5 | 9 = 59%

Abbotsford Avengers		East Melbourne Eagles
6	3	
1	4	
7 3	5	9
8 6 4 2	6	
7 4 0	7	5 6 8
	8	2 3 3 7 8
	9	0 1

Continues →

Back to contents

b. Which team had a higher median *successful passes*?

Explanation

Find the median *successful passes* for both teams.

Each team has 11 players so the median will be in the 6^{th} position.

Key: 5 | 9 = 59%

Abbotsford Avengers		East Melbourne Eagles
6	3	
1	4	
7 3	5	9
8 6 **4** 2	6	
7 4 0	7	5 6 8
	8	2 **3** 3 7 8
	9	0 1

Abbotsford Avengers: 64%

East Melbourne Eagles: 83%

Answer

East Melbourne Eagles

Displaying data using parallel dot plots

Parallel dot plots are a sequence of dot plots that display and compare the distribution of two or more categories.

A separate dot plot is constructed for each category, but they must use the same scale so the distribution can be compared. There is no restriction on the number of categories.

Parallel dot plots can be used to identify an association between two or more categories.

Worked example 2

A grade 5/6 class took a maths test on fractions. The *number of marks*, by grade, was recorded.

Grade 5:	2	4	5	5	5	6	6	7	7	7	7	9
Grade 6:	5	6	6	7	8	8	8	8	9	9	9	10

a. Construct parallel dot plots to display this data.

Explanation

Step 1: Consider the most appropriate scale.

The data (from both grades) ranges from 2 to 10.

An appropriate scale ranges from 2 to 10 with tick marks for each whole number.

Step 2: Construct the parallel dot plots.

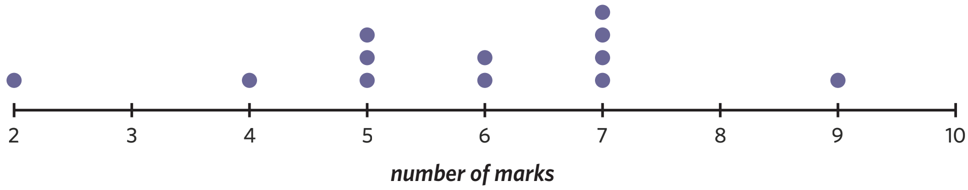
Create two identical axes that are vertically aligned. Mark a dot above the number on the number line each time a value appears in the data set of the associated grade.

Continues →

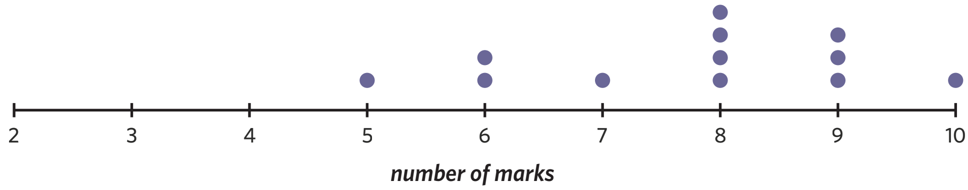
Back to contents

Answer

Grade 5



Grade 6



b. Which grade has a larger range of *number of marks*?

Explanation

Find the range of *number of marks* for each grade.

Grade 5: $9 - 2 = 7$ marks

Grade 6: $10 - 5 = 5$ marks

Answer

Grade 5

Displaying data using parallel boxplots

Parallel boxplots are a sequence of boxplots that display and compare the distribution of two or more categories. They are best used for large data sets.

A separate boxplot is constructed for each category, however, the boxplots share the same axis so the distribution can be compared. There is no restriction on the number of categories.

Parallel boxplots can be used to identify an association between two or more categories.

Worked example 3

Daniel went to his local cafe every Sunday for a year and counted the *number of people* wearing Salomon shoes and the *number of people* wearing Crocs. The five-number summaries for both shoe brands are recorded.

Salomons: 2, 4, 5, 9, 15

Crocs: 0, 2, 4, 7, 8

a. Construct parallel boxplots to display this data.

Explanation

Step 1: Consider the most appropriate scale.

The data (for both shoe brands) ranges from 0 to 15.

An appropriate scale ranges from 0 to 15 with tick marks for each whole number.

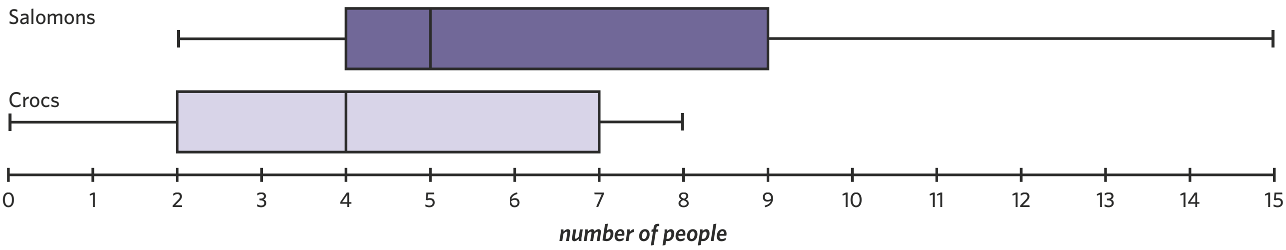
Step 2: Construct the parallel boxplots.

Create one axis and use the five-number summary to construct the boxplots for each shoe brand. There are no outliers, so the whiskers will reach the minimum and maximum *number of people* in both boxplots.

Continues →

Back to contents

Answer



b. For which shoe brand was the *number of people* positively skewed?

Explanation

A positively skewed distribution trails off in a positive direction on the horizontal axis.

Only the distribution of the *number of people* for Salomons is positively skewed.

Answer

Salomons

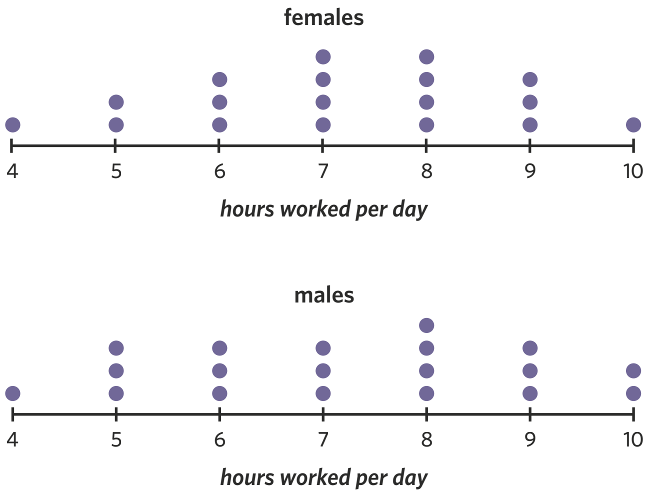
Describing the association between numerical and categorical variables

An association between the categories of the categorical variable can be identified and described by comparing:

- Shape (positively skewed, negatively skewed, approximately symmetrical)
- Centre (median)
- Spread (range, IQR)

If the distributions are similar between the categories of the categorical variable, there is likely to be no association between the variables.

The following parallel dot plots show the number of *hours worked per day* for a group of males and females.

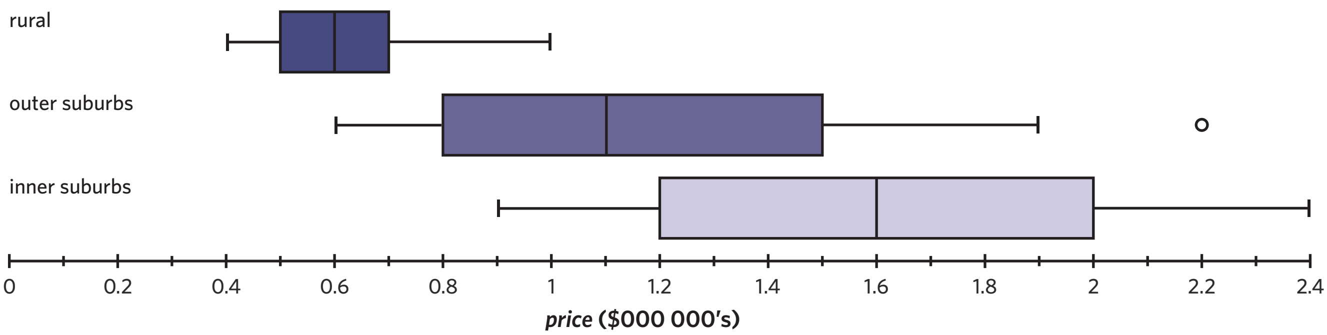


There is no significant difference between the dot plots. Therefore, it is likely that there is no association between *sex* and the number of *hours worked per day*.

Back to contents

Worked example 4

The following parallel boxplots display the distribution of prices for a 3-bedroom house in three different locations in NSW: 'rural', 'outer suburbs' and 'inner suburbs'.



Identify the association between *price* and *location*.

Write a report referencing shape, centre and spread.

Explanation

Step 1: Determine the shapes of the distributions.

'Rural': Positively skewed

'Outer suburbs': Positively skewed

'Inner suburbs': Approximately symmetric

Step 2: Determine the median *price* of each *location*.

'Rural': \$600 000

'Outer suburbs': \$1 100 000

'Inner suburbs': \$1 600 000

Step 3: Calculate the range and *IQR* of each *location*.

range = *maximum value* − *minimum value*

'Rural': 1 000 000 − 400 000 = \$600 000

'Outer suburbs': 2 200 000 − 600 000 = \$1 600 000

'Inner suburbs': 2 400 000 − 900 000 = \$1 500 000

$IQR = Q_3 - Q_1$

'Rural': 700 000 − 500 000 = \$200 000

'Outer suburbs': 1 500 000 − 800 000 = \$700 000

'Inner suburbs': 2 000 000 − 1 200 000 = \$800 000

Step 4: Write a report using shape, centre and spread to explain the association.

Answer

The parallel boxplots show that *price* and *location* are associated. The closer a 3-bedroom house is to a city in NSW, the higher the price of the house.

In relation to shape, the distribution of *price* is positively skewed in locations further from a city in NSW and approximately symmetrical in the inner suburbs.

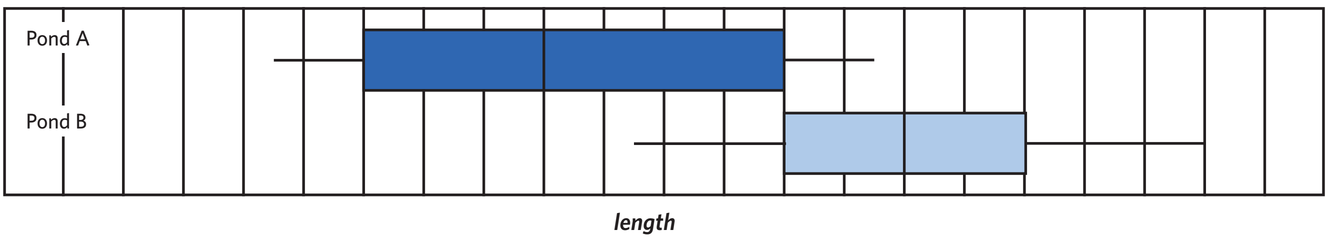
In relation to centre, the median *price* increases as the *location* is closer to a city in NSW.

In relation to spread, the range and *IQR* of *price* tends to be larger in locations closer to a city in NSW.

Exam question breakdown

VCAA 2021 Exam 1 Data analysis Q4

The following boxplots show the distribution of the *length* of fish caught in two different ponds, Pond A and Pond B.



Continues →

Based on the boxplots, it can be said that

A. 50% of the fish caught in Pond A are the same length as the fish caught in Pond B.

B. 50% of the fish caught in Pond B are longer than all of the fish caught in Pond A.

C. 50% of the fish caught in Pond B are shorter than all of the fish caught in Pond A.

D. 75% of the fish caught in Pond A are shorter than all of the fish caught in Pond B.

E. 75% of the fish caught in Pond B are longer than all of the fish caught in Pond A.

Explanation

To solve this question, check whether each option is true or false.

A: This is false. Boxplots do not tell us the exact values of individual data points. ✗

B: This is true. The median of Pond B is larger than the maximum of Pond A. ✓

C: This is false. The median of Pond B is not smaller than the maximum of Pond A. ✗

D: This is false. Q_3 of Pond A is not smaller than the minimum of Pond B. ✗

E: This is false. Q_1 of Pond B is not larger than the maximum of Pond A. ✗

60% of students answered this question correctly.

15% of students incorrectly answered E. This is likely because they misinterpreted the Pond A boxplot and did not realise that the rightmost whisker contains 25% of the data.

Answer

B

2B Questions

Displaying data using back-to-back stem plots

1. A popular local bakery has recently reduced the size of its workforce, and as such needs to reduce their menu.

They are deciding between removing either ‘muffins’ or ‘scones’ from their selection based on the number of *sales* for each.

They have recorded the number of *sales* each day for the month of June in the back-to-back stem plot.

Key: 1 | 9 = 19

muffins		**scones**
	1	9
3	2	0 1 2 2 4 4
9 8	2	6 6 6 7 8 8 9 9 9
5 4 4 3 3 2 1	3	0 0 1 1 1 2 2 3
9 9 8 8 7 6 6 5	3	5 5 6 7 8
4 4 3 2 2 2 1	4	
9 8 7 6	4	5
2	5	

Use the stem plot to fill in the gaps in the following sentences.

The median number of ‘scones’ sold per day is __________ while the median number of ‘muffins’ sold per day is __________. In order to maximise revenue, the bakery should remove __________ from their selection.

Back to contents

2. The *points scored* by the NSW Blues and Queensland Maroons in the first 20 State of Origin rugby league games are recorded.

NSW: 20 7 5 12 10 22 12 2 22 18 21 6 22 24 18 20 6 8 26 6

Queensland: 16 11 10 24 6 43 29 14 12 2 14 20 16 20 16 16 12 10 18 16

a. Construct a back-to-back stem plot to display this data, with the NSW Blues on the left.

b. Which team has the larger range of *points scored*?

c. Which team has the lower median number of *points scored*?

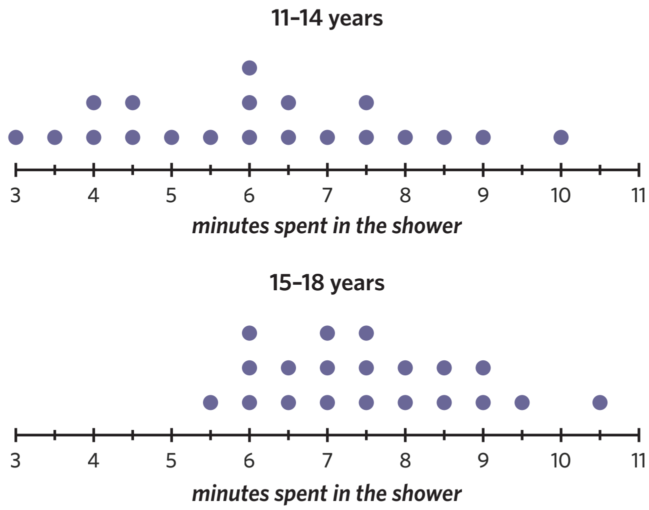
d. In rugby league, a try is worth 4 points, or 6 points if the place kick is accurate, a penalty kick is worth 2 points and a drop goal is worth 1 point. Why do you think most of the leaves in the back-to-back stem plot are even numbers?

Displaying data using parallel dot plots

3. The number of *minutes spent in the shower* by each participant at band camp was recorded correct to the nearest 30 seconds.

The participants were then categorised by age into two groups: '11–14 years' and '15–18 years'.

The results are displayed in the following dot plots.



a. In this circumstance, *age* is classified as what kind of variable?

A. Discrete numerical variable

B. Nominal categorical variable

C. Continuous numerical variable

D. Ordinal categorical variable

b. Which of these statements is not true?

A. The median number of *minutes spent in the shower* increases as age increases from 11–14 years to 15–18 years.

B. The modal number of *minutes spent in the shower* for 11–14 year olds is 6 minutes.

C. The middle 50% of data for 15–18 year olds is less variable than the middle 50% of data for 11–14 year olds.

D. The number of *minutes spent in the shower* is more variable for 15–18 year olds than for 11–14 year olds.

---

4. A fortnight worth of *daily sales* of Aesop and Sukin hand moisturiser brands at a local shop were recorded.

Aesop: 12 11 14 12 10 12 13 11 14 16 11 12 11 11

Sukin: 11 5 11 14 12 18 9 10 13 10 8 17 11 12

a. Construct parallel dot plots to display this data.

b. Which brand has *daily sales* that are positively skewed?

c. Which brand has a greater *IQR*?

Displaying data using parallel boxplots

5. A Unit 3&4 General Mathematics teacher monitored the results of her class in the Edrolo AOS tests. At the end of the week, the teacher was expected to present the data to the Head of Mathematics. In her report, she displayed the data using parallel boxplots.

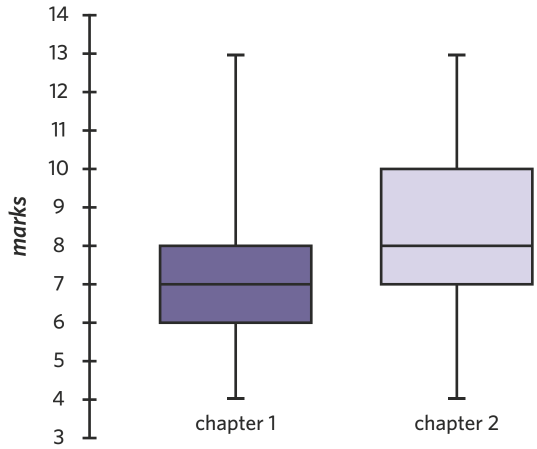
   Which of the following statements relating to this data is not correct?

   A. The median chapter 2 result was higher than the median chapter 1 result.

   B. The lowest 25% of results in the chapter 2 test was better than the lowest 25% of results in the chapter 1 test.

   C. The *IQR* for the chapter 1 test was greater than the *IQR* for the chapter 2 test.

   D. The range for the chapter 1 test was the same as the range for the chapter 2 test.



---

6. The number of *wins* that Essendon and North Melbourne had in each of the seasons in the 2010's is represented in the following five-number summaries.

   Essendon: 3, 7, 11.5, 12, 14

   North Melbourne: 6, 10, 11.5, 13, 14

   a. Construct parallel boxplots to display this data.

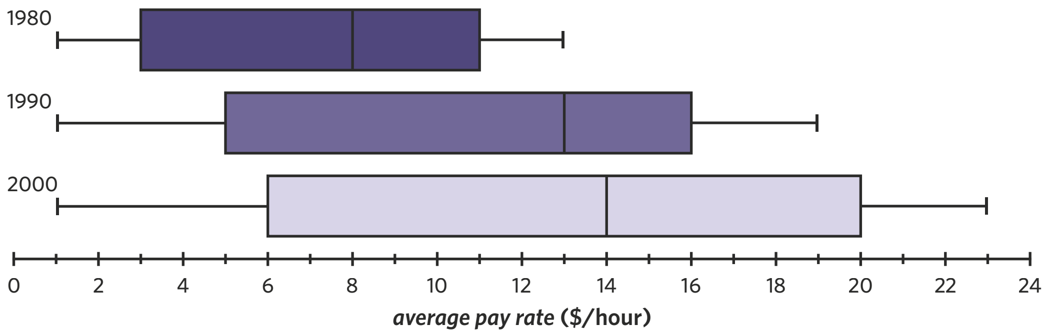
   b. Which team has a greater *IQR* of *wins* in a season?

   c. Which team has a stronger negative skew in the number of *wins* in a season?

Describing the association between numerical and categorical variables

7. The *average pay rate* ($/hour) of workers from 50 countries were tracked in the years '1980', '1990' and '2000'.

   The results are displayed in the following parallel boxplots.



   a. Which of the following statements is not true?

   A. In 1980, over 75% of countries had an *average pay rate* lower than the median *average pay rate* in 1990.

   B. In 1990, 75% of the countries had an *average pay rate* lower than the median *average pay rate* in 2000.

   C. In 1990, there was more variation in *average pay rate* in the middle 50% of countries than the middle 50% of countries in 1980.

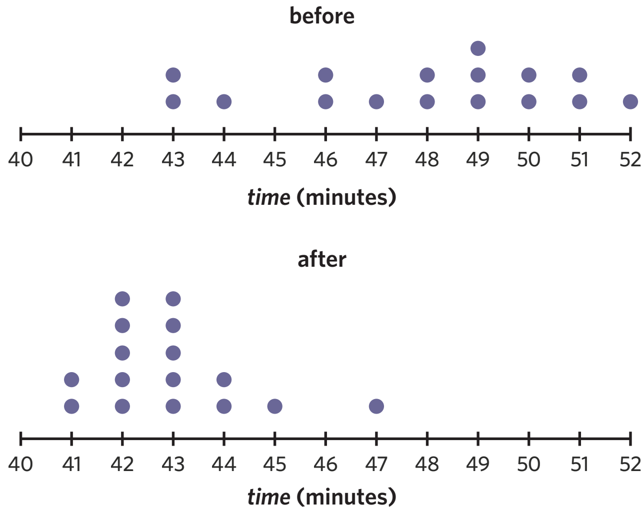
   D. In 2000, the top 50% of countries had an *average pay rate* higher than any of the countries in 1980.

   b. Is there an association between the *average pay rate* in these countries and *year*? Write a brief explanation, referencing centre.

8. Charlotte is a long distance runner. Her coach formulated an intense training program to prepare her for a big competition next year.

In order to see if the training program was effective, Charlotte tracked the *time* it took her to run 10 km (correct to the nearest minute) before and after she completed the program.

Her results are shown in the following parallel dot plots.



a. Did the training program improve Charlotte's running time? Write a brief explanation, referencing centre.

b. Did the training program improve Charlotte's consistency? Write a brief explanation, referencing spread.

9. The daily *maximum temperature* (°C) was recorded in the month of October for Hobart and Launceston.

Key: 1 | 4 = 14 °C

Hobart		Launceston
8	0	8 9
4 4 3 3 2 2 2 2 2 1	1	0 1 1 2 2 3 3 4 4 4
8 8 7 7 7 7 7 6 6 6 5 5 5 5 5 5	1	5 5 5 5 6 6 7 7 8 8 9 9 9 9
2 1 1 0	2	0 2 4 4
	3	9

a. Is there an association between the daily *maximum temperature* (°C) and *location*? Write a brief explanation, referencing centre.

b. Is there an association between the variability of the daily *maximum temperature* (°C) and *location*? Write a brief explanation, referencing spread.

c. Is there an association between the distribution of the daily *maximum temperature* (°C) and *location*? Write a brief explanation, referencing shape.

Joining it all together

10. The number of *years of education* of 1000 random adults was recorded in Argentina, Germany and China, and the five-number summaries for each country are calculated.

Argentina: 3, 7, 10, 12, 18

Germany: 5, 11, 14, 16, 20

China: 3, 5, 8, 11, 18

a. Construct parallel boxplots to display this data.

b. Fill in the blanks:

75% of adults in Germany have more *years of education* than 75% of adults in __________.

The *IQR* of the *years of education* of adults in Argentina is equal to the *IQR* of adults in __________.

c. Is there an association between *years of education* and *country*? Write a brief explanation, referencing centre.

d. Is there an association between the variability of *years of education* and *country* of residence? Write a brief explanation, referencing spread.

Back to contents

Exam practice

11. 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.

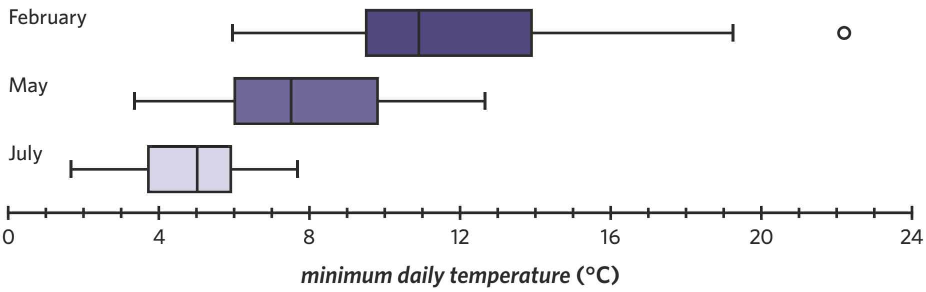
VCAA 2016 Exam 1 Data analysis Q8

45% of students answered this question correctly.

---

12. The five-number summary for the distribution of *minimum daily temperature* for the months of February, May and July in 2017 is shown in the table. The associated boxplots are shown following the table.

month	minimum	Q_1	median	Q_3	maximum
February	5.9	9.5	10.9	13.9	22.2
May	3.3	6.0	7.5	9.8	12.7
July	1.6	3.7	5.0	5.9	7.7



Explain why the information given supports the contention that *minimum daily temperature* is associated with the *month*. Refer to the values of an appropriate statistic in your response. (2 MARKS)

VCAA 2019 Exam 2 Data analysis Q3

The average mark on this question was **1**.

---

13. The following back-to-back stem plot displays the *wingspan*, in millimetres, of 32 moths and their *place of capture* (forest or grassland).

Key: 1 | 8 = 18

***wingspan* (mm)**

forest (n = 13)		grassland (n = 19)
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	

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)

VCAA 2017 Exam 2 Data analysis Q2e

The average mark on this question was **1**.

Back to contents

Questions from multiple lessons

Data analysis

14. In 2022, Yohan Blake recorded his time running 100 m weekly, relative to 9.80 seconds. For example, a time of 9.85 seconds would be recorded as 0.05 seconds. Over the year, he had a mean time of 0.03 seconds, and a standard deviation of 0.16 seconds. What is the standardised value of one of his sprints, which was recorded as −0.10 seconds? Round your answer to two decimal places.

A. 0.81 **B.** 0.72 **C.** −8.67 **D.** −0.72 **E.** −0.81

Adapted from VCAA 2017 Exam 1 Data analysis Q10

Recursion and financial modelling *Year 11 content*

15. The number of coins added to a coin collection each month follows a geometric sequence.

4 coins were added to the collection in the first month.

8 coins were added to the collection in the second month.

16 coins were added to the collection in the third month.

Assuming this sequence continues, how many coins would there be in total after five months?

A. 28 **B.** 60 **C.** 64 **D.** 124 **E.** 252

Adapted from VCAA 2015 Exam 1 Number patterns Q4

Data analysis

16. IQ is normally distributed with a mean of 100 and a standard deviation of 15.

a. Calculate the standardised IQ (z-score) of an individual with an IQ of 73. (1 MARK)

b. What percentage of people are expected to have an IQ over 115? (1 MARK)

c. Estimate the number of people with an IQ between 85 and 130 in a sample of 850 people. Give your answer correct to the nearest whole number. (1 MARK)

Adapted from VCAA 2018NH Exam 2 Data analysis Q2

Back to contents

2C Associations between two numerical variables

STUDY DESIGN DOT POINTS

- response and explanatory variables and their role in investigating associations between variables
- 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)

2A 2B 2C 2D

KEY SKILLS

During this lesson, you will be:

- identifying the response and explanatory variables
- using technology to construct scatterplots
- describing the relationship between two numerical variables.

KEY TERMS

- Response variable
- Explanatory variable
- Scatterplot
- Strength
- Direction
- Form

Once data has been collected for two numerical variables, it is useful to examine them graphically to determine if any associations exist. When doing this, it is important to identify which of the two variables is the response variable and which is the explanatory variable. From here, the data can be plotted on a scatterplot which allows conclusions to be drawn about the relationship between the two variables.

Identifying the response and explanatory variables

The **response variable**, *RV*, may be explained or predicted by changes in the explanatory variable. It can also be called the dependent variable.

See worked example 1

The **explanatory variable**, *EV*, is used to explain or predict the changes observed in the response variable. It can also be called the independent variable.

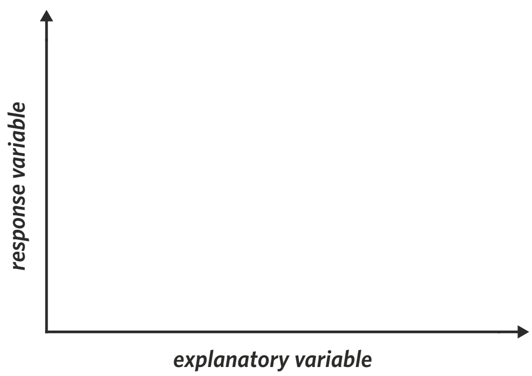
For example, consider the *age* of children and their *height*.

The *height* of children is the response variable, as increases in *height* may be predicted from increases in *age*.

The *age* of children is the explanatory variable, as increases in *age* can predict increases in *height*.

When comparing the relationship between a response variable and its explanatory variable, the data may be represented on a graph. In this case, the response variable is positioned on the vertical axis, and the explanatory variable on the horizontal axis.

See worked example 2



Back to contents

Worked example 1

The following question was posed: 'Can the *number of ice creams sold* be predicted from the *temperature*?'
Identify the response variable, *RV*, and the explanatory variable, *EV*.

Explanation

Step 1: Assess whether each variable is predicting an outcome or being predicted.

The *number of ice creams sold* is being predicted from changes in *temperature*.

Step 2: Classify each variable as either response or explanatory.

The *RV* is predicted from changes in the *EV*.

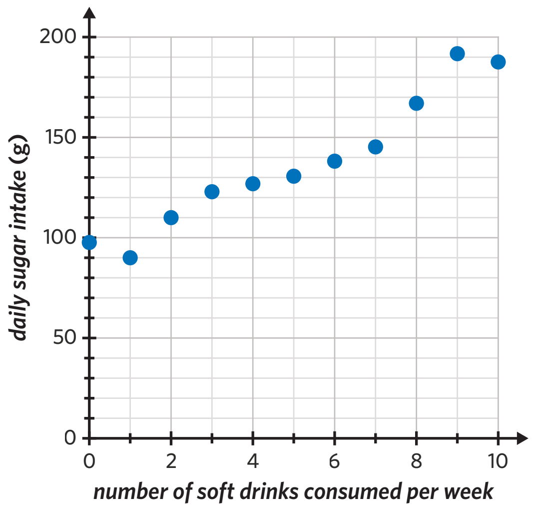
Answer

RV: *number of ice creams sold*

EV: *temperature*

Worked example 2

A study was conducted on the association between *daily sugar intake* and the *number of soft drinks consumed per week*. The following scatterplot displays the data that was collected.



Identify the response variable, *RV*, and the explanatory variable, *EV*.

Explanation

Step 1: Recall which variable is positioned on each axis for a scatterplot.

The *RV* is positioned on the vertical axis, and the *EV* on the horizontal axis.

Step 2: Identify the *RV* and the *EV* from the graph.

Answer

RV: *daily sugar intake*

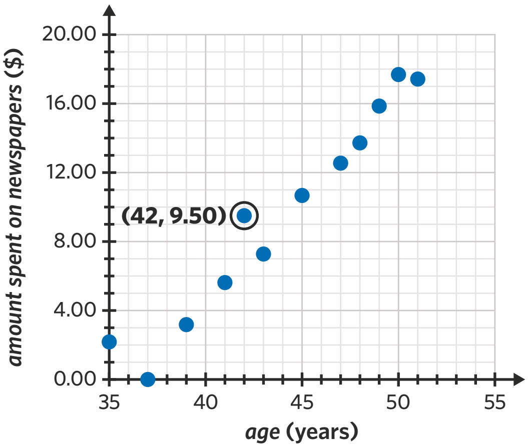
EV: *number of soft drinks consumed per week*

Using technology to construct scatterplots

A **scatterplot** is a display used to represent data relating two numerical variables. Each point represents an individual data entry with the axes providing the numerical measurements. As mentioned previously, the response variable is positioned on the vertical axis, and the explanatory variable is positioned on the horizontal axis.

Back to contents

For example, the point circled on the following scatterplot represents a 42-year-old individual who spent $9.50 on newspapers.



When provided with a table of data, calculators can be used to construct scatterplots.

Worked example 3

10 students were competing in the school cross-country. Each student's house would be awarded *points* based on their placing in the race. The *time* that it took each student to complete the race and the number of house *points* they received were recorded.

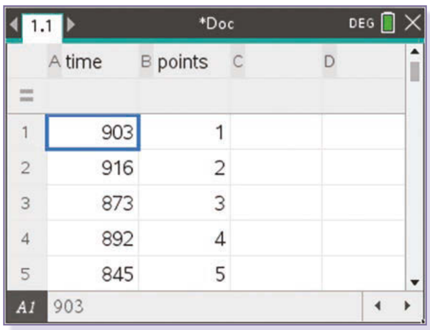
time (seconds)	903	916	873	892	845	823	794	702	740	715
points	1	2	3	4	5	6	7	8	9	10

With a calculator, use the data in the table to construct a scatterplot.

Explanation – Method 1: TI-Nspire

Step 1: From the home screen, select '1: New' → '4: Add Lists & Spreadsheet'.

Step 2: Name a list 'time' and another list 'points' and enter the data from the table.



Step 3: Identify the response and explanatory variables.

Students are awarded *points* based on their *time*, so *time* is used to predict *points*.

RV: *points*

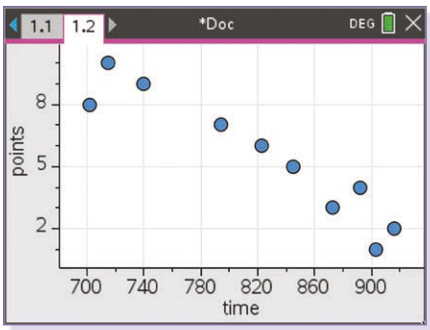
EV: *time*

Step 4: Press ctrl + doc▸, and select 'Add Data & Statistics'.

Step 5: Add the variables on each axis using the 'Click to add variable' function.

The *RV* will be positioned on the vertical axis and the *EV* will be positioned on the horizontal axis.

Answer



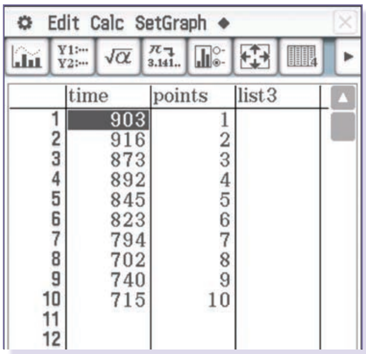
Continues →

Back to contents

Explanation – Method 2: Casio ClassPad

Step 1: From the main menu, tap Statistics.

Step 2: Name a list 'time' and another list 'points' and enter the data from the table.



	time	points	list3
1	903	1	
2	916	2	
3	873	3	
4	892	4	
5	845	5	
6	823	6	
7	794	7	
8	702	8	
9	740	9	
10	715	10	
11			
12			

Step 3: Identify the response and explanatory variables.

Students are awarded *points* based on their *time*, so *time* is used to predict *points*.

RV: *points*

EV: *time*

Step 4: Configure the settings of the graph by tapping the icon in the icon bar.

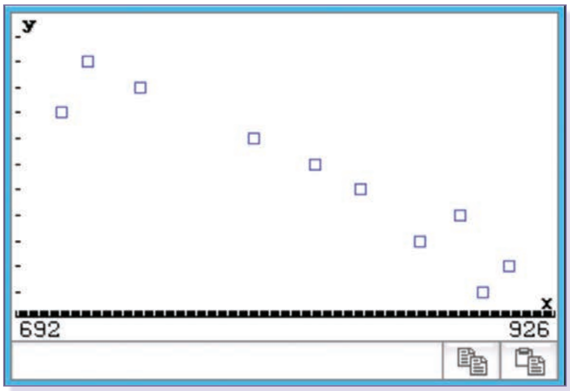
Create a scatterplot by keeping 'Type' as 'Scatter'.

Specify the data set by changing 'XList:' to 'main\time' and 'YList:' to 'main\points'. Tap 'Set' to confirm.

Step 5: Tap the icon bar to plot the graph.

To analyse the graph, tap the icon in the toolbar. Use the left and right arrow keys to navigate along the data points.

Answer



Describing the relationship between two numerical variables

When looking at scatterplots, the relationship between the two numerical variables can be described in terms of strength, direction and form.

Strength refers to how close the data points are to the general trend of the scatterplot. An association can be described as weak, moderate or strong.

Direction refers to the relationship between the two variables.

Positive relationships occur if the response variable increases as the explanatory variable increases.

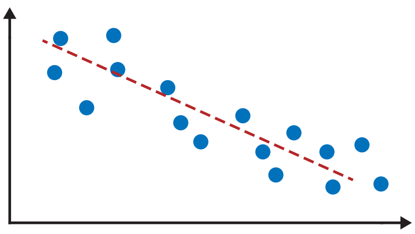
Negative relationships occur if the response variable decreases as the explanatory variable increases.

Form refers to whether the relationship is linear or non-linear.

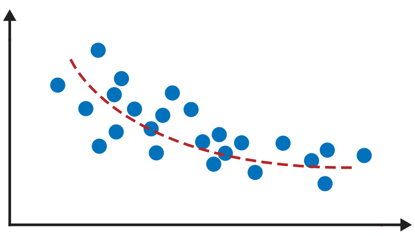
Linear relationships occur when the distribution of data resembles a straight line.

The distribution of data can be described as non-linear if there is a clear association that does not follow a straight line.

Linear relationships will be the primary focus of this chapter.

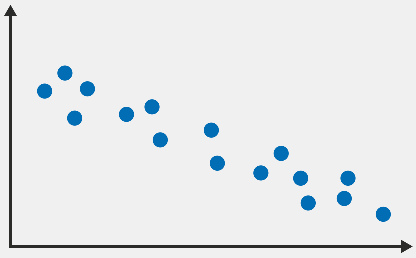
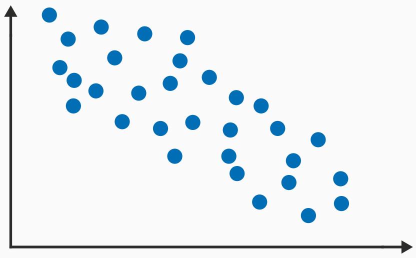
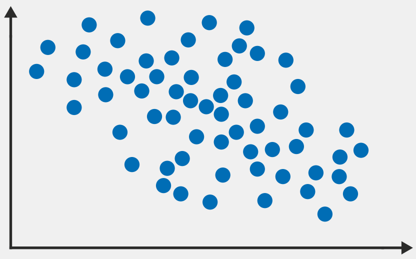
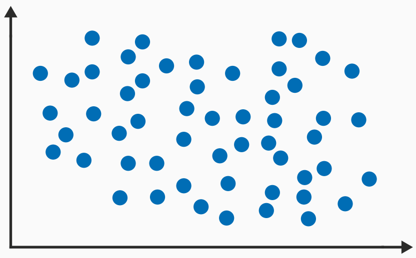
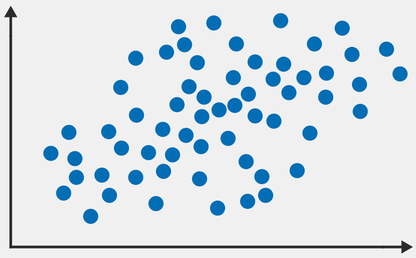
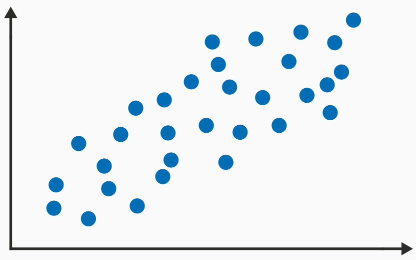
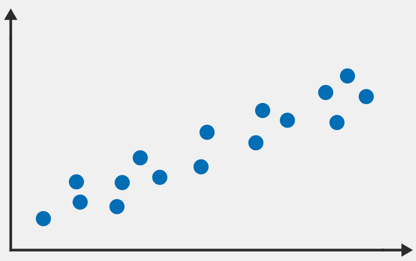


negative linear association

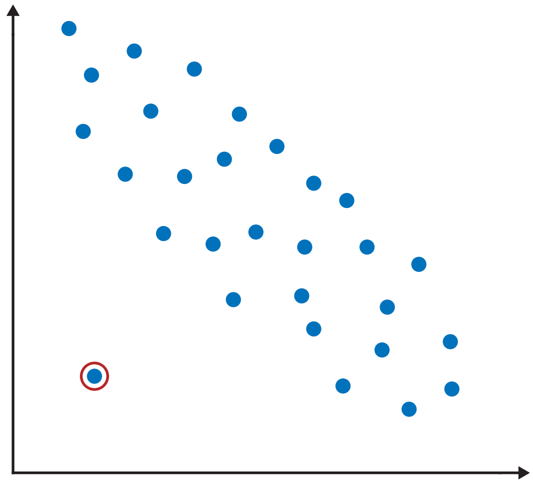


negative non-linear association

Back to contents

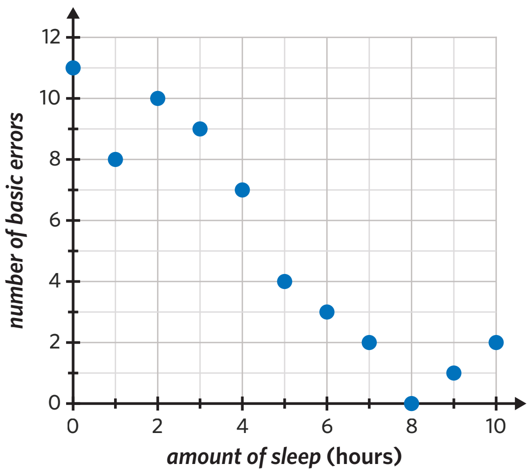
Strength	Direction	Form	Scatterplot
Strong	Negative	Linear	
Moderate	Negative	Linear	
Weak	Negative	Linear	
No association			
Weak	Positive	Linear	
Moderate	Positive	Linear	
Strong	Positive	Linear	

Outliers can also be identified visually, since they exist outside the normal range of the underlying trend.



Worked example 4

An investigation into the association between sleep and work performance was conducted. The following scatterplot displays the collected data.



Describe the relationship in terms of strength, direction, form and potential outliers.

Explanation

Step 1: Determine the strength.

All data points appear to be very close to the underlying trend. The association is strong.

Step 2: Determine the direction.

As the explanatory variable increases, the response variable decreases. The association is negative.

Step 3: Determine the form.

The distribution of data best resembles a straight line. The form is linear.

Step 4: Identify any outliers.

There are no data points that clearly lie outside the normal range of the underlying trend.

Answer

The scatterplot displays a strong, negative, linear relationship, with no visible outliers.

Back to contents

Exam question breakdown

Adapted from VCAA 2020 Exam 2 Data analysis Q4bi

The *age*, in years, *top speed*, in kilometres per hour, and *weight*, in kilograms, of a sample of 12 pandas aged 13 to 15 years are shown in the following table.

age (years)	*top speed* (km/h)	*weight* (kg)
13	26.7	70.1
13	25.1	90.4
13	26.5	73.2
13	25.8	85.0
14	26.1	84.3
14	23.5	95.6
14	28.3	71.7
14	23.8	95.0
15	27.3	80.2
15	25.4	87.4
15	24.1	94.9
15	29.6	65.3

A line of best fit is to be fitted to the data with the aim of predicting *top speed* from *weight*.

Name the explanatory variable for this line of best fit. (1 MARK)

Explanation

Step 1: Assess whether each variable is predicting an outcome or being predicted.

The *top speed* is being predicted from changes in *weight*.

Step 2: Classify each variable as either response or explanatory.

The *RV* is predicted from changes in the *EV*.

Answer

weight

78% of students answered this type of question correctly.

Many of the students who answered this type of question incorrectly selected a variable unrelated to the question, such as *age*.

2C Questions

Identifying the response and explanatory variables

1. Fill in the blanks for the following statement.

The __________ variable is used to predict the changes observed in the __________ variable. When representing the variables on a scatterplot, the response variable is positioned on the __________ axis and the explanatory variable is positioned on the __________ axis.

A. response, explanatory, horizontal, vertical

B. response, explanatory, vertical, horizontal

C. explanatory, response, horizontal, vertical

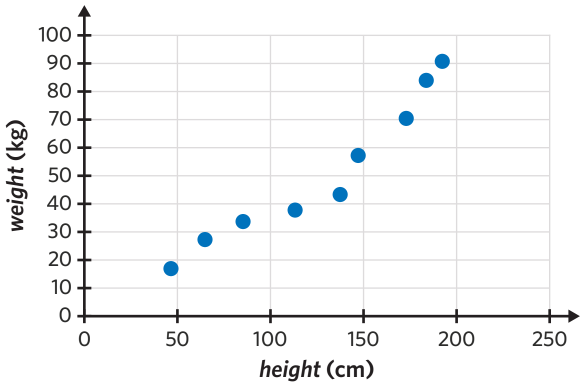
D. explanatory, response, vertical, horizontal

Back to contents

2. Identify the response variable, *RV*, and the explanatory variable, *EV*, in each of the following questions.
   a. Can the *number of umbrellas sold* on a day be predicted from the *amount of rainfall*?
   b. Can the *age* of a second-hand TV predict its *selling price*?
   c. Does the *price* of a watermelon depend on the *season*?
   d. Do *years of experience* affect an individual's *salary*?

3. Identify the response variable, *RV*, and the explanatory variable, *EV*, in each of the following pairs of variables.
   a. *high score* on a game and *time spent playing*
   b. *distance* from work and *money spent* on petrol
   c. *amount of rainfall* and *month*
   d. *age* and *number of homes owned*

4. From the following graph, identify the response variable and explanatory variable.

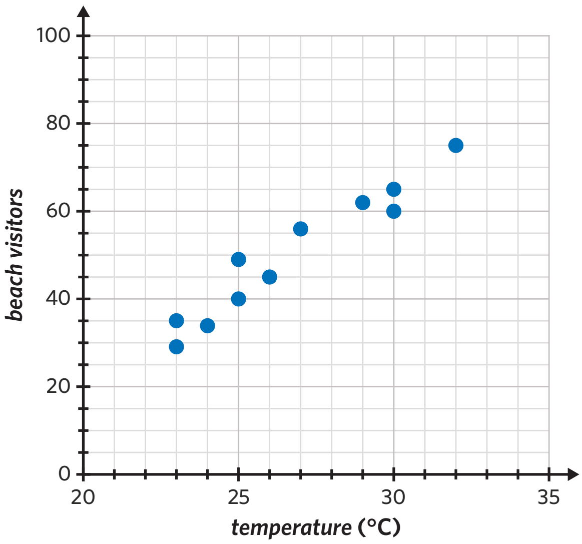


5. Darren plans on collecting data to determine whether the *amount of sleep* can be predicted from *time spent watching Grey's Anatomy*.

   Which variable will be positioned on the horizontal axis if Darren wants to display the data on a scatterplot?

Using technology to construct scatterplots

6. Consider the following scatterplot.



Back to contents

The table used to construct this scatterplot is

A.

temperature (°C)	23	23	24	25	25	26	27	29	30	30	32
beach visitors	40	30	35	50	50	45	55	75	60	65	75

B.

temperature (°C)	23	23	24	25	25	26	27	29	30	30	32
beach visitors	35	29	34	40	49	45	56	62	60	65	75

C.

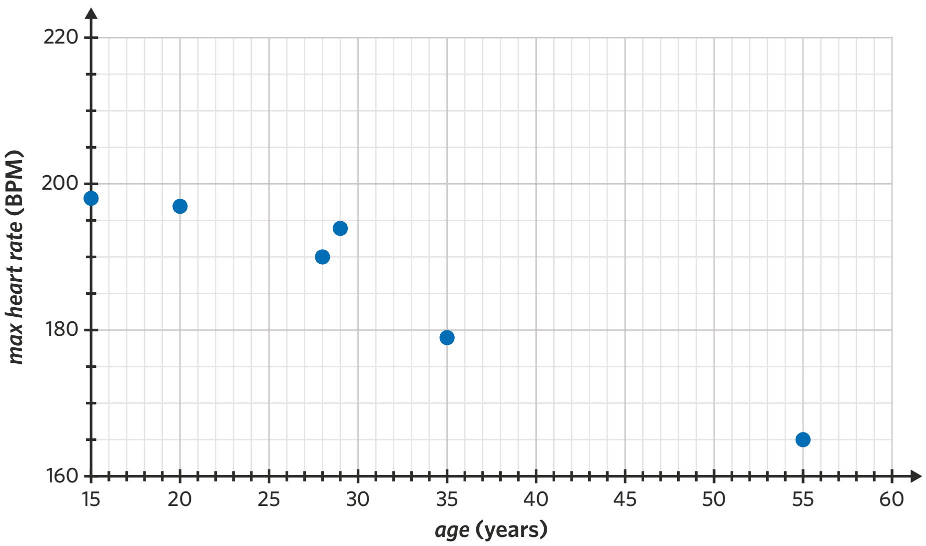
temperature (°C)	35	29	34	40	49	45	56	62	60	65	75
beach visitors	23	23	24	25	25	26	27	29	30	30	32

D.

temperature (°C)	25	25	30	30	27	30	35	30	30	30	35
beach visitors	30	30	35	43	55	40	62	65	70	65	75

7. The following table was used to make the scatterplot shown. Fill in the missing data points that haven't been added to the scatterplot.

age (years)	18	15	42	35	55	29	28	20	31
max heart rate (BPM)	205	198	186	179	165	194	190	197	188



8. The *musical performance* and *mathematical performance* of 10 students were measured on two 100-mark tests. A teacher wanted to determine whether a student's *musical performance* can be predicted from their *mathematical performance*.

Use a CAS to construct a scatterplot from the following data showing the association between *musical performance* and *mathematical performance*. The explanatory variable is *mathematical performance*.

mathematical performance	43	62	84	97	39	53	67	71	78	91
musical performance	57	64	91	93	37	60	69	79	75	82

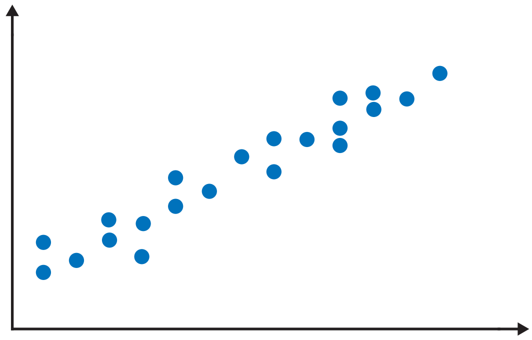
9. An English teacher wished to reduce the number of spelling mistakes her students made. She documented the *number of spelling mistakes* made in each student's essay and asked each student how many books they had read over the past year. Use a CAS to construct a scatterplot from the following data.

number of spelling mistakes	1	2	7	4	9	10	14	19	23	28
number of books read	8	10	7	5	4	3	4	1	2	0

Describing the relationship between two numerical variables

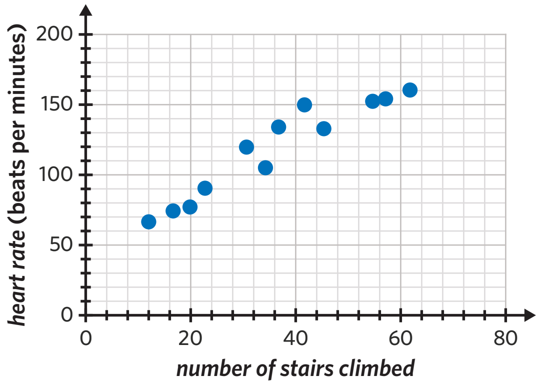
10. The relationship in the scatterplot shown can be described as

A. weak, positive and linear.
B. strong, positive and linear.
C. weak, negative and linear.
D. strong, negative and linear.

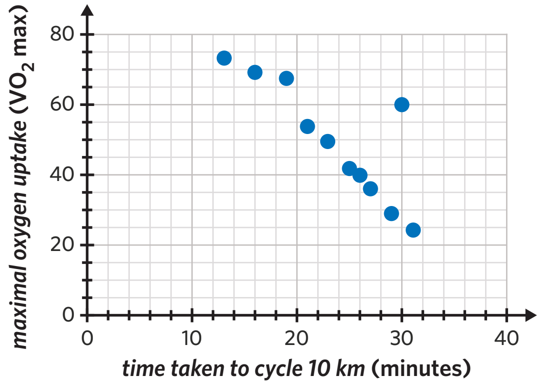


11. For each of the following scatterplots, describe the association between the variables in terms of strength, direction, form and potential outliers.

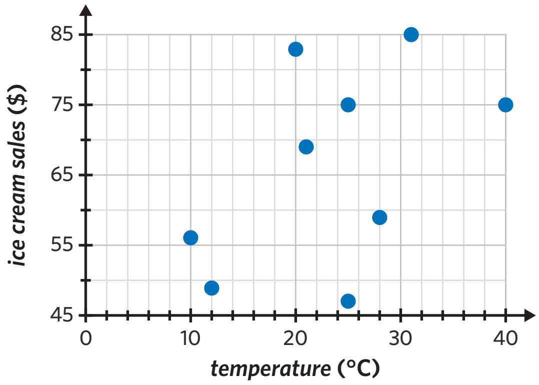
a.



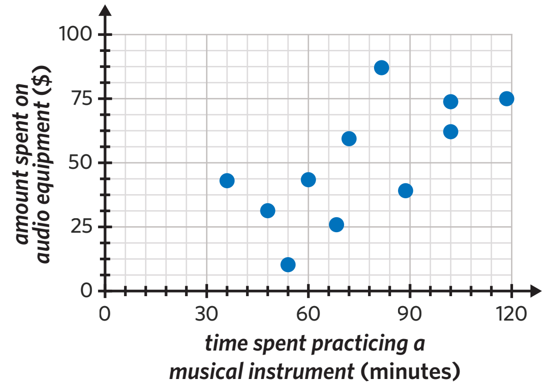
b.



c.



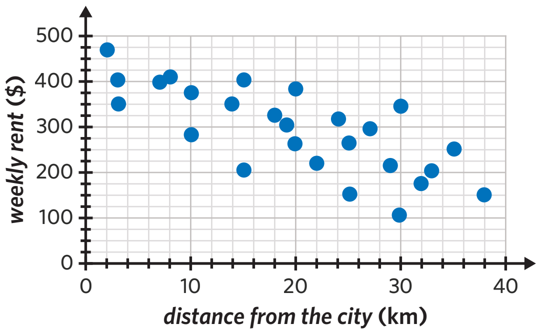
d.



Joining it all together

12. Mavis collected data on how close her friends live to the city and how much they pay in *weekly rent*, and found that *distance from the city*, in km, was a useful predictor of *weekly rent*, in $. She constructed a scatterplot which displayed a moderate, negative, linear relationship. Which of the following graphs could be the scatterplot that Mavis constructed from her data?

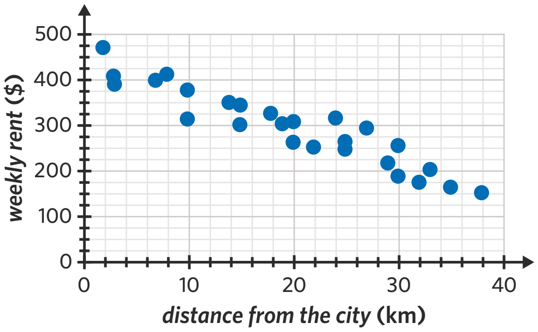
A.



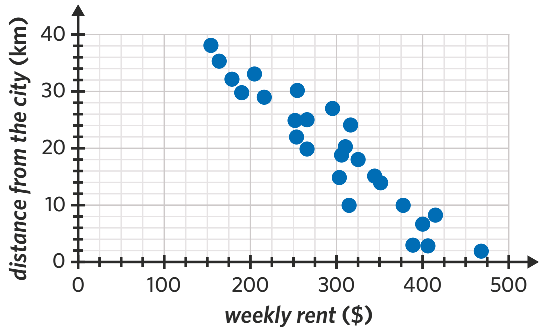
B.

Back to contents

C.



D.



13. In Victoria, if an individual is under 21 years old, they need to have 120 hours of logged practice with a qualified driver before attempting the driving test to obtain their Probationary Driver's Licence (P's). VicRoads wants to review this legislation and has conducted a study.

The study tested ten 18-year-olds who had no prior driving experience, but were then permitted to do up to 150 hours of practice before taking the driving test. Their number of mistakes on the test were recorded.

a. The association between *hours of practice* and *number of mistakes* was investigated. Identify the response and explanatory variables.

b. VicRoads has provided the following data. Use a CAS to construct a scatterplot representing this data.

hours of practice	80	137	22	77	61	110	6	51	150	30
number of mistakes	3	1	5	3	4	2	8	4	0	6

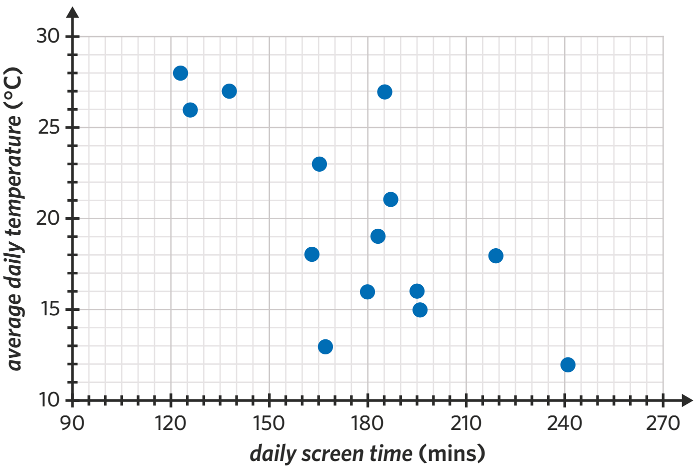
c. Describe the association between the variables in terms of strength, direction and form.

14. Francis wants to determine whether his *daily screen time*, in minutes, on his phone can be predicted by the *average daily temperature*. He assumes that he will spend more time on his phone if it is colder outside.

Francis collects two weeks of data and presents them in the following table.

average daily temperature (°C)	16	23	18	21	28	12	13	16	27	26	27	19	15	18
daily screen time (min)	195	165	219	187	123	241	167	180	138	126	185	183	196	163

He uses the data to construct the following scatterplot.



a. Identify Francis' mistake in the scatterplot he has constructed.

b. Use your CAS to reconstruct the scatterplot from the table.

c. Describe the association between the variables in terms of strength, direction and form.

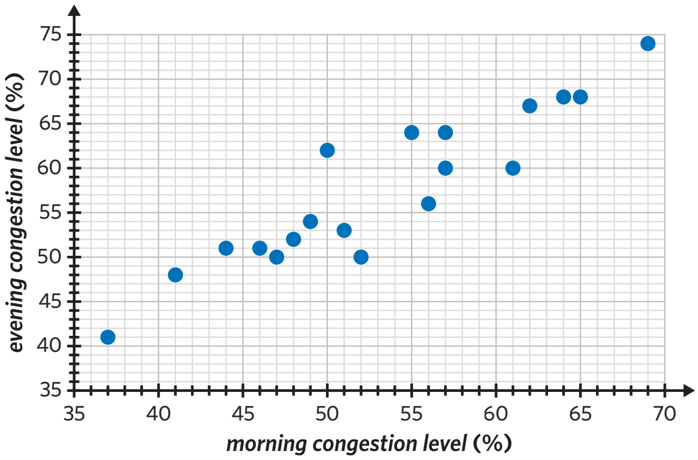
Back to contents

Exam practice

15. 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 following scatterplot.



A line of good fit is to be fitted to the data with the aim of predicting *evening congestion level* from *morning congestion level*.

Name the response variable in this line of good fit. (1 MARK)

Adapted from VCAA 2018 Exam 2 Data analysis Q2b

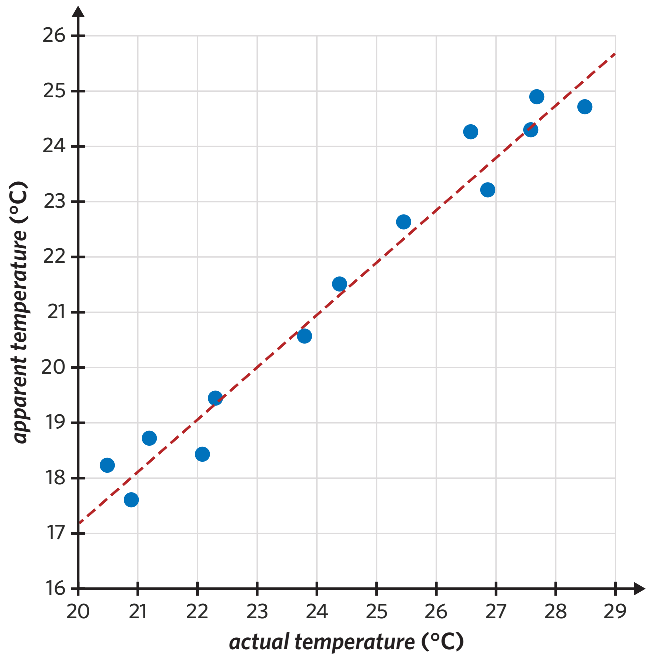
90% of students answered this type of question correctly.

---

16. The data in the following table shows a sample of actual temperatures and apparent temperatures recorded at the 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



Use the scatterplot to describe the association between *apparent temperature* and *actual temperature* in terms of strength, direction and form. (1 MARK)

Adapted from VCAA 2016 Exam 2 Data analysis Q3a

78% of students answered this type of question correctly.

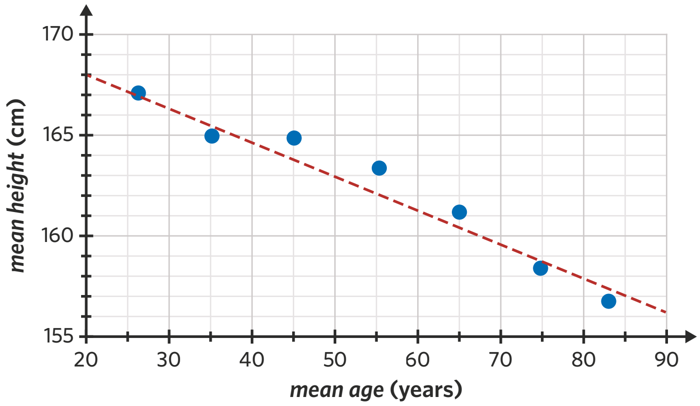
Back to contents

17. The following table shows the *mean age*, in years, and the *mean height*, in centimetres, of 648 women from seven different age groups.

	age group						
	Twenties	**Thirties**	**Fourties**	**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

Data: J Sorkin 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

A scatterplot displaying this data shows an association between the *mean height* and *mean age* of these women. In an initial analysis of the data, a line is fitted by eye, as shown.



Describe this association in terms of strength and direction. (1 MARK)

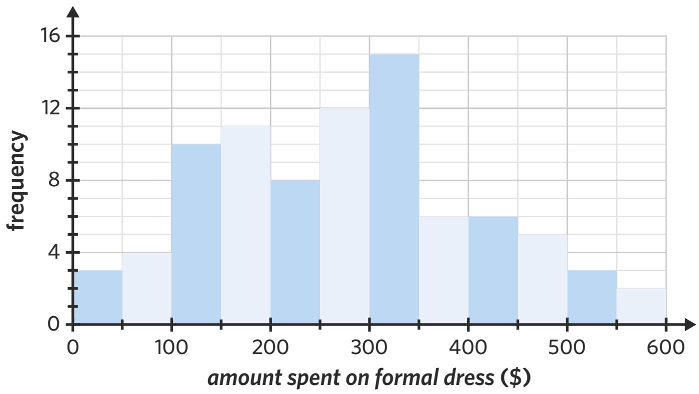
VCAA 2020 Exam 2 Data analysis Q6b

51% of students answered this question correctly.

Questions from multiple lessons

Data analysis *Year 11 content*

18. An elite private school, Edrollington Ladies' Academy has just had their Year 12 Formal. The following histogram displays the distribution of the *amount spent on formal dress*, in dollars, of a sample of 85 students at the school.



The *amount spent on formal dress* for this sample is most frequently

A. greater than or equal to \$150 and less than \$200.

B. greater than or equal to \$200 and less than \$250.

C. greater than or equal to \$250 and less than \$300.

D. greater than or equal to \$300 and less than \$350.

E. greater than or equal to \$350 and less than \$400.

Adapted from VCAA 2018NH Exam 1 Data analysis Q3

Back to contents

Data analysis

19. A survey was conducted with a sample of workers in a Melbourne office to investigate the association between *age* (under 30 years, 30 years or over) and *transport to work* (car, bike, public transport). The distribution is shown in the following percentage segmented bar chart.

There are 80 people under 30 years of age and 120 people 30 years or over at the office.

What is the number of workers aged 30 years or older that ride their bike to work?

A. 20

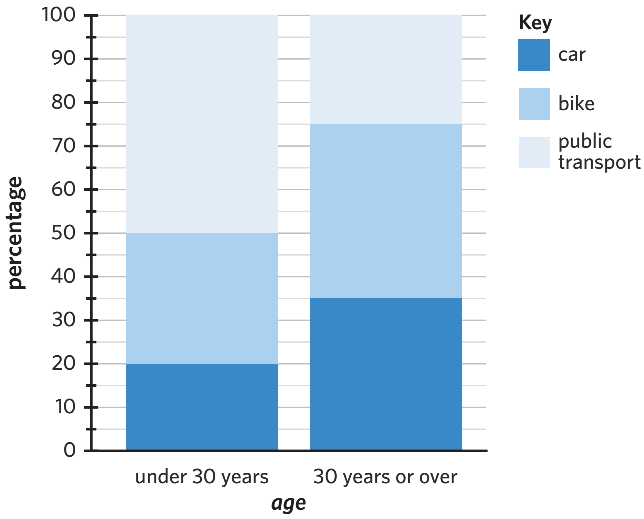
B. 30

C. 40

D. 42

E. 48

Adapted from VCAA 2017 Exam 1 Data analysis Q5



Recursion and financial modelling *Year 11 content*

20. Roger recently purchased a commercial coffee machine.

The value of his coffee machine is depreciated with each coffee made using the unit cost method of depreciation.

The value of his coffee machine, in dollars, after n coffees are made, V_n, can be modelled by the recurrence relation $V_0 = 8000, \quad V_{n+1} = V_n - 0.45$

a. What is the value of the machine after one coffee has been made? (1 MARK)

b. What is the value of the machine after 58 coffees have been made? (1 MARK)

Adapted from VCAA 2017 Exam 2 Recursion and financial modelling Q5a

Back to contents

2D Correlation and causation

STUDY DESIGN DOT POINTS

- Pearson correlation coefficient, *r*, 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
- answering statistical questions that require a knowledge of the associations between pairs of variables

2A 2B 2C **2D**

KEY SKILLS

During this lesson, you will be:

- calculating and interpreting the Pearson correlation coefficient
- distinguishing between correlation and causation.

KEY TERMS

- Pearson's correlation coefficient
- Causation
- Common response
- Confounding variables
- Coincidence
- Observation
- Experimentation

When looking at associations between two numerical variables, it can be helpful to use statistics such as the Pearson correlation coefficient to interpret the data. This is done to make definitive statements about observed associations. While the correlation coefficient provides information on associations, it is important to use appropriate experimentation to determine if an observed association demonstrates causation.

Calculating and interpreting the Pearson correlation coefficient

Pearson's correlation coefficient, r (also known simply as the correlation coefficient), is used to determine the strength and direction of a linear relationship between two numerical variables. When calculating and interpreting the correlation coefficient, the following assumptions are made:

- Data distribution is linear
- Data is numeric
- No outliers are present

The correlation coefficient can be calculated from the following formula:

$$r = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{(n - 1) \times s_x \times s_y}$$

where x and y represent the two numerical variables and n is the number of data values. However it is more efficient to calculate using a calculator.

The Pearson correlation coefficient ranges from -1 to 1. The value of the correlation coefficient determines the strength and direction of a linear association.

A positive r value indicates a positive association, while a negative r value indicates a negative association. The closer r is to -1 or 1, the stronger the association. The closer r is to 0, the weaker the association.

$0.75 \le r \le 1$	Strong, positive, linear association
$0.5 \le r < 0.75$	Moderate, positive, linear association
$0.25 \le r < 0.5$	Weak, positive, linear association
$-0.25 < r < 0.25$	No association
$-0.5 < r \le -0.25$	Weak, negative, linear association
$-0.75 < r \le -0.5$	Moderate, negative, linear association
$-1 \le r \le -0.75$	Strong, negative, linear association

2D THEORY

Worked example 1

A survey followed the study habits of a group of first-year uni students with the aim of predicting their *grade*, as a percentage, from the number of *lectures attended*. The results are shown in the following table.

***grade* (%)**	75	61	92	53	47	86	74
lectures attended	11	6	10	5	2	8	9

a. Is the Pearson correlation coefficient, r, an appropriate statistic for this data set?

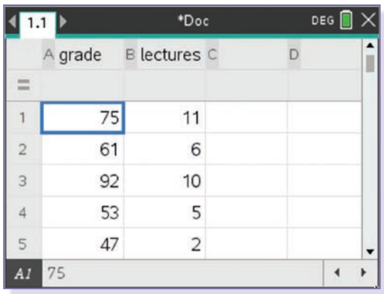
Explanation – Method 1: TI-Nspire

Step 1: From the home screen, select ‘1: New’ → ‘4: Add Lists & Spreadsheet’.

Step 2: Name column A 'grade' and column B 'lectures'.

Enter the *grade* values into column A, starting from row 1.

Enter the *lectures attended* values into column B, starting from row 1.



Step 3: Determine the response and explanatory variables.

The variable *grade* is being predicted from the variable *lectures attended*, so *grade* is the response variable.

RV: *grade*

EV: *lectures attended*

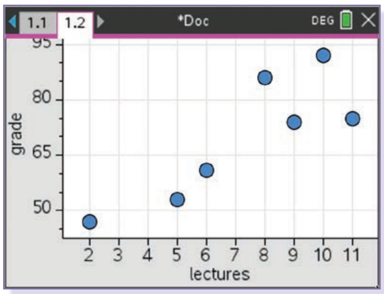
Step 4: Press ctrl + doc, and select ‘5: Add Data & Statistics’.

Move the cursor to the horizontal axis and select ‘Click to add variable’.

Select ‘lectures’.

Move the cursor to the vertical axis and select ‘Click to add variable’.

Select ‘grade’.



Step 5: Determine whether the scatterplot shows a linear or non-linear relationship.

There is no definitive curve or non-linear relationship in this scatterplot. Hence, r is an appropriate statistic.

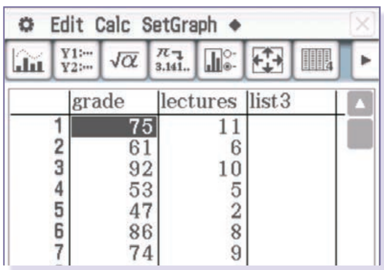
Explanation – Method 2: Casio ClassPad

Step 1: From the main menu, tap Statistics.

Step 2: Name the first list ‘grade’ and the second list ‘lectures’.

Enter the *grade* values into list ‘grade’, starting from row 1.

Enter the *lectures attended* values into list ‘lectures’, starting from row 1.



Continues →

Back to contents

Step 3: Determine the response and explanatory variables.

The variable *grade* is being predicted from the variable *lectures attended*, so *grade* is the response variable.

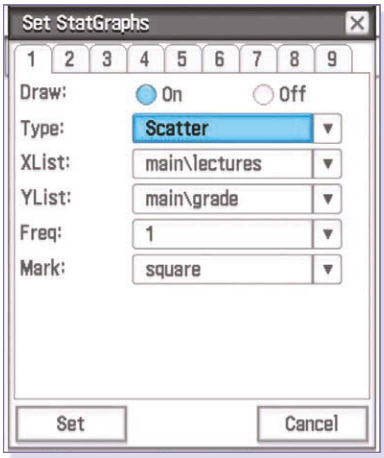
RV: *grade*

EV: *lectures attended*

Step 4: Configure the settings of the graph by tapping .

Create a scatterplot by changing 'Type' to 'Scatter'.

Specify the data set by changing 'XList:' to 'main\lectures' and 'YList:' to 'main\grade'.



Tap 'Set' to confirm.

Step 5: Tap in the icon bar to plot the graph.

Step 6: Determine whether the scatterplot shows a linear or non-linear relationship.

There is no definitive curve or non-linear relationship in this scatterplot. Hence, r is an appropriate statistic.

Answer – Method 1 and 2

Yes, the Pearson correlation coefficient is an appropriate statistic for this data set.

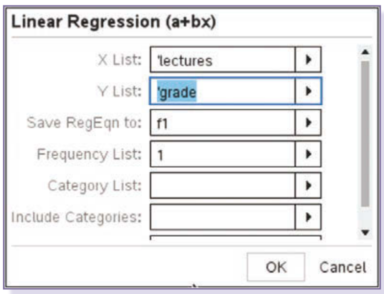
---

b. Determine the value of the correlation coefficient. Give the value correct to two decimal places.

Explanation – Method 1: TI-Nspire

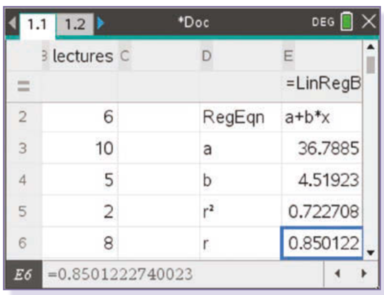
Step 1: From the 'Lists & Spreadsheet' page, press menu and select '4: Statistics' → '1: Stat Calculations' → '4: Linear Regression (a+bx)'.

The variable *lectures attended* is the explanatory variable and the variable *grade* is the response variable, so select 'lectures' in 'X List:' and 'grade' in 'Y List:'



Select 'OK'.

Step 2: Read the r value from the screen and round to two decimal places.



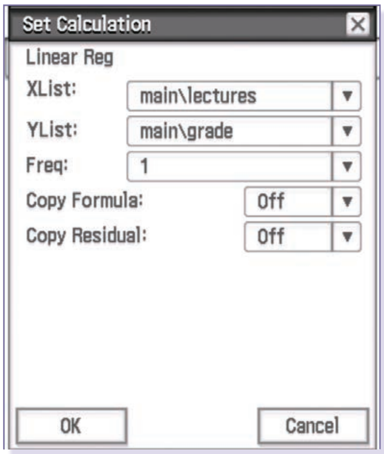
Continues →

Back to contents

Explanation – Method 2: Casio ClassPad

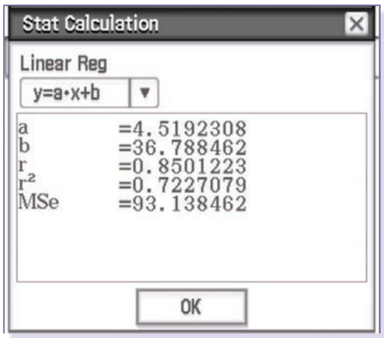
Step 1: Tap 'Calc' → 'Regression' → 'Linear Reg'.

The variable *lectures attended* is the explanatory variable and the variable *grade* is the response variable, so select 'main\lectures' in 'XList:' and 'main\grade' in 'YList:'.



Tap 'OK' to confirm.

Step 2: Read the r value from the screen and round to two decimal places.



Answer – Method 1 and 2

$r = 0.85$

---

c. If possible, describe the association between the variables in terms of strength, direction, and form.

Explanation

Step 1: Recall the r value from part b.

$r = 0.85$

Step 2: Determine the strength of the association.

$0.75 \leq 0.85 \leq 1$

This indicates a strong association.

Step 3: Determine the direction of the association.

0.85 is positive, indicating a positive association.

Answer

Strong, positive, linear association

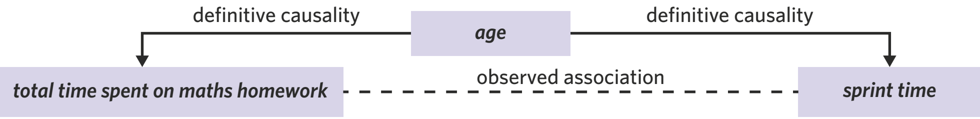
Distinguishing between correlation and causation

While two variables may have a strong correlation, this doesn't necessarily mean that the association implies causation. **Causation** occurs when a change in the explanatory variable definitively causes the observed change in the response variable. A correlation between two variables that do not have a causal relationship can occur due to three different circumstances.

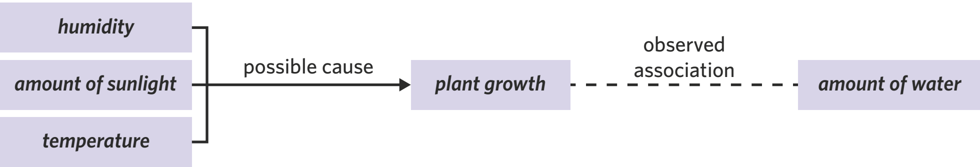
A **common response** to a third variable may be present in the two variables with the observed association. For example, in a primary school, there is a negative correlation between *total time spent on maths homework* and *sprint time*. This means that children who have spent more time on maths homework are associated with faster sprint times. However, it is unlikely that increasing the *total time spent on maths homework* actually decreases a child's *sprint time*.

A common response to a third variable, *age*, is the likely cause of the correlation. Older students will have accumulated more hours on maths homework and they will also have become faster, resulting in lower sprint times.

Back to contents



There could be **confounding variables**, or external variables, that also produce a change in the response variable. For example, a positive association is found between *plant growth* and the *amount of water* poured on the plant per day. Although the *plant growth* could have been affected by a change in *amount of water*, other related factors such as *amount of sunlight*, *temperature* and *humidity* could also have an effect.



Lastly, a correlation may exist by pure **coincidence**. For example, a positive association between *number of siblings* and *number of mistakes* on a spelling test could be purely coincidental.

While **observation** (collecting data to identify an association) is sufficient to identify a correlation, more investigation is required to determine if causation between the variables exist. **Experimentation** is required in order to determine causation in an association. This involves observing or modifying changes in the explanatory variable and recording the changes in the response variable, while all possible external variables remain constant.

If the changes in the explanatory variable result in a persistent change in the response variable, it is reasonable to conclude causation.

Worked example 2

Gunther runs a cafe and collects data throughout the year on the daily number of *hot chocolates purchased*. He finds that there is a negative association between the daily number of *hot chocolates purchased* and the number of *daylight hours* on a particular day.

a. Identify a potential third variable that could have caused this association.

Explanation

During colder seasons, the sun rises later and sets earlier, causing a decreased number of *daylight hours*.

During colder seasons, hot chocolate is more popular, causing an increased daily number of *hot chocolates purchased*.

Therefore, a potential third variable is the *season*.

Answer

season

b. Has Gunther conducted an appropriate experiment to determine causation? Why or why not?

Explanation

Step 1: Recall the requirements for an experiment to determine causation.

All possible external variables must remain constant for an experiment to determine causation.

Step 2: Identify any external variables in the experiment.

From part **a**, it was determined that the *season* could cause a common response. Gunther has collected data across the whole year, so this variable does not remain constant.

Answer

No. The variable *season* does not remain constant across the experiment.

Back to contents

Exam question breakdown

VCAA 2016 Exam 1 Data analysis Q12

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.

Explanation

Step 1: Interpret the association.

A strong positive association means that an increase in the explanatory variable generally results in an increase in the response variable. Similarly, a decrease in the explanatory variable also generally results in a decrease in the response variable. The only options that follow this relationship are A and D.

Step 2: Determine if causation is present.

An experiment needs to be conducted to determine causation instead of just observing an association. In this situation, no experiment has been conducted, so we are unable to conclude that an increase in the explanatory variable definitively causes an increase in the response variable.

Answer

D

67% of students answered this question correctly.

11% of students incorrectly answered A. These students interpreted the association correctly but didn't identify that the association doesn't necessarily mean causation. A further **11%** of students incorrectly answered E. These students recognised that observing the association only shows correlation rather than causation, but misinterpreted the direction of the association.

2D Questions

Calculating and interpreting the Pearson correlation coefficient

1. An r value of 1.00 indicates an association that is

A. strong, positive, and linear

B. moderate, positive, and linear

C. moderate, negative, and linear

D. strong, negative, and linear

---

2. In which of the following circumstances would calculating the correlation coefficient, r, be most useful?

A. Bradley is collecting data on the *favourite sport* of the students in his class.

B. The *weight* (g) and *leg length* (mm) of 10 different mice was measured.

C. Susie records how long it takes her to get to school on each day of the school week.

D. Raj measures the distance between his house and the five closest shopping centres.

Back to contents

3. Max and Penelope collect data on the *goals conceded* by their soccer team each week and the *average sleep*, in hours, of all members of the soccer team the night before each game. They calculated the value of the correlation coefficient to be $r = -0.21$.

   Max says that *goals conceded* and *average sleep* have a negative association.

   Penelope says there is no association.

   Who is correct and why?

4. A group of people were surveyed on the number of hours they spent playing Pokémon GO over one weekend and how many Pokémon they caught in that time. The results are shown in the following table.

hours spent playing Pokémon GO	7	3	4	8	16	9	11	10	9
Pokémon caught	13	2	6	10	38	11	23	16	12

   a. Is the Pearson correlation coefficient, r, an appropriate statistic for this data set? Assume that *hours spent playing Pokemon GO* is the explanatory variable.

   b. Determine the value of Pearson's correlation coefficient, r, for the data shown correct to two decimal places.

   c. If possible, describe the association between the variables in terms of strength, direction, and form.

5. A select group of rural towns were investigated through the census. For each rural town, the *male population* and *average height* was recorded, with *average height* as the response variable. The data is shown in the following table.

male population	3158	7562	9901	1875	2238	10 027	8723	5267	4547
***average height* (m)**	1.76	1.83	1.79	1.86	1.78	1.76	1.84	1.80	1.77

   a. Determine the value of Pearson's correlation coefficient, r, for the data shown correct to two decimal places.

   b. If possible, describe the association in terms of strength, direction, and form.

Distinguishing between correlation and causation

6. A strong, positive, linear association was observed between two variables that are seemingly unrelated. A circumstance that could have caused this association is

   A. a common response to a third variable.

   B. an external variable(s) that also causes a change in the response variable.

   C. a coincidence.

   D. all of the above.

7. The total number of *Nobel Prize winners* from each country was found to have a strong, positive correlation ($r = 0.84$) with the number of *space rock impacts* per year within the country. This strong correlation is likely due to

   A. a common response to a third variable.

   B. an external variable(s) that also causes a change in the response variable.

   C. a coincidence.

   D. a direct causal relationship.

8. Axel runs a bakery with a wide range of products and has found a positive correlation between the number of *cinnamon doughnuts sold* and the *total profit* each day. As a result, he decides to discontinue some popular items to make more cinnamon doughnuts and further increase his profit.

   Has Axel made a logical decision? If not, identify another variable that may also influence the *total profit* to justify your answer.

Back to contents

9. In a study of American cities, there was found to be a positive correlation between the number of *gyms* and the number of *crimes committed* in each city. What variable could cause a common response that would explain this association?

10. Robbie looks at the screen time statistics on his phone to determine whether there is an association between the *time spent on Facebook* and *time spent on Instagram* per day, both in minutes. He collects the data every day for 6 months, and finds that there is a strong association.
    a. Identify three confounding variables that could explain the association.
    b. How could Robbie redesign his experiment so that he would be better able to determine causation?

Joining it all together

11. An association has been observed between the weekly *cream cheese sales* in Forest Hill and the weekly number of *parking tickets* handed out in Sunshine. Eight weeks of data are shown in the following table.

cream cheese sales	46	55	53	61	78	64	95	79
parking tickets	25	34	28	29	21	27	19	18

    a. Assuming *cream cheese sales* is the explanatory variable, is the Pearson correlation coefficient, r, an appropriate statistic for this data set?
    b. Determine the value of the correlation coefficient, correct to two decimal places.
    c. If possible, describe the association between *cream cheese sales* and *parking tickets* in terms of strength, direction and form.
    d. Is it reasonable to conclude that there is a causal relationship between *cream cheese sales* and *parking tickets*? If not, identify what might be causing the observed association.

12. Mikayla plays netball and is trying to determine whether her recovery during the week impacts positively on her performance. For the first 13 games of the season, she collects data on the *time spent on recovery*, in minutes, between games, and the number of *goals* she scores each week. Her results are shown in the following table.

***time spent on recovery* (mins)**	45	62	35	105	96	137	48	56	71	105	60	29	135
goals	5	9	8	9	13	9	8	7	9	10	12	7	15

    a. Is the Pearson correlation coefficient, r, an appropriate statistic for this data set? Why or why not?
    b. Determine the value of the correlation coefficient. Give the value correct to two decimal places.
    c. If possible, describe the association between *time spent on recovery* and number of *goals* in terms of strength, direction, and form.
    d. Identify five confounding variables that could also influence the *number of goals* that Mikayla scores each week.
    e. Has Mikayla conducted an appropriate experiment to determine causation? Why or why not?

Back to contents

Exam practice

13. Data collected over a period of 10 years indicated a strong, positive association between the number of stray cats and the number of stray dogs reported each year ($r = 0.87$) in a large, regional city.

A positive association was also found between the population of the city and both the number of stray cats ($r = 0.61$) and the number of stray dogs (0.72).

During the time that the data was collected, the population of the city grew from 34 564 to 51 055.

From this information, we can conclude that

A. if cat owners paid more attention to keeping dogs off their property, the number of stray cats reported would decrease.

B. the association between the number of stray cats and stray dogs reported cannot be causal because only a correlation of $+1$ or -1 shows causal relationships.

C. there is no logical explanation for the association between the number of stray cats and stray dogs reported in the city so it must be a chance occurrence.

D. because larger populations tend to have both a larger number of stray cats and stray dogs, the association between the number of stray cats and the number of stray dogs can be explained by a common response to a third variable, which is the increasing population size of the city.

E. more stray cats were reported because people are no longer as careful about keeping their cats properly contained on their property as they were in the past.

VCAA 2017 Exam 1 Data analysis Q12

81% of students answered this question correctly.

---

14. The relative humidity (%) at 9 am and 3 pm on 14 days in November 2017 is shown in the following table.

***relative humidity* (%)**	**9 am**	100	99	95	63	81	94	96	81	73	53	57	77	51	41
	3 pm	87	75	67	57	57	74	71	62	53	54	36	39	30	32

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

A least squares line is to be fitted to the data with the aim of predicting the relative humidity at 3 pm (*humidity* 3 *pm*) from the relative humidity at 9 am (*humidity* 9 *am*).

Determine the value of the correlation coefficient for this data set. Round your answer to three decimal places. (1 MARK)

VCAA 2019 Exam 2 Data analysis Q4c

76% of students answered this question correctly.

Questions from multiple lessons

Data analysis *Year 11 content*

15. Parallel boxplots are to be used to display the association between two chosen variables. One of these variables is *height* (cm). The second variable could be

A. *weight* (kg).

B. *arm span* (cm).

C. *ATAR*.

D. *hair colour* (brown, blonde, black, red, other).

E. *foot length* (cm).

Adapted from VCAA 2016 Exam 1 Data analysis Q8

Back to contents

Recursion and financial modelling *Year 11 content*

16. Arlo borrowed a decent sum of money from his friend Sid. They work out that to cover the amount of money Arlo owes Sid, he should pay Sid $235 per month for one year. However, Arlo agrees to pay an extra 25% on top of what he already owes Sid for taking so long to pay him back. How much money in total does Arlo end up giving Sid at the end of the one-year period?

A. $4875 **B.** $3981.25 **C.** $3525 **D.** $3102 **E.** $2878.75

Adapted from VCAA 2014 Exam 1 Business-related mathematics Q4

Data analysis *Year 11 content*

17. The following back-to-back stem plot displays the *weight* (kg) of 38 cavoodles and their *size*, toy ($n = 17$) or mini ($n = 21$).

Key: 9 | 3 = 9.3

***weight* (kg)**

toy cavoodle **mini cavoodle**

				4					
			9	5					
		9	4	6					
	5	4	2	7					
	6	3	1	8	8				
5	3	2	2	9	3	5	7		
		6	4	10	1	4	9		
			3	11	2	3	8	8	8
			5	12	1	1	4		
				13	5	6	6	8	
				14	1	2			

a. Which variable, *weight* or *size*, is a categorical variable? (1 MARK)

b. What is the modal weight of the mini cavoodles? (1 MARK)

c. Find the values of a and b in the following table. (2 MARKS)

	***weight* (kg)**				
size	*minimum*	Q_1	*median*	Q_3	*maximum*
toy cavoodle	5.9	7.3	8.6	9.95	a
mini cavoodle	8.8	10.25	b	13.55	14.2

Adapted from VCAA 2017 Exam 2 Data analysis Q2

Back to contents

3

CHAPTER 3

Investigating and modelling linear associations

LESSONS

KEY KNOWLEDGE

- 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.

Image: ChristianChan/Shutterstock.com

Back to contents

3A Fitting a least squares regression line

STUDY DESIGN DOT POINTS

- 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 $a = \overline{y} - b\overline{x}$
- 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

3A 3B 3C 3D 3E

KEY SKILLS

During this lesson, you will be:

- using technology to determine the least squares regression equation
- determining the least squares regression equation from a graph
- calculating the least squares regression equation from summary statistics
- sketching a least squares regression line from its equation.

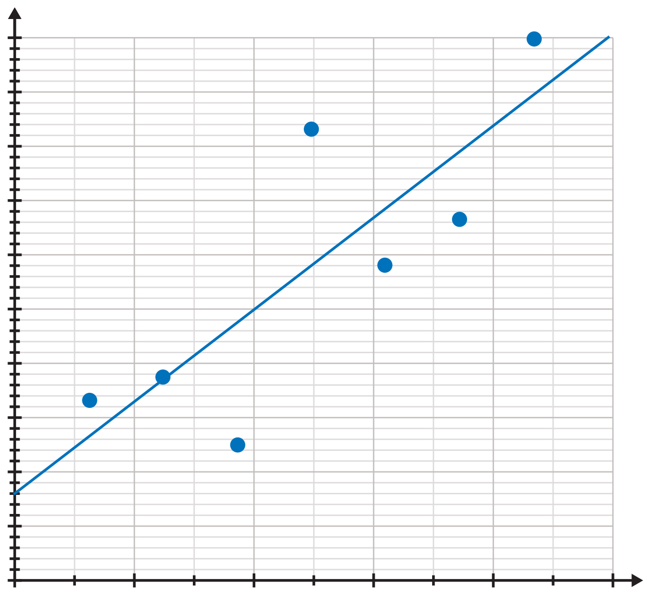
KEY TERMS

- Residual
- Least squares regression line

When bivariate numerical data is represented using scatterplots, there is often an underlying linear trend that can be observed. In these cases, it is useful to show this trend by fitting a straight line onto the scatterplot. While the least squares line will rarely be a perfect fit for the data, it can be useful for making inferences and predictions later on.

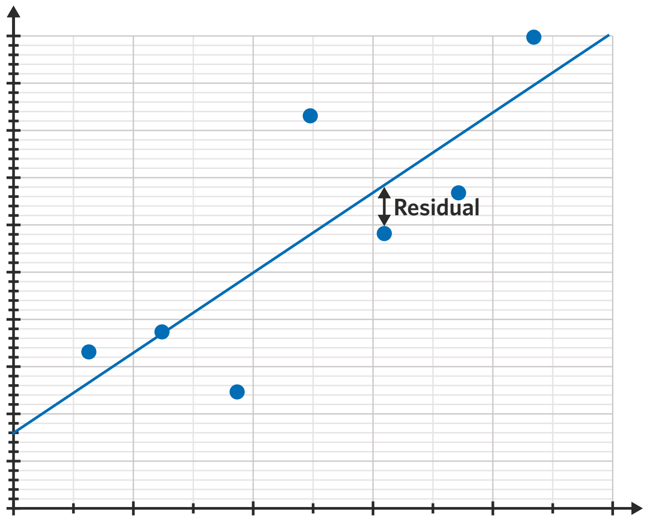
Using technology to determine the least squares regression equation

To fit a line to a scatterplot representing a set of data, a visual approach is often taken. Through this process, the aim is to construct the line to be as close as possible to all of the data points collectively, and to represent the direction of the scatterplot.

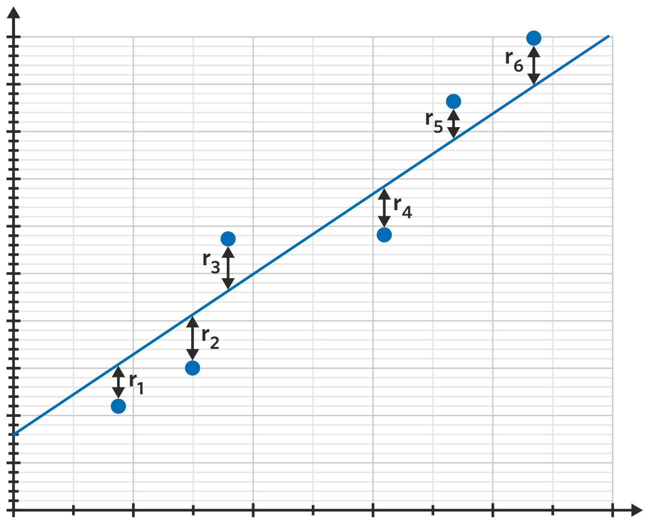


Back to contents

By hand, a reasonable line can usually be achieved. However, there is a mathematical process by which this line can be optimised. This process involves the minimisation of residual values, or residuals. A **residual** is the vertical distance between the straight line and any given point of the scatterplot, as shown.



The **least squares regression line** is the line which creates the minimum sum of the squares of the residuals. It is highly accurate because it can be determined mathematically rather than by eye. Visually, this will represent the underlying trend of the corresponding scatterplot.



The equation of the least squares regression line is generally in the form $y = a + bx$, where b represents the slope of the line, and a represents the y-intercept. This line is used to show the general trend of numerical bivariate data. When using a least squares regression line, some assumptions about the data are made:

- The data is numerical
- The relationship between the variables is linear
- There are no clear outliers present

A calculator can be used to determine the least squares regression equation.

Worked example 1

The following table gives the heights, in cm, of fathers and their sons when they were the same age.

height of father **(cm)**	182	187	174	177	162	171	188	165	170	185
height of son **(cm)**	184	179	176	168	161	173	187	163	170	181

Determine the equation of the least squares regression line that will allow the *height of son* to be predicted from the *height of father*. Give values correct to three significant figures.

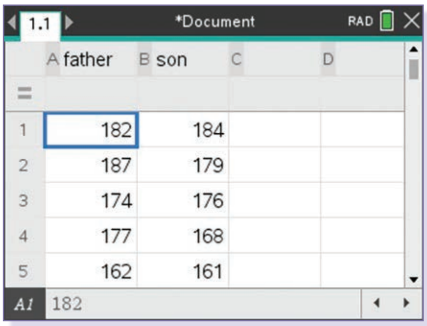
Continues →

Back to contents

Explanation – Method 1: TI-Nspire

Step 1: From the home screen, select '1: New' → '4: Add Lists & Spreadsheet'.

Step 2: Name a list 'father' and another list 'son' and enter the data as shown.



	father	son
1	182	184
2	187	179
3	174	176
4	177	168
5	162	161

Step 3: Identify the explanatory and response variables.

As *height of father* is being used to predict *height of son*, *height of father* is the explanatory variable.

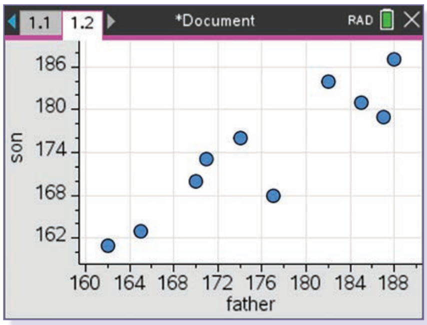
EV: *height of father*

RV: *height of son*

Step 4: Press ctrl + doc▸ and select '5: Add Data & Statistics'.

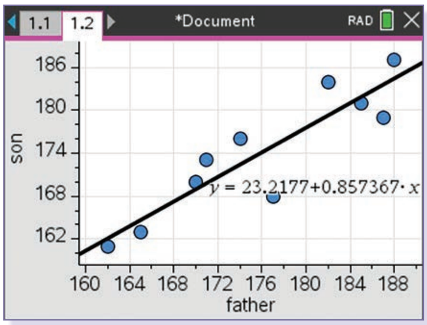
Step 5: Add the variables on each axis using the 'Click to add variable' function.

The *RV* will be positioned on the vertical axis and the *EV* will be positioned on the horizontal axis.



Step 6: Press menu. Select → '4: Analyse' → '6: Regression' →'2: Show Linear (a+bx)' to plot the least squares regression line.

The least squares regression line and its equation in the form $y = a + bx$ will appear.

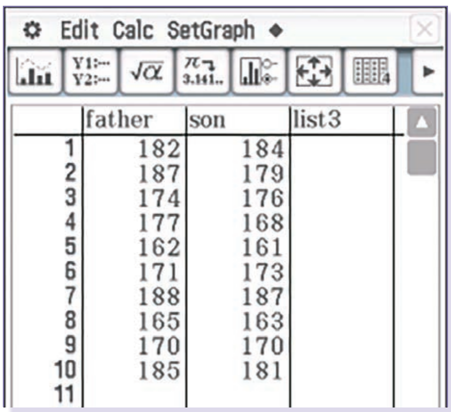


Step 7: Rewrite the equation in terms of the variables in the question and round as specified.

Explanation – Method 2: Casio ClassPad

Step 1: From the main menu, tap Statistics.

Step 2: Name a list 'father' and another list 'son' and enter the data as shown.



	father	son	list3
1	182	184	
2	187	179	
3	174	176	
4	177	168	
5	162	161	
6	171	173	
7	188	187	
8	165	163	
9	170	170	
10	185	181	
11			

Step 3: Identify the explanatory and response variables.

As *height of father* is being used to predict *height of son*, *height of father* is the explanatory variable.

EV: *height of father*

RV: *height of son*

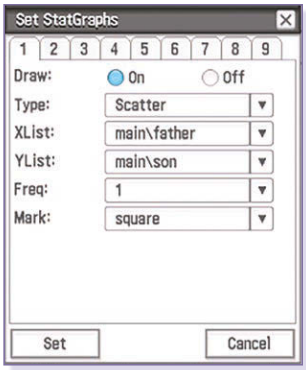
Continues →

Back to contents

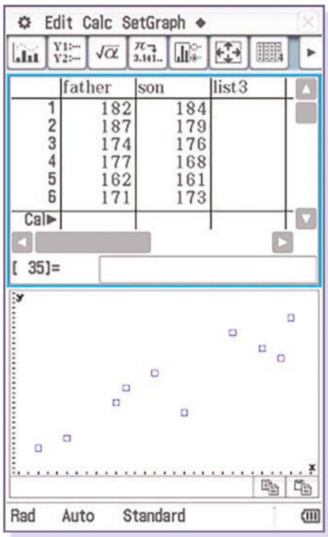
Step 4: Configure the settings of the graph by tapping in the icon bar.

Create a scatterplot by selecting 'Type' as 'Scatter'.

Specify the data set by changing 'XList:' to 'main\father' and 'YList:' to 'main\son'.



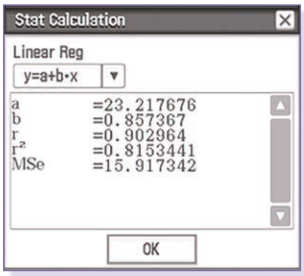
Step 5: Tap 'Set' to confirm and then to plot the scatterplot.



Step 6: Fit a least squares regression line to the scatterplot by tapping 'Calc' → 'Regression' → 'Linear Reg'. Specify the data set by changing 'XList:' to 'main\father' and 'YList:' to 'main\son'.

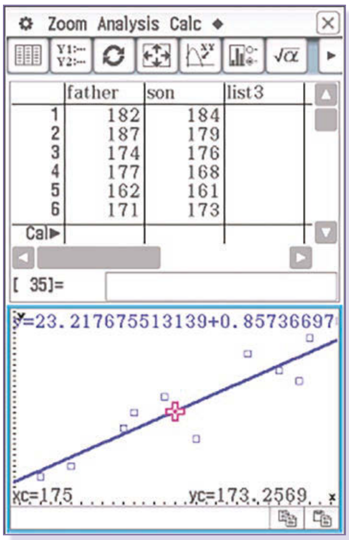
Tap 'OK' to confirm.

Step 7: The Stat Calculation window shows the values of a and b which will be used to write the least squares regression line.



Note: If $y = a \bullet x + b$ is selected, the a and b values need to be switched when writing the equation in $y = a + bx$ form.

Step 8: To visualise the data, tap 'OK', and the regression line will be generated on the scatterplot.



Step 9: Rewrite the equation in terms of the variables in the question and round as specified.

Answer – Method 1 and 2

height of son $= 23.2 + 0.857 \times$ *height of father*

Back to contents

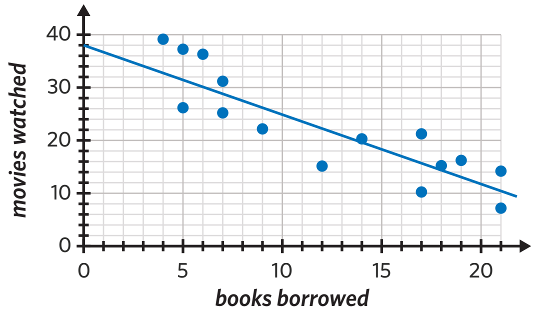
Determining the least squares regression equation from a graph

When the least squares regression line is already sketched on a scatterplot, its equation $y = a + bx$ can be found by determining the y-intercept (a) and the slope of the line (b).

Worked example 2

Consider the scatterplot shown.

From the scatterplot, estimate the equation of the least squares regression line. Give values correct to 3 significant figures.



Explanation

Step 1: Determine the explanatory and response variable.

On a scatterplot, the explanatory variable is positioned on the horizontal axis and the response variable on the vertical axis.

EV: *books borrowed*

RV: *movies watched*

Step 2: Write the equation in terms of the variables in the question.

movies watched $= a + b \times$ *books borrowed*

Step 3: Determine the vertical axis intercept of the line, a.

The point at which the line intersects the vertical axis, *movies watched*, is (0, 38).

$a = 38$

Step 4: Determine the slope of the line, b.

The slope of the line can be determined using two points that the line clearly passes through.

Two points that can be used are (0, 38) and (21, 10).

The formula for the gradient of a straight line between two points is $b = \frac{rise}{run} = \frac{y_2 - y_1}{x_2 - x_1}$.

$$b = \frac{10 - 38}{21 - 0} = \frac{-28}{21}$$

$$= -1.333...$$

Step 5: Write the equation using the values for a and b, rounded to 3 significant figures.

Answer

movies watched $= 38.0 - 1.33 \times$ *books borrowed*

Calculating the least squares regression equation from summary statistics

For a least squares line of the form $y = a + bx$, a and b can also be found using the formulas

- $b = r \times \frac{s_y}{s_x}$
- $a = \bar{y} - b\bar{x}$

where:

- a is the y-intercept
- b is the gradient
- r is Pearson's correlation coefficient
- $\bar{x}$ is the mean of the explanatory variable (x)
- $\bar{y}$ is the mean of the response variable (y)
- s_x is the standard deviation of the explanatory variable (x)
- s_y is the standard deviation of the response variable (y)

Back to contents

Worked example 3

Consider the following summary statistics for a set of bivariate data involving the variables x and y.

$r = 0.845, \quad \bar{x} = 11.0, \quad \bar{y} = 29.2, \quad s_x = 6.06, \quad s_y = 16.8$

Find the equation of the least squares regression line that allows y to be predicted from x. Give all values correct to three significant figures.

Explanation

Step 1: Calculate the slope, b in the equation $y = a + bx$, using the formula $b = r \times \frac{s_y}{s_x}$

$b = 0.845 \times \frac{16.8}{6.06}$

$= 2.342...$

Step 2: Calculate the y-intercept, a using the formula

$a = \bar{y} - b\bar{x}$

$a = 29.2 - 2.342... \times 11.0$

$= 3.431...$

Step 3: Write the equation of the least squares regression line, correct to three significant figures.

Answer

$y = 3.43 + 2.34x$

Sketching a least squares regression line from its equation

When the equation of a least squares regression line is given, it can be sketched onto a scatterplot using the two-point method.

Worked example 4

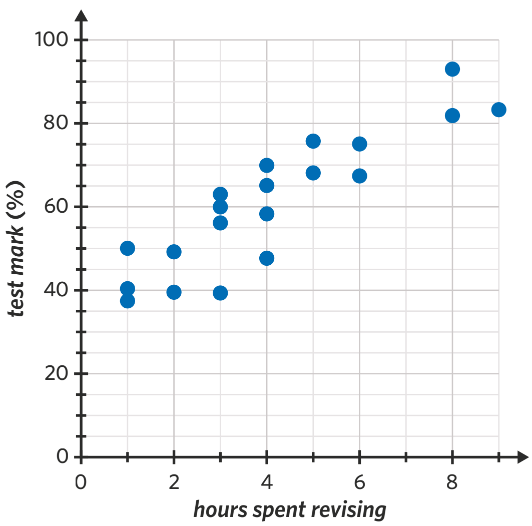
20 students in Ms Benton's geography class took their end of year test. Before the test, each member of the class was surveyed to find out how many hours they spent revising for the test. Each student's *hours spent revising* and *test mark* (%) are depicted in the following scatterplot.

Ms Benton noticed a strong linear association between the two variables.

The least squares regression line for this data is approximated by the following equation:

test mark $= 36.0 + 6.11 \times$ *hours spent revising*

From the equation, sketch the least squares line for the scatterplot.



Explanation

Step 1: Decide on two horizontal axis values to substitute into the equation.

It is best to use two values for *hours spent revising* that are on either end of the visible plane for accuracy.

The best values to use are 0 and 9.

Step 2: Substitute both values for *hours spent revising* into the equation to find their corresponding *test mark* values.

hours spent revising $= 0$

test mark $= 36.0 + 6.1 \times 0 = 36$

hours spent revising $= 9$

test mark $= 36.0 + 6.1 \times 9 = 90.9$

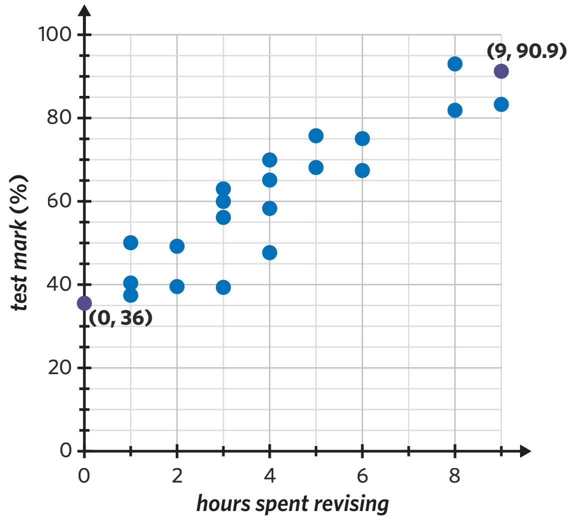
Continues →

Back to contents

Step 3: Determine the points to plot on the graph.

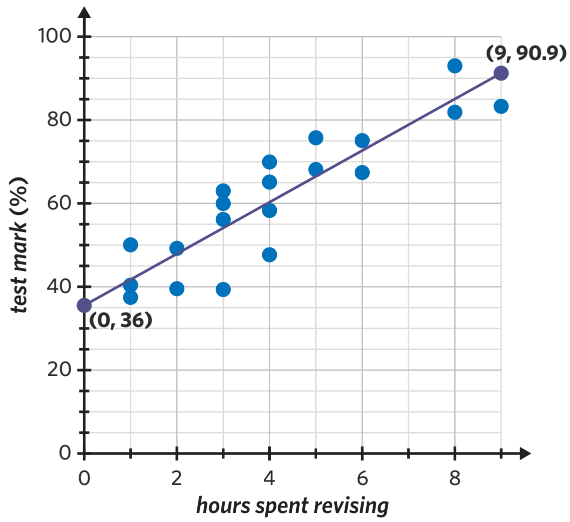
Point 1: (0, 36)

Point 2: (9, 90.9)



Step 4: Draw a straight line passing through the two points.

Answer



Exam question breakdown

VCAA 2017 Exam 1 Data analysis Q8

The following 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.

The equation of the least squares line is closest to

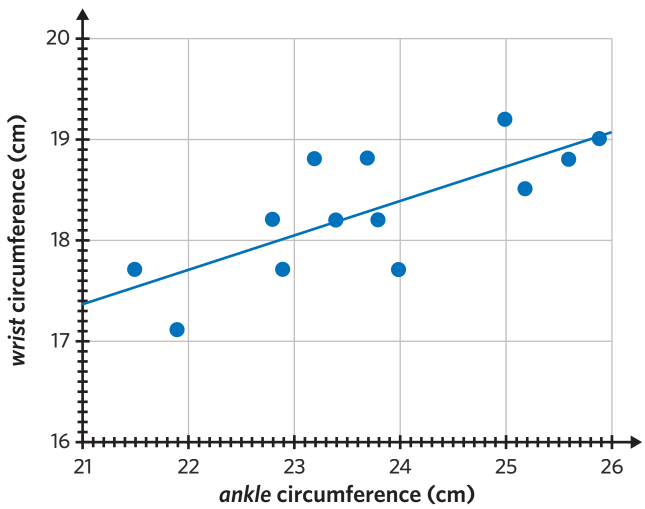
A. *ankle* = 10.2 + 0.342 × *wrist*

B. *wrist* = 10.2 + 0.342 × *ankle*

C. *ankle* = 17.4 + 0.342 × *wrist*

D. *wrist* = 17.4 + 0.342 × *ankle*

E. *wrist* = 17.4 + 0.731 × *ankle*



Continues →

Back to contents

Explanation

Step 1: Identify the explanatory and response variable.

The question specifies that *ankle* circumference is the explanatory variable, which means *wrist* circumference is the response variable.

EV: *ankle*

RV: *wrist*

Step 2: Write the equation in terms of the variables in the question.

$wrist = a + b \times ankle$

Step 3: Determine the vertical axis intercept, a, by inspecting the graph.

At first, the intercept value appears to be 17.4. However, the horizontal axis begins at 21, not 0. Therefore, the actual intercept occurs further to the left, and will be a lower value.

Observing the available options, since a is less than 17.4, a must equal 10.2.

Step 4: Determine the slope, b, by inspecting the graph.

As there are only two possible b values, an approximation of the slope will be sufficient.

The least squares regression line shows that *wrist* circumference increases by approximately 1 cm for every 3 cm increase in *ankle* circumference. $\frac{rise}{run} \approx \frac{1}{3}$

$\frac{1}{3}$ is closest to 0.342 so $b = 0.342$

Step 5: Write the final equation.

$wrist = 10.2 + 0.342 \times ankle$

Answer

B

43% of students answered this question correctly.

39% of students incorrectly answered D, due to the assumption that the horizontal axis begins at 0. This led to an incorrectly obtained vertical axis intercept of 17.4.

3A Questions

Using technology to determine the least squares regression equation

1. Consider the following table.

number of beaches	1	3	4	4	4	5	5	7	7	8
number of surf shops	2	3	4	18	5	6	6	7	9	10

A least squares regression line would not be appropriate for this data because

A. the data is categorical.

B. there are not enough data points.

C. both variables are numerical.

D. there is a clear outlier.

---

2. Consider the following table.

number of shops	8	12	15	19	20	21	25	27	30	34
population (000's)	4.7	6.9	8.4	7.8	10.8	8.1	13.0	11.6	19.0	16.2

A least squares line is fitted to the data, with the aim of predicting *population* from *number of shops*. Which of the following statements is true?

A. $a = 0.91$ and $b = 0.82$

B. $a = 0.49$ and $b = 0.24$

C. $a = 0.24$ and $b = 0.49$

D. $a = 0.82$ and $b = 0.91$

3. A group of students were asked how many hours they had slept the night before a General Mathematics exam.

Their *number of hours slept* and *result* (%) are shown in the table.

number of hours slept	7.2	7.8	8.3	5.3	6	9	2.3	6.8	7.7	8	8.6	9.2
result **(%)**	88	79	83	67	70	92	56	74	63	84	98	85

A least squares regression line has been fitted to the data to predict *result*, from *number of hours slept*.

Write the equation of the regression line, giving values correct to three decimal places.

Determining the least squares regression equation from a graph

4. The least squares line fitted to the scatterplot shown is of the form $y = a + bx$.

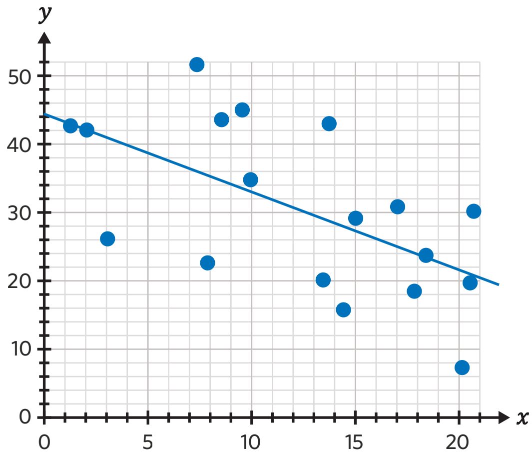
Which of the following statements is false?

A. The value of a is positive.

B. The value of b is positive.

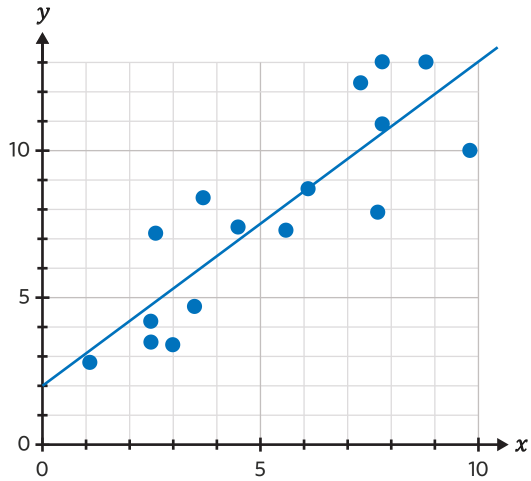
C. x is the explanatory variable.

D. y is the response variable.

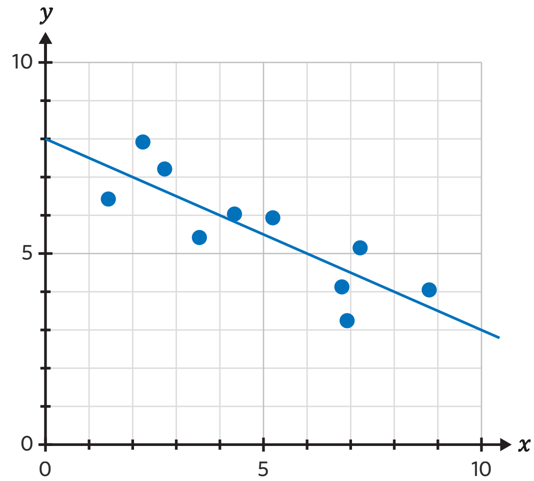


5. Estimate the equation of the least squares regression line for the following scatterplots.

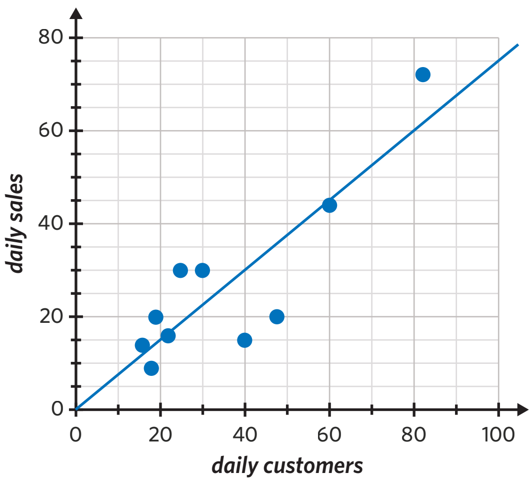
a.



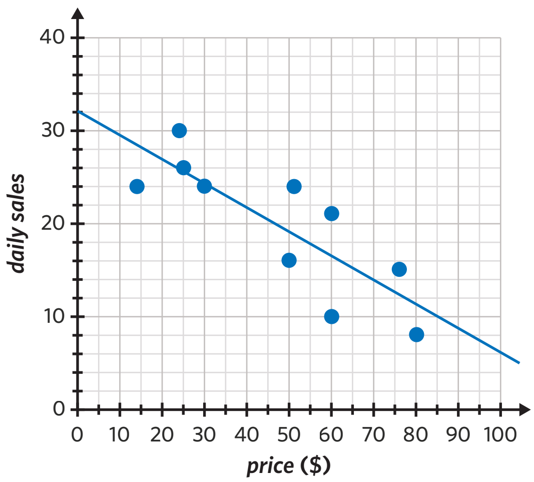
b.



c.



d.



Back to contents

Calculating the least squares regression equation from summary statistics

6. Consider the following summary statistics for a set of bivariate data involving the variables x and y.

$r = -0.96,\quad \bar{x} = 15.3,\quad \bar{y} = 43.8,\quad s_x = 12.2,\quad s_y = 13.0$

The equation of the least squares regression line that allows y to be predicted from x, with values given correct to 3 significant figures, is closest to

A. $y = 59.5 + 1.02x$

B. $y = 1.02 + 59.5x$

C. $y = 59.5 - 1.02x$

D. $y = 1.02 - 59.5x$

---

7. Given the following information:

$a = 31.4,\quad b = 1.7,\quad r = 0.96,\quad s_x = 9.2,\quad \bar{x} = 15.1$

Calculate s_y and $\bar{y}$ correct to one decimal place.

---

8. Joe the farmer wants to investigate the relationship between *time* (days) and the *height* (cm) of his growing wheat crops. Joe hires a data analyst, who is able to provide certain statistics based on the data gathered from a patch of growing wheat crops.

$\bar{x} = 55.0,\quad \bar{y} = 63.8,\quad s_x = 30.3,\quad s_y = 37.4,\quad r = 0.99$

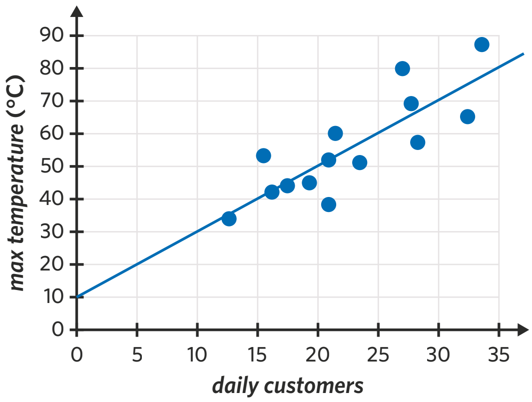
Joe wants to model a least squares regression line to his data, so that he can predict the *height* of any given wheat plant based on the amount of *time* that has passed. Determine the equation of this line, giving values correct to one decimal place.

Sketching a least squares regression line from its equation

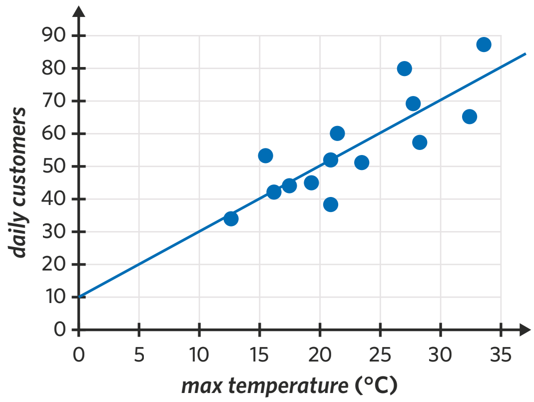
9. A least squares line is fitted to data with the aim of predicting the number of *daily customers* of an ice-cream shop from the *max temperature* (°C). The equation of the line is *daily customers* = 10 + 2 × *max temperature*.

A possible scatterplot that this equation could be obtained from is

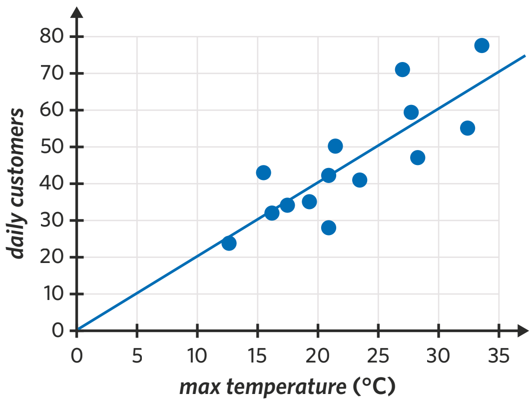
A.



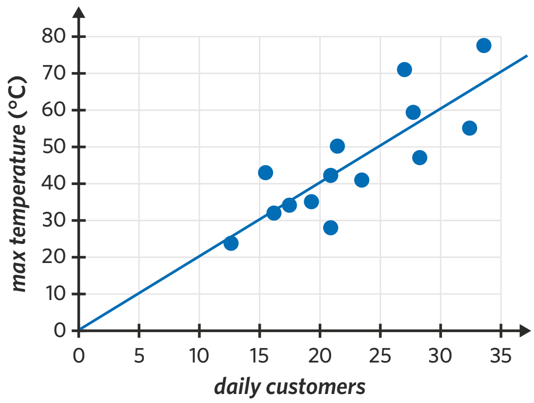
B.



C.



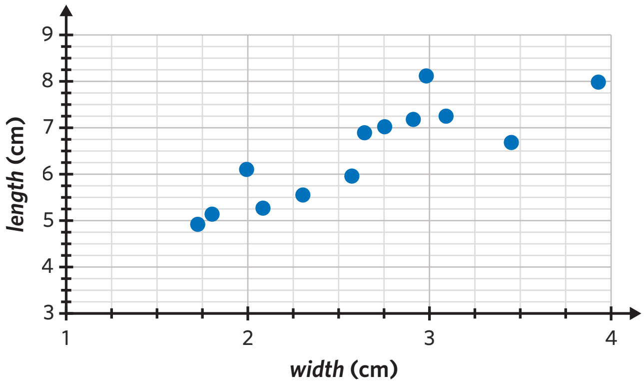
D.



10. The *width* (cm) and *length* (cm) of 13 insects were plotted on a scatterplot.

The least squares regression line for this data is given by the equation $length = 2.85 + 1.37 \times width$

From this equation, plot the line on the scatterplot.



Joining it all together

11. Which of the following statements regarding the least squares regression line is false?

A. A least squares regression line minimises the sum of the squares of the residuals.

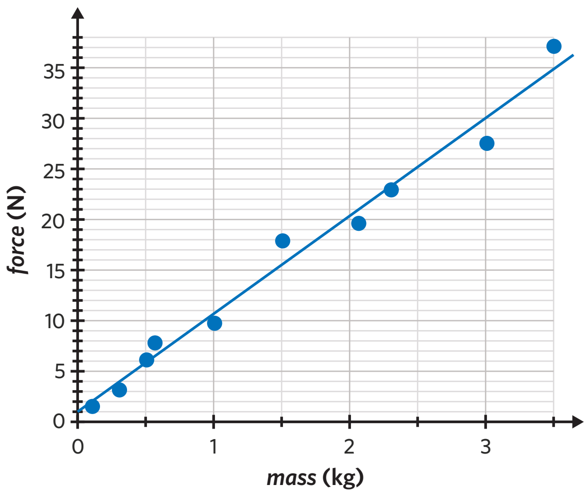
B. A least squares regression line is not appropriate if there is a clear outlier.

C. A least squares regression line is generally written in the form $y = a + bx$.

D. A least squares regression line can be fitted to the variables *day of the week* and *hours spent watching TV*.

---

12. During a high school physics experiment, the *mass* (kg) and gravitational *force* (N) acting on 10 different objects were recorded and displayed using a scatterplot.



a. Estimate the equation of the least squares regression line from the graph, giving values correct to 1 decimal place.

b. Use the corresponding data provided to determine the least squares equation, giving values accurate to 2 decimal places.

mass (kg)	0.1	0.3	0.5	0.56	1.0	1.5	2.06	2.3	3.0	3.5
force (N)	1.5	3.1	6.1	7.8	9.7	17.9	19.6	22.9	27.5	37.1

c. Using the values for *a* and *b* from the equation obtained in part **b**, and knowing that $\bar{x} = 1.482$, show that $\bar{y} = 15.32$.

Back to contents

Exam practice

13. The statistical analysis of a set of bivariate data involving variables x and y resulted in the information displayed in the table shown.

mean	$\bar{x} = 27.8$	$\bar{y} = 33.4$
standard deviation	$s_x = 2.33$	$s_y = 3.24$
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

VCAA 2018 Exam 1 Data analysis Q13

49% of students answered this question correctly.

---

14. The following table shows the *weight* in kilograms, and the *height*, in centimetres, of 10 adults.

***weight* (kg)**	59	67	69	84	64	74	76	56	58	66
***height* (cm)**	173	180	184	195	173	180	192	169	164	180

A least squares line is fitted to the data.

The least squares line enables an adult's *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

A. 3 **B.** 4 **C.** 5
D. 6 **E.** 7

VCAA 2021 Exam 1 Data analysis Q11

40% of students answered this question correctly.

---

15. 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 following 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. 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.

Write the values of the intercept and slope of this least squares line in the appropriate boxes provided.

Round your answers to one decimal place.

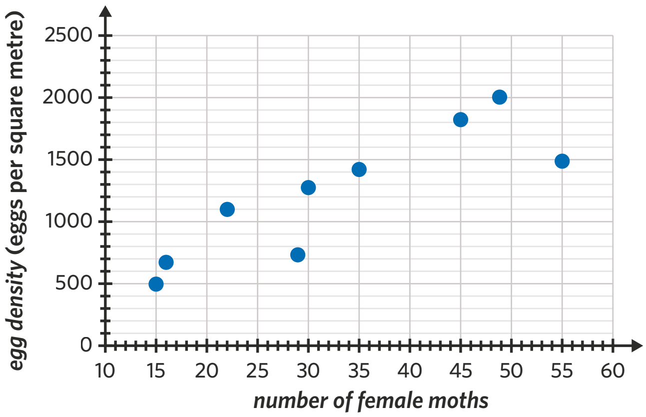
egg density = ☐ + ☐ × *number of male moths* (2 MARKS)

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.

The equation of the least squares line is
egg density = 191 + 31.3 × *number of female moths*

Draw the graph of this least squares line on the scatterplot. (1 MARK)



VCAA 2017 Exam 2 Data analysis Q3a, bi

Part **a**: The average mark on this question was **1.5**.
Part **b**: **26%** of students answered this question correctly.

Back to contents

Questions from multiple lessons

Data analysis *Year 11 content*

16. The marks obtained by 59 students in a maths test are shown in the following histogram.

The median mark for the test was

A. greater than or equal to 15 but less than 20.

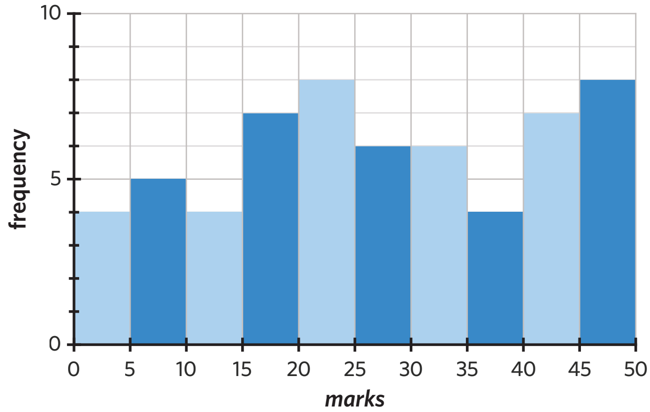
B. greater than or equal to 20 but less than 25.

C. greater than or equal to 25 but less than 30.

D. greater than or equal to 30 but less than 35.

E. greater than or equal to 35 but less than 40.

Adapted from VCAA 2017NH Exam 1 Data analysis Q3



Recursion and financial modelling *Year 11 content*

17. Victoria wants to buy a jumbo trampoline from her friend, Shaun. She agrees to pay Shaun back over several months. In the first month, she will pay Shaun $300, and in each subsequent month she will pay $50 less than she did the month before.

Let p_n be the amount of money that Victoria pays Shaun during the n^{th} month.

A recurrence relation that can be used to model this situation for $1 \leq n \leq 6$ is

A. $p_1 = 300, \quad p_{n+1} = 0.83p_n$

B. $p_1 = 250, \quad p_{n+1} = p_n - 50$

C. $p_0 = 300, \quad p_{n+1} = p_n - 50$

D. $p_0 = 250, \quad p_{n+1} = 300 - 50p_n$

E. $p_1 = 300, \quad p_{n+1} = p_n - 50$

Adapted from VCAA 2014 Exam 1 Number patterns Q4

Data analysis

18. The weight of ducks in a large population is approximately normally distributed with a mean of 1.2 kg and a standard deviation of 120 g.

Which one of the following statements relating to this population of ducks is **not** true?

A. Approximately half of the ducks will weigh greater than 1.2 kg.

B. Approximately 16% of the ducks will weigh less than 1.08 kg.

C. Approximately 81.5% of the ducks will weigh between 1.08 kg and 1.44 kg.

D. Approximately 13.5% of the ducks will weigh between 1.32 kg and 1.44 kg.

E. No duck will weigh less than 1 kg.

Adapted from VCAA 2017NH Exam 1 Data analysis Q8

Back to contents

3B Interpreting a least squares regression line

STUDY DESIGN DOT POINT

- 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

3A 3B 3C 3D 3E

KEY SKILLS

During this lesson, you will be:

- interpreting a least squares regression line
- making predictions using a least squares regression line.

KEY TERMS

- y-intercept
- Slope
- Interpolation
- Extrapolation

A least squares regression line is often used to make predictions. It is therefore useful to understand and be able to interpret the least squares regression line and its equation. When making predictions, it is important to be aware of the reliability of these predictions.

Interpreting a least squares regression line

The **y-intercept** is the approximate value of the response variable (y) when the explanatory variable (x) is equal to 0. When the value of the explanatory variable cannot equal 0, this value has no useful meaning.

The **slope** is the average change in the response variable for every one-unit increase in the explanatory variable.

For a least squares regression line with the rule: $y = a + bx$,

- a is the y-intercept, and
- b is the slope

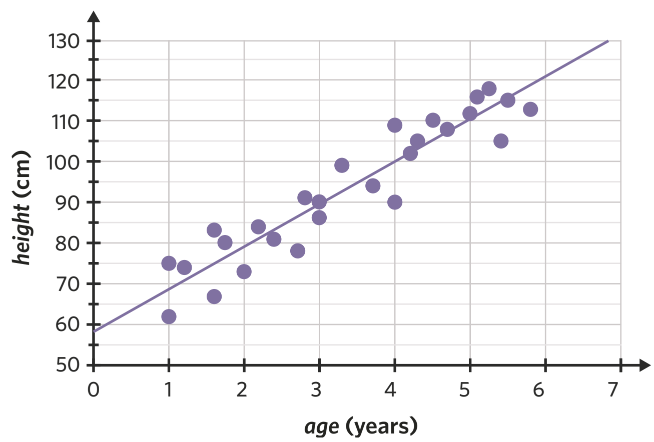
Worked example 1

The heights of a group of children aged between one and six years old were measured.

The scatterplot and regression equation show the relationship between *height* (cm) and *age* (years).

Based on the regression equation,

$height = 58.22 + 10.46 \times age$



Continues →

Back to contents

a. How much on average does a child grow each year, correct to two decimal places?

Explanation

Step 1: Identify which part of the equation requires interpretation.

The question is asking about the change in the response variable (*height*) for each one-unit (one year) increase in the explanatory variable (*age*). This is the slope.

Step 2: Determine the slope.

In the rule $y = a + bx$, the slope is represented by b.

$b = 10.46$

Step 3: Interpret the slope.

For every one year increase, there will be an increase in *height* of 10.46 cm.

Answer

On average, a child grows 10.46 cm each year.

---

b. Interpret the y-intercept in this regression equation.

Explanation

Step 1: Identify the y-intercept.

In the rule $y = a + bx$, a is the y-intercept.

$a = 58.22$

Step 2: Interpret the y-intercept.

The y-intercept is the approximate *height* when *age* equals 0.

Answer

On average, the height of children when they are born is 58.22 cm.

Making predictions using a least squares regression line

Least squares regression lines can be used to predict the value of the response variable from the explanatory variable or vice versa. **Interpolation** involves making predictions that are within the range of the data set, and are the most reliable predictions. **Extrapolation** involves making predictions that are outside the range of the data set. These predictions have limited reliability because an assumption is made that the relationship between the two variables continues outside the range of the data set. However, it is unknown whether this relationship will in fact continue.

Suppose the heights of children aged between 5 and 15 years old are measured and the height of a 7 year old is then predicted. This is interpolation. If the height of a 3 year old is predicted, this is extrapolation since it is outside the range of the data set.

Predictions can be made by eye using a graph of the regression line on a scatterplot.

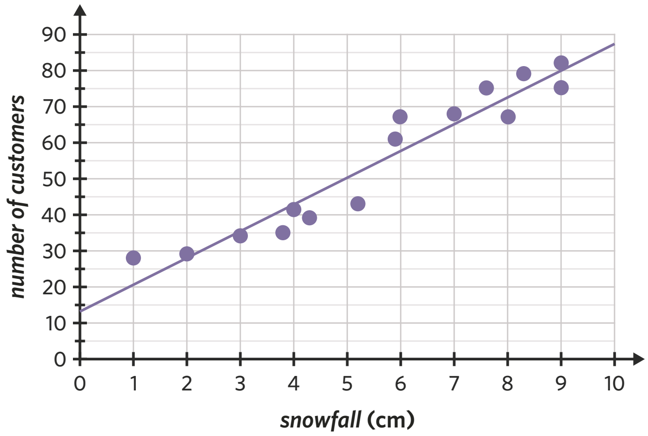
See worked example 2

Predictions can also be made by substituting the value of the explanatory variable or response variable into the regression equation and solving. These predictions are more accurate than ‘by eye’ predictions made from the regression line on a scatterplot.

See worked example 3

Worked example 2

A ski hire shop has recorded the daily *snowfall* (cm) and the corresponding *number of customers* that hire from them. The relationship between *number of customers* and *snowfall* is displayed on the scatterplot.



a. What is the predicted *number of customers* when there is 5 cm of *snowfall*? Is this interpolation or extrapolation?

Explanation

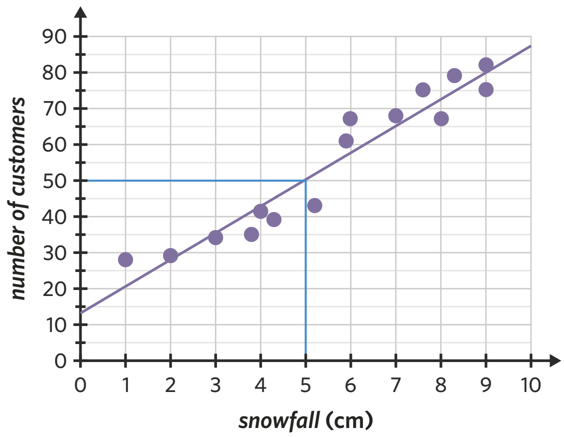
Step 1: Identify whether the prediction is interpolation or extrapolation.

The data set shows the *number of customers* for varying levels of *snowfall* ranging from 1 to 9 cm.

5 cm of *snowfall* is within this range. This is interpolation.

Step 2: Predict the response variable.

Use the regression line to predict the *number of customers* when *snowfall* $= 5$.



Answer

50 customers

Interpolation

b. How many customers are expected to hire from the ski hire shop when there is no snowfall?

Explanation

Determine the *number of customers* when *snowfall* $= 0$. This is the y-intercept.

Note that 0 cm of *snowfall* lies outside the range of the data set, so this prediction involves extrapolating the data.

Each interval on the vertical axis represents 5 customers. The y-intercept is just below the third interval, which is approximately 14.

Answer

14 customers

Worked example 3

The following table shows data collected on the *distance from the CBD* of Melbourne apartments and their *weekly rent*.

***distance from the CBD* (km)**	3.5	7	2.5	4	10	9.5	6.5	1.5	3	2.75	1	8.25	5.5	11.5	4.75
***weekly rent* ($)**	350	275	325	290	195	230	245	425	335	375	440	220	310	170	345

The least squares regression equation from the data is:

$\textit{weekly rent} = 427.92 - 23.25 \times \textit{distance from the CBD}$

a. If an apartment has a *weekly rent* of $250, predict its *distance from the CBD*, correct to two decimal places.

Explanation

Step 1: Substitute *weekly rent* = 250 into the regression equation.

$250 = 427.92 - 23.25 \times \textit{distance from the CBD}$

Step 2: Solve the equation.

In this question, the value of the response variable is given. The equation needs to be solved for the explanatory variable.

$250 = 427.92 - 23.25 \times \textit{distance from the CBD}$

$-177.92 = -23.25 \times \textit{distance from the CBD}$

$\textit{distance from the CBD} = 7.6524...$

Answer

7.65 km

b. Predict the *weekly rent* of an apartment that is 12 km from the CBD, correct to the nearest dollar.

Explanation

Step 1: Substitute *distance from the CBD* = 12 into the regression equation.

$\textit{weekly rent} = 427.92 - 23.25 \times 12$

Step 2: Solve the equation.

$$\begin{aligned}\textit{weekly rent} &= 427.92 - 279\\ &= 148.92\\ &\approx 149\end{aligned}$$

Answer

$149

c. Is the prediction from part **b** reliable? Justify your answer.

Explanation

Identify whether the prediction is an interpolation or extrapolation.

The range of the data set is apartments between 1 and 11.5 km from the CBD.

12 km falls outside the range of the data set, meaning it is an extrapolation.

Answer

No, it may not be reliable because the prediction is an example of extrapolation. The regression equation is not always reliable for values outside the data set.

Back to contents

Exam question breakdown

VCAA 2018 Exam 1 Data analysis Q10

In a study of the association between a person's *height*, in centimetres, and *body surface area*, in square metres, the following least squares line was obtained.

$body\ surface\ area = -1.1 + 0.019 \times height$

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 person's *body surface area*, in square metres, can be determined by adding 1.1 cm to their *height*.
E. A person's *height*, in centimetres, can be determined by subtracting 1.1 from their *body surface area*, in square metres.

Explanation

To solve this question, check whether each option is true or false.

A: This is false. 0.019 is the value of the slope which indicates the average change in the response variable for every one-unit increase in the explanatory variable. The response variable is *body surface area* and the explanatory variable is *height*. ✗

B: This is true. The slope indicates the average change in *body surface area* for every one-unit (1 cm) increase in *height*. ✓

C: This is false. There is not enough information to determine the correlation coefficient. ✗

Answer

B

D: This is false. A person's *body surface area* can be predicted by multiplying their *height* by 0.019 and subtracting 1.1. ✗

E: This is false. A person's *height* can be predicted by adding 1.1 to their *body surface area* and dividing by 0.019. ✗

51% of students answered this question correctly.

20% of students incorrectly answered option A, as they misinterpreted the slope as the average change in the explanatory variable (*height*) for a one-unit increase in the response variable (*body surface area*) instead of the other way around.

3B Questions

Interpreting a least squares regression line

1. The following regression equation shows the relationship between the *number of bedrooms* and *price* (\$000's) of houses in Melbourne:

$price = 593.2 + 98.4 \times number\ of\ bedrooms$

By how much, on average, do house prices increase as the *number of bedrooms* increases?

A. \$98.40 **B.** \$593.20 **C.** \$98 400.00 **D.** \$593 200.00

2. The *price* (\$) of a second-hand car depends on its *age* (years).

An approximate relationship between the car's *price* and *age* is given in the following equation.

$price = 24\,000 - 3500 \times age$

a. What was the price of the car when it was brand new?

b. By how much does the car's value decrease each year?

3. A group of runners are training for the next Melbourne Marathon. During their training, they record the *time* and *distance* of each run. They then calculated the equation of the regression line that could predict the *time* (minutes) taken to complete a run from the *distance* (km).

$time = 0.32 + 5.61 \times distance$

a. How long on average does it take a runner to complete an additional kilometre?

b. Interpret the value of the *y*-intercept. Does this make sense?

Back to contents

Making predictions using a least squares regression line

4. The relationship between the daily *maximum temperature* (°C) and the daily *minimum temperature* (°C) is given by the following least squares regression line:

maximum temperature = 10.42 + 0.89 × *minimum temperature*

What is the expected maximum temperature on Tuesday if the minimum temperature is 18 °C?

A. 16.02 °C **B.** 25.29 °C **C.** 26.44 °C **D.** 28.42 °C

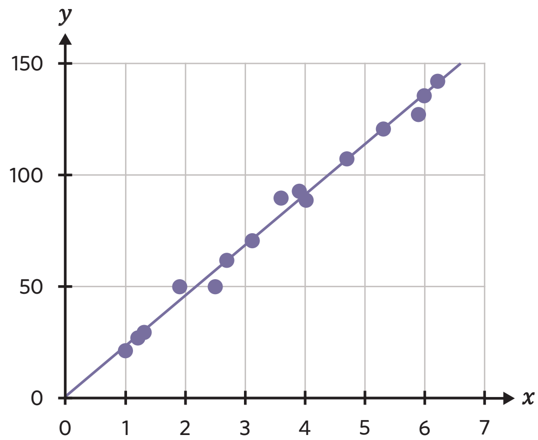
5. The scatterplot shown has been fitted with a regression equation:

$y = -1.7 + 23.2x$

Find the value of y when x equals

a. 2

b. 9

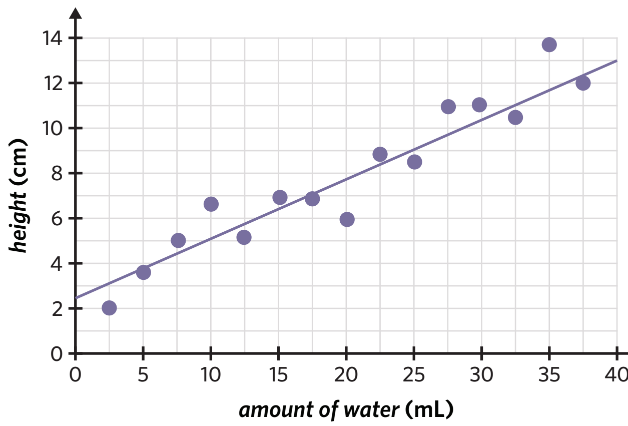


6. 15 seedlings were planted and given different amounts of water. The heights of the seedlings were then recorded after two weeks. The scatterplot shows the *amount of water* (mL) each seedling received when planted and the *height* (cm) after two weeks.
A regression line has been fitted to the data.

a. What is the expected amount of water that a seedling received initially if it is 5 cm tall after two weeks?

b. How tall is a seedling predicted to be after 2 weeks if it receives 40 mL of water initially?

c. Why might the prediction in part **b** not be reliable?



7. The manager of a swimming pool investigated the effect that the *temperature* has on *attendance* at his pool. He recorded the weather and the number of people that attended the pool for two weeks.

The results are displayed in the scatterplot.

The regression equation that allows *attendance* to be predicted from *temperature* (°C) is as follows.

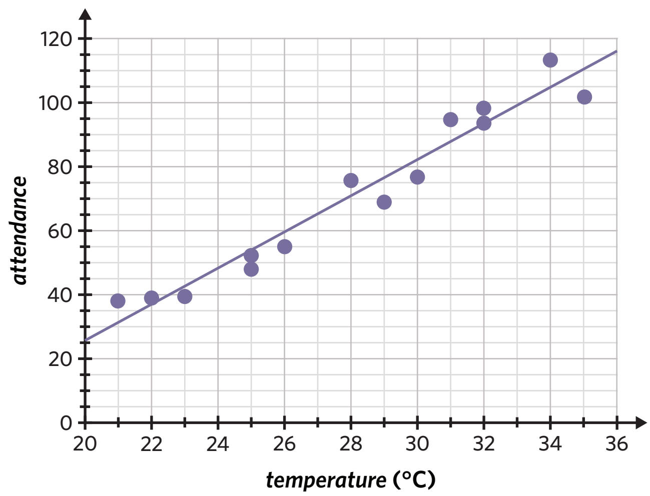
attendance = −88.53 + 5.7 × *temperature*

a. Use the regression equation to predict the number of people that will attend the pool if the *temperature* is 31 °C. Round to the nearest whole number.

b. Is the prediction from part **a** reliable? Explain briefly.

c. Use the regression equation to predict the number of people that will attend the pool if the *temperature* is 16 °C. Round to the nearest whole number.

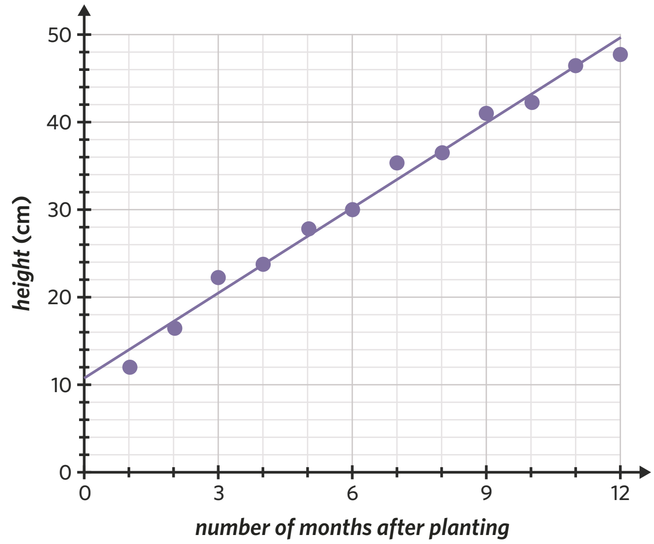
d. Is the prediction from part **c** reliable? Explain briefly.



Joining it all together

8. A group of gardeners have developed a special soil to make cacti grow faster. In order to test it, they planted a small cactus in their soil and measured its *height* each month for a year. The gardeners displayed their results in a scatterplot and calculated the equation of a regression line that could predict the *height* (cm) of the cactus from the *number of months after planting*.

height = 10.89 + 3.23 × *number of months after planting*



a. The gardeners forgot to measure the *height* of the cactus when they planted it into their soil.

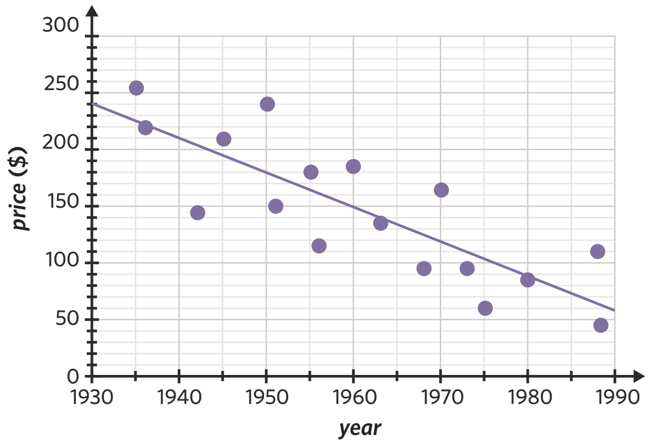
Use the regression equation to find its predicted *height* when it was planted. Give the value correct to two decimal places.

b. On average, how much does the cactus grow in *height* each month? Give the value correct to two decimal places.

c. Predict what the *height* of the cactus was five and a half months after being planted. Give the value correct to two decimal places.

---

9. Anna owns a vintage dress shop and is trying to create an efficient way of pricing her dresses by basing the selling price on the year they were made. She usually prices the older dresses higher because they are more valuable. She has recorded the *year* that some of her dresses were made and their corresponding *price* ($). The results are shown in the scatterplot provided.



The equation of the regression line that Anna calculated is given.

price = 6079.28 − 3.03 × *year*

a. Complete the following sentence:

On average, for each unit increase in *year*, the *price* of the dress ___________ by ___________.

b. Predict the *price* of a dress that was made in 1955, correct to the nearest dollar.

c. Predict the *price* of a dress that was made in 1982, correct to the nearest dollar.

d. Is the prediction from part **b** reliable? Explain briefly.

---

10. The *maximum daily temperature* (°C) and *precipitation* (mm) were recorded over a two-week period. The results are shown in the following table.

***maximum daily temperature* (°C)**	20	15	11	19	22	16	14	18	20	21	10	22	15	16
***precipitation* (mm)**	2.2	4.9	5.7	1.2	1.3	3.4	2.2	1.1	1.2	0.5	6.2	0.4	2.3	2.1

The regression equation that allows *precipitation* to be predicted from *maximum daily temperature* is:

precipitation = 9.77 − 0.43 × *maximum daily temperature*

a. Interpret the slope of the least squares regression equation.

b. What is the *maximum daily temperature* that would result in no precipitation? Round to two decimal places.

c. Predict the *precipitation*, correct to two decimal places, if the daily temperature reaches a maximum of 25 °C. Is this prediction reliable?

Back to contents

d. Which of the following predictions would be most reliable?

A. *precipitation* when *maximum daily temperature* is 25 °C.

B. *precipitation* when *minimum daily temperature* is 15 °C.

C. *maximum daily temperature* when *precipitation* is 7 mm.

D. *maximum daily temperature* when *precipitation* is 1 mm.

Exam practice

11. The scatterplot provided shows the atmospheric pressure, in hectopascals (hPa), at 3 pm (*pressure 3 pm*) plotted against the atmospheric pressure, in hectopascals, at 9 am (*pressure 9 am*) for 23 days in November 2017 at a particular weather station.

A least squares line has been fitted to the scatterplot as shown.

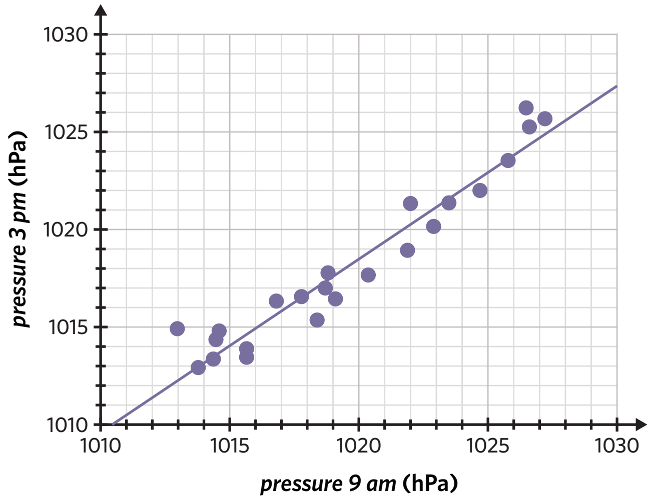
The equation of this line is

pressure 3 pm = 111.4 + 0.8894 × *pressure 9 am*

The equation of the least squares line is used to predict the atmospheric pressure at 3 pm when the atmospheric pressure at 9 am is 1025 hPa.

Is this prediction an example of extrapolation or interpolation? (1 MARK)

Adapted from VCAA 2019 Exam 2 Data analysis Q5c



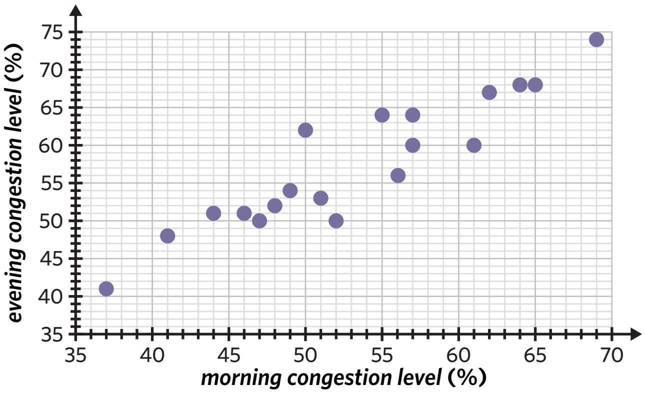
Data: Australian Government, Bureau of Meteorology, <www.bom.gov.au/>

87% of students answered this type of question correctly.

12. 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 shown.



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 × *morning congestion level*

Use the equation of the least squares line to predict the *evening congestion level* when the *morning congestion level* is 60%. (1 MARK)

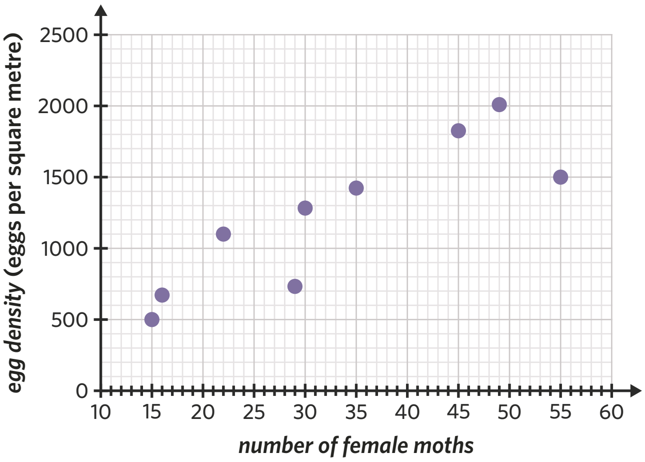
VCAA 2018 Exam 2 Data analysis Q2c

78% of students answered this question correctly.

Back to contents

13. The *number of female moths* caught in a trap set in a forest and the *egg density* (eggs per square metre) in the forest are examined.

A scatterplot of the data is shown.



A least squares regression line is fitted to the data. The equation of the least squares line is

egg density $= 191 + 31.3 \times$ *number of female moths*

Interpret the slope of the regression line in terms of the variables *egg density* and *number of female moths* caught in the trap. (1 MARK)

Adapted from VCAA 2017 Exam 2 Data analysis Q3bii

39% of students answered this type of question correctly.

---

14. The following table shows the yearly average traffic congestion levels in two cities, Melbourne and Sydney, during the period 2008 to 2016.

	***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 least squares line is used to model the trend in the time series plot for Sydney. The equation is

congestion level $= -2280 + 1.15 \times$ *year*

The equation when a least squares line is used to model the trend in the data for Melbourne is

congestion level $= -1515 + 0.7667 \times$ *year*

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)

Adapted from VCAA 2018 Exam 2 Data analysis Q3e

The average mark on this type of question was **0.4**.

Questions from multiple lessons

Data analysis

15. Which of the following graphical representations could be used to best identify and describe the association between the variables *age*, in years, and *favourite colour* (red, green, blue)?

A. a histogram

B. a scatterplot

C. a bar graph

D. a back-to-back stem plot

E. parallel boxplots

Adapted from VCAA 2017NH Exam 1 Data analysis Q6

Back to contents

Data analysis *Year 11 content*

16. A group of students is conducting research on the participation numbers of sports at their school, and decide to focus on the two following variables.

- *number of participants* (less than 25, 25–75, more than 75)
- *sport* (football, soccer, hockey, netball)

These variables are

A. both nominal variables.

B. both ordinal variables.

C. a nominal variable and ordinal variable respectively.

D. an ordinal variable and nominal variable respectively.

E. a numerical variable and a categorical variable respectively.

Adapted from VCAA 2017 Exam 1 Data analysis Q7

Recursion and financial modelling *Year 11 content*

17. Jack decides to deposit some money into a savings account that will pay interest every day so that he can go on a holiday.

The balance in Jack's account, in dollars after n days, V_n, can be modelled by the recurrence relation shown.

$V_0 = 15\,000, \quad V_{n+1} = 1.000041 \times V_n$

A rule of the form $V_n = a \times b^n$ can be used to determine the balance of Jack's account after n days.

a. Determine the values of a and b. (1 MARK)

b. What would be the value of n if Jack wants to determine the value of his investment after six years? Assume that there are no leap years. (1 MARK)

Adapted from VCAA 2018 Exam 2 Recursion and financial modelling Q4ci,ii

Back to contents

3C Performing a regression analysis

STUDY DESIGN DOT POINT

- 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

3A 3B **3C** 3D 3E

KEY SKILLS

During this lesson, you will be:

- calculating and interpreting the coefficient of determination
- performing residual calculations and constructing a residual plot
- performing a residual analysis.

KEY TERMS

- Coefficient of determination
- Residual plot

Although a linear model can fit all data sets, it is not always the best fit. This can result in inaccurate predictions that impact the reliability of long-term modelling for a data set. The coefficient of determination and a residual analysis are tools that can help determine whether the current linear model is a quality fit for the data set being analysed.

Calculating and interpreting the coefficient of determination

The **coefficient of determination**, r^2, is the degree to which the variation in the response variable can be predicted from the variation in the explanatory variable for a given linear relationship. It is calculated by squaring the correlation coefficient, r. As it is the square of the correlation coefficient, the r^2 value will always be between 0 and 1.

$0 \leq r^2 \leq 1$

In order to interpret the coefficient of determination, it must first be converted into a percentage.

A lower coefficient of determination could indicate that a linear model might not be the best fit for the data set.

Worked example 1

A study was conducted to explore the effect of the time in minutes spent playing *video games* on time spent *studying* in minutes. It was determined that the correlation coefficient, r, for the data collected during the study is 0.6584.

a. Determine the coefficient of determination, r^2, correct to four decimal places.

Explanation

Calculate the coefficient of determination.

$(0.6584)^2 = 0.43349...$

Continues →

Back to contents

Answer

0.4335

b. Interpret the coefficient of determination in terms of time spent *studying* and time spent playing *video games*.

Explanation

Step 1: Determine the explanatory and response variables.

Since the time spent *studying* is affected by the time spent playing *video games*, *studying* is the response variable.

EV: *video games*

RV: *studying*

Step 2: Convert the coefficient of determination into a percentage.

$0.4335 \times 100 = 43.35$

Step 3: Interpret the coefficient of determination in terms of time spent playing *video games* and time spent *studying*.

Answer

43.35% of the variation in the time spent *studying* can be explained by the variation in the time spent playing *video games*.

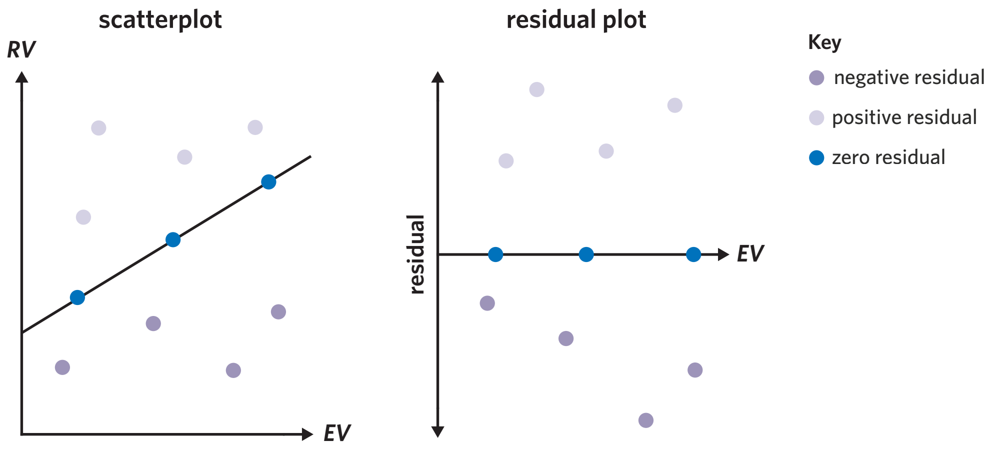
Performing residual calculations and constructing a residual plot

Recall that a residual is the vertical distance between the actual data value and the value predicted from the regression line. A residual can be calculated using the formula

residual = *actual data value* − *predicted data value*

Data points that sit above the regression line will have a positive residual value. Those that fall below the regression line will have a negative residual value. Those that lie on the regression line will have a residual value of zero.

A **residual plot** is a graph of the residual values against the explanatory variable. It can be created once all of the residual values are calculated.



Back to contents

Worked example 2

A least squares regression equation has been used to predict the *total cost* ($) of a table bill at a restaurant from the *number of people* eating together on the table, such that

total cost = 7 + 29 × *number of people*

a. If the actual *total cost* for a table of six people is $166, calculate the residual value when the *total cost* is predicted using the regression equation.

Explanation

Step 1: Determine the actual data value.

The actual data value is the cost given in the question.

actual data value = 166

Step 2: Calculate the predicted data value.

Substitute *number of people* = 6 into the regression equation.

total cost = 7 + 29 × 6

= 181

predicted data value = 181

Step 3: Calculate the residual.

Substitute *actual data value* = 166 and *predicted data value* = 181 into the residual formula.

residual = 166 − 181

Answer

−15

b. There were 8 other groups of people at the restaurant. The *number of people* at each table and the *total cost* for each group were recorded and displayed in a table.

number of people	3	4	1	6	5	2	2	4
actual *total cost* ($)	90	133	29	174	154	73	62	129
predicted *total cost* ($)								
residual value ($)								

Use the regression equation to complete the table by filling in the predicted *total cost* and residual values.

Explanation

Step 1: Fill in the table with the predicted data values.

Substitute each value for *number of people* from the table into the regression equation.

3 people: *total cost* = 7 + 29 × 3 = 94

4 people: *total cost* = 7 + 29 × 4 = 123

number of people	3	4	1	6	5	2	2	4
actual *total cost* ($)	90	133	29	174	154	73	62	129
predicted *total cost* ($)	94	123	36	181	152	65	65	123
residual value ($)								

Step 2: Fill in the table with the residual values.

Substitute the actual and predicted *total cost* for each group into the residual formula.

3 people: *residual* = 90 − 94 = −4

4 people: *residual* = 133 − 123 = 10

Continues →

Back to contents

Answer

number of people	3	4	1	6	5	2	2	4
actual *total cost* ($)	90	133	29	174	154	73	62	129
predicted *total cost* ($)	94	123	36	181	152	65	65	123
residual value ($)	−4	10	−7	−7	2	8	−3	6

---

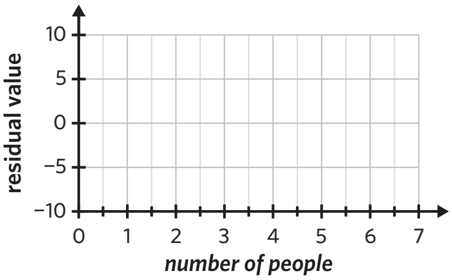
c. Construct a residual plot for this data set.

Explanation – Method 1: By hand

Step 1: Draw a set of axes with an appropriate scale and labels.

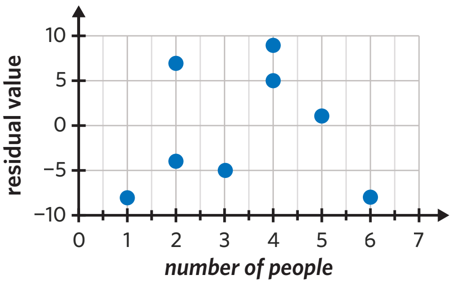
The label for the horizontal axis will be the explanatory variable, *number of people*.

The label for the vertical axis will be 'residual value'.



Step 2: Plot each residual value against the number of people.

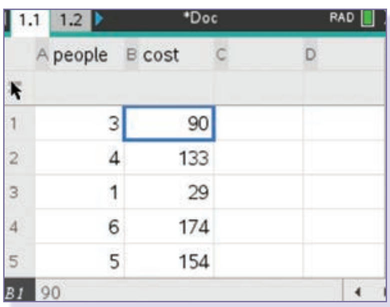
Answer



Explanation – Method 2: TI-Nspire

Step 1: From the home screen, select '1: New' → '4: Add Lists & Spreadsheet'.

Step 2: Name column A 'people' and column B 'cost' and enter the data from the table.

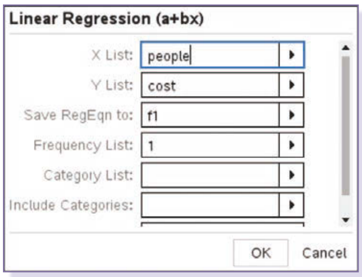


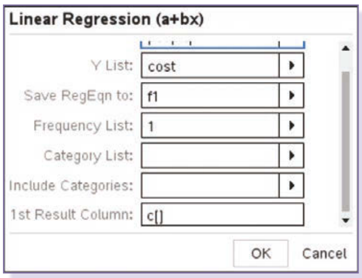
Continues →

Back to contents

Step 3: Determine the least squares regression line. Press menu → '4: Statistics' → '1: Stat Calculations' → '4: Linear Regression (a+bx)'.

Step 4: On the settings window, change 'X List' to 'people' and 'Y List' to 'cost'. Check that '1st Result Column' is set to 'c[ ]'. Select 'OK'.

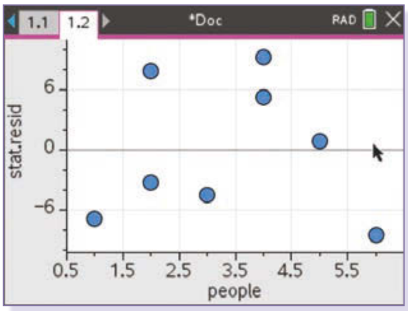




Step 5: Press ctrl + doc and select 'Add Data & Statistics'. Add the variables to each axis using the 'Click to add variable' function.

Select 'people' on the horizontal axis and 'stat.resid' on the vertical axis.

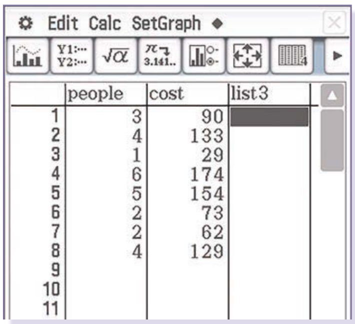
Answer



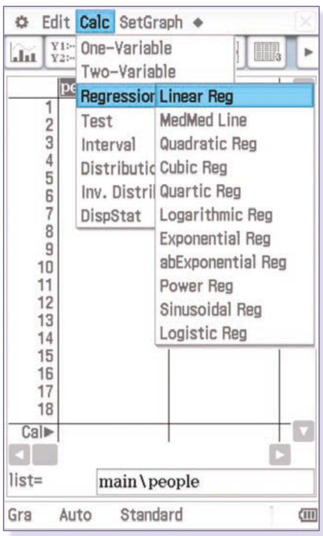
Explanation – Method 3: Casio ClassPad

Step 1: From the main menu, tap Statistics.

Step 2: Name list1 'people' and list2 'cost' and enter the data from the table.

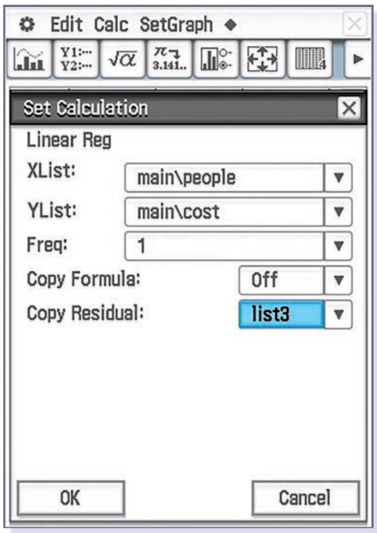


Step 3: Tap 'Calc' → 'Regression' → 'Linear Reg'.



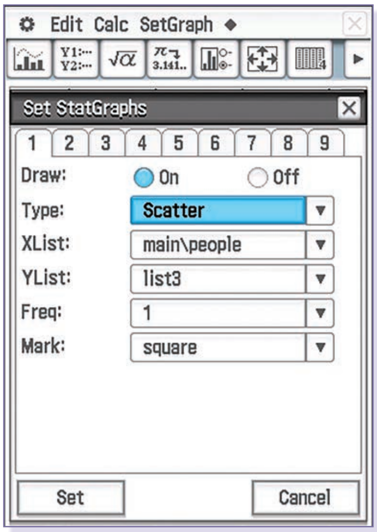
Continues →

Step 4: Specify the data set by changing 'XList:' to 'main\people' and 'YList:' to 'main\cost'. Change 'Copy Residual' to 'list3'.



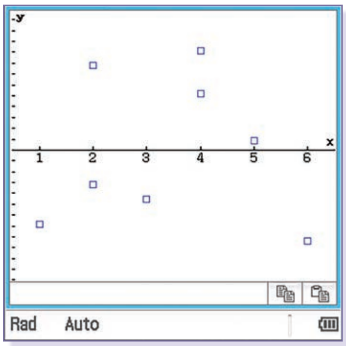
Step 5: Tap 'OK' and the 'Stat Calculations' window will appear. Tap 'OK' to close this window.

Step 6: Tap [icon] and the 'Set StatGraphs' window will appear. Change 'Type 'to 'Scatter'. Change 'XList:' to 'main\people' and 'YList:' to 'list3'. Tap set to confirm.



Step 7: Tap [icon] to display the residual plot.

Answer

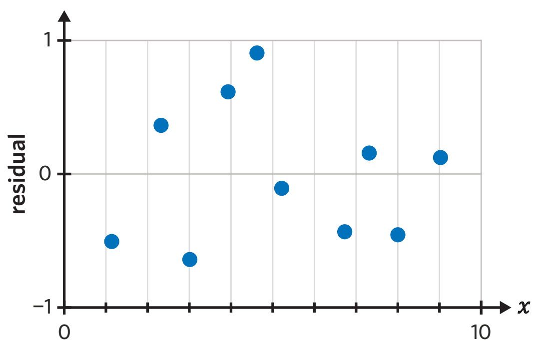


Performing a residual analysis

Residual plots can be analysed to help determine if a linear relationship exists between the explanatory and response variables.

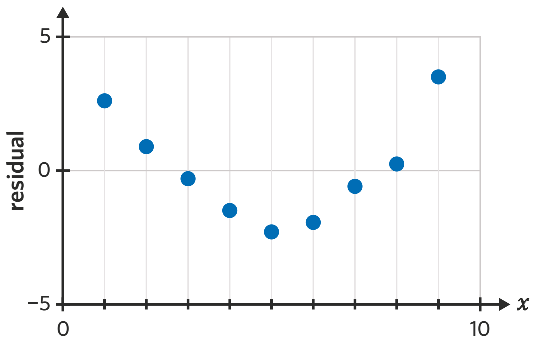
If a linear relationship is present, the residual plot will show an approximately equal number of points randomly scattered above and below the horizontal axis, with no clear pattern. These plots support the assumption of a linear relationship between the variables.

If a linear relationship is not present, the residual plot might show an unequal number of points above and below the horizontal axis or an ordered scattering of the points, showing a clear pattern. These plots do not support the assumption of a linear relationship between the variables.



This residual plot supports the assumption of linearity.

Back to contents

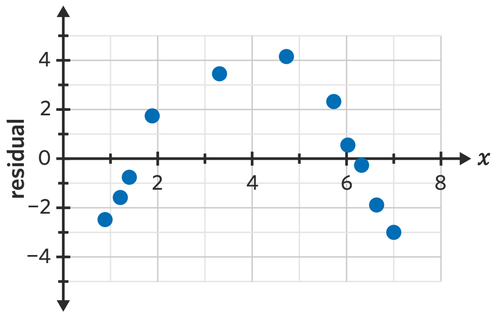


This residual plot does not support the assumption of linearity.

Worked example 3

Determine if the following residual plots support the assumption of linearity.

a.



Explanation

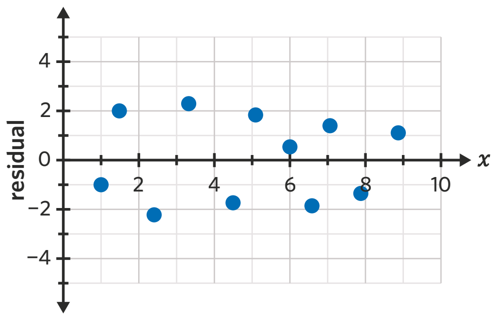
Check if the residual plot meets the conditions to support the assumption of linearity.

The residual plot appears to show a clear curved pattern.

Answer

The residual plot does not support the assumption of linearity.

b.



Explanation

Check if the residual plot meets the conditions to support the assumption of linearity.

The residual plot shows an approximately equal number of points above and below the horizontal axis. The points also appear to be randomly scattered and do not show a pattern.

Answer

The residual plot supports the assumption of linearity.

Back to contents

Exam question breakdown

VCAA 2021 Exam 2 Data analysis Q2b

In the sport of heptathlon, athletes compete in seven events. The two running events in the heptathlon are the 200 m and the 800 m run. The times taken by the athletes in these two events, *time200* and *time800*, are linearly related.

The mean and 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

$time800 = 0.03931 + 5.2756 \times time200$

Use this information to calculate the coefficient of determination as a percentage.

Round to the nearest percentage. (2 MARKS)

Explanation

Step 1: Identify the correct method to calculate the coefficient of determination.

As the coefficient of determination is r^2, the only way to calculate it from the given information is to calculate the correlation coefficient (r) first and square it.

The equation that can be used is

$$b = \frac{r \times s_y}{s_x}$$

Step 2: Identify the values needed to calculate the slope using summary statistics.

*time*800 is the response variable and *time*200 is the explanatory variable. This will help in determining which statistic is s_x and which is s_y.

$s_y = 8.2910$, $s_x = 0.96956$, $b = 5.2756$

Step 3: Substitute the known values and solve for r.

$$b = \frac{r \times s_y}{s_x}$$

$$5.2756 = \frac{r \times 8.2910}{0.96956}$$

$$r = 0.6169...$$

Step 4: Calculate the coefficient of determination.

$$r^2 = (0.6169...)^2$$
$$= 0.3806...$$

Step 5: Convert the coefficient of determination into a percentage.

$0.3806... \times 100 = 38.06...$

Answer

38%

The average mark on this question was **0.6**.

A number of students struggled to calculate Pearson's correlation coefficient from the summary statistics formulas. This is likely because they did not remember how the correlation coefficient is related to the slope of a least squares regression equation. Errors could have also been made by misinterpreting which variables were explanatory and response, which could have led to the incorrect use of the summary statistics given.

3C Questions

Calculating and interpreting the coefficient of determination

1. The *price of burgers* and *revenue* (per annum) for seven different burger stores are collected and studied. The r value for the collected data is 0.785.

The coefficient of determination for the study is closest to

A. 0.616 **B.** 0.617 **C.** 0.785 **D.** 0.886

Back to contents

2. A study was conducted to research the impacts of *hours worked* on *hours slept*. The results are summarised in the following table.

hours worked	4	5	6	7	8	9	10
hours slept	9.0	8.5	8.3	8.0	7.5	7.0	6.8

a. What is the coefficient of determination, rounded to two decimal places?

b. Interpret the coefficient of determination in terms of the amount of *hours worked* and *hours slept*.

Performing residual calculations and constructing a residual plot

3. Rachel's Health and Human Development class is studying the average growth in the *height* of children and young teenagers. They determine the following least squares regression equation, which can be used to predict *height* (cm) from *age* (years).

$height = 47.32 + 7.13 \times age$

Rachel is 14 years old and 152 cm tall. What is the residual value when her *height* is predicted from the least squares regression equation, correct to two decimal places?

A. −4.86 **B.** 0.68 **C.** 4.86 **D.** 152.00

4. A group of scientists collected the IQ of 5 participants. The scientists then determined a least squares regression equation that can be used to predict *test score* on a particular intelligence test from *IQ*. The predicted test scores were calculated and recorded.

The participants then completed the intelligence test and the scientists recorded their actual *test scores*, before calculating the residual values of each participant. However, they lost the predicted *test score* and residual value for the participant with an *IQ* of 91.

IQ	101	87	94	106	91
actual *test score* (%)	76	54	60	80	54
predicted *test score* (%)	72.90	50.92	61.91	80.75	
residual value (%)	3.10	3.08	−1.91	−0.75	

a. Find the regression equation used by the scientists. Round values to two decimal places.

b. Use the regression equation from part **a** to find the predicted *test score* and residual value for the participant with an *IQ* of 91.

5. The following table shows the *length* (cm) and *width* (cm) of seven rectangular cakes at 'Chloe's Cake Shop'.

***length* (cm)**	32	45	50	28	40	35	56
***width* (cm)**	26	35	50	20	25	35	44

A regression equation has been calculated based on these results and can be used to predict the width of a cake from its length.

$width = -3.64 + 0.91 \times length$

Use the regression equation to fill in the predicted widths and residual values in the following table.

Give values correct to two decimal places.

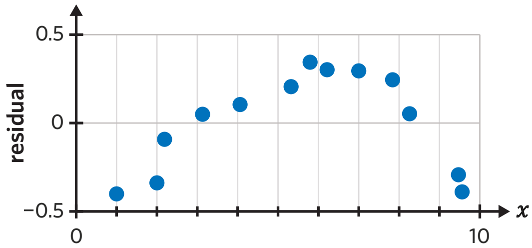
***length* (cm)**	32	45	50	28	40	35	56
***width* (cm)**	26	35	50	20	25	35	44
predicted *width* (cm)							
residual *value* (cm)							

Back to contents

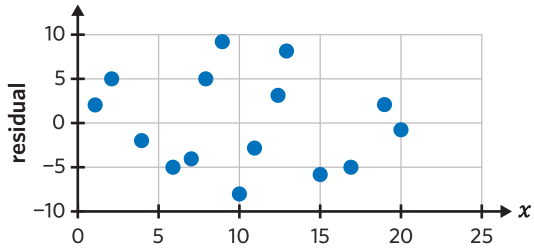
Performing a residual analysis

6. For each of the following residual plots, determine whether the assumption of linearity is supported.

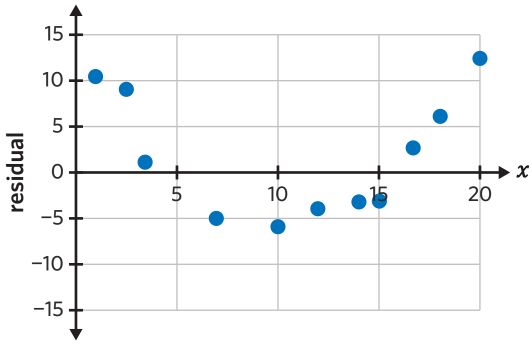
a.



b.



c.



7. The *height* (m) of a small tree was recorded at the end of every year from the time it was planted. The results are displayed in a table.

years after planting	1	2	3	4	5	6	7	8
height **(m)**	0.76	1.58	2.15	2.54	2.98	3.21	3.45	3.55

Use a calculator to construct a residual plot and use it to test the assumption of linearity.

Joining it all together

8. A Year 10 Maths class recently completed a statistics test and a calculus test. Six students recorded their results from both tests. They are displayed in the following table.

	Sophie	**Isaac**	**Mayu**	**Deanne**	**Selby**	**Emily**
statistics test result **(%)**	75	61	88	43	94	64
calculus test result **(%)**	64	53	95	40	92	59

The students calculated the following least squares regression equation that could predict a *calculus test result* from a *statistics test result*.

calculus test result $= -12.66 + 1.13 \times$ *statistics test result*

a. Use the least squares regression equation to determine each student's predicted *calculus test result* (%), correct to one decimal place.

b. Using the rounded answers from part **a**, calculate the residual values when the *calculus test result* (%) is predicted from the least squares regression equation. Give values correct to one decimal place.

c. Which student(s) didn't perform as well as was predicted on the calculus test?

d. Construct a residual plot by hand and use it to determine if the assumption of linearity is supported for the relationship between *statistics test result* and *calculus test result* (%).

Back to contents

9. A scientific study was conducted measuring *age* (years) against *reaction time* (milliseconds).

The results of ten people are displayed in the following table.

age **(years)**	19	42	73	56	23	85	61	49	34	58
reaction time **(ms)**	3.2	5.5	9.6	9.1	6.0	14.1	11.7	8.3	6.6	8.0

a. What is the value of the coefficient of determination, correct to two decimal places?

b. Interpret the coefficient of determination in terms of *age* and *reaction time*.

c. Determine the least squares regression equation that can be used to predict *reaction time* (milliseconds) from *age* (years). Give values correct to two decimal places.

d. Fill in the following table by calculating the residual values. Give values correct to two decimal places.

age **(years)**	19	42	73	56	23	85	61	49	34	58
reaction time **(ms)**	3.2	5.5	9.6	9.1	6.0	14.1	11.7	8.3	6.6	8.0
residual value (ms)										

e. Construct a residual plot on a calculator and use it to determine if the assumption of linearity for the relationship between *age* and *reaction time* is supported.

10. Sebastian thinks that there may be a relationship between the amount of time he spends on his phone and the amount of sleep that he gets. He records his results for 8 days and is summarised in the table.

time spent on phone **(min)**	175	115	75	205	155	220	80	165
sleep **(hours)**	6.0	8.0	9.5	5.0	6.2	5.4	9.0	6.1

a. What is the value of the coefficient of determination, correct to three significant figures?

b. Interpret the coefficient of determination in terms of *time spent on phone* and *sleep*.

c. Determine the least squares regression line that can be used to predict *sleep* (hours) from *time spent on phone* (min). Give values correct to four significant figures.

d. Fill in the following table by calculating the residual values. Give values correct to 2 decimal places.

time spent on phone **(min)**	175	115	75	205	155	220	80	165
sleep **(hours)**	6.0	8.0	9.5	5.0	6.2	5.4	9.0	6.1
residual value (hours)								

e. Construct a residual plot on a calculator and use it to determine if the assumption of linearity for the relationship between *time spent on phone* and *sleep* is supported.

Exam practice

11. The congestion level in a city can be recorded as the percentage increase in travel due to traffic congestion in peak periods (compared to non-peak periods).

This is called the percentage congestion level.

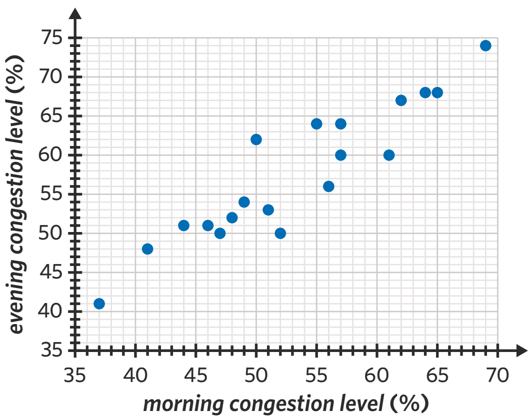
The percentage congestion levels for the morning and evening peak periods for 19 large cities are plotted on the scatterplot.

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 to the nearest percentage. (1 MARK)

VCAA 2018 Exam 2 Data analysis Q2e

59% of students answered this question correctly.



Back to contents

12. The following 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.

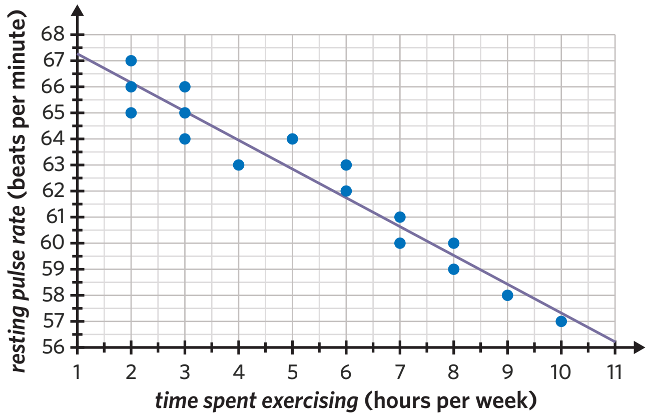
B. −1.0 beats per minute.

C. −0.3 beats per minute.

D. 1.0 beats per minute.

E. 2.0 beats per minute.

VCAA 2018 Exam 1 Data analysis Q7

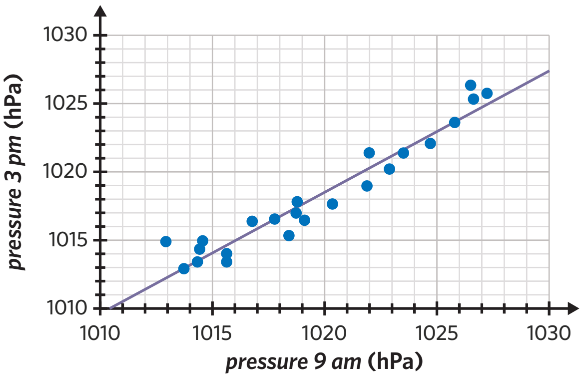


73% of students answered this question correctly.

---

13. The scatterplot shows the atmospheric pressure, in hectopascals (hPa), at 3 pm (*pressure 3 pm*) plotted against the atmospheric pressure, in hectopascals, at 9 am (*pressure 9 am*) for 23 days in November 2017 at a particular weather station.

A least squares line has been fitted to the scatterplot as shown.

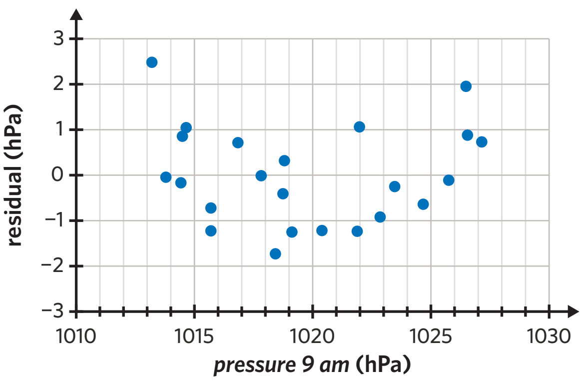


Data: Australian Government, Bureau of Meteorology, <www.bom.gov.au/>

The equation of this line is

pressure 3 pm = 111.4 + 0.8894 × *pressure 9 am*

The residual plot associated with the least squares line is shown.



a. The residual plot 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)

b. The residual plot shown does not support this assumption.

Explain why. (1 MARK)

VCAA 2019 Exam 2 Data analysis Q5f

Part **a**: **41%** of students answered this question correctly.

Part **b**: **36%** of students answered this question correctly.

Back to contents

14. The following plot shows the *winning time*, in seconds, for the women's 100 m freestyle swim plotted against *year*, for each year that the Olympic Games were held during the period 1956 to 2016.

A least squares line has been fitted to the plot to model the decreasing trend in the *winning time* over this period.

The equation of the least squares line is

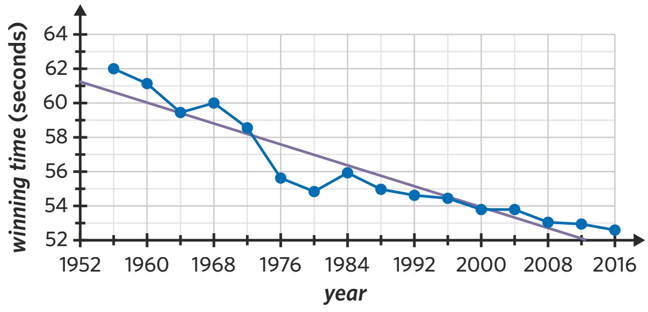
winning time = 357.1 − 0.1515 × *year*

The coefficient of determination is 0.8794.

Determine the value of the correlation coefficient (r).

Round to three decimal places. (1 MARK)

Adapted from VCAA 2021 Exam 2 Data analysis Q3b



Data: International Olympic Committee, <https://olympics.com/en/olympic-games/olympic-results>

19% of students answered this type of question correctly.

Questions from multiple lessons

Data analysis *Year 11 content*

15. The variables *time* (less than 8 minutes, 8–10 minutes, over 10 minutes) and *size* (small, medium, large) are

A. both numerical variables.

B. both nominal variables.

C. both ordinal variables.

D. a numerical and ordinal variable respectively.

E. an ordinal and nominal variable respectively.

Adapted from VCAA 2018NH Exam 1 Data analysis Q5

Data analysis *Year 11 content*

16. The following scatterplot displays the association between the *size of house*, in square metres, and *distance from CBD*, in kilometres for 15 similarly-priced Melbourne houses.

A least squares regression line has been fitted to the data with *distance from CBD* as the explanatory variable.

The equation of the least squares regression line could be

A. *distance from CBD* = 74.5 + 29.2 × *size of house*

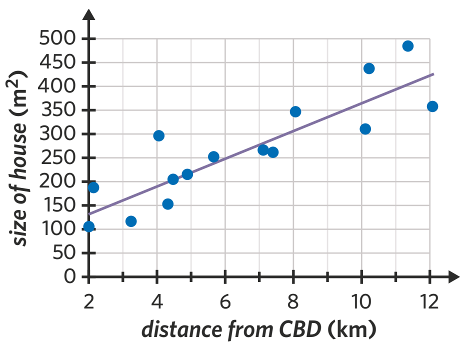
B. *distance from CBD* = 133 + 29.2 × *size of house*

C. *size of house* = 74.5 + 29.2 × *distance from CBD*

D. *size of house* = 133 + 29.2 × *distance from CBD*

E. *size of house* = 74.5 + 85.1 × *distance from CBD*

Adapted from VCAA 2017 Exam 1 Data analysis Q8



Data analysis

17. The following segmented bar chart displays the sport played by the students at three schools.

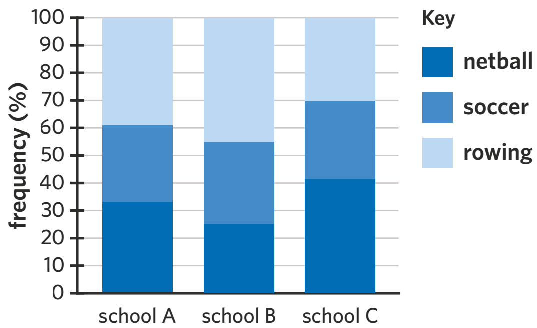
a. What percentage of students from school B play netball? (1 MARK)

b. There are 350 students at school C who play a sport. How many of these students do rowing? (1 MARK)

c. From this bar chart, it can be concluded that there is no association between the percentage of students who play soccer and the school which they attend.

Explain why this can be concluded and quote any appropriate percentages. (1 MARK)

Adapted from VCAA 2014 Exam 2 Data analysis Q1



3D Data transformations

STUDY DESIGN DOT POINT

- 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)

3A 3B 3C **3D** 3E

KEY SKILLS

During this lesson, you will be:

- choosing an appropriate data transformation
- applying a squared transformation
- applying a log transformation
- applying a reciprocal transformation.

KEY TERMS

- Linearise
- x-squared transformation
- y-squared transformation
- $\log x$ transformation
- $\log y$ transformation
- Reciprocal
- x-reciprocal transformation
- y-reciprocal transformation

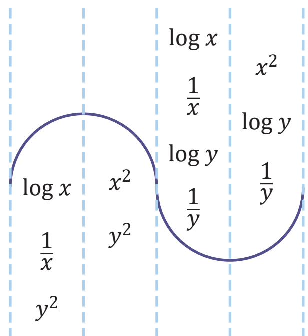
A least squares regression line should not be fitted to data if it is not linear, as any interpretations or predictions will not be accurate. If data is not linear, it may be possible to linearise it by applying a transformation to one of the variables. Three possible transformations are a squared transformation, a log transformation, and a reciprocal transformation.

Choosing an appropriate data transformation

To **linearise** data is to use a transformation to make non-linear data linear. There are three main types of transformations used to linearise data:

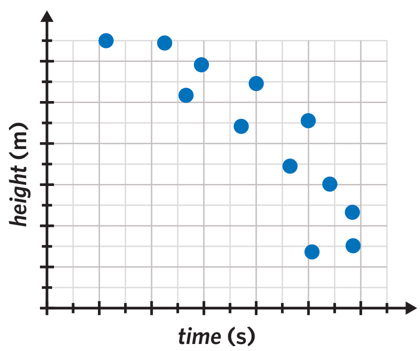
- squared transformations (x^2 and y^2)
- log (base 10) transformations ($\log x$ and $\log y$)
- reciprocal transformations $\left(\frac{1}{x} \text{ and } \frac{1}{y}\right)$

The transformation wave can help identify which transformations are most appropriate to linearise a distribution of data. The shape of the relationship between the variables should be compared to each of the segments of the transformation wave. All transformation options provided within the most similarly shaped segment can be used to linearise the data.



Worked example 1

Determine the transformations that could be used to linearise the data in the scatterplot.

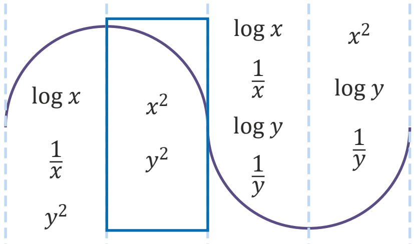


Continues →

Back to contents

Explanation

Step 1: Identify which segment of the transformation wave this scatterplot resembles.



The scatterplot most closely resembles the second segment of the transformation wave.

Step 2: Identify which transformations may be applied to this segment.

The x^2 or y^2 transformations may be applied.

Step 3: Express the transformations in terms of the variables given.

The x variable is *time* and the y variable is *height* in this instance. The transformations become *time*^{2} and *height*^{2}.

Answer

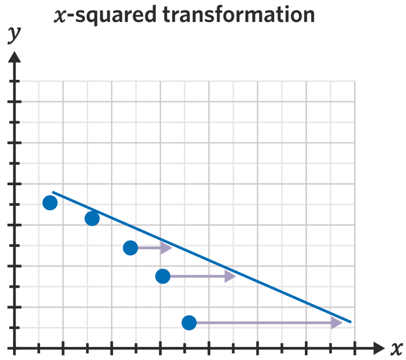
The *time*^{2} or *height*^{2} transformations could be applied.

Applying a squared transformation

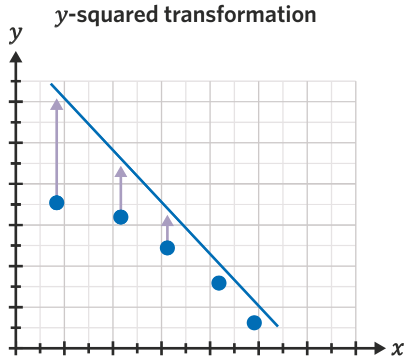
The ***x*-squared transformation** involves 'stretching' the larger x values more than the smaller x values. The y values remain the same.

The ***y*-squared transformation** involves 'stretching' the larger y values more than the smaller y values. The x values remain the same.

x-squared transformation



y-squared transformation



Worked example 2

A scatterplot was constructed from the following data.

x	3	4	5	6	7
y	8	17	24	33	52

Apply an x-squared transformation and plot the transformed data.

Explanation – Method 1: By hand

Step 1: Calculate the square of all the x values.

x	3	4	5	6	7
y	8	17	24	33	52
x^2	9	16	25	36	49

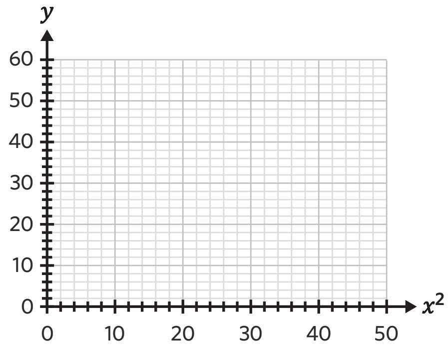
Continues →

Back to contents

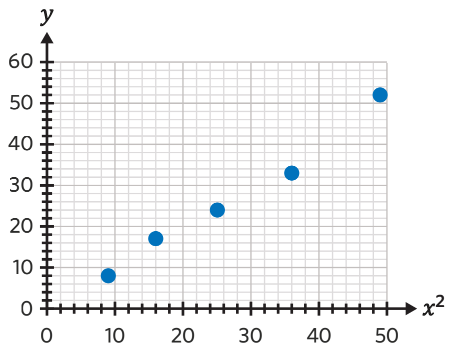
Step 2: Construct a set of axes.

A scale from 0 to 50 is appropriate for the horizontal axis, while the vertical axis needs to extend from 0 to at least 52.

The horizontal axis should be labelled x^2.



Answer



Step 3: Plot the data points using the x^2 values rather than the x values.

Explanation – Method 2: TI-Nspire

Step 1: From the home screen, select '1: New' → '4: Add Lists & Spreadsheet'.

Step 2: Name column A 'x' and column B 'y'.

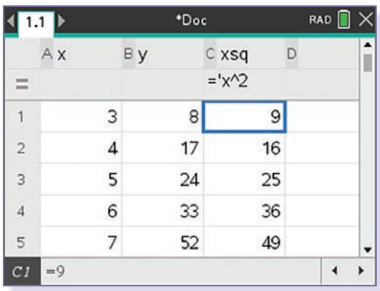
Enter the x values into column A, starting from row 1.

Enter the y values into column B, starting from row 1.

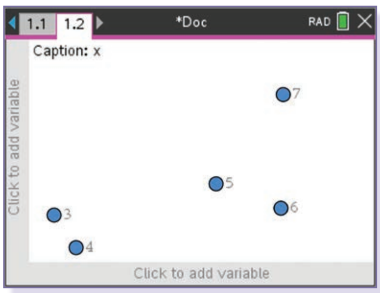
Step 3: Name column C 'xsq' (short for x squared).

Enter '=x^2' into the cell below the 'xsq' heading.

Select 'Variable Reference' → 'OK'.



Step 4: Press ctrl + doc, and select '5: Add Data & Statistics'.



Step 5: Move the cursor to the horizontal axis and select 'Click to add variable'.

Select 'xsq'.

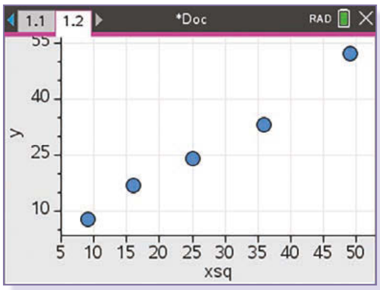
Move the cursor to the vertical axis and select 'Click to add variable'.

Select 'y'.

Continues →

Back to contents

Answer



Explanation – Method 3: Casio ClassPad

Step 1: From the main menu, tap Statistics.

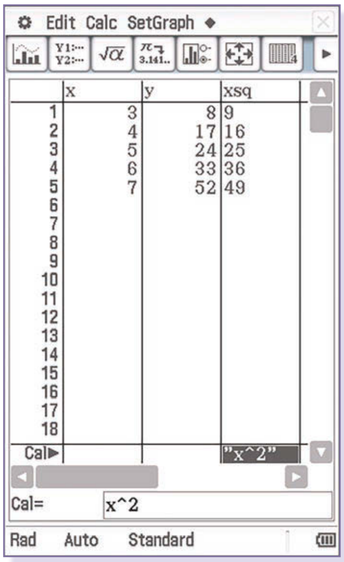
Step 2: Name the first list ‘x’ and the second list ‘y’.

Enter the x values into list ‘x’, starting from row 1.

Enter the y values into list ‘y’, starting from row 1.

Step 3: Name the third list ‘xsq’ (short for x squared).

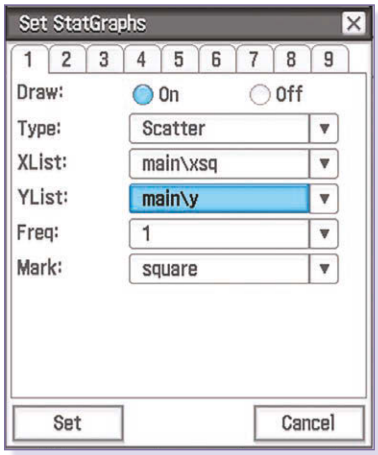
In the third list, go down to the calculation cell Cal▶ and enter ‘x^2’.



Step 4: Configure the settings of the graph by tapping .

Create a scatterplot by changing ‘Type’ to ‘Scatter’.

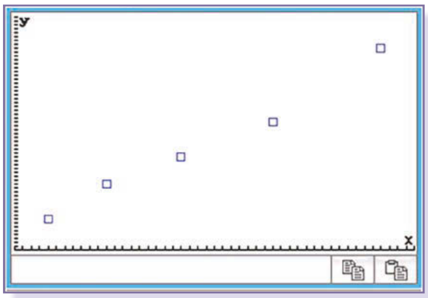
Specify the data set by changing ‘XList:’ to ‘main\xsq’ and ‘YList:’ to ‘main\y’.



Tap ‘Set’ to confirm.

Step 5: Tap to plot the graph.

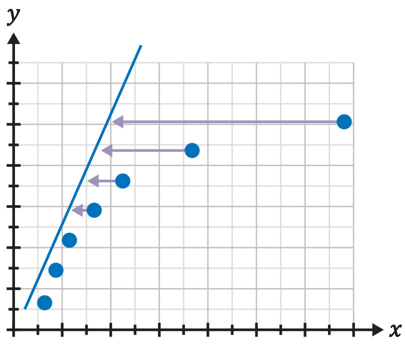
Answer

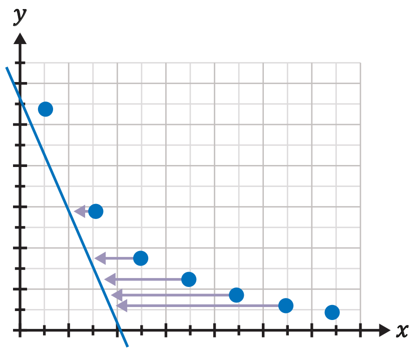


Back to contents

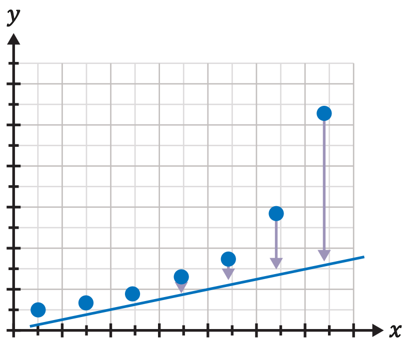
Applying a log transformation

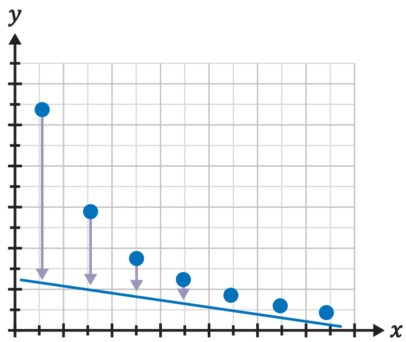
The **log x transformation** involves 'compressing' the larger x values more than the smaller x values. The y values remain the same.





The **log y transformation** involves 'compressing' the larger y values more than the smaller y values. The x values remain the same.





Worked example 3

Augustus participated in a sushi eating competition and noted his progress at certain stages of the 14 minute event. His results are shown in the table.

time **(min)**	2	3	5	7	8	11	13	14
sushi rolls eaten	21	35	51	57	62	64	67	67

Apply a log x transformation and plot the transformed data.

Explanation – Method 1: TI-Nspire

Step 1: From the home screen, select '1: New' → '4: Add Lists & Spreadsheet'.

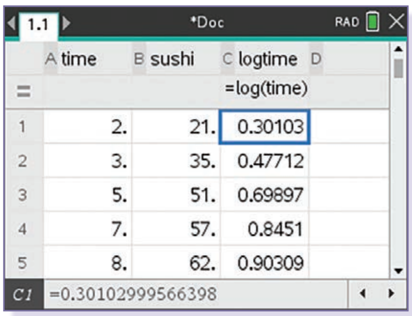
Step 2: Name column A 'time' and column B 'sushi'.

Enter the *time* values into column A, starting from row 1.

Enter the *sushi rolls eaten* values into column B, starting from row 1.

Step 3: Name column C 'logtime'.

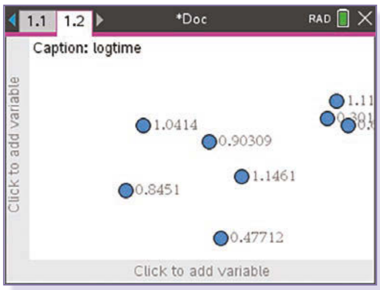
Enter '=log(time)' into the cell below the 'logtime' heading.



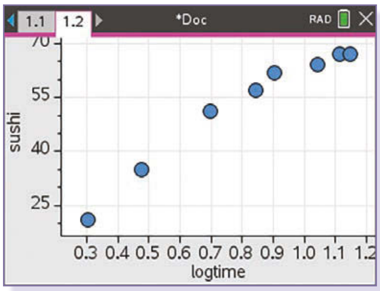
Continues →

Back to contents

Step 4: Press ctrl + doc ▾, and select '5: Add Data & Statistics'.



Answer



Step 5: Move the cursor to the horizontal axis and select 'Click to add variable'.

Select 'logtime'.

Move the cursor to the vertical axis and select 'Click to add variable'.

Select 'sushi'.

Explanation – Method 2: Casio ClassPad

Step 1: From the main menu, tap Statistics.

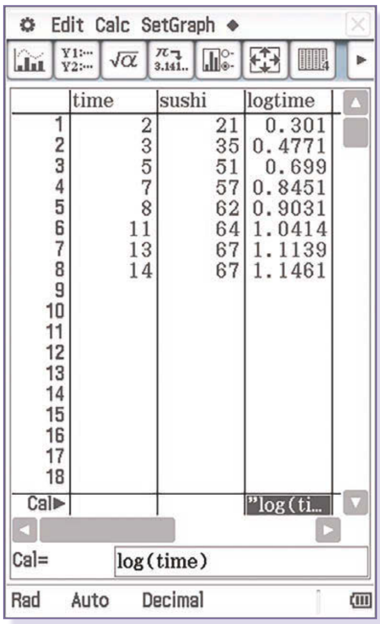
Step 2: Name the first list 'time' and the second list 'sushi'.

Enter the *time* values into list 'time', starting from row 1.

Enter the *sushi rolls eaten* values into list 'sushi', starting from row 1.

Step 3: Name the third list 'logtime'.

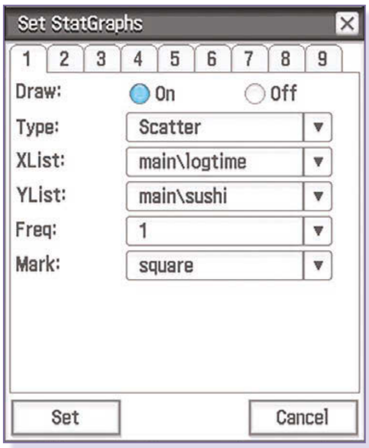
In the third list, go down to the calculation cell Cal▶ and enter 'log(time)'.



Step 4: Configure the settings of the graph by tapping the SetGraph icon.

Create a scatterplot by changing 'Type' to 'Scatter'.

Specify the data set by changing 'XList:' to 'main\logtime' and 'YList:' to 'main\sushi'.

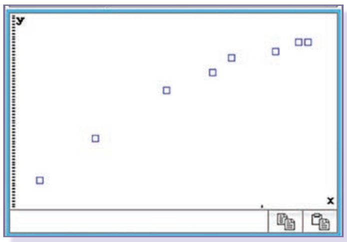


Tap 'Set' to confirm.

Step 5: Tap the graph icon to plot the graph.

Continues →

Answer



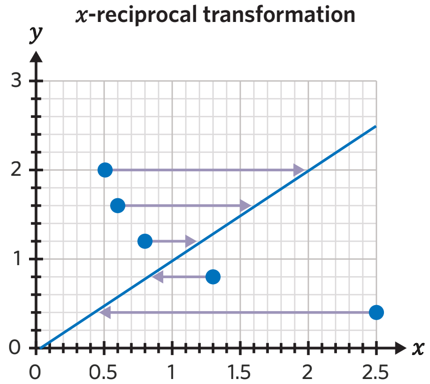
Applying a reciprocal transformation

A **reciprocal** of a number is the value found by dividing one by that number. The reciprocal of a number between zero and one becomes larger than the original number, whereas the reciprocal of a number greater than one is smaller than the original number.

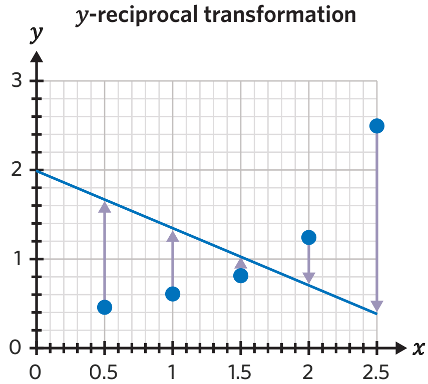
The ***x*-reciprocal transformation** involves 'compressing' x values that are greater than one, whilst 'stretching' x values that are less than one. Values closer to one are compressed/stretched less than numbers that are further away. The y values remain the same.

The ***y*-reciprocal transformation** involves 'compressing' y values that are greater than one, whilst 'stretching' y values that are less than one. Values closer to one are compressed/stretched less than numbers that are further away. The x values remain the same.

x-reciprocal transformation



y-reciprocal transformation



Worked example 4

A class of ten students recorded the time they spent studying for their psychology test. The number of questions they got incorrect are shown in the table.

hours studied	1	2	2.5	3	4	4	5	6	6.5	7
questions incorrect	63	32	23	17	16	14	8	7	5	5

Apply a y-reciprocal transformation $\left(\frac{1}{y}\right)$ and plot the transformed data.

Explanation – Method 1: TI-Nspire

Step 1: From the home screen, select '1: New' → '4: Add Lists & Spreadsheet'.

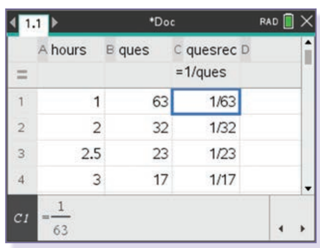
Step 2: Name column A 'hours' and column B 'ques'.

Enter the *hours studied* values into column A, starting from row 1.

Enter the *questions incorrect* values into column B, starting from row 1.

Step 3: Name column C 'quesrec' (short for *questions incorrect* reciprocal).

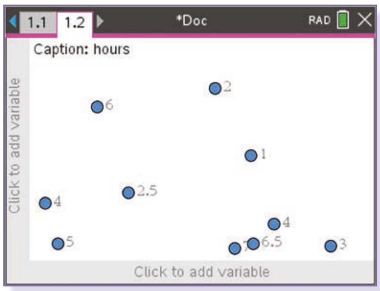
Enter '=1/ques' into the cell below the 'quesrec' heading.



Continues →

Back to contents

Step 4: Press ctrl + doc, and select '5: Add Data & Statistics'.



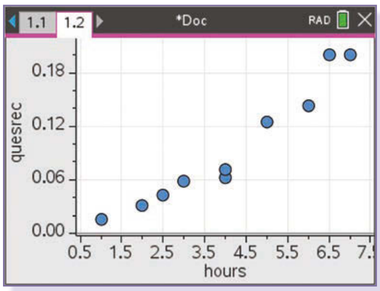
Step 5: Move the cursor to the horizontal axis and select 'Click to add variable'.

Select 'hours'.

Move the cursor to the vertical axis and select 'Click to add variable'.

Select 'quesrec'.

Answer



Explanation – Method 2: Casio ClassPad

Step 1: From the main menu, tap Statistics.

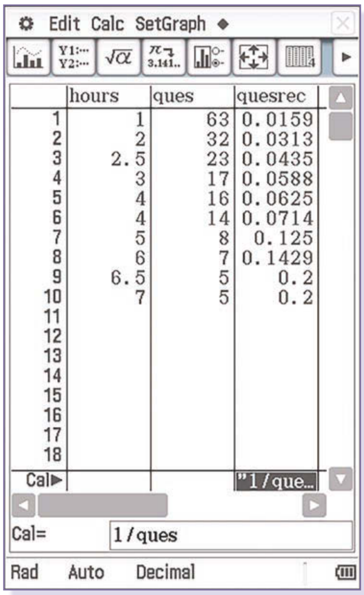
Step 2: Name the first list 'hours' and the second list 'ques'.

Enter the *hours studied* values into list 'hours', starting from row 1.

Enter the *questions incorrect* values into list 'ques', starting from row 1.

Step 3: Name the third list 'quesrec' (short for *questions incorrect* reciprocal).

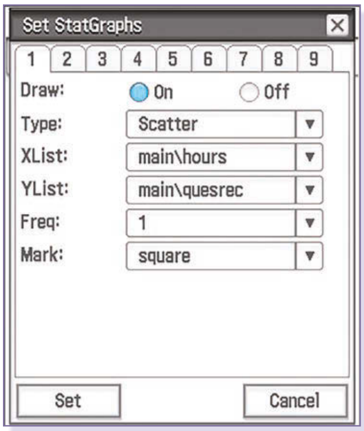
In the third list, go down to the calculation cell Cal▶ and enter '1/ques'.



Step 4: Configure the settings of the graph by tapping the SetGraph icon.

Create a scatterplot by changing 'Type' to 'Scatter'.

Specify the data set by changing 'XList:' to 'main\hours' and 'YList:' to 'main\quesrec'.



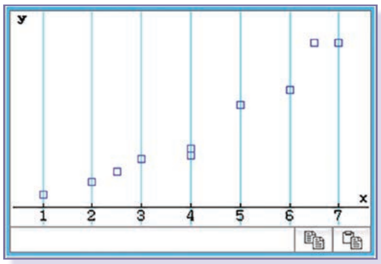
Tap 'Set' to confirm.

Step 5: Tap the graph icon to plot the graph.

Continues →

Back to contents

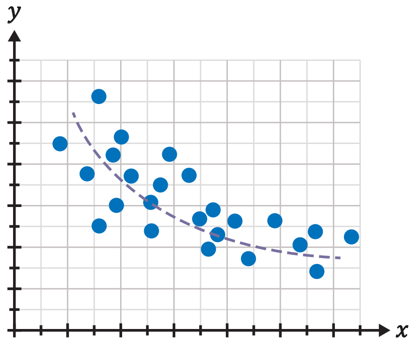
Answer



3D Questions

Choosing an appropriate data transformation

1. Which transformations could be used to linearise the data in the scatterplot?



A. log x, x-reciprocal and y-squared

B. x-squared and y-squared

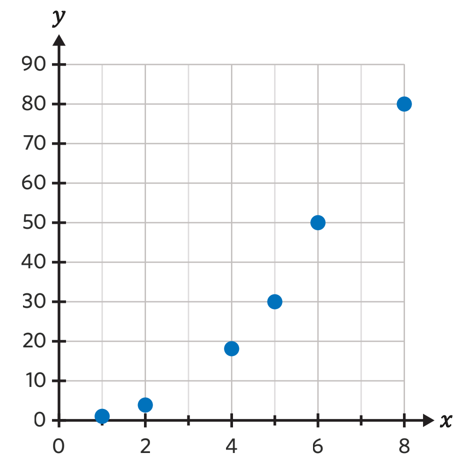
C. log x, x-reciprocal, log y and y-reciprocal

D. x-squared, log y and y-reciprocal

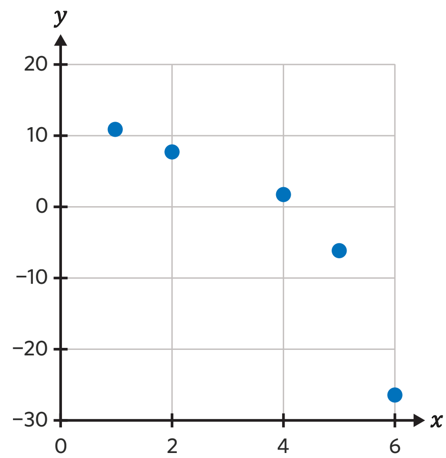
---

2. Which square transformation(s) can be used to linearise each data set?

a.

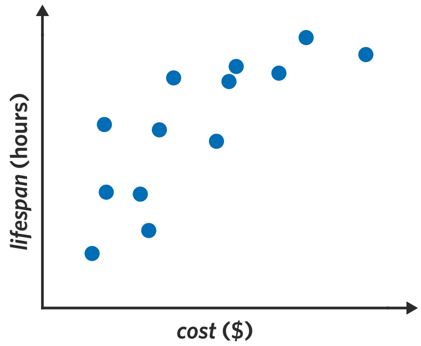


b.



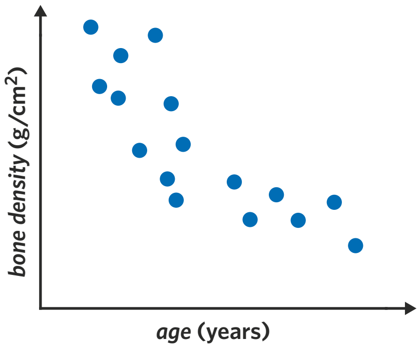
3. Determine the transformations that could be used to linearise the data in each of the following scatterplots.

a.



A. $lifespan^2$ **B.** $\log_{10}(cost)$ **C.** $\frac{1}{cost}$ **D.** All of the above

b.



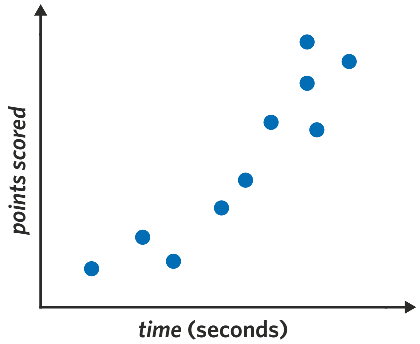
A. age^2 and $(bone\ density)^2$

B. $\frac{1}{bone\ density}$, $\log_{10}(bone\ density)$, $\frac{1}{age}$ and $\log_{10}(age)$

C. age^2, $\log_{10}(bone\ density)$ and $\frac{1}{age}$

D. $(bone\ density)^2$, $\log_{10}(age)$ and $\frac{1}{bone\ density}$

c.



A. $time^2$ **B.** $\log_{10}(points\ scored)$ **C.** $\frac{1}{points\ scored}$ **D.** All of the above

Applying a squared transformation

4. An x-squared transformation was applied to a data set, but one value is missing.

x	1	3	6	8	12
y	3	12	32	81	132
x^2	1	9	36		144

What is the missing value?

A. 8 **B.** 9 **C.** 64 **D.** 81

5. Consider the following table.

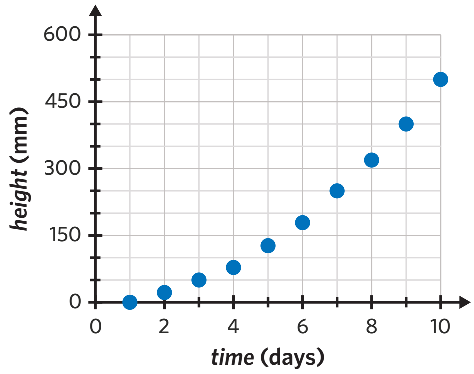
x	1	2	3	4	5
y	1.44	1.36	1.30	1.20	1.00

a. Apply an x-squared transformation to the data in the table by hand, and plot the transformed data.

b. Use a calculator to apply a y-squared transformation to the data in the table, and plot the transformed data.

6. One keen young science student monitored the growth of a plant over 10 days. He recorded the results and constructed a scatterplot from the data shown in the table.

time **(days)**	1	2	3	4	5	6	7	8	9	10
height **(mm)**	7	20	50	80	130	180	250	320	400	500



Use a calculator to apply a squared transformation to the variable *time*, and plot the transformed data.

Applying a log transformation

7. A log y transformation was applied to a data set, but one value is missing.

x	10	20	30	40	50
y	10	84	732	5045	38 720
log y	1	1.92	2.86	3.70	

What is the missing value?

A. 4.59 **B.** 5.34 **C.** 7.81 **D.** 10.56

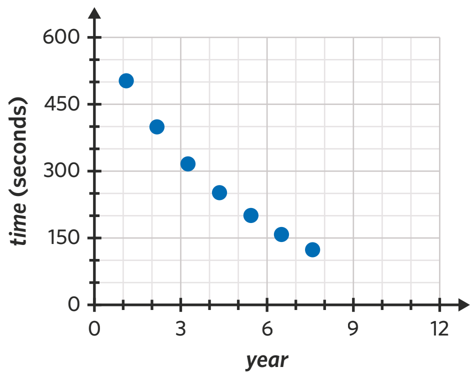
8. Use a calculator to apply a log x transformation to the data in the table, and plot the transformed data.

x	1	2	3	4	5	6	7
y	24	16	10	7	4	3	2

9. Ed has participated in an annual cheese rolling competition for the past seven years. The *time* taken for him to reach the bottom of the hill was recorded each *year* and the results are shown in the table and scatterplot.

year	1	2	3	4	5	6	7
time **(seconds)**	501	398	316	251	200	159	126

Back to contents



A log transformation can be applied to the variable *time* to linearise the scatterplot. Use a calculator to apply the transformation and plot the transformed data.

Applying a reciprocal transformation

10. An x-reciprocal transformation was applied to a data set, but one value is missing.

x	12	16	20	24	28
y	43	94	111	118	120
$1/x$	0.0833	0.0625	0.0500		0.0357

What is the missing value?

A. 0.0042 **B.** 0.0085 **C.** 0.0406 **D.** 0.0417

11. Consider the following table.

x	2	4	6	8	10
y	0.53	0.27	0.16	0.13	0.10

a. Use a calculator to apply an x-reciprocal transformation to the data in the table, and plot the transformed data.

b. Use the same sheet to apply a y-reciprocal transformation to the data in the table, and plot the transformed data.

12. Donald Trump started his career with a 'small loan' of $1 million dollars. The following table shows his *net worth* at the start of each *year* for his first ten years in the business world.

year	1	2	3	4	5	6	7	8	9	10
***net worth* ($mil)**	1.0	1.2	1.5	1.7	2.2	2.6	4.0	5.6	10.4	31.7

Use a calculator to apply a reciprocal transformation to the variable *net worth*, and plot the transformed data.

Joining it all together

13. Danny, an Olympic discus thrower, recorded the distance of 15 of his throws at a training session in the following table and scatterplot. The relationship between the *throw number* and *distance* (m) is non-linear.

throw number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
***distance* (m)**	14.7	15.4	14.9	16.3	18.9	19.6	21.7	21.5	24.7	26.8	31.6	32.4	35.9	43.6	65.2

Back to contents

a. Which transformation **cannot** linearise the scatterplot if applied?

A. x-reciprocal transformation

B. y-reciprocal transformation

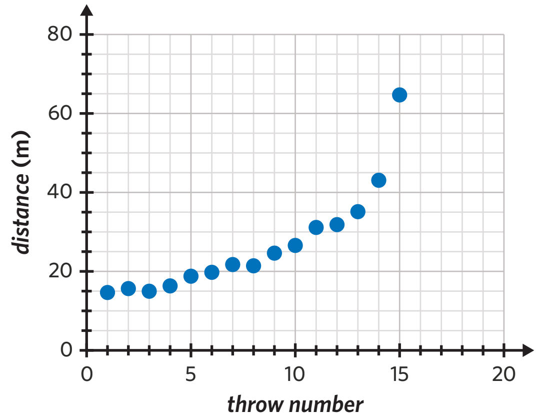
C. x-squared transformation

D. $\log y$ transformation

b. Use a calculator to apply a squared transformation to the variable *throw number*, and plot the transformed data.

c. Use a calculator to apply a log transformation to the variable *distance*, and plot the transformed data.

d. Use a calculator to apply a reciprocal transformation to the variable *distance*, and plot the transformed data.



Exam practice

14. The *mathematics achievement* level (TIMSS score) for grade 8 students and the general rate of *internet use* (%) for 10 countries are displayed in the following scatterplot.

To linearise the data, it would be best to plot

A. *mathematics achievement* against *internet use*.

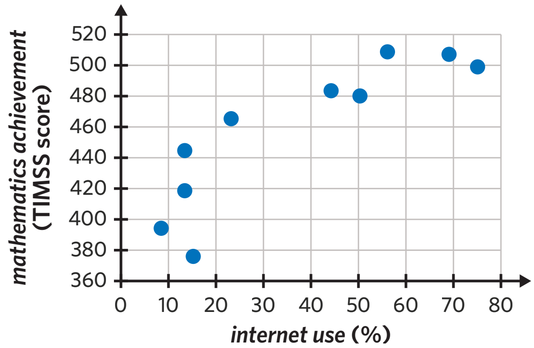
B. $\log_{10}$(*mathematics achievement*) against *internet use*.

C. *mathematics achievement* against $\log_{10}$(*internet use*).

D. *mathematics achievement* against (*internet use*)^2.

E. $\dfrac{1}{\textit{mathematics achievement}}$ against *internet use*.

VCAA 2009 Exam 1 Data analysis Q12

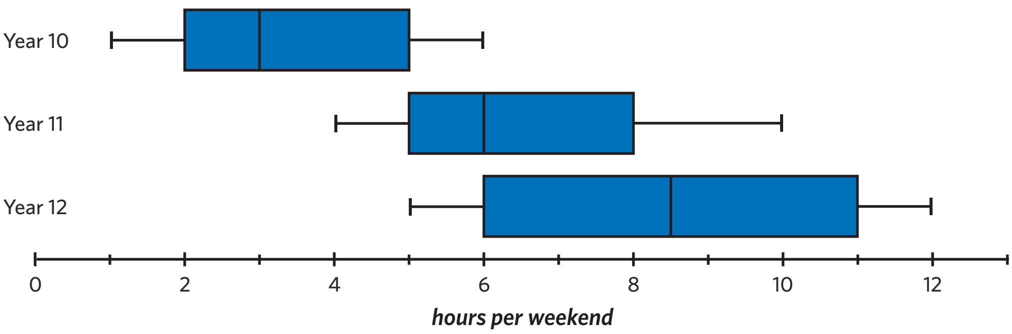


59% of students answered this question correctly.

Questions from multiple lessons

Data analysis *Year 11 content*

15. The following parallel boxplots display the distribution of hours spent studying for three groups of students (Year 10s, Year 11s and Year 12s).



Which one of the following statements is **not** true?

A. Year 12s have the most variable hours spent studying.

B. 75% of Year 10s study less than or equal to all Year 12s.

C. More than 50% of Year 10s study less than all of the other students.

D. In terms of the median, Year 10 students spend the least amount of hours studying.

E. 50% of Year 10s study the same amount as 75% of Year 11s.

Adapted from VCAA 2018 Exam 1 Data analysis Q6

Back to contents

Recursion and financial modelling *Year 11 content*

16. Chris plans to complete his collection of game tokens in four years.

Each year, he collects 120 more game tokens than the previous year.

There are a total of 2000 game tokens to collect.

If he is going to finish collecting in four years, the number of game tokens he needs to collect in the first year is

A. 200 **B.** 230 **C.** 320 **D.** 760 **E.** 1320

Adapted from VCAA 2014 Exam 1 Number patterns Q5

Data analysis

17. Eddie's monthly salary of $2500 during winter was divided according to the type of *expenditure* (food, drinks, clothes, savings).

The percentage of his salary divisions were calculated and are displayed in the following table.

expenditure	**%**
drinks	16
food	23
clothes	11
savings	50
total	100

a. How much of his salary, in dollars, was spent on clothes? (1 MARK)

b. Use the percentages to construct a percentage segmented bar chart. Use a key to indicate the segment of the bar chart that corresponds to each type of *expenditure*. (3 MARKS)

In order to investigate a possible association between his spending habits in summer and winter, Eddie's salary is divided according to the type of *expenditure*, and the season in which it was earned.

	season	
expenditure	**winter**	**summer**
food	23	16
drinks	16	31
clothes	11	29
savings	50	24
total	100	100

c. Does the information support the theory that *expenditure* is associated with *season*? Justify your answer by quoting appropriate percentages. Note that it is sufficient to quote only one type of *expenditure* in your answer. (2 MARKS)

Adapted from VCAA 2018NH Exam 2 Data analysis Q4

Back to contents

3E Data transformations – applications

STUDY DESIGN DOT POINT

- interpretation and use of the equation of the least squares line fitted to the transformed data to make predictions

3A 3B 3C 3D 3E

KEY SKILLS

During this lesson, you will be:

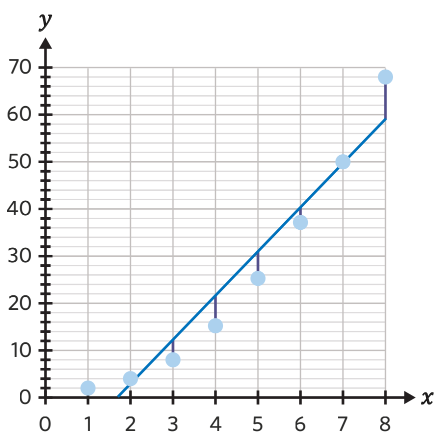
- calculating the equation of the least squares regression line for transformed data
- making predictions using the regression equation of transformed data.

Least squares regression lines are most accurate when fitted to linear data. This means that when data is non-linear, a transformation should be applied to a variable before fitting the least squares regression line. A least squares regression line that is fitted to transformed data can be used to make predictions about non-linear relationships.

Calculating the equation of the least squares regression line for transformed data

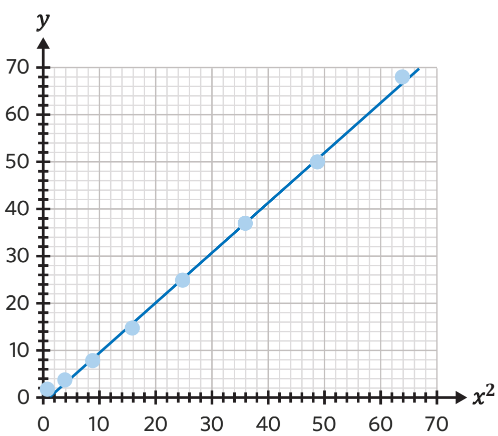
The process of fitting a least squares regression line to transformed data is no different to when fitting to linear data, except that one of the original variables is replaced by a transformed variable in the least squares regression equation. Appropriately transforming data before fitting a least squares regression line will improve accuracy for non-linear relationships.

Non-linear data



$y = -16.14 + 9.39x$

Linearised data (x^2 transformation)



$y = -0.73 + 1.05x^2$

The residuals are smaller for the least squares regression line fitted on the linearised data compared to the non-linear data. This means that the least squares regression line is more accurate after the x^2 transformation.

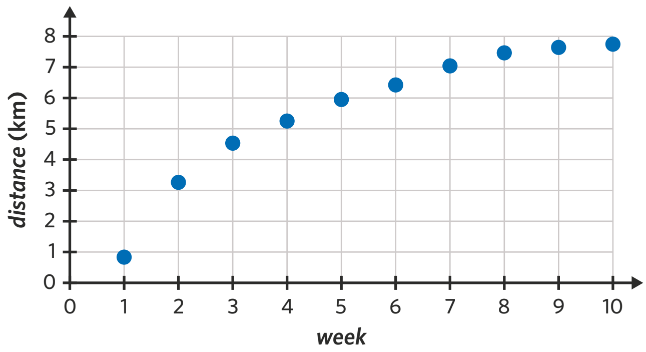
As there is more than one way to linearise a data set, it is important to be able to determine which transformation is best. This can be achieved by determining the transformation with the largest coefficient of determination, or r^2, value.

Back to contents

Worked example 1

Akin is training for an Ironman Triathlon. Each Sunday, for ten weeks, he woke up early and ran ten kilometres. He recorded the *distance*, in kilometres, he ran before his first break. The results are shown in the following table and scatterplot.

week	1	2	3	4	5	6	7	8	9	10
distance (km)	0.8	3.2	4.5	5.2	5.9	6.4	7	7.4	7.6	7.7



a. Apply a squared transformation to the variable *distance* and calculate the equation of the least squares regression line for the transformed data. Round the values of the intercept and slope to three decimal places.

Explanation – Method 1: TI-Nspire

Step 1: From the home screen, select '1: New' → '4: Add Lists & Spreadsheet'.

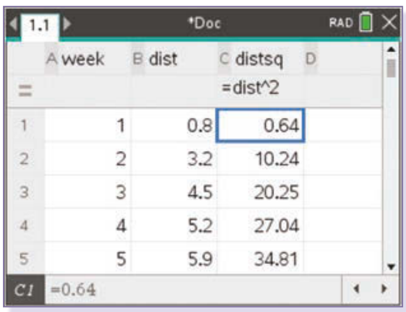
Step 2: Name column A 'week' and column B 'dist'.

Enter the *week* values into column A, starting from row 1.

Enter the *distance* values into column B, starting from row 1.

Step 3: Name column C 'distsq' (short for *distance* squared).

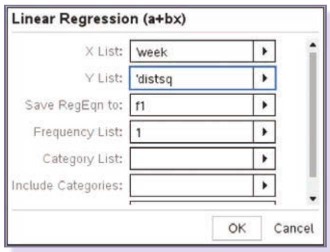
Enter '=dist^2' into the cell below the 'distsq' heading.



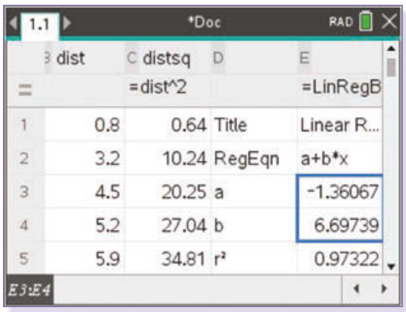
Select 'OK'.

Step 4: Press menu and select '4: Statistics' → '1: Stat Calculations' → '4: Linear Regression (a+bx)'.

Select 'week' in 'X List:' and 'distsq' in 'Y List:'



Step 5: Round the values of a and b to three decimal places.



$a = -1.361$

$b = 6.697$

Step 6: Write the equation in terms of the variables in the question.

The explanatory variable is *week*.

The response variable is $(distance)^2$.

Continues →

Back to contents

Explanation – Method 2: Casio ClassPad

Step 1: From the main menu, tap Statistics.

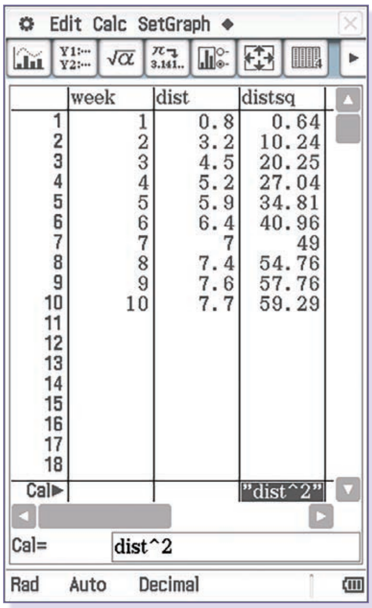
Step 2: Name the first list 'week' and the second list 'dist'.

Enter the *week* values into list 'week', starting from row 1.

Enter the *distance* values into list 'dist', starting from row 1.

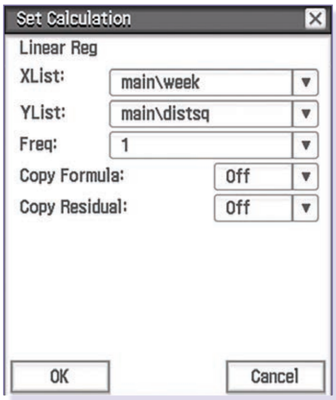
Step 3: Name the third list 'distsq' (short for *distance* squared).

In the third list, go down to the calculation cell Cal▶ and enter 'dist^2'.



Step 4: Tap 'Calc' → 'Regression' → 'Linear Reg'.

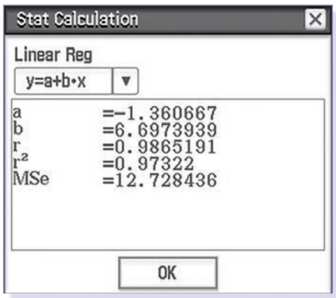
Specify the data set by changing 'XList:' to 'main\week' and 'YList:' to 'main\distsq'.



Tap 'OK' to confirm.

Step 5: Round the values of a and b to three decimal places.

Change the form of the regression equation to $y = a + bx$.



$a = -1.361$

$b = 6.697$

Step 6: Write the equation in terms of the variables in the question.

The explanatory variable is *week*.

The response variable is $(distance)^2$.

Answer – Method 1 and 2

$(distance)^2 = -1.361 + 6.697 \times week$

---

b. Apply a log transformation to the variable *week* and calculate the equation of the least squares regression line for the transformed data. Round the values of the intercept and slope to three decimal places.

Explanation – Method 1: TI-Nspire

Step 1: From the home screen, select '1: New' → '4: Add Lists & Spreadsheet'.

Step 2: Name column A 'week' and column B 'dist'.

Enter the *week* values into column A, starting from row 1.

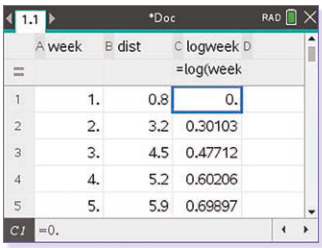
Enter the *distance* values into column B, starting from row 1.

Note: The data can be reused from part **a**.

Continues →

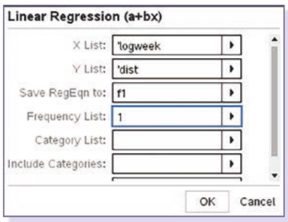
Step 3: Name column C 'logweek'.

Enter '=log(week)' into the cell below the 'logweek' heading.



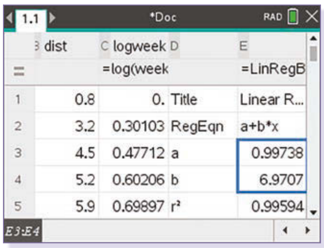
Step 4: Press menu and select '4: Statistics' → '1: Stat Calculations' → '4: Linear Regression (a+bx)'.

Select 'logweek' in 'X List:' and 'dist' in 'Y List:'



Select 'OK'.

Step 5: Round the values of a and b to three decimal places.



$a = 0.997$

$b = 6.971$

Step 6: Write the equation in terms of the variables in the question.

The explanatory variable is $\log_{10}(week)$.

The response variable is *distance*.

Explanation – Method 2: Casio ClassPad

Step 1: From the main menu, tap Statistics.

Step 2: Name the first list 'week' and the second list 'dist'.

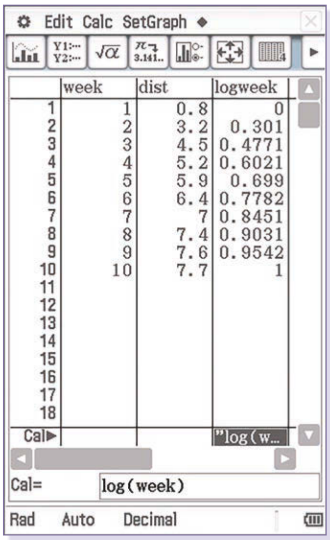
Enter the *week* values into list 'week', starting from row 1.

Enter the *distance* values into list 'dist', starting from row 1.

Note: The data can be reused from part a.

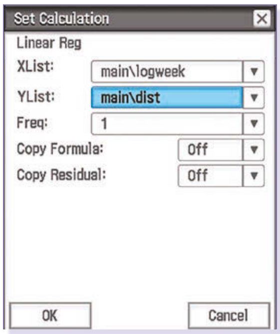
Step 3: Name the third list 'logweek'.

In the third list, go down to the calculation cell Cal▶ and enter 'log(week)'.



Step 4: Tap 'Calc' → 'Regression' → 'Linear Reg'.

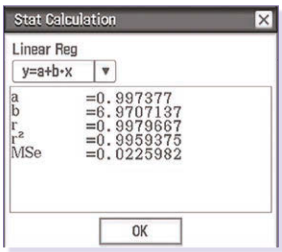
Specify the data set by changing 'XList:' to 'main\logweek' and 'YList:' to 'main\dist'.



Tap 'OK' to confirm.

Step 5: Round the values of a and b to three decimal places.

Change the form of the regression equation to $y = a + bx$.



$a = 0.997$

$b = 6.971$

Continues →

Back to contents

Step 6: Write the equation in terms of the variables in the question.

The explanatory variable is $\log_{10}(week)$.

The response variable is *distance*.

Answer – Method 1 and 2

$distance = 0.997 + 6.971 \times \log_{10}(week)$

---

c. Apply a reciprocal transformation to the variable *week* and calculate the equation of the least squares regression line for the transformed data. Round the values of the intercept and slope to three decimal places.

Explanation – Method 1: TI-Nspire

Step 1: From the home screen, select '1: New' → '4: Add Lists & Spreadsheet'.

Step 2: Name column A 'week' and column B 'dist'.

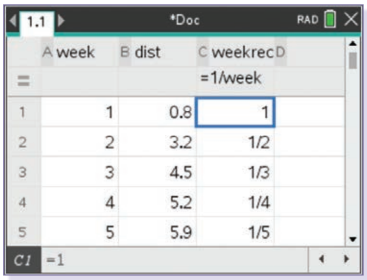
Enter the *week* values into column A, starting from row 1.

Enter the *distance* values into column B, starting from row 1.

Note: The data can be reused from part a.

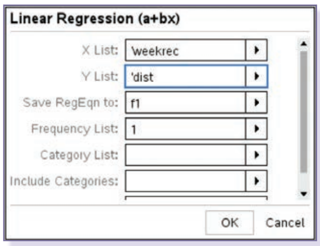
Step 3: Name column C 'weekrec' (short for *week* reciprocal).

Enter '=1/week' into the cell below the 'weekrec' heading.



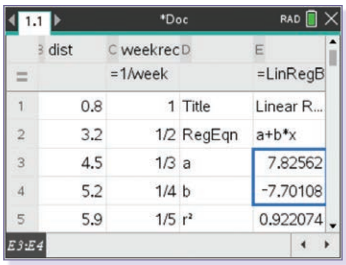
Step 4: Press menu and select '4: Statistics' → '1: Stat Calculations' → '4: Linear Regression (a+bx)'.

Select 'weekrec' in 'X List:' and 'dist' in 'Y List:'



Select 'OK'.

Step 5: Write down the equation for the least squares regression line as displayed on the screen, and round the values of a and b to three decimal places.



$a = 7.826$

$b = -7.701$

Step 6: Write the equation in terms of the variables in the question.

The explanatory variable is $\dfrac{1}{week}$.

The response variable is *distance*.

Explanation – Method 2: Casio ClassPad

Step 1: From the main menu, tap Statistics.

Step 2: Name the first list 'week' and the second list 'dist'.

Enter the *week* values into list 'week', starting from row 1.

Enter the *distance* values into list 'dist', starting from row 1.

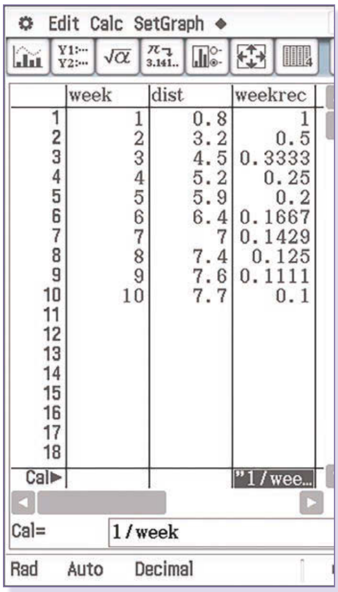
Note: The data can be reused from part **a**.

Continues →

Back to contents

Step 3: Name the third list 'weekrec' (short for *week* reciprocal).

In the third list, go down to the calculation cell Cal▶ and enter '1/week'.



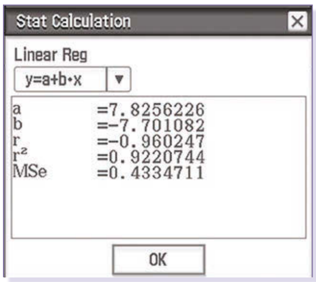
Step 4: Tap 'Calc' → 'Regression' → 'Linear Reg'.

Specify the data set by changing 'XList:' to 'main\weekrec' and 'YList:' to 'main\dist'.

Tap 'OK' to confirm.

Step 5: Round the values of a and b to three decimal places.

Change the form of the regression equation to $y = a + bx$.



$a = 7.826$

$b = -7.701$

Step 6: Write the equation in terms of the variables in the question.

The explanatory variable is $\frac{1}{week}$.

The response variable is *distance*.

Answer – Method 1 and 2

$distance = 7.826 - 7.701 \times \frac{1}{week}$

d. Determine which transformation is the best fit for the data set.

Explanation

Step 1: Determine the r^2 value for the least squares regression line of each transformation.

This can be done using the data entered in parts **a**–**c**.

Remember that the r^2 value can be located on the same screen as the a and b values.

$(distance)^2 = -1.361 + 6.697 \times week \qquad r^2 = 0.9732$

$distance = 0.997 + 6.971 \times \log_{10}(week) \qquad r^2 = 0.9959$

$distance = 7.826 - 7.701 \times \frac{1}{week} \qquad r^2 = 0.9221$

Step 2: Identify the largest value of r^2.

The largest value of r^2 is 0.9959.

Therefore, the log transformation on the variable *week* has the best fit.

Continues →

Back to contents

Answer

$distance = 0.997 + 6.971 \times \log_{10}(week)$

Making predictions using the regression equation of transformed data

A least squares regression line on transformed data can still be used to interpolate and extrapolate. However, when making predictions it is important to consider the transformation that has been applied to one of the variables.

The limitations of extrapolation are also present with a least squares regression line fitted to a transformed data set. When extrapolating, it is assumed that the shape of the relationship between the variables will continue outside of the range of the data set. This assumption has limited reliability.

Worked example 2

Akin recorded the *distance*, in kilometres, he ran before his first break for the last ten weeks, whilst training for an Ironman Triathlon. As the data he recorded was non-linear, a transformation had to be applied to one of the variables before estimating the equation of the least squares regression line. Use the following regression equations to predict how many kilometres Akin will be able to run before his first break in his 20^{th} week of training. Give answers correct to three decimal places.

a. $(distance)^2 = -1.361 + 6.697 \times week$

Explanation

Step 1: Substitute the known value into the regression equation.

Akin wants to predict the value of *distance* in the 20^{th} week, so let $week = 20$.

Step 2: Solve for the unknown value.

$(distance)^2 = -1.361 + 6.697 \times 20$

$(distance)^2 = 132.579$

$distance = 11.514$

Answer

11.514 km

b. $distance = 0.997 + 6.971 \times \log_{10}(week)$

Explanation

Step 1: Substitute the known value into the regression equation.

Akin wants to predict the value of *distance* in the 20^{th} week, so let $week = 20$.

Step 2: Solve for the unknown value.

$distance = 0.997 + 6.971 \times \log_{10}(20)$

$distance = 10.066$

Answer

10.066 km

Continues →

Back to contents

c. $distance = 7.826 - 7.701 \times \frac{1}{week}$

Explanation

Step 1: Substitute the known value into the regression equation.

Akin wants to predict the value of *distance* in the 20th week, so let $week = 20$.

Step 2: Solve for the unknown value.

$distance = 7.826 - 7.701 \times \frac{1}{20}$

$distance = 7.441$

Answer

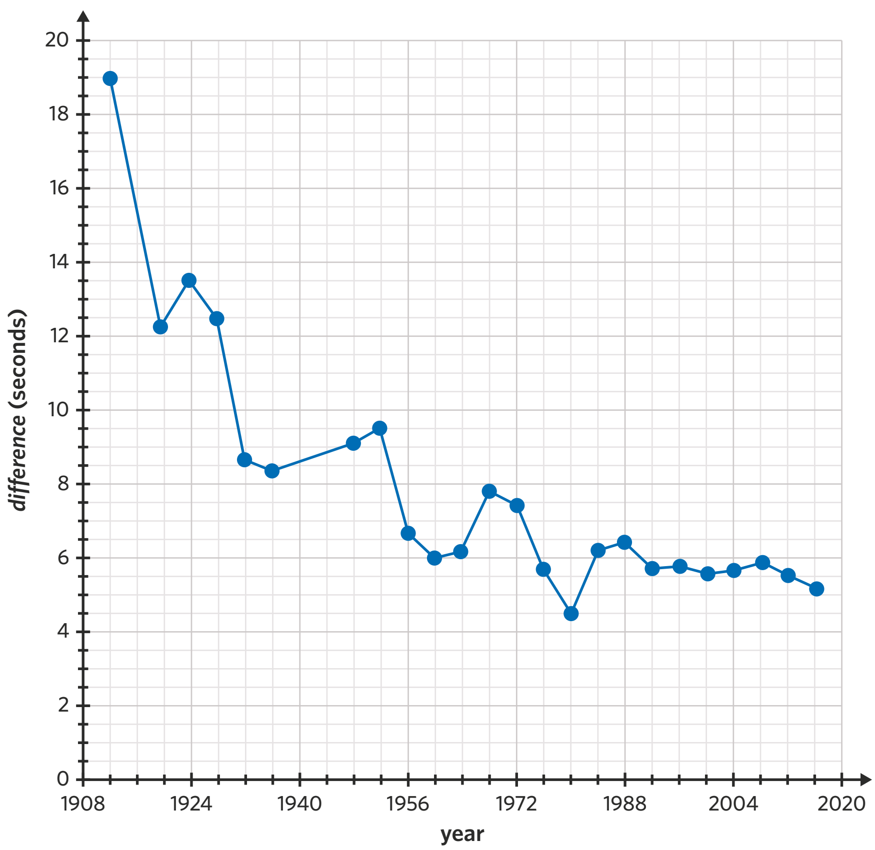
7.441 km

Exam question breakdown

VCAA 2021 Exam 2 Data analysis Q5a

A method for predicting future time differences in the 100 m freestyle swim is to use the formula $difference = winning\ time\ women - winning\ time\ men$

The resulting data and time series plot are shown. The plot is clearly non-linear.



year	*difference* (seconds)
1912	18.8
1920	12.2
1924	13.4
1928	12.4
1932	8.6
1936	8.3
1948	9.0
1952	9.4
1956	6.6
1960	6.0
1964	6.1
1968	7.8
1972	7.4
1976	5.7
1980	4.4
1984	6.1
1988	6.3
1992	5.6
1996	5.8
2000	5.5
2004	5.7
2008	5.9
2012	5.5
2016	5.1

Note: No Olympic Games were held in 1916, 1940 and 1944.

Apply a reciprocal transformation to the variable *difference* 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)

Explanation – Method 1: TI-Nspire

Step 1: From the home screen, select '1: New' → '4: Add Lists & Spreadsheet'.

Step 2: Name column A 'year' and column B 'dif'.

Enter the *year* values into column A, starting from row 1.

Enter the *difference* values into column B, starting from row 1.

Continues →

Step 3: Name column C 'difrec' (short for *difference* reciprocal).

Enter '=1/diff' into the cell below the 'difrec' heading.

Step 4: Press menu and select '4: Statistics' → '1: Stat Calculations' → '4: Linear Regression (a+bx)'.

Select 'year' in 'X List:' and 'difrec' in 'Y List:'

Select 'OK'.

Step 5: Round the values of a and b to four significant figures.

$a = -2.234$

$b = 0.001209$

Step 6: Write the equation in terms of the variables in the question.

The explanatory variable is *year*.

The response variable is $\frac{1}{difference}$.

Explanation – Method 2: Casio ClassPad

Step 1: From the main menu, tap Statistics.

Step 2: Name the first list 'year' and the second list 'dif'.

Enter the *year* values into list 'year', starting from row 1.

Enter the *difference* values into list 'dif', starting from row 1.

Step 3: Name the third list 'difrec' (short for *difference* reciprocal).

In the third list, go down to the calculation cell Cal▶ and enter '1/dif'.

Step 4: Tap 'Calc' → 'Regression' → 'Linear Reg'.

Specify the data set by changing 'XList:' to 'main\year' and 'YList:' to 'main\difrec'.

Tap 'OK' to confirm.

Step 5: Round the values of a and b to four significant figures.

$a = -2.234$

$b = 0.001209$

Step 6: Write the equation in terms of the variables in the question.

The explanatory variable is *year*.

The response variable is $\frac{1}{difference}$.

Answer – Method 1 and 2

$$\frac{1}{difference} = -2.234 + 0.001209 \times year$$

The average mark on this question was **0.5**.

Few students received full marks on this question. Students who calculated the regression equation correctly often wrote the variables incorrectly, or rounded the slope and intercept values incorrectly.

3E Questions

Calculating the equation of the least squares regression line for transformed data

1. An x-squared transformation was applied to the data in the following table, and a least squares regression line was fitted.

x	5	9	14	20	24
y	22	23	25	30	35

The equation of the least squares regression line is closest to

A. $y = 17.269 + 0.676x^2$

B. $y = 20.975 + 0.024x^2$

C. $y = 206.701 + 37.910x^2$

D. $y = -880.917 + 43.093x^2$

Back to contents

2. Apply a $\log_{10}y$ transformation to the data in the following table, and calculate the equation of the least squares regression line for the transformed data. Round the values of the intercept and slope to three decimal places.

x	2	4	5	8	10	14	15	17	19	20
y	32	34	38	53	67	102	118	134	168	182

3. Loki uploads soccer videos to Youtube daily. The number of new views on his latest video, on each day in the first week after uploading, are shown in the table.

day	1	2	3	4	5	6	7
views	321	113	54	36	28	24	21

Apply a reciprocal transformation to the variable *views*, and calculate the equation of the least squares regression line for the transformed data. Round the values of the intercept and slope to four decimal places.

4. Amanda competed in a 400 m swimming race. Her time for each lap of the 50-metre pool was recorded and the results are shown in the table.

lap	1	2	3	4	5	6	7	8
time (seconds)	31	40	44	50	53	55	58	60

a. Apply a log transformation to the variable *lap* and calculate the equation of the least squares regression line for the transformed data. Round the values of the intercept and slope to one decimal place.

b. Apply a squared transformation to the variable *time* and calculate the equation of the least squares regression line for the transformed data. Round the values of the intercept and slope to one decimal place.

c. Which transformation is the best fit for the data set?

Making predictions using the regression equation of transformed data

5. An x-reciprocal 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

$y = 843 - 165 \times \frac{1}{x}$

This equation is used to predict the value of y when $x = 5$.

The value of y is

A. 0

B. 18

C. 810

D. 876

6. The relationship between the *value* ($) and *age* (years) of an antique vase has been linearised using a log transformation. A least squares regression line is fitted to the transformed data and its equation is

$value = 37\,000 + 13\,000 \times \log_{10}(age)$

If the value of the vase is $49 400, its age is closest to

A. 1 year

B. 5 years

C. 8 years

D. 9 years

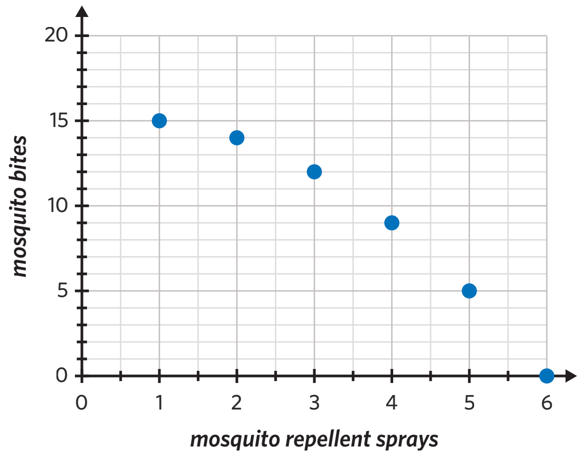
7. A group of medicine students went camping with a limited amount of mosquito repellent. The number of *mosquito repellent sprays* and the number of *mosquito bites* received for each student is shown in the scatterplot.

A squared transformation was applied to the variable *mosquito bites* to linearise the data. A least squares regression line was then fitted to the transformed data and its equation is

$(\textit{mosquito bites})^2 = 282 - 49 \times \textit{mosquito repellent sprays}$

a. Using the equation, estimate the number of *mosquito repellent sprays* a student would have applied if they received 6 mosquito bites. Round your answer to the nearest whole number.

b. Is the estimate in part **a** reliable? Justify your answer.



8. In an attempt to investigate claims that the sea levels are rising, a climate scientist decides to measure the increase in the depth of the sea just off a remote island every year, for ten years. Her results are shown in the following table.

year	1	2	3	4	5	6	7	8	9	10
sea level rise **(cm)**	0.21	0.10	0.13	0.25	0.12	0.20	0.19	0.27	0.20	0.42

A reciprocal transformation was applied to the variable *sea level rise* to linearise the data. A least squares regression line was then fitted to the transformed data and its equation is

$$\frac{1}{\textit{sea level rise}} = 8.14 - 0.46 \times \textit{year}$$

a. Fill out the following table using the equation to predict the *sea level rise* in each of the next five years. Round each estimate to two decimal places.

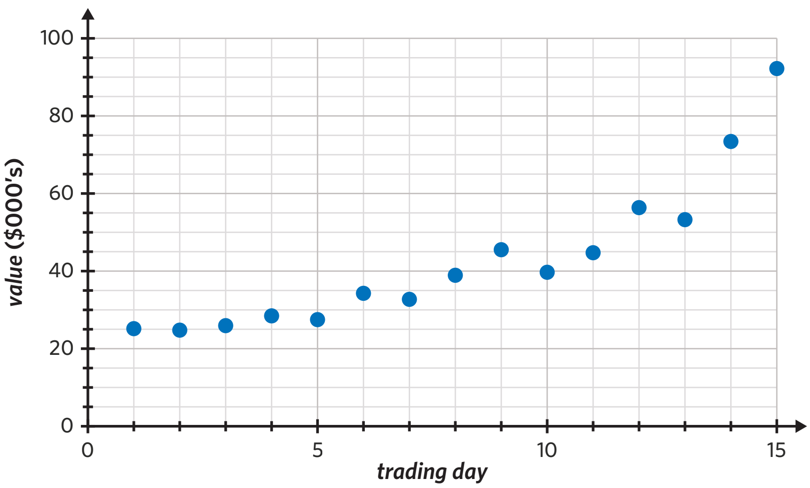
year	11	12	13	14	15
estimated sea level rise **(cm)**					

b. Are the predictions in part **a** reliable? Justify your answer.

Joining it all together

9. Johnny is a day trader who buys and sells futures contracts. He tracks the *value* ($000's) of his portfolio at the end of each *trading day*, over three weeks. The data is shown in the following table and scatterplot.

trading day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
value **($000's)**	25.0	24.5	26.0	28.4	27.3	34.1	32.5	38.7	45.2	39.4	44.8	56.2	53.2	73.4	92.4



Back to contents

a. Apply all appropriate transformations for the data set. Give the least squares regression equation for each transformation, rounding the values of the intercept and slope to three decimal places.

b. Which transformation is the best fit for the data set?

c. Consider the least squares regression equation for the transformation chosen in part **b**.

   i. Predict the *value*, in dollars, of Johnny's portfolio at the end of the 20^{th} *trading day*. Round your answer to the nearest dollar.

   ii. Predict the *trading day* in which the *value* of Johnny's portfolio first closes at more than $200 000.

   iii. Estimate the *value* ($) of Johnny's portfolio at the end of the *trading day* 4 days before he started recording. Round your answer to the nearest dollar.

d. Which of the previous predictions made are reliable? Justify your answer.

Exam practice

10. 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

$\log_{10}(\textit{number of tasks}) = 1.160 + 0.03617 \times \textit{time}$

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

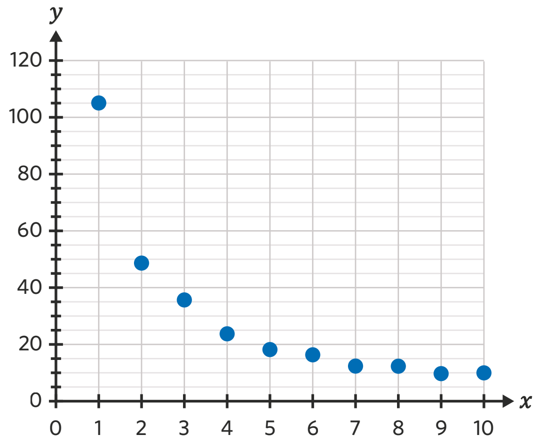
E. 26

VCAA 2020 Exam 1 Data analysis Q14

74% of students answered this question correctly.

11. Freya uses the data in the table to generate the following 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. $\frac{1}{y} = -0.025 + 1.1x$

C. $\frac{1}{y} = 7.8 - 0.082x$

D. $y = 45.3 + 59.7 \times \frac{1}{x}$

E. $y = 59.7 + 45.3 \times \frac{1}{x}$

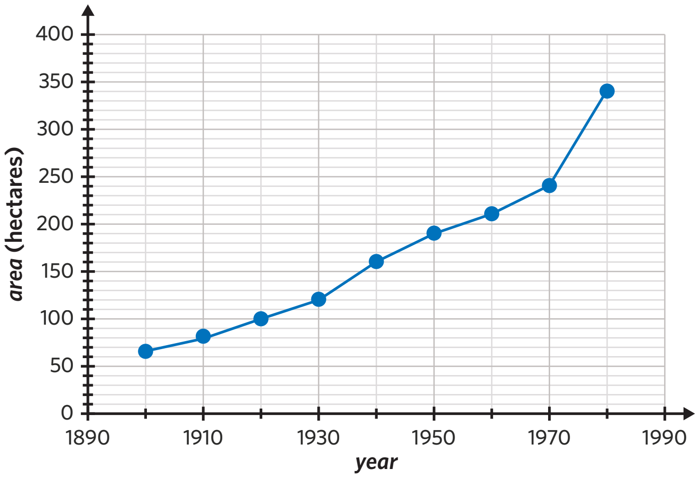
VCAA 2018 Exam 1 Data analysis Q11

58% of students answered this question correctly.

3E QUESTIONS

Back to contents

12. The following time series plot shows the total *area*, in hectares, of forest eaten by the 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. 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*.

Round the values of the intercept and slope to three significant figures. (2 MARKS)

b. 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)

c. Give a reason why this prediction may have limited reliability. (1 MARK)

VCAA 2017 Exam 2 Data analysis Q4b-cii

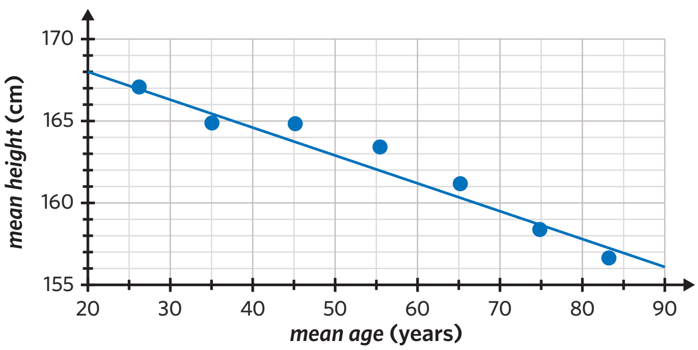
Part **a**: The average mark on this question was **0.9**.
Part **b**: **29%** of students answered this question correctly.
Part **c**: **53%** of students answered this question correctly.

13. The following table shows the *mean age*, in years, and the *mean height*, in centimetres, of 648 women from seven different age groups.

	age group						
	twenties	**thirties**	**fourties**	**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

Data: J Sorkin 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

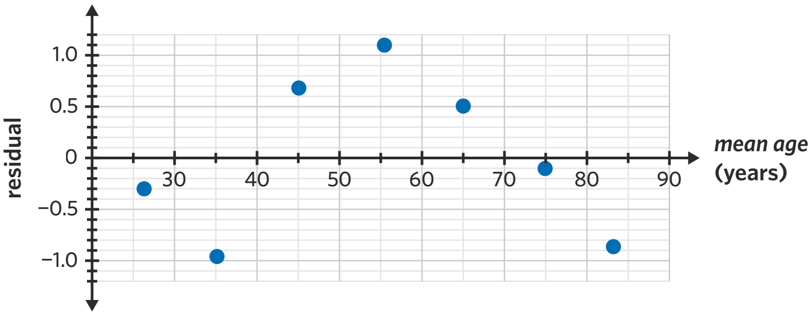
A scatterplot displaying this data shows an association between the *mean height* and the *mean age* of these women. In an initial analysis of the data, a line is fitted to the data by eye, as shown.



Back to contents

In a further analysis of the data, a least squares line was fitted.

The associated residual plot that was generated is shown.



The residual plot indicates that the association between the *mean height* and the *mean age* of women is non-linear.

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)

VCAA 2020 Exam 2 Data analysis Q6d

The average mark on this question was **0.6**.

Questions from multiple lessons

Data analysis

14. A least squares line is fitted to a set of bivariate data. If the response and explanatory variables are reversed on the set of axes, which of the following statistics would not change?

A. The slope of the least squares line

B. The residual values

C. The equation of the least squares line

D. The correlation coefficient

E. The y-intercept of the least squares line

Adapted from VCAA 2018 Exam 1 Data analysis Q14

Recursion and financial modelling *Year 11 content*

15. A sequence can be generated by the following recurrence relation.

$V_{n+1} = 1.5 \times V_n, \quad V_0 = 10$

Which of the following rules can be used to find the term in the sequence after n iterations of the rule?

A. $V_n = 1.5 \times 10$ n

B. $V_n = 1.5 + 10^n$

C. $V_n = 10 \times 1.5n$

D. $V_n = 10 \times 1.5^n$

E. $V_n = 10 + 1.5^n$

Adapted from VCAA 2013 Exam 1 Number patterns Q5

Back to contents

Data analysis

16. The following table shows the number of *words written* and *time spent*, in minutes, on eight English essays.

***time spent* (mins)**	93	144	58	220	195	138	87	104
words written	811	997	440	1543	1230	1114	690	758

a. A least squares regression line is fitted to the data and has a y-intercept of 159.058781.
Round this value to four significant figures. (1 MARK)

b. Use the values in the table to calculate the equation of the least squares regression line, where *time spent* is the explanatory variable. Round all values correct to four significant figures. (2 MARKS)

Adapted from VCAA 2018 Exam 2 Data analysis Q3c,d

Back to contents

4

CHAPTER 4

Investigating and modelling time series data

LESSONS

KEY KNOWLEDGE

- 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.

Back to contents

4A Time series data and their graphs

STUDY DESIGN DOT POINT

- 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

4A 4B 4C 4D 4E

KEY SKILLS

During this lesson, you will be:

- constructing time series plots
- identifying characteristics of time series data.

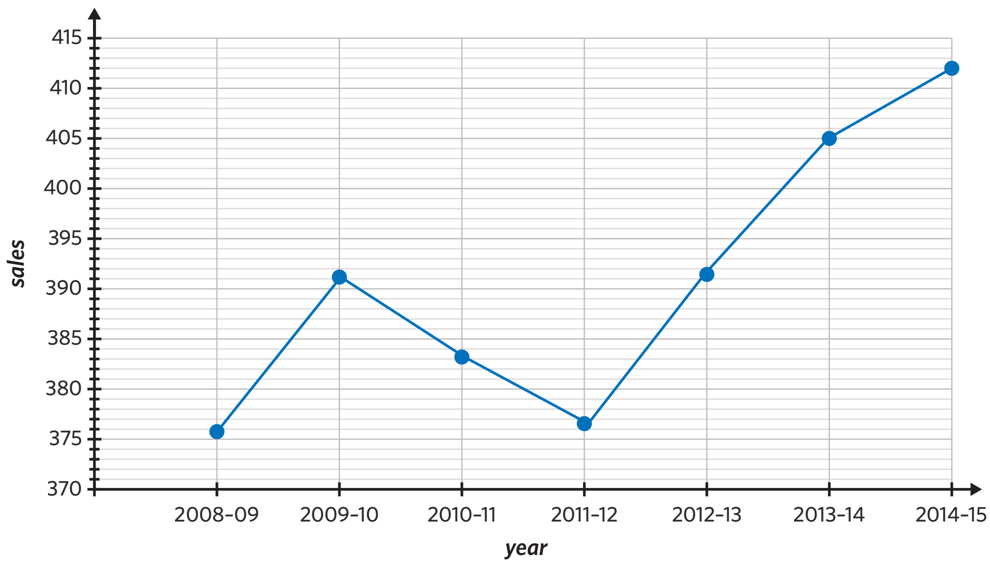
KEY TERMS

- Time series
- Coded time
- Trend
- Period
- Cyclical variation
- Seasonality
- Structural change
- Irregular fluctuations

The changes in a variable over time can be represented by a time series graph. Time can be expressed in many ways – as hours, days, weeks, months, years, seasons, etc. By graphing time series data, any patterns or changes in the response variable over time can be observed.

Constructing time series plots

Time series data is a subset of bivariate data where the explanatory variable is always time. Since time is the explanatory variable, it is always plotted on the horizontal axis. When plotting a time series, the known data points are connected by straight lines. These lines are useful in identifying any trends or changes in the data.



It can sometimes be inefficient to use the full dates when plotting time series data. Instead, simple numerical representations such as 1, 2, 3, 4... can be used. The simplified numerical representation of time is referred to as the **coded time**. Calculations to the time series can also be applied. For example, summer 2022, autumn 2022, winter 2022, and spring 2022 can be represented by the codes 1, 2, 3, and 4.

Back to contents

Worked example 1

The number of *deaths and serious injuries* caused by road accidents in the UK for each month from October 1982 until the end of 1983 are shown in a table.

month	Oct 1982	Nov 1982	Dec 1982	Jan 1983	Feb 1983	Mar 1983	Apr 1983	May 1983	Jun 1983	Jul 1983	Aug 1983	Sep 1983	Oct 1983	Nov 1983	Dec 1983
deaths and serious injuries	1850	1998	2079	1494	1057	1218	1168	1236	1076	1174	1139	1427	1487	1483	1513

a. Use the table to construct a time series plot.

Explanation – Method 1: By hand

Step 1: Create a time code.

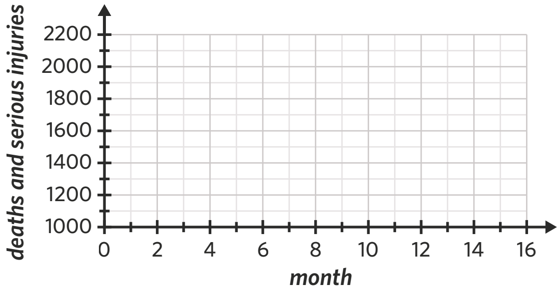
month	*coded month*
Oct 1982	1
Nov 1982	2
Dec 1982	3
Jan 1983	4
Feb 1983	5
Mar 1983	6
Apr 1983	7
May 1983	8
Jun 1983	9
Jul 1983	10
Aug 1983	11
Sep 1983	12
Oct 1983	13
Nov 1983	14
Dec 1983	15

Step 2: Draw a set of axes and label the horizontal axis '*month*' and the vertical axis '*deaths and serious injuries*'.

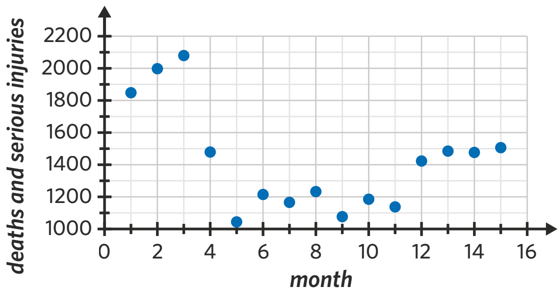
Step 3: Construct an appropriate scale for each axis.

The horizontal axis should range from at least 0 to 15.

The vertical axis should range from at least 1000 to 2100.

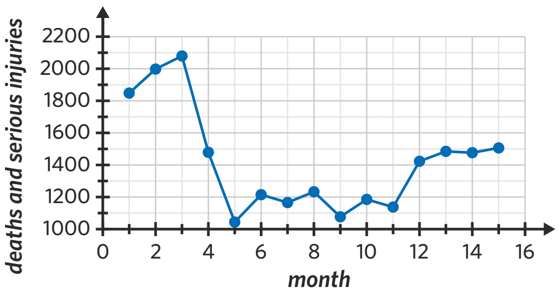


Step 4: Plot each data point on the graph.



Step 5: Connect the data points in order with straight lines.

Answer



Continues →

Back to contents

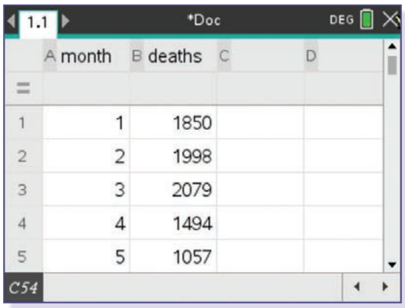
Explanation – Method 2: TI-Nspire

Step 1: Create a time code.

month	*coded month*
Oct 1982	1
Nov 1982	2
Dec 1982	3
Jan 1983	4
Feb 1983	5
Mar 1983	6
Apr 1983	7
May 1983	8
Jun 1983	9
Jul 1983	10
Aug 1983	11
Sep 1983	12
Oct 1983	13
Nov 1983	14
Dec 1983	15

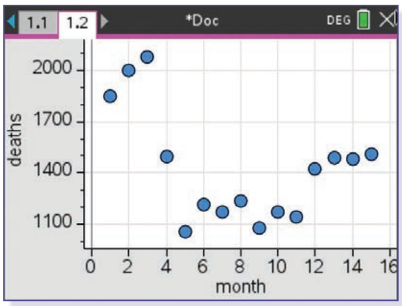
Step 2: From the home screen, select ‘1: New’ → ‘4: Add Lists & Spreadsheet’.

Step 3: Name column A ‘month’ and enter the data values starting from row 1 into the column below. Name column B ‘deaths’ and enter the data values starting from row 1 into the column below.



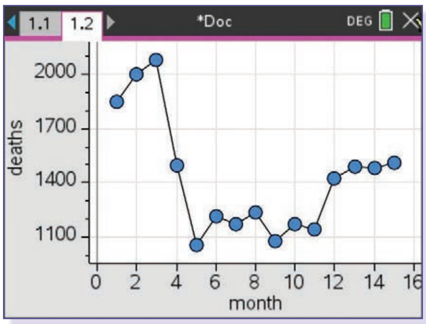
Step 4: Press ctrl + doc▾ and select ‘5: Add Data & Statistics’.

Step 5: Move to the horizontal axis and select ‘Click to add variable’ → ‘month’. Move to the vertical axis and select ‘Click to add variable’ → ‘deaths’.



Step 6: Press menu. Select ‘2: Plot Properties’ → ‘1: Connect Data Points’.

Answer



Continues →

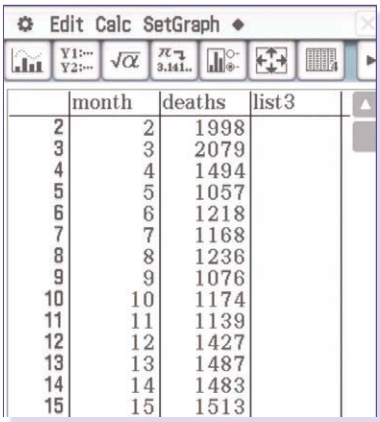
Explanation – Method 3: Casio ClassPad

Step 1: Create a time code.

month	*coded month*
Oct 1982	1
Nov 1982	2
Dec 1982	3
Jan 1983	4
Feb 1983	5
Mar 1983	6
Apr 1983	7
May 1983	8
Jun 1983	9
Jul 1983	10
Aug 1983	11
Sep 1983	12
Oct 1983	13
Nov 1983	14
Dec 1983	15

Step 2: From the main menu, tap Statistics.

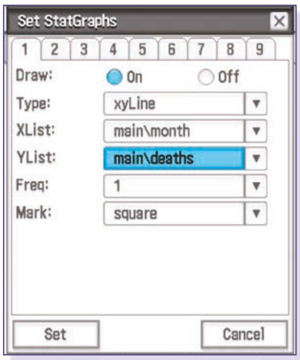
Step 3: Name list1 ‘month’ and enter the data values starting from row 1 into the column below. Name list2 ‘deaths’ and enter the data values starting from row 1 into the column below.



Step 4: Configure the settings of the graph by tapping .

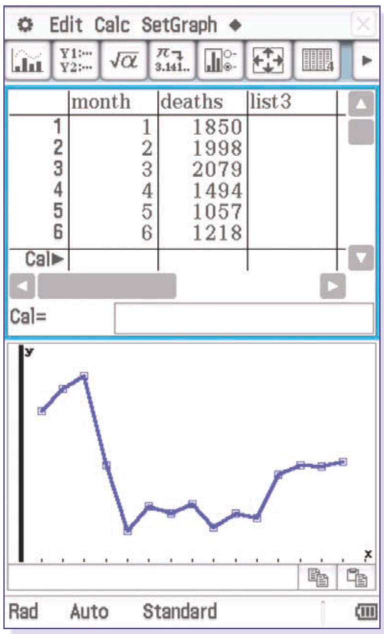
Step 5: Create a time series plot by changing ‘Type’ to ‘xyLine’.

Step 6: Specify the data set by changing ‘XList:’ to ‘main\month’ and ‘YList:’ to ‘main\deaths’.



Step 7: Tap ‘Set’ to confirm and then to plot the graph.

Answer



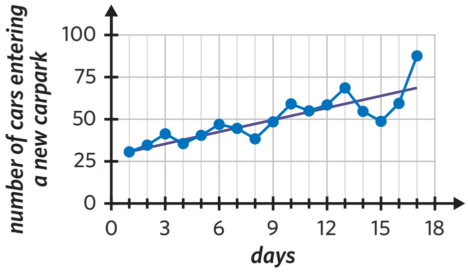
Identifying characteristics of time series data

There are many types of fluctuations or patterns that can be observed in time series data.

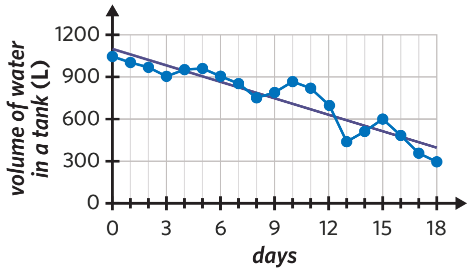
Trends

A **trend** is a general upwards (increasing) or downwards (decreasing) movement over time. This movement can be represented using a trend line.

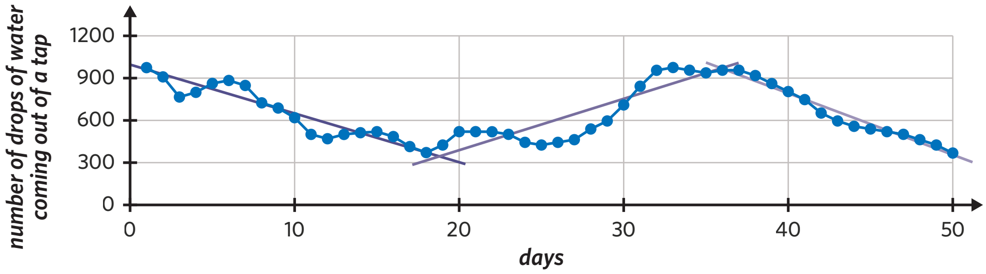
Increasing trend:



Decreasing trend:



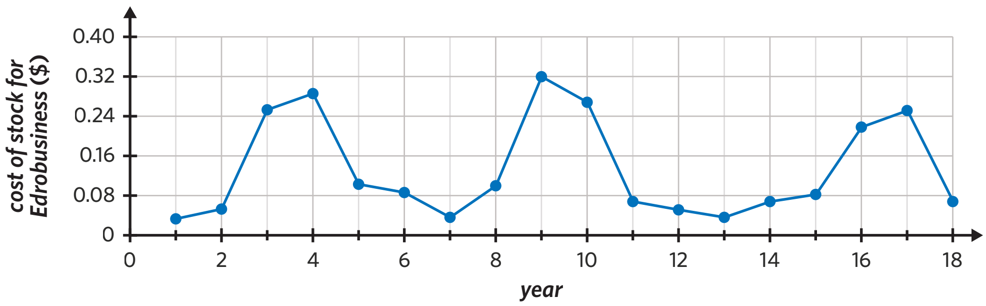
Sometimes, a time series may have multiple trends that change over time.



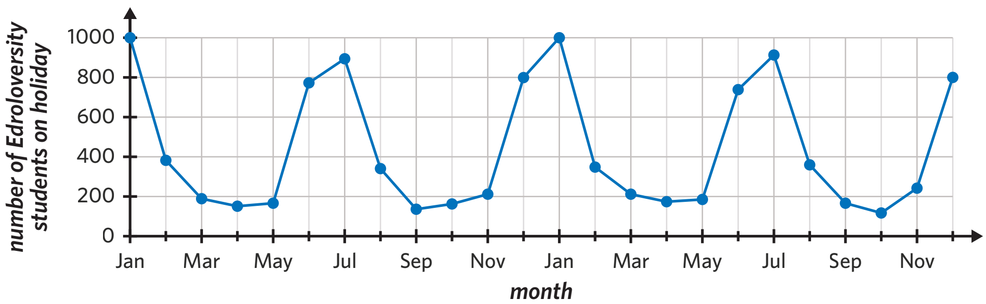
Seasonality & cyclical variation

When a graph shows numerous fluctuations, or peaks and troughs, the **period** is defined as the distance between two adjacent peaks. A period may be regular or irregular. The length of a period can be useful in identifying whether a graph demonstrates cyclical variation or seasonality.

When a time series plot rises and falls with a regular pattern over some time period, it demonstrates **cyclical variation**, or cycles. Whilst the peaks of cycles occur at approximately the same intervals, cycles can have a period which changes slightly between peaks.



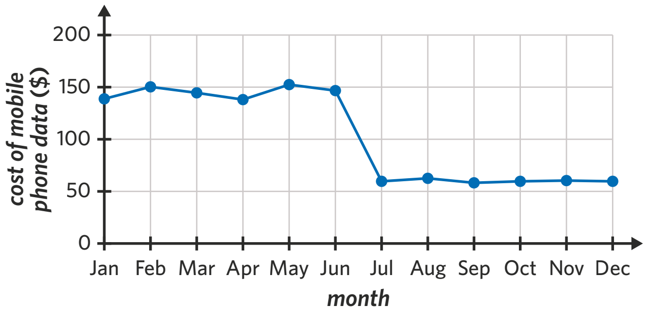
When a time series plot shows cyclical variation within a calendar-related period (e.g. week, month, quarter), the plot demonstrates **seasonality**. Note that the name seasonal does not have to mean the seasons of the year. A seasonal time series plot has regular peaks and troughs that occur at the same time each period, and the length of the period must be a year or less.



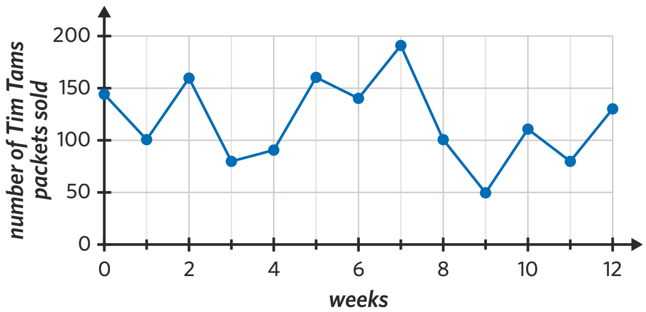
Back to contents

Structural change, irregular fluctuations & outliers

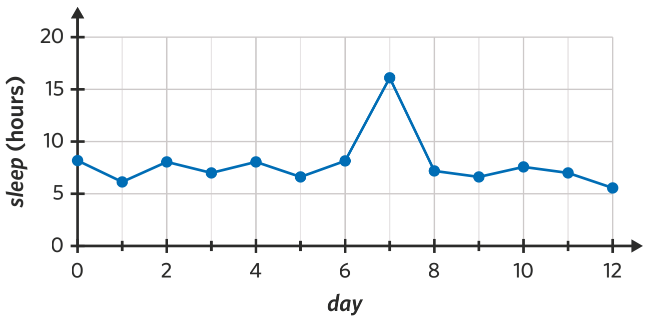
A **structural change** occurs when the established pattern of a time series plot is suddenly altered significantly for some reason.



Irregular fluctuations are random variations in a time series plot that cannot be explained by trend, seasonality, cycles or structural change. Irregular fluctuations cannot be predicted.



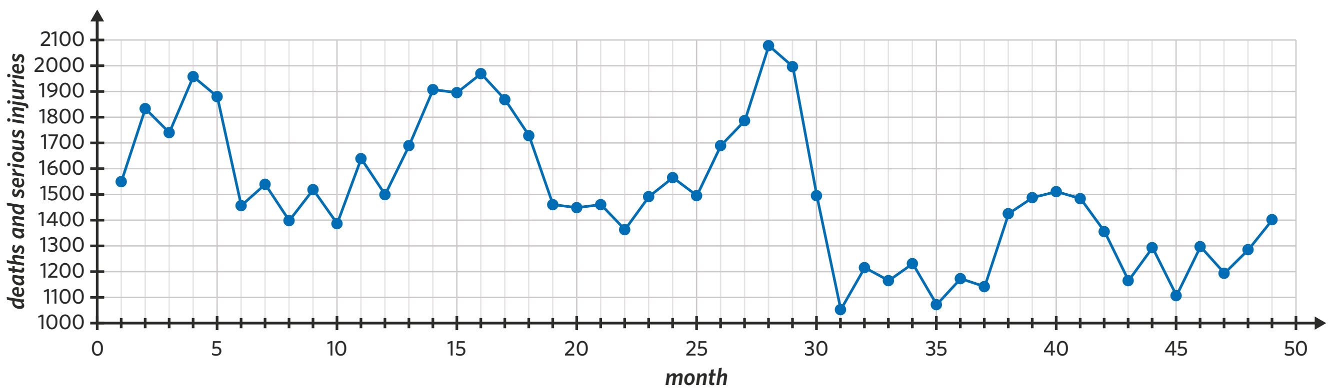
Outliers are values which fall outside of what looks normal or reasonable. There may be a situational explanation for the presence of an outlier.



Worked example 2

The number of *deaths and serious injuries* caused by road accidents in the UK for each month over a four-year period, from September 1980 until September 1984, are shown in the time series plot.

Note that seat belt use became mandatory by law in the UK in February 1983.



Describe any patterns present in the time series plot.

Continues →

Back to contents

Explanation

Step 1: Look for a general trend in the data.

There is no general trend in this time series plot. The peaks and troughs in the data show no clear increasing or decreasing trend.

Step 2: Look for any patterns in the data.

This time series plot has a clear pattern of regular peaks and troughs. The peaks occur every 12 months and are generally at their highest point in December (note that the time series plot begins in September). The fluctuations follow a calendar-related period suggesting that the pattern shows seasonality and not cyclical variation.

The time series plot also has a clear structural change shown by a decrease in *deaths and serious injuries* from month 29 to month 31. This structural change can be explained by the introduction of seatbelt laws in February 1983 (month 30).

No notable irregular fluctuations or outliers can be found in the time series.

Step 3: Present your findings as a brief report.

Answer

The time series plot has no obvious trend. The plot shows seasonality with a period of 12 months peaking every December. There is also a structural change in the plot that occurs between month 29 and month 31. This structural change can be explained by the introduction of mandatory seat belt usage laws in February 1983.

Exam question breakdown

VCAA 2017 Exam 1 Data analysis Q13

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.

The time series is best described as having

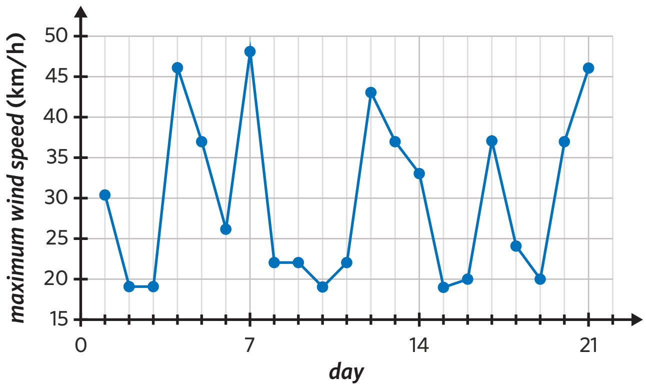
A. seasonality only.

B. irregular fluctuations only.

C. seasonality with irregular fluctuations.

D. a decreasing trend with irregular fluctuations.

E. an increasing trend with irregular fluctuations.



Explanation

Step 1: Look for a general trend in the data.

There is no general trend in this time series plot. The peaks and troughs in the data show no clear pattern of increase or decrease.

Step 2: Look for any seasonality.

Since we have been given data spanning 21 days, the only feasible calendar-related period we could identify would be weekly. In order for there to be weekly seasonality, the peaks and troughs would need to occur at the same time each week (for example, day 1, day 8, and day 15). They do not occur at the same time each week, so there is no seasonality present.

Step 3: Describe the time series.

The time series has no general trend or seasonality. It also has no clear outliers, structural change, or cyclical variation. The time series can be described as having irregular fluctuations only.

Answer

B

51% of students answered this question correctly.

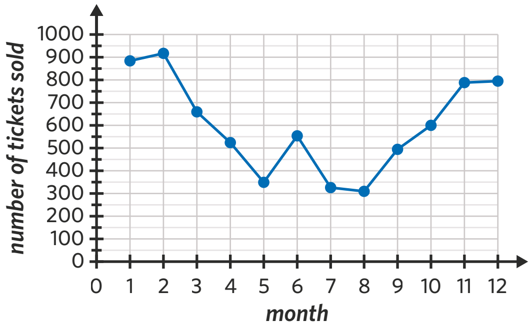
32% of students incorrectly chose option C. These students incorrectly identified the time series graph as showing seasonality. Although there are peaks and troughs in the time series, they do not exist with regular calendar-related intervals of time between them, so seasonality cannot be concluded.

Back to contents

4A Questions

Constructing time series plots

1. Consider the following time series showing the number of tickets sold to Melbourne Zoo's hippopotamus show 'Hypnotic Hippos' each month this year. The *month* data has been represented by a time code, so that January is 1, February is 2, etc.



How many tickets were sold in June?

A. 320 tickets

B. 345 tickets

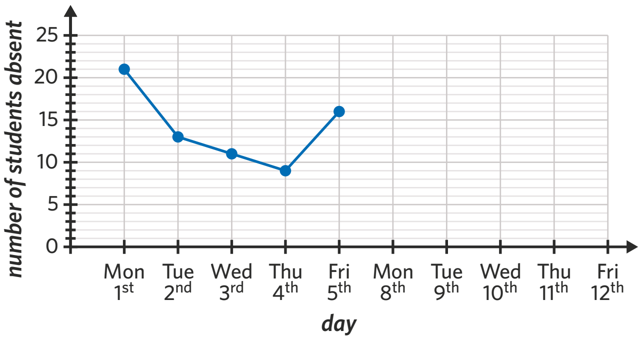
C. 550 tickets

D. 910 tickets

2. The *number of students absent* from school each day was recorded over two weeks in April.

day	Mon 1st	Tue 2nd	Wed 3rd	Thu 4th	Fri 5th	Mon 8th	Tue 9th	Wed 10th	Thu 11th	Fri 12th
number of students absent	21	13	11	9	16	20	8	14	11	18

The data from the first week has been plotted onto the following graph. Complete the time series by plotting the data from the second week of April.



3. The number of *climate change sceptics* (%), as a percentage of the total population, each year from 2000–2013 is displayed in the table. Use the table to construct a time series plot using a calculator.

year	*coded year*	*climate change sceptics* (%)
2000	1	62
2001	2	58
2002	3	57
2003	4	54
2004	5	51
2005	6	52
2006	7	40
2007	8	39
2008	9	37
2009	10	36
2010	11	37
2011	12	34
2012	13	35
2013	14	32

4. The *earnings* ($000's) of a casino were recorded for each month in 1983–84 and are given in the table. Construct a time series plot displaying this data by hand.

month	*earnings* ($000's)
Jan 1983	2400
Feb 1983	1600
Mar 1983	3200
Apr 1983	2100
May 1983	1400
Jun 1983	2200
Jul 1983	2700
Aug 1983	3200
Sep 1983	2500
Oct 1983	1900
Nov 1983	2400
Dec 1983	1600

month (cont'd)	*earnings* ($000's)
Jan 1984	1300
Feb 1984	2100
Mar 1984	2300
Apr 1984	2000
May 1984	1500
Jun 1984	1700
Jul 1984	2600
Aug 1984	3100
Sep 1984	2700
Oct 1984	2900
Nov 1984	3200
Dec 1984	2400

Identifying characteristics of time series data

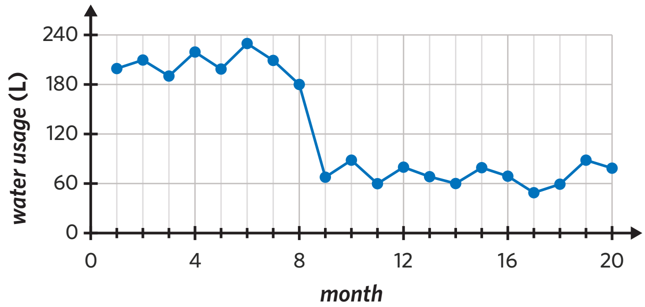
5. Identify which one of the following features is present in the time series plot.

A. Structural change

B. Outlier

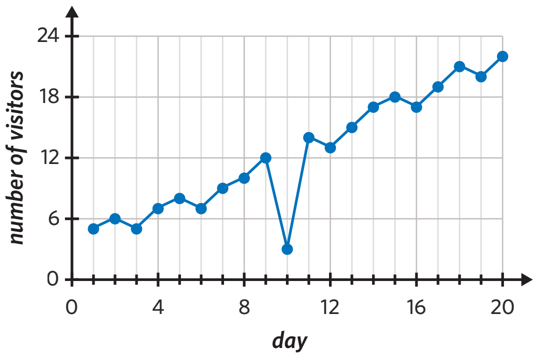
C. Seasonality

D. A decreasing trend

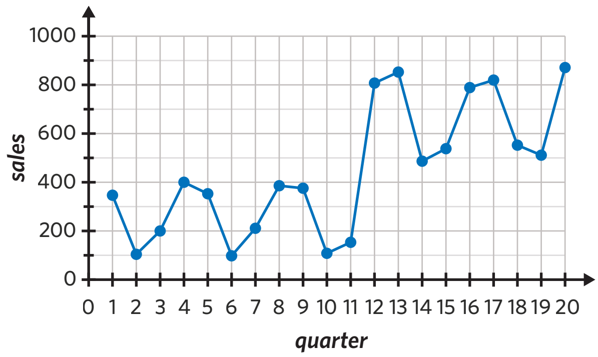


6. Identify two features that are present in each of the following time series plots.

a.

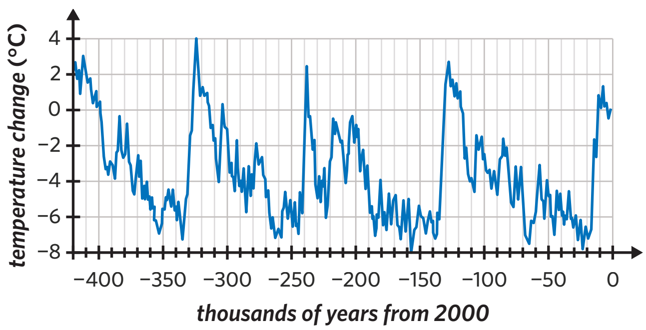


b.



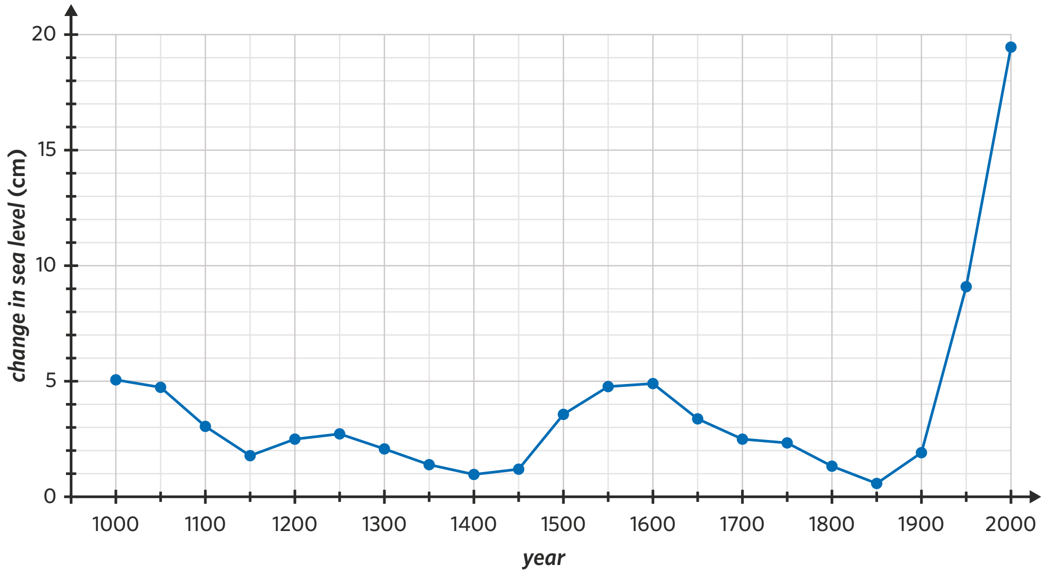
7. The following graph shows the change in global temperature (°C) since approximately 400 000 years before the year 2000.

Identify any trends or patterns in the data.



Back to contents

8. The following graph shows the *change in sea level* (cm) since the year 1000.



Identify any trends or patterns in the data.

Joining it all together

9. The following table shows the yearly *revenue* of a restaurant from 2012 to 2022.

year	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
revenue ($000's)	950	890	935	830	850	760	810	740	350	710	690

a. Construct a time series plot by hand displaying the restaurant's *revenue* between 2012 and 2022.

b. Briefly describe any patterns or characteristics in the time series plot.

10. The profits between 2018 and 2021 of a buy-now pay-later business are shown in the following table.

a. Complete the time code by filling in the 'coded season' column.

b. Construct a time series plot using a calculator to display the data.

c. Describe any trends or patterns in the time series plot between Summer 2018 and Summer 2020, inclusive.

d. Describe any characteristics present from Summer 2020 to Autumn 2020.

e. Describe any trends or patterns in the time series plot after Autumn 2020.

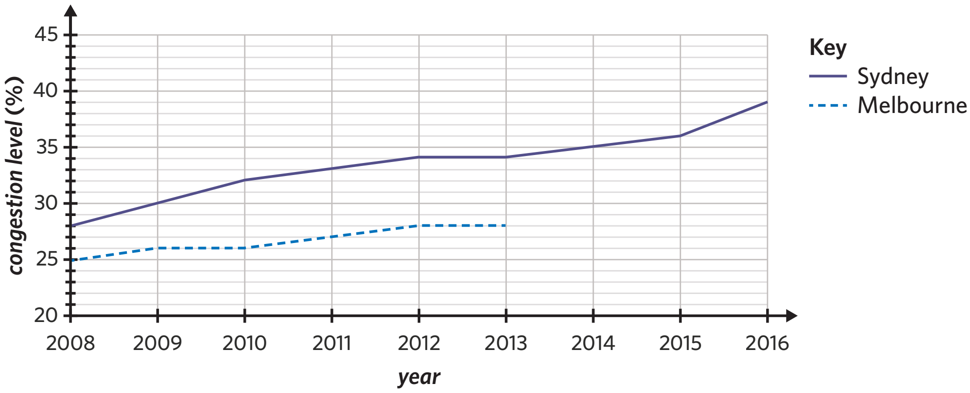
season	*coded season*	*profit* ($000 000's)
Summer 2018		18
Autumn 2018		12
Winter 2018		13
Spring 2018		17
Summer 2019		22
Autumn 2019		16
Winter 2019		15
Spring 2019		21
Summer 2020		27
Autumn 2020		47
Winter 2020		50
Spring 2020		48
Summer 2021		47
Autumn 2021		43
Winter 2021		41
Spring 2021		40

Back to contents

Exam practice

11. The following 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



Use the data in the table to complete the time series plot for Melbourne. (1 MARK)

VCAA 2018 Exam 2 Data analysis Q3a

88% of students answered this question correctly.

---

12. Consider the time series plot.

The pattern in the time series plot shown is best described as having

A. irregular fluctuations only.

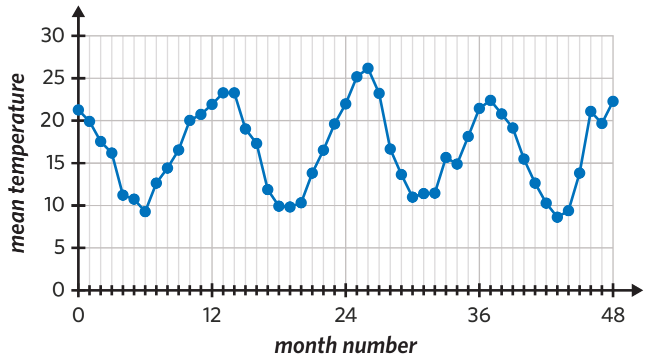
B. an increasing trend with irregular fluctuations.

C. seasonality with irregular fluctuations.

D. seasonality with an increasing trend and irregular fluctuations.

E. seasonality with a decreasing trend and irregular fluctuations.

VCAA 2016 Exam 1 Data analysis Q13



72% of students answered this question correctly.

---

13. 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.

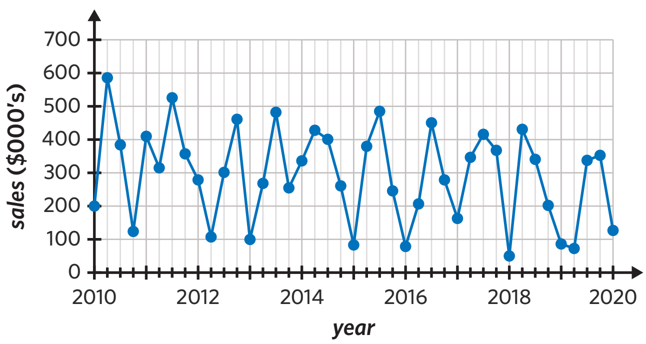
B. irregular fluctuations only.

C. seasonality with irregular fluctuations.

D. a decreasing trend with irregular fluctuations.

E. a decreasing trend with seasonality and irregular fluctuations.

VCAA 2021 Exam 1 Data analysis Q12



34% of students answered this question correctly.

Back to contents

Questions from multiple lessons

Data analysis *Year 11 content*

14. There is found to be a strong negative association between the time spent on social media and average test score of a sample of students. Which of the following conclusions can be drawn from this information?

A. A decrease in the average test score of a student will cause a decrease in their time spent on social media.

B. An increase in the average test score of a student will cause a decrease in their time spent on social media.

C. Students who spend more time on social media tend to have lower average test scores.

D. Students who spend less time on social media tend to have lower average test scores.

E. The association must be due to coincidence.

Adapted from VCAA 2016 Exam 1 Data analysis Q12

Recursion and financial modelling *Year 11 content*

15. Alessandro deposited \$4400 into a savings account with an interest rate of 3.5% per annum, compounding annually. Which of the following recurrence relations models the balance of the savings account, B_n, after n years?

A. $B_0 = 4400, \quad B_{n+1} = 3.5 \times B_n$

B. $B_0 = 4400, \quad B_{n+1} = B_n + 3.5 \times 154$

C. $B_0 = 4400, \quad B_{n+1} = B_n + 154$

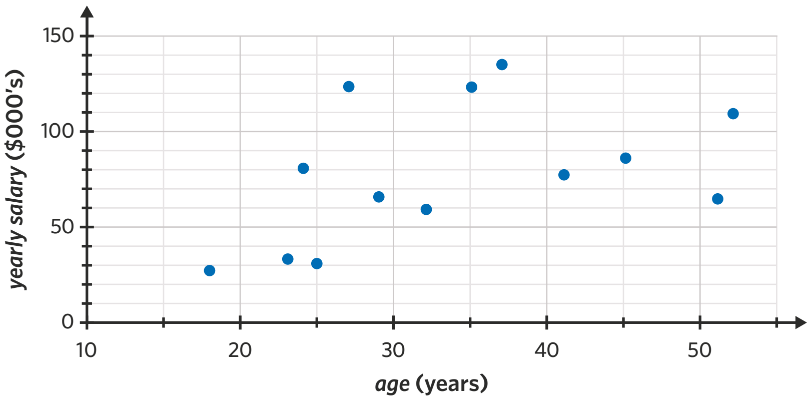
D. $B_0 = 4400, \quad B_{n+1} = 154 \times B_n$

E. $B_0 = 4400, \quad B_{n+1} = 1.035 \times B_n$

Adapted from VCAA 2017NH Exam 1 Recursion and financial modelling Q18

Data analysis

16. The *age*, in years, and *yearly salary*, in dollars, of 13 office workers are shown in the following scatterplot and table.



age (years)	*yearly salary* (\$)
18	27 000
23	33 500
24	81 000
25	31 000
27	123 000
29	66 000
32	59 000
35	122 500
37	134 500
41	77 000
45	86 000
51	64 500
52	109 000

In this sample, the relationship between *age* and *yearly salary* is non-linear.

In order to linearise the data, a log transformation can be applied to the variable *age*.

a. Apply the log transformation and calculate the equation of the least squares regression line that predicts *yearly salary* from $\log_{10}(age)$. Give values correct to one decimal place. (1 MARK)

b. Using this rounded regression equation, predict the *yearly salary* of someone 33 years of age. Give your answer correct to the nearest dollar. (1 MARK)

Adapted from VCAA 2014 Exam 2 Data analysis Q3

Back to contents

4B Smoothing – moving means

STUDY DESIGN DOT POINT

- 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

4A 4B 4C 4D 4E

KEY SKILLS

During this lesson, you will be:

- smoothing over an odd number of data points using moving means
- smoothing over an even number of data points using moving means
- plotting and interpreting a mean smoothed time series.

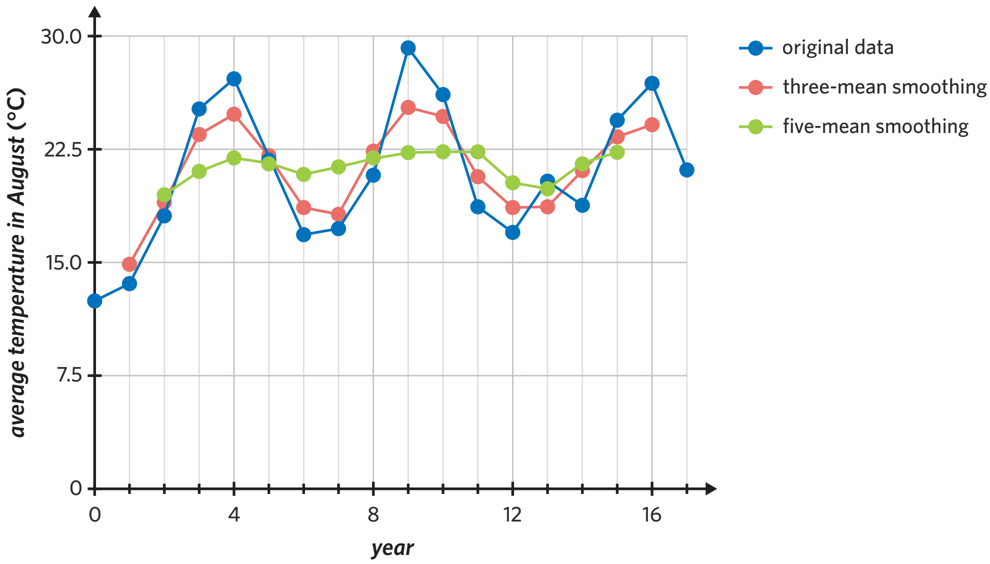
KEY TERMS

- Smoothing
- Moving mean smoothing
- Centring

When a time series contains many fluctuations, it is often difficult to identify the underlying trend. In order to reveal the trend, smoothing can be implemented to remove some of the fluctuations. Moving mean smoothing is one method used to smooth time series data.

Smoothing over an odd number of data points using moving means

Smoothing is the process of removing fluctuations in time series data to reveal any underlying trends. It is possible to smooth using either moving means or moving medians. Smoothing using means is known as **moving mean smoothing**.



Three-mean smoothing involves replacing each data value with the mean of itself and each adjacent value.

Back to contents

The smoothed value of y_2:

$$smoothed\ y_2 = \frac{y_1 + y_2 + y_3}{3}$$

There are no smoothed values for the first and last data entries because they do not have a value on either side.

day	***temp.* (°C)**	***calculation***	**three-mean smoothed *temperature* (°C)**
Mon	24	–	–
Tue	27	$\frac{24 + 27 + 21}{3}$	24
Wed	21	$\frac{27 + 21 + 18}{3}$	22
Thu	18	$\frac{21 + 18 + 15}{3}$	18
Fri	15	$\frac{18 + 15 + 15}{3}$	16
Sat	15	$\frac{15 + 15 + 12}{3}$	14
Sun	12	–	–

Five-mean smoothing is similar to three-mean smoothing but uses five values. Each data value is replaced with the mean of itself and the two values on each side of it.

The smoothed value of y_3:

$$smoothed\ y_3 = \frac{y_1 + y_2 + y_3 + y_4 + y_5}{5}$$

There are no smoothed values for the first two and last two data entries because they do not have two values on either side.

day	***temp.* (°C)**	***calculation***	**five-mean smoothed *temperature* (°C)**
Mon	24	–	–
Tue	27	–	–
Wed	21	$\frac{24 + 27 + 21 + 18 + 15}{5}$	21
Thu	18	$\frac{27 + 21 + 18 + 15 + 15}{5}$	19.2
Fri	15	$\frac{21 + 18 + 15 + 15 + 12}{3}$	16.2
Sat	15	–	–
Sun	12	–	–

Moving mean smoothing can be extended over any number of data points. Smoothing over any odd number of data points is done in the same way as three-mean and five-mean smoothing.

Worked example 1

The following table shows the daily *rainfall* in Dunedin for a week.

day	Mon	Tue	Wed	Thu	Fri	Sat	Sun
***rainfall* (mm)**	2	3	5	5	1	0	2
three-mean smoothed *rainfall*	–						–
five-mean smoothed *rainfall*	–	–				–	–

Continues →

Back to contents

a. Find the three-mean smoothed *rainfall* for Wednesday, correct to one decimal place.

Explanation

Step 1: Find the *rainfall* for Wednesday as well as its adjacent values, and write them in the order they appear in the time series.

3 5 5

Step 2: Find the mean of these three values.

$$mean = \frac{3 + 5 + 5}{3}$$
$$= 4.333...$$

Answer

4.3 mm

---

b. Find the three-mean smoothed values, correct to one decimal place, for the entire time series and fill in the table.

Explanation – Method 1: TI-Nspire

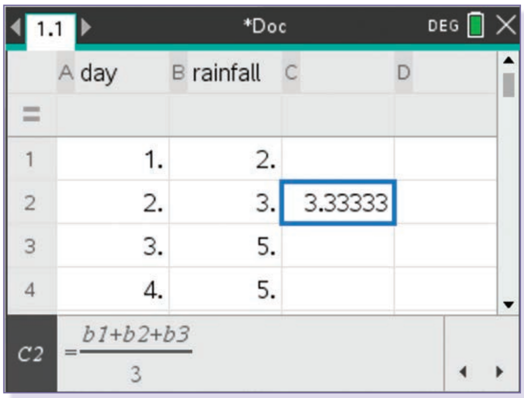
Step 1: From the home screen, select '1: New' → '4: Add Lists & Spreadsheet'.

Step 2: Enter the numbers from 1 to 7 in the first column. Each number will represent a day of the week. Name the column 'day'.

Step 3: Enter the *rainfall* values into the second column and name the column 'rainfall'.

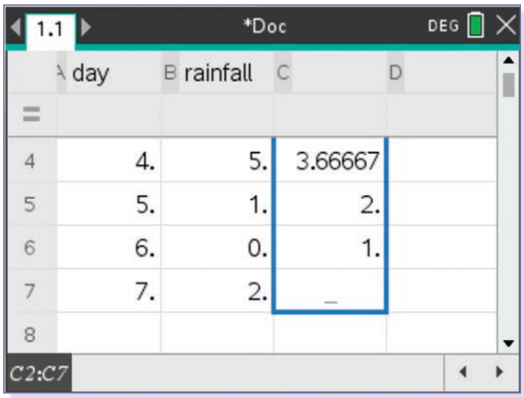
Step 4: Select cell C2. Type '=(b1+b2+b3)/3' to find the smoothed *rainfall* for day 2 (Tuesday).

Press enter.



Note: C1 is left blank as there is no three-mean smoothed value for the first data point.

Step 5: With cell C2 still selected, move the cursor to the bottom right corner of the cell and click and drag downwards to apply the formula to the remaining rows.



Explanation – Method 2: Casio ClassPad

Step 1: From the main menu, tap Spreadsheet.

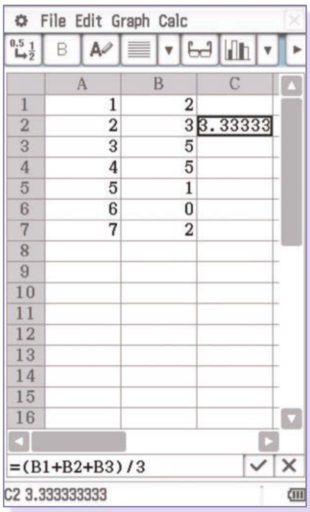
Step 2: Enter the numbers from 1 to 7 in the first column. Each number will represent a day of the week.

Step 3: Enter the *rainfall* values into the second column.

Continues →

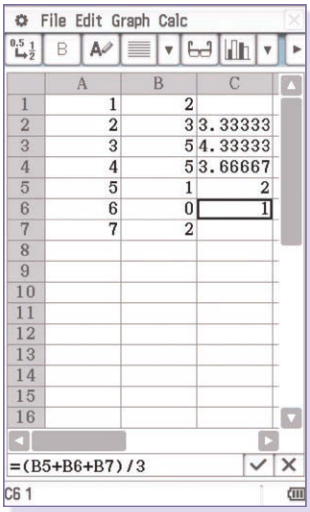
Back to contents

Step 4: Select cell C2. Type '=(b1+b2+b3)/3' to find the smoothed *rainfall* for day 2 (Tuesday). Press **EXE**.



Note: C1 is left blank as there is no three-mean smoothed value for the first data point.

Step 5: With cell C2 still selected, drag from C2 down to C3. This will copy the formula down to the next cell. Repeat this until cell C6.



Answer – Method 1 and 2

day	Mon	Tue	Wed	Thu	Fri	Sat	Sun
***rainfall* (mm)**	2	3	5	5	1	0	2
three-mean smoothed *rainfall*	–	3.3	4.3	3.7	2.0	1.0	–

Note: Alternatively, this question could be solved manually using the technique from part **a**.

c. Find the five-mean smoothed *rainfall* for Wednesday.

Explanation

Step 1: Find the *rainfall* for Wednesday as well as the two values on each side of it, and write them in the order they appear in the time series.

2 3 5 5 1

Step 2: Find the mean of these five values.

$$mean = \frac{2 + 3 + 5 + 5 + 1}{5} = 3.2$$

Answer

3.2 mm

d. Find the five-mean smoothed values, correct to one decimal place, for the entire time series and fill in the table.

Explanation – Method 1: TI-Nspire

Step 1: From the home screen, select '1: New' → '4: Add Lists & Spreadsheet'.

Step 2: Enter the numbers 1 to 7 in the first column. Each number will represent a day of the week. Name the column 'day'.

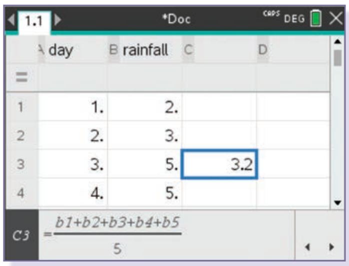
Step 3: Enter the *rainfall* values into the second column and name the column 'rainfall'.

Continues →

Back to contents

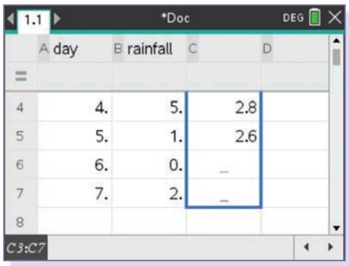
Step 4: Select cell C3. Type '=(b1+b2+b3+b4+b5)/5' to find the smoothed *rainfall* for day 3 (Wednesday).

Press enter.



Note: C1 and C2 are left blank as there are no five-mean smoothed values for the first two data points.

Step 5: With cell C3 still selected, move the cursor to the bottom right corner of the cell and click and drag downwards to apply the formula to the remaining rows.



Explanation – Method 2: Casio ClassPad

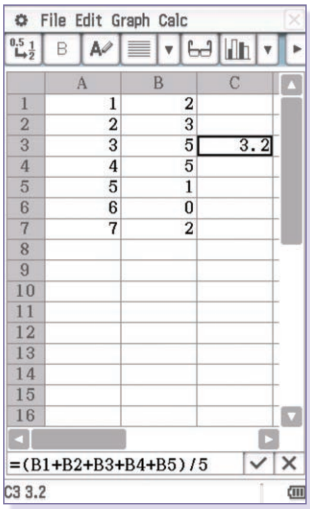
Step 1: From the main menu, tap Spreadsheet.

Step 2: Enter the numbers 1 to 7 in the first column. Each number will represent a day of the week.

Step 3: Enter the *rainfall* values into the second column.

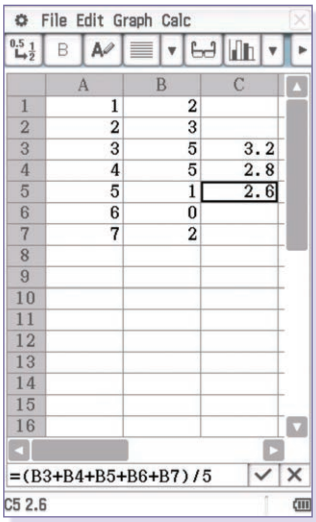
Step 4: Select cell C3. Type '=(b1+b2+b3+b4+b5)/5' to find the smoothed *rainfall* for day 3 (Wednesday).

Press EXE.



Note: C1 and C2 are left blank as there are no five-mean smoothed values for the first two data points.

Step 5: With cell C3 still selected, drag from C3 down to C4. This will copy the formula down to the next cell. Repeat this for cell C5.



Answer – Method 1 and 2

day	Mon	Tue	Wed	Thu	Fri	Sat	Sun
rainfall **(mm)**	2	3	5	5	1	0	2
five-mean smoothed *rainfall*	–	–	3.2	2.8	2.6	–	–

Note: Alternatively, this question could be solved manually using the technique from part **c**.

Back to contents

Smoothing over an even number of data points using moving means

Moving mean smoothing over an even number of data points is slightly more complicated because the centre of each set of points is halfway between two data points, and can therefore not be plotted on the original time series. A process called **centring** must be applied in order to align the data with the time series.

Two-mean smoothing with centring involves first calculating two non-centred means. These are the mean of the original data point and each of its adjacent values, in two separate calculations. The centred mean of the original data point is then found from the mean of the two non-centred means.

To find the smoothed value of y_2:

$$\textit{smoothed } y_2 = \frac{\frac{y_1 + y_2}{2} + \frac{y_2 + y_3}{2}}{2}$$

This formula can be simplified to:

$$\textit{smoothed } y_2 = \frac{y_1 + 2y_2 + y_3}{4}$$

There are no smoothed values for the first and last data entries because they do not have a value on either side.

day	***temp. (°C)***	***before centring***	***after centring***
Mon	24		–
		$\frac{24 + 27}{2} = 25.5$	
Tue	27		$\frac{25.5 + 24}{2} = 24.75$
		$\frac{27 + 21}{2} = 24$	
Wed	21		$\frac{24 + 19.5}{2} = 21.75$
		$\frac{21 + 18}{2} = 19.5$	
Thu	18		–

Four-mean smoothing with centring is similar to two-mean smoothing but uses four values in the mean calculations. The non-centred means are each calculated from four adjacent values surrounding the original data point. The centred mean of the original data point is then found from the mean of the two non-centred means.

To find the smoothed value of y_3:

$$\textit{smoothed } y_3 = \frac{\frac{y_1 + y_2 + y_3 + y_4}{4} + \frac{y_2 + y_3 + y_4 + y_5}{4}}{2}$$

This formula can be simplified to:

$$\textit{smoothed } y_3 = \frac{y_1 + 2y_2 + 2y_3 + 2y_4 + y_5}{8}$$

There are no smoothed values for the first two and last two data entries because they do not have two values on either side.

Back to contents

day	*temp.* (°C)	*before centring*	*after centring*
Mon	24		–
Tue	27		–
		$\frac{24 + 27 + 21 + 18}{4} = 22.5$	
Wed	21		$\frac{22.5 + 20.25}{2} = 21.375$
		$\frac{27 + 21 + 18 + 15}{4} = 20.25$	
Thu	18		$\frac{20.5 + 17.25}{2} = 18.75$
		$\frac{21 + 18 + 15 + 15}{4} = 17.25$	
Fri	15		$\frac{17.25 + 15}{2} = 16.125$
		$\frac{18 + 15 + 15 + 12}{4} = 15$	
Sat	15		–
Sun	12		–

Smoothing over any even number of data points is done in the same way as two-mean and four-mean smoothing.

Worked example 2

Amelia recorded the distance, in km, she ran each month for a year.

month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
***distance run* (km)**	42	47	39	41	37	35	36	28	31	28	35	41
two-mean smoothed *distance run* with centring (km)	–											–
four-mean smoothed *distance run* with centring (km)	–	–									–	–

a. Find the two-mean smoothed *distance run* for July.

Explanation

Step 1: Find the *distance run* for July as well as its adjacent values and write them in the order they appear in the time series.

35 36 28

Step 2: Find the mean of the first two values.

$$mean_1 = \frac{35 + 36}{2} = 35.5$$

Step 3: Find the mean of the last two values.

$$mean_2 = \frac{36 + 28}{2} = 32$$

Step 4: Calculate the centred mean.

$$mean_{centred} = \frac{35.5 + 32}{2} = 33.75$$

Answer

33.75 km

Continues →

Back to contents

b. Find the two-mean smoothed values, correct to one decimal place, for the entire time series and fill in the table.

Explanation – Method 1: TI-Nspire

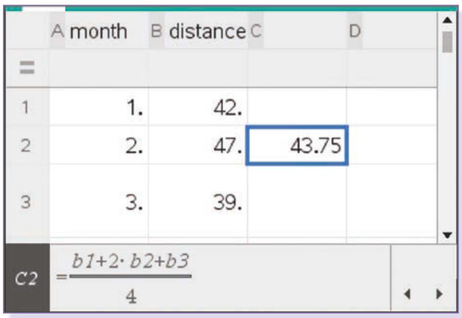
Step 1: From the home screen, select '1: New' → '4: Add Lists & Spreadsheet'.

Step 2: Enter the numbers from 1 to 12 in the first column. Each number will represent a month. Name the column 'month'.

Step 3: Enter the *distance run* values into the second column and name the column 'distance'.

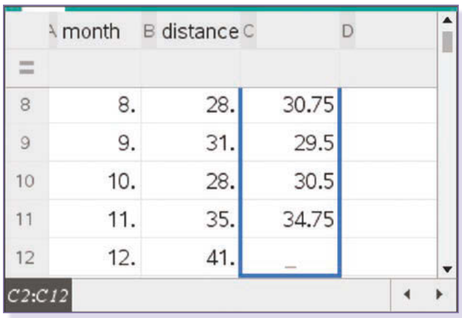
Step 4: Select cell C2. Type '=(b1+2b2+b3)/4' to find the smoothed *distance run* for month 2 (February).

Press enter.



Note: C1 is left blank as there is no two-mean smoothed value for the first data point.

Step 5: With cell C2 still selected, move the cursor to the bottom right corner of the cell and click and drag downwards to apply the formula to the remaining rows.



Explanation – Method 2: Casio ClassPad

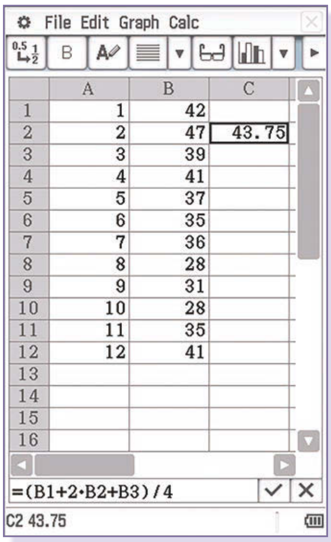
Step 1: From the main menu, tap Spreadsheet.

Step 2: Enter the numbers from 1 to 12 in the first column. Each number will represent a month.

Step 3: Enter the *distance run* values into the second column.

Step 4: Select cell C2. Type '=(b1+2b2+b3)/4' to find the smoothed *distance run* for month 2 (February).

Press EXE.

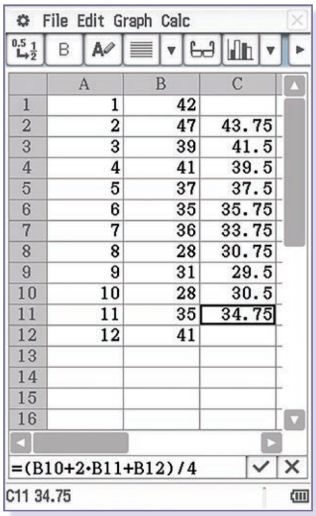


Note: C1 is left blank as there is no two-mean smoothed value for the first data point.

Continues →

Back to contents

Step 5: With cell C2 still selected, drag from C2 down to C3. This will copy the formula down to the next cell. Repeat this until cell C11.



Answer – Method 1 and 2

month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
***distance run* (km)**	42	47	39	41	37	35	36	28	31	28	35	41
two-mean smoothed *distance run* with centring (km)	–	43.8	41.5	39.5	37.5	35.8	33.8	30.8	29.5	30.5	34.8	–

Note: Alternatively, this question could be solved manually using the technique from part **a**.

---

c. Find the four-mean smoothed *distance run* for July.

Explanation

Step 1: Find the *distance run* for July as well as the two values on either side and write them in the order they appear in the time series.

37 35 36 28 31

Step 2: Find the mean of the first four values.

$$mean_1 = \frac{37 + 35 + 36 + 28}{4} = 34$$

Step 3: Find the mean of the last four values.

$$mean_2 = \frac{35 + 36 + 28 + 31}{4} = 32.5$$

Step 4: Calculate the centred mean.

$$mean_{centred} = \frac{34 + 32.5}{2} = 33.25$$

Answer

33.25 km

---

d. Find the four-mean smoothed values, correct to one decimal place, for the entire time series and fill in the table.

Explanation – Method 1: TI-Nspire

Step 1: From the home screen, select ‘1: New’ → ‘4: Add Lists & Spreadsheet’.

Step 2: Enter the numbers from 1 to 12 in the first column. Each number will represent a month. Name the column ‘month’.

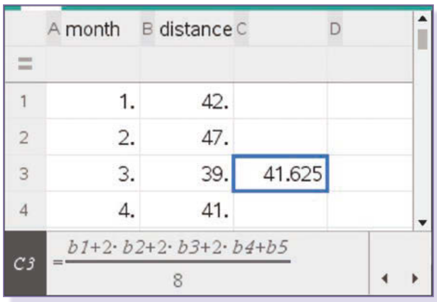
Step 3: Enter the *distance run* values into the second column and name the column ‘distance’.

Continues →

Back to contents

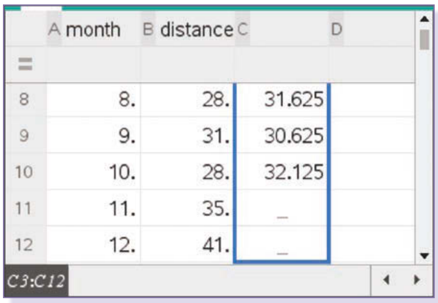
Step 4: Select cell C3. Type '=(b1+2b2+2b3+2b4+b5)/8' to find the smoothed *distance run* for month 3 (March).

Press enter .



Note: C1 and C2 are left blank as there are no four-mean smoothed values for the first two data points.

Step 5: With cell C3 still selected, move the cursor to the bottom right corner of the cell and click and drag downwards to apply the formula to the remaining rows.



Explanation – Method 2: Casio ClassPad

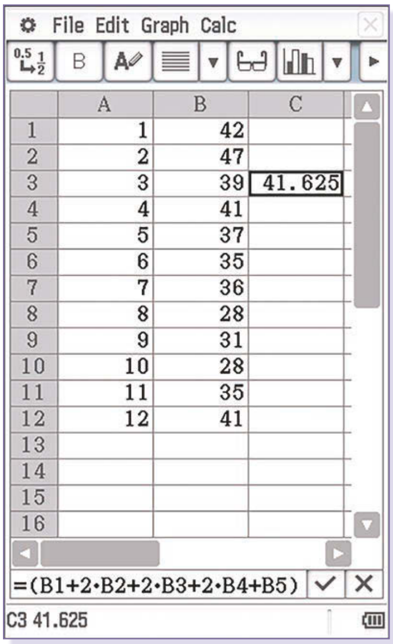
Step 1: From the main menu, tap Spreadsheet.

Step 2: Enter the numbers from 1 to 12 in the first column. Each number will represent a month.

Step 3: Enter the *distance run* values into the second column.

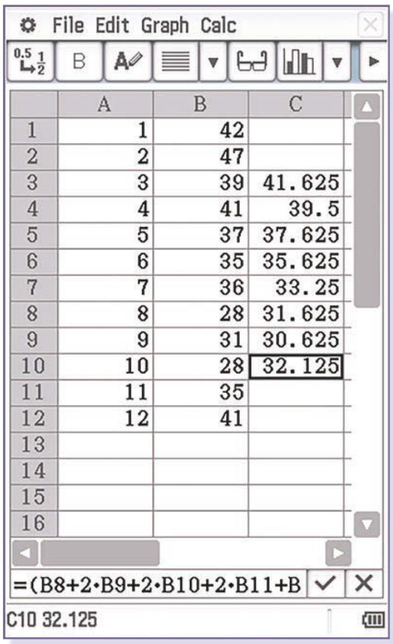
Step 4: Select cell C3. Type '=(b1+2b2+2b3+2b4+b5)/8' to find the smoothed *distance run* for month 3 (March).

Press **EXE**.



Note: C1 and C2 are left blank as there are no four-mean smoothed values for the first two data points.

Step 5: With cell C3 still selected, drag from C3 down to C4. This will copy the formula down to the next cell. Repeat this until cell C10.



Continues →

Back to contents

Answer – Method 1 and 2

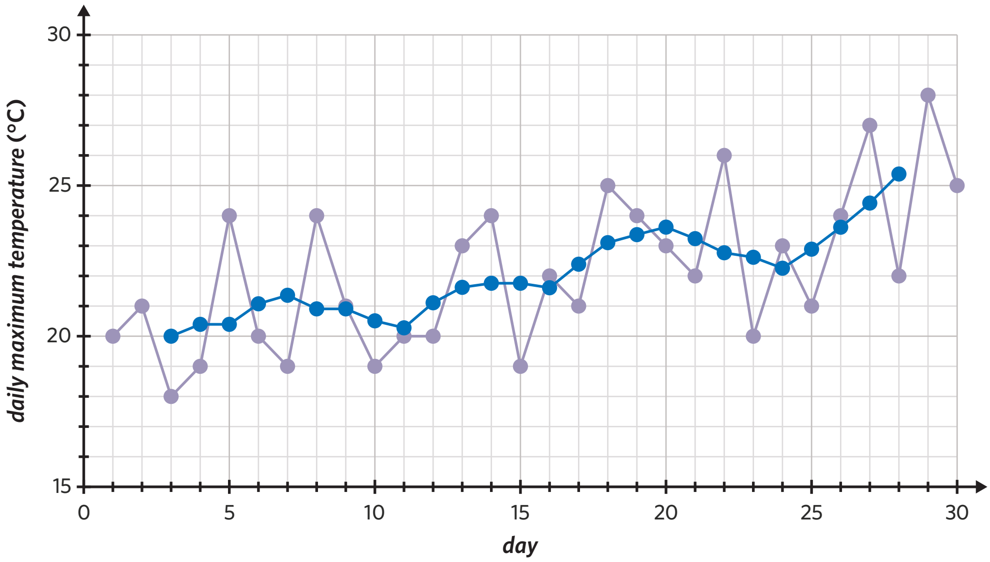
month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
distance run (km)	42	47	39	41	37	35	36	28	31	28	35	41
four-mean smoothed *distance run* with centring (km)	–	–	41.6	39.5	37.6	35.6	33.3	31.6	30.6	32.1	–	–

Note: Alternatively, this question could be solved manually using the technique from part **c**.

Plotting and interpreting a mean smoothed time series

Once data values have been smoothed, they can then be plotted against the original time series. By doing so, any underlying trends become more visible and conclusions can be made about the data.

In the following time series, the *daily maximum temperature* (°C) in November was recorded and then smoothed using four-mean smoothing. The smoothed values show an increasing trend in the data.



Worked example 3

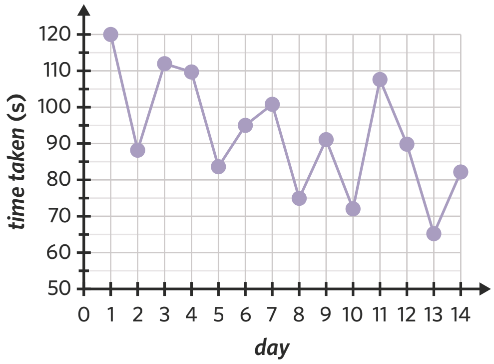
Tobias wants to get faster at solving his Rubik's Cube. He solves it once a day and records the *time taken* in seconds. The data taken over two weeks is shown in the following table. Tobias has also calculated the five-mean smoothed values for the data set, correct to one decimal place.

day	1	2	3	4	5	6	7	8	9	10	11	12	13	14
time taken (s)	120	88	112	110	83	95	101	75	91	72	108	90	65	82
five-mean smoothed *time taken* (s)	–	–	102.6	97.6	100.2	92.8	89.0	86.8	89.4	87.2	85.2	83.4	–	–

Continues →

Back to contents

Tobias has constructed the following graph to display the *time taken*.

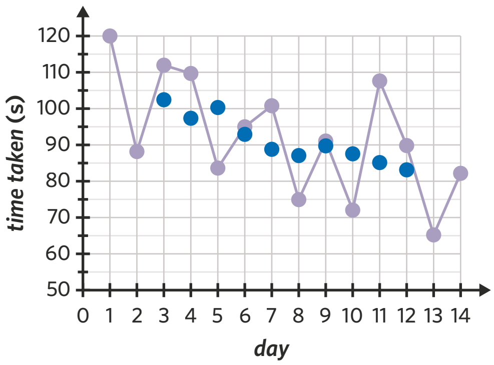


a. Plot the five-mean smoothed values onto the same graph.

Explanation

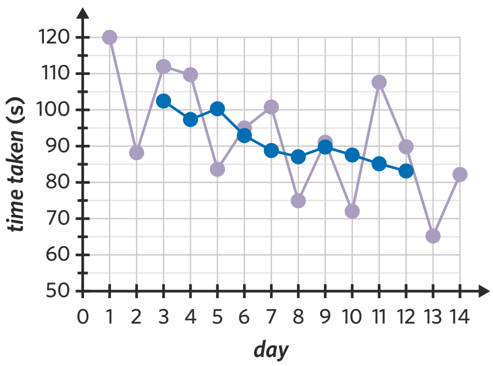
Step 1: Plot each five-mean smoothed data point on the graph.

There are no data values for the first two or last two days.



Step 2: Connect the data points from left to right.

Answer



b. What, if any, trend can be seen from the five-mean smoothed data?

Explanation

From the graph, it can be observed that the five-mean smoothed data values gradually decrease from left to right.

Answer

The five-mean smoothed data shows a decreasing trend in *time taken* (s).

Back to contents

Exam question breakdown

VCAA 2020 Exam 1 Data analysis Q18

The following 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	Jun	Jul	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

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

Explanation

Step 1: Find the *monthly rainfall* for August 2019 as well as the three values on either side and write them in the order they appear in the time series.

92.6 77.2 80.0 86.8 93.8 55.2 97.3

Step 2: Find the mean of the first six values.

$$mean_1 = \frac{92.6 + 77.2 + 80.0 + 86.8 + 93.8 + 55.2}{6}$$

$$= 80.93...$$

Step 3: Find the mean of the last six values.

$$mean_2 = \frac{77.2 + 80.0 + 86.8 + 93.8 + 55.2 + 97.3}{6}$$

$$= 81.71...$$

Step 4: Calculate the centred mean.

$$mean_{centred} = \frac{80.93... + 81.71...}{2} = 81.325$$

Answer

C

73% of students answered this question correctly.

4B Questions

Smoothing over an odd number of data points using moving means

1. Consider the following table.

month	1	2	3	4	5	6
value	75	83	104	97	71	64

The three-mean smoothed value for month 4 is closest to

A. 91 **B.** 92 **C.** 93 **D.** 94

2. Alexia recorded the *time taken*, in minutes, to travel to school over a period of two weeks.

She displayed the data in the following table.

day	Mon 1st Oct	Tue 2nd Oct	Wed 3rd Oct	Thu 4th Oct	Fri 5th Oct	Mon 8th Oct	Tue 9th Oct	Wed 10th Oct	Thu 11th Oct	Fri 12th Oct
***time taken* (mins)**	37	33	44	40	45	31	38	41	37	50

a. Calculate the three-mean smoothed *time taken* for Tuesday 2nd October.

b. Calculate the five-mean smoothed *time taken* for Wednesday 10th October.

Back to contents

3. Each day, Mrs Stinton asks her third grade class if they have eaten fruit. The results for the last ten days are given.

day	1	2	3	4	5	6	7	8	9	10
number of students	14	12	19	22	17	12	19	15	12	21
three-mean smoothed *number of students*	–									–
five-mean smoothed *number of students*	–	–							–	–

a. Fill in the three-mean smoothed values for the entire time series, correct to one decimal place.

b. Fill in the five-mean smoothed values for the entire time series, correct to one decimal place.

4. Crime rates in the city of Ashton are on the rise. The figures for the *number of thefts* recorded in the past year are shown in the table.

month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
number of thefts	3	8	7	12	10	17	18	22	26	25	32	35
three-mean smoothed *number of thefts*	–											–
five-mean smoothed *number of thefts*	–	–									–	–

a. Find the three-mean smoothed *number of thefts* for July correct to the nearest whole number.

b. Find the three-mean smoothed values for the entire time series and fill in the table correct to one decimal place.

c. Find the five-mean smoothed values for the entire time series and fill in the table correct to one decimal place.

5. Rachel is a leading financial analyst based in the US. Upon analysing exchange rate data for 2003 and 2004, she noticed that for one particular USD exchange rate, the data for the first quarter of 2004 was missing.

year	2003				2004			
quarter	1	2	3	4	1	2	3	4
***exchange rate* (USD)**	0.97	0.99	1.04	1.01		0.89	0.86	0.88
three-mean smoothed *exchange rate* (USD)	–							–
five-mean smoothed *exchange rate* (USD)	–	–					–	–

a. After applying her financial skills, Rachel discovers that the three-mean smoothed *exchange rate* for the first quarter of 2004 is 0.94 USD.

Use this information to help Rachel find the exchange rate for the first quarter of 2004 correct to two decimal places.

b. Find the seven-mean smoothed *exchange rate* for the fourth quarter of 2003, correct to two decimal places.

c. Find the three-mean smoothed *exchange rate* for the entire time series, correct to two decimal places.

d. Find the five-mean smoothed *exchange rate* for the entire time series, correct to two decimal places.

Smoothing over an even number of data points using moving means

6. Consider the following table.

year	2016	2017	2018	2019	2020	2021	2022
sales	203	226	241	188	203	261	250

The two-mean smoothed *sales* for 2020 is closest to

A. 214 **B.** 217 **C.** 224 **D.** 232

7. The following table shows the *number of points* scored by the winning team of the AFL grand final between 2010 and 2021.

year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
number of points	108	119	91	77	137	107	89	108	79	114	81	140

a. Calculate the two-mean smoothed *number of points* for 2020.

b. Calculate the four-mean smoothed *number of points* for 2013.

c. Calculate the six-mean smoothed *number of points* for 2015.

8. The following table shows the number of people who visited a popular tourist attraction over the course of a year.

month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
number of tourists (000's)	1.9	2.1	1.8	1.7	1.4	1.1	1.2	1.4	1.7	1.6	1.8	2.0
two-mean smoothed *number of tourists* (000's)	–											–
four-mean smoothed *number of tourists* (000's)	–	–									–	–

a. Find the two-mean smoothed *number of tourists* (000's) centred at July, correct to one decimal place.

b. Find the four-mean smoothed *number of tourists* (000's) centred at July, correct to one decimal place.

c. Find the two-mean smoothed and centred *number of tourists* (000's), correct to one decimal place, for the entire time series and fill in the table.

d. Find the four-mean smoothed and centred *number of tourists* (000's), correct to one decimal place, for the entire time series and fill in the table.

9. Brie recently opened a fromagerie called 'Really Grate Cheese'. She recorded the total *weight*, in kg, of cheese sold each week since the opening of the store.

week	1	2	3	4	5	6	7	8	9	10	11
weight (kg)	18.2	21.5	23.3	22.0	25.4	24.9	26.6		25.8	22.2	24.7
four-mean smoothed *weight* (kg)											

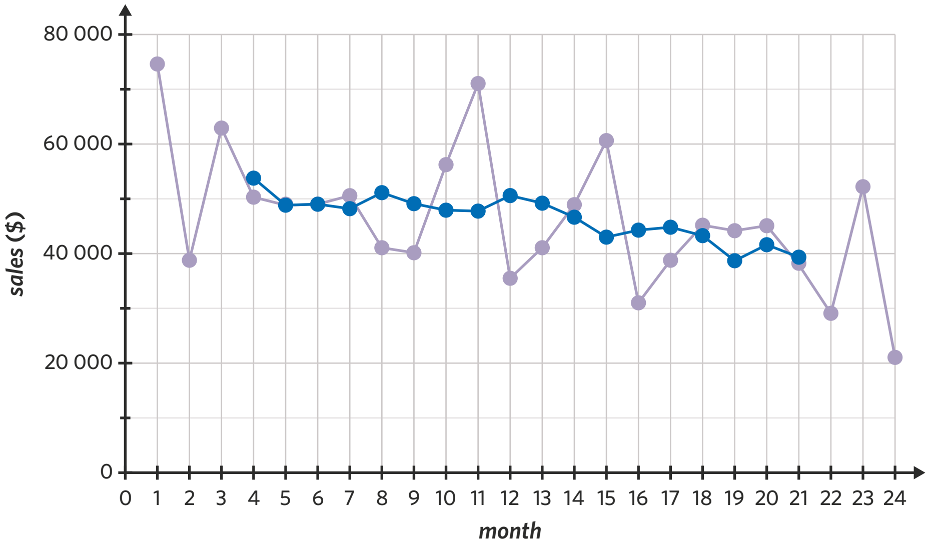
a. Brie accidentally lost her sales records from week 8. Brie knows that the four-mean smoothed *weight* of cheese sold, centred on week 8, is 24.7 kg.
Using this information, find the exact actual *weight* of cheese sold in week 8.

b. Find the four-mean smoothed and centred *weight* for the entire time series, correct to one decimal place.

Back to contents

Plotting and interpreting a mean smoothed time series

10. The following graph shows a retail shop's *sales* and their seven-mean smoothed *sales* over 24 months.



What trend is visible in the smoothed values?

A. Increasing

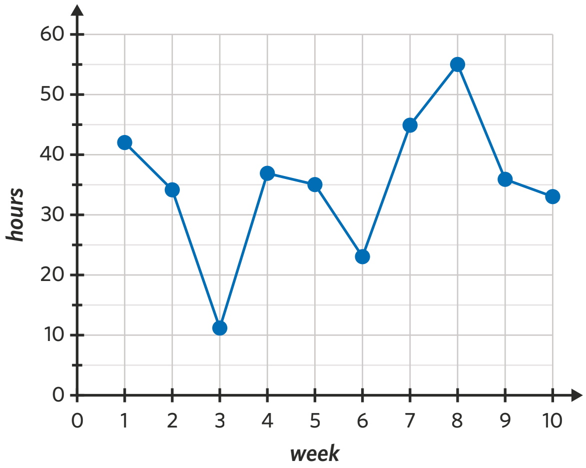
B. Decreasing

C. No trend

D. Structural change

---

11. Winston is a freelance photographer and recorded the number of *hours* he worked each week for 10 weeks, and plotted the data onto the following graph.



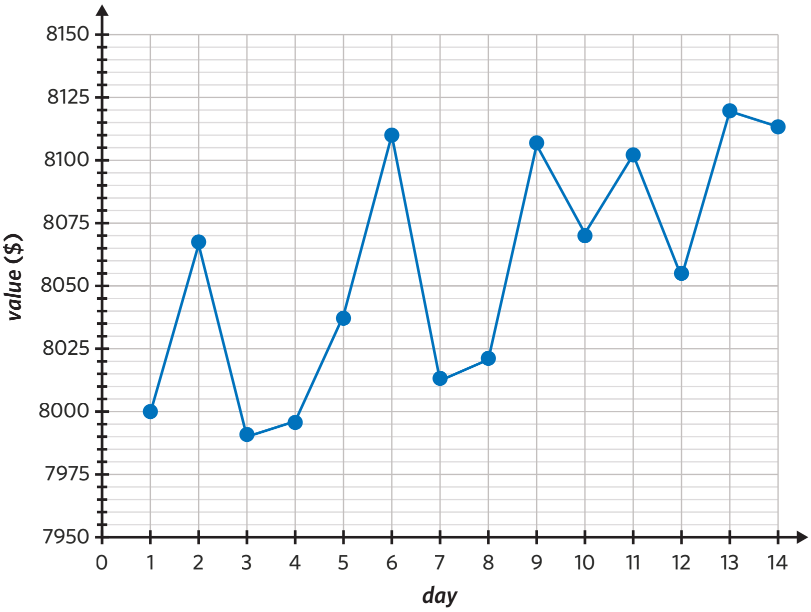
He also calculated the two-mean and five-mean smoothed values for the data, correct to one decimal place.

week	1	2	3	4	5	6	7	8	9	10
hours	42	34	11	37	35	23	45	55	36	33
two-mean smoothed *hours*	–	30.3	23.3	30.0	32.5	31.5	42.0	47.8	40.0	–
five-mean smoothed *hours*	–	–	31.8	28.0	30.2	39.0	38.8	38.4	–	–

a. In one colour, plot the two-mean smoothed values onto the graph.

b. In another colour, plot the five-mean smoothed values onto the same graph.

12. Two weeks ago, Amira invested $8000 in various cryptocurrencies and has since monitored its *value*, in dollars, daily.



Amira also calculated the four-mean smoothed *value* of her crypto portfolio with centring, as shown in the following table. Values are given correct to the nearest whole number.

day	1	2	3	4	5	6	7	8	9	10	11	12	13	14
value **($)**	8000	8067	7991	7996	8037	8110	8013	8021	8106	8070	8102	8055	8119	8113
four-mean smoothed ***value*** **($)**	–	–	8018	8028	8036	8042	8054	8058	8064	8079	8085	8092	–	–

a. Plot the smoothed values onto the graph.

b. What trend is visible in the smoothed values? Interpret this in terms of the context given.

Joining it all together

13. An ice cream shop recorded the *sales* for their brand new bubblegum flavoured ice cream over the course of a year. Their findings are shown in the following table.

month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
sales **($)**	450	510	370	360	290	100	110	90	150	230	310	420

a. Find the three-mean smoothed *sales* figure for May.

b. Find the five-mean smoothed *sales* figure for August.

c. Find the two-mean smoothed *sales* figure with centring for September.

d. Find the four-mean smoothed *sales* figure with centring for March.

e. Find the seven-mean smoothed *sales* figure for July.

f. Find the six-mean smoothed *sales* figure with centring for June, correct to one decimal place.

Back to contents

14. Alfredo is a passionate yogi and recorded the number of *yoga classes* that he has attended each year since 2010.

year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
yoga classes	52	56	48	61	66	59	69	73	65	81	53	88	79

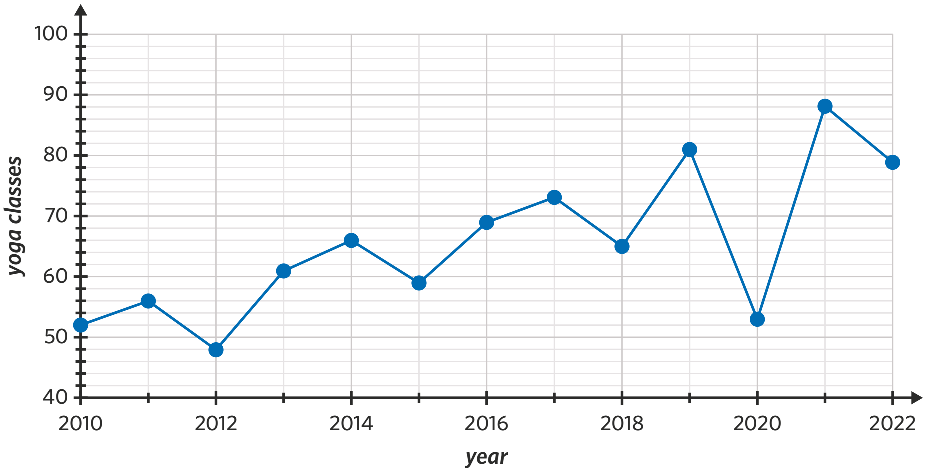
a. Alfredo considers using nine-mean smoothing to smooth the data. If he were to do this, which year would be the first to have a smoothed value?

b. Calculate the nine-mean smoothed value for the year specified in part **a**.

c. Alfredo decides that it would be more useful to use two-mean smoothing so he retains more data points. Complete the following table with the two-mean smoothed and centred values, to the nearest whole number.

year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
yoga classes	52	56	48	61	66	59	69	73	65	81	53	88	79
two-mean smoothed *yoga classes*													

d. Alfredo has already plotted the original data onto the following time series. Plot the two-mean smoothed data onto the same graph.



e. What trend, if any, is visible in the smoothed data? Interpret this with reference to the given variables.

Exam practice

15. The following time series plot shows the *monthly rainfall* at a weather station, in millimetres, for each month in 2017.

If seven-mean smoothing is used to smooth this time series plot, the number of smoothed data points would be

A. 3

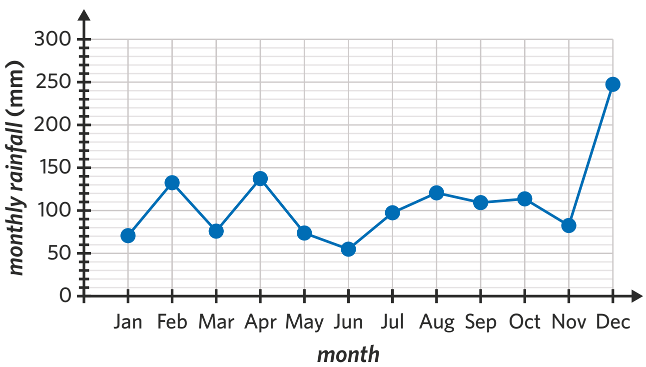
B. 5

C. 6

D. 8

E. 10

VCAA 2019 Exam 1 Data analysis Q16



74% of students answered this question correctly.

16. The wind speed at a city location is measured throughout the day over a three-week period. The following table 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. 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

VCAA 2017 Exam 1 Data analysis Q15

64% of students answered this question correctly.

---

17. A garden centre sells garden soil. The following table shows the daily *quantity* of garden soil sold, in cubic metres, over a one-week period.

day	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
quantity **(m^3)**	234	186				346	346

The *quantity* of garden soil sold on Wednesday, Thursday and Friday is not shown. The five-mean smoothed *quantity* of garden soil sold on Thursday is 206 m^3. The three-mean smoothed *quantity* of garden soil sold on Thursday, in cubic metres, is

A. 143

B. 166

C. 206

D. 239

E. 403

VCAA 2021 Exam 1 Data analysis Q14

48% of students answered this question correctly.

---

18. The maximum daily rainfall each month was recorded at a weather station. The following table shows the *maximum daily rainfall* each month for a period of one year.

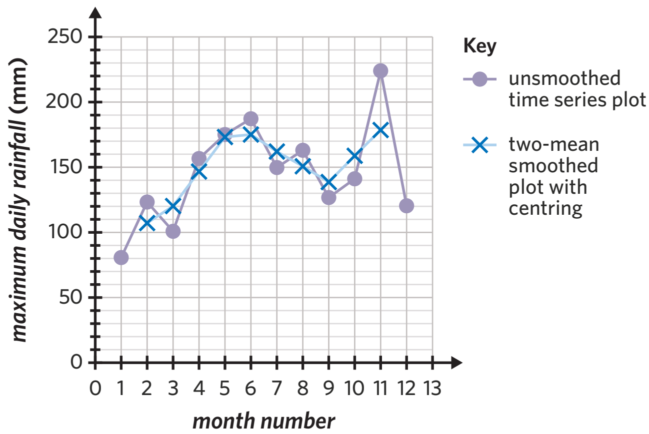
month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
month number	1	2	3	4	5	6	7	8	9	10	11	12
maximum daily rainfall **(mm)**	79	123	100	156	174	186	149	162	124	140	225	119

The data in the table has been used to plot *maximum daily rainfall* against *month number* in the following time series plot.

Two-mean smoothing with centring has been used to smooth the time series plot. The smoothed values are marked with crosses (×).

Using the data given in the table, show that the two-mean smoothed rainfall centred on October is 157.25 mm. (2 MARKS)

VCAA 2016 Exam 2 Data analysis Q4b



The average mark on this question was **1**.

Back to contents

Questions from multiple lessons

Data analysis

19. The following histogram shows the distribution of the *number* of skyscrapers per 1000 m^2 for 61 different cities plotted on a $\log_{10}$ scale.

How many cities have at least one skyscraper per 1000 m^2?

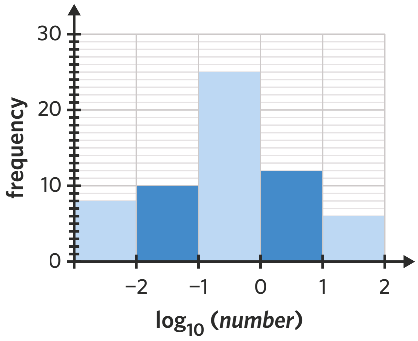
A. 6

B. 12

C. 18

D. 25

E. 43



Adapted from VCAA 2016 Exam 1 Data analysis Q7

Data analysis

20. The following scatterplot shows the number of bubble teas sold versus the temperature in degrees celsius on 12 days. A least squares regression line has been fitted to the scatterplot.

The least squares line is used to predict the number of bubble teas sold on a 15 degree day. The residual is closest to

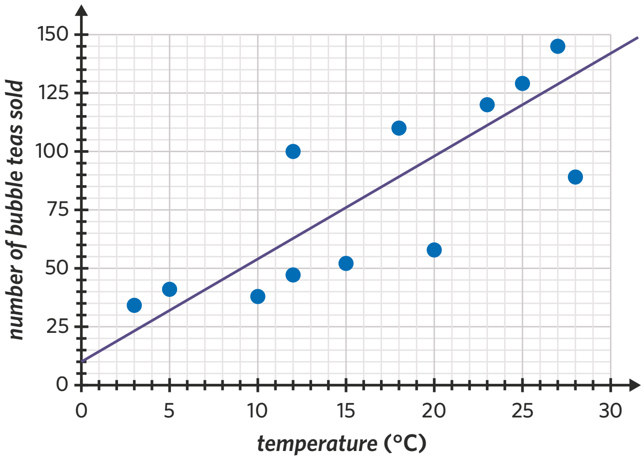
A. −38

B. −23

C. −12

D. 23

E. 38



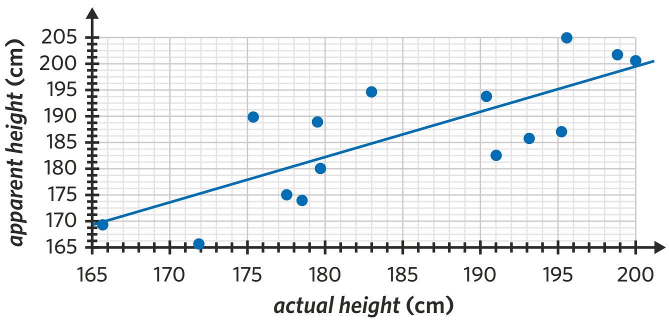
Adapted from VCAA 2017 Exam 1 Data analysis Q9

Data analysis

21. The data in the following table shows a sample of *apparent height* and *actual height* recorded from a group of Year 12 students, where *apparent height* is the height that the students claim to be.

A scatterplot of the data is also shown.

This data will be used to investigate the association between *apparent height* and *actual height*.



$n = 15$
$r^2 = 0.64$

apparent *height* (cm)	*actual* *height* (cm)
200.9	199.7
182.7	190.9
194.2	183.0
193.2	190.3
202.1	198.8
166.8	171.9
189.1	175.5
204.9	195.6
187.6	195.2
179.9	179.5
173.0	178.6
168.4	166.5
185.7	193.2
174.6	177.8
188.0	179.6

a. Use the scatterplot to describe the association between *apparent height* and *actual height* in terms of strength, form and direction. (1 MARK)

b. A least squares regression line can be used to predict the *apparent height* from the *actual height* in the form *apparent height* $= a + b \times$ *actual height*. Determine the values of a and b. Round your answers to two decimal places. (2 MARKS)

Adapted from VCAA 2016 Exam 2 Data analysis Q3a-bi

Back to contents

4C Smoothing – moving medians

STUDY DESIGN DOT POINT

- 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

4A 4B 4C 4D 4E

KEY SKILLS

During this lesson, you will be:

- smoothing using three-moving medians
- smoothing using five (or more)-moving medians.

KEY TERMS

- Moving median smoothing

Similar to moving mean smoothing, moving median smoothing is used to remove large fluctuations in time series data in order to identify long-term trends. However, unlike moving mean smoothing, moving median smoothing can be done directly onto the graph. Additionally, moving median smoothing is not impacted by outlier values. In General Mathematics, only moving median smoothing for an odd number of points is explored.

Smoothing using three-moving medians

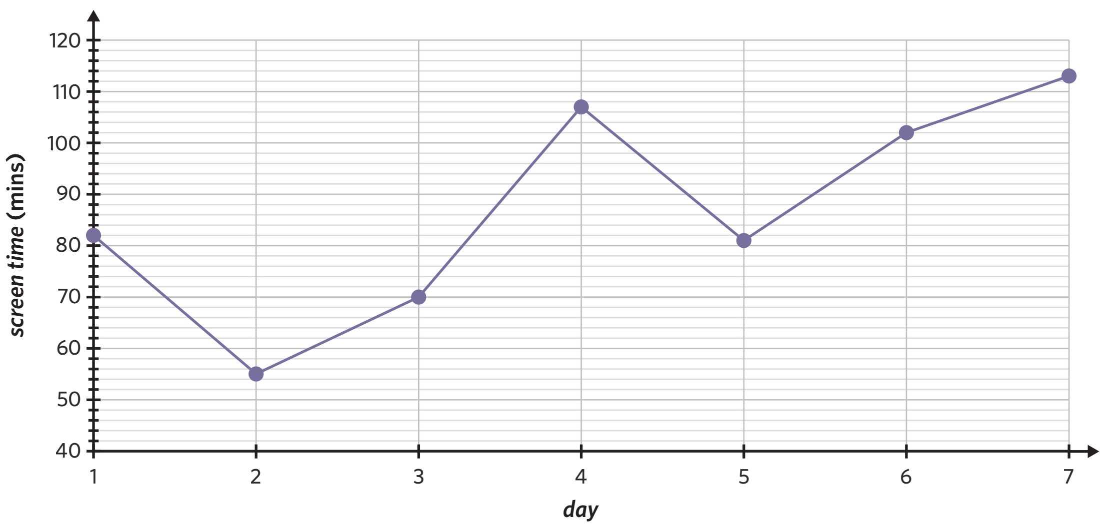
Smoothing using medians is known as **moving median smoothing**. Three-median smoothing involves replacing each data value with the median of itself and each adjacent value.

For example, the three-median smoothed value for day 2 is the median of the values for days 1, 2, and 3.

There are no smoothed values for the first and last data points because they do not have a value on both sides.

Worked example 1

Bailey recorded their *screen time*, in minutes, on their phone over the past week. Their data is presented in the following time series plot.



Continues →

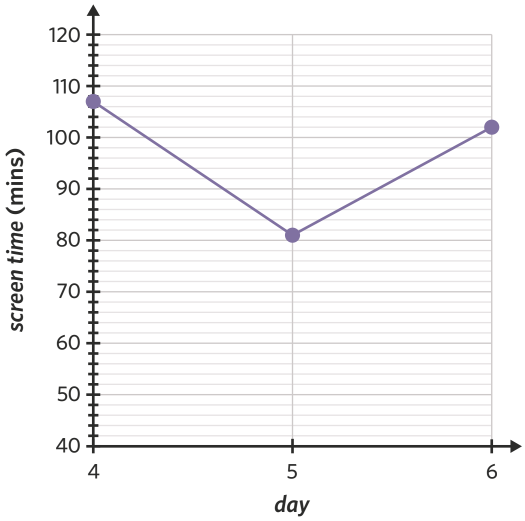
Back to contents

a. Determine the three-median smoothed *screen time* for day 5.

Explanation

Step 1: Identify the necessary data points.

The three-median smoothed value for day 5 is the median of day 5 and the values on each side of it (days 4 and 6).

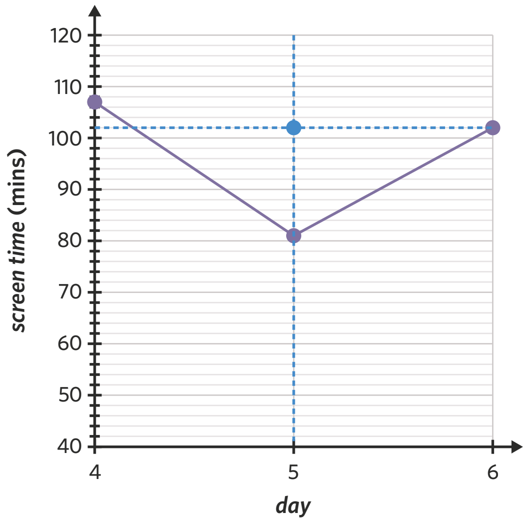


Step 2: Identify the median *screen time* from days 4 to 6.

The variable *screen time* is represented by the vertical height of the points.

As there are three values, the median is the second value from smallest to largest.

The median *screen time* is 102 minutes.



Answer

102 minutes

---

b. Construct a three-median smoothed plot for the entire time series.

Explanation

Step 1: Identify the necessary data points for the first smoothed value.

The first smoothed value will be for day 2 since it has a value on each side.

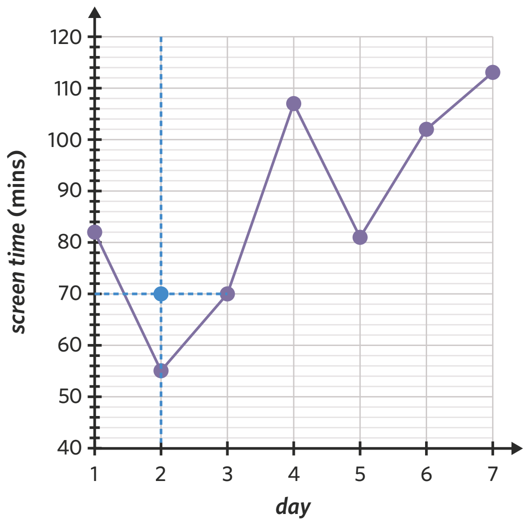
The three-median smoothed value for day 2 is the median of days 1 to 3.

Step 2: Identify the median *screen time* for days 1 to 3.

The *screen time* values for days 1, 2 and 3 are 82, 55, and 70 minutes.

The median *screen time* is 70 minutes.

Step 3: Mark the smoothed value for day 2 on the graph.



Continues →

Back to contents

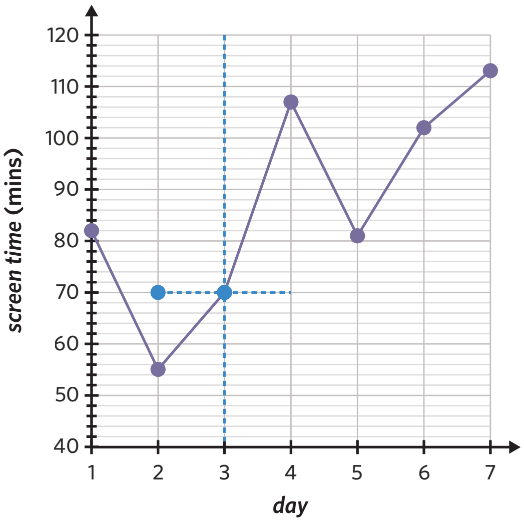
Step 4: Repeat steps 1 to 3 for day 3.

The smoothed value for day 3 is the median *screen time* for days 2 to 4.

The *screen time* values for days 2 to 4 are 55, 70, and 107 minutes.

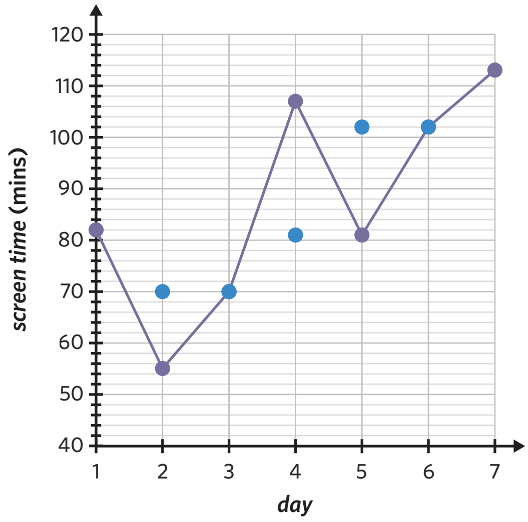
The median *screen time* is 70 minutes.

Make sure the original data points are used and not the previously found smoothed points when finding the three-moving median.



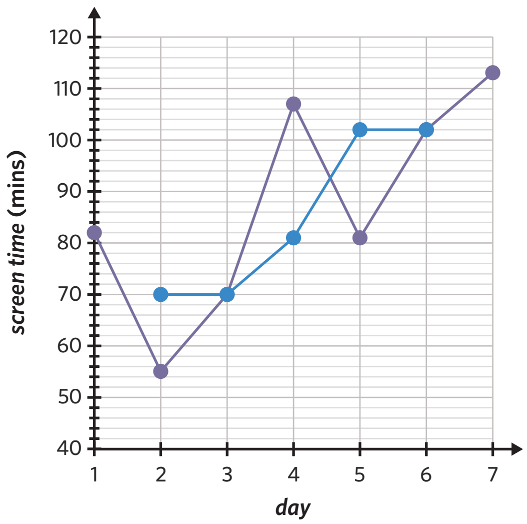
Step 5: Apply this technique for the rest of the time series.

The smoothing will stop at day 6, as this is the last day with a data value on each side.



Step 6: Connect the smoothed points with straight lines.

Answer



c. What, if any, trend can be seen from the three-median smoothed time series?

Explanation

From the graph, it can be observed that the three-median smoothed data values increase from left to right.

Answer

The three-median smoothed data shows an increasing trend in Bailey's *screen time* over the week.

Back to contents

Smoothing using five (or more)-moving medians

Five-median smoothing is similar to three-median smoothing but uses five values. Each data value is replaced with the median of itself and the two values on each side of it.

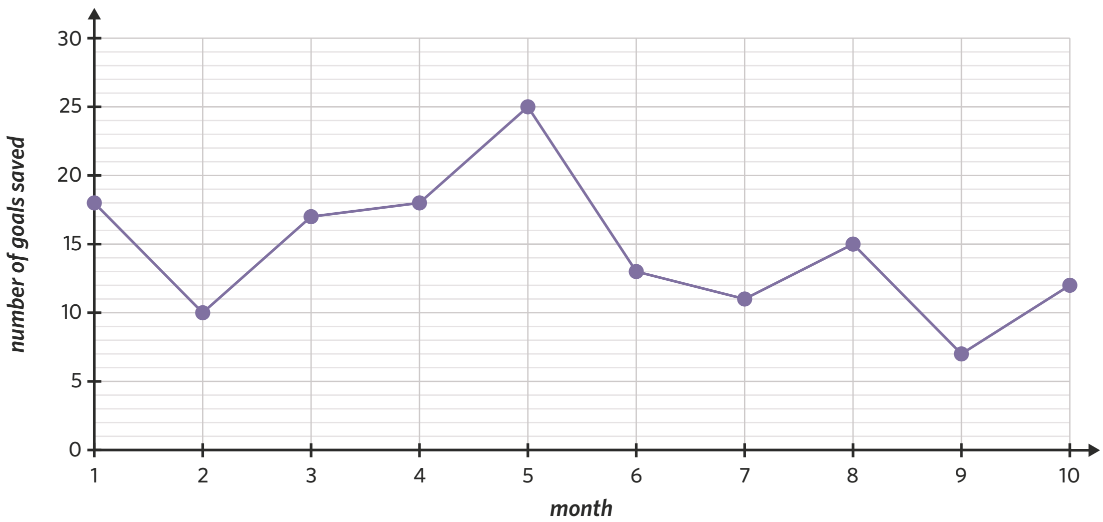
For example, the five-median smoothed value for day 3 is the median of the values for days 1, 2, 3, 4, and 5.

There are no smoothed values for the first two and last two data entries because they do not have two values on both sides.

Moving median smoothing can be done over any odd number of values, in the same way as three and five-median smoothing.

Worked example 2

Emanuel is the goalkeeper for his soccer team and has recorded the *number of goals saved* each month for the past ten months. His data has been presented in the following time series plot.

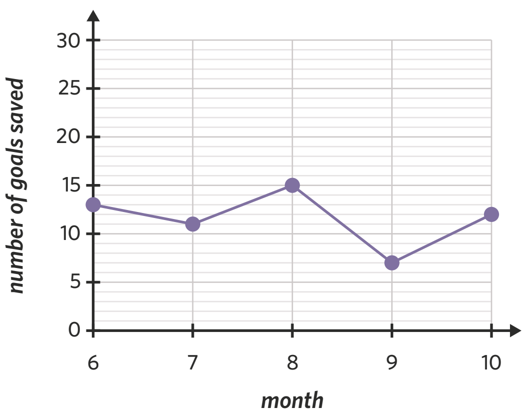


a. Determine the five-median smoothed *number of goals saved* for month 8.

Explanation

Step 1: Identify the necessary data points.

The five-median smoothed value for month 8 is the median of month 8 and the two values on each side of it (months 6, 7, 9 and 10).

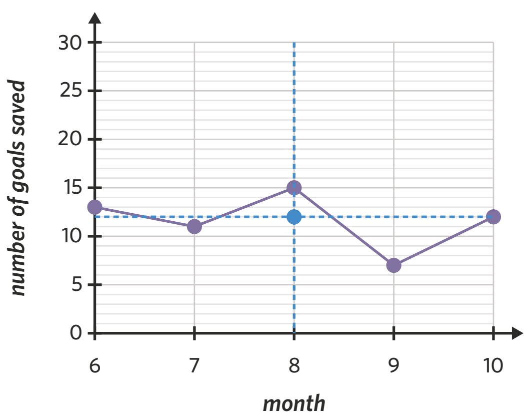


Step 2: Identify the median *number of goals saved* for months 6 to 10.

The variable *number of goals saved* is represented by the vertical height of the points.

As there are five values, the median is the third value from smallest to largest.

The median *number of goals saved* is 12.



Continues →

Back to contents

Answer

12 goals saved

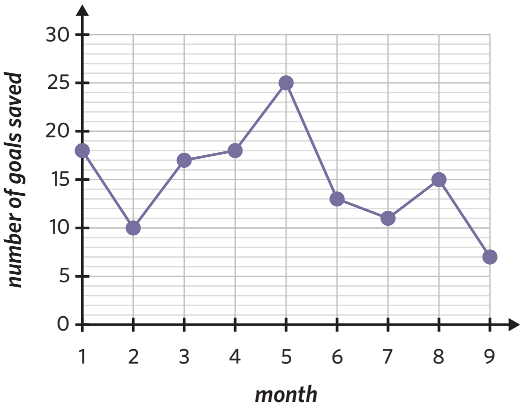
---

b. Determine the nine-median smoothed *number of goals saved* for month 5.

Explanation

Step 1: Identify the necessary data points.

The nine-median smoothed value for month 5 is the median of month 5 and the four values on each side of it (months 1 to 4 and 6 to 9).

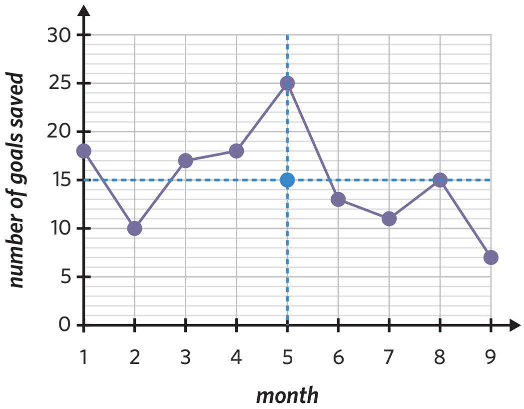


Step 2: Identify the median *number of goals saved* for months 1 to 9.

The variable *number of goals saved* is represented by the vertical height of the points.

As there are nine values, the median is the fifth value from smallest to largest.

The median *number of goals saved* is 15.



Answer

15 goals saved

---

c. Construct a five-median smoothed plot for the entire time series.

Explanation

Step 1: Identify the necessary data points for the first smoothed value.

The first smoothed value will be for month 3 since it has two values on each side.

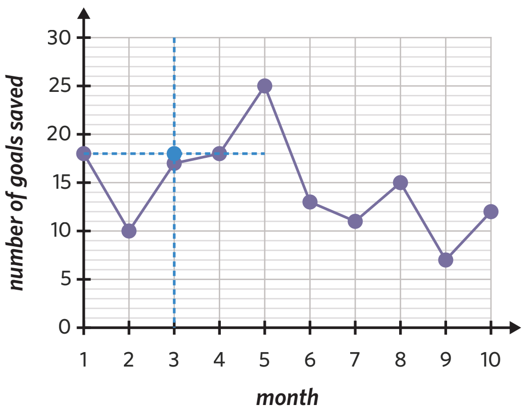
The five-median smoothed value for month 3 is the median of months 1 to 5.

Step 2: Identify the median *number of goals saved* for months 1 to 5.

The *number of goals saved* for months 1 to 5 are 18, 10, 17, 18, and 25.

The median *number of goals saved* is 18.

Step 3: Mark the smoothed value for month 3 on the graph.



Continues →

Back to contents

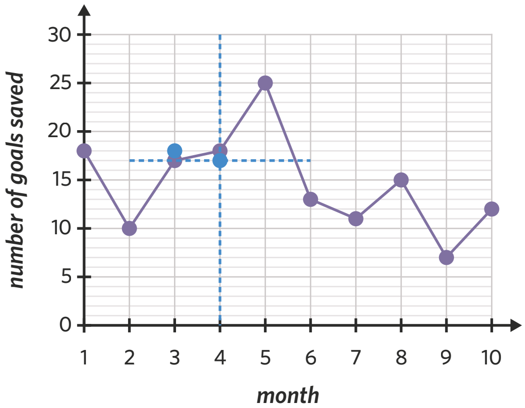
Step 4: Repeat steps 1 to 3 for month 4.

The smoothed value for month 4 is the median *number of goals saved* from months 2 to 6.

The *number of goals saved* for months 2 to 6 are 10, 17, 18, 25 and 13.

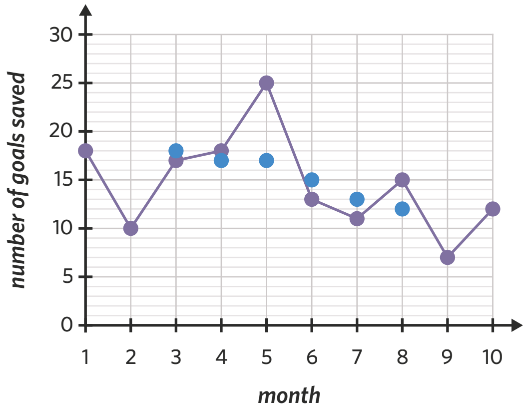
The median *number of goals saved* is 17.

Make sure the original data points are used and not the previously found smoothed points when finding the five-moving median.



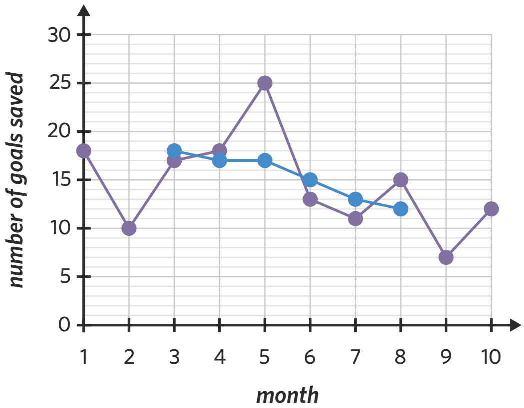
Step 5: Apply this technique for the rest of the time series.

The smoothing will stop at month 8, as this is the last month with two values on each side.



Step 6: Connect the smoothed points with straight lines.

Answer



d. What, if any, trend can be seen from the five-median smoothed time series?

Explanation

From the graph, it can be observed that the five-median smoothed data values gradually decrease from left to right.

Answer

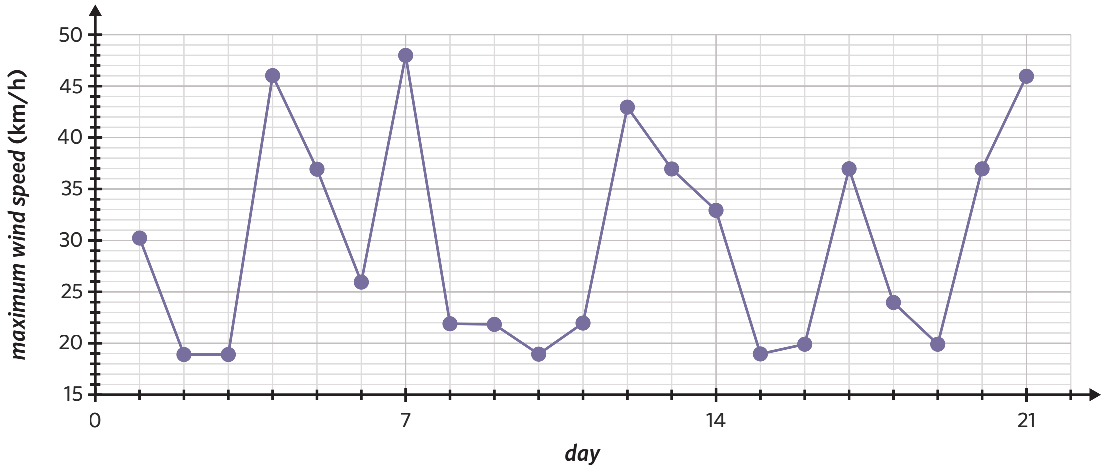
The five-median smoothed data shows a decreasing trend in *number of goals saved*.

Back to contents

Exam question breakdown

VCAA 2017 Exam 1 Data analysis Q14

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.



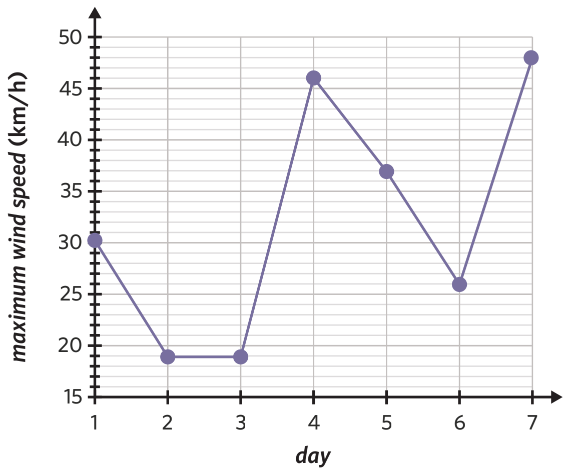
The seven-median smoothed *maximum wind speed*, in kilometres per hour, for day 4 is closest to

A. 22
B. 26
C. 27
D. 30
E. 32

Explanation

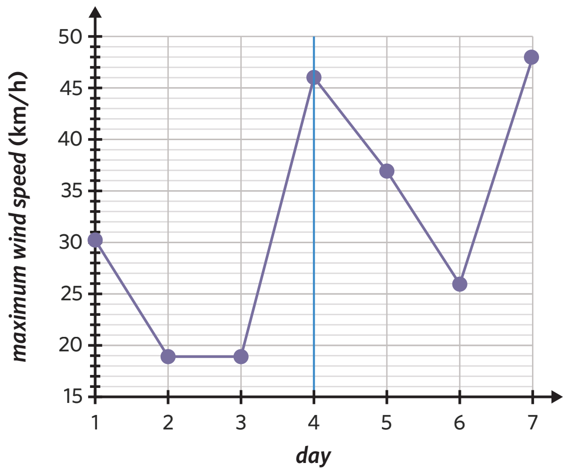
Step 1: Identify the necessary data points.

The seven-median smoothed value for day 4 is the median of day 4 and the three values on each side of it (days 1 to 3 and 5 to 7).



Step 2: Draw a vertical line at day 4.

The smoothed value for day 4 will fall along this line.



Continues →

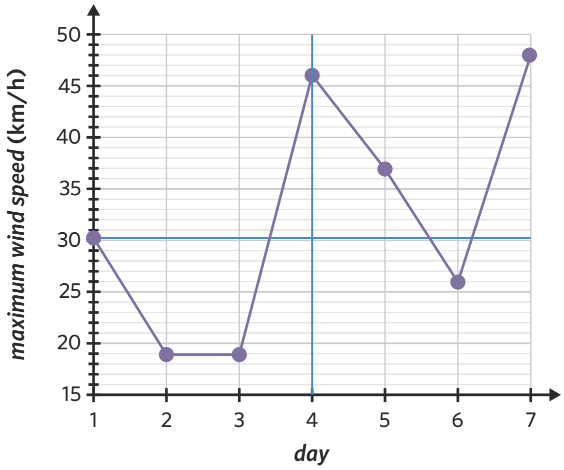
Back to contents

Step 3: Identify the median value for *maximum wind speed* and draw a horizontal line.

The variable *maximum wind speed* is represented by the vertical height of the points.

As there are seven values, the median is the fourth value from smallest to largest.

The median *maximum wind speed* is just over 30 km/h on day 1. Draw a horizontal line from this point. There should be an equal number of points above and below the line.

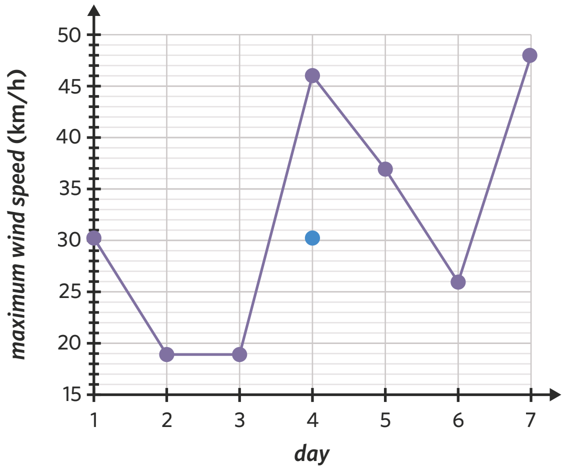


Answer

D

Step 4: Mark the smoothed value.

The seven-median smoothed value for day 4 is at the intersection of these two lines.



The seven-median smoothed *maximum wind speed* for day 4 is closest to 30 km/h.

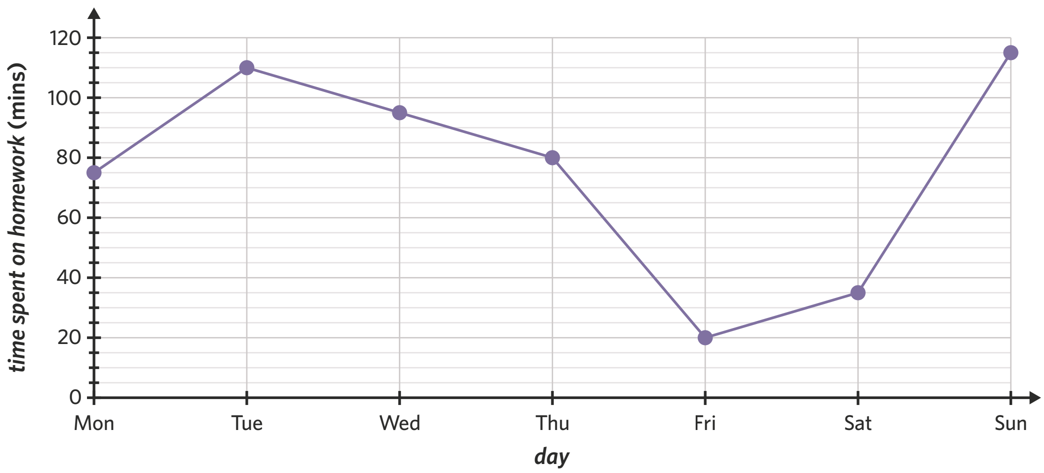
53% of students answered this question correctly.

26% of students incorrectly chose option E. This is likely because they calculated the seven-mean smoothed value, rather than the seven-median smoothed value. The seven-mean smoothed value for day 4 is closest to 32 km/h.

4C Questions

Smoothing using three-moving medians

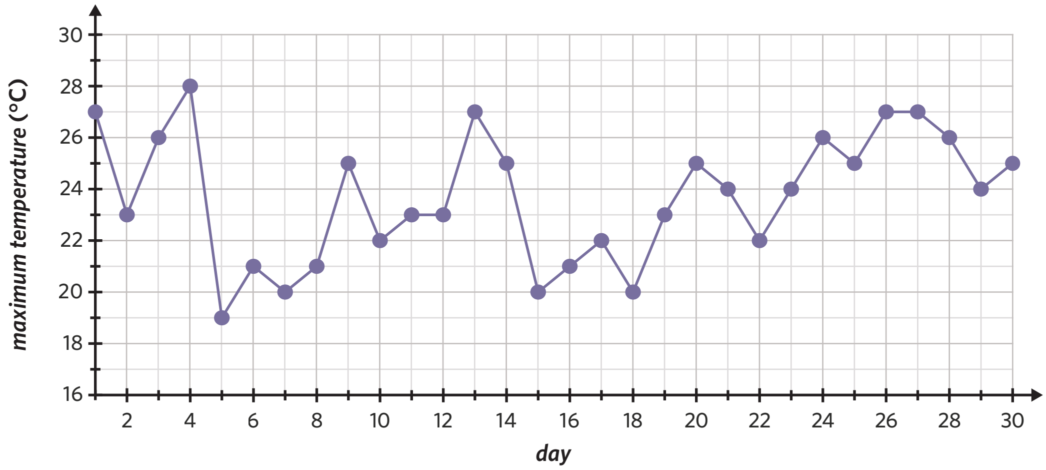
1. Consider the following time series.



The three-median smoothed *time spent on homework* for Friday is

A. 20 minutes **B.** 35 minutes **C.** 45 minutes **D.** 80 minutes

2. The daily *maximum temperature*, in °C, for November is shown.

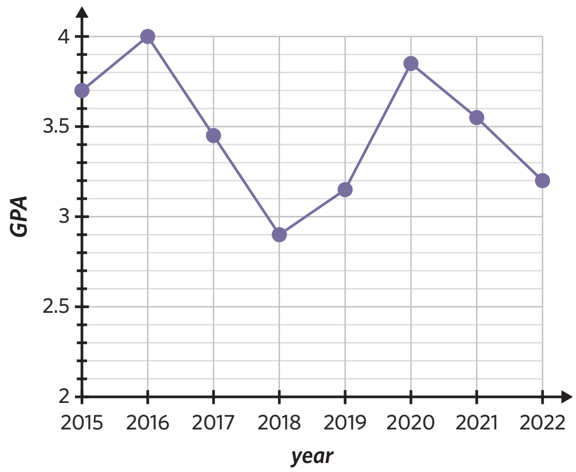


Determine the three-median smoothed *maximum temperature* for the

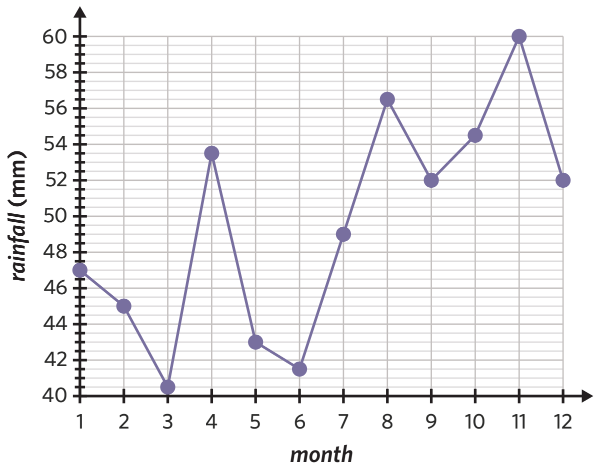
a. 2nd of November. **b.** 7th of November. **c.** 13th of November. **d.** 26th of November.

3. Smooth each of the following time series using three-median smoothing.

a.



b.

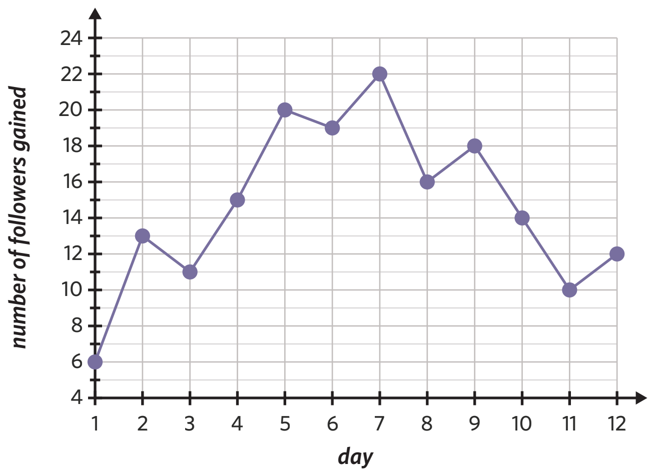


4. The time series plot shows the number of followers Annabel gained on Instagram each day over a 12-day period.

a. Determine the three-median smoothed *number of followers gained* for day 9.

b. Smooth the time-series graph using three-median smoothing.

c. Describe any trends in the smoothed data.

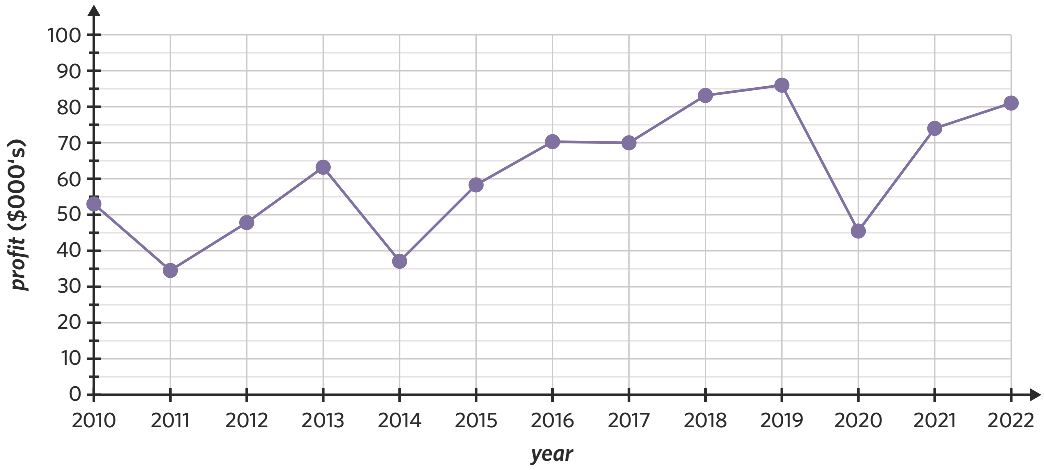


Back to contents

Smoothing using five (or more)-moving medians

5. Consider the following time series.

The five-median smoothed *profit* for 2019 is closest to



A. $70 000 **B.** $74 000 **C.** $83 000 **D.** $86 000

6. Last year, Nikki decided to plant a large crop of potatoes in her veggie garden. This year, Nikki recorded the number of potatoes she harvested each month, as shown.

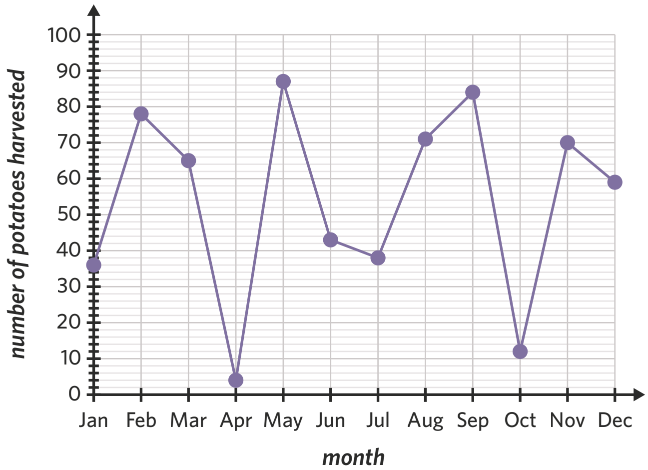
Determine

a. the five-median smoothed *number of potatoes harvested* for March.

b. the seven-median smoothed *number of potatoes harvested* for July.

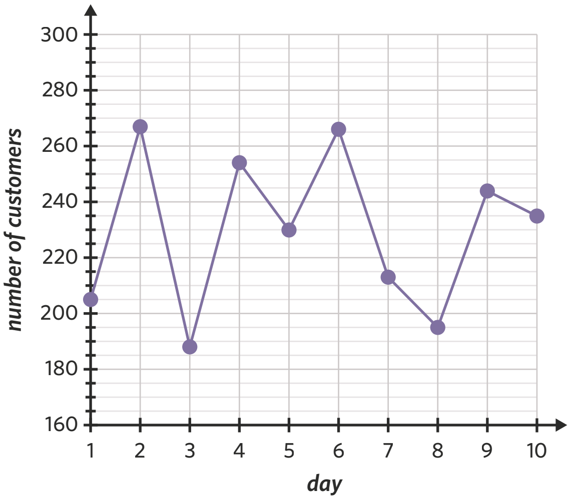
c. the nine-median smoothed *number of potatoes harvested* for August.

d. the eleven-median smoothed *number of potatoes harvested* for June.

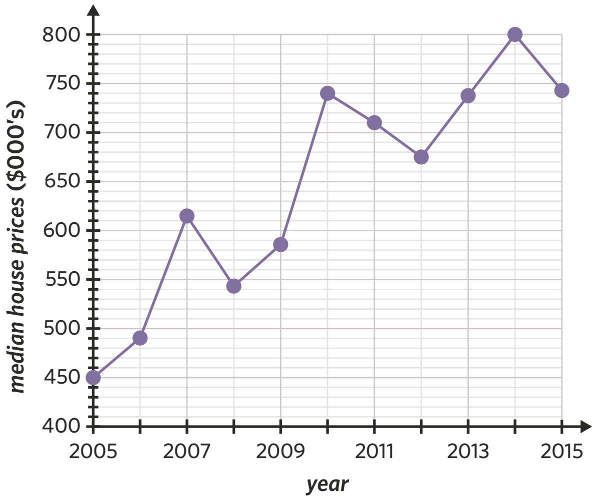


7. Smooth each of the following time series using five-median smoothing.

a.



b.

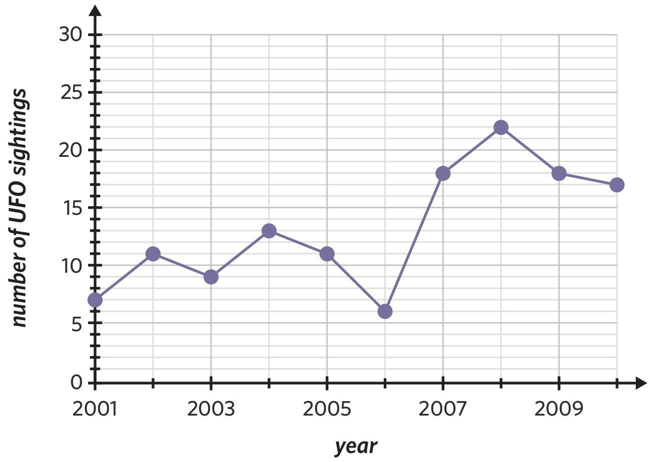


8. The *number of UFO sightings* in a small town was recorded over a ten-year period. The results are shown in the time series plot.

   a. Determine the nine-median smoothed *number of UFO sightings* for 2006.

   b. Smooth the entire time series using five-median smoothing.

   c. What, if any, trend in the *number of UFO sightings* can be seen from the five-median smoothed time series?



Joining it all together

9. Jae is a small business owner. She recorded her *daily income* over the 28 days of February. Jae then smoothed the data using moving median smoothing in order to see any underlying trends. The smoothed data consisted of 22 smoothed data values.

   The smoothing Jae has used is

   A. three-moving median smoothing.

   B. five-moving median smoothing.

   C. seven-moving median smoothing.

   D. nine-moving median smoothing.

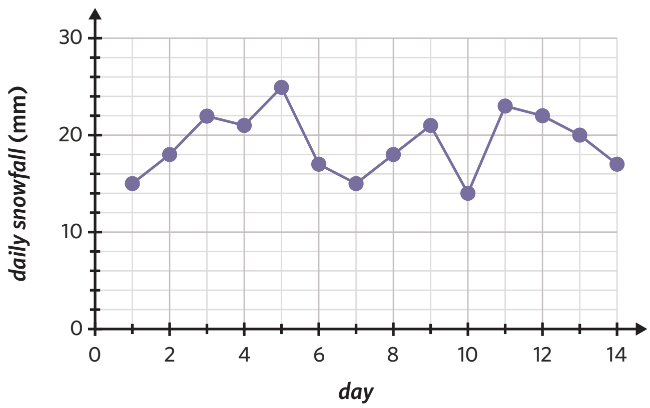
10. The *daily snowfall* (mm) on Mount Buller was recorded over a two-week period. The results are shown in the following time series plot.

   a. Using three-median smoothing, find the smoothed *daily snowfall* for day 4.

   b. Using five-median smoothing, find the smoothed *daily snowfall* for day 7.

   c. Using seven-median smoothing, find the smoothed *daily snowfall* for day 11.

   d. Using nine-median smoothing, find the smoothed *daily snowfall* for day 8.

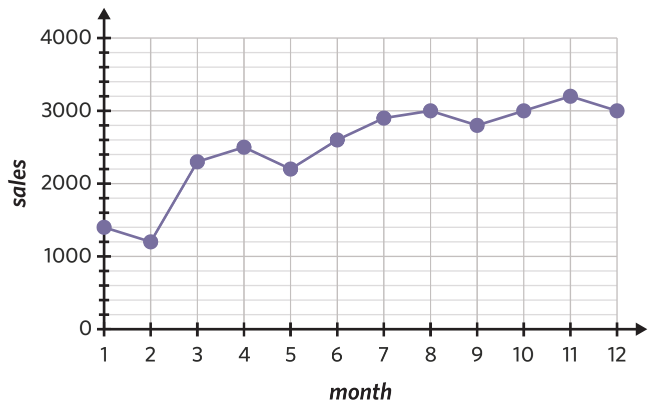


11. The *sales* figures for a new type of tennis racquet were recorded over a 12-month period. The results are shown in a time series plot.

   a. Using three-median smoothing, find the smoothed *sales* for month 2.

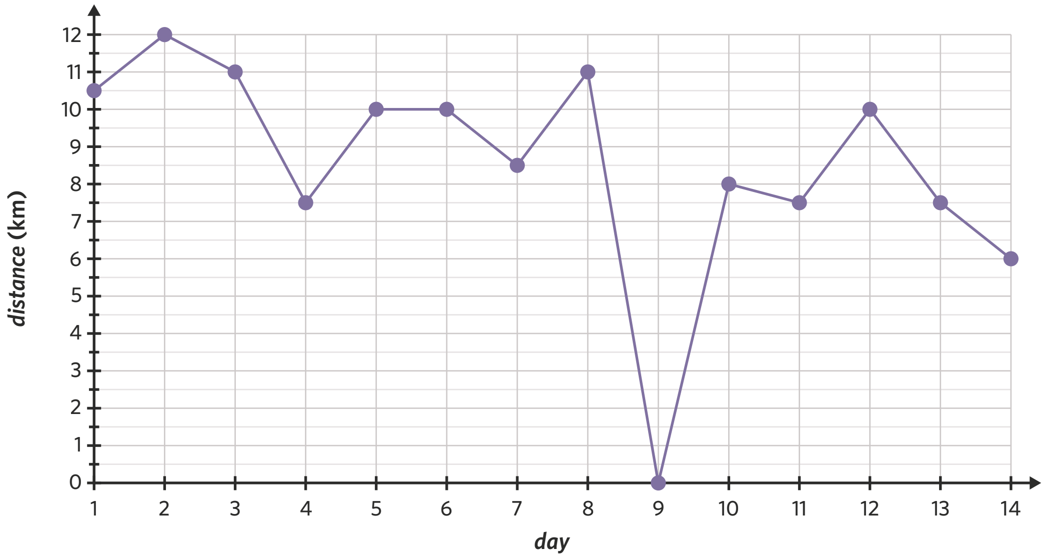
   b. Smooth the time series using five-median smoothing.

   c. Describe the general pattern in *sales* that is displayed by the five-median smoothed plot.



Back to contents

12. Amelia just finished a 14-day hike in Nepal and measured the *distance*, in km, she hiked each day. She displayed her data in the following time series.



a. Amelia had altitude sickness on day 9 so didn't hike at all. To smooth over this outlier, determine the five-median smoothed *distance* for day 9.

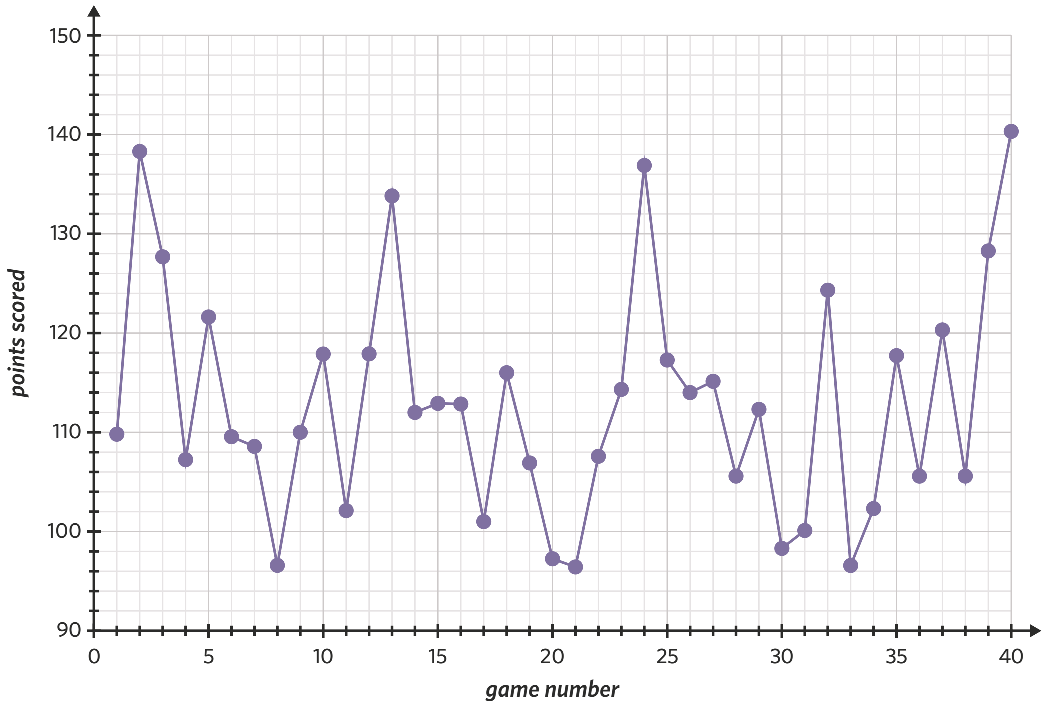
b. Smooth the entire time series using three-median smoothing.

c. Smooth the entire time series using seven-median smoothing.

d. What, if any, trend can be seen from the smoothed time series?

Exam practice

13. The time series plot shows the *points scored* by a basketball team over 40 games.



The nine-median smoothed *points scored* for game number 10 is closest to

A. 102 **B.** 108 **C.** 110
D. 112 **E.** 117

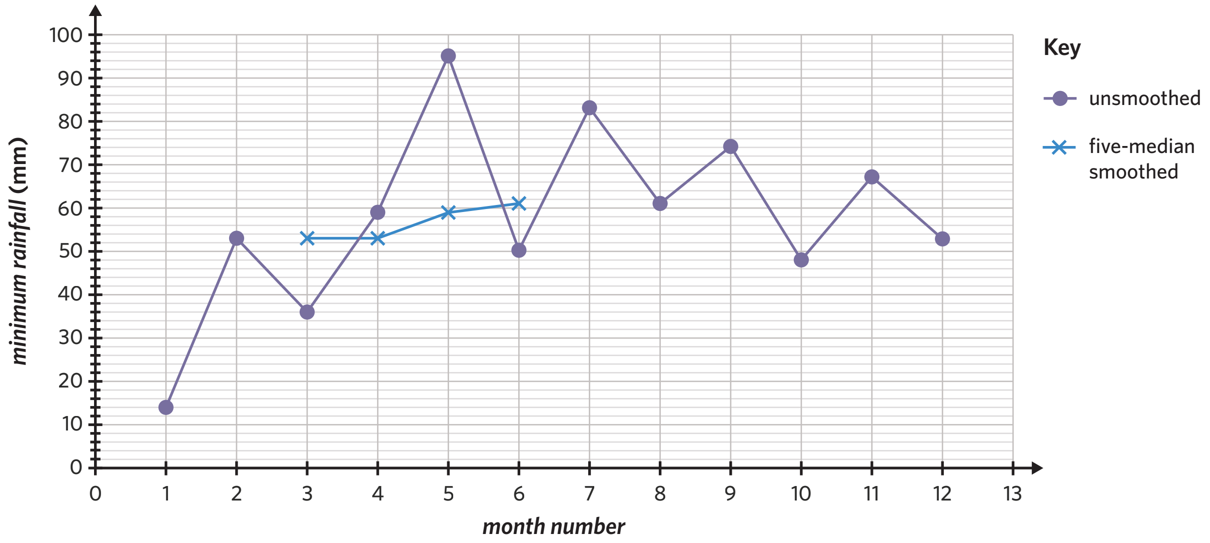
VCAA 2021 Exam 1 Data analysis Q13

53% of students answered this question correctly.

14. The time series plot shows the *minimum rainfall* recorded at the 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.



Five-median smoothing has been used to smooth the time series plot.

The first four smoothed points are shown as crosses (×).

Complete the five-median smoothing by marking smoothed values with crosses (×) on the time series plot. (2 MARKS)

VCAA 2016 Exam 2 Data analysis Q4a

The average mark on this question was **1**.

---

15. The *time*, in minutes, that Liv ran each day was recorded for nine days.

The following 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*

A. 3

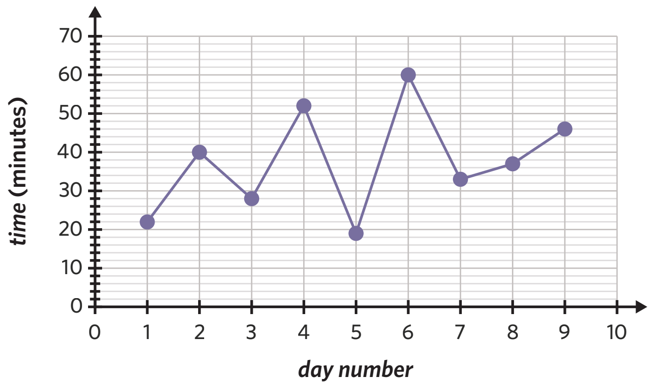
B. 4

C. 5

D. 6

E. 7

VCAA 2019 Exam 1 Data analysis Q13



45% of students answered this question correctly.

Questions from multiple lessons

Data analysis

16. The relationship between *vaccination rate* (%) and *illness rate* (%) for a particular illness can be modelled by a least squares regression line. The equation of the line is

illness rate = 66.41 − 0.61 × *vaccination rate*

The coefficient of determination for this relationship is 0.7012.

The Pearson correlation coefficient, *r*, is closest to

A. 0.8884 **B.** 0.8374 **C.** 0.4917 **D.** −0.4917 **E.** −0.8374

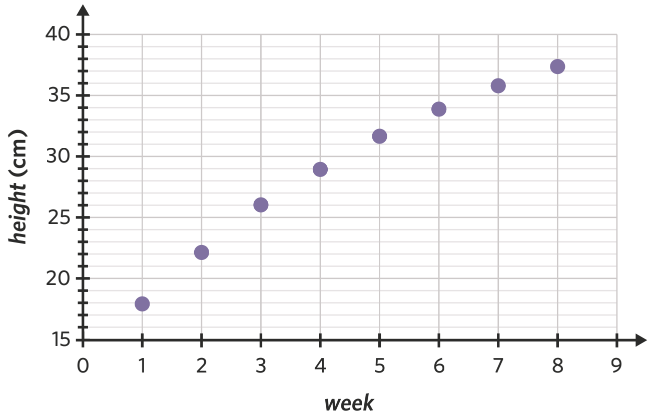
Adapted from VCAA 2018NH Exam 1 Data analysis Q13

Back to contents

Data analysis

17. The following table and scatterplot show the *height* of a plant after a number of weeks.

week	*height* (cm)
1	18.0
2	22.2
3	26.1
4	29.0
5	31.7
6	33.9
7	35.8
8	37.4



The scatterplot shows the relationship is non-linear.

A squared transformation is applied to the variable *height* to linearise the data.

A least squares regression line is then fitted to the data with *week* as the explanatory variable.

The equation of this line is closest to

A. $height^2 = 16.9 + 2.7 \times week$

B. $height^2 = 22.1 + 0.3 \times week$

C. $height^2 = 198.4 + 155.2 \times week$

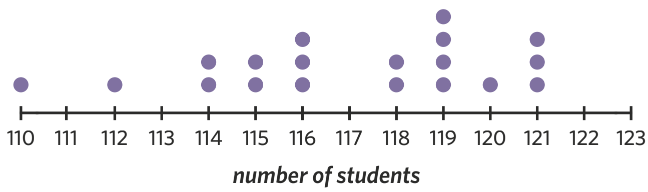
D. $week = -1.3 + 0.006 \times height^2$

E. $week = -64.8 + 3.1 \times height^2$

Adapted from VCAA 2017NH Exam 1 Data analysis Q12

Data analysis

18. The number of prep students at a school over the last 19 years was recorded. The data is displayed in the dot plot.



a. Determine the number of years in which there were more than 118 prep students. (1 MARK)

b. Determine if there is an outlier for the data, and if so, what is it? (2 MARKS)

Adapted from VCAA 2017NH Exam 2 Data analysis Q1

Back to contents

4D Seasonal adjustments

STUDY DESIGN DOT POINT

- seasonal adjustment including the use and interpretation of seasonal indices and their calculation using seasonal and yearly means

4A 4B 4C 4D 4E

KEY SKILLS

During this lesson, you will be:

- calculating and interpreting seasonal indices
- deseasonalising a time series
- reseasonalising a time series
- plotting and interpreting a deseasonalised time series.

KEY TERMS

- Seasonal index
- Deseasonalising
- Reseasonalising

Some time series are affected by seasonality, where there is predictable variation in the data over regular intervals, such as days, months, or quarters. Seasonality can be distracting and make it difficult to observe the overall trends. By deseasonalising a data set, it is possible to minimise the effect of seasonality.

Calculating and interpreting seasonal indices

A **seasonal index** is a measure of how a particular season compares to the average season.

See worked example 1

When only one cycle of seasonal data is given, a seasonal index can be calculated using the formula:

$$\textit{seasonal index} = \frac{\textit{season value}}{\textit{mean of all seasons}}$$

Once the seasonal index has been calculated, it can then be interpreted in relation to the average season. For example, a seasonal index of 1.25 indicates that the value recorded for that season is typically 125% of, or 25% greater than, the average season. A seasonal index of 0.65 indicates that the value recorded for that season is typically 65% of, or 35% less than, the average season.

Because a seasonal index is the comparison of one season against an average season, the sum of the seasonal indices is equal to the number of seasons, and their average is 1.

If the data spans multiple cycles, the seasonal index for each season is the average of its seasonal indices in each cycle.

See worked example 2

Worked example 1

Consider the following table.

	summer	autumn	winter	spring
umbrella sales	205	377	528	442

a. Use the table to calculate the seasonal index for *umbrella sales* in summer, correct to two decimal places.

Explanation

Step 1: Calculate the mean of all seasons.

$$\frac{205 + 377 + 528 + 442}{4} = 388$$

Continues →

Back to contents

Step 2: Calculate the seasonal index for summer.

$$\textit{seasonal index} = \frac{\textit{season value}}{\textit{mean of all seasons}}$$

$$= \frac{205}{388}$$

$$= 0.5283...$$

Answer

0.53

---

b. What does the rounded seasonal index for summer mean in terms of the *umbrella sales* in summer compared to the average season?

Explanation

Convert the seasonal index to a percentage.

$0.53 = 53\%$

Answer

On average, the number of umbrella sales in summer is 53% of, or 47% less than, the average season.

Worked example 2

For the following table, calculate the seasonal indices for summer, autumn, winter, and spring, correct to two decimal places.

	season			
year	summer	autumn	winter	spring
2019	45	30	22	51
2020	53	29	26	60
2021	54	41	29	72

Explanation

Step 1: Calculate the seasonal indices for each season in 2019.

$$\textit{mean of all seasons} = \frac{45 + 30 + 22 + 51}{4}$$

$$= 37$$

$$\textit{seasonal index} = \frac{\textit{season value}}{\textit{mean of all seasons}}$$

Summer 2019: $\textit{seasonal index} = \frac{45}{37} = 1.216...$

Autumn 2019: $\textit{seasonal index} = \frac{30}{37} = 0.810...$

Winter 2019: $\textit{seasonal index} = \frac{22}{37} = 0.594...$

Spring 2019: $\textit{seasonal index} = \frac{51}{37} = 1.378...$

Step 2: Calculate the seasonal indices for each season in 2020.

$$\textit{mean of all seasons} = \frac{53 + 29 + 26 + 60}{4}$$

$$= 42$$

$$\textit{seasonal index} = \frac{\textit{season value}}{\textit{mean of all seasons}}$$

Summer 2020: $\textit{seasonal index} = \frac{53}{42} = 1.261...$

Autumn 2020: $\textit{seasonal index} = \frac{29}{42} = 0.690...$

Winter 2020: $\textit{seasonal index} = \frac{26}{42} = 0.619...$

Spring 2020: $\textit{seasonal index} = \frac{60}{42} = 1.428...$

Continues →

Back to contents

Step 3: Calculate the seasonal indices for each season in 2021.

$$\textit{mean of all seasons} = \frac{54 + 41 + 29 + 72}{4}$$
$$= 49$$

$$\textit{seasonal index} = \frac{\textit{season value}}{\textit{mean of all seasons}}$$

Summer 2021: $\textit{seasonal index} = \frac{54}{49} = 1.102...$

Autumn 2021: $\textit{seasonal index} = \frac{41}{49} = 0.836...$

Winter 2021: $\textit{seasonal index} = \frac{29}{49} = 0.591...$

Spring 2021: $\textit{seasonal index} = \frac{72}{49} = 1.469...$

Step 4: Find the average seasonal indices by calculating the mean for each season.

Summer:
$\textit{seasonal index} = \frac{1.216...+ 1.261...+ 1.102...}{3} \approx 1.19$

Autumn:
$\textit{seasonal index} = \frac{0.810....+ 0.690...+ 0.836...}{3} \approx 0.78$

Winter:
$\textit{seasonal index} = \frac{0.594...+ 0.619...+ 0.591...}{3} \approx 0.60$

Spring:
$\textit{seasonal index} = \frac{1.378...+ 1.428...+ 1.469...}{3} \approx 1.43$

Answer

summer: 1.19, autumn: 0.78, winter: 0.60, spring: 1.43

Deseasonalising a time series

Deseasonalising a time series removes the seasonal variation from the data. Any underlying trends can then be observed.

See worked example 3

Deseasonalised values can be calculated using the formula:

$$\textit{deseasonalised value} = \frac{\textit{actual value}}{\textit{seasonal index}}$$

The concept of deseasonalisation can also be applied when analysing how data can be adjusted to correct for seasonality.

See worked example 4

Worked example 3

Use the given seasonal indices to deseasonalise the following time series data. Give values correct to the nearest whole number.

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
seasonal index	0.70	1.35	0.80	0.75	1.20	1.70	0.50
number of steps	7500	14 000	6000	8000	13 000	17 000	4000
deseasonalised *number of steps*							

Explanation

Step 1: Identify the relevant values for Monday.

$\textit{seasonal index} = 0.70$

$\textit{actual value} = 7500$

Step 2: Substitute the values into the formula.

$$\textit{deseasonalised value} = \frac{\textit{actual value}}{\textit{seasonal index}}$$
$$= \frac{7500}{0.70}$$
$$\approx 10\,714$$

Step 3: Repeat the process for the remaining days and fill in the table.

Answer

Mon: 10 714, Tue: 10 370, Wed: 7500, Thu: 10 667, Fri: 10 833, Sat: 10 000, Sun: 8000

Worked example 4

The seasonal index for January is 0.8. How should data for this season be adjusted to correct for seasonality?

Explanation

Step 1: Identify the seasonal index.

seasonal index $= 0.8$

Step 2: Find the reciprocal of the seasonal index.

Divide 1 by the seasonal index.

$1 \div 0.8 = 1.25$

Step 3: Interpret the value of the reciprocal.

The reciprocal value of 1.25 suggests that deseasonalised values are 125% of the actual value. This means that values for the season need to be increased by 25%.

Answer

The data from January should be increased by 25% to correct for seasonality.

Reseasonalising a time series

Reseasonalising a time series restores the normal seasonal fluctuations of previously deseasonalised data. This results in data that is representative of actual values in the data set.

Reseasonalised values can be calculated using the formula:

actual value $=$ *seasonal index* $\times$ *deseasonalised value*

Worked example 5

The seasonal index for Australian *Oodie sales* in January is 0.65.

Last January, a store in Sydney sold a deseasonalised number of 540 Oodies.

How many Oodies did the store actually sell in January?

Explanation

Step 1: Identify the relevant values.

seasonal index $= 0.65$

deseasonalised value $= 540$

Step 2: Substitute the values into the formula.

actual value $=$ *seasonal index* $\times$ *deseasonalised value*

$= 0.65 \times 540$

$= 351$

Answer

351 Oodies

Plotting and interpreting a deseasonalised time series

A deseasonalised time series can be plotted on a graph to visualise the underlying trends in seasonal data. It can be beneficial to deseasonalise a time series before fitting a regression line or equation.

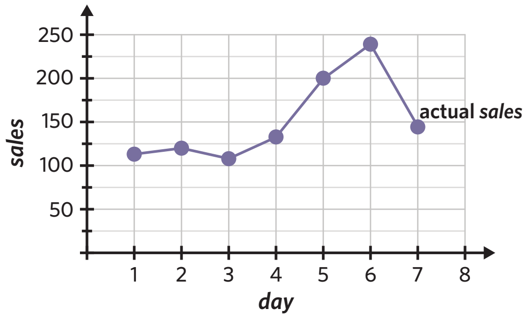
Back to contents

Worked example 6

Consider the following table.

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
coded *day*	1	2	3	4	5	6	7
sales	113	120	108	133	201	240	145
seasonal index	0.62	0.67	0.71	0.85	1.42	1.63	1.1
deseasonalised *sales*	182	179	152	156	142	147	132

The *sales* have been plotted on the following graph.

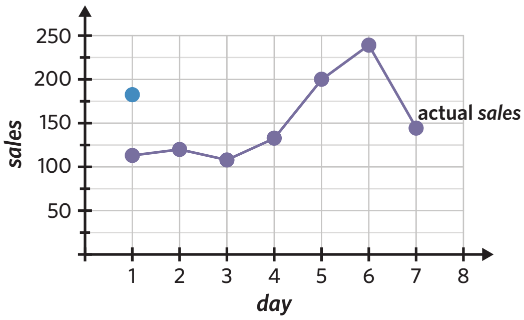


a. Plot the deseasonalised data on the graph.

Explanation

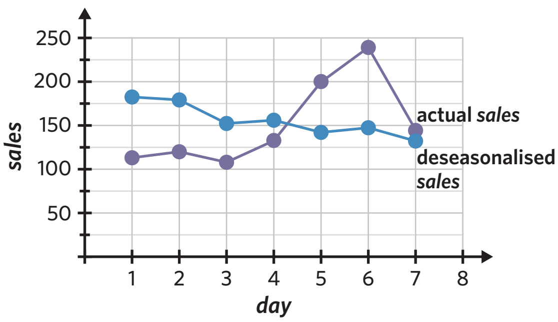
Step 1: Plot the deseasonalised sales for day 1.

Day 1 will have a horizontal axis value of 1 and a vertical axis value of 182.



Step 2: Plot the deseasonalised data for the rest of the days and join the points with a line.

Answer



Continues →

Back to contents

b. Identify any underlying trends using the plotted deseasonalised data.

Explanation

The data is decreasing from left to right.

Answer

There is a decreasing trend in *sales* over time.

Exam question breakdown

VCAA 2018 Exam 1 Data analysis Q16

The quarterly sales figures for a large suburban garden centre, in millions of dollars, for 2016 and 2017 are displayed in the following 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

Explanation

Step 1: Calculate the 2016 seasonal index for quarter 3.

$$\frac{1.73 + 2.87 + 3.34 + 1.23}{4} = 2.2925$$

$$\begin{aligned} \textit{seasonal index} &= \frac{3.34}{2.2925} \\ &= 1.456... \end{aligned}$$

Step 2: Calculate the 2017 seasonal index for quarter 3.

$$\frac{1.03 + 2.45 + 2.05 + 0.78}{4} = 1.5775$$

$$\begin{aligned} \textit{seasonal index} &= \frac{2.05}{1.5775} \\ &= 1.299... \end{aligned}$$

Step 3: Calculate the average seasonal index for quarter 3.

$$\frac{1.456... + 1.299...}{2} = 1.378...$$

Answer

C

51% of students answered this question correctly.

19% of students incorrectly answered option D. These students found the seasonal index for 2016 only, rather than the average for both 2016 and 2017.

4D Questions

Calculating and interpreting seasonal indices

1. The seasonal index for *sales* in June is 0.73.

Which statement is correct?

A. On average, *sales* in June are 73% less than the monthly average.

B. On average, *sales* in June are 27% less than the monthly average.

C. On average, *sales* in June are 27% more than the monthly average.

D. On average, *sales* in June are 73% more than the monthly average.

2. The quarterly *sales* ($000's) of a canine clothing shop are displayed in the following table.

	quarter 1	quarter 2	quarter 3	quarter 4
***sales* ($000's)**	33.5	61.3	52.0	43.4

a. Calculate the seasonal index for quarter 1, correct to two decimal places.

b. What does the rounded seasonal index for quarter 1 mean in terms of the amount of *sales* in quarter 1 compared to the average quarter?

---

3. Consider the following table.

	summer	autumn	winter	spring
seasonal index	0.65	1.15		0.80

What is the seasonal index for winter?

A. 0.87 **B.** 1.20 **C.** 1.25 **D.** 1.40

---

4. Calculate the seasonal indices for the following time series data. Give values correct to two decimal places.

a.

	summer	autumn	winter	spring
customers	55	78	96	75

b.

	day						
week	Mon	Tue	Wed	Thu	Fri	Sat	Sun
1	11	22	25	32	22	15	6
2	10	25	24	30	24	14	8

c.

	quarter			
year	1	2	3	4
2001	10	15	20	15
2002	11	17	19	15
2003	12	18	24	17

Deseasonalising a time series

5. Consider the following table.

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
seasonal index	0.87	0.94	0.91	0.85	0.96	1.26	1.21
value	603	665	509	670	702	931	950

The deseasonalised value for Thursday is closest to

A. 570 **B.** 583 **C.** 771 **D.** 788

---

6. The *number of weddings* held in the Botanical Gardens is seasonal, with the *number of weddings* in spring having a seasonal index of 1.32. If there were 58 weddings held in the Botanical Gardens this spring, the deseasonalised *number of weddings* is closest to

A. 39 **B.** 43 **C.** 44 **D.** 77

Back to contents

7. Use the seasonal indices to deseasonalise the following time series data. Give values correct to one decimal place.

a.

	summer	autumn	winter	spring
seasonal index	1.20	0.75	1.01	1.04
number of thunderstorms	37	30	29	33
deseasonalised *number of thunderstorms*				

b.

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
seasonal index	1.25	0.80	1.23	0.72	0.86	1.45	0.69
kilometres cycled	55	26	47	26	32	63	15
deseasonalised *kilometres cycled*							

8. The seasonal index for winter is 0.8. How should data from this season be adjusted to correct for seasonality?

A. Data from winter should be increased by 20%.

B. Data from winter should be increased by 25%.

C. Data from winter should be decreased by 20%.

D. Data from winter should be decreased by 25%.

9. The *number of holidays* booked by Essendon supporters in September has a seasonal index of 2.5.

a. To correct for seasonality, by what percentage does the *number of holidays* need to be decreased?

b. Calculate the deseasonalised *number of holidays* booked if a total of 28 400 holidays were booked by Essendon supporters in September 2022.

Reseasonalising a time series

10. The deseasonalised value for a particular season is 927. If the seasonal index is 0.75, the actual value is closest to

A. 695 B. 864 C. 927 D. 1236

11. Consider the following table.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
seasonal index	0.65	0.55	0.60	0.80	1.00	1.15	1.50	1.45	1.40	1.25	0.90	0.75
deseasonalised *value*	86	79	80	75	78	72	66	71	72	65	66	61

The actual value for July is

A. 44 B. 66 C. 83 D. 99

Back to contents

12. Baxter has collected the following data regarding the deseasonalised number of *reservations* at his restaurant during the week. The original data that contained the actual number of *reservations* on each day is no longer available. Use the given seasonal indices to reseasonalise the time series data. Give values correct to the nearest whole number where necessary.

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
seasonal index	0.65	0.60	0.54	0.75	1.40	1.85	1.21
deseasonalised *reservations*	64.62	81.67	53.70	74.67	60.00	83.24	104.13
actual *reservations*							

Plotting and interpreting a deseasonalised time series

13. The following graph shows the actual *sales* and deseasonalised *sales*.

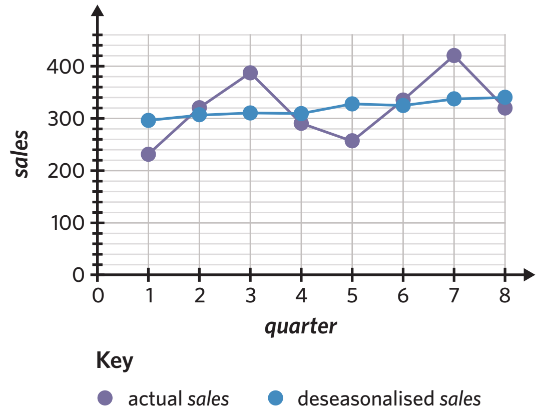
Which of the following statements is true?

A. When adjusted for seasonality, *sales* are increasing over time.

B. When adjusted for seasonality, *sales* are decreasing over time.

C. When adjusted for seasonality, *sales* are increasing over time with two large fluctuations.

D. When adjusted for seasonality, there is no underlying trend in *sales*.



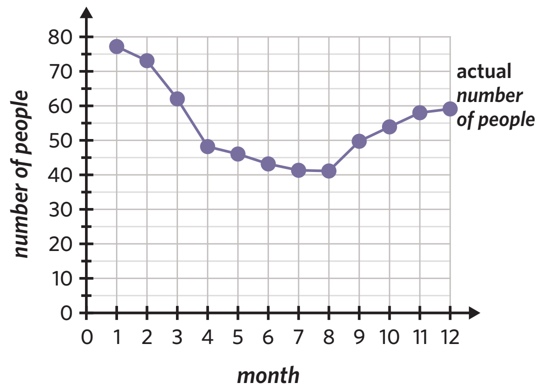
14. Hilaria is a fashion trend forecaster and for years has been researching the popularity of her least favourite sandal, the Birkenstock. The *number of people* wearing Birkenstocks at her workplace displays monthly seasonality, and her data from last year is shown in the following table.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
coded *month*	1	2	3	4	5	6	7	8	9	10	11	12
long-term seasonal index	1.25	1.21	1.06	0.86	0.82	0.81	0.79	0.77	0.96	1.10	1.16	1.21
actual *number of people*	77	73	62	48	46	43	41	41	50	54	58	59
deseasonalised *number of people*	61.6	60.3	58.5	55.8	56.1	53.1	51.9	53.2	52.1	49.1	50.0	48.8

Hilaria has constructed a graph to display the actual *number of people* seen wearing Birkenstocks each month last year.

a. Plot the deseasonalised values.

b. Identify any underlying trend visible in the deseasonalised data.



Joining it all together

15. The *sales* ($000's) for a churro company are seasonal, with the sales for 2020, 2021, and 2022 shown in the following table. The season index for autumn is 1.06 and the seasonal index for winter is 1.32.

	sales ($000's)			
year	summer	autumn	winter	spring
2020	2050	3650	4300	3125
2021	1900	3075	4150	3300
2022	1625	3300	4050	3250

a. Calculate the average seasonal indices of *sales* for summer and spring. Give values correct to two decimal places.

b. What does the seasonal index for autumn tell us about the amount of *sales* in autumn compared to the average season?

c. Fill in the gap in the following sentence, giving the value correct to the nearest percent.

To correct for seasonality, the amount of *sales* in winter should be decreased by __________%.

16. The *number of tourists* visiting a small patisserie in Paris each month is seasonal. The *number of tourists* who visited in 2022 is shown in the following table.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
coded *month*	1	2	3	4	5	6	7	8	9	10	11	12
long-term seasonal index	1.05	0.83	0.76	0.81	0.92	1.03	1.21	1.25	1.10		0.97	1.12
actual *number of tourists*	830	644		703	795	911	1123	1160	1045	910	934	1095
deseasonalised *number of tourists*	790.5	775.9	790.8	867.9	864.1	884.5	928.1	928.0	950.0	957.9	962.9	

a. What is the long-term seasonal index for October?

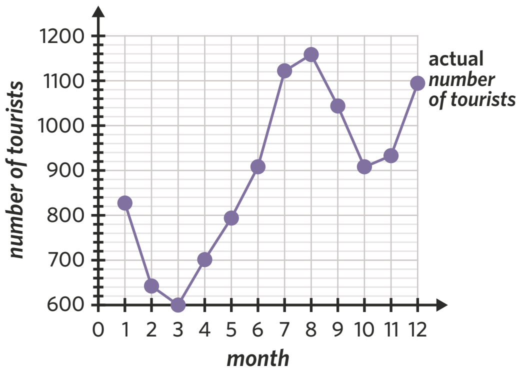
b. Calculate the actual *number of tourists* who visited the patisserie in March, correct to the nearest whole number.

c. Calculate the deseasonalised *number of tourists* who visited the patisserie in December, correct to one decimal place.

d. The actual *number of tourists* who visited the patisserie in 2022 are shown in the following graph.

Plot the deseasonalised *number of tourists*.

e. Describe any underlying trend in the data, as shown by the deseasonalised plot.



Exam practice

17. The following 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	Jun	Jul	Aug	Sep	Oct	Nov	Dec
***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 deseasonalised *rainfall* for May 2019 is closest to

A. 71.3 mm
B. 75.8 mm
C. 86.1 mm
D. 88.1 mm
E. 113.0 mm

VCAA 2020 Exam 1 Data analysis Q17

88% of students answered this question correctly.

Back to contents

18. 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. The seasonal index for winter is 1.41.

Use the values in Table 1 to find the seasonal indices for summer, autumn, and spring. (2 MARKS)

b. The *total rainfall* for each of the four seasons in 2017 is shown in Table 2.

Table 2

	***total rainfall* (mm)**			
year	summer	autumn	winter	spring
2017	141	156	262	120

Use the appropriate seasonal index to deseasonalise the *total rainfall* for winter in 2017. Round your answer to the nearest whole number. (1 MARK)

VCAA 2019 Exam 2 Data analysis Q6a,b

Part **a**: The average mark for this question was **0.9**.
Part **b**: **58%** of students answered this question correctly.

---

19. 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%.

VCAA 2017 Exam 1 Data analysis Q16

32% of students answered this question correctly.

Questions from multiple lessons

Data analysis *Year 11 content*

20. The following histogram displays the distribution of *yearly income* for 50 adults.

The shape of this distribution can be described as

A. symmetric.

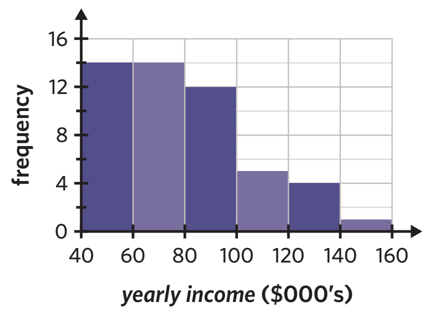
B. approximately symmetric.

C. approximately normal.

D. negatively skewed.

E. positively skewed.

Adapted from VCAA 2017NH Exam 1 Data analysis Q1



Back to contents

Data analysis *Year 11 content*

21. Displayed in the following table is the *length*, in centimetres, and *weight*, in kilograms, of 10 ferrets.

length **(cm)**	38	32	35	39	30	29	37	33	34	36
weight **(kg)**	1.1	0.8	1.5	1.4	1.1	0.9	1.6	1.6	1.2	1.3

Find the value of the Pearson correlation coefficient, r, between *height* and *weight*, correct to two decimal places.

A. -0.53 **B.** -0.28 **C.** 0.27 **D.** 0.28 **E.** 0.53

Adapted from VCAA 2018NH Exam 1 Data analysis Q9

Data analysis

22. The *height* and *width*, in centimetres, of 15 houseplants were recorded and a least squares regression line was fitted to the data. The equation of the line is $width = 4.8 + 0.484 \times height$.

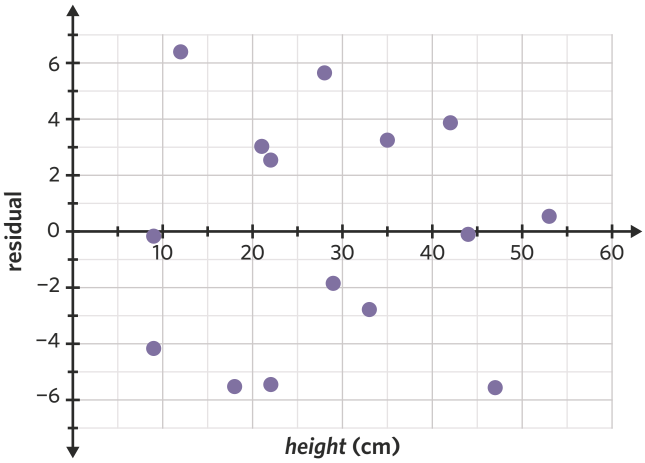
a. The regression equation is used to predict the *width* of a plant with a *height* of 47 cm. The residual value is calculated to be -5.55.

What is the actual *width* of this plant, correct to one decimal place? (1 MARK)

b. Pearson's correlation coefficient, r, for this data is 0.8549.

What percentage of variation in *width* can **not** be explained by the variation in *height*? Give your answer correct to one decimal place. (1 MARK)

c. When the regression line is fitted to the data, the following residual plot is acquired.



What information can be deduced from the residual plot about the association between *height* and *width* for these houseplants? (1 MARK)

Adapted from VCAA 2018NH Exam 2 Data analysis Q5d-f

Back to contents

4E Time series data and least squares regression modelling

STUDY DESIGN DOT POINT

- 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

4A 4B 4C 4D 4E

KEY SKILLS

During this lesson, you will be:

- modelling time series data
- modelling seasonal data.

It can be helpful to fit trend lines to time series data in order to make predictions about the future. If the data is seasonal, it must first be deseasonalised before fitting a trend line, and then reseasonalised after making predictions.

Modelling time series data

Trend lines for time series data are treated the same as least squares regression lines. After a trend line has been fitted, it can be used to make predictions outside the range of the data set. However, the limitations of extrapolation are still present. When extrapolating, it is assumed that the shape of the relationship between the variables will continue outside of the range of the data set. This assumption has limited reliability.

Worked example 1

A public library introduced a record collection to their catalogue. The *number of visitors* to the library each *day* following its introduction was recorded.

The day that the record collection was introduced is noted as day 0 and the first day after the introduction is day 1.

day	1	2	3	4	5	6	7
number of visitors	94	114	120	153	178	191	221

a. Determine the equation of the least squares regression line for the data. Give values correct to two decimal places where necessary.

Explanation – Method 1: TI-Nspire

Step 1: From the home screen, select '1: New' → '4: Add Lists & Spreadsheet'.

Step 2: Name column A 'day' and column B 'visitors'.

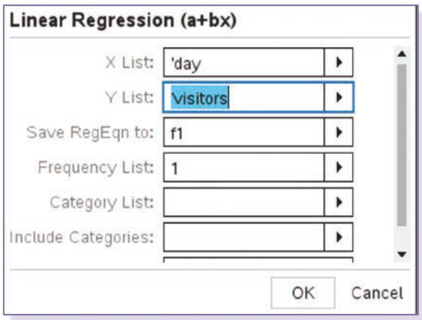
Enter the *day* values into column A, starting from row 1.

Enter the *number of visitors* values into column B, starting from row 1.

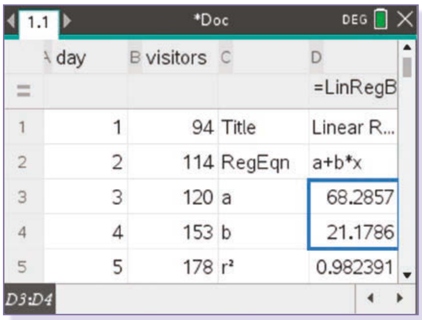
Continues →

Back to contents

Step 3: Press menu and select '4: Statistics' → '1: Stat Calculations' → '4: Linear Regression (a+bx)'. Select 'day' in 'X List:' and 'visitors' in 'Y List:'. Select 'OK'.



Step 4: Write down the equation for the least squares regression line, rounding the values of a and b to two decimal places.



$y = 68.29 + 21.18 \times x$

Step 5: Rewrite the equation in terms of the variables in the question.

The explanatory variable is *day*.

The response variable is *number of visitors*.

Explanation – Method 2: Casio ClassPad

Step 1: From the main menu, tap Statistics.

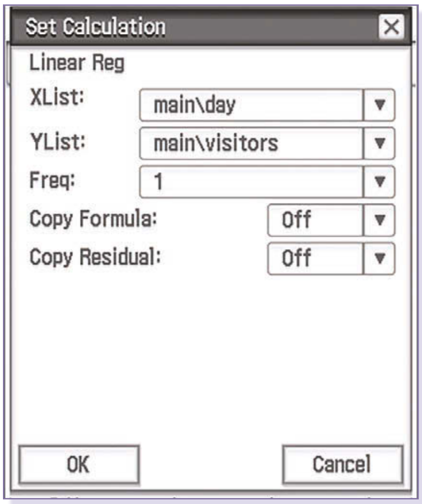
Step 2: Name the first list 'day' and the second list 'visitors'.

Enter the *day* values into list 'day', starting from row 1.

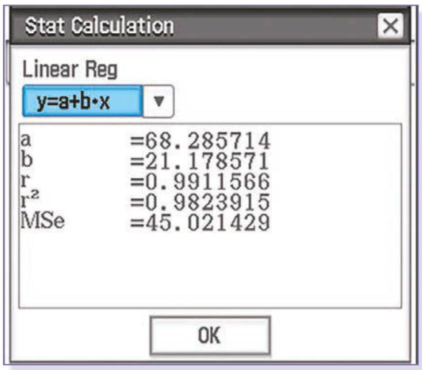
Enter the *number of visitors* values into list 'visitors', starting from row 1.

Step 3: Tap 'Calc' → 'Regression' → 'Linear Reg'.

Specify the data set by changing 'XList:' to 'main\day' and 'YList:' to 'main\visitors'. Tap 'OK'.



Step 4: Change the form of the equation in the drop down box to 'y=a+b·x'. Write down the equation for the least squares regression line, rounding the values of a and b to two decimal places.



$y = 68.29 + 21.18 \times x$

Step 5: Rewrite the equation in terms of the variables in the question.

The explanatory variable is *day*.

The response variable is *number of visitors*.

Answer – Method 1 and 2

number of visitors $= 68.29 + 21.18 \times$ *day*

Continues →

Back to contents

b. Use the rounded regression equation to estimate the *number of visitors* on the 8th day after the introduction. Round to the nearest whole number.

Explanation

Substitute *day* $= 8$ into the regression equation and evaluate.

$$\begin{aligned} \textit{number of visitors} &= 68.29 + 21.18 \times \textit{day} \\ &= 68.29 + 21.18 \times 8 \\ &= 237.73 \end{aligned}$$

Answer

238 visitors

---

c. Use the regression equation to estimate the *number of visitors* to the library on the day that the record collection was introduced. Round to the nearest whole number.

Explanation

The *number of visitors* on day 0 is equal to the y-intercept of the regression line.

The y-intercept is 68.29.

Answer

68 visitors

Modelling seasonal data

A trend line can be fitted to seasonal data in a similar way to regular time series data. However, additional steps are needed to improve the accuracy of predictions.

As seasonal data has many fluctuations, it is important to deseasonalise it before fitting a trend line. Recall that data can be deseasonalised using the formula:

$$\textit{deseasonalised value} = \frac{\textit{actual value}}{\textit{seasonal index}}$$

This means that any predictions made using the trend line will result in a deseasonalised value. Predictions must be reseasonalised afterwards using the formula:

$$\textit{actual value} = \textit{seasonal index} \times \textit{deseasonalised value}$$

Worked example 2

The *sales* of NRL jerseys at a merchandise store are seasonal. The *sales* for the first six months of 2022 are shown in the following table.

month	Jan	Feb	Mar	Apr	May	Jun
sales	479	513	1127	894	800	802

The long-term *seasonal index* for each month is also provided.

month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
seasonal index	0.55	0.63	1.40	1.21	1.11	1.19	1.23	1.19	1.25	0.58	0.52	1.14

Continues →

Back to contents

a. Determine the equation of the least squares regression line to predict the deseasonalised *sales* for a month. Give values correct to one decimal place.

Note: When deseasonalising *sales*, round values correct to one decimal place.

Explanation – Method 1: TI-Nspire

Step 1: Deseasonalise the *sales* values from January to June.

Divide each month's *sales* values by its corresponding seasonal index. Give values correct to one decimal place.

Jan: $\frac{479}{0.55} = 870.9$

Feb: $\frac{513}{0.63} = 814.3$

Mar: $\frac{1127}{1.40} = 805.0$

Apr: $\frac{894}{1.21} = 738.8$

May: $\frac{800}{1.11} = 720.7$

Jun: $\frac{802}{1.19} = 673.9$

Step 2: From the home screen, select '1: New' → '4: Add Lists & Spreadsheet'.

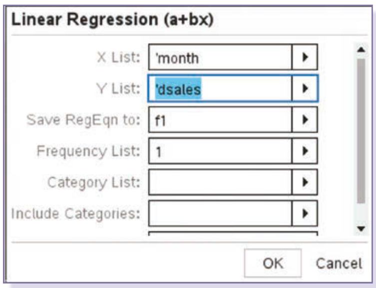
Step 3: Name column A 'month' and column B 'dsales'.

Enter the *month* values into column A, starting from row 1.

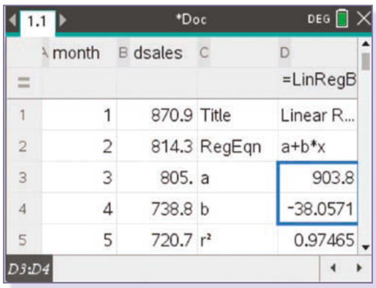
Note: Assign a value of 1 to January, 2 to February, and so on.

Enter the *deseasonalised sales* values into column B, starting from row 1.

Step 4: Press menu and select '4: Statistics' → '1: Stat Calculations' → '4: Linear Regression (a+bx)'. Select 'month' in 'X List:' and 'dsales' in 'Y List:'. Select 'OK'.



Step 5: Write down the equation for the least squares regression line, rounding the values of *a* and *b* correct to one decimal place.



$y = 903.8 - 38.1 \times x$

Step 6: Rewrite the equation in terms of the variables in the question.

The explanatory variable is *month*.

The response variable is *deseasonalised sales*.

Explanation – Method 2: Casio ClassPad

Step 1: Deseasonalise the *sales* values from January to June.

Divide each month's *sales* values by its corresponding seasonal index. Give values correct to one decimal place.

Jan: $\frac{479}{0.55} = 870.9$

Feb: $\frac{513}{0.63} = 814.3$

Mar: $\frac{1127}{1.40} = 805.0$

Apr: $\frac{894}{1.21} = 738.8$

May: $\frac{800}{1.11} = 720.7$

Jun: $\frac{802}{1.19} = 673.9$

Step 2: From the main menu, tap Statistics.

Step 3: Name the first list 'month' and the second list 'dsales'.

Enter the *month* values into list 'month', starting from row 1.

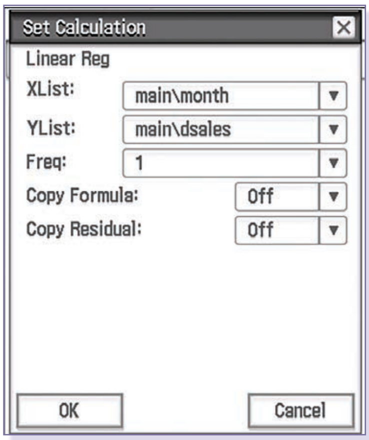
Note: Assign a value of 1 to January, 2 to February, and so on.

Enter the *deseasonalised sales* values into list 'dsales', starting from row 1.

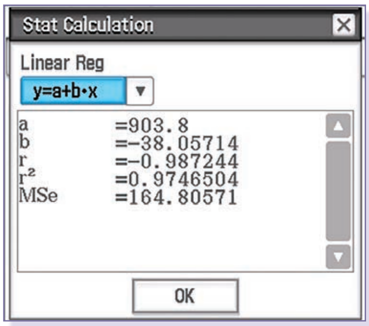
Continues →

Step 4: Tap 'Calc' → 'Regression' → 'Linear Reg'.

Specify the data set by changing 'XList:' to 'main\month' and 'YList:' to 'main\dsales'. Tap 'OK'.



Step 5: Change the form of the equation in the drop down box to 'y=a+b·x'. Write down the equation for the least squares regression line, rounding the values of *a* and *b* correct to one decimal place.



$y = 903.8 - 38.1 \times x$

Step 6: Rewrite the equation in terms of the variables in the question.

The explanatory variable is *month*.

The response variable is *deseasonalised sales*.

Answer – Method 1 and 2

$\textit{deseasonalised sales} = 903.8 - 38.1 \times \textit{month}$

b. Use the rounded regression equation to predict the *sales*, correct to the nearest whole number, for October 2022.

Explanation

Step 1: Calculate the *deseasonalised sales* for October.

October is the 10th month.

Substitute $\textit{month} = 10$ into the regression equation and evaluate.

$$\begin{aligned}\textit{deseasonalised sales} &= 903.8 - 38.1 \times \textit{month}\\ &= 903.8 - 38.1 \times 10\\ &= 522.8\end{aligned}$$

Step 2: Reseasonalise the value.

$$\begin{aligned}\textit{actual value} &= \textit{seasonal index} \times \textit{deseasonalised value}\\ &= 522.8 \times 0.58\\ &= 303.224\end{aligned}$$

Answer

303 sales

c. Interpret the slope of the rounded regression line in terms of the change in *deseasonalised sales* each month.

Explanation

Step 1: Identify the slope.

The slope of the regression line is 38.1.

Step 2: Interpret the slope.

The slope indicates the average change in the response variable for every one-unit increase in the explanatory variable.

Answer

On average, deseasonalised sales decrease by 38.1 each month.

Back to contents

Exam question breakdown

VCAA 2019 Exam 1 Data analysis Q14

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* (mins)**	22	40	28	51	19	60	33	37	46

The following time series plot was generated from this data.

A least squares line is to be fitted to the time series plot shown.

The equation of this least squares line, with *day number* as the explanatory variable, is closest to

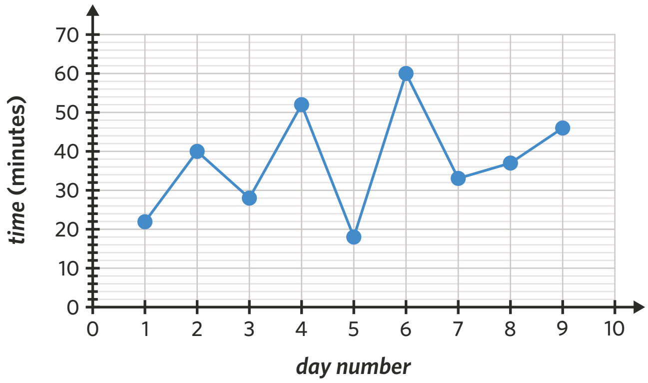
A. *day number* $= 23.8 + 2.29 \times$ *time*

B. *day number* $= 28.5 + 1.77 \times$ *time*

C. *time* $= 23.8 + 1.77 \times$ *day number*

D. *time* $= 23.8 + 2.29 \times$ *day number*

E. *time* $= 28.5 + 1.77 \times$ *day number*



Explanation – Method 1: TI-Nspire

Step 1: From the home screen, select '1: New' → '4: Add Lists & Spreadsheet'.

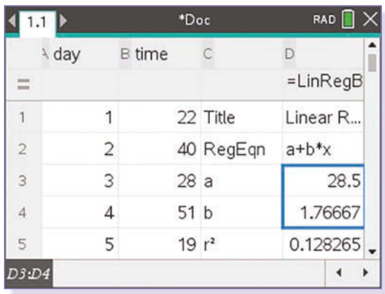
Step 2: Name column A 'day' and column B 'time'.

Enter the *day number* values into column A, starting from row 1.

Enter the *time* values into column B, starting from row 1.

Step 3: Press menu and select '4: Statistics' → '1: Stat Calculations' → '4: Linear Regression (a+bx)'. Select 'day' in 'X List:' and 'time' in 'Y List:'. Select 'OK'.

Step 4: Write down the equation for the least squares regression line, and round the values of a and b to the required number of decimal places.



$y = 28.5 + 1.77 \times x$

Step 5: Rewrite the equation in terms of the variables in the question.

The explanatory variable is *day number*.

The response variable is *time*.

time $= 28.5 + 1.77 \times$ *day number*

Explanation – Method 2: Casio ClassPad

Step 1: From the main menu, tap Statistics.

Step 2: Name the first list 'day' and the second list 'time'.

Enter the *day number* values into list 'day', starting from row 1.

Enter the *time* values into list 'time', starting from row 1.

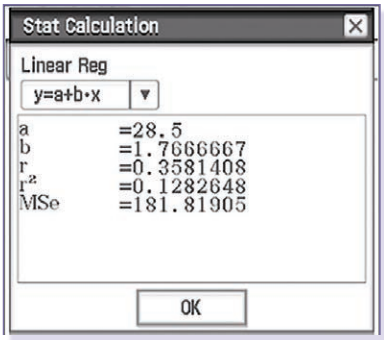
Step 3: Tap 'Calc' → 'Regression' → 'Linear Reg'.

Specify the data set by changing 'XList:' to 'main\day' and 'YList:' to 'main\time'. Tap 'OK'.

Continues →

Back to contents

Step 4: Change the form of the equation in the drop down box to 'y=a+b·x'. Write down the equation for the least squares regression line, and round the values of a and b to the required number of decimal places.



$y = 28.5 + 1.77 \times x$

Step 5: Rewrite the equation in terms of the variables in the question.

The explanatory variable is *day number*.

The response variable is *time*.

$time = 28.5 + 1.77 \times day\ number$

Answer – Method 1 and 2

E

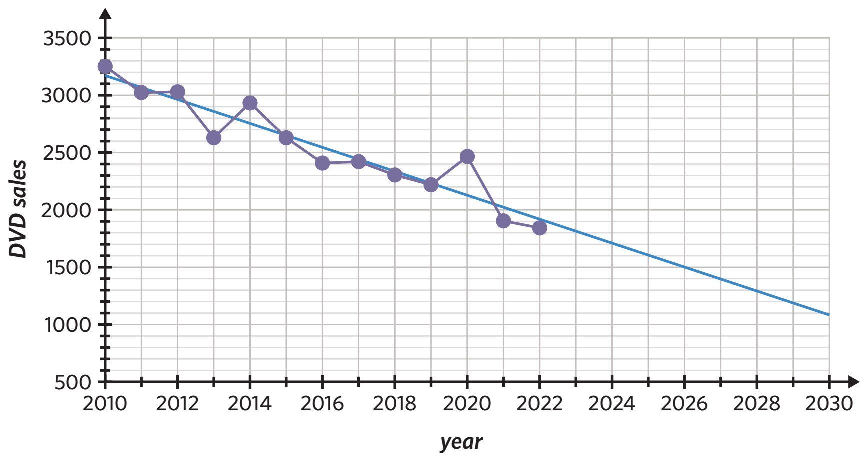
69% of students answered this question correctly.

12% of students incorrectly chose option B. This is likely because they mixed up the placement of the explanatory and response variables when replacing x and y with the actual variables.

4E Questions

Modelling time series data

1. An electronics store recorded their *DVD sales* between 2010 and 2022. They then fitted a regression line to the data, as shown.



From the regression line, *DVD sales* in 2030 are predicted to be closest to

A. 900 **B.** 1100 **C.** 1900 **D.** 2100

2. A new escape room has determined a regression equation to estimate their predicted *number of customers* each month for the upcoming year.

$number\ of\ customers = 316 + 32.4 \times month$

Note: In January, $month = 1$.

Predict the *number of customers*, correct to the nearest whole number, in

a. March. **b.** June. **c.** September. **d.** December.

Back to contents

3. Rafaella and Jeremiah recorded the daily *number of steps* they each walked for a week.

For each of them:

- find the regression equation used to predict the *number of steps* from the *day*. Round values correct to one decimal place.
- use the rounded regression equation to predict their *number of steps*, correct to the nearest whole number, on day 14.

a. Rafaella:

day	1	2	3	4	5	6	7
number of steps	7700	6550	9100	8600	8950	10 200	9850

b. Jeremiah:

day	1	2	3	4	5	6	7
number of steps	7450	8600	7000	6350	7150	6100	5200

4. Leonard conducted research into the number of *alien sightings* from 1975 to 2015 in the northern and southern hemispheres. He then used this data to construct regression equations to predict the number of *alien sightings* in each hemisphere for any given year.

Note: Leonard has used *year* $= 0$ to represent 1975.

The least squares regression equations are:

Southern hemisphere: *alien sightings* $= 208 + 23.4 \times$ *year*

Northern hemisphere: *alien sightings* $= 361 + 41.2 \times$ *year*

a. Estimate the difference between the number of *alien sightings* in each hemisphere in the year 2030, correct to the nearest whole number.

b. Explain why this prediction may be of limited reliability, even if Leonard's original data was correct.

c. Interpret the slope of the regression line in terms of the change in the number of *alien sightings* in the northern hemisphere each year.

5. Daniel is currently 6 weeks into a 16-week exchange program in the Netherlands.

He recorded his *bank balance* at the end of each *week* for the first 6 weeks, as shown in the table.

week	***bank balance*** **($)**
1	9058
2	8624
3	7580
4	7305
5	6617
6	6303

a. Daniel wants to keep track of his finances by calculating the least squares regression equation to predict his *bank balance* from *week*. Determine the regression equation, giving values correct to two decimal places.

b. Daniel is unsure whether he has enough money to fund the entire exchange. Use the rounded regression equation to predict Daniel's bank balance at the end of his exchange, correct to the nearest cent.

c. After how many full weeks on exchange will Daniel first have a *bank balance* under $3000?

d. On average, how much money does the rounded regression equation estimate that Daniel spends each week?

e. Using the rounded regression equation, estimate Daniel's bank balance at the start of his exchange.

Modelling seasonal data

6. A regression equation has predicted the *deseasonalised value* in September to be 3750.

If September has a long-term seasonal index of 1.09, the actual predicted *value* for September is closest to

A. 3413 **B.** 3440 **C.** 4088 **D.** 4121

7. The following equation can be used to forecast the *deseasonalised sales* ($) of a small business.

deseasonalised sales = 20 005 + 15 000 × *quarter number*

quarter number = 1 in quarter 1 of 2014 and *quarter number* = 2 in quarter 2 of 2014.

The seasonal index for the third quarter of each year is 0.65.

What is the actual value of *sales* in the third quarter of 2015?

8. The *average price*, in dollars, for a punnet of strawberries at Preston market was recorded for each season over 2021 and 2022. This is shown in the following table.

season	summer 2021	autumn 2021	winter 2021	spring 2021	summer 2022	autumn 2022	winter 2022	spring 2022
coded season	1	2	3	4	5	6	7	8
average price **($)**	2.15	2.80	4.15	3.95	2.60	3.65	5.20	5.25

The long-term seasonal indices for the seasons are also shown.

season	summer	autumn	winter	spring
seasonal index	0.68	0.9	1.26	1.16

The data was deseasonalised and a least squares regression line was fitted to the deseasonalised data.

Calculate the equation of the regression line for *deseasonalised average price* and *coded season*, giving values correct to two decimal places.

9. Calliope recorded the monthly *profit* ($) for her small business in 2022. She noticed that her *profit* was seasonal. She then calculated the *deseasonalised profit* ($) for each month in 2022, as shown in the following table.

month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
coded month	1	2	3	4	5	6	7	8	9	10	11	12
deseasonalised profit **($)**	8703	7105	7132	6810	7304	5923	6670	6543	5539	6435	6118	5605

a. Calculate the equation of the least squares regression line to predict *deseasonalised profit* ($) from *coded month*. Give values correct to one decimal place.

b. Calliope has calculated the long-term seasonal index for August to be 1.19. Predict Calliope's *profit*, correct to the nearest dollar, in August 2023 using the rounded regression equation from part **a**.

c. Is this prediction completely reliable? Explain briefly.

10. Eleanor is a freelance dog-walker. She recorded her quarterly *earnings*, in dollars, for 2021 and 2022.

quarter	Q1 2021	Q2 2021	Q3 2021	Q4 2021	Q1 2022	Q2 2022	Q3 2022	Q4 2022
coded quarter	1	2	3	4	5	6	7	8
earnings **($)**	8540	7515	7020	8890	9625	8145	8020	9950

Over the years, Eleanor has noticed that her *earnings* are seasonal.

The seasonal index for each quarter is shown in the following table.

quarter	Q1	Q2	Q3	Q4
seasonal index	1.18	0.92	0.89	1.06

a. Fill in the following table with Eleanor's *deseasonalised earnings*, correct to the nearest dollar.

quarter	Q1 2021	Q2 2021	Q3 2021	Q4 2021	Q1 2022	Q2 2022	Q3 2022	Q4 2022
coded quarter	1	2	3	4	5	6	7	8
deseasonalised earnings (\$)								

b. Plot the *deseasonalised earnings* against *coded quarter*. Comment on any visible trends and state whether a least squares regression line would be suitable.

c. Calculate the equation of the least squares regression line for *deseasonalised earnings* and *coded quarter*. Give values correct to the nearest whole number.

d. Use the rounded regression equation to predict Eleanor's *earnings* in Q2 2025.

e. In what quarter will Eleanor's *deseasonalised earnings* first be over \$10 500?

Joining it all together

11. Fatima is studying mathematics at university, and decided to record the *number of students* that attended her university library each day for the first two weeks of the semester.

	week 1							week 2						
day	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun
coded day	1	2	3	4	5	6	7	8	9	10	11	12	13	14
number of students	65	73	66	71	57	49	70	81	76	101	82	61	65	87

a. Fatima then fitted a least squares regression line to this data, to predict the *number of students* from *coded day*. Write down the equation of this line, giving values correct to two decimal places.

b. Each semester is 12 weeks long. Use the rounded regression to predict the *number of students*, correct to the nearest whole number, in the library on the last teaching day of the semester (Friday of Week 12).

Fatima later did some further research and discovered that the *number of students* in the library each day typically displays weekly seasonality. She calculates the long-term seasonal index for each day of the week, as shown.

day	Mon	Tue	Wed	Thu	Fri	Sat	Sun
seasonal index	1.14	1.07	1.12	1.06	0.81	0.77	1.03

c. Use the given seasonal indices to deseasonalise the original data. Fill in the following table, giving values correct to two decimal places.

	week 1							week 2						
day	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun
coded day	1	2	3	4	5	6	7	8	9	10	11	12	13	14
deseasonalised number of students														

d. Calculate the equation of the least squares regression line for the deseasonalised data, to predict the *deseasonalised number of students* from *coded day*. Give values correct to two decimal places.

e. Using the rounded regression equation from part **d**, calculate the predicted actual *number of students*, correct to the nearest whole number, in the library on the last teaching day of the semester.

f. Interpret the slope of the rounded regression line in terms of the change in *deseasonalised number of students* in the library each day of the semester.

Back to contents

Exam practice

12. The following table shows the yearly average traffic *congestion levels* in two cities, Melbourne and Sydney, during the period 2008 to 2016.

	***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 least squares line is used to model the trend in the time series plot for Sydney. The equation is

$\textit{congestion level} = -2280 + 1.15 \times \textit{year}$

a. i. 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.

Write in the box provided. (1 MARK)

☐ % per year

ii. Use the least squares line to predict when the percentage congestion level in Sydney will be 43%. (1 MARK)

b. 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.

Write the values of the intercept and the slope of this least squares line in the appropriate boxes provided.

Round both values to four significant figures. (2 MARKS)

$\textit{congestion level} = \square + \square \times \textit{year}$

VCAA 2018 Exam 2 Data analysis Q3bii,biii,d

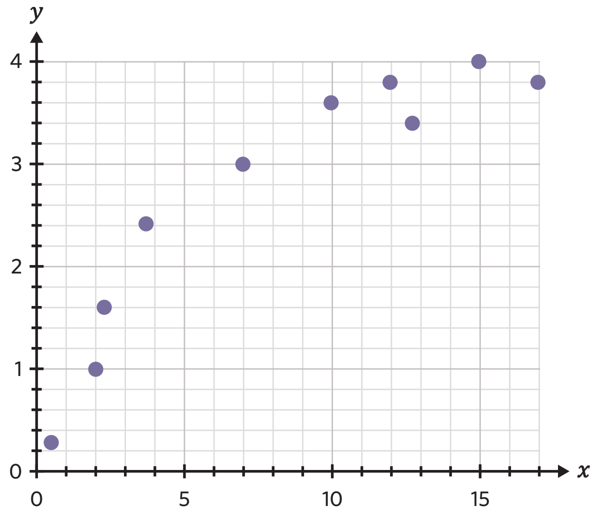
Part **ai**: **36%** of students answered this question correctly.
Part **aii**: **74%** of students answered this question correctly.
Part **b**: The average mark on this question was **1.4**.

Questions from multiple lessons

Data analysis

13. Annabelle uses the following data to generate the scatterplot shown.

x	y
15	4
7	3
3.7	2.4
0.5	0.3
17	3.8
2.3	1.6
12.7	3.4
12	3.8
2	1
10	3.6



The scatterplot demonstrates that the data is not linear, so to linearise the data, Annabelle performs a log transformation to the variable x.

Subsequently, she fits a least squares regression line to the transformed data.

Back to contents

With y as the response variable, the equation of this least squares regression line is closest to

A. $y = 0.78 + 2.59 \times \log(x)$

B. $y = -0.26 + 0.37 \times \log(x)$

C. $y = 1.03 + 0.20 \times \log(x)$

D. $\log(y) = 0.96 + 0.98 \times x$

E. $\log(y) = 1.75 + 3.49 \times x$

Adapted from VCAA 2018 Exam 1 Data analysis Q11

Data analysis

14. Disneyland records its long-term average number of visitors each day of the week.

The seasonal index for Tuesday is 0.79.

This means that, on average, the number of visitors to Disneyland on Tuesday is

A. 79% less than the daily average.

B. 21% less than the daily average.

C. the same as the daily average.

D. 21% more than the daily average.

E. 79% more than the daily average.

Adapted from VCAA 2016 Exam 1 Data analysis Q14

Data analysis

15. The least squares regression line that can predict a company's revenue, in dollars, based on its number of employees is

$\textit{revenue} = 51.3 + 48.6 \times \textit{number of employees}$

The correlation coefficient, r, for the relationship is 0.792.

a. Fizz Wizz is considered a small start-up company that produces confectionery for children. The company makes a total of $25 000 in revenue and has 400 employees. Calculate the residual when the least squares regression line is used to predict the revenue of Fizz Wizz from its number of employees. (1 MARK)

b. What percentage of variation in the amount of revenue can be explained by the variation in the number of employees? Round to one decimal place. (1 MARK)

Adapted from VCAA 2014 Exam 2 Data analysis Q2di,ii

Back to contents

AOS 2

Recursion and financial modelling

CALCULATOR QUICK LOOK-UP GUIDE

Back to contents

5

CHAPTER 5

Recurrence relations and basic financial applications

LESSONS

KEY KNOWLEDGE

- use of a first-order linear recurrence relation of the form: $u_0 = a$, $u_{n+1} = Ru_n + d$ 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
- 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.

Image: ChristianChan/Shutterstock.com

Back to contents

5A Recurrence relations and their graphs

STUDY DESIGN DOT POINT

- use of a first-order linear recurrence relation of the form: $u_0 = a$, $u_{n+1} = Ru_n + d$ where a, R and d are constants to generate the terms of a sequence

5A 5B 5C 5D 5E 5F 5G

KEY SKILLS

During this lesson, you will be:

- interpreting a recurrence relation
- constructing a recurrence relation
- generating a sequence of terms using a recurrence relation.

KEY TERMS

- Sequence
- Term
- Recurrence relation
- Common difference
- Common ratio
- Linear growth/decay
- Geometric growth/decay
- Iteration
- Arithmetic
- Geometric

Recurrence relations can be used to model different number patterns. They show how the value of a number in a sequence can be dependent on the previous number. Recurrence relations are useful for generating sequences of numbers and predicting future values.

Interpreting a recurrence relation

A **sequence** is a list of numbers written in succession. Each number in a sequence is a **term**. For example, a sequence might be described by u_n where n describes that term's position in the sequence. Often, an ellipsis (...) is used to show a sequence continues indefinitely.

For example, in the following sequence, the third term, u_2, is 2.

$u_n = 0, 1, 2, 3, \ldots$

A pattern-based sequence is where each successive term follows the same change. For example, in the previous sequence, each term increases by 1 so it is pattern-based. If there is no pattern determining the sequence, it is called random.

A **recurrence relation** is a formula that links each term in a pattern-based sequence to the next. It allows predictions to be made about the value of any term in a sequence. Recurrence relations comprise two parts: the initial value (u_0) and the pattern.

The pattern is written as $u_{n+1} = \textit{an expression involving } u_n$, where u is the sequence, u_n is the current term, and u_{n+1} is the next term.

The general form of a recurrence relation is:

$u_0 = a, \quad u_{n+1} = R \times u_n + d$, where

- a is the initial value of the recurrence relation
- d is the **common difference** which shows linear growth/decay
- R is the **common ratio** which shows geometric growth/decay

Linear growth/decay is when each term in a sequence increases or decreases by a constant amount. If $d > 0$ there is linear growth and if $d < 0$ there is linear decay.

Geometric growth/decay is when each term in a sequence increases or decreases by a constant ratio. When $R > 1$ there is geometric growth and when $R < 1$ there is geometric decay.

Worked example 1

Determine the initial value for the sequence from the following recurrence relations and describe the pattern in words.

a. $R_0 = 7, \quad R_{n+1} = R_n + 1$

Explanation

Step 1: Identify the initial value.

The subscript zero denotes the initial value, R_0, which is 7.

Step 2: Identify the pattern.

The pattern is given by $R_{n+1} = R_n + 1$.

Step 3: Describe the pattern.

In the recurrence relation, $d = 1$ which shows there is linear growth.

Each term increases by 1.

Answer

The initial value is 7. The next term is equal to the current term plus one.

b. $u_0 = 2, \quad u_{n+1} = 0.6 \times u_n$

Explanation

Step 1: Identify the initial value.

The subscript zero denotes the initial value, u_0, which is 2.

Step 2: Identify the pattern.

The pattern is given by $u_{n+1} = 0.6 \times u_n$.

Step 3: Describe the pattern.

In the recurrence relation, $R = 0.6$ which shows there is geometric decay.

Each term is 60% of the previous term or 40% less than the previous term.

Answer

The initial value is 2. The next term is 40% less than the current term.

c. $S_0 = 6, \quad S_{n+1} = 2 \times S_n - 4$

Explanation

Step 1: Identify the initial value.

The subscript zero denotes the initial value.

The initial value is S_0, which is 6.

Step 2: Identify the pattern.

The pattern is given by $S_{n+1} = 2 \times S_n - 4$.

Step 3: Describe the pattern.

In the recurrence relation, $R = 2$ and $d = -4$ so there is both geometric growth and linear decay.

Answer

The initial value is 6. The next term is found by multiplying the current value by two and subtracting four.

Back to contents

Constructing a recurrence relation

See worked example 2

See worked example 3

Recurrence relations can be constructed from a description of the pattern found in a sequence.

They can also be constructed from a number sequence by identifying whether the terms have a common difference, a common ratio or a combination of both.

A recurrence relation has the general form:

$u_0 = a, \quad u_{n+1} = R \times u_n + d$

If the terms increase or decrease by a constant amount, the sequence has a common difference (d), and $R = 1$.

Each term in the following sequence increases by 3, so there is a common difference, d, of 3.

$u_n = 4, 7, 10, 13, \ldots$

If the terms increase or decrease by a different amount but by a constant ratio, the sequence has a common ratio (R), and $d = 0$.

Each term in the sequence decreases by half, so there is a common ratio, R, of 0.5.

$v_n = 24, 12, 6, 3, \ldots$

If the sequence is pattern-based and has neither a common difference nor a common ratio, the recurrence relation will include both R and d ($R \neq 1$ and $d \neq 0$).

Worked example 2

Write down the recurrence relation for the following patterns in terms of u.

a. Each successive term is three times as big as the previous term, starting with 1.5.

Explanation

Step 1: Identify the initial value.

The recurrence relation starts with 1.5 so the initial value is 1.5.

$u_0 = 1.5$

Step 2: Identify the pattern.

The next term is three times as big as the current term. This means the next term is the current term multiplied by 3.

$u_{n+1} = 3 \times u_n$

Answer

$u_0 = 1.5, \quad u_{n+1} = 3 \times u_n$

b. Each successive term decreases by seven, starting with 34.

Explanation

Step 1: Identify the initial value.

The recurrence relation starts with 34 so the initial value is 34.

$u_0 = 34$

Step 2: Identify the pattern.

The next term is seven less than the current term. This means the next term is found by subtracting 7 from the current term.

$u_{n+1} = u_n - 7$

Answer

$u_0 = 34, \quad u_{n+1} = u_n - 7$

Back to contents

Worked example 3

Construct a recurrence relation to represent the following sequences in terms of u.

a. 23, 17, 11, 5, −1, ...

Explanation

Step 1: Identify the initial value (a).

The first term in the sequence is 23.

$a = 23$

Step 2: Identify if there is a common difference (d) or a common ratio (R).

Each successive term decreases by 6.

Therefore there is a common difference of −6, so $d = -6$.

Step 3: Determine the recurrence relation.

Substitute $a = 23$ and $d = -6$ into the recurrence relation: $u_0 = a,\ u_{n+1} = u_n + d$.

Answer

$u_0 = 23, \quad u_{n+1} = u_n - 6$

b. 5, 6, 7.2, 8.64, ...

Explanation

Step 1: Identify the initial value (a).

The first term in the sequence is 5.

$a = 5$

Step 2: Identify if there is a common difference (d) or a common ratio (R).

The initial term increases by 1, the second term increases by 1.2, and the third term increases by 1.44. Each term increases by a different amount so there is not a common difference.

To identify the common ratio, calculate the ratio of each term to the previous term.

$\frac{6}{5} = 1.2$

$\frac{7.2}{6} = 1.2$

$\frac{8.64}{7.2} = 1.2$

The common ratio is 1.2, so $R = 1.2$.

Step 3: Determine the recurrence relation.

Substitute $a = 5$ and $R = 1.2$ into the recurrence relation: $u_0 = a, \quad u_{n+1} = R \times u_n$.

Answer

$u_0 = 5, \quad u_{n+1} = 1.2 \times u_n$

c. 2, 5, 11, 23, 47

Explanation

Step 1: Identify the initial value (a).

The first term in the sequence is 2.

$a = 2$

Continues →

Back to contents

Step 2: Identify if there is a common difference (d) or a common ratio (R).

Each term increases by a different amount so there is not a common difference.

The ratio of the second term to the first term is 2.5 and the ratio of the third term to the second term is 2.2, so there is no common ratio.

Step 3: Determine the recurrence relation that generates the sequence.

Since there is no common difference or common ratio, check to see if there is a combination of both R and d.

Each successive term is one more than double the previous term. Therefore each term is multiplied by 2 and then added by 1.

This pattern is expressed as: $u_{n+1} = 2 \times u_n + 1$.

Answer

$u_0 = 2, \quad u_{n+1} = 2 \times u_n + 1$

Generating a sequence of terms using a recurrence relation

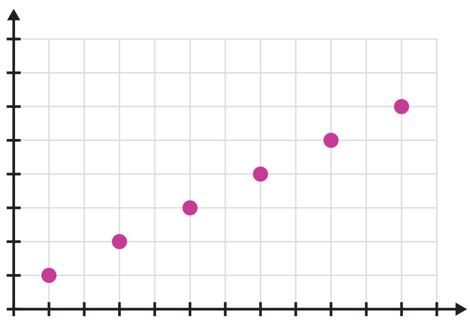
A recurrence relation of the form: $u_0 = a, \quad u_{n+1} = R \times u_n + d$ can be used to generate a sequence of numbers. A sequence is generated by substituting the initial value into the recurrence relation, and then repeating this for each term to determine the next term in the sequence. Each term in a repeated process is known as an **iteration**, where u_n is the value of the term in the sequence after n iterations.

See worked example 4

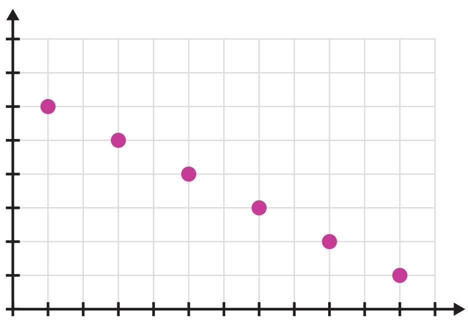
There are two types of sequences.

A recurrence relation that only has linear growth or decay, i.e. where $R = 1$, is **arithmetic**. It has the form: $u_0 = a, u_{n+1} = u_n + d$. It generates a sequence with a common difference.

Linear growth

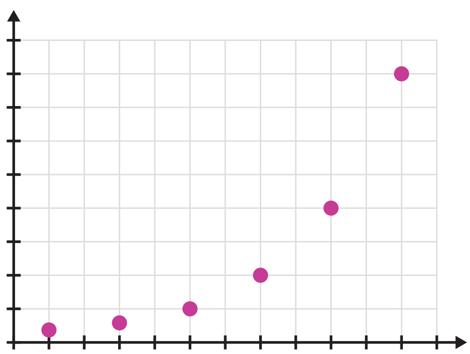


Linear decay

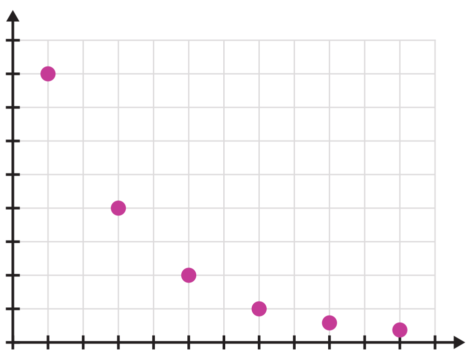


A recurrence relation that only has geometric growth or decay, i.e. where $d = 0$, is **geometric**. It has the form: $u_0 = a, \quad u_{n+1} = R \times u_n$. It generates a sequence with a common ratio.

Geometric growth



Geometric decay



A sequence can be generated from a recurrence relation either by hand or using a calculator.

See worked example 5

Back to contents

Worked example 4

Find the first five terms of the sequence given by the following recurrence relations.

a. $H_0 = 42, \quad H_{n+1} = H_n - 3.5$

Explanation

Step 1: Identify the initial value.

The initial value is denoted by the subscript zero, H_0, which is 42.

Step 2: Use the recurrence relation to find the second term, H_1.

$$H_1 = H_0 - 3.5$$
$$= 42 - 3.5$$
$$= 38.5$$

Step 3: Use the recurrence relation to find the third term, H_2.

$$H_2 = H_1 - 3.5$$
$$= 38.5 - 3.5$$
$$= 35$$

Step 4: Repeat for the remaining terms.

Step 5: Write the values in sequential order, separated by commas.

Answer

42, 38.5, 35, 31.5, 28

b. $Y_0 = 2, \quad Y_{n+1} = 4Y_n - 3$

Explanation

Step 1: Identify the initial value.

The initial value is denoted by the subscript zero, Y_0, which is 2.

Step 2: Use the recurrence relation to find the second term, Y_1.

$$Y_1 = 4 \times Y_0 - 3$$
$$= 4 \times 2 - 3$$
$$= 5$$

Step 3: Use the recurrence relation to find the third term, Y_2.

$$Y_2 = 4 \times Y_1 - 3$$
$$= 4 \times 5 - 3$$
$$= 17$$

Step 4: Repeat for the remaining terms.

Step 5: Write the values in sequential order, separated by commas.

Answer

2, 5, 17, 65, 257

Worked example 5

Use a calculator to find the first five terms of the sequence given by the following geometric recurrence relation, correct to two decimal places.

$B_0 = 2.12, \quad B_{n+1} = -3.76 \times B_n$

Explanation – Method 1: TI-Nspire

Step 1: Open a new page by pressing ctrl + doc▸ and select '1: Add Calculator'.

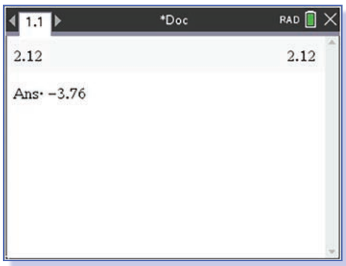
Step 2: Enter the initial value by typing '2.12'. Press enter.

Continues →

Back to contents

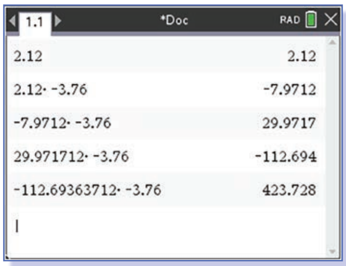
Step 3: Press '×'.

'Ans·' will appear to show that it is multiplying by the previous answer. Type '− 3.76'. Press enter.



Step 4: To find the next value in the sequence, press enter.

Step 5: Repeat step 4 for each of the remaining first five terms.



Explanation – Method 2: Casio ClassPad

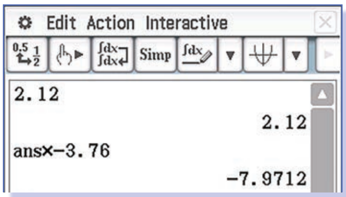
Step 1: From menu, tap √α Main.

Step 2: Enter the initial value by typing '2.12'. Press EXE.

Step 3: To calculate the second term B_1, type '×'.

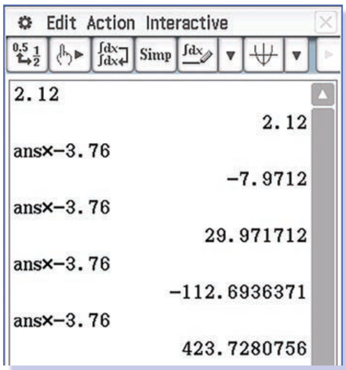
'ans' will appear to show that the calculator is using the previous answer.

Type '− 3.76'. Press EXE.



Step 4: To find the next value in the sequence, press EXE.

Step 5: Repeat step 4 for each of the remaining first five terms.



Answer – Method 1 and 2

2.12, −7.97, 29.97, −112.69, 423.73

Exam question breakdown

VCAA 2016 Exam 1 Recursion and financial modelling Q17

Consider the recurrence relation

$A_0 = 2, \quad A_{n+1} = 3A_n + 1$

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 …

Explanation

Step 1: Identify the initial value.

The initial value is denoted by the subscript zero.
The initial value is 2.

Step 2: Apply the pattern to find the second term, A_1.

$A_1 = 3 \times A_0 + 1$

$A_1 = 3 \times 2 + 1$

$A_1 = 7$

Step 3: Repeat for the remaining terms.

Answer

E

89% of students answered this question correctly.

5A Questions

Interpreting a recurrence relation

1. What is the initial term of the following recurrence relation?

$M_0 = 2, \quad M_{n+1} = M_n - 4$

A. -4 **B.** 0 **C.** 2 **D.** 4

---

2. Which statement is true of the following recurrence relation?

$W_0 = 15, \quad W_{n+1} = W_n + 2$

A. Each term is twice the previous term.

B. Each term is equal to the previous term plus 15.

C. Each term is 15.

D. Each term is equal to the previous term plus 2.

---

3. Determine the initial value for the sequence from the following recurrence relations. State whether the sequence has linear growth or decay, geometric growth or decay, or both.

a. $S_0 = 4, \quad S_{n+1} = 0.4 \times S_n$

b. $V_0 = 3.2, \quad V_{n+1} = 1.34 \times V_n - 6$

c. $u_0 = 14, \quad u_{n+1} = u_n + 1.3$

Constructing a recurrence relation

4. The first five terms of a sequence are 3, 7, 15, 31, 63 ...

The recurrence relation that generates this sequence is

A. $H_0 = 3, \quad H_{n+1} = H_n + 4$

B. $H_0 = 3, \quad H_{n+1} = 2 \times H_n + 1$

C. $H_0 = 3, \quad H_{n+1} = 2.333 \times H_n$

D. $H_0 = 3, \quad H_{n+1} = 4 \times H_n - 5$

---

5. Identify the common difference or common ratio in each of the following sequences.

a. 2, 4, 8, 16, 32, 64, ...

b. $-17, -10, -3, 4, 11, 18, \ldots$

c. 10, 9, 8.1, 7.29, 6.561, ...

d. $3, -3.6, 4.32, -5.184, 6.2208, \ldots$

---

6. Construct a recurrence relation for each of the following patterns.

a. Each successive term increases by four, starting with 22.

b. Each successive term is five times smaller than the previous term, starting with 100.

c. Each successive term is three less than double the previous term, starting with 14.

d. Each successive term is six more than 40% of the previous term, starting with 3.

7. Construct a recurrence relation for each of the following sequences in terms of u.

a. 4, 6.5, 9, 11.5, 14, 16.5, 19

b. 768, 192, 48, 12, 3

c. $\frac{1}{4}, \frac{5}{4}, \frac{25}{4}, \frac{125}{4}, \frac{625}{4}$

d. 3, 11, 35, 107, 323

8. The number of people, P_n, on day n eating at a restaurant as it gets more and more popular follows the following sequence. Note that this sequence starts at $n = 1$.

day (n)	1	2	3	4	5
people (P_n)	3	6	12	24	48

a. Assuming the sequence continues, how many people will eat at the restaurant on day 6?

b. Write a recurrence relation for this sequence.

Generating a sequence of terms using a recurrence relation

9. Using the following recurrence relation, what is the value of S_3?

$S_0 = 15, \quad S_{n+1} = S_n + 2$

A. 15 **B.** 19 **C.** 21 **D.** 23

10. Find the terms in each recurrence relation when $n = 2$, 4 and 7. Round to three decimal places where necessary.

a. $U_0 = 11, \quad U_{n+1} = U_n + 3$

b. $V_0 = \frac{1}{27}, \quad V_{n+1} = 3 \times V_n$

c. $W_0 = 3, \quad W_{n+1} = 2 \times W_n - 4$

d. $X_0 = 21, \quad X_{n+1} = -0.8 \times X_n + 2$

11. Generate the first five terms for the following recurrence relations.

a. $A_0 = 4.5, \quad A_{n+1} = A_n - 0.25$

b. $B_0 = 10, \quad B_{n+1} = 1.1 \times B_n$

c. $C_0 = 17, \quad C_{n+1} = 3 \times C_n - 22$

d. $D_0 = 12.5, \quad D_{n+1} = 0.8 \times D_n + 5$

12. Use a calculator to find the term G_{15} in the following sequence. Round to the nearest whole number if necessary.

$G_0 = 5, \quad G_{n+1} = 1.2 \times G_n + 3$

13. A sequence follows a pattern given by $M_{n+1} = M_n + 3$, where $M_4 = 23$. What is the initial value of the sequence?

Joining it all together

14. Consider the following recurrence relation:

$u_0 = 64, \quad u_{n+1} = 0.5 \times u_n + 4$

a. What is the initial value of the sequence generated by this recurrence relation?

b. Calculate the value of u_5.

15. Theo bought his first car when he got his learner permit two years ago. He bought it for \$3500. After a year, the car was revalued at \$2950. His insurance company inspected the car today and valued it at \$2400.

a. Assuming the arithmetic sequence continues, how many years after Theo purchased the car will it be worth \$750?

b. Let V_n be the value of the car n years after it was purchased. Write a recurrence relation to show the arithmetic sequence of the value of the car.

16. The height, h, in centimetres, of a ripple in a pond can be measured. The height of the ripple depends on the distance, d, in whole number metres, from the centre of the ripple. This relationship can be modelled using the geometric sequence $h_{d+1} = 0.9h_d$.

If the ripple was 1.458 cm high three metres from the centre of the ripple, what was the height of the ripple at the centre?

Exam practice

17. The first five terms of a sequence are 2, 6, 22, 86, 342 ...

The recurrence relation that generates this sequence could be

A. $P_0 = 2, \quad P_{n+1} = P_n + 4$

B. $P_0 = 2, \quad P_{n+1} = 2P_n + 2$

C. $P_0 = 2, \quad P_{n+1} = 3P_n$

D. $P_0 = 2, \quad P_{n+1} = 4P_n - 2$

E. $P_0 = 2, \quad P_{n+1} = 5P_n - 4$

VCAA 2017 Exam 1 Recursion and financial modelling Q18

92% of students answered this question correctly.

18. The following recurrence relation can generate a sequence of numbers.

$T_0 = 10, \quad T_{n+1} = T_n + 3$

The number 13 appears in this sequence as

A. T_1 **B.** T_2 **C.** T_3

D. T_{10} **E.** T_{13}

VCAA 2020 Exam 1 Recursion and financial modelling Q21

85% of students answered this question correctly.

19. The following recurrence relation can generate a sequence of numbers.

$L_0 = 37, \quad L_{n+1} = L_n + C$

The value of L_2 is 25.

The value of C is

A. -6 **B.** -4 **C.** 4

D. 6 **E.** 37

VCAA 2021 Exam 1 Recursion and financial modelling Q17

83% of students answered this question correctly.

20. Consider the following recurrence relation.

$A_0 = 3, \quad A_{n+1} = 2A_n + 4$

The value of A_3 in the sequence generated by this recurrence relation is given by

A. $2 \times 3 + 4$ **B.** $2 \times 4 + 4$ **C.** $2 \times 10 + 4$

D. $2 \times 24 + 4$ **E.** $2 \times 52 + 4$

VCAA 2019 Exam 1 Recursion and financial modelling Q17

68% of students answered this question correctly.

Back to contents

Questions from multiple lessons

Data analysis

21. A reciprocal transformation was used to linearise a set of non-linear bivariate data. Subsequently, a least squares line was fitted to the transformed data.

The equation is this least squares line is

$y = 7.25 + 1.27 \times \frac{1}{x}$

The equation was used to predict the value of y when x was set to a certain value. This gave a y value of 9.24.

The value of x is closest to

A. 1.59 **B.** 1.57 **C.** 0.64 **D.** 7.39 **E.** 18.98

Adapted from VCAA 2018 Exam 1 Data analysis Q12

Recursion and financial modelling *Year 11 content*

22. The following sequences are either five consecutive terms of an arithmetic sequence or five consecutive terms of a geometric sequence.

Which one of these sequences could **not** include 15 as a term?

A. 99, 82, 65, 48, 31...

B. −12, −9, −6, −3, 0...

C. 0.46875, 0.9375, 1.875, 3.75, 7.5...

D. −15, −10, −5, 0, 5...

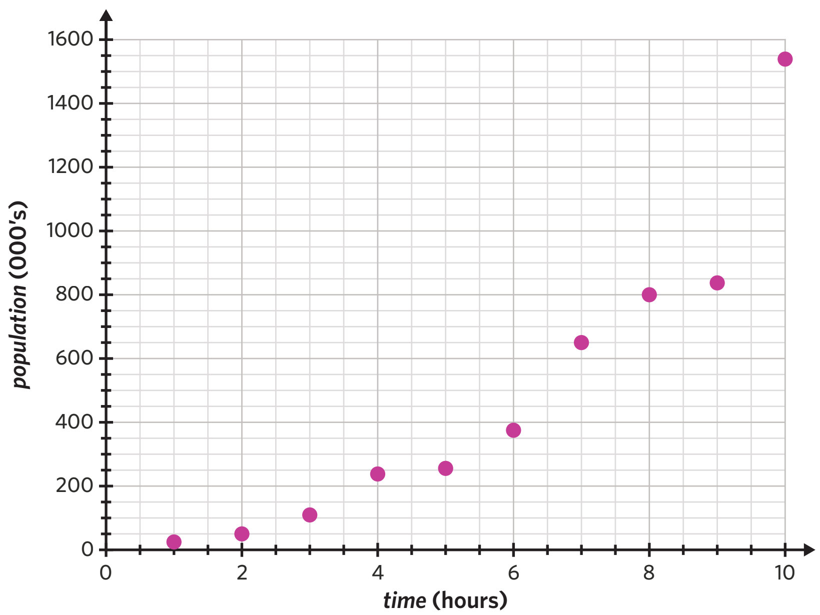
E. None of these.

Adapted from VCAA 2013 Exam 1 Number patterns Q7

Data analysis

23. Bacteria replicate and multiply over time.

The following time series plot shows the *population* of bacteria in a culture dish every hour for 10 hours. The data used to generate this plot is also given.



time (hours)	1	2	3	4	5	6	7	8	9	10
population (000's)	14.51	42.68	107.33	235.95	251.92	371.9	645.77	800.9	832.42	1533.93

Back to contents

The association between the *population* of the bacteria in the culture dish and *time* is non-linear.

A square transformation was performed on the variable *time* to linearise the data.

a. When the equation of the least squares regression line is calculated, the slope of this line is approximately 13.519406. Round this number to four significant figures. (1 MARK)

b. Perform the square transformation to the variable *time* to determine the equation of the least squares regression line that can be used to predict *population* from $time^2$. Write down the values of the intercept and slope of this equation. Round to four significant figures. (2 MARKS)

Adapted from VCAA 2017 Exam 2 Data analysis Q4a,b

Back to contents

5B Flat rate and unit cost depreciation – recurrence relations

STUDY DESIGN DOT POINT

- 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

5A 5B 5C 5D 5E 5F 5G

KEY SKILLS

During this lesson, you will be:

- modelling flat rate and unit cost depreciation using recurrence relations
- using an arithmetic recurrence relation to determine the value of an asset after n depreciation periods.

KEY TERMS

- Depreciation
- Flat rate depreciation
- Principal
- Unit cost depreciation

Recurrence relations can be applied to real-life cases such as the loss in value of an asset. That is, they can be used to model depreciation. The value of an asset after a certain number of periods, or units of use, can then be determined using a process of iteration on the recurrence relation.

Modelling flat rate and unit cost depreciation using recurrence relations

The loss of value of an asset is known as **depreciation**. Depreciation can either be arithmetic (linear) or geometric. There are two types of depreciation that can be represented by arithmetic relations; flat rate and unit cost.

See worked example 1

Flat rate depreciation is used when the value of an asset decreases by a constant amount each period. This amount is a percentage of the **principal**, or initial value, of the asset. For example, flat rate depreciation could describe the value of a computer as it gets outdated each year. A computer that had an initial value of \$2000 and depreciates at a flat rate of 20% per year, will depreciate by \$400 per year. It will take 5 years before it has no value.

A recurrence relation can be used to model flat rate depreciation and calculate the value of the asset after each period.

$V_0 = principal, \quad V_{n+1} = V_n - d$, where

- d is the depreciation per period, calculated by $d = \frac{r}{100} \times V_0$
- r is the depreciation rate (%) per period
- V_n is the value after n periods

See worked example 2

Unit cost depreciation is used when an asset loses value after each unit of use. For example, an air fryer might lose 5 cents in value for every hour of use. The unit of use is often different between cases. The unit of use for an air fryer could be hours of use, while the unit of use for a car could be kilometres driven.

A recurrence relation can be used to model unit cost depreciation and calculate the value of the asset after each unit of use.

$V_0 = principal, \quad V_{n+1} = V_n - d$, where

- d is the depreciation amount per unit of use
- V_n is the value after n uses

Back to contents

Worked example 1

Determine the recurrence relation that models the value of the assets, V_n, after n years, in each of the scenarios.

a. Karl does not believe in banks. Instead of investing his money, he buries his \$20 000 in his backyard. Unfortunately for Karl, inflation causes his buried cash to depreciate at a constant amount per year, equal to 3% of the amount of money that he buried.

Explanation

Step 1: Determine the initial value, V_0.

Initially Karl buries \$20 000.

$V_0 = 20\ 000$

Step 2: Calculate the depreciation amount, d.

$$d = \frac{r}{100} \times V_0$$

The rate of depreciation is 3%, so $r = 3$.

$$d = \frac{3}{100} \times 20\ 000 = 600$$

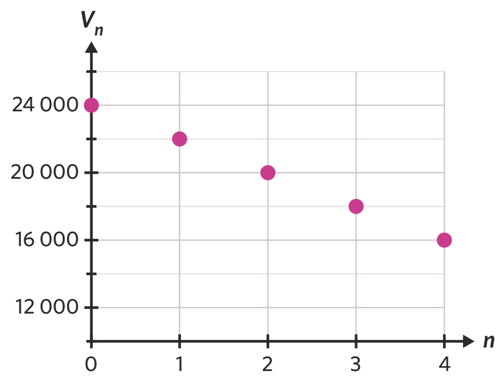
Step 3: Substitute these values into the recurrence relation.

$V_0 = principal, \quad V_{n+1} = V_n - d$

Answer

$V_0 = 20\ 000, \quad V_{n+1} = V_n - 600$

b. The graph shows the value of a car, purchased for \$24 000, over a period of four years.



Explanation

Step 1: Determine the initial value, V_0.

The value of the car when purchased was \$24 000. This is the vertical axis intercept of the graph.

$V_0 = 24\ 000$

Step 2: Calculate the depreciation amount, d.

$V_1 = 22\ 000$

This means that the value of the car depreciates by \$2000 each year.

$d = 2000$

Step 3: Substitute these values into the recurrence relation.

$V_0 = principal, \quad V_{n+1} = V_n - d$

Answer

$V_0 = 24\ 000, \quad V_{n+1} = V_n - 2000$

Back to contents

Worked example 2

Determine the recurrence relation that models the value of the assets, V_n after n units of use, in each of the following scenarios.

a. David spends \$33 on a disposable underwater camera and estimates that it will depreciate by \$1.50 for each photo that he takes.

Explanation

Step 1: Determine the initial value, V_0.

The camera was purchased for \$33.

$V_0 = 33$

Step 2: Determine the depreciation amount, d.

The value of the camera depreciates by \$1.50 per unit of use.

$d = 1.5$

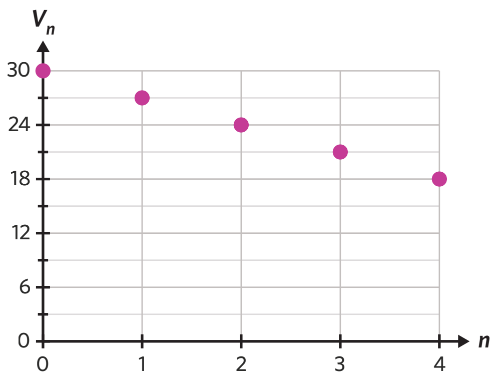
Step 3: Substitute these values into the recurrence relation.

$V_0 = principal, \quad V_{n+1} = V_n - d$

Answer

$V_0 = 33, \quad V_{n+1} = V_n - 1.5$

b. Chloe purchases a book for \$30. The book depreciates each time it is read. The value of the book is shown in the graph.



Explanation

Step 1: Determine the initial value, V_0.

The book is purchased for \$30.

$V_0 = 30$

Step 2: Determine the depreciation amount, d.

$V_1 = 27$

This means that the value of the book depreciates by \$3 each time it is read.

$d = 3$

Step 3: Substitute these values into the recurrence relation.

$V_0 = principal, \quad V_{n+1} = V_n - d$

Answer

$V_0 = 30, \quad V_{n+1} = V_n - 3$

Back to contents

Using an arithmetic recurrence relation to determine the value of an asset after *n* depreciation periods

A recurrence relation that models depreciation can be used to determine the value of an asset after a given number of periods, or a given number of units of use. The value is found by the process of iteration.

When modelling flat rate depreciation, the initial value, V_0, is substituted into the recurrence relation to find the value of the asset after one period, V_1. V_1 is then substituted into the relation to find the value after two periods, V_2, and so on.

See worked example 3

When modelling unit cost depreciation, the initial value, V_0, is substituted into the recurrence relation to find the value of the asset after one unit of use, V_1. V_1 is then substituted into the relation to find the value after two units of use, V_2, and so on.

See worked example 4

Worked example 3

The value of a brand new laptop, V_n after n years, is depreciated using flat rate depreciation. This is represented by the following recurrence relation.

$V_0 = 2100, \quad V_{n+1} = V_n - 350$

a. What is the value of the laptop two years after it is purchased?

Explanation

Step 1: Identify the initial value, V_0.

$V_0 = 2100$

Step 2: Calculate V_1, the value after one year.

$$\begin{aligned} V_1 &= V_0 - 350 \\ &= 2100 - 350 \\ &= 1750 \end{aligned}$$

Step 3: Calculate V_2, the value after two years.

$$\begin{aligned} V_2 &= V_1 - 350 \\ &= 1750 - 350 \\ &= 1400 \end{aligned}$$

Note: These calculations can also be performed using a calculator as shown in 5A Worked example 5.

Answer

$1400

b. Graph the value of the laptop over the first five years.

Explanation

Step 1: Calculate V_1 to V_5, the values over the first five years.

$V_0 = 2100$

$V_1 = V_0 - 350 = 2100 - 350 = 1750$

$V_2 = V_1 - 350 = 1750 - 350 = 1400$

$V_3 = V_2 - 350 = 1400 - 350 = 1050$

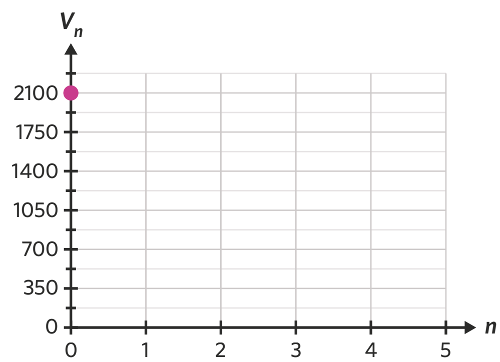
$V_4 = V_3 - 350 = 1050 - 350 = 700$

$V_5 = V_4 - 350 = 700 - 350 = 350$

Note: These calculations can also be performed using a calculator as shown in 5A Worked example 5.

Step 2: Plot V_0 on a graph.

V_0 will be the vertical axis intercept.

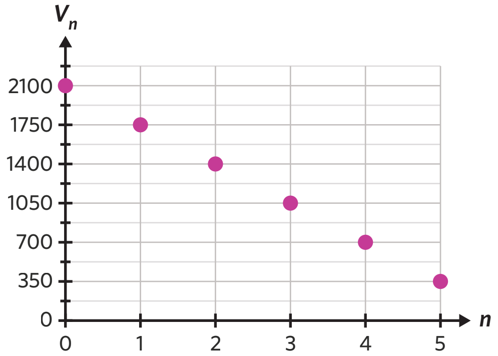


Step 3: Plot the remaining values on the graph.

Continues →

Back to contents

Answer



Worked example 4

A new piece of equipment was purchased for \$300 000. The equipment depreciates by \$10 000 for every 10 uses, which means that it depreciates by \$1000 for each use.

The value, V_n, of the equipment after n uses can be modelled by the following recurrence relation.

$V_0 = 300\ 000, \quad V_{n+1} = V_n - 1000$

a. After how many uses will the value of the equipment be \$297 000?

Explanation

Step 1: Determine the initial value, V_0.

$V_0 = 300\ 000$

Step 2: Calculate the value after each use until it reaches \$297 000.

Value after one use:

$$\begin{aligned} V_1 &= V_0 - 1000 \\ &= 300\ 000 - 1000 \\ &= 299\ 000 \end{aligned}$$

Value after two uses:

$$\begin{aligned} V_2 &= V_1 - 1000 \\ &= 299\ 000 - 1000 \\ &= 298\ 000 \end{aligned}$$

Value after three uses:

$$\begin{aligned} V_3 &= V_2 - 1000 \\ &= 298\ 000 - 1000 \\ &= 297\ 000 \end{aligned}$$

Note: These calculations can also be performed using a calculator as shown in 5A Worked example 5.

Answer

3 uses

Continues →

Back to contents

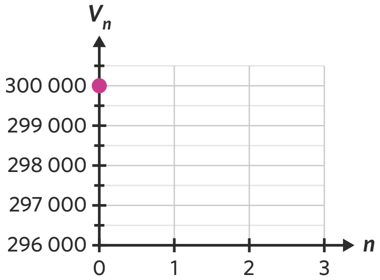
b. Using the values found in part **a**, graph the unit cost depreciation of the equipment after three uses.

Explanation

Step 1: Determine the initial value, V_0, and plot this on a graph.

$V_0 = 300\ 000$

This will be the vertical axis intercept.



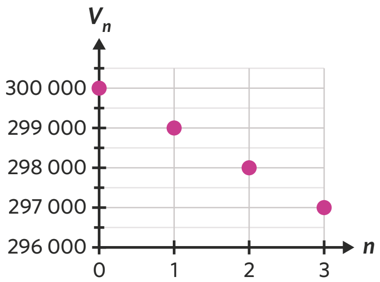
Step 2: Plot the values found in part **a** on the graph.

$V_1 = 299\ 000$

$V_2 = 298\ 000$

$V_3 = 297\ 000$

Answer



Exam question breakdown

VCAA 2017 Exam 1 Recursion and financial modelling Q21

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.

Explanation

Step 1: Identify the initial value.

The printer was purchased for \$680.

Step 2: Calculate the depreciation per year.

After four years, the printer is worth \$125.

$$\textit{depreciation over four years} = 680 - 125$$
$$= 555$$

$$\textit{depreciation per year} = \frac{555}{4}$$
$$= 138.75$$

Step 3: Calculate the depreciation per page.

On average, 1920 pages were printed each year.

$$\textit{depreciation per page} = \frac{138.75}{1920}$$
$$= 0.0722...$$

Step 4: Convert the depreciation to cents.

\$0.0722 is 7.22 cents.

Answer

E

53% of students answered this question correctly.

22% of students incorrectly chose option A. This is likely because the students divided the total depreciation over 4 years by the annual amount of pages that were printed.

Back to contents

5B Questions

Modelling flat rate and unit cost depreciation using recurrence relations

1. Courtney's yacht loses its value each time she uses it. The following recurrence relation models the depreciation of her yacht using the unit cost method.

$V_0 = 100\,000, \quad V_{n+1} = V_n - 1500$

Fill in the following gaps.

Courtney's yacht was worth __________ when brand new. The value of her yacht decreases by __________ every time she uses it. This amount can be expressed as __________% of the principal.

2. Write a recurrence relation to model each of the following scenarios.

a. A car is depreciating at \$2000 per year. It was bought for \$18 000.

b. A laptop was bought for \$800, and depreciates at a rate of 16% of its sale price per year.

c. A bike bought for \$575 depreciates at a rate of 10% of its sale price per year.

3. Write a recurrence relation to model each of the following scenarios.

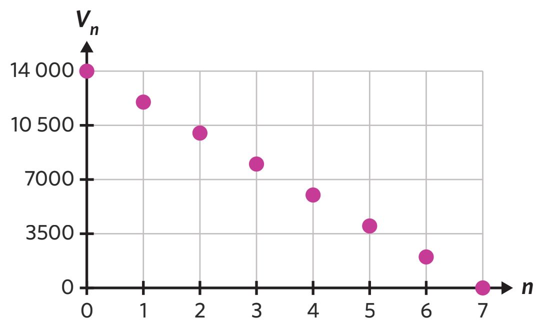
a. A cricket bat worth \$300 depreciates at \$2 per innings played with it.

b. A \$20 000 car depreciates at five cents per kilometre driven.

c. A printer worth \$500 depreciates at 0.002% of its sale price for every page printed.

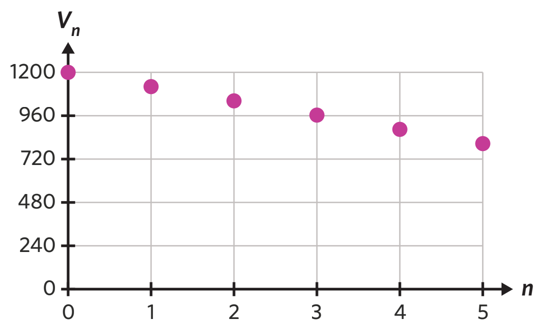
4. A brand new tractor was purchased for \$14 000. The tractor will be worthless after it is used 700 times. The graph shown models the depreciation of the tractor using the unit cost method.

Write a recurrence relation that models the value, V_n, of the tractor where n represents every 100 uses.



5. The following graph models the flat rate depreciation of a mobile phone after n years.

Determine the recurrence relation that models the depreciation of the mobile phone.



Using an arithmetic recurrence relation to determine the value of an asset after *n* depreciation periods

6. Isaac bought new DJ decks for \$740 which depreciate by 20% of the initial value each year. This is represented by the following recurrence relation.

$V_0 = 740, \quad V_{n+1} = V_n - 148$

How much will the DJ decks be worth after 2 years?

A. \$296 **B.** \$444 **C.** \$592 **D.** \$740

Back to contents

7. The following recurrence relation models flat rate depreciation.

$V_0 = 10, \quad V_{n+1} = V_n - 2$

a. Find the values of V_1, V_2, V_3 and V_4.

b. Plot the depreciation over four periods.

---

8. Let V_n represent the value of an item after n uses, in dollars. The value of the item depreciating using the unit cost method could be modelled using the following recurrence relation.

$V_0 = 184, \quad V_{n+1} = V_n - 14$

a. What is the initial value of the item?

b. What will the value of the item be after four uses?

c. After how many uses will the item be worth \$100?

---

9. A brand new refrigerator costs \$930. The refrigerator depreciates by \$132.86 every 10 000 times it is opened and closed. Let V_n represent the value of the refrigerator after n units of use, in dollars, where n represents every 10 000 uses.

$V_0 = 930, \quad V_{n+1} = V_n - 132.86$

a. How much is the refrigerator worth after it is opened and closed 20 000 times?

b. How many times will the refrigerator need to be opened and closed to be worthless? Round to the nearest ten thousand.

Joining it all together

10. A microwave bought for \$600 depreciates by 5.5% of its sale price each year.

a. Write a recurrence relation that describes its value, V_n, after n years.

b. How much will it be worth in five years?

---

11. Tristan's mobile tablet depreciates according to the following recurrence relation, where V_n is its value after it has been dropped n times.

$V_0 = 1200, \quad V_{n+1} = V_n - 210$

a. What is the tablet worth after it has been dropped twice?

b. After how many times being dropped will the tablet be worth less than half of its original value?

c. Express the loss in value each time the tablet is dropped as a percentage of the original sale price.

---

12. A brand new sedan made by Luxury Motors, that was bought for \$84 000, is worth \$44 000 four years later. Its value depreciates over time using the flat rate method.

a. Determine the recurrence relation that models this scenario, where V_n represents the value of the sedan after n years.

b. Plot the value of the sedan on a graph, from the time of purchase up to $n = 5$.

c. After how many years will the value drop below \$56 000?

---

13. Rachel bought a new watch for \$280. The watch depreciates by \$22 for every 1000 rotations of the hour hand.

a. Write a recurrence relation that describes the unit cost depreciation of the watch, where n represents 1000 rotations of the hour hand.

b. What is the value of the watch after 4000 rotations of the hour hand?

c. The hour hand rotates twice in one day. After how many days will the watch be worth \$236?

Back to contents

Exam practice

14. 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_n, is

A. $V_0 = 4500, \quad V_{n+1} = V_n - 450$

B. $V_0 = 4500, \quad V_{n+1} = V_n + 450$

C. $V_0 = 4500, \quad V_{n+1} = 0.9V_n$

D. $V_0 = 4500, \quad V_{n+1} = 1.1V_n$

E. $V_0 = 4500, \quad V_{n+1} = 0.1(V_n - 450)$

VCAA 2019 Exam 1 Recursion and financial modelling Q19

76% of students answered this question correctly.

---

15. Samuel owns a printing machine.

The printing machine, in dollars, after n years, V_n, can be modelled by the recurrence relation

$V_0 = 120\,000, \quad V_{n+1} = V_n - 15\,000$

Showing recursive calculations, determine the value of the machine, in dollars, after two years. (1 MARK)

VCAA 2020 Exam 2 Recursion and financial modelling Q7b

72% of students answered this question correctly.

---

16. Phil is a builder who has purchased a large set of tools.

The initial value of the tools was \$60 000.

He can depreciate 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. (1 MARK)

Adapted from VCAA 2019 Exam 2 Recursion and financial modelling Q7d

42% of students answered this type of question correctly.

---

17. Ken's first caravan had a purchase price of \$38 000.

After eight years, the value of the caravan was \$16 000.

The caravan has travelled an average of 5000 km 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? (1 MARK)

Adapted from VCAA 2016 Exam 2 Recursion and financial modelling Q6c

29% of students answered this type of question correctly.

Questions from multiple lessons

Data analysis

18. The number of passengers on a particular 7:30 am Melbourne train was continuously monitored.

The following table displays the long-term average number of *passengers* and seasonal index for every day of the week.

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
long-term average *passengers*	175	163	158	166	150	83	64
seasonal index	1.28	1.19	1.15	1.21	1.09	0.61	0.47

On Sunday last week, there was a second-hand market in the morning. As a result, the train contained 125 *passengers*. The deseasonalised number of *passengers* last Sunday was closest to

A. 184 **B.** 191 **C.** 205 **D.** 236 **E.** 266

Adapted from VCAA 2016 Exam 1 Data analysis Q15

Back to contents

Recursion and financial modelling *Year 11 content*

19. The following graph displays the growth in the value of a $200 investment over five years.

Ferdinand invested a different sum of money under the same investment conditions and received a total of $450 in interest only over a period of nine years.

How much money did Ferdinand invest?

A. $400

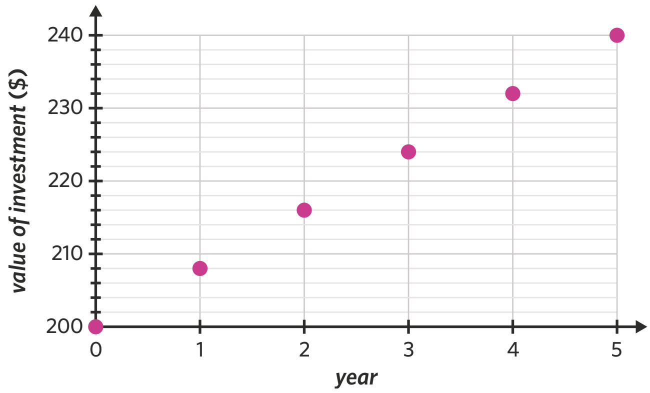
B. $612

C. $1125

D. $1250

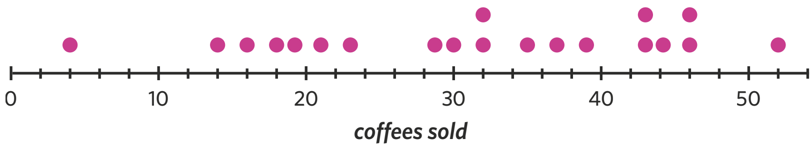
E. $2250

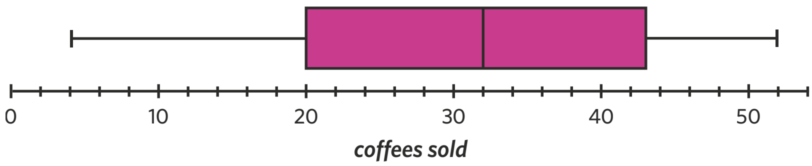
Adapted from VCAA 2013 Exam 1 Business-related mathematics Q7



Data analysis

20. Advertising for a cafe can lead to more coffees being sold. The following dot plot and boxplot both show the increase of *coffees sold*, per day for 20 days, due to an increased number of advertisements displayed around the city.





a. Describe the shape of the distribution of the increase in coffees sold for the 20 days. (1 MARK)

b. The value of 4 is above the lower fence and is not an outlier.

Determine the value of the lower fence. (1 MARK)

Adapted from VCAA 2018 Exam 2 Data analysis Q1f,g

Back to contents

5C Reducing balance depreciation – recurrence relations

STUDY DESIGN DOT POINT

- 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

5A 5B 5C 5D 5E 5F 5G

KEY SKILLS

During this lesson, you will be:

- modelling reducing balance depreciation using recurrence relations
- using a geometric recurrence relation to determine the value of an asset after n depreciation periods.

KEY TERM

- Reducing balance depreciation

Recurrence relations can be used to model the loss in the value of an asset over time. As seen previously, arithmetic depreciation is where the value of an asset decreases by a constant amount. However, some assets instead depreciate by a fixed ratio each period, and are represented by geometric recurrence relations.

Modelling reducing balance depreciation using recurrence relations

Reducing balance depreciation occurs when the value of an asset depreciates by a constant ratio, rather than a constant amount, each time period. Rather than the depreciation amount being a set value, it will be a percentage of the previous value. This amount changes each period.

For example, a computer that had an initial value of \$2000 and depreciates by 5% per year according to the reducing balance method, will depreciate by $0.05 \times 2000 = \$100$ in the first year, $0.05 \times 1900 = \$95$ in the second year, and so on.

A geometric recurrence relation can be used to model reducing balance depreciation and calculate the value of the asset from one period to the next.

$V_0 = principal, \quad V_{n+1} = R \times V_n$, where

- $R = 1 - \frac{r}{100}$
- r is the depreciation rate (%) per period
- V_n is the value after n periods

Back to contents

Worked example 1

Determine the recurrence relation that models the depreciation in each of the following scenarios.

a. A tablet is worth \$550 and is depreciated using a reducing balance depreciation method with a rate of 4.3% per annum.

Explanation

Step 1: Determine the initial value, V_0.

The tablet is initially worth \$550.

$V_0 = 550$

Step 2: Calculate the value of R.

$R = 1 - \frac{r}{100}$

The reducing balance rate is 4.3%, so $r = 4.3$.

$$R = 1 - \frac{4.3}{100}$$
$$= 0.957$$

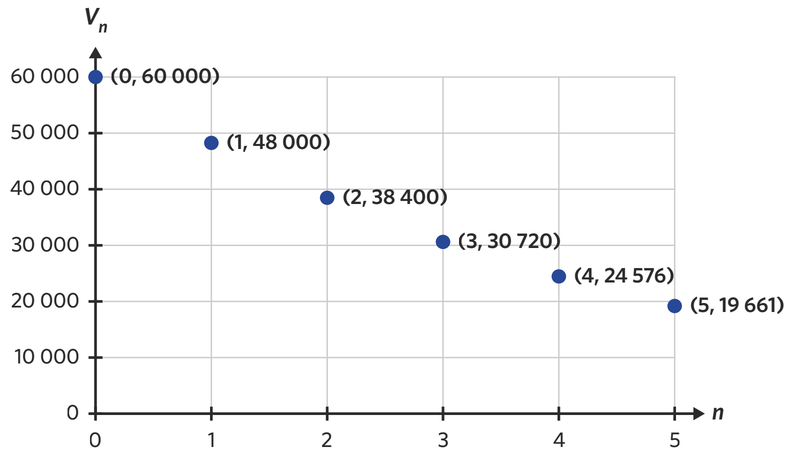
Step 3: Substitute these values into the recurrence relation.

$V_0 = principal, \quad V_{n+1} = R \times V_n$

Answer

$V_0 = 550, \quad V_{n+1} = 0.957 \times V_n$

b. The graph shows the value of a boat, purchased for \$60 000, over a period of four years.



Explanation

Step 1: Determine the initial value, V_0.

The boat was purchased for \$60 000.

$V_0 = 60\,000$

Step 2: Calculate the ratio, R, of the value of the boat from each period to the next.

Initial value to the end of the first year:

$\frac{48\,000}{60\,000} = 0.8$

First year to second year:

$\frac{38\,000}{48\,000} = 0.8$

By repeating this for the remaining years, it can be seen that there is a constant ratio, R, of 0.8.

Step 3: Substitute these values into the recurrence relation.

$V_0 = principal, \quad V_{n+1} = R \times V_n$

Answer

$V_0 = 60\,000, \quad V_{n+1} = 0.8 \times V_n$

Back to contents

Using a geometric recurrence relation to determine the value of an asset after *n* depreciation periods

A recurrence relation that models depreciation can be used to determine the value of an asset after a given number of periods. The value is found by the process of iteration.

When modelling reducing balance depreciation, the initial value, V_0, is substituted into the recurrence relation to find the value of the asset after one period, V_1. V_1 is then substituted into the relation to find the value after two periods, V_2, and so on.

Worked example 2

A tablet is worth $550 and is depreciated using a reducing balance depreciation method with a rate of 4.3% per annum.

A recurrence relation is constructed to represent the depreciation of the tablet:

$V_0 = 550, \quad V_{n+1} = 0.957 \times V_n$

a. After how many years will the value of the tablet drop below $505?

Explanation

Step 1: Determine the initial value, V_0.

$V_0 = 550$

Step 2: Calculate the value at the end of each year until the tablet is less than $505.

Value after 1 year:

$$\begin{aligned} V_1 &= 0.957 \times V_0 \\ &= 0.957 \times 550 \\ &= 526.35 \end{aligned}$$

Value after 2 years:

$$\begin{aligned} V_2 &= 0.957 \times V_1 \\ &= 0.957 \times 526.35 \\ &= 503.716... \end{aligned}$$

Note: These calculations can also be performed using a calculator as shown in 5A Worked example 5.

Answer

2 years

b. After three years, what is the value of the tablet?

Explanation

Calculate the value after 3 years.

The value after 2 years, V_2, was found in part **a**.

$$V_2 = 503.716...$$

$$\begin{aligned} V_3 &= 0.957 \times V_2 \\ &= 0.957 \times 503.716... \\ &= 482.057... \end{aligned}$$

Answer

$482.06

Continues →

Back to contents

c. What is the amount of depreciation in the third year?

Explanation

Calculate the amount of depreciation in year 3.

Use the value of the tablet after year 2 and year 3.

$V_2 = 503.72$

$V_3 = 482.06$

The amount of depreciation is the difference between the value after year 2 and 3.

$503.72 - 482.06 = 21.66$

Answer

$21.66

---

d. Graph the value of the tablet over the first four years.

Explanation

Step 1: Identity the values over the first four years, V_1 to V_4.

$V_0 = 550$

$V_1 = 526.35$

$V_2 = 503.72$

$V_3 = 482.06$

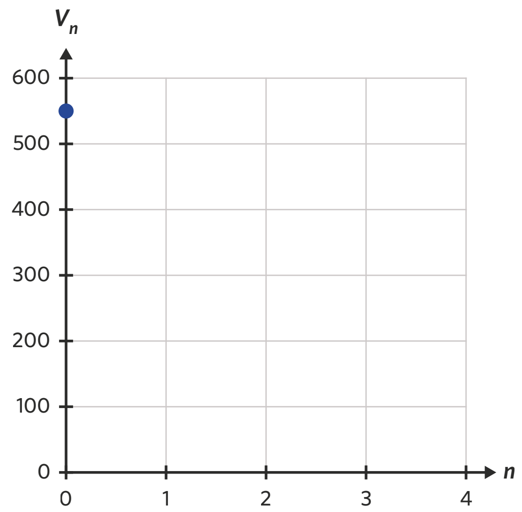
$V_4 = 0.957 \times V_3$

$= 0.957 \times 482.06$

≈ 461.33

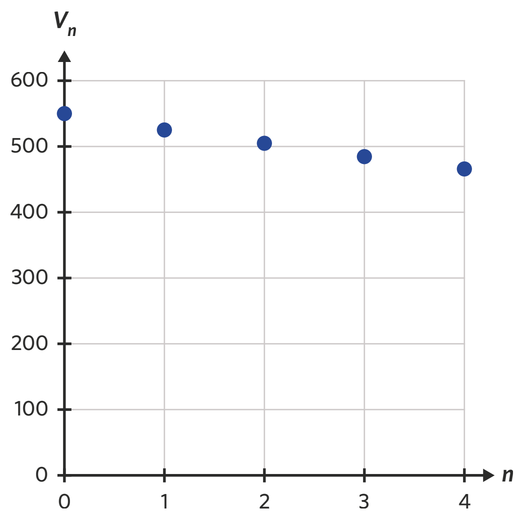
Step 2: Plot V_0 on a graph.

V_0 will be the vertical axis intercept.



Step 3: Plot the remaining values on the graph.

Answer



Back to contents

Exam question breakdown

VCAA 2016 Exam 1 Recursion and financial modelling Q19

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. $C_0 = 26\,000, \quad C_{n+1} = 0.92C_n$

B. $C_0 = 26\,000, \quad C_{n+1} = 1.08C_n$

C. $C_0 = 26\,000, \quad C_{n+1} = C_n + 8$

D. $C_0 = 26\,000, \quad C_{n+1} = C_n - 8$

E. $C_0 = 26\,000, \quad C_{n+1} = 0.92C_n - 8$

Explanation

Step 1: Determine the initial value, C_0.

The purchase price of the car was \$26 000.

$C_0 = 26\,000$

Step 2: Calculate the value of R.

$R = 1 - \frac{r}{100}$

The reducing balance rate is 8%, so $r = 8$.

$R = 1 - \frac{8}{100}$

$= 0.92$

Step 3: Substitute these values into the recurrence relation.

$C_0 = \textit{principal}, \quad C_{n+1} = R \times C_n$

$C_0 = 26\,000, \quad C_{n+1} = 0.92 \times C_n$

74% of students answered this question correctly.

14% of students incorrectly answered option B. The recurrence relation in option B is modelling geometric growth, since $R > 1$. Depreciation is a decrease in an asset's value, therefore R must be less than 1.

Answer

A

5C Questions

Modelling reducing balance depreciation using recurrence relations

1. Alex recently purchased a new tractor to use on his farm for \$13 000. The value of the tractor depreciates on a reducing balance basis at a rate of 10% per annum. Which of the following recurrence relations models the depreciation of Alex's tractor?

A. $V_0 = 13\,000, \quad V_{n+1} = V_n - 1300$

B. $V_0 = 13\,000, \quad V_{n+1} = V_n + 1300$

C. $V_0 = 13\,000, \quad V_{n+1} = 1.1V_n$

D. $V_0 = 13\,000, \quad V_{n+1} = 0.9V_n$

---

2. The value of Ella's car, V_n, at year n, is depreciating according to the following recurrence relation.

$V_0 = 2000, \quad V_{n+1} = \left(1 - \frac{6}{100}\right) \times V_n$

a. How much did Ella pay for her car?

b. What is the depreciation rate per period of Ella's car, as a percentage?

---

3. Write the recurrence relations, O_n and A_n, to model the reducing balance depreciation of the following items, n years after purchase.

a. Oliver's camera was bought for \$600, and depreciates at a rate of 8% per year.

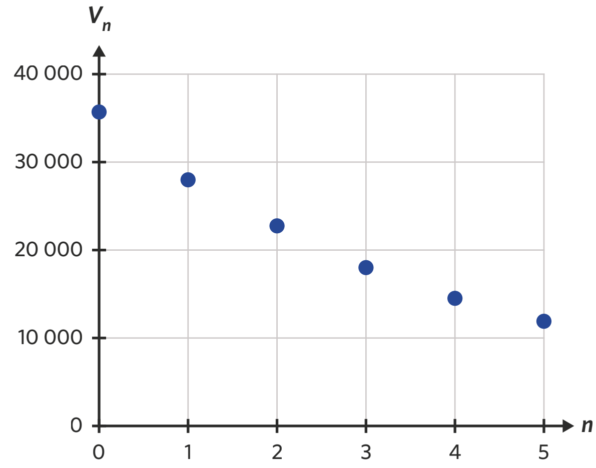
b. Ava's tablet was bought for \$500, and depreciates at a rate of 14% per year.

Back to contents

4. Grace purchased a brand new grand piano for $36 000. The piano loses value according to reducing balance depreciation, as shown in the graph.

The value of the grand piano after year 1 is $28 800.

Construct a recurrence relation to represent the depreciation of the piano.



Using a geometric recurrence relation to determine the value of an asset after *n* depreciation periods

5. The value of Caitlyn's new TV, V_n, at year n, is given by the following recurrence relation.

$V_0 = 2000, \quad V_{n+1} = 0.98V_n$

What is the value of the TV, to the nearest dollar, after two years?

A. $1882 **B.** $1920 **C.** $1921 **D.** $1960

---

6. Maddy has saved up to buy a second hand car from her cousin. The car is worth $11 800 and is depreciated by 9% per annum according to the reducing balance method.

The following recurrence relation models the depreciation of Maddy's car.

$M_0 = 11\ 800, \quad M_{n+1} = 0.91M_n$

a. What is the value of the car after Maddy has owned it for three years?

b. What is the total amount of depreciation for the three years after Maddy buys the car?

---

7. The following recurrence relation models the monthly depreciation of a brand new jet ski that was purchased for $28 750.

$J_0 = 28\ 750, \quad J_{n+1} = 0.985J_n$

a. How much is the jet ski worth after one year?

b. How long will it be until the jet ski is worth $26 657.47?

c. What is the amount of depreciation in the sixth month to the nearest dollar?

---

8. The following recurrence relation models the depreciation per year of a computer that was purchased for $3250.

$V_0 = 3250, \quad V_{n+1} = 0.9V_n$

a. Determine the values of V_1, V_2, V_3 and V_4.

b. Plot the depreciation of the computer over four years.

c. After how many years does the value of the computer fall below $2500?

---

9. Suppose a car is depreciated by the reducing balance method at a rate of 17% per annum. The value of the car, $V\,n$, after n years, is given by the following recurrence relation.

$V_0 = 54\ 000, \quad V_{n+1} = 0.83V_n$

a. Find the value of the car, to the nearest dollar, after three years.

b. Suppose that the car depreciates by 17% each year for the first three years of its life, but this rate then reduces to 11% for the remainder of the car's life. What is the value of the car five years after it is purchased? Round to the nearest dollar.

Back to contents

Joining it all together

10. Tracy and her brother Dean both bought a laptop. Tracy's laptop is depreciated according to the recurrence relation

$T_0 = 800, \quad T_{n+1} = 0.98T_n,$

where n represents the number of months since purchase.

a. What is the reducing balance rate of depreciation per period for Tracy's laptop?

b. How much value will Tracy's laptop lose over the first three months? Round to the nearest cent.

Dean's laptop was bought for \$900 and lost \$22.50 in value in the first month. His laptop depreciates according to the reducing balance method.

c. Find the monthly rate of depreciation.

d. Write a recurrence relation for the value of Dean's laptop, D_n, in terms of months, n.

---

11. Sam recently bought a new phone for \$1700. It will be depreciated at a reducing balance rate of 16% per annum.

a. Write a recurrence relation to model the value of Sam's phone, S_n, where n is the number of years after it was purchased.

b. What will the value of Sam's phone be after four years?

c. Plot the future value of Sam's phone on a graph over the first four years.

d. After how many years will the value of the phone fall below \$1200?

---

12. Courtney was given a designer handbag for her birthday which cost \$3250. The handbag loses its value each year as the style becomes more outdated. It is depreciated at a reducing balance rate of 10% per annum.

a. Write a recurrence relation that represents the value of Courtney's handbag, C_n, over n years.

b. How many years will it take for the handbag to lose half its value?

c. Suppose that once the handbag is worth less than half of the initial purchase price, the rate of depreciation reduces to 5% per annum. Calculate the value of the handbag, to the nearest dollar, two years after it loses half its value.

Exam practice

13. 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 following recurrence relation.

$V_0 = 60\,000, \quad V_{n+1} = 0.9V_n$

Use recursion to show that the value of the tools after two years, V_2, is \$48 600. (1 MARK)

VCAA 2019 Exam 2 Recursion and financial modelling Q7a

74% of students answered this question correctly.

---

14. Consider the following graph.

This graph could show the value of

A. a piano depreciating at a flat rate of 6% per annum.

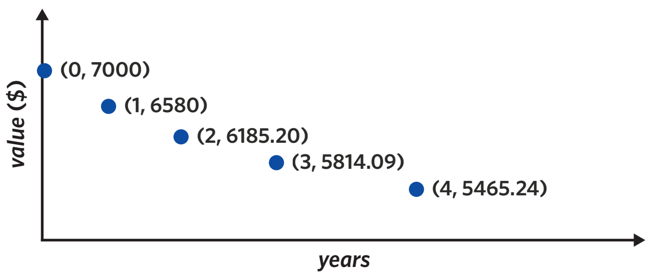
B. a car depreciating with a reducing balance rate of 6% per annum.

C. a compound interest investment earning interest at the rate of 6% per annum.

D. an item of equipment depreciating with a reducing balance rate of 10% per annum.

E. a boat depreciating at a flat rate of 10% per annum.

Adapted from VCAA 2017 Exam 1 Recursion and financial modelling Q22



65% of students answered this type of question correctly.

Back to contents

15. Julie withdraws \$14 000 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 following relation shown.

$C_0 = 14\,000, \quad C_{n+1} = R \times C_n$

a. 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. 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 to the nearest cent. (2 MARKS)

VCAA 2018 Exam 2 Recursion and financial modelling Q5a,b

Part **a**: **55%** of students answered this question correctly.

Part **b**: The average mark on this question was **0.8**.

Questions from multiple lessons

Data analysis

16. The following scatterplot displays the distribution of the *number of absences* in a semester and *average mark* (%) of 30 university students. A least squares regression line has been fitted to the data.

The correlation coefficient of the distribution is $r = -0.8259$.

Which of the following statements is false?

A. There is a strong negative association between *number of absences* and *average mark* for these students.

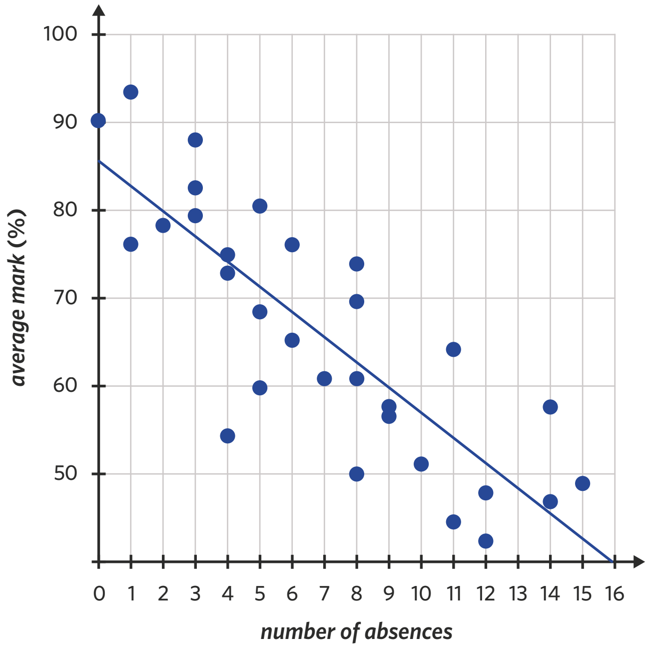
B. Approximately 68% of the variation in *average mark* can be explained by the variation in *number of absences*.

C. Using the regression line to predict the *average mark* of a student with 16 absences is an example of interpolation.

D. The least squares line has a negative slope.

E. Students with higher numbers of absences tend to have lower average marks.

Adapted from VCAA 2017NH Exam 1 Data analysis Q11



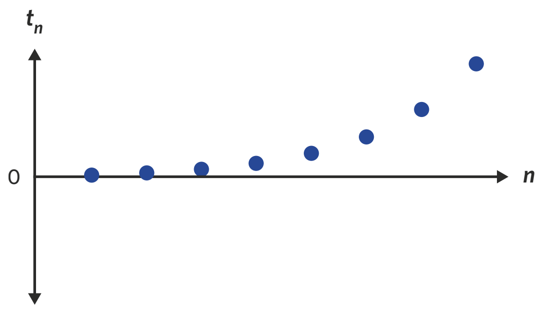
Recursion and financial modelling *Year 11 content*

17. The first term in a geometric sequence is x, where $x > 0$.

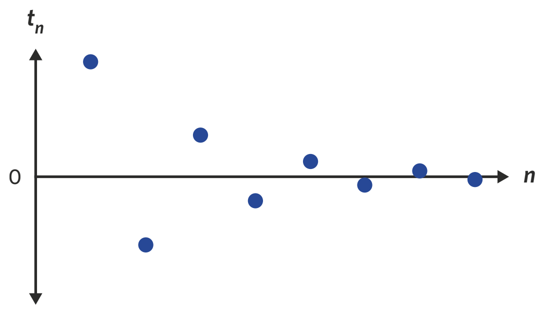
The common ratio of the sequence is R, where $-1 < R < 0$.

Which of the following graphs could show the first eight terms of the sequence?

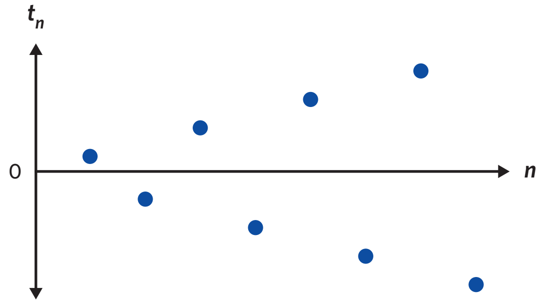
A.



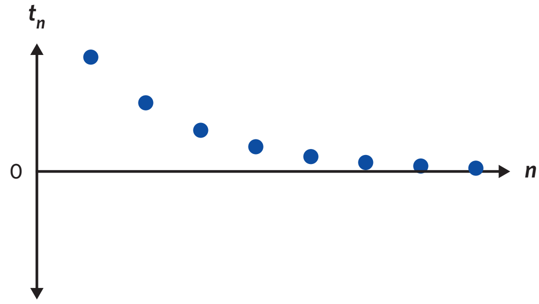
B.



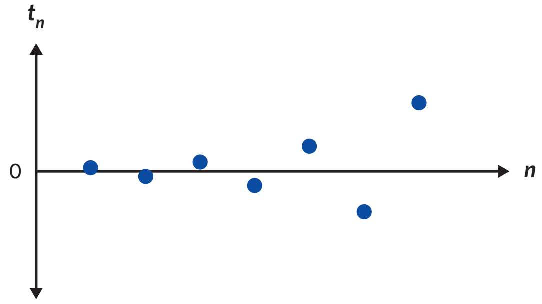
C.



D.



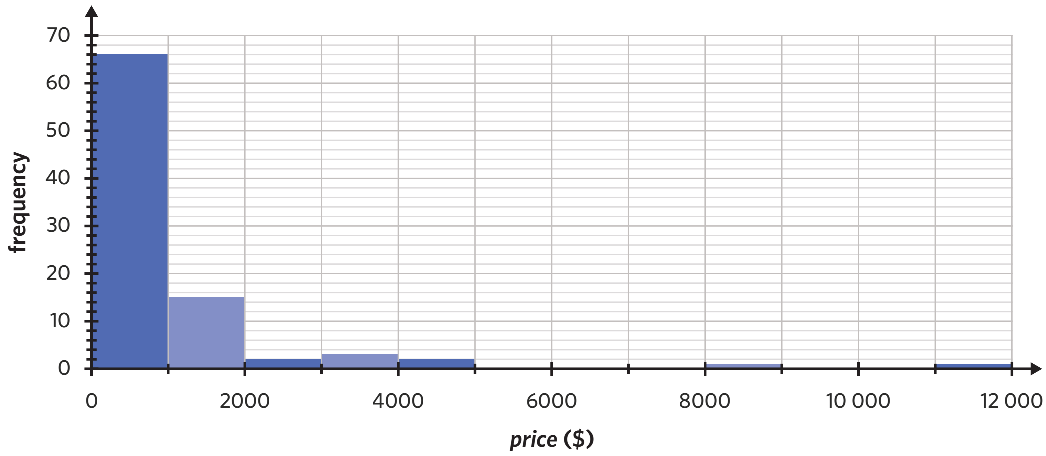
E.



Adapted from VCAA 2014 Exam 1 Number patterns Q8

Data analysis

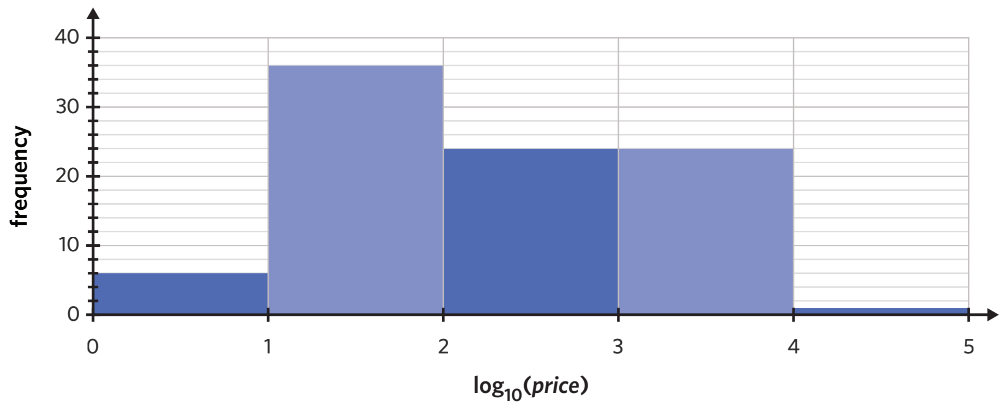
18. The following histogram shows the price of 90 handbags from various shops in Melbourne.



a. Describe the shape of the distribution, including any possible outliers. (1 MARK)

b. What percentage of the handbags cost under \$2000? (1 MARK)

c. The following histogram displays the distribution of $\log_{10}(price)$ for the same sample of 90 handbags.



How many of the handbags cost between \$100 and \$1000? (1 MARK)

Adapted from VCAA 2018NH Exam 2 Data analysis Q3

Back to contents

5D Depreciation – finding the rule for the *n*th term

STUDY DESIGN DOT POINT

- 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

5A 5B 5C **5D** 5E 5F 5G

KEY SKILLS

During this lesson, you will be:

- creating and using a rule for flat rate depreciation
- creating and using a rule for unit cost depreciation
- creating and using a rule for reducing balance depreciation.

Depreciation can be modelled accurately using recurrence relations. However, a recurrence relation can only be used to calculate the next term in the sequence. It is useful to create rules that model depreciation so that any term in the sequence can be determined from one calculation, instead of multiple iterations.

Creating and using a rule for flat rate depreciation

Recall that flat rate depreciation is used when the value of an asset decreases by a constant amount each period, calculated from the initial value.

The following rule can be used to model V_n, the value of an asset after n periods, depreciated by the flat rate method.

$V_n = V_0 - nd$, where

- V_n is the value after n periods
- V_0 is the principal
- d is the depreciation amount per period, calculated by $d = \frac{r}{100} \times V_0$
- r is the depreciation rate (%) per period

Once the rule has been created, the value of the asset after n periods can be found by substituting the value of n into the rule.

Worked example 1

Kwyjibo bought an apartment for \$670 000 in what he thought was a high-growth suburb. Unfortunately, the growth stopped once he bought the property and it is now depreciating yearly by 1.5% of its initial value.

a. Write the rule to determine V_n, the value of the apartment after n years.

Explanation

Step 1: Determine the initial value, V_0.

Kwyjibo bought the real estate for \$670 000.

$V_0 = 670\ 000$

Continues →

Back to contents

Step 2: Calculate the depreciation amount, d.

$d = V_0 \times \frac{r}{100}$

The rate of depreciation per period is 1.5%, so $r = 1.5$.

$$\begin{aligned} d &= 670\,000 \times \frac{1.5}{100} \\ &= 10\,050 \end{aligned}$$

Step 3: Substitute these values into the rule for V_n.

$V_n = V_0 - nd$

Answer

$V_n = 670\,000 - 10\,050n$

b. Find the value of Kwyjibo's apartment after 9 years.

Explanation

Calculate V_9, the value of the apartment after 9 years.

Substitute $n = 9$ into the rule for V_n.

$$\begin{aligned} V_9 &= 670\,000 - 10\,050 \times 9 \\ &= 670\,000 - 90\,450 \\ &= 579\,550 \end{aligned}$$

Answer

$579 550

Creating and using a rule for unit cost depreciation

Recall that unit cost depreciation is used when the value of an asset decreases by a constant amount for each unit of use.

The following rule can be used to model V_n, the value of an asset after n units of use, depreciated using the unit cost method.

$V_n = V_0 - nd$

- V_n is the value after n units of use
- V_0 is the principal
- d is the depreciation amount per unit of use

Once a rule has been created, the value of the asset after n units of use can be found by substituting the value of n into the rule.

Worked example 2

Samson purchased a computer worth $2000. The research he did online told him that the model of computer he purchased depreciates by $0.32 for every hour of use.

a. Write the rule to determine V_n, the value of the computer after n hours of use.

Explanation

Step 1: Determine the initial value, V_0.

Samson purchased the computer for $2000.

$V_0 = 2000$

Step 2: Calculate the depreciation amount, d.

The computer depreciates by $0.32 for every hour of use.

$d = 0.32$

Continues →

Back to contents

Step 3: Substitute these values into the rule for V_n.

$V_n = V_0 - nd$

Answer

$V_n = 2000 - 0.32n$

b. What is the value of Samson's computer after 300 hours of use?

Explanation

Calculate V_{300}, the value of the computer after 300 hours of use.

Substitute $n = 300$ into the rule for V_n.

$$\begin{aligned} V_{300} &= 2000 - 0.32 \times 300 \\ &= 2000 - 96 \\ &= 1904 \end{aligned}$$

Answer

\$1904

Creating and using a rule for reducing balance depreciation

Recall that reducing balance depreciation occurs when the value of an asset depreciates by a percentage, calculated from the value at the previous time period.

The following rule can be used to model V_n, the value of an asset after n periods, depreciated by reducing balance depreciation.

$V_n = V_0 \times R^n$, where

- V_n is the value after n periods
- V_0 is the principal
- $R = 1 - \frac{r}{100}$
- r is the depreciation rate (%) per period

Once the rule has been created, the value of the asset after n periods can be found by substituting the value of n into the rule.

Worked example 3

Gemima purchased a \$10 000 lawnmower that depreciates at a rate of 10% each year, using the reducing balance method.

a. Write the rule to determine V_n, the value of the lawnmower after n years.

Explanation

Step 1: Determine the initial value, V_0.

Gemima purchased the lawnmower for \$10 000.

$V_0 = 10\,000$

Step 2: Calculate the value of R.

$R = 1 - \frac{r}{100}$

The rate of depreciation is 10% per period, so $r = 10$.

$$\begin{aligned} R &= 1 - \frac{10}{100} \\ &= 0.9 \end{aligned}$$

Continues →

Back to contents

Step 3: Substitute these values into the rule for V_n.

$V_n = V_0 \times R^n$

Answer

$V_n = 10\,000 \times 0.9^n$

b. Find the value of the lawnmower after 5 years.

Explanation

Calculate V_5, the value of the lawnmower after 5 years.

Substitute $n = 5$ into the rule for V_n.

$V_5 = 10\,000 \times 0.9^5$
$= 5904.9$

Answer

\$5904.90

Exam question breakdown

VCAA 2019 Exam 1 Recursion and financial modelling Q22

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 \$18 480.

A rule for the value of the machine after n units are produced, V_n, is

A. $V_n = 0.872n$
B. $V_n = 24\,000n - 3840$
C. $V_n = 30\,000 - 24\,000n$
D. $V_n = 30\,000 - 0.872n$
E. $V_n = 30\,000 - 0.16n$

Explanation

Step 1: Identify the initial value, V_0.

The machine is purchased for \$30 000.

$V_0 = 30\,000$

Step 2: Calculate the depreciation per year.

After three years, the machine is worth \$18 480.

$\textit{depreciation over three years} = 30\,000 - 18\,480$
$= 11\,520$

$\textit{depreciation per year} = \dfrac{11\,520}{3}$
$= 3840$

Step 3: Calculate the depreciation per unit produced, d.

24 000 items are produced each year.

$d = \dfrac{3840}{24\,000}$
$= 0.16$

Step 4: Substitute these values into the rule for V_n.

$V_n = V_0 - nd$

$V_n = 30\,000 - 0.16n$

53% of students answered this question correctly.

17% of students incorrectly answered D. These students calculated the average yearly depreciation (11 520 ÷ 3 = 3840) as a percentage of the initial value (3840 ÷ 30 000 = 0.128 or 12.8%). They then applied the rule for reducing balance depreciation, rather than for unit cost depreciation.

Answer

E

Back to contents

5D Questions

Creating and using a rule for flat rate depreciation

1. A $15 000 asset depreciates by a flat rate of 9% each year. By what dollar amount does the asset depreciate each year?

A. $900 **B.** $1250 **C.** $1350 **D.** $1500

2. Rodney purchases a $400 scientific calculator that depreciates at a flat rate. He creates the rule $C_n = 400 - 45n$ to calculate C_n, the value of the calculator after n years. The value of the calculator after two years is

A. $310 **B.** $355 **C.** $445 **D.** $490

3. For each of the following scenarios:

- write the rule to determine V_n, the value of the asset after n periods.
- calculate the value of the asset after nine periods.

a. A $1 000 000 property that depreciates by 6% of the initial value each year.

b. A $550 designer t-shirt that depreciates by 3% of the sale price every month.

c. A $380 fridge that depreciates by 6.5% of the initial value each year.

d. A $1545 pair of shoes that depreciates by 4% of its sale price every month.

4. Steve buys a smartphone for $550. It depreciates by 10% of the sale price every year.

a. Write a rule to determine S_n, the value of the smartphone after n years.

b. What is the value of the smartphone after the third year?

c. After how many years will the smartphone have no value?

5. Riya buys a computer for $1225. Two years later she gets it valued, and finds it is worth $955.50.

a. What is the annual flat rate percentage depreciation for her computer?

b. Write the flat rate depreciation rule to determine C_n,the value of her computer after n years.

c. What will the value of the computer be after 7 years?

6. Hessa buys a car for $8600. She hopes that in three years it will still be worth at least $6000. What is the highest rate of flat rate depreciation per year, to the nearest percent, the car can have for this to happen?

Creating and using a rule for unit cost depreciation

7. Hassan bought a SodaStream for $100 that depreciates by $0.25 each time it is used. He wants to create a rule to determine S_n, the value of the SodaStream after n uses. Which of the following is correct?

A. $S_0 = 0.25, \quad d = 100$

B. $S_n = 100, \quad d = 0.25$

C. $S_0 = 100, \quad d = 0.25$

D. $S_0 = 100, \quad d = 25$

Back to contents

8. For each of the following scenarios:

- write the rule to determine V_n, the value of the asset after n units of use.
- calculate the value of the asset after five units of use.

a. A \$65 air fryer that depreciates by \$0.08 for each hour of use.

b. A \$315 printer depreciates by \$0.05 for each page printed.

c. A \$20 000 car that depreciates by \$245 for every 1000 km driven.

d. An \$825 dishwasher that depreciates by \$7.50 after each wash.

---

9. A parachute costs \$400 and depreciates by \$8 per jump.

a. Write a rule to determine P_n, the value of the parachute after n jumps.

b. What is the value of the parachute after 20 jumps?

c. After how many jumps will the parachute be worth \$200?

---

10. The following table shows the value of a jet at various points during its service life.

***distance travelled* (km)**	400 000	500 000	600 000
***value* (\$)**	4 358 200	4 017 800	3 677 400

a. What is the depreciation of the jet, per kilometre travelled, rounded to the nearest cent?

b. Using the rounded answer from part **a**, what was the original value of the jet?

---

11. Anakin owns a cafe and purchases an industrial coffee machine for \$9300. The coffee machine depreciates for each coffee made. On average, his cafe sells 215 coffees per day. After 120 days, the coffee machine is worth \$8655. Assuming the cafe continues selling coffees at the same average daily rate, how much will the coffee machine be worth after 350 days?

Creating and using a rule for reducing balance depreciation

12. Max uses the following rule to calculate V_n, the value of his photocopier after n years.

$V_n = 2500 \times 0.95^n$

By how much does the photocopier depreciate each year?

A. 0.05% **B.** 0.5% **C.** 0.95% **D.** 5%

---

13. For each of the following scenarios:

- write the rule to determine V_n, the value of the asset after n periods.
- calculate the value of the asset, to the nearest cent, after the specified number of periods.

a. A car, originally valued at \$13 000, depreciating at a rate of 6% per annum for three years.

b. A television with an initial value of \$3100, depreciating at a rate of 10% per annum for seven years.

c. A printer with an initial value of \$2000, depreciating at a rate of 3.5% every six months for four years.

d. A drum set purchased for \$6500, depreciating at a rate of 2% every fortnight for one year.

---

14. Marko wants to buy a new smartwatch, which is currently priced at \$779, but he only has \$700. After doing some market research, he finds that this exact model of watch depreciates at a rate of 1.75% every three months, using a reducing balance depreciation method. How long will Marko have to wait before he can buy the new smartwatch?

Back to contents

Joining it all together

15. Baxter purchases a brand new portable electric heater for $170.

After doing some research he finds out that depending on who he tries to sell the heater to, it can depreciate using one of three methods.

- The flat rate method at 2.8% each month.
- The unit cost method at $0.22 for every 10 hours of use.
- The reducing balance method of 3% each month.

a. Write a rule to determine H_n, the value of the heater after n periods or units of use for each of the depreciation methods.

b. Baxter works out that he will use the heater for an average of 240 hours each month. After two months of owning the heater, he decides to travel overseas and wants to sell the heater. Which depreciation method will allow Baxter to sell the heater for the largest amount, and how much would he be able to sell it for?

c. Due to travel restrictions, Baxter has to stay home and ends up using the heater for a total of six months before he tries to sell it. Which depreciation method will allow Baxter to sell the heater for the highest amount, and how much would he be able to sell it for?

---

16. Mildred buys a high pressure washer for $1200. She is told that the washer depreciates by $2.50 for every 10 litres of water used, and estimates that she will use 15 litres of water each month.

a. Assuming Mildred uses an average of 15 litres of water each month, what is the annual rate of depreciation that will ensure its depreciated value using the flat rate method will be equivalent to the unit cost method after one year?

b. Using the reducing balance method, what is the **annual** rate of depreciation that will ensure the value of the washer will be equivalent to the unit cost method after two years? Round to two decimal places.

c. Mildred wants to sell the washer after 10 years. Calculate the depreciated value of the washer using the flat rate and reducing balance methods, using the answers calculated in parts **a** and **b**. Which method will allow Mildred to sell the washer for the highest amount?

Exam practice

17. 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 30 000 km in this fourth year is $26 166.24.

The value of k is

A. 9 **B.** 12 **C.** 14
D. 16 **E.** 18

VCAA 2020 Exam 1 Recursion and financial modelling Q29

54% of students answered this question correctly.

---

18. Sammy purchased a boat for $72 000.

The value of the boat is depreciated each year by 10% using the reducing balance method.

In the third year, the boat will depreciate in value by 10% of

A. $47 239.20 **B.** $52 488.00 **C.** $58 320.00
D. $64 800.00 **E.** $72 000.00

VCAA 2021 Exam 1 Recursion and financial modelling Q20

44% of students answered this question correctly.

19. Sienna owns a coffee shop.

A coffee machine, purchased for \$12 000, is depreciated in value using the unit cost method.

The rate of depreciation is \$0.05 per cup of coffee made.

The recurrence relation that models the year-to-year value, in dollars, of the coffee machine is

$M_0 = 12\,000, \quad M_{n+1} = M_n - 1440$

a. The recurrence relation could also represent the value of the coffee machine depreciating at a flat rate.

What annual flat rate percentage of depreciation is represented? (1 MARK)

b. Complete the following rule that gives the value of the coffee machine, M_n, in dollars, after n cups have been produced.

$M_n = \square + \square \times n$ (1 MARK)

VCAA 2021 Exam 2 Recursion and financial modelling Q7b,c

Part **a**: **49%** of students answered this question correctly.

Part **b**: **21%** of students answered this question correctly.

Questions from multiple lessons

Data analysis

20. The seasonal indices for the number of tickets sold at a tourist attraction each month, excluding August, are shown in the following table.

month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
seasonal index	1.48	1.34	1.07	0.89	0.66	0.68	0.63		0.75	0.91	1.24	1.68

If 4592 tickets were sold last August then the deseasonalised number of tickets is closest to

A. 3076 **B.** 3077 **C.** 3168 **D.** 6655 **E.** 6854

Adapted from VCAA 2017NH Exam 1 Data analysis Q15

Recursion and financial modelling *Year 11 content*

21. Find the first five terms of the following recurrence relation.

$L_0 = -4, \quad L_{n+1} = 8 - 6 \times L_n$

A. $-4, -32, -184, -1112, -6664...$

B. $-4, -24, -144, 864, -5184...$

C. $-4, -16, -88, -520, -3112...$

D. $-4, 32, -184, 1112, -6664...$

E. $32, -184, 1112, -6664, 39\,992...$

Adapted from VCAA 2016 Exam 1 Recursion and financial modelling Q17

Recursion and financial modelling *Year 11 content*

22. Argus realises that he is allergic to his dog, so he decides to sell it to his friend Houston. In order to pay for the dog, Houston agrees to make eight payments over eight months. He will pay \$32 in the first month, January, and in each subsequent month he will pay 50% more than the previous month. This follows the geometric sequence $V_{n+1} = 1.5V_n$.

a. How much money does Houston pay Argus in the sixth month? (1 MARK)

b. In which month will Houston first pay over \$150? (1 MARK)

c. How much does Houston end up paying Argus in total? (1 MARK)

Adapted from VCAA 2013 Exam 2 Number patterns Q2

Back to contents

5E Simple interest

STUDY DESIGN DOT POINT

- the concepts of simple and compound interest

5A 5B 5C 5D 5E 5F 5G

KEY SKILLS

During this lesson, you will be:

- modelling simple interest using recurrence relations
- modelling simple interest using a rule.

KEY TERMS

- Interest
- Simple interest

Simple interest is the most straightforward type of interest. It is often found in the real world with short-term loans, such as car loans or small personal loans. Simple interest is often the most preferred type of interest by those paying it, as it is usually the cheapest option when borrowing money.

Modelling simple interest using recurrence relations

Interest is the cost of borrowing money. For loans, it is the fee that banks and lenders charge the borrowers for borrowing their money. On the other hand, for investments, the investor will receive interest from the investment location, such as a bank, since they are effectively borrowing money from them.

Simple interest is a constant amount of interest that is charged or earned each period. It is calculated as a percentage of the principal.

An investment or loan with simple interest can be modelled by a recurrence relation which can then be used to calculate the value of the interest or loan after any amount of time periods.

The value of these loans and investments, V_n, after n periods can be modelled by an arithmetic recurrence relation of the form

$V_0 = principal, \quad V_{n+1} = V_n + d$, where

- d is the interest paid each period, calculated by $d = \frac{r}{100} \times V_0$
- r is the interest rate (%) per period

Note: This recurrence relation is only valid if no loan repayments are made or if no additional payments are added to an investment.

Worked example 1

Sally borrows \$300 000 where simple interest is charged at a rate of 0.5% per year.

a. Construct a recurrence relation to model the value of this loan, V_n after n years.

Explanation

Step 1: Determine the principal and interest rate.

$V_0 = 300\ 000$

$r = 0.5\%$ per annum

Step 2: Calculate the interest, d.

$$d = \frac{0.5}{100} \times 300\ 000$$
$$= 1500$$

Step 3: Construct the recurrence relation.

Answer

$V_0 = 300\ 000, \quad V_{n+1} = V_n + 1500$

b. What is the value of the loan after three years?

Explanation

Step 1: Calculate V_1, the value of the loan after one year.

$$V_1 = V_0 + 1500$$
$$= 300\ 000 + 1500$$
$$= 301\ 500$$

Step 2: Calculate V_2, the value of the loan after two years.

$$V_2 = V_1 + 1500$$
$$= 301\ 500 + 1500$$
$$= 303\ 000$$

Step 3: Calculate V_3, the value of the loan after three years.

$$V_3 = V_2 + 1500$$
$$= 303\ 000 + 1500$$
$$= 304\ 500$$

Answer

\$304 500.00

c. How much interest has been charged after three years?

Explanation

The interest charged is the difference between the current value and the principal.

$$\textit{interest} = V_3 - V_0$$
$$= 304\ 500 - 300\ 000$$
$$= 4500$$

Answer

\$4500.00

Back to contents

Modelling simple interest using a rule

In VCE General Mathematics, a rule can be used to calculate the value of an investment or loan accumulating simple interest. This rule is of the form

$V_n = V_0 + nd$, where

- V_0 is the principal
- d is the interest paid each period, calculated by $d = \frac{r}{100} \times V_0$
- r is the interest rate (%) per period

This is used to calculate V_n for any value of n.

Worked example 2

Joan invests \$5000 in an account paying a simple interest rate of 4.5% per year.

a. Construct a rule that will calculate the value of the investment, V_n, after n years.

Explanation

Step 1: Determine the principal and interest rate.

$V_0 = 5000$

$r = 4.5$

Step 2: Calculate the interest, d.

$$d = \frac{4.5}{100} \times 5000$$
$$= 225$$

Step 3: Construct the rule.

Answer

$V_n = 5000 + 225n$

b. Use the rule to calculate the value of the investment after five years.

Explanation

Calculate V_5, the value of the investment after 5 years.

$$V_5 = 5000 + 225 \times 5$$
$$= 6125$$

Answer

\$6125.00

5E Questions

Modelling simple interest using recurrence relations

1. A \$13 000 investment earns simple interest at a rate of 1.5% per period. Which of the following recurrence relations can be used to model the investment?

A. $V_0 = 195, \quad V_{n+1} = V_n + 13\,000$

B. $V_0 = 13\,000, \quad V_{n+1} = V_n + 19\,500$

C. $V_0 = 13\,000, \quad V_{n+1} = V_n + 1.5 \times V_0$

D. $V_0 = 13\,000, \quad V_{n+1} = V_n + 195$

Back to contents

2. Write a recurrence relation to model each of the following scenarios.
   a. Tina invests $1500 in a savings account that pays interest at a flat rate of $30 every month.
   b. Ahmed puts $500 in a bank account. The account pays simple interest once a year, at a rate of 8% of the opening balance.

3. The following recurrence relations describe the balance of three savings accounts, V_n, X_n, Y_n after n months. What will the balance of each account be after five months?
   a. $V_0 = 700, \quad V_{n+1} = V_n + 15$
   b. $X_0 = 1200, \quad X_{n+1} = X_n + 60$
   c. $Y_0 = 820, \quad Y_{n+1} = Y_n + 43$

4. The balance of Mala's loan, M_n, at month n, follows the given recurrence relation. The loan charges simple interest at a monthly rate.
   $M_0 = x, \quad M_{n+1} = M_n + 43$
   a. If the balance owing when $n = 4$ is $1772, what is x?
   b. What is the simple rate of interest per month for this account? Round the answer to two decimal places.

Modelling simple interest using a rule

5. For the following rule, what is V_{19}?
   $V_n = 223 + 19n$
   A. 223 B. 242 C. 565 D. 584

6. If $9354 was invested in an account that earns simple interest at a rate of 0.7% per month, what is the total amount of simple interest paid at the end of a 22-month period, rounded to a whole number?
   A. 120 B. 1200 C. 1440 D. 1441

7. A $200 loan charges simple interest at a rate of 9% per annum. What is the total amount outstanding on the loan after 8 years, assuming no repayments have been made?

8. The balance of Alonso's simple interest savings account n months after the initial deposit is given by the following rule.
   $A_n = 1200 + 50n$
   a. What was Alonso's initial deposit?
   b. How much interest does the account earn every month?
   c. Correct to two decimal places, what is the monthly interest rate of the account?

9. Manish takes out a loan of $3000 from EasyShark Loans. The loan accumulates simple interest at a rate of 3.75% per month.
   a. How much interest is charged every month?
   b. What will the balance of the loan be after the first six months, assuming no repayments have been made?

10. Alex wants to open a simple interest bank account and has two options. An account with the Bank of Edrolo earns interest at a rate of 1% paid out at the end of each month, whereas an account with The Bank of Victoria earns interest at a rate of 8% paid each year. She has $765 to invest in the account.

a. Which bank pays more interest every year?

b. After 15 years, how much more money will she earn from the bank with the higher interest amount per year?

---

11. Janine put $400 into her savings account when she opened it. It pays simple interest at a rate of 3% per month.

a. Construct a rule to calculate the balance of the account, J_n, at month n.

b. How much money will be in the account after six months?

c. After how many months will the balance exceed $500?

---

12. In order to save up to buy Christmas gifts for her family, Olivia requires $30 of interest to be paid to her bank account each month, for the remaining 5 months of the year. Olivia opens a savings account with The Bank of Edrolo, with a simple interest rate of 6% per month.

a. What principal amount must she place into the account to achieve her goal?

b. What will the balance of her bank account be at the end of the year?

Joining it all together

13. Tao borrowed $3650 from his local loan agency, and won't make any repayments for the first 6 years. For each of these 6 years, the value of this loan that must be repaid increases by $83.

a. Given this information, find the simple interest rate charged per annum, rounded to two decimal places.

b. Write down a rule to model this scenario, T_n at year n.

c. If Tao wishes to repay the full loan 6 years later, how much does he owe the loan agency?

---

14. Charlee sets herself a savings goal of $700, so she opens a brand new savings account and moves all the money from her piggy bank into it. Each year she earns simple interest of $22 with an interest rate of 4% per annum.

a. How much money did she initially move into the savings account from her piggy bank?

b. Write down a recurrence relation to model this scenario.

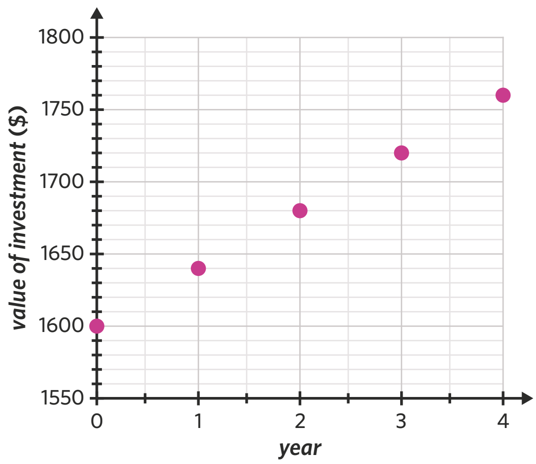
c. After how many years has Charlee reached her savings goal?

---

15. Sally invested $1600 in savings account four years ago and plotted the value of the investment every year, shown in the following graph. Three years ago, Harry invested in the same account and received $270 total in simple interest.

a. How much did Harry invest?

b. Construct a rule that can be used to calculate H_n, the balance of Harry's account after n years.



Back to contents

Exam practice

16. \$3000 is invested at a simple interest rate of 6.5% per annum.

The total interest earned in three years is

A. \$195.00 **B.** \$580.50 **C.** \$585.00
D. \$3623.85 **E.** \$3585.00

VCAA 2012 Exam 1 Business-related mathematics Q2

90% of students answered this question correctly.

17. A sum of money is invested in an account paying simple interest at a rate of 8% per annum.

The total interest earned on this investment over 6 years is \$27 000.

The sum of money invested is

A. \$12 960 **B.** \$45 000 **C.** \$56 250
D. \$202 500 **E.** \$337 500

VCAA 2007 Exam 1 Business-related mathematics Q3

68% of students answered this question correctly.

Questions from multiple lessons

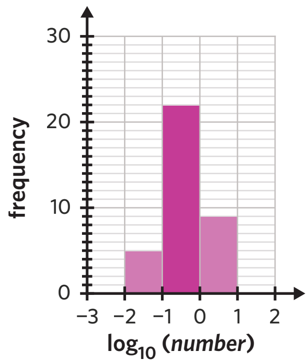
Data analysis

18. The following histogram shows the distribution of the average *number* of pets owned per household for 36 different suburbs plotted on a $\log_{10}$ scale.

The number of suburbs where the average *number* of pets is one or more per household is

A. 0
B. 5
C. 9
D. 22
E. 31

Adapted from VCAA 2016 Exam 1 Data analysis Q7



Recursion and financial modelling *Year 11 content*

19. Consider the following recurrence relation.

$T_0 = 5, \quad T_{n+1} = -2T_n + 6$

The first four terms of the recurrence relation are

A. 5, −10, 4, −8 **B.** 5, −4, 14, −22 **C.** 5, −4, 2, 10 **D.** 6, 4, 2, 0 **E.** 6, −6, 18, −30

Adapted from VCAA 2016 Exam 1 Recursion and financial modelling Q17

Data analysis *Year 11 content*

20. The humidity level in the morning and in the evening on 15 different days in Melbourne was recorded. The data is displayed on the following scatterplot.

The least squares regression line for the given data is:

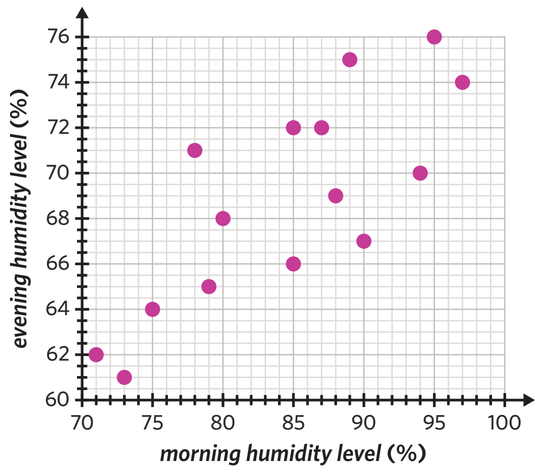
evening humidity level = 30.7 + 0.45 × *morning humidity level*

a. Use the least squares line to predict the humidity level in the evening when the morning humidity level is 76%. (1 MARK)

b. What is the residual value when the least squares line is used to predict the evening humidity level when the morning humidity level is 80%? Round to the nearest one decimal place. (2 MARKS)

c. The correlation coefficient *r* is 0.79. What percentage of the variation in the evening humidity level can be explained by the variation in the morning humidity level? Round to the nearest percent. (1 MARK)

Adapted from VCAA 2018 Exam 2 Data analysis Q2



Back to contents

5F Compound interest

STUDY DESIGN DOT POINTS

- 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 future value of a compound interest investment or loan after n compounding periods and its use to solve practical problems

5A 5B 5C 5D 5E **5F** 5G

KEY SKILLS

During this lesson, you will be:

- modelling compound interest using recurrence relations
- modelling compound interest using a rule.

KEY TERMS

- Compound interest

While simple interest increases a value at a fixed rate, compound interest is more adaptive and responsive to changes over time. For this reason, compound interest can be very practical, and is used for a wide range of real life situations, such as investments and loans. Compound interest is the most common form of interest and most people deal with it at certain points in their lives, making it an important concept to understand.

Modelling compound interest using recurrence relations

As opposed to simple interest, **compound interest** takes into account both the principal and any accumulated interest earned in previous compounding periods. This means that as the value of the investment or loan grows, its rate of growth increases, whilst keeping the same interest rate (%).

The value of these loans and investments, V_n, after n compounding periods, can be modelled by a geometric recurrence relation of the form

$V_0 = principal, \quad V_{n+1} = R \times V_n$, where

- $R = 1 + \frac{r}{100}$
- r is the interest rate (%) per compounding period

When interest compounds more than once a year, the annual interest rate must be converted to the interest rate per compounding period. This can be done by dividing the annual interest rate by the number of compounding periods per year. Some common compounding periods are shown in the following table.

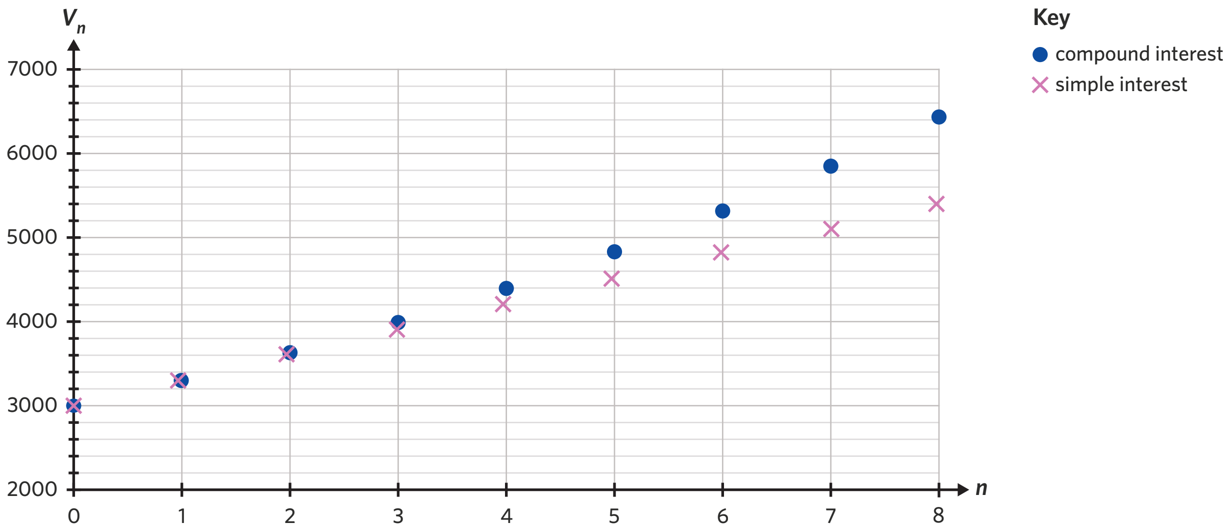
type	compounding periods per year
annual	1
biannual	2
quarterly	4
monthly	12
fortnightly	26
weekly	52

For example, if the interest rate is 6% per annum compounding monthly, the interest rate per month is given by $r = \frac{6}{12}\% = 0.5\%$.

Additionally, if the interest rate is converted to the interest rate per compounding period, the number of compounding periods, n, should be updated accordingly. For example, 5 years is equivalent to $5 \times 12 = 60$ months.

Compound interest can also be modelled graphically, by plotting each balance, V_n, for each compounding period, n. As compound interest will lead to increased growth as the balance increases, the graph will be non-linear.

The graph shown compares simple and compound interest for two investments with the same starting value of \$3000 and an interest rate of 10% per annum. The compound interest investment compounds annually, whereas the simple interest investment adds the same amount every year.



Worked example 1

Duncan has an investment of \$10 000 that is earning interest at a rate of 8% per annum, compounding fortnightly.

a. Construct a recurrence relation that can be used to calculate value of the investment, V_n, after n fortnights. Round to three decimal places where necessary.

Explanation

Step 1: Identify the principal and interest rate.

$V_0 = 10\ 000$

As interest is compounding fortnightly, the annual interest rate must be divided by 26 to give r, the interest rate per fortnight.

$r = \frac{8}{26}\%$

$= 0.307...\%$ per compounding period

Step 2: Calculate R.

$R = 1 + \frac{r}{100}$

$R = 1 + \frac{0.307...}{100}$

$= 1.00307...$

≈ 1.003

Step 3: Construct the recurrence relation.

Answer

$V_0 = 10\ 000, \quad V_{n+1} = 1.003 \times V_n$

---

b. Calculate the value of Duncan's investment after 4 fortnights.

Explanation

Step 1: Calculate V_1, the value of the investment after 1 fortnight.

$V_1 = 1.003 \times V_0$

$= 1.003 \times 10\ 000$

$= 10\ 030$

Step 2: Calculate V_2, the value of the investment after 2 fortnights.

$V_2 = 1.003 \times V_1$

$= 1.003 \times 10\ 030$

$= 10\ 060.09$

Continues →

Step 3: Calculate V_3, the value of the investment after 3 fortnights.

$$V_3 = 1.003 \times V_2$$
$$= 1.003 \times 10\,060.09$$
$$= 10\,090.270...$$

Step 4: Calculate V_4, the value of the investment after 4 fortnights.

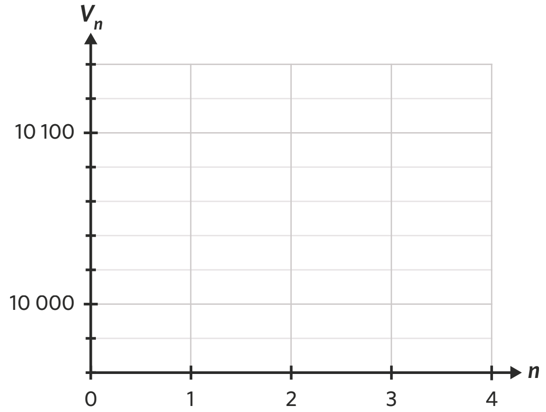
$$V_4 = 1.003 \times V_3$$
$$= 1.003 \times 10\,090.270...$$
$$= 10\,120.541...$$

Answer

$10 120.54

---

c. Plot the value of the investment for the first four fortnights on the set of axes shown.



Explanation

Step 1: Determine the values over the first year.

From part **b**,

$V_0 = 10\,000$

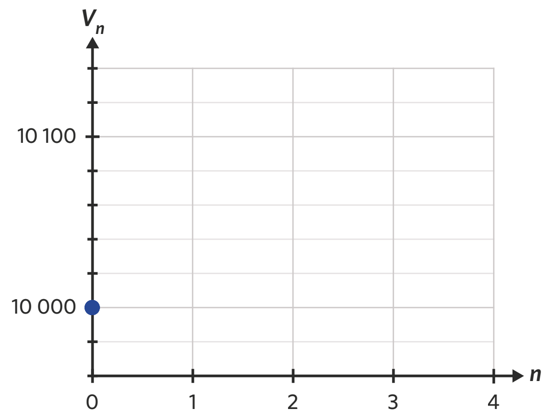
$V_1 = 10\,030$

$V_2 = 10\,060.09$

$V_3 \approx 10\,090.27$

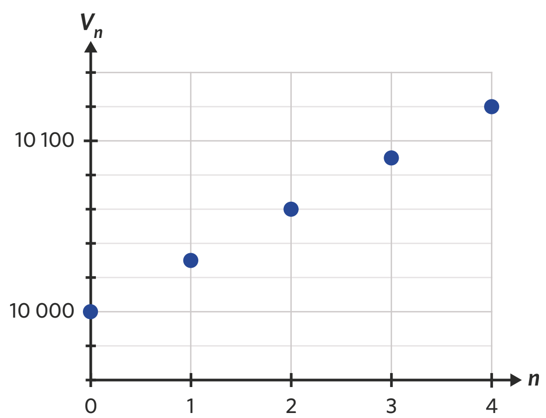
$V_4 \approx 10\,120.54$

Step 2: Plot V_0 on the graph.



Step 3: Plot the remaining values on the graph.

Answer



Back to contents

Modelling compound interest using a rule

A rule can be used to calculate the value of an investment or loan earning compound interest. This rule is of the form

$V_n = V_0 \times R^n$, where

- V_n is the value after n periods
- V_0 is the principal
- $R = 1 + \frac{r}{100}$
- r is the interest rate (%) per compounding period

This is useful for calculating V_n for larger values of n.

Note: This rule format is very similar to the compound interest formula, $A = P \times \left(1 + \frac{r}{100}\right)^n$.
One useful aspect of recurrence relations is that the term 'V_n' can be used to keep track of the amount of time that is associated with an amount of money. For example, V_3 represents the amount of money after 3 periods, V_4 represents the amount of money after 4 periods, and so on.

Worked example 2

$20 000 is invested into an account earning compound interest at a rate of 9% per annum, compounding monthly.

a. Construct a rule that can be used to calculate the balance of the account, V_n after n months.

Explanation

Step 1: Identify the principal and interest rate.

$V_0 = 20\,000$

As interest is compounding monthly, the annual interest rate must be divided by 12 to give r, the interest rate per month.

$r = \frac{9}{12}\%$

$= 0.75\%$ per compounding period

Step 2: Calculate R.

$R = 1 + \frac{r}{100}$

$R = 1 + \frac{0.75}{100}$

$= 1.0075$

Step 3: Construct the rule.

Answer

$V_n = 20\,000 \times 1.0075^n$

b. How much interest would the account have earned after 4 years?

Explanation

Step 1: Determine the value of n by converting 4 years into months.

$n = 4 \times 12$

$= 48$

Step 2: Calculate V_{48}, the value of the account after 48 months.

$V_{48} = 20\,000 \times 1.0075^{48}$

$= 28\,628.106...$

Step 3: Calculate the interest earned after 48 months.

This is the difference between V_{48} and V_0.

$interest = 28\,628.106... - 20\,000.00$

$= 8628.106...$

Answer

$8628.11

Back to contents

Exam question breakdown

VCAA 2020 Exam 2 Recursion and financial modelling Q9a

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 shown.

$B_0 = 5000, \quad B_{n+1} = 1.003 \times B_n$

Write down the value of B_4, the balance of the savings account after four months.

Round the answer to the nearest cent. (1 MARK)

Explanation

Step 1: Calculate B_1, the balance of the account after 1 month.

$$B_1 = 1.003 \times B_0$$
$$= 1.003 \times 5000$$
$$= 5015$$

Step 2: Calculate B_2, the balance of the account after 2 months.

$$B_2 = 1.003 \times B_1$$
$$= 1.003 \times 5015$$
$$= 5030.045$$

Step 3: Calculate B_3, the balance of the account after 3 months.

$$B_3 = 1.003 \times B_2$$
$$= 1.003 \times 5030.045$$
$$= 5045.1351...$$

Step 4: Calculate B_4, the balance of the account after 4 months.

$$B_4 = 1.003 \times B_3$$
$$= 1.003 \times 5045.1351...$$
$$= 5060.2705...$$

Answer

$5060.27

64% of students answered this question correctly.

Some students rounded to the nearest ten cents to give an answer of $5060.30. The question required them to round to the nearest cent instead.

5F Questions

Modelling compound interest using recurrence relations

1. Sarah invests $2400 in a bank account which pays 4% interest per annum, compounding annually.
Which of the following recurrence relations could describe the growth of Sarah's account?

A. $V_0 = 0, \quad V_{n+1} = 0.04 \times V_n$

B. $V_0 = 0, \quad V_{n+1} = 1.04 \times 2400$

C. $V_0 = 2400, \quad V_{n+1} = 1.04 \times V_n$

D. $V_0 = 2400, \quad V_{n+1} = V_n + 0.04$

---

2. The amount of money in Said's bank account, V_n, after n months, is given by the following recurrence relation.

$V_0 = 3000, \quad V_{n+1} = 1.03 \times V_n$

a. What was the initial balance in Said's account?

b. What is the monthly interest rate for this account?

3. The amount of money in Nicole's bank account, V_n, after n months, is given by the following recurrence relation.

$V_0 = 1500, \quad V_{n+1} = 1.05 \times V_n$

a. What will the balance of Nicole's bank account be after four months, correct to the nearest cent?

b. When will the account reach $2000?

---

4. Write recurrence relations, L_n and S_n, to model the following situations.

a. Leyla deposits $2800 when opening a new savings account. The account pays 2% interest per month, compounding monthly. Use n to represent months.

b. Shaun opens a new savings account with a deposit of $800. The account has 16% interest per annum, compounded quarterly. Use n to represent quarters.

---

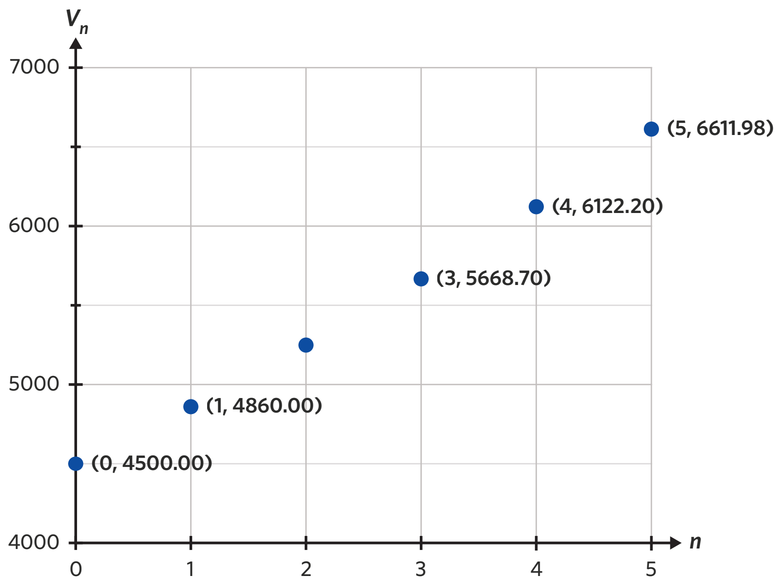
5. Harold is saving up to build a swimming pool in his backyard. He currently has $29 000 but the pool will cost $30 000 to build. Harold decides to invest the $29 000 in a savings account that has an interest rate of 9.6% per annum, compounding monthly.

a. Write a recurrence relation to model the value of Harold's investment month to month.

b. How many months does Harold have to wait until he can afford to build the pool?

---

6. The graph shown displays the value, V_n, of an investment after n compounding periods. One value, V_2, is missing from the graph.



Using the graph, determine the value of V_2.

---

7. Elise is trying to decide whether the Super Saver account or the Save Big account is best for her. She has $1200 to invest. The recurrence relation for the Super Saver account from month to month is given.

$S_0 = 1200, \quad S_{n+1} = 1.015 \times S_n$

a. How much interest will she earn with the Super Saver account over six months? Round to the nearest cent.

b. The Save Big account has 24% interest per annum, compounding monthly, but has a $100 account start-up fee. Deduct the fee from Elise's initial deposit and write a recurrence relation for the balance of the Save Big account, B_n, after n months.

Back to contents

Modelling compound interest using a rule

8. The rule $V_n = 6000 \times 1.09^n$ could be used to model which of the following scenarios?

A. An investment of $6000 with an interest rate of 0.09% p.a, compounding monthly.

B. An investment of $6000 with an interest rate of 0.09% p.a, compounding quarterly.

C. An investment of $6000 with an interest rate of 9% p.a, compounding monthly.

D. An investment of $6000 with an interest rate of 9% p.a, compounding annually.

9. Find the final value of each of the following investments, compounding yearly, correct to the nearest cent.

a. $120 invested for three years with compound interest of 5% per annum.

b. $9 000 000 invested in a compound interest account for five years at 8.2% per annum.

c. $230 000 invested in a compound interest account for twenty years at 2.31% per annum.

10. Andy, Bec and Charlie each invested $3000 into compound interest savings accounts three years ago. Find the current balance of each of their savings accounts, correct to the nearest cent.

a. Andy: compound interest rate of 4.5% per annum, compounding yearly.

b. Bec: compound interest rate of 6.2% per annum, compounding quarterly.

c. Charlie: compound interest rate of 3.9% per annum, compounding weekly.

11. Find the value of the following investment accounts correct to the nearest dollar.

a. An initial deposit of $40 000 compounding monthly at an annual interest rate of 6.4% for ten months.

b. An account with a starting balance of $1200 compounding quarterly at an annual interest rate of 8.1% for one and a half years.

c. An investment of $650 was made for two years at 3.3% annual interest, compounding half-yearly.

12. Eliza and Frankie win $1000 each and invest their winnings into savings accounts with an interest rate of 8% per annum. Eliza's account compounds monthly whereas Frankie's account compounds quarterly. After five years, who will have more money in their account and how much more will they have, correct to the nearest cent?

Joining it all together

13. Which recurrence relation, corresponding to the rule $V_n = 2500 \times 1.082^n$, is written correctly?

A. $V_0 = 2500, \quad V_{n+1} = 1.082 \times V_0$

B. $V_0 = 2500, \quad V_{n+1} = 1.082 \times V_n$

C. $V_n = 2500, \quad V_{n+1} = 1.082 \times V_n$

D. $V_n = 2500, \quad V_{n+1} = 1.082 \times V_0$

14. Which of the following rules corresponds to the recurrence relation

$S_n = 10\,000, \quad S_{n+1} = 1.002 \times S_0$?

A. $S_n = 10\,000 \times 0.2$

B. $S_n = 10\,000 \times 1.02$

C. $S_n = 10\,000 \times 1.002$

D. $S_n = 10\,000 \times 1.002^n$

Back to contents

15. For each of the following scenarios, model the scenario by

- generating a recurrence relation
- generating a rule.

a. \$1000 invested in an account that earns interest at a rate of 1.8%, compounding annually (let V_n be the account's value after n years).

b. \$3500 invested in an account that earns interest at a rate of 2.4%, compounding monthly (let V_n be the account's value after n months).

c. \$5000 invested in an account that earns interest at a rate of 6%, compounding quarterly (let V_n be the account's value after n quarters).

---

16. Rachel and Naomi are arguing about which bank offers a better savings account, regarding the amount of interest it pays in 3 years, from an initial deposit of \$5000.

- Rachel believes that account A is more profitable, as it provides 6.8% interest per annum, compounding annually.
- Naomi believes that account B is better, because it provides 6.6% interest per annum, but compounds monthly.

Which account is more profitable? Justify your answer using relevant calculations.

Exam practice

17. 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

$B_0 = 3000, \quad B_{n+1} = 1.0048 \times B_n$

The total interest earned by Manu's investment after the first five months is closest to

A. \$57.60 **B.** \$58.02 **C.** \$72.00
D. \$72.69 **E.** \$87.44

VCAA 2020 Exam 1 Recursion and financial modelling Q24

80% of students answered this question correctly.

---

18. Alex 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)

VCAA 2017 Exam 2 Recursion and financial modelling Q6

Part **a**: **61%** of students answered this question correctly.
Part **b**: The average mark on this question was **1**.
Part **c**: **26%** of students answered this question correctly.

Back to contents

Questions from multiple lessons

Data analysis

19. A least squares regression line

A. maximises the number of data points that lie on the line.

B. minimises the sum of the vertical distances from the data points to the line.

C. minimises the sum of the squares of the vertical distances from the data points to the line.

D. maximises the sum of the vertical distances from the data points to the line.

E. maximises the sum of the squares of the vertical distances from the data points to the line.

Adapted from VCAA 2017NH Exam 1 Data analysis Q7

Recursion and financial modelling *Year 11 content*

20. A family recently planted an apple tree. The number of apples grown each year follows a geometric sequence.

During the first year, the tree grew 3 apples.

During the second year, the tree grew 6 apples.

During the third year, the tree grew 12 apples.

How many apples are expected to grow in the sixth year?

A. 48 **B.** 96 **C.** 192 **D.** 384 **E.** 768

Adapted from VCAA 2015 Exam 1 Number patterns Q5

Recursion and financial modelling *Year 11 content*

21. A family living in the country decide to start a hobby farm.

Each year, a number of new animals are added to the farm.

In the third year, 11 new animals are added to the farm.

The number of new animals, A_n, added to the farm after n years, is modelled by the recurrence relation

$A_0 = x, \quad A_{n+1} = A_n + 3$

a. What is the value of x? (1 MARK)

b. What type of sequence does the recurrence relation generate? (1 MARK)

The number of new animals, A_n, added to the farm after n years can also be modelled by the rule $A_n = b + c \times n$.

c. Determine the values of b and c. (2 MARKS)

Adapted from VCAA 2013 Exam 2 Number patterns Q3

Back to contents

5G Nominal and effective interest rates

STUDY DESIGN DOT POINT

- 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

5A 5B 5C 5D 5E 5F 5G

KEY SKILLS

During this lesson, you will be:

- calculating effective interest rates
- comparing nominal and effective interest rates and returns.

KEY TERMS

- Nominal interest rate
- Effective interest rate

The nominal and effective interest rates of a loan or investment both describe the interest on the same loan or investment. However, interest can be earned or charged at different compounding periods such as yearly or monthly. Comparing the effective interest rate of loans or investments with different compounding periods can help determine which options are the most profitable.

Calculating effective interest rates

The **nominal interest rate** is the quoted interest rate for a loan or investment, and is also known as the annual interest rate.

The **effective interest rate** is the adjusted or 'real' rate, respective to the number of compounding periods in the year. It can be used to compare the total amount of interest earned or charged on different investment or loan options with different compounding periods.

The effective interest rate, $r_{effective}$, can be calculated by hand using the formula

$$r_{effective} = \left(\left(1 + \frac{r}{100n}\right)^n - 1\right) \times 100, \text{ where}$$

- r is the nominal interest rate, per annum
- n is the number of compounding periods per year.

However, to save time and ensure accuracy, the effective interest should be calculated on either the TI-Nspire or the Casio ClassPad calculator.

Worked example 1

Convert an interest rate of 12.7% per annum, compounding monthly, to an effective interest rate, correct to two decimal places.

Explanation – Method 1: By hand

Step 1: Determine the nominal interest rate and number of compounding periods per year.

$r = 12.7\%$ per annum

$n = 12$ compounding periods in a year, since interest compounds monthly.

Step 2: Calculate the effective interest rate.

$$r_{eff} = \left(\left(1 + \frac{12.7}{100 \times 12}\right)^{12} - 1\right) \times 100$$

$$= (1.01058...^{12} - 1) \times 100$$

$$= 13.465...\%$$

Continues →

Back to contents

Explanation – Method 2: TI-Nspire

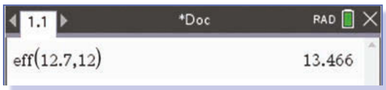
Step 1: From the home screen, select ‘1: New’ → ‘1: Add Calculator’.

Step 2: Press menu. Select ‘8: Finance’ → ‘5: Interest Conversion’ → ‘2: Effective Interest Rate’.

The function is in the format of eff(nominal interest rate, number of compounding periods per year).

Step 3: Enter the nominal interest rate, followed by the number of compounding periods per year.

The nominal interest rate is 12.7% p.a. and there are 12 compounding periods per year, so this should be ‘12.7, 12’. Press enter.



Explanation – Method 3: Casio ClassPad

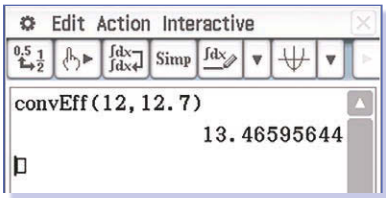
Step 1: From the main menu, tap √α Main.

Step 2: Tap ‘Action’ → ‘Financial’ → ‘Interest Conversion’ → ‘convEff’.

The function is in the format of convEff(number of compounding periods, nominal interest rate).

Step 3: Enter the number of compounding periods per year, followed by the nominal interest rate.

There are 12 compounding periods per year and the nominal interest rate is 12.7% p.a., so this should be ‘12, 12.7’. Press EXE.



Answer – Method 1, 2 and 3

13.47%

Comparing nominal and effective interest rates and returns

A nominal interest rate does not give the full picture on how much interest is being calculated. For example, it is difficult to see if an interest rate of 4.5% per annum compounding monthly will be more profitable than a rate of 4.3% per annum compounding weekly.

The effective interest rate is used to compare different loan or investment options, and determine the option that is more financially beneficial.

If interest is calculated more frequently, given the same nominal interest rate, then the effective interest rate of the investment or loan will be greater. This means that the account will earn or be charged more interest over time.

For a loan, a borrower would typically prefer a lower effective interest rate so that they can pay less interest. On the other hand, for an investment, an investor would generally prefer a higher effective interest rate so that they can receive more interest.

Worked example 2

Pete is saving up money for schoolies and is considering investing his savings into either of two accounts with the following interest rates:

- Superior Saver: 4.5% per annum, compounding quarterly
- Supreme Saver: 4.4% per annum, compounding weekly

Which account should Pete choose?

Continues →

Explanation

Step 1: Calculate the effective interest rate for Superior Saver.

$r = 4.5\%$ p.a. and $n = 4$ compounding periods per year.

$r_{effective} = 4.5765...\%$

Step 2: Calculate the effective interest rate for Supreme Saver.

$r = 4.4\%$ p.a. and $n = 52$ compounding periods per year.

$r_{effective} = 4.4962...\%$

Step 3: Determine the best option.

The question involves an investment so the option with a higher effective interest rate will be the preferred choice.

Superior Saver has a greater effective interest rate.

Answer

Superior Saver

5G Questions

Calculating effective interest rates

1. For a loan with a nominal interest rate of 4.30% per annum, compounding fortnightly, the effective interest rate is closest to

A. 4.19% **B.** 4.30% **C.** 4.39% **D.** 4.51%

2. Convert the following nominal interest rates to effective interest rates, correct to two decimal places.

a. 10% per annum, compounding monthly.

b. 8% per annum, compounding weekly.

c. 9.2% per annum, compounding daily.

d. 7.32% per annum, compounding yearly.

e. 2.01% per annum, compounding fortnightly.

3. Brigitte decides to put $550 in a savings account that has an annual interest rate of 11.75%, compounding monthly. What is the effective interest rate, correct to one decimal place?

Comparing nominal and effective interest rates and returns

4. Which of the following nominal interest rates returns the lowest effective interest rate?

A. 5.5% per annum, compounding weekly

B. 5.5% per annum, compounding fortnightly

C. 5.5% per annum, compounding monthly

D. 5.5% per annum, compounding annually

5. In each of the following parts, state which of the two interest rates is lower.

a. An effective interest rate of 8.5% or a nominal rate of 7.9% p.a., compounding monthly.

b. An effective interest rate of 11.3% or a nominal rate of 10.9% p.a., compounding quarterly.

c. An effective interest rate of 9.5% or a nominal rate of 8.7% p.a., compounding weekly.

Back to contents

6. Emily is looking to invest her savings into an account, in the hope that the interest she earns will support her art business. Her bank offers her two options:
   - The SuperAccount with an interest rate of 5.91% per annum, compounding weekly
   - The SmartAccount with an interest rate of 6.00% per annum, compounding yearly

   Which account should Emily choose?

7. Liam wants to buy a Hilux in time for summer. He needs to take out a loan of $30 000 to cover the costs. Liam shops around for the best loan option and narrows it to two.
   - ANB is offering a loan with an interest rate of 7.10% per annum, compounding monthly
   - Westbank is offering a loan with an interest rate 7.05% per annum, compounding weekly

   Which bank should he choose?

Joining it all together

8. If an account has an effective interest rate of 14.6% and the interest compounds monthly, then the nominal interest rate is closest to

   A. 12.9

   B. 13.5

   C. 13.7

   D. 13.9

9. Rex wins $5000 in a snooker competition and wants to put it into a savings account. VicBank is offering an interest rate of 4.12% per annum, compounding weekly, and AussieBank is offering an interest rate of 4.21% per annum, compounding quarterly. Which account would be the most profitable for Rex?

10. Jess currently has $30 000 invested in an account with an interest rate of 5.6% p.a. compounding quarterly.

    For a limited time, there is a special offer from another bank for an account with an interest rate of 5.54%, compounding daily.

    Should Jess switch banks? Explain why.

Exam practice

11. 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.

    VCAA 2020 Exam 1 Recursion and financial modelling Q28

58% of students answered this question correctly.

Back to contents

12. 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 shown:

- 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.

VCAA 2018 Exam 1 Recursion and financial modelling Q19

54% of students answered this question correctly.

Questions from multiple lessons

Data analysis

13. The following table displays the long-term average *number of rainy days* each month in Melbourne.

month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
number of rainy days	8.7	6.6	9.3	10.5	12.2	13.5	14.4	15.3	14.0	13.3	11.3	10.0

The data in the table is used to calculate the seasonal indices for each month. The seasonal index for August is closest to

A. 0.68 **B.** 0.76 **C.** 1.10 **D.** 1.24 **E.** 1.32

Adapted from VCAA 2018NH Exam 1 Data analysis Q16

Recursion and financial modelling

14. A company has had an increase in profit of 4% per year for the past two years. This year, the company's profit was $123 400. The company's profit 2 years ago is closest to

A. $114 090 **B.** $116 728 **C.** $118 653 **D.** $128 336 **E.** $133 469

Adapted from VCAA 2013 Exam 1 Business-related mathematics Q6

Recursion and financial modelling

15. Aidan bought an industrial printer for a purchase price of $4000. After six years, the printer was worth $2200.

a. What was the average depreciation in the value of the printer per year? (1 MARK)

b. Let V_n be the value of the printer n years after it was purchased.

One way of calculating the depreciation of the printer's value is the flat-rate method of depreciation.

Determine the recurrence relation, in terms of V_0, V_{n+1} and V_n, that models the value of the printer under flat-rate depreciation. (1 MARK)

c. The printer has printed 2000 pages each year since it was purchased.

The value of the printer can also be depreciated using the unit cost method of depreciation.

Using this method, by how much does the value of the printer reduce for each page printed? (1 MARK)

Adapted from VCAA 2016 Exam 2 Recursion and financial modelling Q6

Back to contents

6

CHAPTER 6

Advanced financial mathematics

LESSONS

KEY KNOWLEDGE

- 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
- 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
- use of a first-order linear recurrence relation to model and analyse (numerically and graphically) annuity investments, 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.

Back to contents

6A Introducing financial applications

STUDY DESIGN DOT POINT

- prerequisite lesson

6A 6B 6C 6D 6E 6F

KEY SKILLS

During this lesson, you will be:

- using an amortisation table to model financial problems
- using financial applications of technology.

KEY TERMS

- Amortisation table
- Financial solver
- PV (present value)
- PMT (payment amount)
- FV (future value)

Calculations for financial scenarios can get complex when there is a combination of payments and interest being calculated on the balance of a loan or investment. Amortisation tables and technology can be used to break down these complex scenarios and assist greatly in completing these calculations.

Using an amortisation table to model financial problems

An **amortisation table** is a table that can be used to determine the reducing balance of a loan or investment. These tables can model a range of financial problems by showing step-by-step calculations for the resultant balances in a series of consecutive compounding periods.
A standard table will have five columns:

- 'Payment number' – the number of payments that have been made to, or received from, the loan or investment. This number will increase by one with each iteration or change in balance, even if the payment value is zero.
- 'Payment' – the amount that has been paid to or from the loan or investment.
- 'Interest' – the amount that has been added to the balance of the loan or investment due to interest. This is calculated using the interest rate, per compounding period, and previous balance.
- 'Principal reduction' – the overall change in the balance of the loan or investment after payment and interest have been taken into account for the payment period.
- 'Balance of loan/investment' – the balance of the loan or investment after the payment period, which considers the previous balance and the principal increase/reduction for the payment period.

A new row is added to the amortisation table each time the loan or investment balance is altered. A '0' payment line is also included to show the establishment of a loan or investment.

For example, the following amortisation table shows the first four lines of a $50 000 account with a fixed amount of interest earned each month. Payments of $1000 are taken out of the account each month.

Calculations have been shown to demonstrate how the principal reduction is calculated.

Back to contents

payment number	payment	interest	principal reduction	balance
0	0.00	0.00	0.00	50 000.00
1	1000.00	50.00	$= 1000.00 - 50.00$ $= 950.00$	$= 50\,000.00 - 950.00$ $= 49\,050.00$
2	1000.00	50.00	$= 1000.00 - 50.00$ $= 950.00$	$= 49\,050.00 - 950.00$ $= 48\,100.00$
3	1000.00	50.00	$= 1000.00 - 50.00$ $= 950.00$	$= 48\,100.00 - 950.00$ $= 47\,150.00$

Worked example 1

Ella has borrowed \$10 000 from her parents to pay for a trip to Europe.

They have decided not to charge interest on the loan and Ella will make equal weekly repayments once she gets back from her trip.

The first five lines of the amortisation table for this loan is shown.

The loan balance for the third week is missing.

payment number	payment	interest	principal reduction	balance of loan
0	0.00	0.00	0.00	10 000.00
1	150.00	0.00	150.00	9850.00
2	150.00	0.00	150.00	9700.00
3	150.00	0.00	150.00	
4	150.00	0.00	150.00	9400.00

a. How much is Ella paying each week?

Explanation

Read the 'payment' column.

The repayments each week are equal in this scenario.

payment number	payment	interest	principal reduction	balance of loan
1	150.00	0.00	150.00	9850.00

Answer

\$150

b. Calculate the loan balance after the third repayment.

Explanation

Step 1: Determine the relevant cells in the amortisation table.

Since the loan is decreasing in value, the new balance is the difference between the previous loan balance and the principal reduction.

payment number	payment	interest	principal reduction	balance of loan
2	150.00	0.00	150.00	9700.00
3	150.00	0.00	150.00	

Step 2: Calculate the loan balance after the third payment.

balance of loan

$= \textit{previous loan balance} - \textit{principal reduction}$

$= 9700.00 - 150.00$

$= 9550.00$

Answer

\$9550

Back to contents

Using financial applications of technology

The **financial solver** is a useful calculator program that can be used to solve complex financial problems.

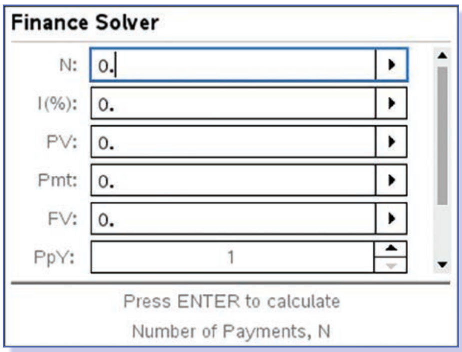
The notation of the financial solver is as follows:

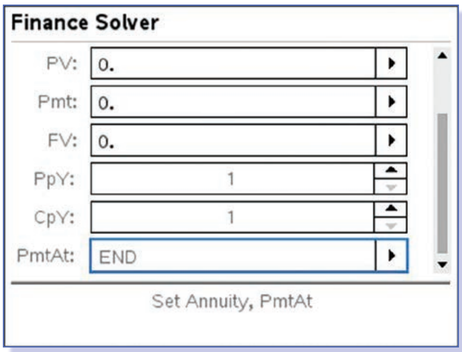
- 'N' – the number of payments made.
- 'I%' – the annual interest rate.
- '**PV**' – the present value of the loan or investment. It is also known as the current value.
- '**PMT**' – the payment amount made per compounding period.
- '**FV**' – the future value of the loan or investment, after *n* compounding periods.
- 'PpY' or 'P/Y' – how often payments are made or received per year.
- 'CpY' or 'C/Y' – how often interest is compounded per year. In General Mathematics, the CpY is usually the same as the PpY.
- 'PmtAT' (TI-Nspire only) – when the payment is made (BEGIN or END). In General Mathematics, all payments occur at the end of each period.

The financial solver can be used to determine unknown figures. For example, the financial solver could be used to calculate how long (N) it would take for an investment to reach a certain value, provided that all of the other variables are known.

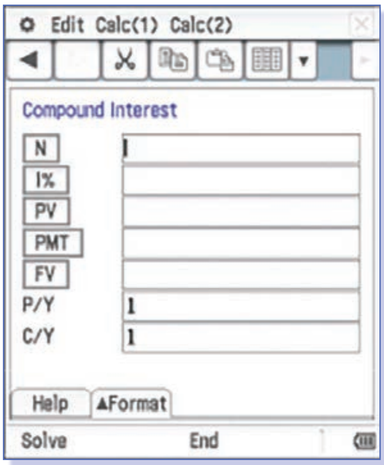
On the TI-Nspire, the Finance Solver can be used to model financial applications. It is accessed by pressing menu and selecting '8: Finance' → '1: Finance Solver'.

When first opening the Finance Solver, most values will be set at zero by default.





On the Casio ClassPad, the financial tool can be used to model financial applications. It is accessed by tapping Financial from the main menu and tapping 'Compound Interest'.



The direction of money exchange must be specified when using a financial calculator. The PV, PMT and FV are either positive or negative to specify whether money is being received or paid.

- The sign is positive if an individual receives money.
- The sign is negative if an individual pays money.

Back to contents

Worked example 2

Isabelle takes out an interest-free loan of $30 000. She makes repayments of $1000 per week to pay off the loan. Isabelle wants to know how much she owes the bank 10 weeks after establishing the loan.

a. Determine the inputs for the financial solver.

Explanation

N: There are 10 weeks and repayments are made weekly, so there are 10 repayments made during this timeframe.

I(%): There is no interest for this loan.

PV: Isabelle borrows $30 000. The PV is positive because she receives the money that she borrowed.

PMT: Isabelle pays back $1000 per week. The PMT is negative because she gives the money to the lender.

FV: This is the value that the financial solver will calculate, so it needs to be left blank.

PpY: The repayments are made weekly, which is equivalent to 52 times in a year.

CpY: Even though there is no interest, this can also be set to 52 to be consistent with the PpY.

Answer

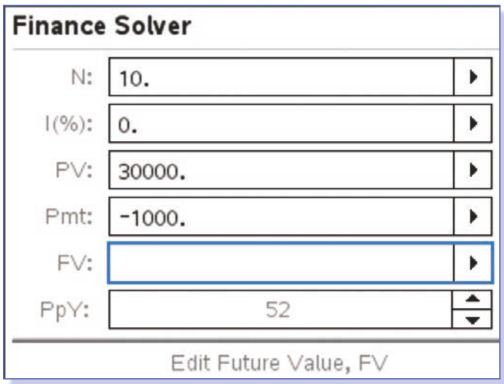
N	10
I(%)	0
PV	30 000
PMT	−1000
FV	
PpY	52
CpY	52

b. Use the financial solver to determine the balance of the loan after 10 weeks. Round to the nearest cent.

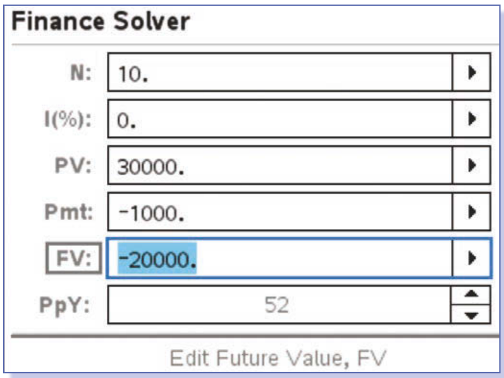
Explanation – Method 1: TI-Nspire

Step 1: Open the Finance Solver application and input the values from the table in part **a**.

Make sure the FV is blank.



Step 2: Move the cursor to the FV input box and press enter.



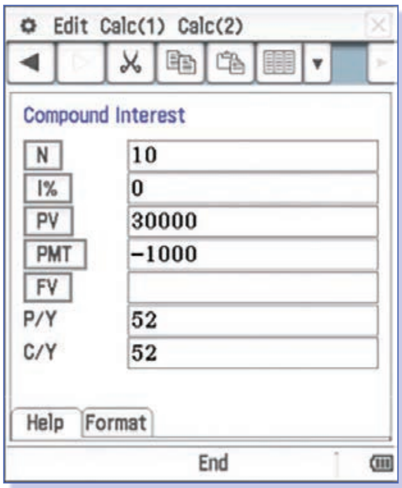
The FV is negative because this represents how much she still needs to pay.

Continues →

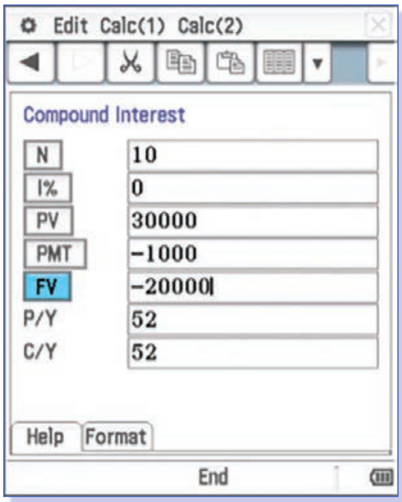
Back to contents

Explanation – Method 2: Casio ClassPad

Step 1: From the main menu, tap Financial and input the values from the table in part **a**. Make sure the FV is blank.



Step 2: Tap 'FV'.



The FV is negative because this represents how much she still needs to pay.

Answer – Method 1 and 2

$20 000

6A Questions

Using an amortisation table to model financial problems

1. Stephanie takes out a loan of $50 000 with her bank who doesn't charge interest for the first three months. She makes monthly repayments of $1200 towards the loan.

She has set up an amortisation table to model the amortisation of her loan. There are a few cells missing.

payment number	payment	interest	principal reduction	balance of loan
0	1200.00	0.00	0.00	50 000.00
1	1200.00	0.00	1200.00	48 800.00
2	1200.00	0.00	1200.00	47 600.00
3	1200.00			

a. Which of the following expressions can be used to calculate the principal reduction in the third month?

A. $principal\ reduction = 1200 - 0$

B. $principal\ reduction = 50\,000 - 1200$

C. $principal\ reduction = 50\,000 + 48\,800$

D. $principal\ reduction = 48\,800 - 1200$

b. What is the balance of Stephanie's loan after 3 months?

A. $40 000

B. $46 400

C. $48 800

D. $50 000

Back to contents

2. Hannes has put $500 into a safe. Each week, he will take out $50 to pay for petrol.

The amortisation table shows the balance of the money in the safe.

payment number	payment	interest	principal reduction	balance
0	0.00	0.00	0.00	500.00
1	50.00	0.00	50.00	450.00
2	50.00	0.00	50.00	400.00
3	50.00	0.00	50.00	350.00

Explain why the balance is decreasing in value each week. Include a calculation in your explanation.

3. Florence has borrowed $500 from his grandmother, Florina.

Florina will charge interest at a flat rate of $15 per week while Florence will make weekly payments of $50.

Florence's weekly payments will go towards paying off the interest before reducing the principal.

payment number	payment	interest	principal reduction	balance of loan
0	0.00	0.00	0.00	500.00
1	50.00	15.00		
2	50.00	15.00		
3	50.00	15.00		
4	50.00	15.00		

a. Show that the principal reduction in the first week is $35.

b. Show that the loan balance is $465 after the first week.

c. Fill out the remaining missing cells in the table.

Using financial applications of technology

4. Alo has established a loan of $15 000 to pay for her car.

She will be charged interest at a rate of 15% per annum, compounding twice a month, with repayments of $100 every half-month.

If she wants to know how much she still owes after one year, the value of N will be

A. 1 **B.** 12 **C.** 24 **D.** 26

5. In 2003, Zo opened up a savings account and put $215 into it.

Her bank offered an interest rate of 1.5% per annum, compounding monthly. However, she has since forgotten about the account.

In 2023, she discovers a document that contains details about this account and finally remembers it. She wants to know how much it is worth and uses a financial solver to help her.

Identify the value of the following:

a. I(%) **b.** PV **c.** PMT

6. Hubert wants to start investing his earnings. He has put $4000 in a savings account.

The interest rate for Hubert's investment is 2% per annum, compounding daily.

Identify the value of the following:

a. I(%) **b.** PV **c.** CpY

Back to contents

Joining it all together

7. The following amortisation table models the balance of a loan worth \$500.

The interest rate is 5% per annum, compounding monthly.

The repayments are also made monthly and go towards paying off the interest before reducing the principal.

payment number	payment	interest	principal reduction	balance of loan
0	0.00	0.00	0.00	500.00
1	15.00	2.08	12.92	487.08
2	15.00	2.03	12.97	474.11
3	15.00	1.98	13.02	461.09
4	15.00	1.92		

a. Fill in the missing cells, giving values to the nearest cent.

b. If this was modelled using the financial solver,

i. explain why the PV value would be positive.

ii. explain why the PMT value would be negative.

---

8. The following amortisation table models the balance of an investment worth \$500.

The interest rate is 10% per annum, compounding quarterly.

Payments of \$50 are also received quarterly.

payment number	payment	interest	principal reduction	balance of investment
0	0.00	0.00	0.00	500.00
1	50.00	12.50	37.50	462.50
2	50.00	11.56	38.44	424.06
3	50.00	10.60	39.40	384.66
4	50.00	9.62		

a. Calculate the balance of the investment after payment number 4, to the nearest cent.

b. If this was modelled using the financial solver,

i. would the PV value be positive or negative? Explain why.

ii. would the PMT value be positive or negative? Explain why.

Questions from multiple lessons

Recursion and financial modelling *Year 11 content*

9. The first four terms of a sequence are:

3, 9, 33, 129

Which of the following recurrence relations could generate this sequence?

A. $V_0 = 3, \quad V_{n+1} = 3V_n$

B. $V_0 = 3, \quad V_{n+1} = 2V_n + 3$

C. $V_0 = 3, \quad V_{n+1} = 4V_n - 3$

D. $V_0 = 3, \quad V_{n+1} = V_n + 6$

E. $V_0 = 3, \quad V_{n+1} = 5V_n - 6$

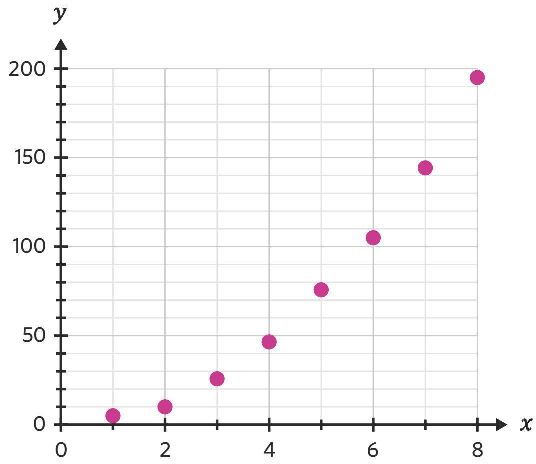
Adapted from VCAA 2017 Exam 1 Recursion and financial modelling Q18

Back to contents

Data analysis

10. The data in the table provided generates the following scatterplot.

x	y
1	3
2	11
3	28
4	48
5	77
6	105
7	143
8	195



A squared transformation is applied to the variable x to linearise the data. What is the equation of the least squares regression line for the transformed data? Round all values to two decimal places.

A. $y = -0.21 + 3.00x$

B. $y = -45.04 + 26.95x$

C. $y = -0.21 + 3.00\,x^2$

D. $y = -45.04 + 26.95x$

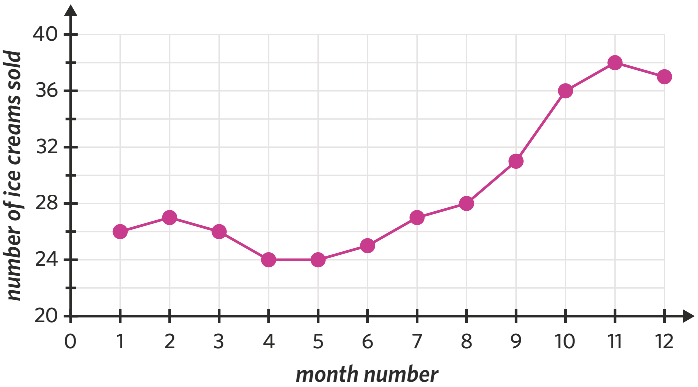
E. $y = 0.10 + 0.33\,x^2$

Adapted from VCAA 2018 Exam 1 Data analysis Q11

Data analysis

11. The average *number of ice creams sold* per day is recorded over 12 months.

The data is plotted for each month from January (month 1) to December (month 12) in the following time series plot.



a. State the overall trend in the time series plot. (1 MARK)

b. The trend in the data can be modelled by a least squares regression line.
The data used to obtain this line is shown.

month number	1	2	3	4	5	6	7	8	9	10	11	12
number of ice creams sold	26	27	26	24	24	25	27	28	31	36	38	37

Calculate the equation of the regression line, using *month number* as the explanatory variable.
Give values correct to four significant figures. (3 MARKS)

Adapted from VCAA 2017NH Exam 2 Data analysis Q3a, b

Back to contents

6B Reducing balance loans

STUDY DESIGN DOT POINTS

- 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

6A 6B 6C 6D 6E 6F

KEY SKILLS

During this lesson, you will be:

- using recurrence relations to model reducing balance loans
- using amortisation tables to solve problems involving reducing balance loans
- using financial applications of technology to solve problems involving reducing balance loans.

KEY TERMS

- Reducing balance loan

Reducing balance loans can be modelled using recurrence relations and amortisation tables. These show the amortisation of a loan from one compounding period to another. In addition to this, a financial solver can be used to solve problems over longer periods of time. There are many examples of reducing balance loans in the real world, such as mortgages and bank loans, where interest is charged and regular repayments are made to reduce the amount that is owed.

Using recurrence relations to model reducing balance loans

A **reducing balance loan** is a compound interest loan with repayments made at regular intervals. The repayments need to be greater than the amount of interest charged at each compounding period. This reduces the loan balance until it reaches zero, and the borrower no longer owes the lender.

The value of these loans, V_n, after n compounding periods, can be modelled by a recurrence relation of the form

$V_0 = principal, \quad V_{n+1} = R \times V_n - d$, where

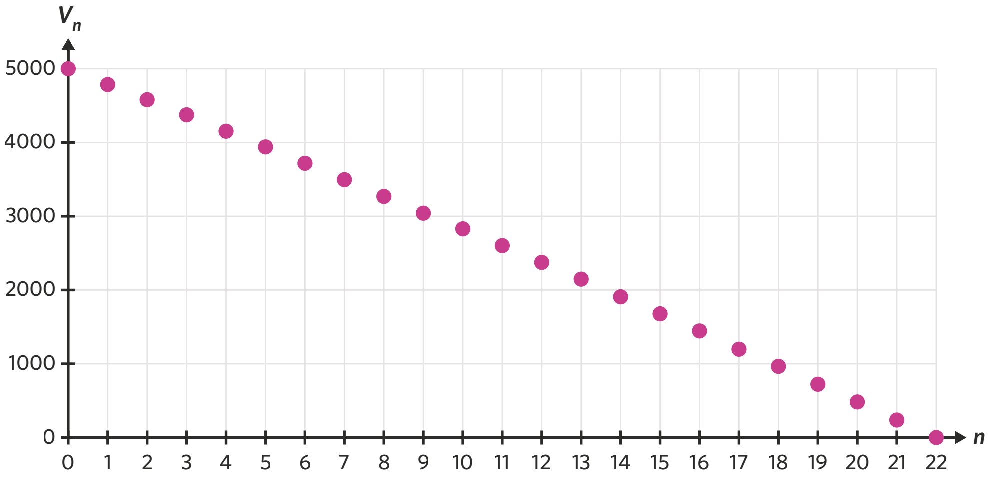
- $R = 1 + \frac{r}{100}$, where r is the interest rate (%) per compounding period
- d is the payment made per compounding period

For example, Rupert takes out a \$5000 loan with an interest rate of 10% per annum, compounding monthly and makes repayments of \$250 every month. A recurrence relation that Rupert can use to determine V_n, the value of the loan after n months is

$V_0 = 5000, \quad V_{n+1} = 1.00833 \times V_n - 250.$

Back to contents

The balance of the loan can also be represented on a graph.



There is a slight curve in the graph. This is because, at the beginning of the loan, a smaller portion of the repayment amount goes towards repaying the principal as the amount of interest owed is larger.

Worked example 1

Dawn has borrowed \$5000 from Barry, who charges interest at a rate of 5% per annum, compounding weekly. Dawn will make weekly \$500 repayments.

a. Construct a recurrence relation that can be used to model V_n, the value of the loan after n weeks. Round R to 5 decimal places.

Explanation

Step 1: Determine the principal, interest rate and payment amount.

$V_0 = 5000$

$r = \frac{5}{52}\%$ per compounding period

$d = 500$

Step 2: Calculate R.

$$R = 1 + \frac{5}{52 \times 100}$$
$$= 1 + \frac{5}{5200}$$
$$= 1.000961...$$

Step 3: Construct the recurrence relation.

Answer

$V_0 = 5000, \quad V_{n+1} = 1.00096 \times V_n - 500$

b. Calculate the amount that Dawn owes Barry after 3 weeks.

Explanation

Step 1: Calculate V_1, the value of the loan after 1 week.

$$V_1 = 1.00096 \times V_0 - 500$$
$$= 1.00096 \times (5000) - 500$$
$$= 4504.80$$

Step 2: Calculate V_2, the value of the loan after 2 weeks.

$$V_2 = 1.00096 \times V_1 - 500$$
$$= 1.00096 \times (4504.80) - 500$$
$$= 4009.124...$$

Step 3: Calculate V_3, the value of the loan after 3 weeks.

$$V_3 = 1.00096 \times V_2 - 500$$
$$= 1.00096 \times (4009.124...) - 500$$
$$= 3512.973...$$

Answer

\$3512.97

Back to contents

Using amortisation tables to solve problems involving reducing balance loans

Amortisation tables are used to keep track of the repayments made and the remaining balance in reducing balance loans.

In reducing balance loans, a portion of each payment repays the full amount of interest owed and the remainder is used to pay off the principal. In turn, this reduces the interest owed for the next compounding period and increases the portion of the payment that reduces the balance of the loan.

For example, the amortisation table shown models a $3500 reducing balance loan with an interest rate of 9% per annum, compounding monthly, and monthly repayments of $600.

payment number	payment	interest	principal reduction	balance of loan
0	0.00	0.00	0.00	3500.00
1	600.00	26.25	573.75	2926.25
2	600.00	21.95	578.05	2348.20
3	600.00	17.61	582.39	1765.81
4	600.00	13.24	586.76	1179.05
5	600.00	8.84	591.16	587.89
6	592.30	4.41	587.89	0.00

Note: Often the final payment is adjusted so that the balance can be fully paid off. The final payment amount does not equal the previous loan balance because there is additional interest that needs to be repaid.

The columns of an amortisation table for a reducing balance loan can be calculated using the following formulas:

$$\textit{interest} = \frac{r}{100} \times \textit{previous loan balance}$$

$$\textit{principal reduction} = \textit{payment} - \textit{interest}$$

$$\textit{balance of loan} = \textit{previous loan balance} - \textit{principal reduction}$$

Worked example 2

The first four lines of an amortisation table for a reducing balance loan of $500 000 are shown.

Interest is charged at a rate of 6.2% per annum, compounding quarterly.

The loan is to be repaid with quarterly payments.

There are a number of cells missing.

payment number	payment	interest	principal reduction	balance of loan
0	0.00	0.00	0.00	500 000.00
1	10 000.00	7750.00	2250.00	497 750.00
2	**a.**	7715.13	2284.87	495 465.13
3	10 000.00	**b.**	**c.**	**d.**

a. Calculate the missing payment value for payment number 2.

Explanation

Step 1: Determine the interest and principal reduction for payment number 2.

$\textit{interest} = 7715.13$

$\textit{principal reduction} = 2284.87$

Step 2: Calculate the payment.

$\textit{principal reduction} = \textit{payment} - \textit{interest}$

$\textit{payment} = \textit{principal reduction} + \textit{interest}$

$= 7715.13 + 2284.87$

$= 10\ 000.00$

Answer

$10 000.00

Continues →

Back to contents

b. Calculate the missing interest for payment number 3.

Explanation

Step 1: Identify the interest rate and previous loan balance.

$r = \frac{6.2}{4}\%$ per compounding period

$\textit{previous loan balance} = 495\,465.13$

Step 2: Calculate the interest.

$$\begin{aligned} \textit{interest} &= \frac{r}{100} \times \textit{previous loan balance} \\ &= \frac{6.2}{400} \times 495\,465.13 \\ &\approx 7679.71 \end{aligned}$$

Answer

$7679.71

c. Calculate the missing principal reduction for payment number 3.

Explanation

Step 1: Identify the payment and interest.

$\textit{payment} = 10\,000.00$

$\textit{interest} = 7679.71$ (from part **b**)

Step 2: Calculate the principal reduction.

$$\begin{aligned} \textit{principal reduction} &= \textit{payment} - \textit{interest} \\ &= 10\,000.00 - 7679.71 \\ &= 2320.29 \end{aligned}$$

Answer

$2320.29

d. Calculate the missing loan balance after payment number 3.

Explanation

Step 1: Identify the previous loan balance and principal reduction.

$\textit{previous loan balance} = 490\,465.13$

$\textit{principal reduction} = 2320.29$ (from part **c**)

Step 2: Calculate the loan balance.

$$\begin{aligned} \textit{balance of loan} &= \textit{previous loan balance} \\ &\quad - \textit{principal reduction} \\ &= 495\,465.13 - 2320.29 \\ &= 493\,144.84 \end{aligned}$$

Answer

$493 144.84

Using financial applications of technology to solve problems involving reducing balance loans

The financial solver is a useful tool that can be used to solve problems involving reducing balance loans.

The present value, PV, is always positive since the borrower receives this amount from the lender.

The payment, PMT, is always negative since the borrower makes a payment to the lender.

The future value, FV, is always negative as it represents money that is owed, or yet to be paid.

A future value of 0 indicates that the loan has been fully repaid.

Back to contents

Worked example 3

Cilian takes out a reducing balance loan of $35 000 with a bank that charges an interest rate of 5.9% per annum, compounding monthly. He makes a payment of $1000 each month to pay off the loan.

a. How much does Cilian still owe after 21 months?

Explanation

Step 1: Determine the financial solver inputs.

N	21	(there are 21 months)
I(%)	5.9	(annual interest rate)
PV	35 000	(this is positive because Cilian receives it from the lender)
PMT	−1000	(this is negative because Cilian pays the lender)
FV		
PpY	12	(payments made monthly)
CpY	12	(interest compounds monthly)

Step 2: Use the financial solver to solve for FV.

FV	−16 731.708...

The negative FV indicates that Cilian still owes the bank.

Answer

$16 731.71

b. The bank changes the interest rate for Cilian's loan after 21 months to 6.4% per annum compounding monthly. Cilian's monthly payments remain unchanged.

To ensure the loan is fully repaid, the final repayment will be lower.

If Cilian manages to pay off the rest of the loan, how much interest will he pay in total?

Use the rounded value found in part **a**.

Explanation

Step 1: Calculate the number of compounding periods it will take Cilian to pay off the loan.

Cilian still owes $16 731.71 when the interest rate changes.

This is now the present value of the loan.

N		
I(%)	6.4	(annual interest rate)
PV	16 731.71	(this is positive since it's the amount owing)
PMT	−1000	(this is negative since Cilian pays the lender)
FV	0	(the loan is to be fully repaid)
PpY	12	(payments made monthly)
CpY	12	(interest compounds monthly)

N	17.572...

This means that after the first 21 months, there will be 17 full payments of $1000 and 1 final payment less than $1000.

Continues →

Back to contents

Step 2: Calculate the value of the final payment.

N	17	
I(%)	6.4	(annual interest rate)
PV	16 731.71	(this is positive since it's the amount owing)
PMT	−1000	(this is negative since Cilian pays the lender)
FV		
PpY	12	(payments made monthly)
CpY	12	(interest compounds monthly)

FV	−570.161...

This means that after 17 payments since the interest rate change, Cilian still owes \$570.16.

This is not equal to the final payment amount because additional interest still needs to be paid.

N	1	(one final payment)
I(%)	6.4	(annual interest rate)
PV	570.161...	(this is positive since it's the amount owing)
PMT		
FV	0	(the loan is to be fully repaid)
PpY	12	(payments made monthly)
CpY	12	(interest compounds monthly)

PMT	−573.201...

The final payment is \$573.20.

Step 3: Calculate the total amount of money that Cilian pays the lender.

The 21 payments before the interest rate was changed need to be included as well.

$$\begin{aligned} \textit{total amount paid} &= 21 \times 1000 + 17 \times 1000 + 573.20 \\ &= 38\,573.20 \end{aligned}$$

Step 4: Calculate the total interest that Cilian pays the lender.

The interest is the difference between the total amount paid and the amount that is borrowed.

$$\begin{aligned} \textit{interest} &= \textit{total amount paid} - \textit{total amount borrowed} \\ &= 38\,573.20 - 35\,000 \\ &= 3573.20 \end{aligned}$$

Answer

\$3573.20

Exam question breakdown

VCAA 2019 Exam 2 Recursion and financial modelling Q9a

Phil would like to purchase a block of land.

He will borrow \$350 000 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 Phil's loan will be \$262 332.33.

What is the value of each fortnightly repayment Phil will make?

Round to the nearest cent. (1 MARK)

Explanation

Step 1: Determine the financial solver inputs.

N	78	(there are 78 fortnights in 3 years)
I(%)	4.9	(annual interest rate)
PV	350 000	(this is positive because Phil receives it from the lender)
PMT		
FV	−262 332.33	(this is negative because Phil still owes the lender)
PpY	26	(payments made fortnightly)
CpY	26	(interest compounds fortnightly)

Step 2: Use the financial solver to solve for PMT.

PMT	−1704.0300...

The PMT is negative because Phil pays the lender.

27% of students answered this question correctly.

A significant number of students incorrectly entered a positive FV value into the financial solver. The future value of a loan needs to be negative as it represents the money that is owed, or yet to be paid. A few students incorrectly rounded to \$1704.05 or \$1704.

Answer

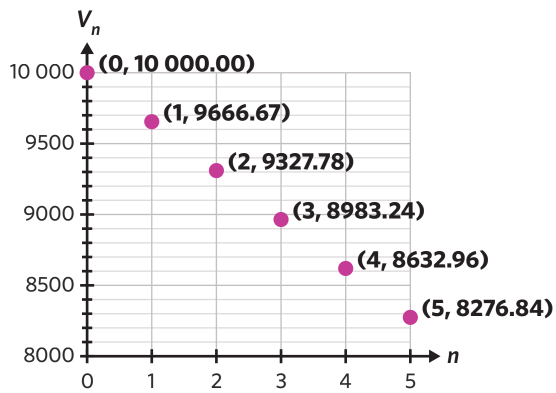
\$1704.03

6B Questions

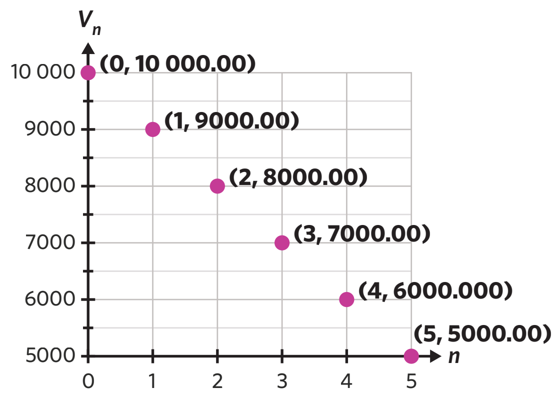
Using recurrence relations to model reducing balance loans

1. Which of the following graphs is most likely to represent the value of a reducing balance loan over 5 years?

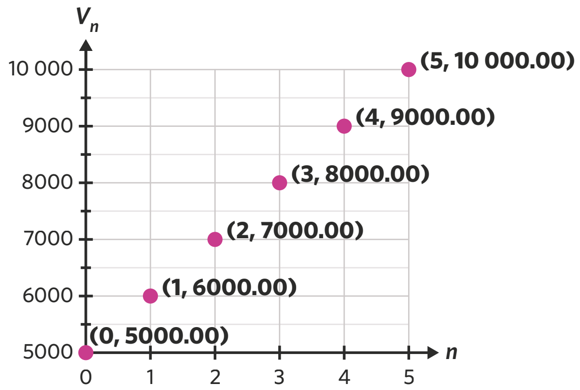
A.



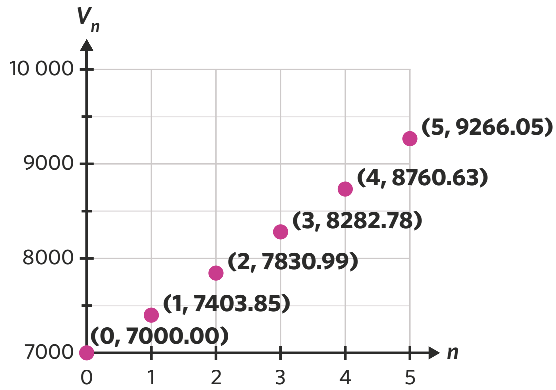
B.



C.



D.



2. Skyla took out a loan that can be modelled by the following recurrence relation, where V_n represents the value of the loan after n months.

$V_0 = 10\,000, \quad V_{n+1} = 1.004 \times V_n - 600$

a. How much did Skyla borrow?

b. How much does Skyla pay the lender each month?

c. What is the annual interest rate of Skyla's loan?

3. Construct a recurrence relation for each scenario, where V_n represents the value of the reducing balance loan after n compounding periods.

a. Brycen takes out a loan of $8500 with an annual interest rate of 4.68% per annum, compounding fortnightly. A payment of $250 is made every fortnight to pay off the loan.

b. Roxie takes out a mortgage of $985 000 with an annual interest rate of 3.12% per annum, compounding weekly, and repayments of $1200 each week.

4. Grant takes out a loan to start his own food truck business.

The loan can be modelled by the recurrence relation

$V_0 = 28\,000, \quad V_{n+1} = 1.0056 \times V_n - 2100$

where n represents the number of months after the loan is established.

a. What is the annual interest rate of Grant's loan? Round to two decimal places.

b. How much will Grant owe after three months? Round to the nearest cent.

c. After how many months will Grant owe less than $20 000?

5. Candice took out a loan of $15 000 to buy her car. Interest is charged at 39% per annum, compounding fortnightly.

Candice has been making repayments of $1800 each fortnight. The current balance of the loan is $10 203.77. How many fortnights has it been since Candice established the loan?

Using amortisation tables to solve problems involving reducing balance loans

6. The first four lines of an amortisation table for a reducing balance loan are shown. Interest is compounded on a fortnightly basis.

payment number	payment	interest	principal reduction	balance of loan
0	0.00	0.00	0.00	94 000.00
1	5600.00	455.54	5144.46	88 855.54
2	5600.00	430.61	5169.39	83 686.15
3	5600.00	405.56	5194.44	78 491.71

a. The amount that was borrowed is

A. $0.00 B. $455.54 C. $5600.00 D. $94 000.00

b. The principal reduction after payment number 2 is

A. $0.00 B. $5144.46 C. $5169.39 D. $5194.44

c. The interest rate, per annum, is closest to

A. 0.48% B. 1.26% C. 4.80% D. 12.60%

7. A loan of $18 000 has an interest rate of 6% per annum, compounding monthly.
Repayments of $1627 are made each month.
Use this information to complete the amortisation table.

payment number	payment	interest	principal reduction	balance of loan
0	0.00	0.00	0.00	18 000.00
1				

8. The following amortisation table models a reducing balance loan. Fill in the empty cells.

payment number	payment	interest	principal reduction	balance of loan
0	0.00	0.00	0.00	1200.00
1	242.42	4.02	238.40	961.60
2	242.42		239.20	722.40
3	242.42	2.42	240.00	
4	242.42	1.62		241.60
5		0.81	241.60	0.00

9. Consider the following amortisation table for a reducing balance loan. Interest is compounded quarterly.

payment number	payment	interest	principal reduction	balance of loan
0	0.00	0.00	0.00	50 000.00
1	1450.00	300.00	1150.00	48 850.00
2	1450.00	293.10	1156.90	47 693.10
3	1450.00	286.16	1163.84	46 529.26
4	1450.00	279.18	1170.82	45 358.44

Calculate the interest rate, per annum, of the loan.

10. Joan has agreed to a loan of $3000 with her bank. The loan has an interest rate of 4% per annum, compounding monthly. Joan makes monthly payments of $500 towards the loan.
Fill out the amortisation table for the first three payments.

payment number	payment	interest	principal reduction	balance of loan
0	0.00	0.00	0.00	3000.00
1				
2				
3				

Back to contents

Using financial applications of technology to solve problems involving reducing balance loans

11. Jasmine takes out a loan of \$560 000 to pay for her mortgage. She makes monthly repayments of \$3006.20 and will pay off the loan after 30 years.

The annual interest rate is closest to

A. 4% p.a. **B.** 5% p.a. **C.** 10% p.a. **D.** 15% p.a.

12. Use a financial solver for each of the following.

a. Clay takes out a \$20 000 loan from the bank to pay for a holiday trip to Abu Dhabi. The interest rate is 10.2% per annum, compounding quarterly. Clay makes quarterly repayments of \$1000.

How long will it take for Clay to owe \$16 866.08? Round to the nearest quarter.

b. Blaine borrows \$42 000 from a loan shark to pay for a car. The loan shark charges interest and forces Blaine to make weekly payments of \$2000 a week. 12 weeks after establishing the loan, Blaine still owes \$21 003.35.

What is the annual interest rate that is being charged? Round to the nearest percentage.

c. Sabrina needs to borrow some money to pay for a house. Her bank offers an interest rate of 6.4% per annum, compounding biannually. The loan is to be paid off after 5 years of biannual payments of \$4974.07.

How much does Sabrina borrow? Round to the nearest thousand dollars.

d. Rinto took out a loan of \$25 000 to pay for a car. Interest is calculated at a rate of 12.4% per annum, compounding fortnightly. He makes repayments on a fortnightly basis and owes \$21 979.13 after four years.

How much is Rinto paying each fortnight? Round to the nearest cent.

13. Drake wants to purchase the latest phone for \$1299 but doesn't have enough money on hand. He places a 10% deposit and borrows the remainder of the money from a tech store that charges an interest rate of 1% p.a. compounding weekly.

Drake is required to pay the tech store \$50 on a weekly basis.

If Drake now owes the tech store \$920.13, how long has it been since he purchased the phone?

14. Miette borrows \$19 000 with an interest rate of 7.8% per annum, compounding monthly. She makes monthly repayments of \$240.

a. How much does she owe after two years? Round to the nearest cent.

The bank decides to increase the interest rate for Miette's loan after two years to 9.5% per annum. Miette's monthly repayments remain unchanged.

To ensure the loan is fully repaid, the final repayment will be a larger amount.

b. After the two years, how many months will it take for Miette to fully repay the loan? Use the rounded value found in part **a**.

c. How much does Miette pay in total?

d. How much interest does Miette pay in total?

Joining it all together

15. An amortisation table for a reducing balance loan is shown. Interest compounds monthly.

payment number	payment	interest	principal reduction	balance of loan
0	0.00	0.00	0.00	5000.00
1	1200.00	30.00	1170.00	3830.00
2	1200.00	22.98	1177.02	2652.98
3	1200.00	15.92	1184.08	1468.90

a. Calculate the annual interest rate.

b. Construct a recurrence relation that can be used to model the value of this loan.

c. If the loan is to be fully repaid in 5 months, calculate the value of the final repayment.

---

16. The value of a reducing balance loan, V_n, after n quarters, can be modelled by the recurrence relation, as shown.

$V_0 = 24\,745, \quad V_{n+1} = R \times V_n - 3816$

The loan is to be fully repaid in 7 quarters. Calculate the value of R, rounded to four decimal places.

---

17. Frank wishes to buy a high-end gaming chair. He borrows \$8000 at 10.98% per annum, compounding weekly. Frank's weekly repayments are chosen so he pays off the chair in one year.

a. How much is Frank paying each week, correct to the nearest cent?

After 10 weeks, Frank decides to increase his repayments by \$50 per week.

b. To ensure the loan is fully repaid, the final repayment will be lower.

In total, how long will it take Frank to pay off the loan?

c. What is the total amount of interest that Frank will pay in the first 16 weeks?

Use the rounded value found in part a.

Exam practice

18. The value of a reducing balance loan, in dollars, after n months, V_n, can be modelled by the recurrence relation shown.

$V_0 = 26\,000, \quad V_{n+1} = 1.003 \times V_n - 400$

What is the value of this loan after five months?

A. \$24 380.31 **B.** \$24 706.19 **C.** \$25 031.10

D. \$25 355.03 **E.** \$25 678.00

VCAA 2017 Exam 1 Recursion and financial modelling Q17

82% of students answered this question correctly.

---

19. 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	0.00	0.00	0.00	250 000.00
1	1500.00	1000.00	500.00	249 500.00
2	1500.00			

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

VCAA 2016 Exam 1 Recursion and financial modelling Q22

54% of students answered this question correctly.

---

20. Bimal has a reducing balance loan.

The balance, in dollars, of the loan from month to month, B_n, is modelled by the recurrence relation shown.

$B_0 = 450\,000, \quad B_{n+1} = R \times B_n - 2633$

Back to contents

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

VCAA 2021 Exam 1 Recursion and financial modelling Q23

31% of students answered this question correctly.

---

21. For renovations to her coffee shop, Sienne took out a reducing balance loan of $570 000, with interest calculated fortnightly.

The balance of the loan, in dollars, after n fortnights, S_n, can be modelled by the recurrence relation

$S_0 = 570\,000, \quad S_{n+1} = 1.001 \times S_n - 1193$

Show that the compound interest rate for this loan is 2.6% per annum. (1 MARK)

VCAA 2021 Exam 2 Recursion and financial modelling Q8b

27% of students answered this question correctly.

---

22. Ken has borrowed $70 000 to buy a new caravan.

He will be charged interest at a rate of 6.9% per annum, compounding monthly.

Ken will make monthly repayments of $800.

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 Ken's loan is fully repaid in a further three years.

Ken's repayment amount and interest remains the same.

What is the value of Ken's lump sum payment, $\$L$? Round to the nearest dollar. (2 MARKS)

VCAA 2016 Exam 2 Recursion and financial modelling Q7b

The average mark on this question was **0.2**.

Questions from multiple lessons

Data analysis

23. Dayne works at a restaurant each day of the week.

The table shows the daily seasonal indices for the amount of money Dayne earns each day.

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
seasonal index	0.68	0.72	0.73	0.93	1.06	1.38	

The deseasonalised amount of money Dayne earned last Sunday was $185.

How much money did Dayne actually make last Sunday?

A. $123.33 **B.** $185.00 **C.** $255.30 **D.** $277.50 **E.** $134.06

Adapted from VCAA 2018NH Exam 1 Data analysis Q14

Recursion and financial modelling

24. A plasma screen TV was purchased for $3500.

The TV depreciates 12% each year using the reducing balance method.

A recurrence relation that can be used to determine the value of the TV after n years, T_n, is

A. $T_n = 3500, \quad T_{n+1} = 1.12T_n$

B. $T_n = 3500, \quad T_{n+1} = 0.88T_n$

C. $T_n = 3500, \quad T_{n+1} = 0.88T_n - 12$

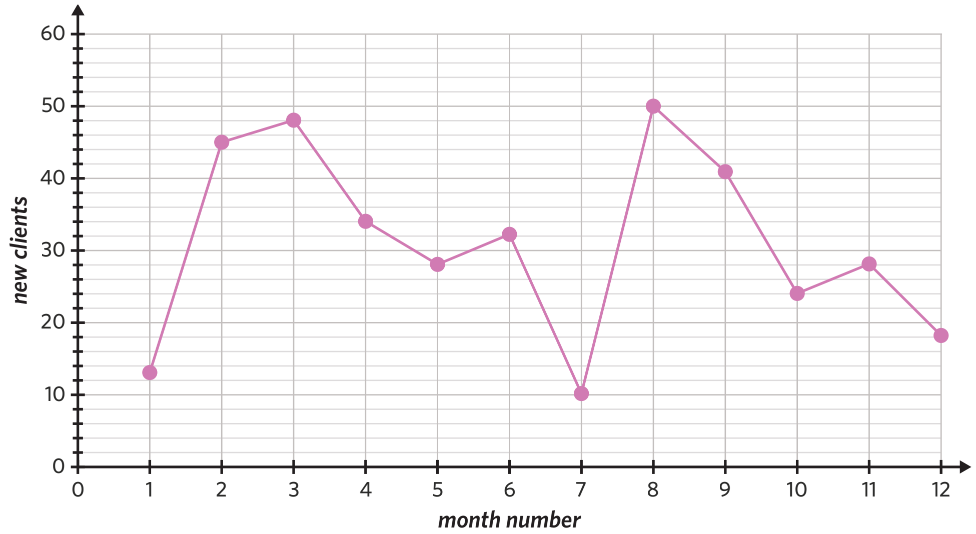
D. $T_n = 3500, \quad T_{n+1} = T_n + 12$

E. $T_n = 3500, \quad T_{n+1} = T_n - 12$

Adapted from VCAA 2016 Exam 1 Recursion and financial modelling Q19

Data analysis

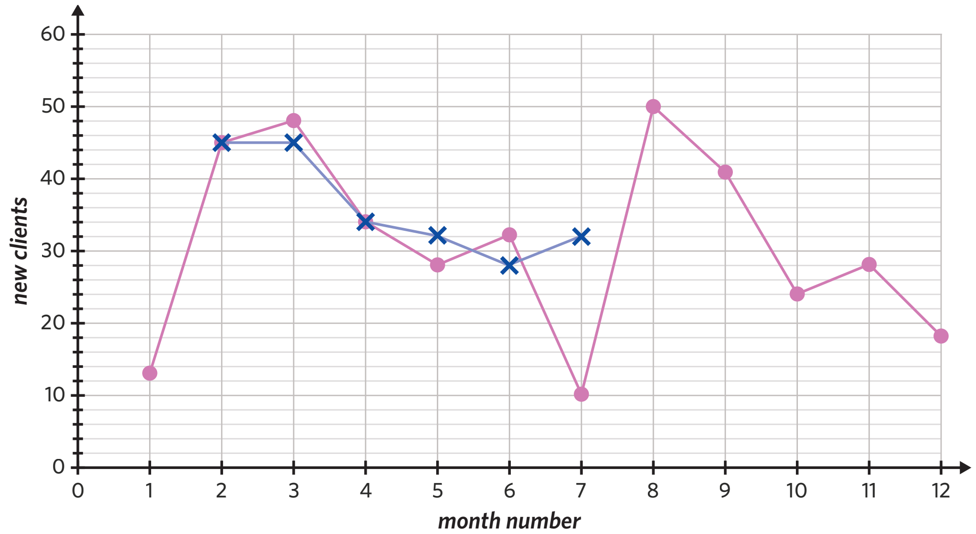
25. The following time series plot shows the total number of *new clients* a consultancy agency acquires each month for 2018.



a. Find the range of *new clients*? (1 MARK)

b. Determine the five-median smoothed number of *new clients* for month 9. (1 MARK)

Three-median smoothing has been used to smooth the time series plot up to month 7, as shown.



c. Complete the three-median smoothing by marking each remaining smoothed point with a cross (×). (2 MARKS)

Adapted from VCAA 2017NH Exam 2 Data analysis Q4

Back to contents

6C Interest-only loans

STUDY DESIGN DOT POINTS

- 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

6A 6B **6C** 6D 6E 6F

KEY SKILLS

During this lesson, you will be:

- using recurrence relations to model interest-only loans
- using financial applications of technology to solve problems involving interest-only loans.

KEY TERMS

- Interest-only loan

When making big purchases, such as a home, it generally requires a large sum of money to be borrowed from the bank. In order to make repayments more affordable, the borrower can opt to make interest-only payments at the start of the loan. This makes the initial regular repayments cheaper as the borrower will not repay any of the principal, however it may increase the total amount paid in the long run.

Using recurrence relations to model interest-only loans

In an **interest-only loan**, the borrower only repays the interest that is charged. As a result of this, the principal that must be paid back remains the same after each compounding period.

See worked example 1

The value of an interest-only loan, V_n, after n compounding periods, can be modelled by a recurrence relation of the same form as a reducing balance loan.

$V_0 = principal, \quad V_{n+1} = R \times V_n - d$, where

- $R = 1 + \frac{r}{100}$, where r is the interest rate (%) per compounding period
- d is the payment made per compounding period

The payment is equal to the interest earned each compounding period and can therefore be calculated as

$d = \frac{r}{100} \times V_0$

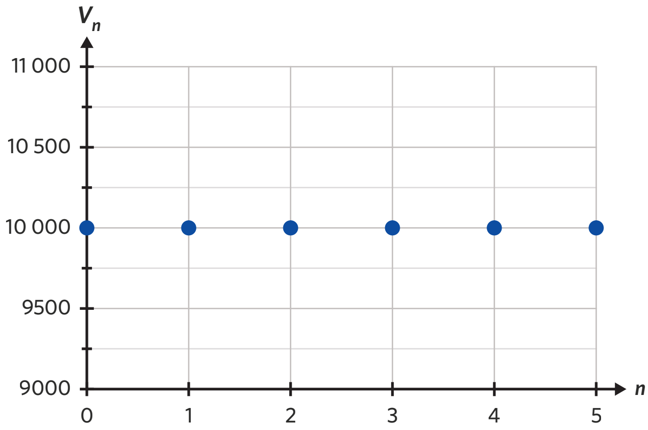
For example, \$10 000 is borrowed at an interest rate of 4% per annum, compounding annually. The interest charged, 4% of \$10 000, is \$400. If the borrower pays back \$400 at the end of each year, this keeps the principal balance at \$10 000 indefinitely. A recurrence relation that can be used to model the value of the loan, V_n, after n years is

$V_0 = 10\,000, \quad V_{n+1} = 1.04 \times V_n - 400.$

Back to contents

The balance of the loan over a period of 5 years can also be represented on a graph.

Notice that it is constant.



Interest-only loans can be modelled using a table as well, similar to an amortisation table. However, there will be no change from one payment to the next because the principal balance will not reduce.

See worked example 2

Worked example 1

Brock borrows \$8000 in an interest-only loan at 3% per annum, compounding monthly.

a. What is the monthly amount that he will be required to pay?

Explanation

Step 1: Determine the principal and interest rate for the interest-only loan.

$V_0 = 8000$

$r = \frac{3}{12}\%$ per compounding period

Step 2: Calculate d.

$$d = \frac{3}{12 \times 100} \times 8000$$
$$= \frac{3}{1200} \times 8000$$
$$= 20$$

Answer

\$20

b. Construct a recurrence relation that can be used to model the value of Brock's loan, V_n, after n months.

Explanation

Step 1: Determine the principal, interest rate and payment amount.

$V_0 = 8000$

$r = \frac{3}{12}\%$ per compounding period

$d = 20$

Step 2: Calculate R.

$$R = 1 + \frac{3}{12 \times 100}$$
$$= 1 + \frac{3}{1200}$$
$$= 1.0025$$

Step 3: Construct the recurrence relation.

Answer

$V_0 = 8000, \quad V_{n+1} = 1.0025 \times V_n - 20$

Worked example 2

Felicity has an interest-only loan of $45 000, interest is charged at a rate of 3% per annum, compounding weekly. This is modelled in the following table similar to an amortisation table.

payment number	payment	interest	principal reduction	balance of loan
0	0.00	0.00	0.00	45 000.00
1	25.96	25.96	0.00	45 000.00
2	25.96	25.96	0.00	45 000.00
3	25.96			

Fill in the missing cells for payment number 3.

Explanation

Step 1: Calculate the interest for payment number 3.

The interest amount is equal to the payment received, which is $25.96.

Step 2: Calculate the principal reduction for payment number 3.

The value of an interest-only loan remains constant, so there is no reduction of the principal.

Step 3: Calculate the loan balance after payment number 3.

The balance does not decrease in an interest-only loan, so it will remain at $45 000.00.

Answer

payment number	payment	interest	principal reduction	balance of loan
0	0.00	0.00	0.00	45 000.00
1	25.96	25.96	0.00	45 000.00
2	25.96	25.96	0.00	45 000.00
3	25.96	25.96	0.00	45 000.00

Using financial applications of technology to solve problems involving interest-only loans

The financial solver is a handy program that can be used to solve problems involving interest-only loans.

The number of payments, N, can be set to any value since the value of the interest-only loan remains constant.

The present value, PV, is always positive since the borrower receives this amount from the lender.

The payment, PMT, is always negative since the borrower makes a payment to the lender.

The future value, FV, is always negative as it represents money that is owed, or yet to be paid. It is also the negative value of the PV.

The financial solver can be used to calculate an unknown interest rate I(%) or payment amount PMT, but not an unknown principal value, since both the PV and FV would be unknown.

Back to contents

Worked example 3

Mark borrows $40 000 in an interest-only loan to help fund his business. He is required to make quarterly payments to offset the interest, compounding quarterly at 4.7% p.a.

a. How much is Mark required to pay each quarter?

Explanation

Step 1: Determine the financial solver inputs.

N	1	(can be any value since the balance is unchanged)
I(%)	4.7	(annual interest rate)
PV	40 000	(this is positive because Mark receives it from the lender)
PMT		
FV	−40 000	(this is negative because Mark still owes the lender)
PpY	4	(payments made quarterly)
CpY	4	(interest compounds quarterly)

Step 2: Use the financial solver to solve for PMT.

PMT	−470

The PMT is negative because Mark pays it to the lender.

Answer

$470

b. Mark decides that after he has paid more than $10 000 in interest, he will close the account and end the interest-only loan. How many quarters is it before he closes the account?

Explanation

Step 1: Determine the payment amount made each quarter.

payment = 470

Step 2: Calculate the number of quarters it takes to reach $10 000.

$$\frac{10\,000}{470} = 21.2765...$$

This must be rounded up. At 21 quarters, Mark will not have exceeded the $10 000 threshold.

Answer

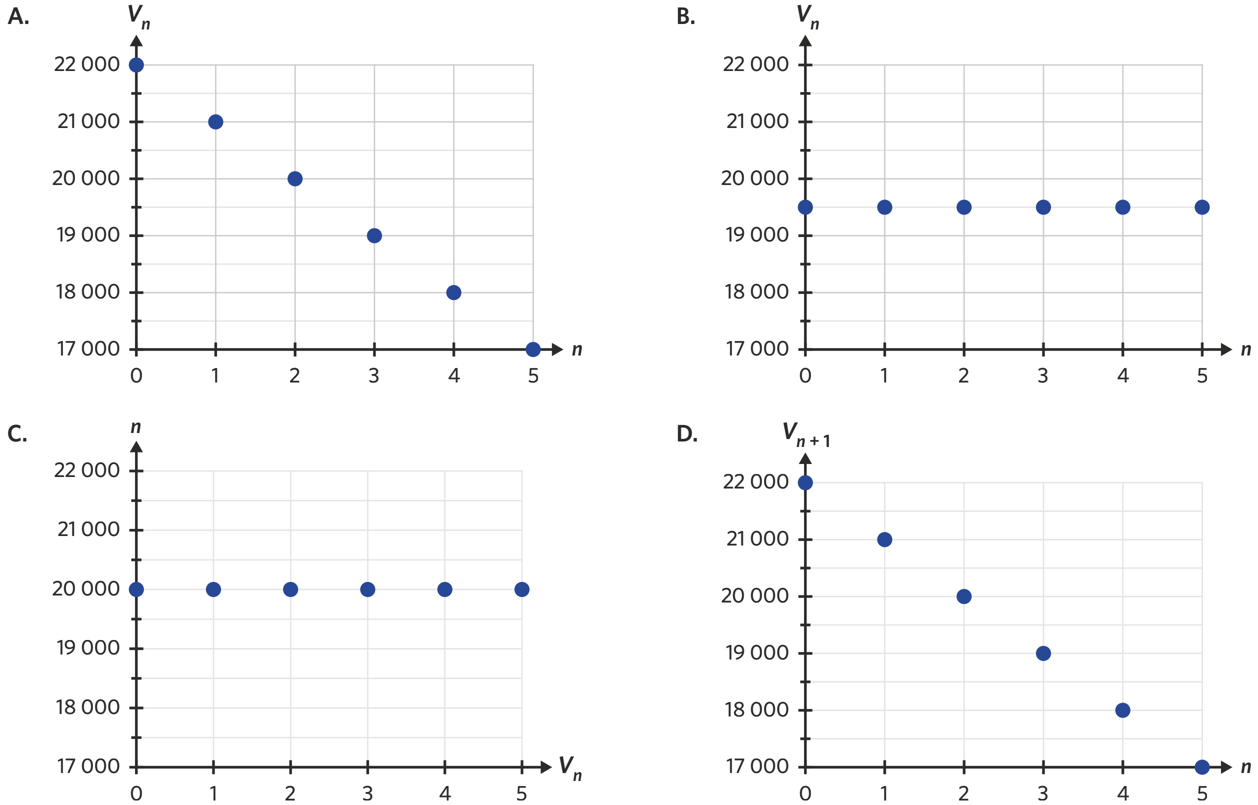
22 quarters

Back to contents

6C Questions

Using recurrence relations to model interest-only loans

1. Which of the following graphs is most likely to represent the value of an interest-only loan across 5 years?



2. The monthly repayments of an interest-only loan are described by the following lines of working.

$$d = \frac{r}{100} \times V_0$$
$$= \frac{36}{1200} \times 8000$$
$$= 240$$

a. What is the principal of this interest-only loan?

b. What is the monthly repayment?

c. What is the annual interest rate?

d. Construct a recurrence relation that can be used to model the value of the interest-only loan, V_n, after n compounding periods.

3. Calculate the regular repayment amount required for each of the following interest-only loans. Round to the nearest cent.

a. A \$12 000 loan with 2.8% interest per month, compounding monthly. Payments are made monthly.

b. A \$22 000 loan with 14% interest per annum, compounding annually. Payments are made annually.

c. A \$1600 loan with 18.9% interest per annum, compounding quarterly. Payments are made quarterly.

d. A \$4850 loan with 2.6% interest per annum, compounding fortnightly. Payments are made fortnightly.

4. Terry has taken out an interest-only loan. Every month, he makes a payment of $465. The interest rate is 24% per annum and interest is compounded monthly.
   a. What is the principal of the loan?
   b. Construct a recurrence relation that can be used to model V_n, the value of the loan after n months.

5. Find the annual interest rates for the following interest-only loans, rounded to one decimal place.
   a. An interest-only loan of $21 000, compounding monthly, with monthly payments of $315.
   b. An interest-only loan of $16 000, compounding quarterly, with quarterly payments of $86.
   c. An interest-only loan of $82 000, compounding fortnightly, with fortnightly payments of $1353.

6. Jennifer borrows $750 000 from the bank to purchase an investment property. She will make monthly interest-only payments for the first 5 years. Her bank charges interest at a rate of 5% per annum, compounding monthly.

   The following table is a snippet of this loan.

payment number	payment	interest	principal reduction	balance of loan
11				
12				
13				
14				

   Fill in the missing cells. Round values to the nearest cent.

Using financial applications of technology to solve problems involving interest-only loans

7. What is the annual interest rate required to establish an interest-only loan of $4500, with monthly repayments of $9? Interest compounds on a monthly basis.

   A. 1.2% p.a. **B.** 2.0% p.a. **C.** 2.3% p.a. **D.** 2.4% p.a.

8. Determine the annual interest rate required for each of the following interest-only loans. Round to one decimal place.
   a. An interest-only loan of $48 000 with interest compounding monthly. Payments of $415.10 are made monthly.
   b. An interest-only loan of $8122.50 with interest compounding fortnightly. Payments of $162.45 are made fortnightly.
   c. An interest-only loan of $12 700 with interest compounding quarterly. Payments of $506.13 are made quarterly.

9. Determine the payment required for each of the following interest-only loans. Give answers correct to the nearest cent.
   a. An interest-only loan of $16 000 at 5.7% interest per annum compounding yearly. Payments are made annually.
   b. An interest-only loan of $250 000 at 7.6% interest per annum compounding fortnightly. Payments are made fortnightly.
   c. An interest-only loan of $550 000 at 5.3% interest per annum compounding monthly. Payments are made monthly.

Back to contents

10. Lee wants to purchase a sports car worth $547 841. He places a 30% deposit and finances the rest of the car. The car dealer offers an interest-only loan for the first 3 years of the loan and Lee wants to take full advantage of this offer. Interest compounds weekly and Lee is required to make weekly payments.
    - a. How much did Lee borrow?
    - b. If Lee is required to pay $221.24 each week, then what is the annual interest rate for this loan? Give the answer correct to two significant figures.
    - c. How much interest has Lee paid by the end of the interest-only period?

Joining it all together

11. Harvey is looking to borrow $500 000 and invest it wisely, in the hope of making a decent profit over the next four years. He likes the idea of an interest-only loan and has been presented with various offers.

    Melbourne Bank offers Harvey an interest rate of 6.9% per annum, compounding monthly. Harvey must make monthly repayments.
    - a. How much interest will Harvey pay across the four years?

    Sydney Bank offers the $500 000 loan with the total interest amount being $150 000 across the four years. Payments are to be made monthly and interest compounds monthly.
    - b. What is the annual interest rate of this option?
    - c. Write a recurrence relation that can be used to model the value of the loan, V_n, after n payment periods using Sydney Bank's offer.
    - d. Which option is better for Harvey? Explain why.

12. Cassandra wants to borrow some money to pay for a beach house in Byron Bay. She places a deposit of 10% and decides to take out an interest-only loan to cover the remaining cost of the house.

    Her bank offers an interest rate of 7.3% per annum, compounding daily. She will be paying $180 every day.
    - a. How much does the beach house cost?
    - b. Write a recurrence relation that can be used to model the value of Cassandra's loan, V_n after n days.

    In three years, her bank changes the interest rate and she is required to increase her daily repayments by $90 to maintain the interest-only loan. Interest continues to compound daily.
    - c. How much interest has Cassandra paid in the first three years? Assume that there were no gap years during this time.
    - d. How much does Cassandra still owe?
    - e. How much has the annual interest rate increased by?

Exam practice

13. Bob borrowed $400 000 to buy an apartment.

    The interest rate for this loan was 3.14% per annum, compounding monthly.

    A scheduled monthly repayment that allowed Bob to fully repay the loan in 20 years was determined.

    Bob decided, however, to make interest-only repayments for the first two years.

    After these two years, the interest rate changed. Bob was still able to pay off the loan in the 20 years by repaying the scheduled amount each month.

    The interest rate, per annum, for the final 18 years of the loan was closest to

    A. 1.85% **B.** 2.21% **C.** 2.79%
    D. 3.14% **E.** 4.07%

    VCAA 2021 Exam 1 Recursion and financial modelling Q24

33% of students answered this question correctly.

Back to contents

Questions from multiple lessons

Recursion and financial modelling

14. Percy has a compound interest investment that earns interest compounding quarterly.

The balance of Percy's compound interest investment was \$27 060.80 after one year.

After three years, the balance of the investment was \$31 706.04.

The value of Percy's initial investment is closest to

A. \$20 000 **B.** \$21 337 **C.** \$24 669 **D.** \$25 000 **E.** \$25 669

Adapted from VCAA 2017NH Exam 1 Recursion and financial modelling Q24

Data analysis

15. Jeff is conducting some research on data from 2016 on how the distance a second-hand car was driven (in 000's of km) impacted the car's price. A least squares line has been fitted to the data and its equation is

$price = 15\,129 - 68 \times distance$

In 2016, the price of a second-hand car that had driven 123 000 km was \$6299. When the least squares line is used to predict the price of this car, the residual is closest to

A. −\$8 348 871 **B.** −\$6765 **C.** −\$466 **D.** \$466 **E.** \$6765

Adapted from VCAA 2016 Exam 1 Data analysis Q10

Recursion and financial modelling

16. A 3D printer was purchased for \$4600.

The value of the 3D printer can be depreciated using the reducing balance method of depreciation. Its value after one year was \$4324 and its value after two years was \$4064.56.

a. Show that the depreciation rate of the 3D printer is 6% per annum. (1 MARK)

b. Let V_n be the value of the 3D printer after n years.

Determine a recurrence relation in terms of V_{n+1}, V_n and V_0 that displays the value of the 3D printer after n years. (1 MARK)

Adapted from VCAA 2017NH Exam 2 Recursion and financial modelling Q5d

Back to contents

6D Amortising annuities

STUDY DESIGN DOT POINTS

- 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

6A 6B 6C 6D 6E 6F

KEY SKILLS

During this lesson, you will be:

- using recurrence relations to model amortising annuities
- using amortisation tables to solve problems involving amortising annuities
- using financial applications of technology to solve problems involving amortising annuities.

KEY TERMS

- Annuity
- Amortising annuity

An amortising annuity can take on the form of a retirement investment fund. This occurs when an individual's superannuation, a large lump sum, is invested with a bank or super fund where interest is earned but regular payments are taken out for the retiree to use. It is also referred to as a retirement income stream.

Using recurrence relations to model amortising annuities

An **annuity** is an investment that involves a fixed sum of money paid to an investor at regular intervals, and typically earns compound interest. In the case of an **amortising annuity**, the balance of the investment will decrease over time until it drops to zero or the account is closed.

The value of an amortising annuity, V_n, after n compounding periods, can be modelled by a recurrence relation of the form

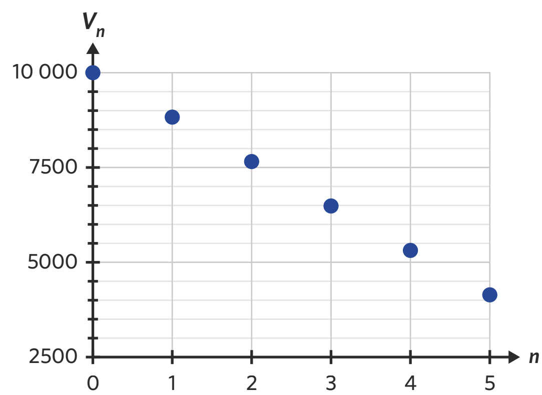
$V_0 = principal, \quad V_{n+1} = R \times V_n - d$, where

- $R = 1 + \frac{r}{100}$, where r is the interest rate (%) per compounding period
- d is the payment received per compounding period

For example, Lara has an annuity of \$10 000 that is opened at an interest rate of 12% per annum, compounding monthly. The annuity provides an income of \$1250 per month. A recurrence relation that can be used to determine V_n, the value of the investment after n months is

$V_0 = 10\,000, \quad V_{n+1} = 1.01 \times V_n - 1250.$

The balance of Lara's investment over a period of 5 months can also be represented on a graph.



There is a slight curve in the graph. This is because, at the beginning of the annuity, a larger amount of interest is earned, so a smaller portion of the payment reduces the principal.

Worked example 1

Tom invests \$25 000 in an annuity with an interest rate of 7% per annum, compounding monthly. This annuity will provide him with a monthly income of \$2500.

a. Construct a recurrence relation that can be used to model V_n, the value of this annuity after n compounding periods. Round R to 5 decimal places.

Explanation

Step 1: Determine the principal, interest rate and payment amount.

$V_0 = 25\,000$

$r = \frac{7}{12}\%$ per compounding period

$d = 2500$

Step 2: Calculate R.

$R = 1 + \frac{7}{12 \times 100}$

$R = 1 + \frac{7}{1200}$

$R = 1.00583...$

Step 3: Construct the recurrence relation.

Answer

$V_0 = 25\,000, \quad V_{n+1} = 1.00583 \times V_n - 2500$

b. Calculate the value of the investment after 4 months.

Explanation

Step 1: Calculate V_1, the value of the investment after 1 month.

$V_1 = 1.00583 \times V_0 - 2500$
$= 1.00583 \times (25\,000) - 2500$
$= 22\,645.75$

Step 2: Calculate V_2, the value of the investment after 2 months.

$V_2 = 1.00583 \times V_1 - 2500$
$= 1.00583 \times (22\,645.75) - 2500$
$= 20\,277.774...$

Step 3: Calculate V_3, the value of the investment after 3 months.

$V_3 = 1.00583 \times V_2 - 2500$
$= 1.00583 \times (20\,277.774...) - 2500$
$= 17\,895.994...$

Step 4: Calculate V_4, the value of the investment after 4 months.

$V_4 = 1.00583 \times V_3 - 2500$
$= 1.00583 \times (17\,895.994...) - 2500$
$= 15\,500.327...$

Answer

\$15 500.33

Using amortisation tables to solve problems involving amortising annuities

Amortisation tables are used to keep track of the payments received by an investor and the remaining balance in an amortising annuity.

In an amortising annuity, a portion of each payment received by the investor consists of the interest calculated each compounding period, and the remainder is taken out of the investment balance.

Back to contents

For example, the amortisation table shown models the duration of an annuity.

payment number	payment	interest	principal reduction	balance of investment
0	0.00	0.00	0.00	10 000.00
1	2000.00	100.00	1900.00	8100.00
2	2000.00	81.00	1919.00	6181.00
3	2000.00	61.81	1938.19	4242.81
4	2000.00	42.43	1957.57	2285.24
5	2000.00	22.85	1977.15	308.09
6	311.17	3.08	308.09	0.00

Note: Often the final payment is adjusted so that the entire balance of the investment is paid out. The final payment amount does not equal the previous balance because there is interest that needs to be calculated.

The columns of an amortisation table for an amortising annuity can be calculated using the following formulas:

$$\textit{interest} = \frac{r}{100} \times \textit{previous investment balance}$$

$$\textit{principal reduction} = \textit{payment} - \textit{interest}$$

$$\textit{balance of investment} = \textit{previous investment balance} - \textit{principal reduction}$$

Worked example 2

Mike invests $5000 in an annuity account paying interest at 10% per annum. Mike wants to receive a payment of $2010 per year. This is modelled using the following amortisation table.

payment number	payment	interest	principal reduction	balance of investment
0	0.00	0.00	0.00	5000.00
1	2010.00	500.00	1510.00	3490.00
2	2010.00	349.00	1661.00	1829.00
3				0.00

Mike wants the final payment amount adjusted so that the annuity will be fully paid out after payment number 3.

a. Calculate the missing interest for payment number 3.

Explanation

Step 1: Identify the interest rate and previous loan balance.

$r = 10\%$ per compounding period

$\textit{previous loan balance} = 1829.00$

Step 2: Calculate the interest amount.

$\textit{interest} = \frac{10}{100} \times 1829.00$

$\textit{interest} = 182.90$

Answer

$182.90

b. Calculate the final payment that Mike will receive for this annuity.

Explanation

Calculate the payment amount for payment number 3.

The final payment will be the sum of the remaining investment balance and the interest calculated.

$$\begin{aligned}\textit{final payment} &= 1829.00 + 182.90 \\ &= 2011.90\end{aligned}$$

Continues →

Back to contents

Answer

$2011.90

c. How much interest has the account earned over three years?

Explanation

Step 1: Calculate the total amount received in payments for three years.

$$\begin{aligned} \textit{total amount} &= 2 \times 2010.00 + 1 \times 2011.90 \\ &= 6031.90 \end{aligned}$$

Step 2: Calculate the total reduction in principal over three years.

The investment was depleted after three payments, so the total principal reduction is the full $5000 investment.

Step 3: Calculate the total interest earned over three years.

This is the difference between the total amount received in payments and the total reduction in principal.

$$\begin{aligned} \textit{interest earned} &= 6031.90 - 5000.00 \\ &= 1031.90 \end{aligned}$$

Note: Adding the values in the 'interest' column can also be used to calculate the total interest earned for this period.

Answer

$1031.90

Using financial applications of technology to solve problems involving amortising annuities

The financial solver is a useful tool for solving problems involving amortising annuities.

The present value, PV, is always negative since the investor gives this amount to the financial institution.

The payment, PMT, is always positive since the investor receives payments, taken out of the investment.

The future value, FV, is always positive as it represents future payments that can be made to the investor.

A future value of 0 indicates that the annuity has been fully paid out.

Worked example 3

Chuck invests $400 000 in an annuity with a bank who offers an interest rate of 4% per annum, compounding monthly. The annuity pays out $5000 per month.

a. What will the annuity be worth after 6 months?

Explanation

Step 1: Determine the financial solver inputs.

N	6	(there are 6 months)
I(%)	4	(annual interest rate)
PV	−400 000	(this is negative because Chuck gives it to the bank)
PMT	5000	(this is positive because Chuck receives it from the bank)
FV		
PpY	12	(payments made monthly)
CpY	12	(interest compounds monthly)

Continues →

Back to contents

Step 2: Use the financial solver to solve for FV.

FV	377 815.849...

This is positive because it represents future payments that can be made to Chuck.

Answer

$377 815.85

b. After 6 months, the bank decreases the interest rate to 3.6% per annum, compounding monthly. Chuck has opted to keep receiving the same payment amount each month.

To ensure the annuity is fully paid out, the final payment will be lower.

In total, how many payments will it take for the annuity to reach a zero balance?

Use the rounded value found in part **a**.

Explanation

Step 1: Determine the financial solver inputs.

Chuck still has $377 815.85 invested with the bank when the interest rate changes. This is now represented as the PV of the annuity.

N		
I(%)	3.6	(annual interest rate)
PV	−377 815.85	(this is negative because the amount is still with the bank)
PMT	5000	(this is positive because Chuck receives this payment)
FV	0	(annuity to be fully paid out)
PpY	12	(payments made monthly)
CpY	12	(interest compounds monthly)

Step 2: Use the financial solver to solve for N.

N	85.820...

This means that after the first six months, there will be 85 full payments of $5000 and 1 final payment less than $5000 before the annuity reaches a balance of $0. Hence, there will be 86 payments after the interest rate is changed.

Step 3: Calculate the total number of payments.

$86 + 6 = 92$

Answer

92 payments

Exam question breakdown

VCAA 2016 Exam 1 Recursion and financial modelling Q24

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 $130 784.93 after five years.

The balance of the annuity will be $66 992.27 after ten years.

The monthly payment that Mai receives from the annuity is closest to

A. $1270 **B.** $1400 **C.** $1500 **D.** $2480 **E.** $3460

Continues →

Explanation

Step 1: Determine the financial solver inputs.

The value of Mai's annuity after 5 years is the present value because this is the earliest time point in the scenario.

The value of Mai's annuity after 10 years can subsequently be treated as the future value, 5 years later.

N	60	(there are 60 months in 5 years)
I(%)	5.2	(annual interest rate)
PV	−130 784.93	(this is negative because Mai invests it with the bank)
PMT		
FV	66 992.27	(this is positive because it represents future payments that can be made to Mai)
PpY	12	(payments made monthly)
CpY	12	(interest compounds monthly)

Step 2: Use the financial solver to solve for PMT.

PMT	1500.000...

This is positive because Mai receives it from the bank.

30% of students answered this question correctly.

23% of students incorrectly answered option E. This was caused by the incorrect input of figures into the financial solver, with the wrong positive and negative signs for the PV and FV. Students may have struggled with this question because of the unconventional method of modelling the value of a loan in the middle of its lifespan, rather than the start or end of its duration.

Answer

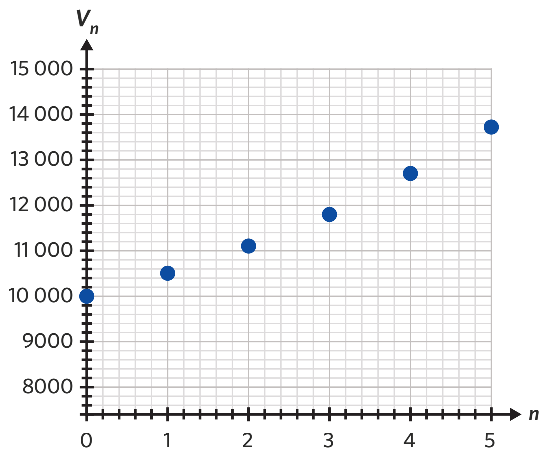
C

6D Questions

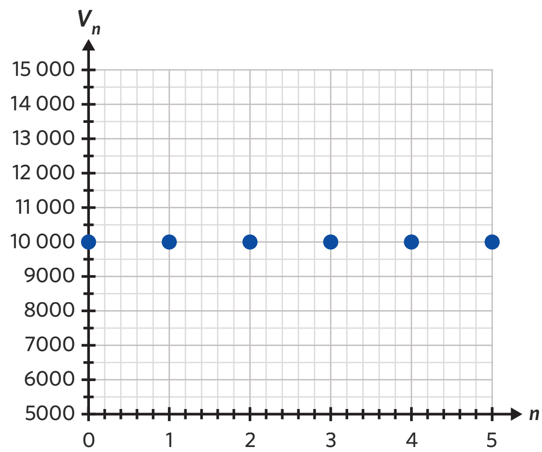
Using recurrence relations to model amortising annuities

1. Which of the following graphs is most likely to represent the value of an amortising annuity?

A.

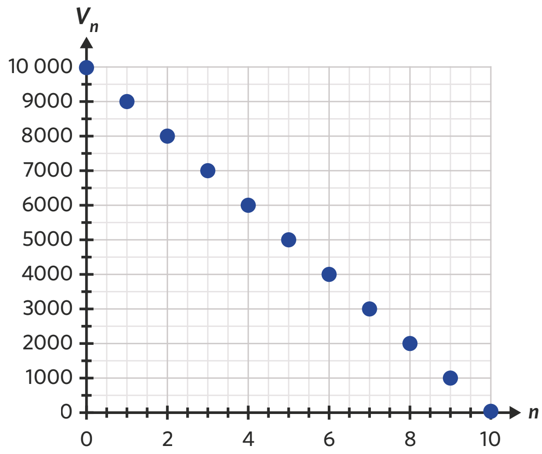


B.

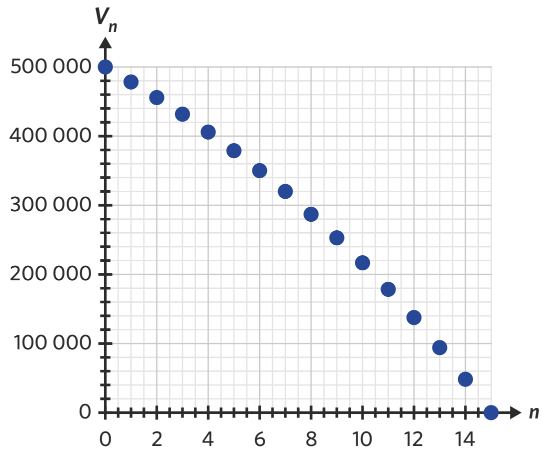


Back to contents

C.



D.



2. The value of an annuity V_n, after n months can be modelled using the recurrence relation

$V_0 = 25\,000, \quad V_{n+1} = 1.012 \times V_n - 1500$

a. How much was originally invested in this annuity?

b. How much does the investor receive as a payment from the investment every month?

c. What is the annual interest rate?

3. Carlos has invested in an annuity modelled by the recurrence relation

$V_0 = 50\,000, \quad V_{n+1} = 1.150 \times V_n - 10\,000$

a. Using recursive calculations, calculate the value of the annuity after Carlos has received four payments.

b. Carlos' friend, Gabby, takes out the same annuity, but receives payments of $9000 instead. What is the balance of her annuity after three payments?

4. Generate a recurrence relation to model the value of the annuity V_n, after n compounding periods for each of the following scenarios. If necessary, round values to 5 significant figures.

a. Amy invests $2500 in an annuity with an interest rate of 6.9% per annum, compounding monthly. The annuity provides her with a monthly income of $100.

b. Bella invests $6200 in an annuity with an interest rate of 4.1% per annum, compounding quarterly. The annuity provides her with a quarterly income of $250.

c. Cameron has an annuity with an interest rate of 5.0% per annum, compounding fortnightly. He initially invests $1500 and receives a fortnightly income of $60.

5. The balance of an annuity after n fortnights, B_n can be modelled by the recurrence relation

$B_0 = 168\,714, \quad B_{n+1} = 1.002 \times B_n - 14\,562$

Calculate the interest, in dollars, the annuity has earned after five payments.

Using amortisation tables to solve problems involving amortising annuities

6. The first four lines of the amortisation table for an annuity are shown.

payment number	payment	interest	principal reduction	balance of investment
0	0.00	0.00	0.00	14 521.00
1	954.00	145.21	808.79	13 712.21
2	954.00	137.12	816.88	12 895.33
3		128.95	825.05	12 070.28
4	954.00	120.70	833.30	11 236.98

The payment that the investor received for payment number 3 is

A. $954.00

B. $1000.00

C. $1150.44

D. $12 091.93

Back to contents

7. John invests his superannuation in an annuity. Interest compounds monthly and he receives a retirement income stream of $2000 each month.

An amortisation table that models the annuity is shown.

payment number	payment	interest	principal reduction	balance of investment
20	2000.00	482.26	1517.74	136 270.19
21	2000.00	476.95	1523.05	134 747.14
22	2000.00	471.61	1528.39	133 218.75
23	2000.00	466.27	1533.73	131 685.02
24	2000.00	460.90	1539.10	130 145.92
25	2000.00			

a. Show that the interest rate is 4.2% per annum.

For payment number 25, calculate

b. the amount of interest earned.

c. the reduction in principal.

d. the resultant balance of the annuity.

8. Find x, y and z in each of the following amortisation tables.

a.

payment number	payment	interest	principal reduction	balance of investment
0	x	0.00	0.00	400.00
1	50.00	y	10.00	390.00
2	50.00	39.00	11.00	z

b.

payment number	payment	interest	principal reduction	balance of investment
0	0	0	0	1600.00
1	150.00	64.00	x	y
2	150.00	z	89.44	1424.56

c.

payment number	payment	interest	principal reduction	balance of investment
0	0.00	0.00	0.00	5000.00
1	2000.00	120.00	1880.00	3120.00
2	2000.00	x	1925.12	1194.88
3	y	28.68	z	0.00

Using financial applications of technology to solve problems involving amortising annuities

9. Ashton's parents invested $25 000 in an annuity that pays out $5000 each year to cover his school expenses. After two years of high school, the annuity is worth $17 455.23. Interest compounds annually.

What is the annual interest rate for this annuity?

A. 5.0%

B. 5.3%

C. 6.0%

D. 6.1%

Back to contents

10. Find the current value of the following annuities.

a. A $20 000 annuity began three years ago and has been paying out $5000 per year. The annuity pays interest at a rate of 8% per annum.

b. A $1200 annuity opened six months ago, with an interest rate of 13.2% per annum, compounding monthly. It pays out $150 per month.

c. A $6500 annuity began one and a half years ago, with an interest rate of 8.8% per annum, compounding every three months, paying $400 each quarter.

d. Five years ago, $81 000 was invested in an annuity. It has an interest rate of 8.24% per annum, compounding weekly, and pays $250 weekly.

---

11. When Jack retires, he wants to have enough superannuation invested in an annuity to pay him $1000 a month for exactly 20 years. His superfund offers an annual interest rate of 5.2% compounding monthly. How much does Jack need to have in his super account before he can retire?

---

12. Dawn has invested $500 000 in an annuity. Her bank offers an interest rate of 4.7% per annum, compounding fortnightly. She will receive fortnightly payments of $1500.

a. How much interest does the annuity earn in the 1st year?

b. What will the annuity be worth in 4 years?

After 4 years, Dawn opts to have the payments increase to $2000 each fortnight for the remainder of the annuity. The interest rate and compounding frequency remains constant.

c. To ensure the annuity is fully paid out, the final payment will be lower.

In total, how many fortnights will the annuity last for?

Joining it all together

13. Dale invests his life savings in an annuity, where interest compounds fortnightly. A recurrence relation that models the value of the annuity, V_n, after n fortnights is shown.

$V_0 = 1\ 250\ 000, \quad V_{n+1} = 1.004 \times V_n - 8500$

a. How much did Dale invest?

b. Calculate the annual interest rate.

Dale's accountant, Randall, has created an amortisation table to model the annuity after n fortnights. However, due to an error with his fax machine, two lines didn't print.

The balance of the investment after payment 9 is $1 217 991.27.

payment number	payment	interest	principal reduction	balance of investment
10				
11				

c. Fill in the missing rows.

d. After 11 fortnights, how much interest has the account earned?

Use the rounded value found in part **c**.

Back to contents

14. Bobby has recently retired after 40 years of hard work. She has \$759 231 invested in an annuity with her superannuation fund who offers an interest rate of 8.9% per annum, compounding weekly. She will receive a weekly retirement income stream. Her grandchild has set up an amortisation table to model the annuity. The first five lines are shown.

payment number	payment	interest	principal reduction	balance of investment
0	0.00	0.00	0.00	759 231.00
1	2000.00	1299.45	700.55	758 530.45
2	2000.00	1298.25	701.75	757 828.70
3	2000.00	1297.05	702.95	757 125.75
4	2000.00	1295.85	704.15	756 421.60
5		1294.64	705.36	755 716.24

a. Determine the payment amount for payment number 5.

b. Construct the recurrence relation that can be used to model the value of the annuity, V_n after n weeks. Round to 4 decimal places where necessary.

After 2 years, the interest rate decreases to 7.4% p.a, compounding weekly.

c. To ensure the annuity is fully paid out, the final payment will be lower.

If everything else remains the same, how many more weeks will the annuity last?

---

15. The value of an annuity, V_n after n months can be modelled by the recurrence relation

$V_0 = 645\,628.00, \quad V_{n+1} = R \times V_n - 5478.47$

If the annuity is to be fully paid out in 10 years, calculate R, rounded to 4 decimal places.

---

16. Ella invests \$100 000 in an annuity. Her bank offers a 4.68% per annum interest rate, compounding fortnightly. She will receive payments of \$1250 every fortnight for 3 years.

a. Construct a recurrence relation that can be used to model the value of the annuity V_n after n fortnights. Round to 4 decimal places where necessary.

b. What is the value of the annuity after three years?

After three years, Ella decreases the fortnightly payments to \$900.

To ensure the annuity is fully paid out, the final payment will be lower.

c. How long, in years and fortnights, does the annuity last in total?
Use the rounded value found in part **b**.

Exam practice

17. The value of an annuity, V_n, after n monthly payments of \$555 have been made, can be determined using the recurrence relation

$V_0 = 100\,000, \quad V_{n+1} = 1.0025 \times V_n - 555$

The value of the annuity after five payments have been made is closest to

A. \$97 225

B. \$98 158

C. \$98 467

D. \$98 775

E. \$100 224

VCAA 2016 Exam 1 Recursion and financial modelling Q18

77% of students answered this question correctly.

Back to contents

18. Julie has retired from work and has received a superannuation payment of $492 800.

Julie could invest the $492 800 in an annuity.

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 would be $480 242.25.

What monthly payment, in dollars, would Julie receive? (1 MARK)

VCAA 2018 Exam 2 Recursion and financial modelling Q6bi

48% of students answered this question correctly.

---

19. Deepa invests $500 000 in an annuity that provides an annual payment of $44 970.55. Interest is calculated annually.

The first five lines of the amortisation table are shown.

payment number	payment	interest	principal reduction	balance of investment
0	0.00	0.00	0.00	500 000.00
1	44 970.55	20 000.00	24 970.55	475 029.45
2	44 970.55	19 001.18	25 969.37	449 060.08
3	44 970.55	17 962.40	27 008.15	422 051.93
4	44 970.55	16 882.08	28 088.47	393 963.46

The number of years, in total, for which Deepa will receive the regular payment of $44 970.55 is closest to

A. 12 **B.** 15 **C.** 16
D. 18 **E.** 20

VCAA 2021 Exam 1 Recursion and financial modelling Q19

44% of students answered this question correctly.

---

20. Sienna invests $152 431 into an annuity from which she will receive a regular monthly payment of $900 for 25 years. The interest rate for this annuity is 5.1% per annum, compounding monthly.

a. Let V_n be the balance of the annuity after n monthly payments. A recurrence relation written in terms of V_0, V_{n+1} and V_n can model the value of this annuity from month to month.

Showing recursive calculations, determine the value of the annuity after two months. Round the answer to the nearest cent. (2 MARKS)

b. After two years, the interest rate for this annuity will fall to 4.6%.

To ensure that she will still receive the same number of $900 monthly payments, Sienna will add an extra one-off amount into the annuity at this time.

Determine the value of this extra amount that Sienna will add.

Round the answer to the nearest cent. (1 MARK)

VCAA 2021 Exam 2 Recursion and financial modelling Q9

Part **a**: The average mark on this question was **0.7**.
Part **b**: **8%** of students answered this question correctly.

Questions from multiple lessons

Recursion and financial modelling

21. Jessie took out a $15 000 loan to renovate her apartment. Interest for the loan was charged at a rate of 13.8% per annum, compounding quarterly.

For the first two years of the loan, Jessie made quarterly repayments of $800.

For the next two years of the loan, Jessie made quarterly repayments of $1000.

The amount Jessie paid in interest only during this four-year period is closest to

A. $6300 **B.** $6700 **C.** $7100 **D.** $7300 **E.** $7700

Adapted from VCAA 2018NH Exam 1 Recursion and financial modelling Q24

Back to contents

Data analysis

22. The population of a bee farm was monitored over a period of time.

The time series plot shows the *population* of the farm, in thousands, over a thirty-month period.

The five-median smoothed population in thousands, for the 15th month is closest to

A. 79

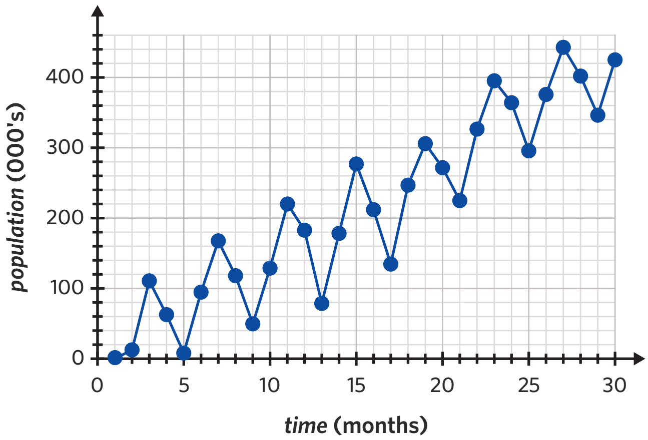
B. 135

C. 178

D. 212

E. 277

Adapted from VCAA 2017 Exam 1 Data analysis Q14



Recursion and financial modelling

23. Igor has just purchased a new jet ski for \$22 000.

The value of his jet ski depreciates using the reducing balance method.

The value of Igor's jet ski, in dollars, after n years, J_n, can be modelled by the recurrence relation shown.

$J_0 = 22\,000, \quad J_{n+1} = R \times J_n$

a. For each of the first two years of the reducing balance depreciation, the value of R is 0.75. What is the annual rate of depreciation during the first two years? (1 MARK)

b. For the next three years of reducing balance depreciation, the annual rate of depreciation is changed to 15.9%. What is the value of the jet ski five years after it was purchased? Round the answer to the nearest cent. (2 MARKS)

Adapted from VCAA 2018 Exam 2 Recursion and financial modelling Q5

Back to contents

6E Perpetuities

STUDY DESIGN DOT POINT

- simple perpetuity as a special case of an annuity that lasts indefinitely

6A 6B 6C 6D 6E 6F

KEY SKILLS

During this lesson, you will be:

- using recurrence relations to model perpetuities
- using financial applications of technology to solve problems involving perpetuities.

KEY TERM

- Perpetuity

An investment can last indefinitely if the investor receives a payment each period that is equal to the interest earned. This type of investment could be useful for an investor who wants a regular stream of income over a long period of time, such as a retiree who has received their superannuation.

Using recurrence relations to model perpetuities

A **perpetuity** is a special type of annuity where the payment received is equal to the interest earned, effectively allowing the annuity to last indefinitely. The value of the perpetuity remains the same for each compounding period.

See worked example 1

The value of a perpetuity, V_n, after n compounding periods, can be modelled by a recurrence relation of the same form as an annuity.

$V_0 = principal, \quad V_{n+1} = R \times V_n - d$, where

- $R = 1 + \frac{r}{100}$, where r is the interest rate (%) per compounding period
- d is the payment received per compounding period

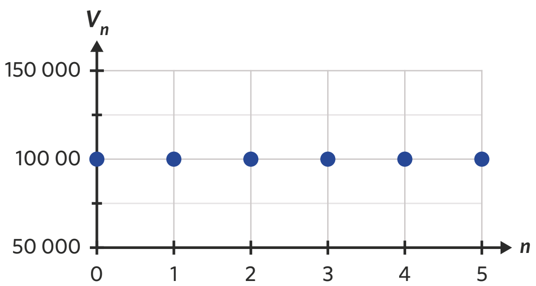
However, the payment is equal to the interest earned each compounding period, and can therefore be calculated as

$d = \frac{r}{100} \times V_0$

For example, \$100 000 is invested in a perpetuity with an interest rate of 4.2% per annum, compounding annually. The interest amount is 4.2% of \$100 000, which is \$4200. If the investor receives a payment of \$4200 at the end of each year, the principal balance will remain constant at \$100 000. A recurrence relation that can be used to model the value of the investment, V_n, after n years is

$V_0 = 100\,000, \quad V_{n+1} = 1.042 \times V_n - 4200.$

The balance of the perpetuity over a period of 5 years can also be represented on a graph. Notice that it is constant.



Back to contents

Perpetuities can also be modelled using a table, similar to an amortisation table. However, there will be no change with each payment because the principal balance remains constant.

See worked example 2

Worked example 1

Elise invests $259 920 in a perpetuity. Her bank offers an interest rate of 4.8% per annum, compounding monthly.

a. What is the monthly amount she will receive from the bank?

Explanation

Step 1: Determine the principal and interest rate for the perpetuity.

$V_0 = 259\,920$

$r = \frac{4.8}{12}\%$ per compounding period

Step 2: Calculate d.

$$d = \frac{4.8}{12 \times 100} \times 259\,920$$
$$= \frac{4.8}{1200} \times 259\,920$$
$$= 1039.68$$

Answer

$1039.68

b. Construct a recurrence relation that can be used to model the value of Elise's investment, V_n, after n months.

Explanation

Step 1: Determine the principal, interest rate and payment amount.

$V_0 = 259\,920$

$r = \frac{4.8}{12}\%$ per compounding period

$d = 1039.68$

Step 2: Calculate R.

$$R = 1 + \frac{4.8}{12 \times 100}$$
$$= 1 + \frac{4.8}{1200}$$
$$= 1.004$$

Step 3: Construct the recurrence relation.

Answer

$V_0 = 259\,920, \quad V_{n+1} = 1.004 \times V_n - 1039.68$

Worked example 2

Sona has $490 000 invested in a perpetuity account. Interest is earned at a rate of 1.04% per annum, compounding weekly. This is modelled in the following table.

payment number	payment	interest	principal reduction	balance of investment
10	98.00	98.00	0.00	490 000.00
11	98.00	98.00	0.00	490 000.00
12	98.00	98.00	0.00	490 000.00
13	98.00	98.00	0.00	490 000.00
14	98.00			

Fill in the missing cells for payment number 14.

Explanation

Step 1: Determine the interest for payment number 14.

The interest amount is equal to the payment received, which is $98.00

Continues →

Back to contents

Step 2: Determine the principal reduction for payment number 14.

The value of a perpetuity remains constant, so there is no reduction of the principal.

Step 3: Determine the loan balance after payment number 14.

The balance doesn't decrease in a perpetuity, so it will remain at $490 000.00.

Answer

payment number	payment	interest	principal reduction	balance of investment
10	98.00	98.00	0.00	490 000.00
11	98.00	98.00	0.00	490 000.00
12	98.00	98.00	0.00	490 000.00
13	98.00	98.00	0.00	490 000.00
14	98.00	98.00	0.00	490 000.00

Using financial applications of technology to solve problems involving perpetuities

The financial solver can be used to solve problems involving perpetuities.

The number of payments, N, can be set to any value since the value of the perpetuity remains constant.

The present value, PV, is always negative since the investor deposits the initial investment.

The payment, PMT, is always positive since the investor receives payments, taken out of the investment.

The future value, FV, is always positive since it represents a future payment that will be made to the investor. It is also the positive value of the PV.

The financial solver can be used to calculate an unknown interest rate I(%) or payment amount PMT, but not an unknown principal value, since both the PV and FV would be unknown.

Worked example 3

May invests $123 201 in a perpetuity. Her bank offers an interest rate of 1.3% per annum, compounding fortnightly.

a. How much will May receive each fortnight?

Explanation

Step 1: Determine the financial solver inputs.

N	1	(can be any value since the balance is unchanged)
I(%)	1.3	(annual interest rate)
PV	−123 201	(this is negative because May invests it with the bank)
PMT		
FV	123 201	(this is positive because it represents a future payment that will be made to May if she closes the perpetuity)
PpY	26	(payments made fortnightly)
CpY	26	(interest compounds fortnightly)

Continues →

Back to contents

Step 2: Use the financial solver to solve for PMT.

PMT	61.60

This is positive because May receives it from the bank.

Answer

$61.60

b. May decides that once she receives a cumulative total of $6400, she will close the perpetuity account. For how many fortnights will the account be open?

Explanation

Step 1: Recall the payment amount made each fortnight.

$payment = 61.60$

Step 2: Calculate the number of fortnights to receive a cumulative total of $6400.

$\frac{6400}{61.40} = 103.896...$

This must be rounded up. At 103 fortnights, May will not have exceeded the $6400 threshold.

Answer

104 fortnights

Exam question breakdown

VCAA 2018 Exam 2 Recursion and financial modelling Q6a

Julie has retired from work and has received a superannuation payment of $492 800.

Julie could invest the $492 800 in a perpetuity. She would then receive $887.04 each fortnight for the rest of her life.

At what annual percentage rate is interest earned by this perpetuity? (1 MARK)

Explanation

Step 1: Determine the financial solver inputs.

N	1	(can be any value since the balance is unchanged)
I(%)		
PV	−492 800	(this is negative because Julie invests it with the bank)
PMT	887.04	(this is positive because Julie receives it from the bank)
FV	492 000	(this is positive because it represents a future payment that will be made to Julie once she closes the perpetuity)
PpY	26	(payments made fortnightly)
CpY	26	(interest compounds fortnightly)

Step 2: Use the financial solver to solve for I(%).

I(%)	4.68

This is the annual interest rate.

Answer

4.68% p.a.

41% of students answered this question correctly.

A number of students incorrectly answered **0.18%**, the fortnightly interest rate. This does not account for the fortnightly payment frequency.

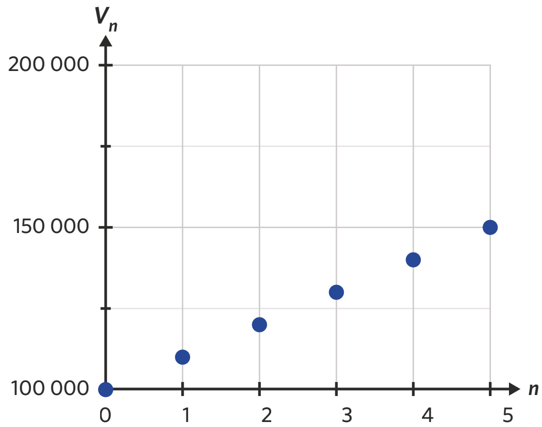
Back to contents

6E Questions

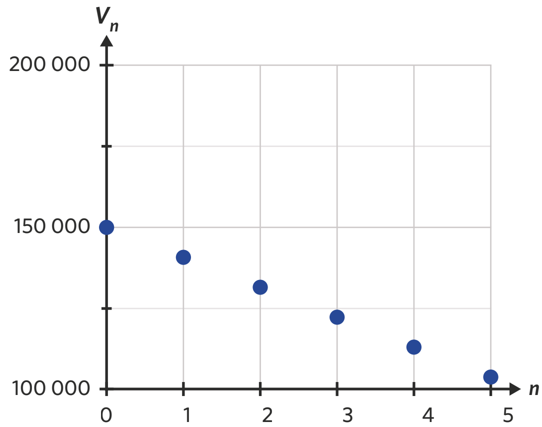
Using recurrence relations to model perpetuities

1. Which of the following graphs is most likely to represent the value of a perpetuity after 5 years?

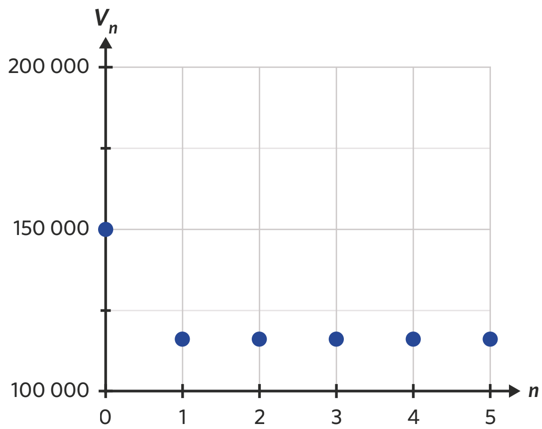
A.



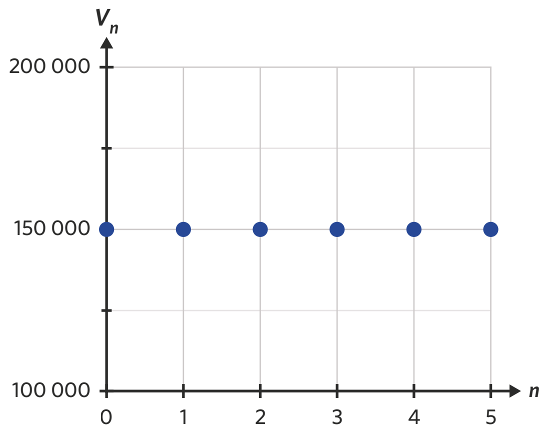
B.



C.



D.



2. The following working calculates the monthly payment from a perpetuity investment.

$$d = \frac{r}{100} \times V_0$$
$$= \frac{0.3}{100} \times 90\,000$$
$$= 270$$

a. What is the value of the initial investment?

b. What is the monthly payment?

c. What is the annual interest rate?

d. Construct a recurrence relation that can be used to model the value of the perpetuity, V_n, after n compounding periods.

3. The following recurrence relation models the value of a perpetuity, V_n, after n months.

$V_0 = 168\,429, \quad V_{n+1} = 1.005 \times V_n - 842.15$

a. How much does the investor receive each month?

b. Calculate the annual interest rate.

c. What is the value of the perpetuity after 100 months?

Back to contents

4. Answer the following questions without using a financial solver. Round to two decimal places where necessary.
   a. A perpetuity with an initial investment of $80 000 has an interest rate of 3.5% per annum, compounding monthly. How much are the monthly payments?
   b. A perpetuity with an initial investment of $250 000 has an interest rate of 2.8% per annum, compounding quarterly. How much are the quarterly payments?
   c. A perpetuity has a monthly payment of $225 to the investor. If the interest rate is 3.4% per annum, how much was the initial investment?
   d. An investor puts $145 000 into a perpetuity. They receive weekly payments of $200. If interest compounds weekly, what is the annual interest rate?

5. Solt has a perpetuity that pays $2153.85 on a fortnightly basis. Interest is calculated at a rate of 4% per annum, compounding fortnightly.
   a. Calculate the amount that Solt invested.
   b. Construct a recurrence relation that can be used to model the value of this perpetuity, V_n, after n fortnights. Round R to 4 decimal places.

6. Peppa has invested her superannuation in a perpetuity. Her super fund offers an interest rate of 7.2% per annum, compounding monthly. She receives a retirement income stream of $4525.62 each month.

   The following table can be used to model this investment over a number of periods.

payment number	payment	interest	principal reduction	balance of investment
20	4525.62			
21				

   a. Calculate the amount that Peppa invested with her super fund.
   b. Fill in the table.

Using financial applications of technology to solve problems involving perpetuities

7. The annual interest rate required to establish a perpetuity with a balance of $64 482 and fortnightly payments of $128.96 is closest to

   A. 0.2% p.a.   B. 2.0% p.a.   C. 5.2% p.a.   D. 6.7% p.a.

8. A perpetuity with an initial investment of $28 500 has an interest rate of 3.16% per annum, compounding monthly. What is the value of the monthly payments?

9. The initial investment in a perpetuity was $1 287 995. The perpetuity pays $1451.27 per fortnight. The interest is compounded fortnightly. What is the perpetuity's annual interest rate, correct to three significant figures?

10. A perpetuity had an initial investment of $323 185. If the interest rate is 2.98% per annum, and interest is compounded quarterly, how much are the quarterly payments?

11. Farah won $236 064 in a lottery and decided to invest it in a perpetuity. She invests 64% of her total lottery winnings. Interest is compounded monthly and she will receive a monthly income stream.
    a. How much was her initial investment?
    b. The annual interest rate of the perpetuity is 3.41% p.a. What are the monthly payments?

    Farah closes the perpetuity after two and a half years.

    c. What is the total interest she received from the perpetuity?

Back to contents

Joining it all together

12. Rebecca has $245 105 invested in a perpetuity. Her bank offers an interest rate of 4.8% per annum, compounding quarterly.

a. How much does Rebecca receive each quarter from the perpetuity?

b. Construct a recurrence relation that can be used to model the value of Rebecca's annuity, V_n, after n quarters.

c. Calculate the value of the perpertuity after three years.

13. The following table shows the value of a perpetuity over a period of time.

payment number	payment	interest	principal reduction	balance of investment
0	0.00	0.00	0.00	135 560.00
1	271.12	271.12	0.00	135 560.00
2	271.12	271.12	0.00	135 560.00

a. Calculate the interest rate per compounding period for this investment.

b. Construct a recurrence relation that can be used to model the value of the annuity, V_n, over n periods.

c. Construct a graph that displays the value of this perpetuity over the first 5 compounding periods.

14. Hart wants to have enough superannuation so that, once invested in a perpetuity, he will receive a retirement income stream of $736.44 per fortnight. His super fund offers an interest rate of 4.2% per annum, compounding fortnightly.

a. How much superannuation does Hart need to have before he can retire?

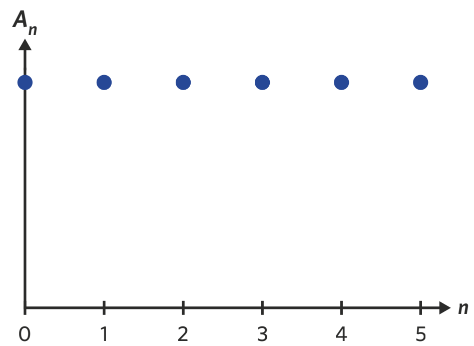
Seven years after Hart retired, his super fund decreased its interest rate to 2.7% per annum, compounding fortnightly.

b. By how much will the fortnightly payments decrease as a result? Use the rounded value found in part **a**.

c. How much interest will Hart have received 20 years after establishing the perpetuity?

Exam practice

15. The following graph represents the value of an annuity investment, A_n, in dollars, after n time periods.



A recurrence relation that could match this graphical representation is

A. $A_0 = 200\,000, \quad A_{n+1} = 1.015 \times A_n - 2500$

B. $A_0 = 200\,000, \quad A_{n+1} = 1.025 \times A_n - 5000$

C. $A_0 = 200\,000, \quad A_{n+1} = 1.03 \times A_n - 5500$

D. $A_0 = 200\,000, \quad A_{n+1} = 1.04 \times A_n - 6000$

E. $A_0 = 200\,000, \quad A_{n+1} = 1.05 \times A_n - 8000$

VCAA 2020 Exam 1 Recursion and financial modelling Q25

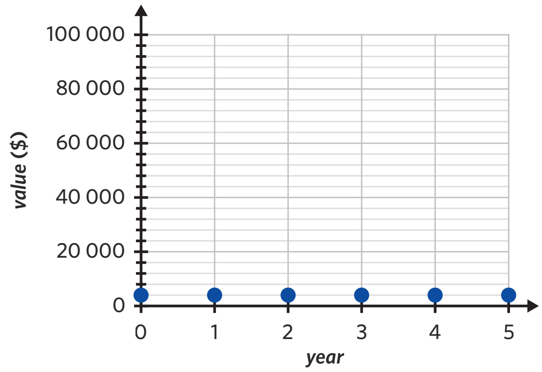
79% of students answered this question correctly.

Back to contents

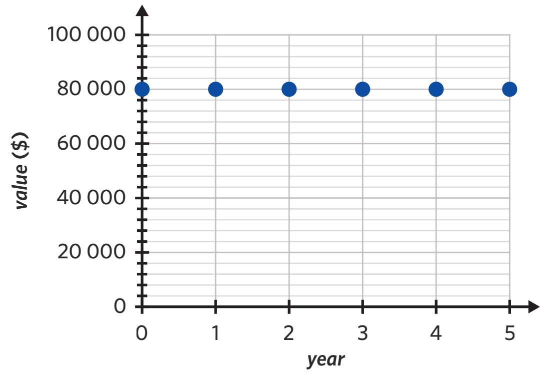
16. Juanita invests \$80 000 in a perpetuity that will provide \$4000 per year to fund a scholarship at a university.

The graph that shows the value of this perpetuity over a period of five years is

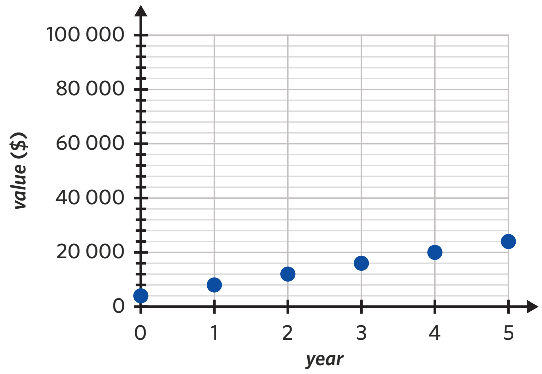
A.



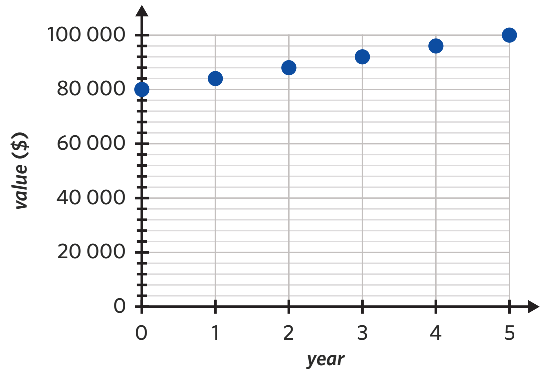
B.



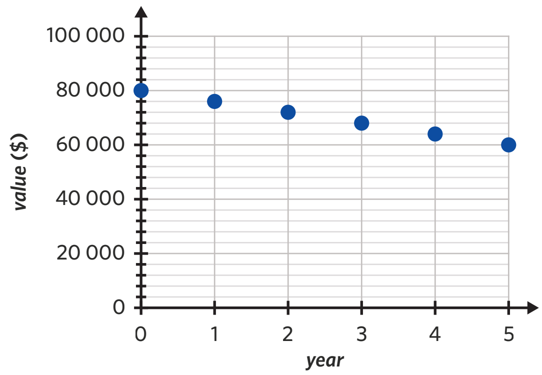
C.



D.



E.



VCAA 2016 Exam 1 Recursion and financial modelling Q21

50% of students answered this question correctly.

---

17. Sienna invests \$420 000 in a perpetuity from which she will receive a regular monthly payment of \$1890.

The perpetuity earns interest at the rate of 5.4% per annum.

a. Determine the total amount, in dollars, that Sienna will receive after one year of monthly payments. (1 MARK)

b. Write down the value of the perpetuity after Sienna has received one year of monthly payments. (1 MARK)

c. Let S_n be the value of Sienna's perpetuity after n months.

Complete the recurrence relation, in terms of S_0, S_{n+1} and S_n, that would model the value of the perpetuity over time. (1 MARK)

$$S_0 = \boxed{}, \quad S_{n+1} = \boxed{} \times S_n - 1890$$

VCAA 2021 Exam 2 Recursion and financial modelling Q6

Part **a**: **59%** of students answered this question correctly.
Part **b**: **47%** of students answered this question correctly.
Part **c**: **38%** of students answered this question correctly.

Back to contents

Questions from multiple lessons

Recursion and financial modelling

18. Josh has a personal loan with a present value of $12 356.89 that he took out to fund his round-the-world gap year trip.

The interest rate for Josh's loan is 15.7% per annum, compounding quarterly.

His quarterly repayment is $800.

The loan is to be fully repaid after six years.

Josh knows that the loan cannot be exactly repaid with 24 repayments of $800.

To solve this problem, Josh will make 23 payments of $800. He will then adjust the value of the final repayment so that the loan is fully repaid on the 24th payment.

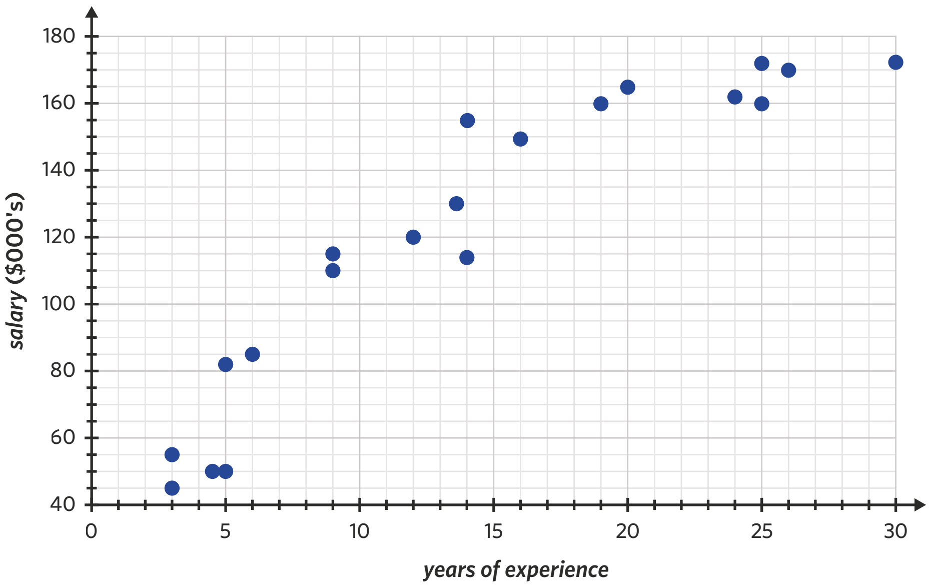
The value of the 24th payment is closest to

A. $127.64 **B.** $164.05 **C.** $635.95 **D.** $927.64 **E.** $964.05

Adapted from VCAA 2018 Exam 1 Recursion and financial modelling Q22

Data analysis

19. The following table provides information on *years of experience* and *salary* ($000's) of 20 workers at a large company. A scatterplot of the data is also shown.



years of experience	*salary* ($000's)
3	45
4.5	50
5	50
3	55
9	110
5	81
6	85
9	115
14	115
12	120
13.5	130
16	150
14	155
19	160
25	160
24	162
20	165
26	170
30	171
25	171

A log transformation is applied to the variable *years of experience* to linearise the scatterplot.

With *salary* as the response variable, the equation of the least squares line fitted to the linearised data is closest to

A. $salary = -22.98 + 136.82 \times years\ of\ experience$

B. $\log_{10}(salary) = -22.98 + 136.82 \times years\ of\ experience$

C. $salary = -22.98 + 136.82 \times \log_{10}(years\ of\ experience)$

D. $\log_{10}(years\ of\ experience) = -61.93 + 37.22 \times salary$

E. $salary = -61.93 + 37.22 \times \log_{10}(years\ of\ experience)$

Adapted from VCAA 2016 Exam 1 Data analysis Q11

Back to contents

Recursion and financial modelling

20. Denzel borrows $35 000 to buy his girlfriend a car.

He will be charged interest at 5.2% per annum, compounding monthly.

Denzel makes monthly repayments of $400 to repay his loan.

a. How much money does Denzel owe after the first 24 payments? (1 MARK)

b. What is the total interest that Denzel will have paid after 12 repayments? (1 MARK)

After four years, Denzel makes a lump sum payment of $$L$ to reduce the balance of the loan.
This payment will ensure that his loan is fully repaid in a further two years.

The repayment amount remains at $400 per month, and the interest rate remains at 5.2% per annum, compounding monthly.

c. What is the value of Denzel's lump-sum payment, $$L$? Round to the nearest dollar. (2 MARKS)

Adapted from VCAA 2016 Exam 2 Recursion and financial modelling Q7

Back to contents

6F Annuity investments

STUDY DESIGN DOT POINTS

- use of a first-order linear recurrence relation to model and analyse (numerically and graphically) annuity investments, 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

6A 6B 6C 6D 6E **6F**

KEY SKILLS

During this lesson, you will be:

- using recurrence relations to model annuity investments
- using tables to solve problems involving annuity investments
- using financial applications of technology to solve problems involving annuity investments.

KEY TERMS

- Annuity investment

An annuity investment occurs when an investor puts away a sum of money into an account that earns compound interest and makes regular additional payments to increase the principal balance. Due to these additional payments, annuity investments tend to grow faster than accounts that only earn compound interest. As such, annuity investments are particularly useful for an investor who wishes to have a sum of money stored away for use later in life.

Using recurrence relations to model annuity investments

An **annuity investment** is a special type of annuity where the investor deposits a fixed sum of money every compounding period in addition to the interest earned on the account. This allows the investment to increase in value. At this point in time, money will not be paid out to the investor until they close the account.

The value of an annuity investment, V_n, after n compounding periods, can be modelled by a recurrence relation of the form

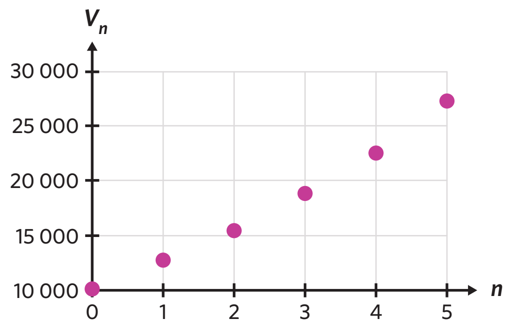
$V_0 = principal, \quad V_{n+1} = R \times V_n + d$, where

- $R = 1 + \frac{r}{100}$, where r is the interest rate (%) per compounding period
- d is the payment made per compounding period

For example, John initially deposited \$10 000 in an annuity that earns 15% interest per annum, compounding annually. He deposits an additional \$1000 to the account at the end of each year. The recurrence relation that can be used to determine V_n, the value of the investment after n years is

$V_0 = 10\ 000, \quad V_{n+1} = 1.15 \times V_n + 1000.$

The balance of the investment over a period of 5 years can also be represented by the following graph.



There is a slight curve in the graph. This is because, as the investment grows, the amount of interest earned increases.

Worked example 1

An annuity investment currently has a balance of \$3600 and earns interest at a rate of 3.6% per annum, compounding monthly. The investor adds an additional \$500 into the account at the end of each month.

a. Construct a recurrence relation that can be used to model the value of the investment, V_n, after n months.

Explanation

Step 1: Determine the principal, interest rate and payment amount.

$V_0 = 3600$

$r = \frac{3.6}{12}\%$ per compounding period

$d = 500$

Step 2: Calculate R.

$$R = 1 + \frac{3.6}{12 \times 100}$$
$$= 1 + \frac{3.6}{1200}$$
$$= 1.003$$

Step 3: Construct the recurrence relation.

Answer

$V_0 = 3600, \quad V_{n+1} = 1.003 \times V_n + 500$

b. How much will the account be worth after three months?

Explanation

Step 1: Calculate V_1, the value of the investment after 1 month.

$V_1 = 1.003 \times V_0 + 500$

$V_1 = 1.003 \times 3600 + 500$

$V_1 = 4110.80$

Step 2: Calculate V_2, the value of the investment after 2 months.

$V_2 = 1.003 \times V_1 + 500$

$V_2 = 1.003 \times 4110.80 + 500$

$V_2 = 4623.132...$

Step 3: Calculate V_3, the value of the investment after 3 months.

$V_3 = 1.003 \times V_2 + 500$

$V_3 = 1.003 \times 4623.132... + 500$

$V_3 = 5137.001...$

Answer

\$5137.00

Back to contents

Using tables to solve problems involving annuity investments

Tables similar to amortisation tables can be used to keep track of the payments added by the investor, and the resultant balance in an annuity investment.

In an annuity investment, the deposited payments and interest earned both contribute to an increase in the principal balance.

The following table models an annuity investment over a period of five years.

payment number	payment	interest	principal addition	balance of investment
0	0.00	0.00	0.00	10 000.00
1	1000.00	1500.00	2500.00	12 500.00
2	1000.00	1875.00	2875.00	15 375.00
3	1000.00	2306.25	3306.25	18 681.25
4	1000.00	2802.19	3802.19	22 483.44
5	1000.00	3372.52	4372.52	26 855.95

The columns of a table for an annuity investment can be calculated using the following formulas:

$\textit{interest} = \frac{r}{100} \times \textit{previous investment balance}$

$\textit{principal addition} = \textit{payment} + \textit{interest}$

$\textit{balance of investment} = \textit{previous investment balance} + \textit{principal addition}$

Worked example 2

The table shown represents an annuity investment with an initial deposit of $5000 and an interest rate of 5.2% per annum, compounding quarterly. Regular deposits of $200 are made at the end of each quarter.

payment number	payment	interest	principal addition	balance of investment
0	0.00	0.00	0.00	5000.00
1	**a.**	65.00	265.00	5265.00
2	200.00	68.45	**b.**	5533.45
3	200.00	71.93	271.93	**c.**
4	200.00	**d.**	275.47	6080.85

There are a few cells missing.

a. Determine the missing payment for payment number 1.

Explanation

Determine the payment made to the annuity investment.

Regular deposits of $200 are made at the end of each quarter.

$\textit{payment} = 200$

Answer

$200

Continues →

b. Calculate the missing principal addition for payment number 2.

Explanation

Step 1: Determine the payment and interest.

$payment = 200.00$

$interest = 68.45$

Step 2: Calculate the principal addition.

$$\begin{aligned} principal\ addition &= payment + interest \\ &= 200.00 + 68.45 \\ &= 268.45 \end{aligned}$$

Answer

$268.45

c. Calculate the missing investment balance after payment number 3.

Explanation

Step 1: Determine the principal addition and previous investment balance.

$previous\ investment\ balance = 5533.45$

$principal\ addition = 271.93$

Step 2: Calculate the investment balance.

$$\begin{aligned} &balance\ of\ investment \\ &= principal\ addition + previous\ investment\ balance \\ &= 5533.45 + 271.93 \\ &= 5805.38 \end{aligned}$$

Answer

$5805.38

d. Calculate the missing interest for payment number 4.

Explanation

Step 1: Determine the interest rate and previous investment balance.

$r = \frac{5.2}{4}\%$ per compounding period

$previous\ investment\ balance = 5805.38$

Step 2: Calculate the interest.

$$\begin{aligned} interest &= \frac{r}{100} \times previous\ investment\ balance \\ &= \frac{5.2}{4 \times 100} \times 5805.38 \\ &= \frac{5.2}{400} \times 5805.38 \\ &= 75.469... \end{aligned}$$

Answer

$75.47

Using financial applications of technology to solve problems involving annuity investments

The financial solver is a useful tool that can be used to solve problems involving annuity investments.

The present value, PV, is always negative since the investor gives this amount to the bank for investing.

The payment, PMT, is always negative since the investor deposits this amount with the bank every compounding period.

The future value, FV, is always positive as it represents a future payment that will be made to the investor once they close their investment.

Back to contents

Worked example 3

Steve is saving up for a cruise holiday by investing \$10 000 in an account that earns interest at a rate of 3.90% per annum, compounding monthly. He deposits \$500 at the end of each month.

a. After 5 years, how much is in Steve's account?

Explanation

Step 1: Determine the financial solver inputs.

N	60	(there are 60 months in 5 years)
I(%)	3.90	(annual interest rate)
PV	−10 000	(this is negative because Steve invests it with the bank)
PMT	−500	(this is negative because Steve deposits it into the account)
FV		
PpY	12	(payments made monthly)
CpY	12	(interest compounds monthly)

Step 2: Use the financial solver to solve for FV.

FV	45 214.925...

This is positive because it represents a future payment that will be made to Steve once he closes the annuity.

Answer

\$45 214.93

b. How much interest has the account earned after 5 years?

Explanation

Step 1: Calculate the growth of the investment after 5 years.

This growth is the difference between the present value and initial value.

From part **a**, Steve's account is worth \$45 214.93.

$$\begin{aligned} \textit{growth} &= 45\ 214.93 - 10\ 000 \\ &= 35\ 214.93 \end{aligned}$$

Step 2: Calculate the total amount made in payments during the 5 year period.

$$\begin{aligned} \textit{total amount} &= 500 \times 60 \\ &= 30\ 000 \end{aligned}$$

Step 3: Calculate the interest.

The interest is the difference between the principal growth and the total amount made in payments.

$$\begin{aligned} \textit{interest} &= 35\ 214.93 - 30\ 000 \\ &= 5214.93 \end{aligned}$$

Answer

\$5214.93

Back to contents

6F QUESTIONS

Exam question breakdown

VCAA 2016 Exam 1 Recursion and financial modelling Q23

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

Explanation

Step 1: Determine the financial solver inputs.

N	8	(there are 8 quarters in 2 years)
I(%)	3.9	(annual interest rate)
PV	−5000	(this is negative because Sarah invests it with the bank)
PMT	−200	(this is negative because Sarah deposits it into the account)
FV		
PpY	4	(payments made quarterly)
CpY	4	(interest compounds quarterly)

Step 2: Use the financial solver to solve for FV.

FV	7059.249...

This is positive because it represents a future payment that will be made to Sarah once she closes the annuity.

58% of students answered this question correctly.

13% of students who chose A did not consider the quarterly compounding periods in their calculations. **12%** of students who chose B calculated the value of the investment as if it was not earning any interest. **11%** of students who chose C would have incorrectly constructed an n^{th} term rule to perform the calculations.

Answer

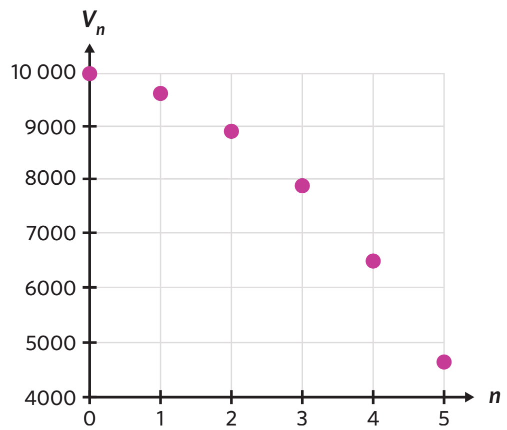
D

6F Questions

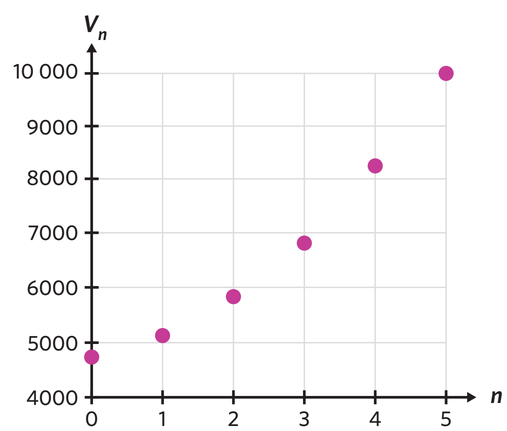
Using recurrence relations to model annuity investments

1. Which of the following graphs is most likely to represent the value of an annuity investment after 5 compounding periods?

A.

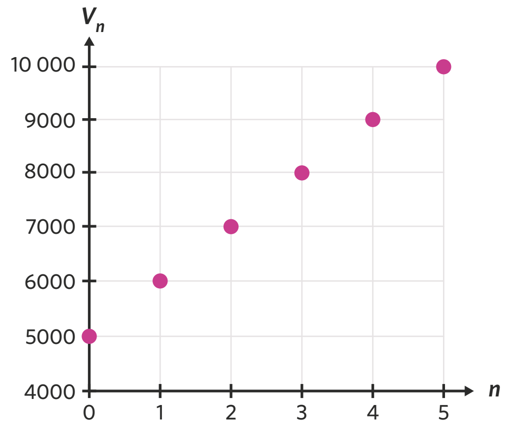


B.

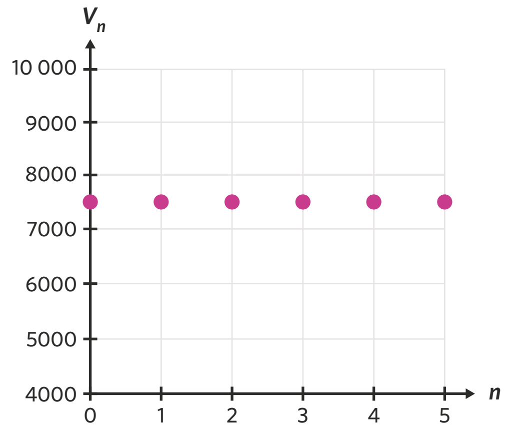


Back to contents

C.



D.



2. Paula's annuity investment can be modelled by the following recurrence relation, where V_n is the value of the account after n months.

$V_0 = 6500, \quad V_{n+1} = 1.0035 \times V_n + 180$

a. How much was the initial investment?

b. How much is Paula depositing into the account every month?

c. What is the annual interest rate?

3. Write a recurrence relation that can be used to model the value of the annuity investments V_n, after n compounding periods, for each of the following scenarios.

a. A principal of \$2000 with a weekly deposit of \$10 made, and an interest rate of 5.2% per annum, compounding weekly.

b. An annuity investment with an annual interest rate of 4.2% per annum, compounding monthly, with an initial deposit of \$500 and additional monthly payments of \$40.

4. Phil wants to predict how much money he will have in his savings account each month. He uses the following recurrence relation to model his account balance from month to month.

$V_0 = 8000, \quad V_{n+1} = 1.0025 \times V_n + 230$

a. Calculate the annual interest rate for this annuity investment.

b. How much would Phil have in his savings account after two months?

c. How many months will it take for Phil's account balance to first reach \$9000?

Using tables to solve problems involving annuity investments

5. The following table can be used to model the value of an annuity investment over a period of five months.

payment number	payment	interest	principal addition	balance of investment
0	0.00	0.00	0.00	15 000.00
1	1000.00	65.00	1065.00	16 065.00
2	1000.00	69.62		17 134.62
3	1000.00	74.25	1074.25	18 208.87
4	1000.00	78.91	1078.91	19 287.78
5	1000.00	83.58	1083.58	20 371.36

The principal addition after the second payment is

A. \$0.00 **B.** \$1065.00 **C.** \$1069.62 **D.** \$1074.25

6. Sheareen has constructed a table to predict the value of her savings for the next six months. Interest compounds monthly and payments are scheduled monthly.

payment number	payment	interest	principal addition	balance of investment
0	0.00	0.00	0.00	100 000.000
1	2500.00	340.00	2840.00	102 840.00
2		349.66	2849.66	105 689.66
3	2500.00	359.34		108 549.00
4	2500.00		2869.07	111 418.07
5	2500.00	378.82	2878.82	
6	2500.00	388.61	2888.61	117 185.50

There are a number of cells missing.

a. Calculate the missing payment value for payment number 2.

b. Calculate the missing principal addition after payment number 3.

c. Calculate the missing interest for payment number 4.

d. Calculate the missing investment balance after payment number 5.

e. Calculate the annual interest rate for this account. Round to three significant figures.

7. An annuity investment has an initial principal of \$5600 and an interest rate of 4.1% per annum, compounding monthly. A regular deposit of \$200 is made every month.

The following table can be used to model the value of the investment for the first two months.

payment number	payment	interest	principal addition	balance of investment
0	0.00	0.00	0.00	5600.00
1				
2				

Fill in the missing cells.

8. Johnathan puts \$104 123 in an annuity investment earning interest at a rate of 1.04% per annum, compounding monthly. He will be adding \$1000 to the account each month.

He wants to set up a table that can model the growth of his investment from month to month.

payment number	payment	interest	principal addition	balance of investment
0				
1				
2				
3				

a. Fill in the missing cells.

b. How much interest was earned in the first three months?

c. Explain why the calculated interest is increasing each month.

d. Explain why the principal addition is increasing each month.

Back to contents

Using financial applications of technology to solve problems involving annuity investments

9. Mariana invests \$55 000 in an annuity investment which earns interest at a rate of 3.9% per annum, compounding monthly. She deposits a further \$450 per month.

After 6 months, the value of the annuity is closest to

A. \$53 359.21 **B.** \$53 359.22 **C.** \$58 803.28 **D.** \$58 803.29

10. Jonah needs \$17 780 to go on a whale watching trip exactly one year from now. He currently has \$8500 that he will put in a savings account earning interest at a rate of 4.1% per annum, compounding monthly. Jonah knows that this alone will not be enough and plans to make monthly deposits into the account.

How much will Jonah have to deposit each month to be able to afford the trip?

11. Jacob wants to have \$51 988 saved up for an extended holiday in Dubai. He plans to go in 1.5 years. He deposits \$12 000 with his bank and makes regular payments of \$1000 to the account each fortnight to save up for the trip. His account earns interest, compounding fortnightly.

Assuming he saves up the exact amount in 1.5 years, what interest rate does he require from the bank? Round your answer to one decimal place.

12. Bel invests \$50 000 in an annuity investment and wishes to have enough for a down payment of a house in 5 years. Her dream house is \$650 000 and she is required to make a 20% down payment. Her bank currently offers an interest rate of 4.2% per annum, compounding quarterly.

a. How much does the annuity investment need to be worth after the 5 years?

b. How much does Bel need to deposit every quarter in order to have enough for a down payment after 5 years?

Bel is able to add \$5000 each quarter to the account instead.

c. How many quarters will it take for her goal to be reached?

13. Rachel wants to purchase a car in four years. She currently has \$20 000 which she will put in an annuity investment that earns interest at a rate of 3.4% per annum, compounding fortnightly. She will make regular deposits of \$161.02 on a fortnightly basis that will allow her investment to reach the exact value of the car she wishes to purchase.

a. How much is the car worth?

b. How much is the annuity investment worth after two years?

After two years, her bank changes the interest rate and she is required to increase her fortnightly payments to \$169.15 in order to reach her goal in the remaining two years. Interest continues to compound fortnightly.

c. What is the new annual interest rate, rounded to two significant figures? Use the rounded values found in parts **a** and **b**.

Joining it all together

14. The following table models an annuity investment from week to week.

payment number	payment	interest	principal addition	balance of investment
0	0.00	0.00	0.00	7000.00
1	30.00	5.81	35.81	7035.81
2	30.00	5.84	35.84	7071.65

a. Calculate the annual interest rate for this annuity investment, rounded to four significant figures.

b. Construct a recurrence relation that can be used to calculate the value of the annuity investment V_n, after n weeks. Round to five decimal places where necessary.

c. How many weeks will it take for the annuity investment to surpass \$7200?

15. Ash wants to retire in 30 years with at least \$430 000 in savings. He currently has \$15 000 which will be used to establish an annuity investment. In addition to this, Ash will need to make equal weekly deposits to ensure that he can retire comfortably. His bank currently offers an interest rate of 2.1% per annum, compounding weekly.

a. What is the minimum amount that Ash needs to deposit each week in order to meet his goals?

b. After 10 years, Ash is offered a pay rise, and therefore is able to increase his weekly payments. His bank continues to offer the same interest rate, compounding weekly. This allows him to retire in the next 15 years.
What is this new payment amount? Use the rounded value found in part **a**.

Exam practice

16. The value of an annuity investment, in dollars, after n years, V_n, can be modelled by the following recurrence relation.

$V_0 = 46\ 000, \quad V_{n+1} = 1.0034 \times V_n + 500$

a. 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

b. 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

VCAA 2018 Exam 1 Recursion and financial modelling Q17, 18

Part **a**: **79%** of students answered this question correctly.
Part **b**: **61%** of students answered this question correctly.

17. Joanna deposited \$12 000 in an investment account earning interest at the rate of 2.8% per annum, compounding monthly.

She would like this account to reach a balance of \$25 000 after five years.

To achieve this balance, she will make an extra payment into the account each month, immediately after the interest is calculated.

The minimum value of this payment is closest to

A. \$113.85 **B.** \$174.11 **C.** \$580.16
D. \$603.22 **E.** \$615.47

VCAA 2021 Exam 1 Recursion and financial modelling Q22

56% of students answered this question correctly.

18. Four lines of a table, similar to 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.

Back to contents

The value of payment number 20 is closest to

A. $29 **B.** $100 **C.** $135
D. $237 **E.** $295

VCAA 2017 Exam 1 Recursion and financial modelling Q23

38% of students answered this question correctly.

19. Alex sold his mechanics' business for $360 000 and invested this amount in 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 $500 000 in his annuity after a further two years.
What is the value of Alex's new monthly payment?
Round the answer to the nearest cent. (2 MARKS)

VCAA 2017 Exam 2 Recursion and financial modelling Q7b

The average mark for this question was **0.6**.

Questions from multiple lessons

Recursion and financial modelling

20. The value, R_n, after n years of a reducing balance loan can be modelled by the following recurrence relation.

$R_0 = 100\ 000, \quad R_{n+1} = 1.04 \times R_n - 6500$

What is the balance of the loan after six years?

A. $81 692.99 **B.** $83 417.56 **C.** $86 459.19 **D.** $87 531.90 **E.** $89 907.56

Adapted from VCAA 2017 Exam 1 Recursion and financial modelling Q17

Data analysis

21. The number of Instagram followers for a group of Year 11 students is approximately normally distributed.

One of the students, Alejandro, has a standardised number of followers of $z = 1$.

The percentage of students with less Instagram followers than Alejandro is closest to

A. 16% **B.** 34% **C.** 68% **D.** 84% **E.** 98%

Adapted from VCAA 2018 Exam 1 Data analysis Q4

Recursion and financial modelling

22. Archibald has recently moved out of home and opened a savings account to save money for a big couch so that his friends can come over.

The amount of money in the savings account after n years, V_n, can be modelled by the recurrence relation

$V_0 = 1000, \quad V_{n+1} = 1.05 \times V_n$

a. How much money did Archibald initially deposit into the savings account? (1 MARK)

b. Use recursion to show the amount of money that will be in Archibald's savings account after 3 years, V_3, to the nearest cent. (1 MARK)

c. 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. Complete the rule by writing the appropriate numbers in the boxes provided. (1 MARK)

$V_n = \boxed{}^{\,n} \times \boxed{}$

ii. How much money, to the nearest cent, will be in Archibald's account after 7 years? (1 MARK)

Adapted from VCAA 2016 Exam 2 Recursion and financial modelling Q5

Back to contents

AOS 2

Matrices

CALCULATOR QUICK LOOK-UP GUIDE

Back to contents

7

CHAPTER 7

Matrices

LESSONS

KEY KNOWLEDGE

- 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 $a_{ij}{}^{\text{th}}$ 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
- use of the matrix recurrence relation: $S_0 =$ initial state matrix, $S_{n+1} = TS_n$ or $S_{n+1} = LS_n$ where T is a transition matrix, L is a Leslie matrix, and S_n 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 $S_0 =$ initial state matrix, $S_{n+1} = TS_n + B$ to extend modelling to populations that include culling and restocking.

Back to contents

7A Introduction to matrices

STUDY DESIGN DOT POINTS

- 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)
- use of matrices to represent numerical information presented in tabular form, and the use of a rule for the $a_{ij}{}^{th}$ element of a matrix to construct the matrix

7A 7B 7C 7D 7E 7F 7G 7H 7I

KEY SKILLS

During this lesson, you will be:

- identifying matrix properties and types
- constructing and interpreting matrices.

KEY TERMS

- Matrix
- Row
- Column
- Order
- Element
- Row matrix
- Column matrix
- Square matrix
- Zero matrix
- Leading diagonal
- Symmetric matrix
- Diagonal matrix
- Upper triangular matrix
- Lower triangular matrix
- Identity matrix

Matrices are useful tools for displaying data. Simple and complex applications can be modelled using matrices. Before this can be conducted, it is important to understand the fundamental properties of matrices.

Identifying matrix properties and types

See worked example 1

A **matrix** is a tool for displaying a collection of numerical values that is sorted into rows and columns depending on what it represents.

A **row** is a horizontal list of numbers and is counted from top to bottom.

$$\begin{array}{cc} \begin{bmatrix} 2 & 3 & 7 & 2 \\ 5 & 11 & 3 & 5 \\ 7 & 7 & 8 & 2 \end{bmatrix} & \begin{array}{c} \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \end{array} \end{array}$$

A **column** is a vertical list of numbers and is counted from left to right.

$$\begin{array}{c} \begin{array}{cccc} \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} \end{array} \\ \begin{bmatrix} 2 & 3 & 7 & 2 \\ 5 & 11 & 3 & 5 \\ 7 & 7 & 8 & 2 \end{bmatrix} \end{array}$$

The **order** of a matrix describes its dimensions, and is expressed as *number of rows* $\times$ *number of columns.*

For example, if a matrix has three rows and four columns, it is referred to as a 'three-by-four matrix', and is expressed as 3×4.

An **element** is an entry in a matrix. The total number of elements can be found by multiplying the number of rows with the number of columns. Matrices are usually defined with a capital letter, such as A. To refer to a particular element in a matrix, the lowercase letter of the matrix is written, followed by the row and column number written in subscript to the right.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{22} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

For matrix A, element a_{mn} refers to the entry in the m^{th} row and n^{th} column.

$$D = \begin{array}{cc} \begin{array}{cc} 1 & \\ \end{array} & \\ \begin{bmatrix} 3 & 6 \\ 8 & 7 \end{bmatrix} & \begin{array}{c} \\ 2 \end{array} \end{array}$$

For matrix D, element d_{21} refers to the entry in the 2nd row and the 1st column, which is 8.

There are various different types of matrices.

See worked example 2

A **row matrix** has only one row and any number of columns.

$\begin{bmatrix} 21 & 31 \end{bmatrix}$ and $\begin{bmatrix} 8 & 17 & 42 & 52 \end{bmatrix}$

A **column matrix** has only one column and any number of rows.

$$\begin{bmatrix} 42 \\ 56 \\ 74 \end{bmatrix} \text{ and } \begin{bmatrix} 19 \\ 63 \\ 17 \\ 42 \end{bmatrix}$$

A **square matrix** has an equal number of rows and columns.

$$\begin{bmatrix} 46 & 29 \\ 62 & 83 \end{bmatrix} \text{ and } \begin{bmatrix} 11 & 8 & 53 \\ 6 & 98 & 5 \\ 23 & 55 & 72 \end{bmatrix}$$

A **zero matrix** is a matrix where every element is '0'.

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

The **leading diagonal** is the series of elements in a square matrix that goes from the top left element to the bottom right element.

$$D = \begin{bmatrix} d_{11} & \cdots & 0 & 0 \\ 0 & d_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_{mn} \end{bmatrix}$$

A **symmetric matrix** is a square matrix that is symmetric along the leading diagonal.

$$\begin{bmatrix} 3 & 6 & 5 \\ 6 & 0 & 1 \\ 5 & 1 & 2 \end{bmatrix}$$

A **diagonal matrix** is a square, symmetric matrix in which all elements not in the leading diagonal are '0'.

$$\begin{bmatrix} \mathbf{25} & 0 & 0 \\ 0 & \mathbf{9} & 0 \\ 0 & 0 & \mathbf{16} \end{bmatrix} \text{ and } \begin{bmatrix} 17 & 0 & 0 & 0 \\ 0 & 32 & 0 & 0 \\ 0 & 0 & 57 & 0 \\ 0 & 0 & 0 & 12 \end{bmatrix}$$

An **upper triangular matrix** is a square matrix where all elements below the leading diagonal are '0'.

$$\begin{bmatrix} 4 & 5 & 7 \\ 0 & 1 & 6 \\ 0 & 0 & 8 \end{bmatrix}$$

A **lower triangular matrix** is a square matrix where all elements above the leading diagonal are '0'.

$$\begin{bmatrix} 3 & 0 & 0 \\ 2 & 3 & 0 \\ 5 & 4 & 2 \end{bmatrix}$$

An **identity matrix** is a square, diagonal matrix in which the leading diagonal consists only of '1', and '0' elsewhere. This matrix is denoted with I.

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Back to contents

Worked example 1

Matrix A contains a list of multiple elements.

$$A = \begin{bmatrix} 101 & 2 & 85 \\ 41 & 56 & 73 \end{bmatrix}$$

a. What is the order of matrix A?

Explanation

Step 1: Count the number of rows.

The matrix has 2 rows.

Step 2: Count the number of columns.

The matrix has 3 columns.

Answer

2×3

b. What is element a_{21}?

Explanation

Step 1: Locate the row.

The 2 in a_{21} refers to the second row.

$$\begin{bmatrix} 101 & 2 & 85 \\ 41 & 56 & 73 \end{bmatrix}$$

Step 2: Locate the column.

The 1 in a_{21} refers to the first column.

$$\begin{bmatrix} 101 & 2 & 85 \\ 41 & 56 & 73 \end{bmatrix}$$

Answer

41

Worked example 2

Matrices P, Q, R, S and T are all 3×3 matrices. Each of the following matrices fall into one or more classifications, namely diagonal, identity, symmetric, upper triangular, and lower triangular.

a. How can matrix P be classified?

$$P = \begin{bmatrix} 6 & 5 & 2 \\ 0 & 5 & 6 \\ 0 & 0 & 5 \end{bmatrix}$$

Explanation

Consider the properties of the matrix.

All elements below the main diagonal are '0'.

Answer

Upper triangular matrix

b. How can matrix Q be classified?

$$Q = \begin{bmatrix} 1 & 0 & 5 \\ 0 & 0 & 2 \\ 5 & 2 & 3 \end{bmatrix}$$

Continues →

Back to contents

Explanation

Consider the properties of the matrix.

Each element is symmetric across the leading diagonal.

$$\begin{bmatrix} 1 & 0 & 5 \\ 0 & 0 & 2 \\ 5 & 2 & 3 \end{bmatrix}$$

Answer

Symmetric matrix

---

c. How can matrix R be classified?

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Explanation

Consider the properties of the matrix.

Each element is symmetric across the leading diagonal.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

All elements not in the leading diagonal are ‘0’.

It is only ‘1’ on the leading diagonal and ‘0’ elsewhere.

All elements below the main diagonal are ‘0’.

All elements above the main diagonal are ‘0’.

Answer

Symmetric, diagonal, identity, upper triangular, lower triangular matrix

---

d. How can matrix S be classified?

$$S = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 7 \end{bmatrix}$$

Explanation

Consider the properties of the matrix.

Each element is symmetric across the leading diagonal.

$$\begin{bmatrix} 5 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 7 \end{bmatrix}$$

All elements not in the leading diagonal are ‘0’.

All elements below the main diagonal are ‘0’.

All elements above the main diagonal are ‘0’.

Answer

Symmetric, diagonal, upper triangular, lower triangular matrix

Continues →

Back to contents

e. How can matrix T be classified?

$$T = \begin{bmatrix} 0 & 0 & 0 \\ 11 & 2 & 0 \\ 2 & 5 & 8 \end{bmatrix}$$

Explanation

Consider the properties of the matrix.

All elements above the main diagonal are '0'.

Answer

Lower triangular matrix

Constructing and interpreting matrices

See worked example 3

Matrices can be used to represent numerical information. In most cases, storing data in a matrix makes it easier to visualise and manipulate. For example, the following table is used to represent the number of days with snow (S) and without snow (N) in July (J) and August (A).

	July (J)	August (A)
number of days with snow (S)	24	17
number of days without snow (N)	7	13

A matrix can be constructed that represents the information in the table.

$$\begin{array}{cc} \begin{array}{cc} J & A \end{array} & \\ \begin{bmatrix} 24 & 17 \\ 7 & 13 \end{bmatrix} & \begin{array}{c} S \\ N \end{array} \end{array}$$

See worked example 4

Matrices can also be constructed using element rules. These element rules often use the row and column number of a particular element, as part of the rule.

For example, matrix A is a 2×2 matrix with the element rule $a_{ij} = i + j$, where a_{ij} is the element in the i^{th} row and j^{th} column. The element rule can be applied to each element in A to give the following.

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad A = \begin{bmatrix} 1+1 & 1+2 \\ 2+1 & 2+2 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 3 & 4 \end{bmatrix}$$

Worked example 3

A group of 100 people were surveyed about their favourite fruit. Their results are shown in the table.

	favourite fruit		
	banana	**dragon fruit**	**kiwi fruit**
adult	14	19	16
child	23	11	17

a. Construct matrix F, to represent the information in the table.

Explanation

Step 1: Set up an empty matrix.

There are 2 rows and 3 columns of data presented in the table. As such, the order of the matrix should be 2×3. Rows can be labelled A and C for 'adult' and 'child' respectively. Columns can be labelled B, D and K for banana, dragon fruit and kiwifruit respectively.

$$F = \begin{array}{cc} \begin{array}{ccc} B & D & K \end{array} & \\ \begin{bmatrix} & & \\ & & \end{bmatrix} & \begin{array}{c} A \\ C \end{array} \end{array}$$

Continues →

Back to contents

Step 2: Fill in the first row.

Element f_{11} corresponds to the number of adults who prefer bananas, which is 14.

Element f_{12} corresponds to the number of adults who prefer dragon fruit, which is 19.

Element f_{13} corresponds to the number of adults who prefer kiwifruit which is 16

$$F = \begin{array}{c} \begin{matrix} \text{B} & \text{D} & \text{K} \end{matrix} \\ \begin{bmatrix} 14 & 19 & 16 \\ & & \end{bmatrix} \end{array} \begin{matrix} \text{A} \\ \text{C} \end{matrix}$$

Step 3: Repeat for the second row.

This row corresponds to the favourite fruit of children.

Answer

$$F = \begin{array}{c} \begin{matrix} \text{B} & \text{D} & \text{K} \end{matrix} \\ \begin{bmatrix} 14 & 19 & 16 \\ 23 & 11 & 17 \end{bmatrix} \end{array} \begin{matrix} \text{A} \\ \text{C} \end{matrix}$$

b. Construct a row matrix to represent the favourite fruit of adults surveyed.

Explanation

Step 1: Identify the row on the table which corresponds to the favourite fruits of adults.

The first row on the table represents the favourite fruit of adults.

Step 2: Construct a row matrix representing the data.

A row matrix will only have one row. As there are three fruits listed on the table, the row matrix will have three columns.

Answer

$\begin{bmatrix} 14 & 19 & 16 \end{bmatrix}$

c. Interpret the sum of the matrix constructed in **b**.

Explanation

Step 1: Sum the elements in the matrix.

$14 + 19 + 16 = 49$

Step 2: Interpret the sum.

The matrix constructed in part **b** corresponds to the adults surveyed and their favourite fruits.

Answer

The total number of adults surveyed, which was 49.

Worked example 4

Matrix C is a matrix with an order of 3×2.

Construct matrix C, with the element rule $c_{ij} = i \times j$, where c_{ij} refers to the entry in the i^{th} row and j^{th} column.

Explanation

Step 1: Set up the matrix.

$$C = \begin{array}{c} \begin{matrix} 1 & 2 \end{matrix} \\ \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \end{bmatrix} \end{array} \begin{matrix} 1 \\ 2 \\ 3 \end{matrix}$$

Step 2: Calculate the matrix elements.

$$C = \begin{array}{c} \begin{matrix} 1 & 2 \end{matrix} \\ \begin{bmatrix} 1 \times 1 & 1 \times 2 \\ 2 \times 1 & 2 \times 2 \\ 3 \times 1 & 3 \times 2 \end{bmatrix} \end{array} \begin{matrix} 1 \\ 2 \\ 3 \end{matrix}$$

$$= \begin{array}{c} \begin{matrix} 1 & 2 \end{matrix} \\ \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \end{bmatrix} \end{array} \begin{matrix} 1 \\ 2 \\ 3 \end{matrix}$$

Continues →

Back to contents

Answer

$$C = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \end{bmatrix}$$

Exam question breakdown

VCAA 2019 Exam 1 Matrices Q3

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 p_{ij}.

The elements in matrix P are determined by the rule

A. $p_{ij} = 4 - j$ **B.** $p_{ij} = 2i + 1$ **C.** $p_{ij} = i + j + 1$ **D.** $p_{ij} = i + 2j$ **E.** $p_{ij} = 2i - j + 2$

Explanation

Step 1: Set up the matrix.

$$P = \begin{array}{cc} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ & \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \end{bmatrix} \begin{array}{c} 1 \\ 2 \end{array} \end{array}$$

Step 2: Calculate the matrix elements for each option and compare it P.

A:

$$P_A = \begin{bmatrix} 4-1 & 4-2 & 4-3 \\ 4-1 & 4-2 & 4-3 \end{bmatrix} = \begin{bmatrix} 3 & 2 & 1 \\ 3 & 2 & 1 \end{bmatrix}$$

$P_A \neq P$ ✗

B:

$$P_B = \begin{bmatrix} 2\times 1+1 & 2\times 1+2 & 2\times 1+3 \\ 2\times 2+1 & 2\times 2+2 & 2\times 2+3 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 4 & 5 \\ 5 & 6 & 7 \end{bmatrix}$$

$P_B \neq P$ ✗

C:

$$P_C = \begin{bmatrix} 1+1+1 & 1+2+1 & 1+3+1 \\ 2+1+1 & 2+2+1 & 2+3+1 \end{bmatrix} = \begin{bmatrix} 3 & 4 & 5 \\ 4 & 5 & 6 \end{bmatrix}$$

$P_C \neq P$ ✗

D:

$$P_D = \begin{bmatrix} 1+2\times 1 & 1+2\times 2 & 1+2\times 3 \\ 2+2\times 1 & 2+2\times 2 & 2+2\times 3 \end{bmatrix} = \begin{bmatrix} 3 & 5 & 7 \\ 4 & 6 & 8 \end{bmatrix}$$

$P_D \neq P$ ✗

E:

$$P_E = \begin{bmatrix} 2\times 1-1+2 & 2\times 1-2+2 & 2\times 1-3+2 \\ 2\times 2-1+2 & 2\times 2-2+2 & 2\times 2-3+2 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 2 & 1 \\ 5 & 4 & 3 \end{bmatrix}$$

$P_E = P$ ✓

59% of students answered this question correctly.

A total of **34%** of students incorrectly selected B, C and D. This is likely because they had only worked through a select number of elements in matrix P. In such questions, it is important to spend the time to calculate each matrix element individually.

Answer

E

Back to contents

7A Questions

Identifying matrix properties and types

1. What matrix has an order of 2×3?

A. $\begin{bmatrix} 3 & 7 & 5 \\ 1 & 8 & 9 \end{bmatrix}$ **B.** $\begin{bmatrix} 2 & 5 & 2 & 9 \\ 0 & 8 & 6 & 4 \\ 6 & 8 & 6 & 3 \\ 6 & 6 & 2 & 5 \end{bmatrix}$ **C.** $\begin{bmatrix} 1 & 4 \\ 7 & 2 \\ 3 & 5 \end{bmatrix}$ **D.** $\begin{bmatrix} 5 & 2 & 1 \\ 1 & 3 & 9 \\ 8 & 1 & 3 \end{bmatrix}$

2. Consider matrix B.

$$B = \begin{bmatrix} 5 & 2 \\ 1 & 3 \\ 8 & 1 \\ 1 & 7 \end{bmatrix}$$

a. How many rows and columns can be found in B?

b. What is the order of B?

c. What entry corresponds with b_{31}?

3. A, B, C and D are all matrices of different orders.

$$A = \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix} \quad B = \begin{bmatrix} 5 & 6 & 7 \\ 1 & 3 & 8 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 7 & 4 \\ 4 & 8 & 0 \\ 12 & 5 & 7 \end{bmatrix} \quad D = \begin{bmatrix} 11 & 77 & 72 & 24 \\ 46 & 26 & 89 & 23 \end{bmatrix}$$

a. What is the order of A?

b. What is the order of B?

c. What is the order of C?

d. What is the order of D?

4. The following matrix is a 3×5 matrix.

$$D = \begin{bmatrix} 1 & 89 & 67 & 4 & 111 \\ 32 & 4 & 46 & 53 & 72 \\ 74 & 3 & 67 & 12 & 47 \end{bmatrix}$$

a. What entry corresponds to d_{23}?

b. What entry corresponds to d_{35}?

c. What entry corresponds to d_{15}?

d. Which entry corresponds to d_{24}?

5. M, N, O, P, and Q are all 3×3 matrices.

$$M = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad N = \begin{bmatrix} 9 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad O = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad P = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 5 \end{bmatrix} \quad Q = \begin{bmatrix} 7 & 0 & 0 \\ 1 & 5 & 0 \\ 3 & 5 & 8 \end{bmatrix}$$

Identify the matrices that can be classified as

a. an upper triangular matrix.

b. a diagonal matrix.

c. the identity matrix.

d. a symmetric matrix.

Back to contents

Constructing and interpreting matrices

6. A group of 100 people were surveyed about their favourite pets. The results are shown in the following table.

	favourite pets		
	cats	**dogs**	**fish**
class A	18	12	16
class B	19	20	15

Choose the matrix that correctly represents the information in the table.

A. $\begin{array}{cc} \text{B} & \\ \begin{bmatrix} 19 \\ 15 \\ 20 \end{bmatrix} & \begin{array}{c} \text{C} \\ \text{D} \\ \text{F} \end{array} \end{array}$

B. $\begin{array}{cc} \begin{array}{ccc} \text{C} & \text{D} & \text{F} \end{array} & \\ \begin{bmatrix} 18 & 16 & 12 \end{bmatrix} & \text{A} \end{array}$

C. $\begin{array}{cc} \begin{array}{ccc} \text{C} & \text{D} & \text{F} \end{array} & \\ \begin{bmatrix} 18 & 12 & 16 \\ 19 & 20 & 15 \end{bmatrix} & \begin{array}{c} \text{A} \\ \text{B} \end{array} \end{array}$

D. $\begin{array}{cc} \begin{array}{cc} \text{A} & \text{B} \end{array} & \\ \begin{bmatrix} 19 & 18 \\ 12 & 20 \\ 16 & 15 \end{bmatrix} & \begin{array}{c} \text{C} \\ \text{D} \\ \text{F} \end{array} \end{array}$

---

7. The following table and matrix both represent the number of residents and cats in three different households.

	number of residents	*number of cats*
household 1	5	2
household 2	3	1
household 3	1	9

$$A = \begin{bmatrix} 5 & 2 \\ 3 & 1 \\ 1 & 9 \end{bmatrix}$$

a. What does element a_{21} represent?

b. What does element a_{32} represent?

c. What does column 1 of matrix A represent?

d. What does column 2 of matrix A represent?

---

8. A group of children were surveyed on their favourite animal between flamingoes (F), porcupines (P) and reindeer (R). The results are shown in the given table.

	favourite animal		
	flamingoes	**porcupines**	**reindeer**
boys	15	8	9
girls	13	14	5

a. Convert the table into a matrix with the following form.

$$\begin{array}{cc} \begin{array}{ccc} \text{F} & \text{P} & \text{R} \end{array} & \\ \begin{bmatrix} \quad & \quad & \quad \\ \quad & \quad & \quad \end{bmatrix} & \begin{array}{l} \text{boys} \\ \text{girls} \end{array} \end{array}$$

b. What does the entry in the second row and second column of the matrix represent?

c. What does the sum of all elements in the first row of the matrix represent?

---

9. Construct a matrix using the given order and element rule, where a_{ij} is the element in the i^{th} row and j^{th} column.

a. 2×2 and $a_{ij} = 2i + 2j$

b. 3×2 and $a_{ij} = 2 \times i \times j$

c. 2×4 and $a_{ij} = (i + j)^2$

Back to contents

Joining it all together

10. Kyle noted how many scoops of each ice cream flavour he sold over three days. The results are in the table shown.

	day 1	day 2	day 3
cookies and cream	23	19	28
salted caramel	13	20	9
strawberry	14	11	26

a. Construct matrix F, a square matrix that represents the number of scoops of each flavour sold each day.

b. What does the sum of all elements in matrix F represent?

c. Construct matrix G, a column matrix that represents the number of scoops of each flavour sold on day 2.

d. What does the sum of all elements in matrix G represent?

---

11. Matrix W shows the total amount of water stored, in litres, in a water tank from Monday to Thursday.

$$\begin{array}{cc} & \begin{array}{cccc} \text{Mon} & \text{Tue} & \text{Wed} & \text{Thu} \end{array} \\ W = & \begin{bmatrix} 300 & 600 & 900 & 1200 \end{bmatrix} \end{array}$$

a. What is the order of this matrix?

b. What type of matrix is this?

c. What does w_{13} represent?

Exam practice

12. 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.

$$C = \begin{array}{c} \begin{bmatrix} 3.58 \\ 2.22 \\ 2.87 \end{bmatrix} \end{array} \begin{array}{c} \text{E} \\ \text{F} \\ \text{G} \end{array}$$

a. What is the cost of one journey on section G? (1 MARK)

b. Write down the order of matrix C. (1 MARK)

c. One day, Kim travels once on section E and twice in section G. Construct a row matrix that shows this. (1 MARK)

VCAA 2018 Exam 2 Matrices Q1

Part **a**: **99%** of students answered this question correctly.
Part **b**: **94%** of students answered this question correctly.
Part **c**: **66%** of students answered this question correctly.

Questions from multiple lessons

Data analysis

13. Data was collected on how employees commuted to work, to investigate the association between the two following variables.

- *rides a bicycle to work* (yes, no)
- *distance from office* (under 2 km, 2–10 km, over 10 km)

Which one of the following is appropriate to use in the statistical analysis of this association?

A. Back to back stem plot
B. Segmented bar chart
C. The coefficient of determination
D. Residual plot
E. Parallel boxplots

Adapted from VCAA 2018 Exam 1 Data analysis Q6

Back to contents

Recursion and financial modelling

14. The value of an investment, in dollars, after n years, V_n , can be modelled by the recurrence relation shown.

$V_0 = 32\ 000, \quad V_{n+1} = 1.0038V_n + 350$

What is the value of the regular payment added to the principal of this investment?

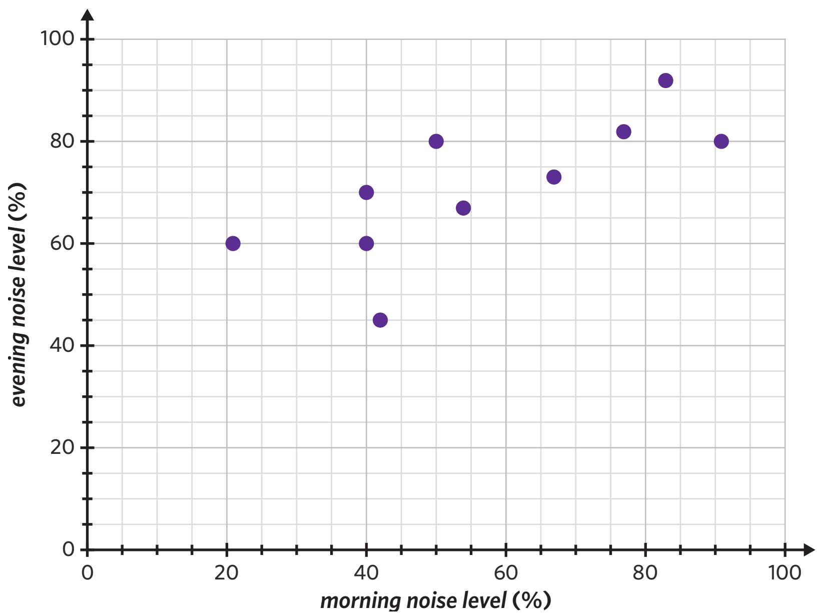
A. $38.00 **B.** $350.00 **C.** $32 000.00 **D.** $325.00 **E.** $700.00

Adapted from VCAA 2018 Exam 1 Recursion and financial modelling Q17

Data analysis

15. The amount of noise in a suburb can be recorded as a percentage of the maximum amount of noise allowed by the council. This is called the percentage noise level.

The percentage noise levels for the morning and evening peak periods for 10 large suburbs are plotted on the following scatterplot.



A least squares line is to be fitted to the data with the aim of predicting evening noise level from the morning noise level.

The equation of this line is

evening noise level = 45.3 + 0.45 × *morning noise level*

a. Use the equation of the least squares line to predict the evening noise level when the morning noise level is 75%. Round to one decimal place. (1 MARK)

b. Determine the residual value when the equation of the least squares line is used to predict the evening noise level when the morning noise level is 50%. Round to one decimal place. (2 MARKS)

c. The value of the correlation coefficient r is 0.74. What percentage of the variation in the evening noise level can be explained by the variation in the morning noise level? Round to the nearest whole number. (1 MARK)

Adapted from VCAA 2018 Exam 2 Data analysis Q2

Back to contents

7B Operations with matrices

STUDY DESIGN DOT POINT

- 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)

7A 7B 7C 7D 7E 7F 7G 7H 7I

KEY SKILLS

During this lesson, you will be:

- adding and subtracting matrices
- multiplying matrices by a scalar
- determining the transpose of a matrix.

KEY TERMS

- Scalar multiplication
- Transpose

Once matrices have been created, it can be helpful to perform operations on them in order to make calculations with the data. It is important to know how each of the basic operations differ when performing them with matrices rather than single values.

Adding and subtracting matrices

Matrix addition or subtraction is only defined if the matrices are of the same order. If they are not, then the solution is undefined.

When adding or subtracting matrices, elements in the same position are added or subtracted from each other.

Matrix addition

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} e & g \\ f & h \end{bmatrix} = \begin{bmatrix} a+e & b+g \\ c+f & d+h \end{bmatrix}$$

Matrix subtraction

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} - \begin{bmatrix} e & g \\ f & h \end{bmatrix} = \begin{bmatrix} a-e & b-g \\ c-f & d-h \end{bmatrix}$$

Worked example 1

Consider the matrices $A = \begin{bmatrix} 2 & 2 \\ 1 & 4 \\ 5 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & 1 \\ 0 & 2 \\ 4 & 5 \end{bmatrix}$.

a. Is $A + B$ defined?

Explanation

Step 1: Identify the order of each matrix.

Matrix A has three rows and two columns. Its order is 3×2.

Matrix B has three rows and two columns. Its order is 3×2.

Step 2: Determine whether the matrix sum is defined.

The matrices are of the same order.

Hence, the matrix sum is defined.

Answer

Yes

Continues →

b. If possible, calculate $A - B$.

Explanation – Method 1: By hand

Step 1: Write down the calculation.

$$A - B = \begin{bmatrix} 2 & 2 \\ 1 & 4 \\ 5 & 3 \end{bmatrix} - \begin{bmatrix} 4 & 1 \\ 0 & 2 \\ 4 & 5 \end{bmatrix}$$

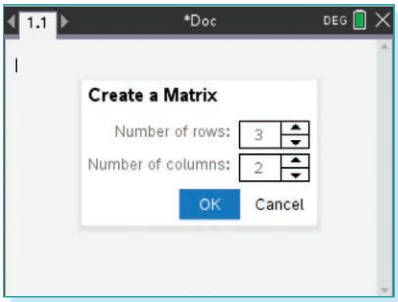
Step 2: Subtract each element in matrix B from its corresponding element in matrix A.

$$= \begin{bmatrix} 2-4 & 2-1 \\ 1-0 & 4-2 \\ 5-4 & 3-5 \end{bmatrix}$$

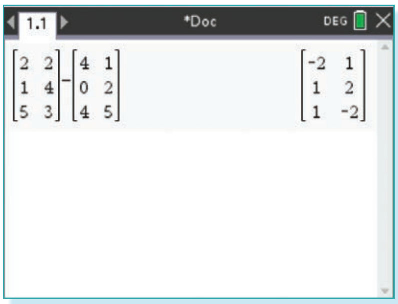
Explanation – Method 2: TI-Nspire

Step 1: From the home screen, select '1: New' → '1: Add Calculator'.

Step 2: Press [templates key] and select [matrix template]. On the settings window, set 'Number of rows' as 3 and 'Number of columns' as 2. Select 'OK'. Enter the values for matrix A.



Step 3: Type '-' and then repeat step 2 for matrix B. Press enter to calculate.

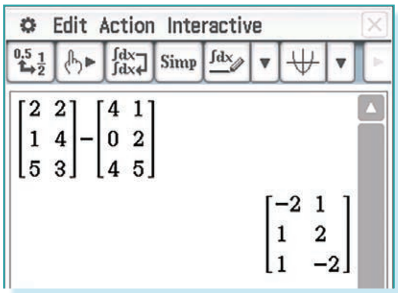


Explanation – Method 3: Casio ClassPad

Step 1: From the main menu, tap √α Main.

Step 2: Press keyboard and tap Math2. Tap [matrix icon] to create a matrix, and [row icon] to add an extra row. Enter the values for matrix A.

Step 3: Type '-' and then repeat step 2 for matrix B. Press EXE to calculate.



Answer – Method 1, 2 and 3

$$\begin{bmatrix} -2 & 1 \\ 1 & 2 \\ 1 & -2 \end{bmatrix}$$

Back to contents

Multiplying matrices by a scalar

Scalar multiplication refers to multiplying a matrix by a number (the scalar). Each element in the matrix is multiplied by the scalar.

$$k \times \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} k \times a & k \times b \\ k \times c & k \times d \end{bmatrix}$$

Worked example 2

Given that $C = \begin{bmatrix} 2 & 1 & 3 \\ 4 & 5 & 2 \\ 1 & 0 & 4 \end{bmatrix}$, calculate $3C$.

Explanation – Method 1: By hand

Multiply each individual element by 3.

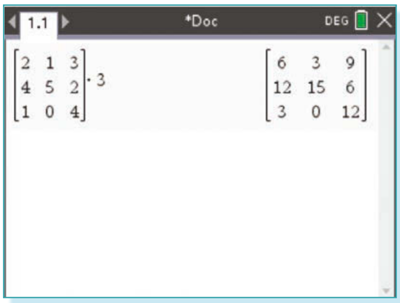
$$3 \times \begin{bmatrix} 2 & 1 & 3 \\ 4 & 5 & 2 \\ 1 & 0 & 4 \end{bmatrix} = \begin{bmatrix} 3 \times 2 & 3 \times 1 & 3 \times 3 \\ 3 \times 4 & 3 \times 5 & 3 \times 2 \\ 3 \times 1 & 3 \times 0 & 3 \times 4 \end{bmatrix}$$

Explanation – Method 2: TI-Nspire

Step 1: From the home screen, select '1: New' → '1: Add Calculator'.

Step 2: Press and select . Select 'OK'. Enter the values for matrix C.

Step 3: Type '× 3' and press enter to calculate.

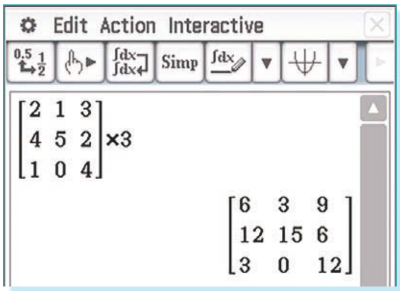


Explanation – Method 3: Casio ClassPad

Step 1: From the main menu, tap Main.

Step 2: Press keyboard and tap Math2. Tap to create a matrix, to add an extra row, and to add an extra column. Enter the values for matrix C.

Step 3: Type '× 3' and press EXE to calculate.



Answer – Method 1, 2 and 3

$$\begin{bmatrix} 6 & 3 & 9 \\ 12 & 15 & 6 \\ 3 & 0 & 12 \end{bmatrix}$$

Back to contents

Determining the transpose of a matrix

The **transpose** of a matrix can be determined by swapping its rows and columns.

The transpose of matrix A is denoted A^T.

If the order of matrix A is 2×4, the order of its transpose A^T will be 4×2.

$$\begin{bmatrix} a & b & c & d \\ e & f & g & h \end{bmatrix}^T = \begin{bmatrix} a & e \\ b & f \\ c & g \\ d & h \end{bmatrix}$$

Worked example 3

If $D = \begin{bmatrix} 1 & 0 & 9 \\ 5 & 2 & 3 \end{bmatrix}$, determine D^T.

Explanation – Method 1: By hand

Step 1: Determine the order of D^T.

The order of matrix D is 2×3.

Hence, the order of matrix D^T will be 3×2.

Step 2: Copy the values from the first row of D into the first column of D^T.

$$D^T = \begin{bmatrix} 1 & \\ 0 & \\ 9 & \end{bmatrix}$$

Step 3: Copy the values from the second row of D into the second column of D^T.

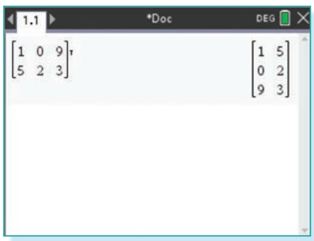
$$D^T = \begin{bmatrix} 1 & 5 \\ 0 & 2 \\ 9 & 3 \end{bmatrix}$$

Explanation – Method 2: TI-Nspire

Step 1: From the home screen, select '1: New' → '1: Add Calculator'.

Step 2: Press and select . On the settings window, set 'Number of rows' as 2 and 'Number of columns' as 3. Select 'OK'. Enter the values for matrix D.

Step 3: Press menu and select '7: Matrices' → '2: Transpose'. Press enter.

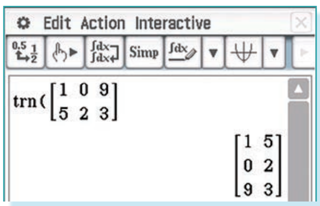


Explanation – Method 3: Casio ClassPad

Step 1: From the main menu, tap √α Main.

Step 2: Tap 'Action' → 'Matrix' → 'Create' → 'trn'.

Step 3: Press keyboard and tap Math2. Tap to create a matrix and to add an extra column. Enter the values for matrix D. Press EXE.



Answer – Method 1, 2 and 3

$$\begin{bmatrix} 1 & 5 \\ 0 & 2 \\ 9 & 3 \end{bmatrix}$$

Back to contents

Exam question breakdown

VCAA 2021 Exam 1 Matrices Q1

If matrix $M = \begin{bmatrix} 3 & 2 \\ 8 & 9 \\ 13 & 7 \end{bmatrix}$, then its transpose, M^T, is

A. $\begin{bmatrix} 2 & 3 \\ 9 & 8 \\ 7 & 13 \end{bmatrix}$

B. $\begin{bmatrix} 2 & 9 & 7 \\ 3 & 8 & 13 \end{bmatrix}$

C. $\begin{bmatrix} 7 & 9 & 2 \\ 13 & 8 & 3 \end{bmatrix}$

D. $\begin{bmatrix} 3 & 8 & 13 \\ 2 & 9 & 7 \end{bmatrix}$

E. $\begin{bmatrix} 13 & 8 & 3 \\ 7 & 9 & 2 \end{bmatrix}$

Explanation – Method 1: By hand

Step 1: Determine the order of M^T.

The order of matrix M is 3×2.

Hence, the order of matrix M^T will be 2×3.

Step 2: Transpose matrix M by swapping its rows and columns.

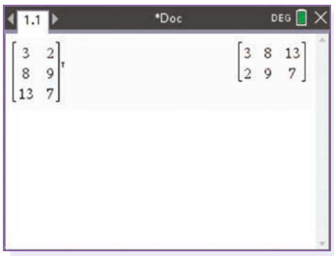
$$M^T = \begin{bmatrix} 3 & 8 & 13 \\ 2 & 9 & 7 \end{bmatrix}$$

Explanation – Method 2: TI-Nspire

Step 1: From the home screen, select ‘1: New’ → ‘1: Add Calculator’.

Step 2: Press [icon] and select [icon]. On the settings window, set ‘Number of rows’ as 3 and ‘Number of columns’ as 2. Select ‘OK’. Enter the values for matrix M.

Step 3: Press menu and select ‘7: Matrices’ → ‘2: Transpose’. Press enter.

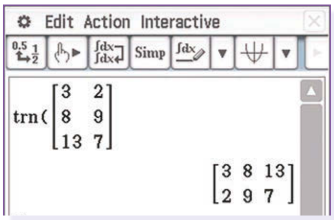


Explanation – Method 3: Casio ClassPad

Step 1: From the main menu, tap √α Main.

Step 2: Tap ‘Action’ → ‘Matrix’ → ‘Create’ → ‘trn’.

Step 3: Press keyboard and tap Math2. Tap [icon] to create a matrix and [icon] to add an extra row. Enter the values for matrix M. Press EXE.



Answer – Method 1, 2 and 3

D

82% of students answered this question correctly.

7B Questions

Adding and subtracting matrices

1. Which of the following matrix operations is defined?

A. $\begin{bmatrix} 4 & 1 & 9 \end{bmatrix} - \begin{bmatrix} 4 \\ 1 \\ 9 \end{bmatrix}$

B. $\begin{bmatrix} 1 & 8 & 9 & 0 & 5 \end{bmatrix} - \begin{bmatrix} 3 & 11 & 6 & 15 \end{bmatrix}$

C. $\begin{bmatrix} 3 & 4 \\ 12 & 19 \\ 5 & 11 \end{bmatrix} + \begin{bmatrix} 20 & -1 \\ 6 & 7 \\ 8 & 15 \end{bmatrix}$

D. $\begin{bmatrix} 2 & 8 \\ 4 & 6 \end{bmatrix} + \begin{bmatrix} 1 & 9 & 5 \\ 3 & 6 & 1 \\ 5 & 3 & 7 \end{bmatrix}$

2. For each of the following matrix operations, determine if the matrix operation is defined. If the operation is defined, evaluate.

a. $\begin{bmatrix} 5 & -9 & 7 & 5 \end{bmatrix} + \begin{bmatrix} 2 & 6 & -3 & 5 \end{bmatrix}$

b. $\begin{bmatrix} 3 & 0 \\ 1 & 7 \\ 5 & 3 \end{bmatrix} + \begin{bmatrix} 7 & 3 \\ 5 & 4 \end{bmatrix}$

c. $\begin{bmatrix} 2 & 3 & 8 \\ 8 & 5 & 9 \end{bmatrix} - \begin{bmatrix} 1 & 2 & 4 \\ 6 & 4 & 0 \end{bmatrix}$

d. $\begin{bmatrix} 9 & 5 \\ -13 & 8 \\ 4 & -3 \\ 11 & 7 \end{bmatrix} - \begin{bmatrix} 10 & 11 \\ 4 & -2 \\ 1 & 8 \\ 9 & 7 \end{bmatrix}$

3. Simon and Darren are both part-time Uber drivers who drive on Mondays, Wednesdays, and Fridays. The average number of passengers driven by each of them on each day are shown in the following table.

day	Simon	Darren
Mon	10	9
Wed	8	8
Fri	15	12

a. Construct two 3×1 matrices, S and D, that represent the average number of passengers driven each day by Simon and Darren respectively.

b. Calculate $S - D$. What does this matrix represent?

Their colleague Theodora works on Mondays, Wednesdays, Fridays, and Saturdays. Her average number of passengers each day is shown.

day	Theodora
Mon	14
Wed	11
Fri	16
Sat	20

c. Construct a 4×1 matrix, T, to represent this information.

d. Is it possible to evaluate $S + D + T$ to find the total number of passengers they drove each day? Explain why.

Multiplying matrices by a scalar

4. If $A = \begin{bmatrix} 4 \\ 6 \\ 8 \end{bmatrix}$, $3A$ is

A. $\begin{bmatrix} 54 \end{bmatrix}$

B. $\begin{bmatrix} 7 \\ 9 \\ 11 \end{bmatrix}$

C. $\begin{bmatrix} 12 \\ 14 \\ 16 \end{bmatrix}$

D. $\begin{bmatrix} 12 \\ 18 \\ 24 \end{bmatrix}$

Back to contents

5. Calculate each of the following scalar multiplications.

a. $2\begin{bmatrix}3 & 6 & 1 & 9 & 5\end{bmatrix}$

b. $4\begin{bmatrix}2 & 5 & -6\\ -3 & 1 & 4\end{bmatrix}$

c. $-3\begin{bmatrix}12\\ -7\\ -4\\ 9\\ 8\\ 11\end{bmatrix}$

d. $1.5\begin{bmatrix}6 & 10 & 5\\ 4 & -3 & -20\\ -1 & 8 & 2\end{bmatrix}$

---

6. The prices of men's and women's garments at a clothing store are displayed in matrix P.

$$P = \begin{array}{c} \begin{array}{cc} \text{men's} & \text{women's} \end{array} \\ \begin{bmatrix}40 & 45\\ 80 & 90\\ 100 & 120\end{bmatrix} \end{array} \begin{array}{l} \\ \text{t-shirt}\\ \text{hoodie}\\ \text{jeans}\end{array}$$

This week, the store is offering a 20% discount on all their stock.

The discounted prices can be calculated by multiplying matrix P by a scalar, k.

a. What is the value of the scalar, k?

b. Calculate $k \times P$.

c. What is the discounted price of a women's hoodie?

Determining the transpose of a matrix

7. Which of the following is the transpose of the matrix $\begin{bmatrix}a & b & c & d\end{bmatrix}$?

A. $\begin{bmatrix}a\\ b\\ c\\ d\end{bmatrix}$

B. $\begin{bmatrix}d\\ c\\ b\\ a\end{bmatrix}$

C. $\begin{bmatrix}d & c & b & a\end{bmatrix}$

D. $\begin{bmatrix}d & c\\ b & a\end{bmatrix}$

---

8. Determine the transpose of each of the following matrices.

a. $\begin{bmatrix}1\\ 2\\ 3\\ 4\end{bmatrix}$

b. $\begin{bmatrix}1 & 5\\ 9 & 12\end{bmatrix}$

c. $\begin{bmatrix}-7 & 6\\ 13 & 5\\ -1 & 15\end{bmatrix}$

d. $\begin{bmatrix}8 & 9 & 12\\ -11 & 1 & 5\\ 4 & 0 & -2\end{bmatrix}$

---

9. Gabrielle owns a patisserie franchise with two locations in Melbourne.

Her Footscray store's sales data over 3 days for 3 of their most popular pastries are shown in matrix F.

$$F = \begin{array}{c} \begin{array}{ccc} \text{Mon} & \text{Tue} & \text{Wed} \end{array} \\ \begin{bmatrix}34 & 28 & 40\\ 20 & 15 & 29\\ 41 & 35 & 37\end{bmatrix} \end{array} \begin{array}{l} \\ \text{almond croissant}\\ \text{brownie}\\ \text{croissant}\end{array}$$

Her Carlton store also recorded their sales data, but displayed it in the opposite way, in matrix C.

$$C = \begin{array}{c} \begin{array}{ccc} \text{almond croissant} & \text{brownie} & \text{croissant} \end{array} \\ \begin{bmatrix}51 & 33 & 60\\ 38 & 21 & 30\\ 50 & 26 & 62\end{bmatrix} \end{array} \begin{array}{l} \\ \text{Mon}\\ \text{Tue}\\ \text{Wed}\end{array}$$

Display the Carlton store's sales information in matrix C^T, in the same format as the Footscray store's.

Joining it all together

10. For each of the following matrix expressions, determine if the matrix operation is defined. If the operation is defined, evaluate.

a. $2\begin{bmatrix} 11 & 1 & 20 \\ 6 & 8 & 10 \\ -9 & 13 & 4 \\ 4 & 2 & -6 \end{bmatrix} - 3\begin{bmatrix} 8 & -5 & 1 & -6 \\ 7 & -4 & 3 & -7 \\ 6 & -3 & 5 & -8 \end{bmatrix}$

b. $[12 \quad 11 \quad 5 \quad -6 \quad 3] + [4 \quad -9 \quad 10 \quad -1 \quad 2]^T$

c. $([12 \quad 11 \quad 5 \quad -6 \quad 3] + [4 \quad -9 \quad 10 \quad -1 \quad 2])^T$

d. $\begin{bmatrix} 5 & 11 \\ 9 & 7 \\ -10 & 2 \end{bmatrix} - 2\begin{bmatrix} 3 & -5 & 8 \\ 6 & 4 & 1 \end{bmatrix}^T$

11. Determine the values of x and y in the following matrix equations.

a. $\begin{bmatrix} 5 & 10 & 11 \\ 9 & x & -8 \\ -3 & 7 & 5 \end{bmatrix} - \begin{bmatrix} 15 & 9 & 13 \\ 4 & -4 & -2 \\ 8 & 2 & y \end{bmatrix} = \begin{bmatrix} -10 & 1 & -3 \\ 5 & 2 & -6 \\ -11 & 5 & -1 \end{bmatrix}$

b. $4\begin{bmatrix} 1 \\ 7 \\ -9 \\ -3 \\ 5 \end{bmatrix} + 2\begin{bmatrix} 6 \\ x \\ 7 \\ 9 \\ -1 \end{bmatrix} = \begin{bmatrix} 16 \\ 12 \\ -22 \\ y \\ 18 \end{bmatrix}$

c. $x\begin{bmatrix} 4 & 10 \\ 7 & 8 \end{bmatrix}^T = \begin{bmatrix} y & -10.5 \\ -15 & -12 \end{bmatrix}$

d. $\left(3\begin{bmatrix} 1 & 6 \\ 5 & -3 \\ -4 & 2 \end{bmatrix} - \begin{bmatrix} 7 & -9 \\ 8 & -10 \\ x & 11 \end{bmatrix}\right)^T = \begin{bmatrix} -4 & 7 & 6 \\ 27 & y & -5 \end{bmatrix}$

12. Oxbridge University retains data on the number of commencing students each year. The number of new local and international students in 2020 for 4 select courses are shown in matrix O_{2020}.

$$O_{2020} = \begin{array}{c} \begin{array}{cccc} \text{medicine} & \text{literature} & \text{law} & \text{philosophy} \end{array} \\ \begin{bmatrix} 230 & 240 & 280 & 150 \\ 220 & 150 & 260 & 120 \end{bmatrix} \end{array} \begin{array}{l} \\ \text{local} \\ \text{international} \end{array}$$

a. In 2021, Oxbridge University's enrolments were 20% higher than in 2020. Using scalar multiplication, calculate the matrix O_{2021}, displaying Oxbridge's 2021 local and international student enrolments in these 4 courses.

b. In total, how many international students started studying law at Oxbridge over 2020 and 2021?

c. Comparable data was also collected by Camford University for their 2021 enrolments. The data is shown in matrix C_{2021}.

$$C_{2021} = \begin{array}{c} \begin{array}{cc} \text{local} & \text{international} \end{array} \\ \begin{bmatrix} 305 & 350 \\ 372 & 231 \\ 320 & 318 \\ 266 & 175 \end{bmatrix} \end{array} \begin{array}{l} \\ \text{medicine} \\ \text{literature} \\ \text{law} \\ \text{philosophy} \end{array}$$

Determine the transpose of C_{2021}, to display this data in the same form as Oxbridge's data.

d. Calculate ${C_{2021}}^T - O_{2021}$ to find the difference between the 2021 enrolment numbers of the two universities.

e. How many more local students in 2021 started studying philosophy at Camford than at Oxbridge?

Back to contents

Exam practice

13. Consider the following four matrix expressions.

$\begin{bmatrix} 8 \\ 12 \end{bmatrix} + \begin{bmatrix} 4 \\ 2 \end{bmatrix}$ $\quad \begin{bmatrix} 8 \\ 12 \end{bmatrix} + \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}$ $\quad \begin{bmatrix} 8 & 0 \\ 12 & 0 \end{bmatrix} + \begin{bmatrix} 4 \\ 2 \end{bmatrix}$ $\quad \begin{bmatrix} 8 & 0 \\ 12 & 0 \end{bmatrix} + \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}$

How many of these four matrix expressions are defined?

A. 0 **B.** 1 **C.** 2
D. 3 **E.** 4

VCAA 2019 Exam 1 Matrices Q1

87% of students answered this question correctly.

---

14. The transpose of $\begin{bmatrix} 2 & 7 & 10 \\ 13 & 19 & 8 \end{bmatrix}$ is

A. $\begin{bmatrix} 13 & 19 & 8 \\ 2 & 7 & 10 \end{bmatrix}$ **B.** $\begin{bmatrix} 10 & 7 & 2 \\ 8 & 19 & 13 \end{bmatrix}$ **C.** $\begin{bmatrix} 2 & 13 \\ 7 & 19 \\ 10 & 8 \end{bmatrix}$

D. $\begin{bmatrix} 13 & 2 \\ 19 & 7 \\ 8 & 10 \end{bmatrix}$ **E.** $\begin{bmatrix} 8 & 10 \\ 19 & 7 \\ 13 & 2 \end{bmatrix}$

VCAA 2016 Exam 1 Matrices Q1

81% of students answered this question correctly.

---

15. The following table shows information about two matrices, A and B.

matrix	order	rule
A	3×3	$a_{ij} = 2i + j$
B	3×3	$b_{ij} = i - j$

The element in row i and column j of matrix A is a_{ij}.
The element in row i and column j of matrix B is b_{ij}.
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.** $\begin{bmatrix} 3 & 6 & 3 \\ 6 & 3 & 9 \\ 3 & 9 & 3 \end{bmatrix}$

VCAA 2017 Exam 1 Matrices Q6

58% of students answered this question correctly.

Questions from multiple lessons

Data analysis

16. The statistical analysis of a set of bivariate data involving variables x and y resulted in the information displayed in the following table.

mean	$\bar{x} = 17.05$	$\bar{y} = 19.92$
standard deviation	$s_x = 1.25$	$s_y = 1.83$
equation of the least squares line	$y = 15.146 + 0.28x$	

Using this information, the value of the correlation coefficient r for this set of bivariate data is closest to

A. 0.15 **B.** 0.19 **C.** 0.33 **D.** 0.41 **E.** 0.44

Adapted from VCAA 2018 Exam 1 Data analysis Q13

Back to contents

Recursion and financial modelling

17. Lindsay took out a loan to buy a new house. Information about her first loan repayment is shown in the following amortisation table.

repayment number	repayment	interest	principal reduction	balance of loan
0	0.00	0.00	0.00	230 000.00
1	1000.00	690.00	310.00	229 310.00
2	1000.00			

How much interest does Lindsay pay in repayment number 2?

A. \$687.00 **B.** \$687.93 **C.** \$688.12 **D.** \$690.00 **E.** \$692.08

Adapted from VCAA 2017NH Exam 1 Recursion and financial modelling Q21

Matrices *Year 11 content*

18. A small ticketing company sells tickets for a variety of events; music concerts (M), theatre (T), sporting events (S), and comedy shows (C).

Matrix N contains the number of each type of booking received last month.

$$N = \begin{bmatrix} 112 \\ 46 \\ 75 \\ 53 \end{bmatrix} \begin{array}{l} \mathrm{M} \\ \mathrm{T} \\ \mathrm{S} \\ \mathrm{C} \end{array}$$

a. What is the order of matrix N? (1 MARK)

b. A booking fee per ticket is collected by the company with each ticket sale.

Matrix F contains the booking fee per ticket for each type of event.

$$F = \begin{array}{c} \begin{array}{cccc} \mathrm{M} & \mathrm{T} & \mathrm{S} & \mathrm{C} \end{array} \\ \begin{bmatrix} 8 & 10 & 7 & 5 \end{bmatrix} \end{array}$$

i. Calculate the matrix product $R = F \times N$. (1 MARK)

ii. What information is represented by matrix R? (1 MARK)

Adapted from VCAA 2016 Exam 2 Matrices Q1

Back to contents

7C Advanced operations with matrices

STUDY DESIGN DOT POINT

- 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)

7A 7B 7C 7D 7E 7F 7G 7H 7I

KEY SKILLS

During this lesson, you will be:

- defining matrix products
- calculating a matrix product
- using a summing matrix
- calculating a matrix power.

KEY TERMS

- Matrix product
- Post-multiplication
- Pre-multiplication
- Summing matrix
- Matrix power

A key component of matrix arithmetic involves multiplying matrices. In order to multiply two matrices, the matrix product must be defined. Matrix multiplication has several applications, including summing the rows and columns in matrices and raising matrices to a power.

Defining matrix products

A **matrix product** is the resulting matrix when two or more matrices are multiplied. **Post-multiplication** is the multiplication of one matrix after another matrix. For example, the product AB can be defined as matrix A post-multiplied by matrix B.

Pre-multiplication is the multiplication of one matrix before another matrix. For example, the product AB can also be defined as matrix B pre-multiplied by matrix A.

Not all matrix multiplications can be performed. For a matrix product to be defined, the number of columns in the first matrix must equal the number of rows in the second matrix.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad B = \begin{bmatrix} e \\ f \end{bmatrix}$$

$$AB = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \times \begin{bmatrix} e \\ f \end{bmatrix} \qquad BA = \begin{bmatrix} e \\ f \end{bmatrix} \times \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$\text{Order: } 2 \times \underbrace{2 \quad 2}_{\text{equal}} \times 1 \qquad \text{Order: } 2 \times \underbrace{1 \quad 2}_{\text{not equal}} \times 1$$

The product AB is defined since the number of columns in matrix A equals the number of rows in matrix B. The product BA is not defined since the number of columns in matrix B is not equal to the number of rows in matrix A.

Matrix multiplication is therefore not commutative (reversible): $AB \neq BA$.

If the matrix product is defined, the order of the matrix product will be equal to the number of rows in the first matrix and the number of columns in the second matrix.

$$AB = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \times \begin{bmatrix} e \\ f \end{bmatrix}$$

$$\text{Order: } \underbrace{2 \times 2 \quad 2 \times 1}_{2 \times 1}$$

The order of the matrix product AB is 2×1.

Worked example 1

Determine whether the following matrix products are defined. If defined, determine its order.

$K = \begin{bmatrix} 3 & -2 \\ 0 & 5 \end{bmatrix} \quad L = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$

a. KL

Explanation

Step 1: Determine the order of each matrix.

Matrix K has an order of 2×2.

Matrix L has an order of 2×1.

Step 2: Determine whether the matrix product is defined.

The number of columns of matrix K must equal the number of rows of matrix L.

$$KL = \begin{bmatrix} 3 & -2 \\ 0 & 5 \end{bmatrix} \times \begin{bmatrix} -1 \\ 3 \end{bmatrix}$$

Order: $2 \times 2 \quad 2 \times 1$

equal

The matrix product is defined.

Step 3: Determine the order of the matrix product.

The order of the matrix product is equal to the number of rows of matrix K and the number of columns of matrix L.

$$KL = \begin{bmatrix} 3 & -2 \\ 0 & 5 \end{bmatrix} \times \begin{bmatrix} -1 \\ 3 \end{bmatrix}$$

Order: $2 \times 2 \quad 2 \times 1$

2×1

Answer

KL is defined. The order of KL is 2×1.

b. LK

Explanation

Step 1: Determine the order of each of the matrices.

Matrix K has an order of 2×2.

Matrix L has an order of 2×1.

Step 2: Determine whether the matrix product is defined.

The number of columns of matrix L must equal the number of rows of matrix K.

$$LK = \begin{bmatrix} -1 \\ 3 \end{bmatrix} \times \begin{bmatrix} 3 & -2 \\ 0 & 5 \end{bmatrix}$$

Order: $2 \times 1 \quad 2 \times 2$

not equal

Answer

LK is not defined.

Calculating a matrix product

Multiplying two matrices involves both multiplication and addition of elements. To multiply two matrices together, elements in specific rows and columns must be multiplied and summed together.

$$AB = C$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \times \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

To find c_{mn}, multiply each element in the m^{th} row of matrix A by its corresponding element in the n^{th} column of matrix B. Add these products together to find the value of c_{mn}.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \times \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} a \times e + b \times g & a \times f + b \times h \\ c \times e + d \times g & c \times f + d \times h \end{bmatrix}$$

Back to contents

For example, to find c_{12}, the elements in the 1st row of matrix A are multiplied by their corresponding elements in the 2nd column of matrix B. These numbers are then added.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \times \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} a \times e + b \times g & a \times f + b \times h \\ c \times e + d \times g & c \times f + d \times h \end{bmatrix}$$

Worked example 2

Consider the following matrices.

$$P = \begin{bmatrix} -1 & 3 \\ 0 & 5 \end{bmatrix} \quad Q = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

Evaluate the matrix product PQ.

Explanation – Method 1: By hand

Step 1: Determine whether the product matrix is defined.

$$PQ = \begin{bmatrix} -1 & 3 \\ 0 & 5 \end{bmatrix} \times \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

Order: $2 \times \underbrace{2 \quad 2}_{\text{equal}} \times 1$

The matrix product is defined.

Step 2: Determine the order of the product matrix.

The order is equal to the number of rows of matrix P and the number of columns of matrix Q.

The order of PQ is 2×1.

Step 3: Set up the equation using matrices.

Ensure PQ has the correct order.

$$PQ = \begin{bmatrix} -1 & 3 \\ 0 & 5 \end{bmatrix}\begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} \\ \end{bmatrix}$$

Step 4: Find the elements of PQ.

$$PQ = \begin{bmatrix} -1 & 3 \\ 0 & 5 \end{bmatrix}\begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} -1 \times 2 + 3 \times 4 \\ 0 \times 2 + 5 \times 4 \end{bmatrix}$$

Step 5: Sum the products to find the final elements of matrix PQ.

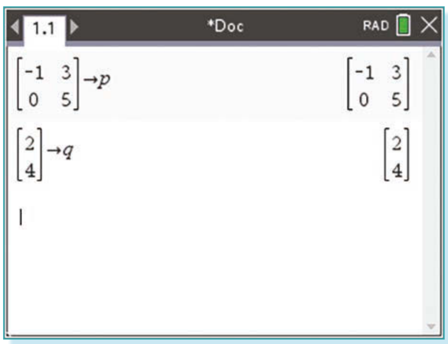
Explanation – Method 2: TI-Nspire

Step 1: From the home screen, select '1: New' → '1: Add Calculator'.

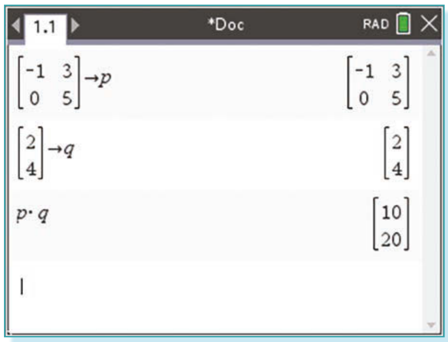
Step 2: Create matrix P and Q.

Create a 2×2 matrix and enter the values for P.
Press ctrl + var + 'p' to store the matrix as 'p'.
Press enter.

Create a 2×1 matrix and enter the values for Q.
Press ctrl + var + 'q' to store the matrix as 'q'.
Press enter.



Step 3: Type '$p \times q$' to calculate PQ and press enter.



Continues →

Back to contents

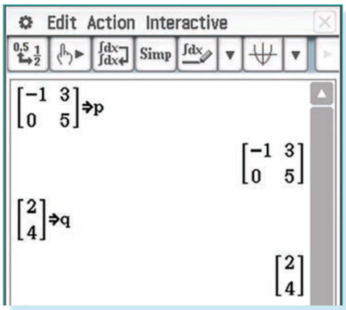
Explanation – Method 3: Casio ClassPad

Step 1: From the main menu, tap √α Main.

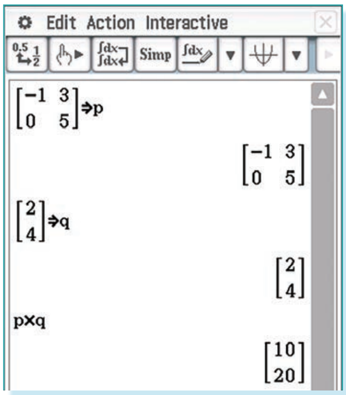
Step 2: Create matrix P and Q.

Create a 2×2 matrix and enter the values for P. On the touch-screen keyboard, go to var and tap ⇒ + 'p' to store the matrix as 'p'. Press EXE.

Create a 2×1 matrix and enter the values for Q. On the touch-screen keyboard, go to var and press ⇒ + 'q' to store the matrix as 'q'. Press EXE.



Step 3: Type '$p \times q$' to calculate PQ.



Answer – Method 1, 2 and 3

$$PQ = \begin{bmatrix}10\\20\end{bmatrix}$$

Using a summing matrix

A **summing matrix** is a row or column matrix that consists of only the number 1 for each element, and is used to find the sum of either the rows or columns of another matrix.

$$\begin{bmatrix}1\\1\end{bmatrix} \quad \begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix}$$

The sum of the rows of a matrix can be found by post-multiplying a column summing matrix. The number of rows in the summing matrix must be equal to the number of columns in the matrix to be summed, otherwise the matrix product will be undefined.

For example, to find the sum of the rows in matrix E, the post-multiplication of a 3×1 column summing matrix would be required, since it has 3 columns.

$$E = \begin{bmatrix}1 & 2 & 3\\4 & 5 & 6\end{bmatrix}$$

$$\begin{bmatrix}1 & 2 & 3\\4 & 5 & 6\end{bmatrix} \times \begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}1 \times 1 + 2 \times 1 + 3 \times 1\\4 \times 1 + 5 \times 1 + 6 \times 1\end{bmatrix}$$

$$= \begin{bmatrix}6\\15\end{bmatrix}$$

The sum of the columns of a matrix can be found by pre-multiplying a row summing matrix. The number of columns in the summing matrix must be equal to the number of rows in the matrix to be summed, otherwise the matrix product will be undefined.

For example, to find the sum of the columns in matrix F, the pre-multiplication of a 1×4 row summing matrix would be required, since it has 4 rows.

$$F = \begin{bmatrix}1 & 2\\4 & 5\\3 & 3\\7 & 6\end{bmatrix}$$

$$\begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix} \times \begin{bmatrix}1 & 2\\4 & 5\\3 & 3\\7 & 6\end{bmatrix} = \begin{bmatrix}1 \times 1 + 1 \times 4 + 1 \times 3 + 1 \times 7 & 1 \times 2 + 1 \times 5 + 1 \times 3 + 1 \times 6\end{bmatrix}$$

$$= \begin{bmatrix}15 & 16\end{bmatrix}$$

Back to contents

Worked example 3

$S = \begin{bmatrix} -1 & 0 & 6 \\ 2 & -3 & 4 \end{bmatrix}$

Use an appropriate summing matrix to

a. sum the rows of matrix S.

Explanation

Step 1: Construct a summing matrix.

To sum the rows of a matrix, a column summing matrix is required.

The number of rows in the summing matrix should equal the number of columns in matrix S.

Matrix S has 3 columns. The summing matrix must be a 3×1 matrix.

$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

Step 2: Post-multiply matrix S with the summing matrix.

$\begin{bmatrix} -1 & 0 & 6 \\ 2 & -3 & 4 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

Answer

$\begin{bmatrix} 5 \\ 3 \end{bmatrix}$

b. sum the columns of matrix S.

Explanation

Step 1: Construct a summing matrix.

To sum the columns of a matrix, a row summing matrix is required.

The number of columns in the summing matrix should equal the number of rows in matrix S.

Matrix S has 2 rows. The summing matrix must be a 1×2 matrix.

$\begin{bmatrix} 1 & 1 \end{bmatrix}$

Step 2: Pre-multiply matrix S with the summing matrix.

$\begin{bmatrix} 1 & 1 \end{bmatrix} \times \begin{bmatrix} -1 & 0 & 6 \\ 2 & -3 & 4 \end{bmatrix}$

Answer

$\begin{bmatrix} 1 & -3 & 10 \end{bmatrix}$

Calculating a matrix power

A **matrix power** is the resultant product when a matrix is raised to an index or power. Only square matrices can be raised to a power, since the number of columns in the first matrix must equal the number of rows in the second matrix for a matrix product to be defined. The order of the resultant matrix product will be the same as the original matrix.

It is important to maintain the normal order of operations that are used with standard numbers when working with matrices. For example, if A and B are square matrices, then $(A + B)^2 = (A + B)(A + B)$ and $(AB)^3 = (AB)(AB)(AB)$.

Worked example 4

Consider the following matrices.

$$A = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad B = \begin{bmatrix} 2 & 1 \\ 8 & 9 \end{bmatrix} \quad C = \begin{bmatrix} 2 & 3 & 5 \end{bmatrix}$$

a. Calculate B^2.

Explanation – Method 1: By hand

Step 1: Determine if the matrix power is defined.

Matrix B has an order of 2×2. It is a square matrix so it can be raised to a power.

Step 2: Evaluate the matrix power.

$$B^2 = \begin{bmatrix} 2 & 1 \\ 8 & 9 \end{bmatrix}^2$$

$$= \begin{bmatrix} 2 & 1 \\ 8 & 9 \end{bmatrix} \times \begin{bmatrix} 2 & 1 \\ 8 & 9 \end{bmatrix}$$

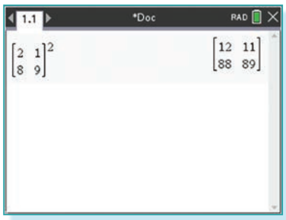
$$= \begin{bmatrix} 2 \times 2 + 1 \times 8 & 2 \times 1 + 1 \times 9 \\ 8 \times 2 + 9 \times 8 & 8 \times 1 + 9 \times 9 \end{bmatrix}$$

Explanation – Method 2: Ti-Nspire

Step 1: From the home screen, select '1: New' → '1: Add Calculator'.

Step 2: Create a 2×2 matrix and enter the values for B.

Step 3: Type ^ + '2' and press enter.

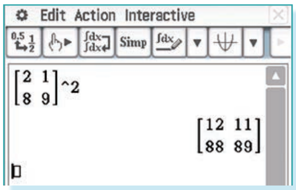


Explanation – Method 3: Casio ClassPad

Step 1: From the main menu, tap √α Main.

Step 2: Create a 2×2 matrix and enter the values for B.

Step 3: Type ^ + '2' and press EXE.



Answer – Method 1, 2 and 3

$$\begin{bmatrix} 12 & 11 \\ 88 & 89 \end{bmatrix}$$

b. Determine whether $(AC)^2$ is defined.

Explanation

Step 1: Determine if the matrix product is defined.

Matrix A has an order of 3×1 and C an order of 1×3.

The matrix product AC is defined since the number of columns in matrix A is equal to the number of rows in matrix C.

Step 2: Determine if the matrix being raised to a power is a square matrix.

The order of the matrix product AC will be 3×3.

This is a square matrix so it can be raised to a power.

Answer

$(AC)^2$ is defined.

Back to contents

Exam question breakdown

VCAA 2016 Exam 1 Matrices Q4

The table shows the number of each type of coin saved in a money box.

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. $\begin{bmatrix} 5 & 10 & 20 & 50 \\ 15 & 32 & 48 & 24 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$

Explanation

Step 1: Determine the information to be displayed in the matrix product.

The matrix product must display the total number of coins and the total value of these coins.

Step 2: Determine the values needed to find the total number of coins.

The number of coins can be represented in the following row matrix.

$\begin{bmatrix} 15 & 32 & 48 & 24 \end{bmatrix}$

The total number of coins is the sum of these values. The sum can be found by post-multiplying the matrix by a 4×1 column summing matrix:

$$\begin{bmatrix} 15 & 32 & 48 & 24 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Step 3: Determine the values needed to calculate the total value of the coins.

The total value of the coins can be found by multiplying the number of each type of coin by its value and then finding the sum.

This can be represented by the following matrix multiplication:

$$\begin{bmatrix} 15 & 32 & 48 & 24 \end{bmatrix}\begin{bmatrix} 5 \\ 10 \\ 20 \\ 50 \end{bmatrix}$$

Step 4: Combine the matrix multiplications.

Both expressions post-multiply the row matrix of the number of coins by a column matrix.

The column matrices in each expression can be combined into a 4×2 matrix.

$$\begin{bmatrix} 15 & 32 & 48 & 24 \end{bmatrix}\begin{bmatrix} 1 & 5 \\ 1 & 10 \\ 1 & 20 \\ 1 & 50 \end{bmatrix}$$

Answer

B

34% of students answered this question correctly.

31% of students incorrectly chose option A. Option A multiplied the value of each type of coin by the number of those coins and summed these values, resulting in the total value of the coins. However, the matrix product needed to calculate two quantities: the total value and the total number of the coins. Only option B calculates the required information.

7C Questions

Defining matrix products

1. Which of the following matrix products is defined?

$A = \begin{bmatrix} 2 \\ 6 \\ 3 \end{bmatrix}$ $B = \begin{bmatrix} 4 & 3 \\ -1 & 2 \end{bmatrix}$ $C = \begin{bmatrix} 3 & 8 & 4 \\ 3 & 1 & 6 \\ 9 & 2 & 6 \end{bmatrix}$ $D = \begin{bmatrix} 2 & 5 & 4 \\ 1 & 2 & 3 \end{bmatrix}$

A. AB **B.** BC **C.** $(BD)C$ **D.** $A(BC)$

2. Consider the following matrices.

$A = \begin{bmatrix} 9 & 7 & -3 \\ 1 & -6 & 5 \end{bmatrix}$ $B = \begin{bmatrix} 1 & 7 \\ -5 & 0 \end{bmatrix}$ $C = \begin{bmatrix} 3 & 8 \end{bmatrix}$ $D = \begin{bmatrix} 4 & 9 & 2 \end{bmatrix}$ $E = \begin{bmatrix} 0 \\ -4 \end{bmatrix}$ $F = \begin{bmatrix} 4 \\ 1 \\ -6 \end{bmatrix}$

Determine if the following matrix products are defined. If defined, state its order.

a. AB **b.** BA **c.** AD **d.** AF

e. CB **f.** EB

3. Using the matrices from question 2, determine the order of the following matrix products.

a. $C(BA)$

A. 1×1
B. 1×3
C. 2×3
D. $C(BA)$ is undefined

b. $(DF)(CA)$

A. 1×1
B. 1×3
C. 2×3
D. $(DF)(CA)$ is undefined

Calculating a matrix product

4. Calculate the following matrix product.

$\begin{bmatrix} -1 & 0 \\ 2 & -3 \end{bmatrix} \times \begin{bmatrix} 2 & 1 \\ 4 & 1 \end{bmatrix}$

A. $\begin{bmatrix} -2 \\ -8 \end{bmatrix}$ **B.** $\begin{bmatrix} -6 & 0 \\ 12 & -6 \end{bmatrix}$ **C.** $\begin{bmatrix} -2 & 0 \\ 8 & -3 \end{bmatrix}$ **D.** $\begin{bmatrix} -2 & -1 \\ -8 & -1 \end{bmatrix}$

5. Calculate the following matrix products.

a. $\begin{bmatrix} 3 & 2 \end{bmatrix} \times \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

b. $\begin{bmatrix} 2 & -1 \end{bmatrix} \times \begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix}$

c. $\begin{bmatrix} 1 & 3 \\ 2 & 5 \end{bmatrix} \times \begin{bmatrix} 2 \\ 0 \end{bmatrix}$

d. $\begin{bmatrix} 2 & 0 \\ 5 & -1 \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 2 & 4 \end{bmatrix}$

e. $\begin{bmatrix} 2 & 7 & 6 \\ 1 & 5 & -4 \end{bmatrix} \times \begin{bmatrix} 3 & 4 \\ 1 & -3 \\ 2 & 5 \end{bmatrix}$

f. $\begin{bmatrix} 1 & 5 & -9 \\ 3 & 4 & 8 \\ 7 & 3 & 2 \end{bmatrix} \times \begin{bmatrix} 6 & 1 & 9 \\ 5 & -2 & 4 \\ 7 & 2 & 0 \end{bmatrix}$

6. In a soccer tournament, team Soccerolo (S) and Edroloball (E) played in 20 games and their wins (W), draws (D) and losses (L) are shown in matrix G. A team received 2 points if they won, 1 point if they drew and 0 points if they lost, as shown in matrix H.

$$G = \begin{array}{c} \begin{array}{ccc} \text{W} & \text{D} & \text{L} \end{array} \\ \begin{bmatrix} 11 & 4 & 5 \\ 8 & 3 & 9 \end{bmatrix} \end{array} \begin{array}{l} \\ \text{S} \\ \text{E} \end{array} \qquad H = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \begin{array}{l} \text{W} \\ \text{D} \\ \text{L} \end{array}$$

a. Calculate GH.

b. What does the matrix GH represent?

c. Which team had the most points at the end of the tournament and how many points did they have?

Back to contents

7. Tess runs a social media page and is paid based on the number of likes and comments she gets on each post. Tess makes 5 cents per like and 10 cents per comment, as shown in matrix J. At the end of the week, the number of likes (L) and comments (C) on each of her four posts is shown in matrix K.

$$J = \begin{bmatrix} 5 & 10 \end{bmatrix} \quad K = \begin{array}{c} \text{post} \\ \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 27 & 31 & 56 & 12 \\ 12 & 19 & 11 & 21 \end{bmatrix} \end{array} \begin{array}{l} \\ \\ \text{L} \\ \text{C} \end{array}$$

a. What does element k_{13} represent?

b. Calculate JK.

c. How much money, in dollars, did Tess make in that week?

d. Which post was most profitable for Tess?

---

8. Mac needs three different cheeses, cheddar (C), gruyere (G) and parmesan (P), to cook his famous pasta dish. The price of each cheese pack from two different stores is shown in matrix L.

$$L = \begin{array}{c} \begin{array}{ccc} \text{C} & \text{G} & \text{P} \end{array} \\ \begin{bmatrix} 1.75 & 6.20 & 3.40 \\ 1.30 & 6.95 & 2.80 \end{bmatrix} \end{array} \begin{array}{l} \\ \text{Cheeseworld} \\ \text{Cheesetopia} \end{array}$$

a. Mac needs 5 packs of cheddar, 2 packs of gruyere and 3 packs of parmesan for his dish. Construct a matrix, M, to represent this information and calculate LM.

b. Which store should Mac buy his cheese from?

c. Mac finds out that Cheeseworld is having a sale on parmesan cheese and the price has dropped to \$2.70. Mac decides he will buy his cheese from Cheeseworld. Did he make the right decision? Justify.

Using a summing matrix

9. Which of the following is the correct summing matrix to sum the columns of matrix A?

$$A = \begin{bmatrix} 5 & 4 & 5 & 1 \\ 2 & 1 & 9 & 2 \\ 3 & 2 & 2 & 4 \end{bmatrix}$$

A. $\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$ **B.** $\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}$ **C.** $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ **D.** $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$

---

10. Construct a summing matrix that

a. sums the columns of $\begin{bmatrix} 5 & 2 \\ 1 & 0 \\ 9 & 3 \\ 7 & 4 \end{bmatrix}$.

b. sums the rows of $\begin{bmatrix} 7 & 0 & 3 & 3 \\ -2 & 6 & 0 & 1 \\ 6 & 1 & 8 & 4 \end{bmatrix}$.

---

11. Heston has four dessert stores and wanted to know which store was doing better in the past month. The matrix shows how many desserts each store sold in the past month.

$$\begin{array}{c} \begin{array}{ccc} \text{strudels} & \text{macaroons} & \text{tarts} \end{array} \\ \begin{bmatrix} 20 & 24 & 24 \\ 16 & 30 & 21 \\ 18 & 15 & 28 \\ 22 & 24 & 19 \end{bmatrix} \end{array} \begin{array}{l} \\ \text{store 1} \\ \text{store 2} \\ \text{store 3} \\ \text{store 4} \end{array}$$

a. Construct a matrix that, when multiplied with the one provided, will find the total number of desserts sold by each store in the past month.

b. Show, using matrix calculations, which store sold the most desserts in the past month.

Back to contents

c. Construct a matrix that, when multiplied with the one provided, will find the total number of each type of dessert sold across all stores.

d. Show, using matrix calculations, which type of dessert is the most popular across all stores.

Calculating a matrix power

12. Which of the following matrices can be raised to a power?

A. $\begin{bmatrix} 4 & 2 & 8 \end{bmatrix}$ **B.** $\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$ **C.** $\begin{bmatrix} 2 & 4 \\ -1 & -3 \end{bmatrix}$ **D.** $\begin{bmatrix} 0 & 2 \end{bmatrix}$

---

13. For the following matrices:

$$A = \begin{bmatrix} 4 & 5 \\ 8 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 3 & 1 & 0 \\ 0 & 6 & 2 \end{bmatrix} \quad C = \begin{bmatrix} 9 & 3 \\ 4 & 7 \end{bmatrix} \quad D = \begin{bmatrix} 2 & 4 \\ 7 & 1 \\ 5 & 3 \end{bmatrix}$$

Determine if each of the following expressions is defined.

a. A^2 b. B^2 c. BD^2 d. $(BD)^2$ e. $(BD)(AC)^3$

---

14. Use the matrices shown to evaluate the following expressions.

$$A = \begin{bmatrix} 3 & 0 \\ 2 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} \quad C = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

a. C^5 b. $A^2 - 3B$ c. $AC^2 - B^2$

Joining it all together

15. Consider the following matrices.

$$J = \begin{bmatrix} 3 & 1 \\ 7 & 1 \end{bmatrix} \quad K = \begin{bmatrix} 2 & 5 & 1 \\ 4 & 3 & 6 \end{bmatrix} \quad L = \begin{bmatrix} 9 & 3 \\ 2 & 5 \\ 4 & 6 \end{bmatrix}$$

a. Is the matrix product KL defined? If defined, state its order.

b. Which of the matrices can be raised to a power?

c. Is the matrix expression $J^2 - (KL)^2$ defined? If defined, evaluate.

d. Evaluate $(J + KL)^2$.

---

16. Fin Demali is an avid bird watcher. He tracks the population of three different species in two areas. The recorded population (in thousands) of mockingjays (M), finches (F) and woodpeckers (W) in 2022 are shown in matrix B.

$$B = \begin{array}{c} \begin{matrix} \text{M} & \text{F} & \text{W} \end{matrix} \\ \begin{bmatrix} 23 & 12 & 7 \\ 18 & 14 & 8 \end{bmatrix} \end{array} \begin{matrix} \\ \text{area 1} \\ \text{area 2} \end{matrix}$$

a. Construct a summing matrix to sum the number of birds in each area and use this to calculate the total number of birds in each area.

b. Fin predicts the bird population will change in 2023. The population in 2023 can be predicted by post-multiplying matrix C with matrix B. Create the matrix showing the new populations (in thousands) in 2023.

$$C = \begin{bmatrix} 0.8 & 0 \\ 0 & 1.3 \end{bmatrix}$$

c. What is the total bird population predicted to be in 2023?

d. By what percentage did the bird populations change in each area between 2022 and 2023?

Back to contents

Exam practice

17. A school canteen sells pies (P), rolls (R) and sandwiches (S).

The number of each item sold over three school weeks is shown in matrix M.

$$\begin{array}{cc} & \begin{array}{ccc} \mathrm{P} & \mathrm{R} & \mathrm{S} \end{array} \\ M = & \begin{bmatrix} 35 & 24 & 60 \\ 28 & 32 & 43 \\ 32 & 30 & 56 \end{bmatrix} \end{array} \begin{array}{l} \\ \text{week 1} \\ \text{week 2} \\ \text{week 3} \end{array}$$

In total, how many sandwiches were sold in these three weeks? (1 MARK)

VCAA 2017 Exam 2 Matrices Q1a

89% of students answered this question correctly.

---

18. The matrix product $[4 \quad 2 \quad 0] \times \begin{bmatrix} 4 \\ 12 \\ 8 \end{bmatrix}$ is equal to

A. $[144]$

B. $\begin{bmatrix} 16 \\ 24 \\ 0 \end{bmatrix}$

C. $4 \times [1 \quad 2 \quad 0] \times \begin{bmatrix} 1 \\ 12 \\ 8 \end{bmatrix}$

D. $2 \times [2 \quad 1 \quad 0] \times \begin{bmatrix} 2 \\ 6 \\ 4 \end{bmatrix}$

E. $4 \times [2 \quad 1 \quad 0] \times \begin{bmatrix} 2 \\ 6 \\ 4 \end{bmatrix}$

VCAA 2018 Exam 1 Matrices Q2

68% of students answered this question correctly.

Questions from multiple lessons

Recursion and financial modelling

19. The following graph shows the value, V_n, of the new Pineapple phone as it depreciates over a period of six months. Which one of the following depreciation situations does this graph best represent?

A. Reducing balance depreciation with a decrease in depreciation rate after 4 months.

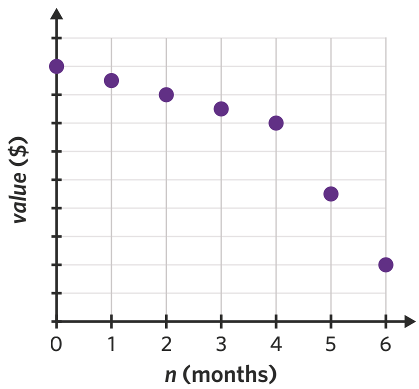
B. Flat rate depreciation with an increase in depreciation rate after 4 months.

C. Unit cost depreciation with constant depreciation rate.

D. Flat rate depreciation with a decrease in depreciation rate after 4 months.

E. Reducing balance depreciation with an increase in depreciation rate after 4 months.

Adapted from VCAA 2018 Exam 1 Recursion and financial modelling Q20



Data analysis

20. Bill has owned and run a small hotel by the beach for 15 years and wants to maximise the number of bookings at his hotel during summer. He collects data on the number of bookings each summer over the last 15 years, which is modelled on the time series plot shown.

The time series plot exhibits

A. seasonality with irregular fluctuations.

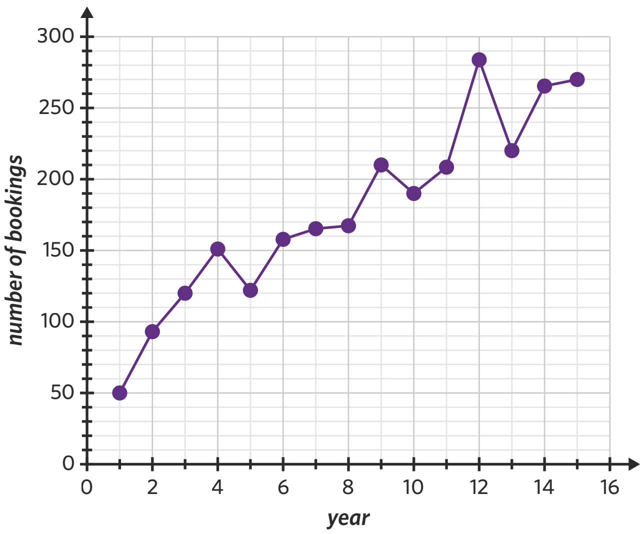
B. seasonality with an increasing trend.

C. seasonality with an increasing trend and irregular fluctuations.

D. irregular fluctuations only.

E. an increasing trend with irregular fluctuations.

Adapted from VCAA 2016 Exam 1 Data analysis Q13



Back to contents

Recursion and financial modelling

21. Dani withdraws \$20 000 from her account to purchase some dragon eggs. For tax purposes, she plans to depreciate the value of her eggs using the reducing balance method. The value of Dani's eggs, in dollars, after n years, D_n, can be modelled by the recurrence relation shown.

$D_0 = 20\ 000, \quad D_{n+1} = R \times D_n$

a. For the first two years of reducing balance depreciation, the value of R is 0.77.

What is the annual rate of depreciation during these two years? (1 MARK)

b. For the next three years of reducing balance depreciation, the annual rate of depreciation of the value of Dani's eggs is changed to 13.5%.

What is the value of the eggs 5 years after they were purchased? Round your answer to the nearest dollar. (2 MARKS)

Adapted from VCAA 2018 Exam 2 Recursion and financial modelling Q5

Back to contents

7D Inverse matrices

STUDY DESIGN DOT POINT

- inverse of a matrix, its determinant, and the condition for a matrix to have an inverse

7A 7B 7C 7D 7E 7F 7G 7H 7I

KEY SKILLS

During this lesson, you will be:

- calculating the determinant of a matrix
- calculating the inverse of a matrix
- solving simultaneous equations using matrix equations.

KEY TERMS

- Determinant
- Inverse matrix
- Singular matrix

An important feature of matrices is their ability to be incorporated into equations. Matrix operations and inverse matrices can be used to solve for unknown variables. Since matrices cannot be divided, an inverse matrix can instead be found and then used when solving equations.

Calculating the determinant of a matrix

The **determinant** is a number associated with a matrix which determines whether the inverse of a matrix is defined. It can only be calculated for square matrices. For a matrix to have an inverse, its determinant must not equal zero.

The determinant of matrix A is denoted as $\det(A)$.

For 2×2 matrices, a formula can be used to find the determinant.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$\det(A) = ad - bc$$

While the formula can be used to find the determinant of 2×2 matrices, a CAS can be used to find the determinant of square matrices of any order.

Worked example 1

$$A = \begin{bmatrix} 4 & 2 & 13 \\ 6 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix} \quad B = \begin{bmatrix} 2 & 2 \\ 3 & 1 \end{bmatrix}$$

a. Find the determinant of B.

Explanation

Calculate the determinant.

$$\begin{aligned} \det(B) &= ad - bc \\ &= 2 \times 1 - 2 \times 3 \end{aligned}$$

Answer

-4

Continues →

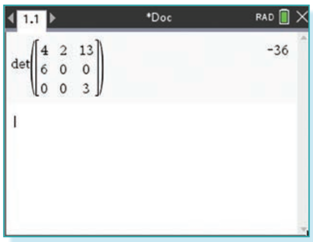
b. Use a CAS to find the determinant of A.

Explanation – Method 1: TI-Nspire

Step 1: From the home screen, select '1: New' → '1: Add Calculator'.

Step 2: Press menu. Select '7: Matrix & Vector' → '3: Determinant'.

Step 3: Enter matrix A. Press enter.

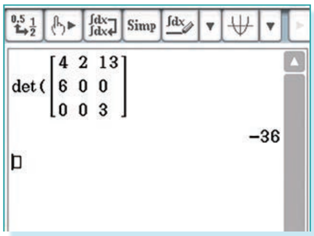


Explanation – Method 2: Casio ClassPad

Step 1: From the main menu, tap √α Main.

Step 2: Tap 'Action' → 'Matrix' → 'Calculation' → 'det'.

Step 3: Enter matrix A. Press EXE.



Answer – Method 1 and 2

-36

Calculating the inverse of a matrix

If a matrix's determinant is not equal to zero, a square matrix will have an inverse matrix. When an **inverse matrix** is pre-multiplied or post-multiplied with its matrix, the result is the identity matrix, I.

The inverse matrix is denoted by A^{-1}.

$A \times A^{-1} = A^{-1} \times A = I$

For example, if

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$AA^{-1} = \begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix} \times \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ -\frac{3}{8} & \frac{1}{8} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$A^{-1}A = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ -\frac{3}{8} & \frac{1}{8} \end{bmatrix} \times \begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

The identity matrix works in a similar way to the number 1 when multiplied by other matrices. Pre-multiplying or post-multiplying a matrix by the identity matrix results in no change to the original matrix.

$A = AI = IA$

The inverse of a 2×2 matrix in the form $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is given by

$$A^{-1} = \frac{1}{\det(A)}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \text{ or } A^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

A **singular matrix** is a matrix with no inverse, where the determinant equals zero. For a 2×2 singular matrix with a determinant of zero, the fraction, $\frac{1}{ad - bc}$ will be equal to $\frac{1}{0}$ which is undefined. Hence there is no inverse.

Worked example 2

$A = \begin{bmatrix} 7 & 5 \\ 2 & 3 \end{bmatrix}$

a. Does the inverse of A exist?

Explanation

Step 1: Calculate the determinant.

$$\begin{aligned} \det(A) &= 7 \times 3 - 5 \times 2 \\ &= 21 - 10 \\ &= 11 \end{aligned}$$

Step 2: Determine if the inverse is defined.

$\det(A) \neq 0$

Since the determinant is not equal to 0, the inverse matrix is defined.

Answer

Yes

b. Calculate A^{-1}.

Explanation – Method 1: By hand

Find the inverse matrix by using the formula for 2×2 matrices.

$$A^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

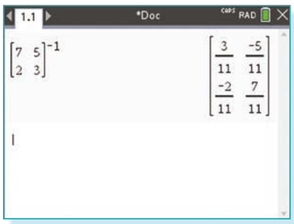
$$A^{-1} = \frac{1}{7 \times 3 - 5 \times 2}\begin{bmatrix} 3 & -5 \\ -2 & 7 \end{bmatrix}$$

$$A^{-1} = \frac{1}{11}\begin{bmatrix} 3 & -5 \\ -2 & 7 \end{bmatrix}$$

Explanation – Method 2: TI-Nspire

Step 1: From the home screen, select '1: New' → '1: Add Calculator'.

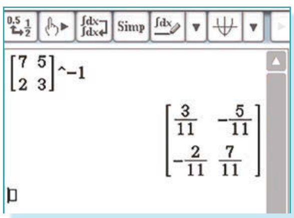
Step 2: Enter matrix A and type '^-1'. Press enter.



Explanation – Method 3: Casio ClassPad

Step 1: From the main menu, tap √α Main.

Step 2: Enter matrix A and type '^-1'. Press EXE.



Answer – Method 1, 2 and 3

$$\begin{bmatrix} \frac{3}{11} & -\frac{5}{11} \\ -\frac{2}{11} & \frac{7}{11} \end{bmatrix}$$

Back to contents

Solving simultaneous equations using matrix equations

If an equation contains two unknown variables, two equations are required to solve for these values. These equations are called 'simultaneous equations'. It is possible to have more than two simultaneous equations where there are more than two unknown variables.

Matrices can be used to solve simultaneous equations.

The simultaneous equations

$ax + by = e$

$cx + dy = f$

can be converted into a matrix equation:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} e \\ f \end{bmatrix}$$

The matrix equation is in the form $AX = B$.

- Matrix A, $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, contains information about the coefficients of x and y.
- Matrix X, $\begin{bmatrix} x \\ y \end{bmatrix}$, contains the unknown variables, x and y.
- Matrix B, $\begin{bmatrix} e \\ f \end{bmatrix}$ contains information about what the matrix equations equate to.

Once simultaneous equations have been represented in matrix form, the equations can be rearranged using matrix operations to solve for x and y.

Matrices cannot be divided. Therefore, when solving for matrix X in a matrix equation, the inverse of A is used.

Assuming A^{-1} exists, $AX = B$ can be solved for X by using the inverse matrix.

$$\begin{aligned} AX &= B \\ A^{-1}AX &= A^{-1}B \\ IX &= A^{-1}B \\ X &= A^{-1}B \end{aligned}$$

The order of multiplication by the inverse matrix needs to be consistent on each side of the equation. If A is pre-multiplied by A^{-1}, then B must also be pre-multiplied by A^{-1}.

Note: Questions involving three or more simultaneous equations should be solved using a CAS instead.

If the determinant of a matrix is zero, its inverse does not exist. This means there is no unique solution for the simultaneous equations.

Worked example 3

Solve the following sets of simultaneous equations.

a. $2x + 3y = 2$

$3x + 4y = 3$

Explanation

Step 1: Set up the matrix equations.

$$\begin{bmatrix} 2 & 3 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

Step 2: Define the matrices in the matrix equation.

Let $A = \begin{bmatrix} 2 & 3 \\ 3 & 4 \end{bmatrix}$

Let $X = \begin{bmatrix} x \\ y \end{bmatrix}$

Let $B = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$

Step 3: Solve the matrix equation.

$$\begin{aligned} AX &= B \\ A^{-1}AX &= A^{-1}B \\ X &= A^{-1}B \end{aligned}$$

Step 4: Calculate A^{-1}.

$$A^{-1} = \begin{bmatrix} -4 & 3 \\ 3 & -2 \end{bmatrix}$$

Continues →

Back to contents

Step 5: Evaluate $X = A^{-1}B$.

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -4 & 3 \\ 3 & -2 \end{bmatrix}\begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Answer

$x = 1, \quad y = 0$

---

b. $3x + 2y - 2z = 26$

$4x + 6y + z = 76$

$-2x + y + 4z = 12$

Explanation – Method 1: TI-Nspire

Step 1: Write a matrix equation representing the simultaneous equations in the form $AX = B$.

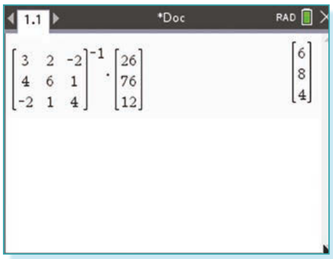
$$\begin{bmatrix} 3 & 2 & -2 \\ 4 & 6 & 1 \\ -2 & 1 & 4 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 26 \\ 76 \\ 12 \end{bmatrix}$$

Step 2: Rearrange the equation using inverse matrices.

$$AX = B$$
$$A^{-1}AX = A^{-1}B$$
$$X = A^{-1}B$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3 & 2 & -2 \\ 4 & 6 & 1 \\ -2 & 1 & 4 \end{bmatrix}^{-1}\begin{bmatrix} 26 \\ 76 \\ 12 \end{bmatrix}$$

Step 3: Enter the equation into the calculator.



Explanation – Method 2: Casio ClassPad

Step 1: Write a matrix equation representing the simultaneous equations in the form $AX = B$.

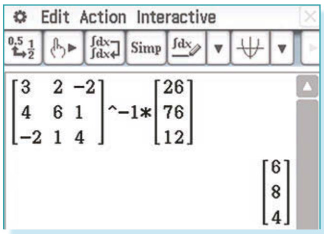
$$\begin{bmatrix} 3 & 2 & -2 \\ 4 & 6 & 1 \\ -2 & 1 & 4 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 26 \\ 76 \\ 12 \end{bmatrix}$$

Step 2: Rearrange the equation using inverse matrices.

$$AX = B$$
$$A^{-1}AX = A^{-1}B$$
$$X = A^{-1}B$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3 & 2 & -2 \\ 4 & 6 & 1 \\ -2 & 1 & 4 \end{bmatrix}^{-1}\begin{bmatrix} 26 \\ 76 \\ 12 \end{bmatrix}$$

Step 3: Enter the equation into the calculator.



Answer – Method 1 and 2

$x = 6, \quad y = 8, \quad z = 4$

Exam question breakdown

VCAA 2021 Exam 1 Matrices Q5

A is a 7×7 matrix.

B is a 10×7 matrix.

Which of the following matrix equations is defined?

A. $AB - 2B$ **B.** $A(BA)^{-1}$ **C.** AB^2 **D.** $A^2 - BA$ **E.** $A(B^T)$

Continues →

Back to contents

Explanation

To solve this question, check whether each option is correct or incorrect.

A: This is incorrect. In order for AB to be defined, the number of columns in A must equal the number of rows in B. ✗

B: This is incorrect. The order of BA will be equal to the number of rows in B and the number of columns in A. The order of BA is 10×7. This is not a square matrix so the inverse does not exist. ✗

C: This is incorrect. B cannot be raised to a power because it is not a square matrix. ✗

D: This is incorrect. The order of BA would be 10×7. The order of A^2 is 7×7. $A^2 - BA$ is undefined since both matrices must have the same order. ✗

E: This is correct. The order of B^T will be 7×10. The number of columns in A must equal the number of rows in B in order for AB to be defined. Therefore AB is defined. ✓

Answer

E

45% of students answered this question correctly.

7D Questions

Calculating the determinant of a matrix

1. What is the determinant of matrix A?

$A = \begin{bmatrix} -2 & 2 \\ 3 & 4 \end{bmatrix}$

A. -14 **B.** -2 **C.** 0 **D.** 14

2. Find the determinant for each of the following matrices by hand.

a. $\begin{bmatrix} -3 & 12 \\ -2 & 8 \end{bmatrix}$ **b.** $\begin{bmatrix} 3 & 2 \\ 2 & 2 \end{bmatrix}$ **c.** $\begin{bmatrix} \frac{1}{2} & -\frac{2}{3} \\ -3 & -\frac{3}{4} \end{bmatrix}$

3. Find the determinant for each of the following matrices using a CAS.

a. $\begin{bmatrix} 4 & 5 & 1 \\ -2 & 1 & 0 \\ 1 & 3 & -3 \end{bmatrix}$ **b.** $\begin{bmatrix} 2 & 1 & 3 \\ 0 & -1 & -1 \\ 2 & 0 & 5 \end{bmatrix}$ **c.** $\begin{bmatrix} 2 & 3 & 1 & -2 \\ 6 & 1 & 4 & 1 \\ 0 & 3 & 7 & 1 \\ 1 & 2 & 4 & 0 \end{bmatrix}$

4. The determinant of matrix A is -4. Find k.

$A = \begin{bmatrix} \frac{2}{3} & 4 \\ k & 3 \end{bmatrix}$

Calculating the inverse of a matrix

5. What is the inverse of matrix P?

$P = \begin{bmatrix} -2 & 7 \\ 3 & 0 \end{bmatrix}$

A. $\begin{bmatrix} 0 & -7 \\ -3 & -2 \end{bmatrix}$ **B.** $\begin{bmatrix} 0 & \frac{1}{3} \\ \frac{1}{7} & \frac{2}{21} \end{bmatrix}$ **C.** $\begin{bmatrix} \frac{2}{21} & -\frac{1}{3} \\ -\frac{1}{7} & 0 \end{bmatrix}$ **D.** $\begin{bmatrix} 0 & \frac{1}{7} \\ \frac{1}{3} & \frac{2}{21} \end{bmatrix}$

Back to contents

6. For each of the following matrices, determine whether the inverse is defined. If it is defined, find the inverse.

a. $A = \begin{bmatrix} 8 & 1 \\ 2 & 3 \end{bmatrix}$ **b.** $B = \begin{bmatrix} 0 & -6 \\ -2 & -9 \end{bmatrix}$ **c.** $C = \begin{bmatrix} 1 & 2 & 0 \\ 4 & 3 & 7 \\ 0 & 5 & 6 \end{bmatrix}$ **d.** $D = \begin{bmatrix} 1 & 0 & 4 \\ 8 & 3 & 16 \\ 0 & 0 & 0 \end{bmatrix}$

7. Determine whether the following pairs of matrices are each other's inverses.

a. $\begin{bmatrix} \frac{1}{2} & \frac{-1}{4} \\ \frac{-1}{4} & \frac{3}{8} \end{bmatrix}$ and $\begin{bmatrix} 3 & 2 \\ 2 & 4 \end{bmatrix}$ **b.** $\begin{bmatrix} 7 & 3 \\ 6 & 2 \end{bmatrix}$ and $\begin{bmatrix} \frac{1}{2} & -\frac{1}{3} \\ -\frac{1}{6} & \frac{1}{7} \end{bmatrix}$ **c.** $\begin{bmatrix} 4 & -\frac{1}{2} \\ 16 & 5 \end{bmatrix}$ and $\begin{bmatrix} 5 & \frac{1}{2} \\ -16 & 4 \end{bmatrix}$

8. $A = \begin{bmatrix} -\frac{1}{3} & u \\ -\frac{3}{2} & 5 \end{bmatrix}$, $A^{-1} = \begin{bmatrix} \frac{15}{22} & -\frac{9}{11} \\ \frac{9}{44} & -\frac{1}{22} \end{bmatrix}$

Find the value of u.

Solving simultaneous equations using matrix equations

9. Which of the following pairs of simultaneous equations is represented by this matrix equation?

$\begin{bmatrix} 4 & 1 \\ -2 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$

A. $4x - 2y = 3$
$x + 2y = 2$

B. $4x + y = 3$
$-2x + 2y = 2$

C. $y = 3x - 4$
$2y = 2x + 1$

D. $y = -4x + 3$
$y = x + 2$

10. Express the following simultaneous equations in matrix form.

a. $x + 2y = 2$ and $2x + 3y = 6$

b. $3x - 2y = 5$ and $x - 3y = 5$

c. $y = 3x - 3$ and $y = -2x + 1$

d. $2x - 4y + 6z = 12$ and $-x + 3y + 5z = 2$ and $-3x + 2y + 5z = 5$.

11. Solve the following matrix equations for the values of x and y.

a. $\begin{bmatrix} 6 & 8 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 60 \\ 14 \end{bmatrix}$ **b.** $\begin{bmatrix} -6 & 2 \\ -14 & 5 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 58 \\ 137 \end{bmatrix}$ **c.** $\begin{bmatrix} 8 & 2 \\ 11 & 3 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 18 \\ 26 \end{bmatrix}$

12. For each of the following sets of simultaneous equations:

- Express the equations in matrix form.
- Calculate the determinant of the coefficient matrix and state whether a unique solution exists.

a. $3x + 5y = -2$ and $3x + 6y = -2$

b. $-2x + 3y = 0$ and $-3x + 4.5y = 0$

c. $x - 3y = 6$ and $2x + 4y = 3$

d. $-3y + 2x = 6$ and $-3x + 2y = 6$

Joining it all together

13. Pearl loves body products and decides to call her favourite store and ask them to create two gift baskets worth \$120 to sell at an auction. The first basket has 7 bottles of body wash (b), and 6 bottles of moisturiser (m). The second basket has 10 bottles of body wash and 4 bottles of moisturiser. This has been represented by the following simultaneous equations.

$7b + 6m = 120$
$10b + 4m = 120$

a. Represent the simultaneous equations in matrix form.

b. What is the determinant of the coefficient matrix from part **a**?

c. The matrix equation $\begin{bmatrix} b \\ m \end{bmatrix} = \frac{1}{p}\begin{bmatrix} 4 & -6 \\ q & 7 \end{bmatrix}\begin{bmatrix} 120 \\ 120 \end{bmatrix}$ can be used to solve the simultaneous equations.
Determine the values of p and q.

d. How much did each body wash and moisturiser cost?

---

14. Yasmine has started a cafe that sells full-cream and soy coffees in both regular and large sizes. The following table contains information about her daily sales.

	regular	**large**
full-cream	78	19
soy	37	5

a. Which of the following matrices, S, accurately represents the information provided in the table?

A. $$\begin{array}{cc} \begin{matrix}\text{full-cream} & \text{soy}\end{matrix} & \\ \begin{bmatrix} 78 & 37 \\ 5 & 19 \end{bmatrix} & \begin{matrix}\text{regular} \\ \text{large}\end{matrix} \end{array}$$

B. $$\begin{array}{cc} \begin{matrix}\text{soy} & \text{full-cream}\end{matrix} & \\ \begin{bmatrix} 78 & 5 \\ 19 & 37 \end{bmatrix} & \begin{matrix}\text{large} \\ \text{regular}\end{matrix} \end{array}$$

C. $$\begin{array}{cc} \begin{matrix}\text{large} & \text{regular}\end{matrix} & \\ \begin{bmatrix} 19 & 37 \\ 78 & 5 \end{bmatrix} & \begin{matrix}\text{full-cream} \\ \text{soy}\end{matrix} \end{array}$$

D. $$\begin{array}{cc} \begin{matrix}\text{regular} & \text{large}\end{matrix} & \\ \begin{bmatrix} 78 & 19 \\ 37 & 5 \end{bmatrix} & \begin{matrix}\text{full-cream} \\ \text{soy}\end{matrix} \end{array}$$

Yasmine aims to have daily revenue of \$500. She expects regular-sized coffees to account for 70% of this revenue, with the remaining 30% being provided by large coffees.

b. Complete matrix R which shows the revenue, in dollars, earned from sales of each coffee size.

$$R = \begin{bmatrix} \quad \\ \quad \end{bmatrix}\begin{matrix}\text{regular} \\ \text{large}\end{matrix}$$

c. Matrix R can be found by post-multiplying matrix S by a 2×1 matrix, C. Matrix C contains the prices of each coffee. Write down a matrix equation showing the calculation for Matrix R.

d. In order to achieve her daily revenue goal, how much should Yasmine charge, correct to the nearest 10 cents, for a regular coffee and for a large coffee?

Exam practice

15. Which 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. $\begin{bmatrix} 4 & 0 \\ 0 & -2 \end{bmatrix}$

VCAA 2018 Exam 1 Matrices Q1

78% of students answered this question correctly.

---

16. The preferred number of cafes (x) and sandwich bars (y) in Grandmall's food court can be determined by solving the following equations written in matrix form.

$$\begin{bmatrix} 5 & -9 \\ 4 & -7 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$$

a. The value of the determinant of the 2×2 matrix is 1.
Use this information to explain why this matrix has an inverse. (1 MARK)

b. Write the three missing values of the inverse matrix that can be used to solve these equations. (1 MARK)

$$\begin{bmatrix} \quad & 9 \\ \quad & \quad \end{bmatrix}$$

c. Determine the preferred number of sandwich bars for Grandmall's food court. (1 MARK)

VCAA 2020 Exam 2 Matrices Q2

Part **a**: **37%** of students answered this question correctly.
Part **b**: **71%** of students answered this question correctly.
Part **c**: **51%** of students answered this question correctly.

Back to contents

17. The following matrix equation represents a pair of simultaneous linear equations.

$$\begin{bmatrix} 12 & 9 \\ m & 3 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 6 \\ 6 \end{bmatrix}$$

These simultaneous linear equations have no unique solution when m is equal to

A. -4 **B.** -3 **C.** 0
D. 3 **E.** 4

VCAA 2016 Exam 1 Matrices Q3

63% of students answered this question correctly.

Questions from multiple lessons

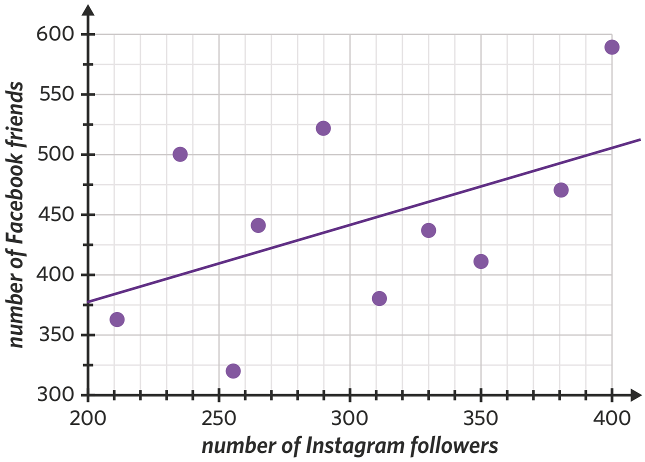
Data analysis

18. The scatterplot shows the *number of Facebook friends* and *number of Instagram followers* for a group of 10 people. A least squares regression line has been fitted to the graph.

The regression line is used to predict the number of Facebook friends for a person with 350 Instagram followers. The residual is closest to

A. -78
B. -62
C. -36
D. 62
E. 78

Adapted from VCAA 2017 Exam 1 Data analysis Q9



Matrices

19. Last night Lena won the powerball jackpot.

Each week, she allocates some of her winnings, in dollars, to either spending (SP) or savings (SV) for four weeks. This is represented in the following matrix:

$$\begin{array}{c} \begin{array}{cc} \text{SP} & \text{SV} \end{array} \\ \begin{bmatrix} 300\,000 & 40 \\ 10\,000 & 100 \\ 98\,000 & 60 \\ 50\,450 & 50 \end{bmatrix} \end{array} \begin{array}{l} \\ \text{week 1} \\ \text{week 2} \\ \text{week 3} \\ \text{week 4} \end{array}$$

How much did Lena spend in week 4?

A. \$50 **B.** \$10 000 **C.** \$50 450 **D.** \$50 500 **E.** \$98 000

Adapted from VCAA 2017 Exam 1 Matrices Q1

Data analysis

20. Mary has a shopping addiction. She goes shopping every single day. The amount of money she spends each day is approximately normally distributed with a mean of \$130 and a standard deviation of \$40.

a. Using the 68–95–99.7% rule, determine:

i. the percentage of days where she spent more than \$170. (1 MARK)

ii. the expected number of days where Mary spends less than \$250 from a sample of 2000 days. (1 MARK)

b. The standardised amount she spent on Boxing Day is given by $z = 2.7$.

Determine the actual amount she spent on Boxing Day. (1 MARK)

Adapted from VCAA 2017 Exam 2 Data analysis Q1

Back to contents

7E Binary and permutation matrices

STUDY DESIGN DOT POINT

- binary and permutation matrices, and their properties and applications

7A 7B 7C 7D 7E 7F 7G 7H 7I

KEY SKILLS

During this lesson, you will be:

- applying permutations to matrices
- constructing permutation matrices.

KEY TERMS

- Binary matrix
- Permutation matrix
- Column permutation
- Row permutation

A permutation matrix is a unique way to solve problems with several possible solutions (called permutations). This particular application of the binary matrix allows for the efficient rearrangement of rows or columns in other matrices.

Applying permutations to matrices

A **binary matrix** is a matrix in which all elements are either 0 or 1. A **permutation matrix** is a square binary matrix in which each row and column must contain the number 1 only once.

Binary matrix

$$\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

Permutation matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

When multiplied with another matrix, a permutation matrix has the effect of rearranging either the rows or columns of the other matrix. A permutation matrix can be post-multiplied (multiplied after another matrix) or pre-multiplied (multiplied before another matrix).

If a permutation matrix is defined as matrix P, and another matrix is defined as matrix Q, then:

- $Q \times P$ is a post-multiplication of the permutation matrix, resulting in a column permutation.

  Note: A **column permutation** is a new matrix formed from the rearranged columns of another matrix.

- $P \times Q$ is a pre-multiplication of the permutation matrix, resulting in a row permutation.

  Note: A **row permutation** is a new matrix formed from the rearranged rows of another matrix.

When applying a column permutation, the permutation matrix P must be an $m \times m$ square matrix, where m is equal to the number of columns in matrix Q.

See worked example 1

For example,

matrix Q matrix P

$$\begin{bmatrix} 3 & 7 & 2 \\ 4 & 1 & 5 \end{bmatrix} \times \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

When interpreting a column permutation matrix, the position of each '1' element becomes extremely important.

Back to contents

For the example $Q \times P$, matrix P contains '1' elements in the following positions:

- p_{13} – This indicates that column 1 in matrix Q (elements 3 and 4) will become column 3.
- p_{21} – This indicates that column 2 in matrix Q (elements 7 and 1) will become column 1.
- p_{32} – This indicates that column 3 in matrix Q (elements 2 and 5) will become column 2.

Because of these permutations, the new matrix will be

$$\begin{bmatrix} 7 & 2 & 3 \\ 1 & 5 & 4 \end{bmatrix}$$

When applying a row permutation, the permutation matrix P must be an $n \times n$ square matrix, where n is equal to the number of rows in matrix Q.

See worked example 2

When interpreting a row permutation matrix, the position of each '1' element becomes extremely important. These are read in the reverse order to column permutations.

For example, consider the following matrix multiplication.

$$\underset{\text{matrix } P}{\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}} \times \underset{\text{matrix } Q}{\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \\ j & k & l \end{bmatrix}}$$

Matrix P contains '1' elements in the following positions:

- p_{14} – This indicates that row 4 in matrix Q (elements j, k and l) will become row 1.
- p_{22} – This indicates that row 2 in matrix Q (elements d, e and f) will stay as row 2 (unchanged).
- p_{31} – This indicates that row 1 in matrix Q (elements a, b and c) will become row 3.
- p_{43} – This indicates that row 3 in matrix Q (elements g, h and i) will become row 4.

Because of these permutations, the new matrix will be

$$\begin{bmatrix} j & k & l \\ d & e & f \\ a & b & c \\ g & h & i \end{bmatrix}$$

Permutations can be more efficiently performed using a calculator.

When multiple iterations of a permutation are to be applied, matrix powers can be used to complete the calculation more efficiently.

For example, a permutation matrix raised to the power of 3 would indicate that this permutation is being applied three times in succession.

Worked example 1

Perform a column permutation of matrix X by using permutation matrix P.

$$X = \begin{bmatrix} u & m & l & p \end{bmatrix}, \quad P = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Explanation – Method 1: By hand

Step 1: Determine the movement of each column in matrix X using the permutation matrix.

p_{13} – column 1 moves to column 3

p_{24} – column 2 moves to column 4

p_{32} – column 3 moves to column 2

p_{41} – column 4 moves to column 1

Step 2: Write the new matrix.

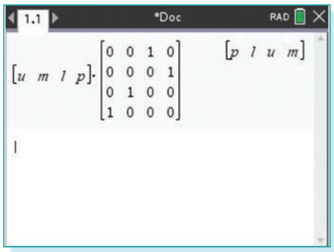
Continues →

Back to contents

Explanation – Method 2: TI-Nspire

Step 1: From the home screen, select '1: New' → '1: Add Calculator'.

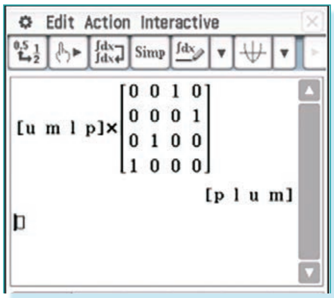
Step 2: Enter the matrix calculation. This is a column permutation, which means the permutation matrix must be post-multiplied to matrix X. Press enter.



Explanation – Method 3: Casio ClassPad

Step 1: From the main menu, tap √α Main.

Step 2: Enter the matrix calculation. This is a column permutation, which means the permutation matrix must be post-multiplied to matrix X. Press EXE.



Answer – Method 1, 2 and 3

$[p \quad l \quad u \quad m]$

Worked example 2

Perform a row permutation of matrix X by using permutation matrix P.

$$X = \begin{bmatrix} f & o & x \\ f & e & d \\ t & h & e \\ w & a & s \end{bmatrix}, \quad P = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

Explanation – Method 1: By hand

Step 1: Determine the movement of each row in matrix X using the permutation matrix.

p_{13} – row 3 moves to row 1

p_{21} – row 1 moves to row 2

p_{34} – row 4 moves to row 3

p_{42} – row 2 moves to row 4

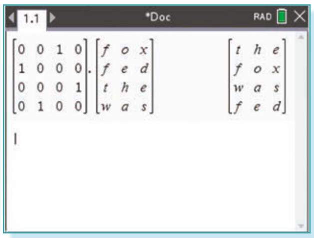
Step 2: Write the new matrix.

Continues →

Explanation – Method 2: TI-Nspire

Step 1: From the home screen, select '1: New' → '4: Add Calculator'.

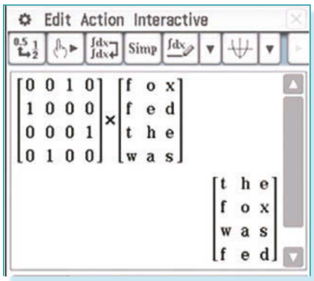
Step 2: Enter the matrix calculation. This is a row permutation, which means the permutation matrix must be pre-multiplied to matrix X. Press enter.



Explanation – Method 3: Casio ClassPad

Step 1: From the main menu, tap √α Main.

Step 2: Enter the matrix calculation. This is a row permutation, which means the permutation matrix must be pre-multiplied to matrix X. Press EXE.



Answer – Method 1, 2 and 3

$$\begin{bmatrix} t & h & e \\ f & o & x \\ w & a & s \\ f & e & d \end{bmatrix}$$

Constructing permutation matrices

When there is a requirement to create a matrix permutation, an understanding of the position of the '1' elements allows for the construction of a permutation matrix.

For example, if column 3 is to be moved to column 1, the resultant permutation matrix P would have a '1' element at p_{31}.

Similarly, if row 3 is to be moved to row 1, the resultant permutation matrix P would have a '1' element at p_{13}.

Worked example 3

Determine the permutation matrix P that would be required to change

$$\begin{bmatrix} 21 & 17 & 42 \\ 31 & 22 & 27 \\ 46 & 33 & 15 \end{bmatrix} \text{ to } \begin{bmatrix} 31 & 22 & 27 \\ 21 & 17 & 42 \\ 46 & 33 & 15 \end{bmatrix}$$

Explanation

Step 1: Determine the type of permutation.

As the rows are staying intact and are being moved up or down, this question represents a row permutation.

Step 2: Determine the order of the permutation matrix.

As the original matrix has 3 rows, the permutation matrix will be a 3 × 3 square matrix.

Continues →

Back to contents

Step 3: Identify the row changes to determine the ‘1’ elements in the permutation matrix.

Row 1 moves to row 2 – there is a ‘1’ element at p_{21}.

Row 2 moves to row 1 – there is a ‘1’ element at p_{12}.

Row 3 stays as row 3 – there is a ‘1’ element at p_{33}.

Step 4: Construct the permutation matrix.

Write ‘1’s in the required positions, with ‘0’s everywhere else. There should only be one ‘1’ element in each row and column.

Answer

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Exam question breakdown

VCAA 2017 Exam 1 Matrices Q4

A permutation matrix, P, can be used to change $\begin{bmatrix} f \\ e \\ a \\ r \\ s \end{bmatrix}$ into $\begin{bmatrix} s \\ a \\ f \\ e \\ r \end{bmatrix}$.

Matrix P is

A. $\begin{bmatrix} 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix}$ **B.** $\begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$ **C.** $\begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$ **D.** $\begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$ **E.** $\begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$

Explanation

Step 1: Determine the type of permutation.

As there is only one column, this will be a row permutation.

The order of the permutation matrix does not need to be determined, as all solutions are 5×5 matrices.

Step 2: Identify the row changes to determine the ‘1’ elements in the permutation matrix.

Row 1 moves to row 3 – there is a ‘1’ element at p_{31}.

Row 2 moves to row 4 – there is a ‘1’ element at p_{42}.

Row 3 moves to row 2 – there is a ‘1’ element at p_{23}.

Row 4 moves to row 5 – there is a ‘1’ element at p_{54}.

Row 5 moves to row 1 – there is a ‘1’ element at p_{15}.

Step 3: Construct the matrix.

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Answer

C

74% of students answered this question correctly.

This question was generally well answered, but the **12%** of students who chose option A or D did not recall that a permutation matrix can only have one ‘1’ element in every column and row.

Back to contents

7E Questions

Applying permutations to matrices

1. For the following matrices

$$A = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

a. Which of these are binary matrices?

b. Which of these are permutation matrices?

2. Perform a column permutation on matrix M using permutation matrix P in the following.

a. $M = \begin{bmatrix} 4 & 7 \\ 3 & 9 \end{bmatrix}, \quad P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

b. $M = \begin{bmatrix} -3 & 0 & 4 \end{bmatrix}, \quad P = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

c. $M = \begin{bmatrix} f & m & v \\ c & q & p \end{bmatrix}, \quad P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

d. $M = \begin{bmatrix} 15 & 27 & 19 & 12 \\ 7 & 15 & 9 & 11 \\ 14 & 8 & 17 & 15 \end{bmatrix}, \quad P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$

3. Perform a row permutation on matrix Q using permutation matrix P in the following.

a. $P = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad Q = \begin{bmatrix} 4 & 9 \\ 3 & 8 \\ 6 & 7 \end{bmatrix}$

b. $P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad Q = \begin{bmatrix} -2 \\ 4 \\ 0 \\ 1 \end{bmatrix}$

c. $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad Q = \begin{bmatrix} a & s & d & f \\ h & j & k & l \end{bmatrix}$

d. $P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad Q = \begin{bmatrix} 22 & 31 \\ 14 & 17 \\ 25 & 19 \end{bmatrix}$

Constructing permutation matrices

4. Which of the following permutation matrices would be required to change $[s \quad l \quad o \quad t]$ to $[l \quad o \quad t \quad s]$?

A. Pre-multiplication of $\begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$

B. Pre-multiplication of $\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$

C. Post-multiplication of $\begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$

D. Post-multiplication of $\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$

5. For matrix, $A = \begin{bmatrix} s \\ l \\ a \\ p \end{bmatrix}$

a. Determine the permutation matrix P that would be required to change $\begin{bmatrix} s \\ l \\ a \\ p \end{bmatrix}$ to $\begin{bmatrix} l \\ a \\ p \\ s \end{bmatrix}$.

b. Would this be a pre-multiplication or post-multiplication of P?

6. Show a matrix calculation that can be used to rearrange the following matrix to spell the word 'MASH':

$[H \quad S \quad M \quad A]$

Back to contents

Joining it all together

7. Wordle is a game where players find words inside a grid of jumbled letters. Samantha is playing a variation of the game where she is tasked with finding four letter words running either horizontally or vertically in the following matrix:

$$Q = \begin{bmatrix} i & r & l & s \\ b & u & h & d \\ n & a & e & m \\ t & c & a & t \end{bmatrix}$$

She decides to apply a series of permutation matrices to rearrange the rows and columns so that the words will be easier for her to see.

a. Apply a column permutation to the original letter matrix, Q, so that a word is spelled in one of the rows. Show the calculation as well as the final answer.

b. Apply a row permutation to the original letter matrix, Q, so that a word is spelled in one of the columns. Show the calculation as well as the final answer.

Exam practice

8. 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.** $\begin{bmatrix} P \\ E \\ A \\ L \\ S \end{bmatrix}$

VCAA 2016 Exam 1 Matrices Q2

86% of students answered this question correctly.

---

9. Matrix P is a 4×4 permutation matrix.

Matrix W is another matrix such that the matrix product $P \times W$ is 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.** $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

VCAA 2018 Exam 1 Matrices Q4

69% of students answered this question correctly.

Back to contents

10. Matrices P and W are defined as shown.

$$P = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad W = \begin{bmatrix} A \\ S \\ T \\ O \\ R \end{bmatrix}$$

If $P^n \times W = \begin{bmatrix} A \\ S \\ T \\ O \\ R \end{bmatrix}$, the value of n could be

A. 1 **B.** 2 **C.** 3
D. 4 **E.** 5

VCAA 2020 Exam 1 Matrices Q3

64% of students answered this question correctly.

---

11. A theme park has four locations, Air World (A), Food World (F), Ground World (G) and Water World (W). The proportion of visitors moving from one location to another each hour on Saturday is shown in matrix T.

$$\begin{array}{c} \text{this hour} \\ T = \begin{array}{c} \begin{array}{cccc} \mathrm{A} & \mathrm{F} & \mathrm{G} & \mathrm{W} \end{array} \\ \begin{bmatrix} 0.1 & 0.2 & 0.1 & 0.2 \\ 0.3 & 0.4 & 0.6 & 0.3 \\ 0.1 & 0.2 & 0.2 & 0.1 \\ 0.5 & 0.2 & 0.1 & 0.4 \end{bmatrix} \end{array} \begin{array}{c} \\ \mathrm{A} \\ \mathrm{F} \\ \mathrm{G} \\ \mathrm{W} \end{array} \text{ next hour} \end{array}$$

The proportion of visitors moving from one location to another each hour on Sunday is different from Saturday, shown in matrix V.

$$\begin{array}{c} \text{this hour} \\ V = \begin{array}{c} \begin{array}{cccc} \mathrm{A} & \mathrm{F} & \mathrm{G} & \mathrm{W} \end{array} \\ \begin{bmatrix} 0.3 & 0.4 & 0.6 & 0.3 \\ 0.1 & 0.2 & 0.1 & 0.2 \\ 0.1 & 0.2 & 0.2 & 0.1 \\ 0.5 & 0.2 & 0.1 & 0.4 \end{bmatrix} \end{array} \begin{array}{c} \\ \mathrm{A} \\ \mathrm{F} \\ \mathrm{G} \\ \mathrm{W} \end{array} \text{ next hour} \end{array}$$

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.

Write down matrix M. (1 MARK)

VCAA 2019 Exam 2 Matrices Q2d

21% of students answered this question correctly.

Questions from multiple lessons

Recursion and financial modelling

12. The value of an annuity investment, in dollars, after n years, V_n, can be modelled by the following recurrence relation.

$V_0 = 58\ 000, \quad V_{n+1} = 1.0027\ V_n + 350$

The increase in value of the investment between the second and third years is closest to

A. \$507 **B.** \$508 **C.** \$509
D. \$1015 **E.** \$1524

Adapted from VCAA 2018 Exam 1 Recursion and financial modelling Q18

Matrices *Year 11 content*

13. The cost of different pastries at a bakery are shown in the table.

croissant	\$5.50
scone	\$3.00
doughnut	\$4.20

Jamie wants to buy two croissants, four scones and three doughnuts.

Which one of the following matrix multiplications will result in a matrix showing the total cost of Jamie's purchase, in dollars?

A. $\begin{bmatrix} 2 \\ 4 \\ 3 \end{bmatrix} \times \begin{bmatrix} 5.50 \\ 3.00 \\ 4.20 \end{bmatrix}$

B. $\begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{bmatrix} \times \begin{bmatrix} 5.50 \\ 3.00 \\ 4.20 \end{bmatrix}$

C. $\begin{bmatrix} 2 & 4 & 3 \end{bmatrix} \times \begin{bmatrix} 5.50 \\ 3.00 \\ 4.20 \end{bmatrix}$

D. $\begin{bmatrix} 5.50 \\ 3.00 \\ 4.20 \end{bmatrix} \times \begin{bmatrix} 2 & 4 & 3 \end{bmatrix}$

E. $\begin{bmatrix} 2 & 4 & 3 \end{bmatrix} \times \begin{bmatrix} 5.50 & 3.00 & 4.20 \end{bmatrix}$

Adapted from VCAA 2017NH Exam 1 Matrices Q2

Matrices *Year 11 content*

14. Three of the most popular streaming websites are Netflix, Hulu and Stan.

The cost, in dollars, for a one month subscription to each of these streaming websites is shown in matrix C.

$$C = \begin{array}{c} \begin{bmatrix} 12.49 \\ 5.99 \\ 10 \end{bmatrix} \end{array} \begin{array}{l} \text{Netflix} \\ \text{Hulu} \\ \text{Stan} \end{array}$$

a. What is the cost of a one month subscription to Stan? (1 MARK)

b. Write down the order of matrix C. (1 MARK)

c. Oscar bought three months of a Netflix subscription and two months of a Hulu subscription.

The total amount of money he spent on subscriptions can be found by the matrix product $S \times C$.

Determine matrix S. (1 MARK)

Adapted from VCAA 2018 Exam 2 Matrices Q1

Back to contents

7F Communication and dominance matrices

STUDY DESIGN DOT POINT

- communication and dominance matrices and their use in analysing communication systems and ranking players in round-robin tournaments

7A 7B 7C 7D 7E 7F 7G 7H 7I

KEY SKILLS

During this lesson, you will be:

- interpreting and constructing communication matrices
- interpreting and constructing dominance matrices.

KEY TERMS

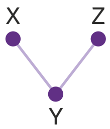
- Communication matrix
- One-step communication links
- Two-step communication links
- Redundant communication links
- Dominance matrix
- One-step dominance
- Two-step dominance

Binary matrices can be used to model everyday life situations. One of these examples is identifying direct and indirect communication pathways between individuals. Additionally, they can be used to analyse and rank teams or individuals in tournaments. In these situations, they are particularly useful when not every team or individual has played each other.

Interpreting and constructing communication matrices

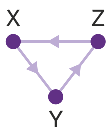
A **communication matrix** is a square binary matrix where the 1's represent the connections in a communication system. In an undirected communication matrix, the connections go in both directions. As a result, undirected communication matrices are always symmetrical about the leading diagonal. In a directed communication matrix, the connections can go in both, or one direction.

The following example is an undirected communication matrix, constructed from the network shown. Element c_{12} indicates a connection from X to Y, as well as from Y to X.



$$C = \begin{array}{c} \begin{array}{ccc} X & Y & Z \end{array} \\ \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ X \\ Y \\ Z \end{array}$$

The following example is a directed communication matrix, constructed from the network shown. Element c_{23} indicates that Y can communicate to Z. Element c_{32}, however, indicates that Z cannot communicate to Y.



$$C = \begin{array}{c} \text{receiver} \\ \begin{array}{ccc} X & Y & Z \end{array} \\ \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ \\ X \\ Y \\ Z \end{array} \text{ sender}$$

One-step communication links are direct connections between two points. They are demonstrated by a 1 in a communication matrix. Element c_{12} demonstrates a one-step communication link from X to Y.

$$C = \begin{array}{c} \text{receiver} \\ \begin{array}{ccc} X & Y & Z \end{array} \\ \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ \\ X \\ Y \\ Z \end{array} \text{ sender}$$

Back to contents

Two-step communication links are connections between two points via another point. Squaring a one-step communication matrix will generate a matrix that shows all two-step communication links.

The original matrix C shows that Z cannot communicate directly to Y. However, Z can communicate to Y via X. Element $(c^2)_{32}$ from matrix C^2 shows this two-step communication link from Z to Y.

$$C^2 = \begin{array}{c} \text{receiver} \\ \begin{array}{ccc} \text{X} & \text{Y} & \text{Z} \end{array} \\ \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ \\ \text{X} \\ \text{Y} \\ \text{Z} \end{array} \ \text{sender}$$

Summing one and two-step communication matrices gives the total number of communication links between two points which use up to two steps.

Matrix T shows the total number of one-step and two-step communication links between the points by summing the one and two-step communication matrices.

$$T = C + C^2 = \begin{array}{c} \text{receiver} \\ \begin{array}{ccc} \text{X} & \text{Y} & \text{Z} \end{array} \\ \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ \\ \text{X} \\ \text{Y} \\ \text{Z} \end{array} \ \text{sender}$$

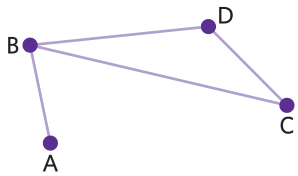
Redundant communication links are links which start and finish at the same point, and are represented by non-zero elements in the leading diagonal. They are 'redundant' as they indicate unchanged information that has returned to the starting point.

Consider a new undirected two-step communication matrix. The element in row 1, column 1, indicates that there is a two-step communication link between J and J. This is redundant as it is sending the information from J back to J.

$$C^2 = \begin{array}{c} \begin{array}{ccc} \text{J} & \text{K} & \text{L} \end{array} \\ \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ \text{J} \\ \text{K} \\ \text{L} \end{array}$$

Worked example 1

The following network diagram shows communication lines between four lighthouses, A, B, C and D.



a. Construct an undirected communication matrix, C, that indicates all one-step communication links between the lighthouses.

Explanation

Step 1: Set up an appropriate square matrix.

As there are four lighthouses, it will be a 4×4 matrix.

Label the rows and columns A to D to represent the lighthouses.

As this network is undirected, each element in the matrix represents a two-way connection, so *sender* and *receiver* labels are not required.

$$C = \begin{array}{c} \begin{array}{cccc} \text{A} & \text{B} & \text{C} & \text{D} \end{array} \\ \begin{bmatrix} & & & \\ & & & \\ & & & \\ & & & \end{bmatrix} \end{array} \begin{array}{c} \\ \text{A} \\ \text{B} \\ \text{C} \\ \text{D} \end{array}$$

Step 2: Fill in the first row of the matrix.

There is a communication line between A and B. There are no communication lines between A and A, A and C or A and D.

$$C = \begin{array}{c} \begin{array}{cccc} \text{A} & \text{B} & \text{C} & \text{D} \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 0 \\ & & & \\ & & & \\ & & & \end{bmatrix} \end{array} \begin{array}{c} \\ \text{A} \\ \text{B} \\ \text{C} \\ \text{D} \end{array}$$

Step 3: Repeat this process for the remaining rows.

Continues →

Back to contents

Answer

$$C = \begin{array}{c} \\ \\ \\ \\ \\ \end{array}\begin{array}{cccc} A & B & C & D \end{array}$$

$$C = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}\begin{array}{c} A \\ B \\ C \\ D \end{array}$$

---

b. Construct a communication matrix, T, that indicates all one-step and two-step communication links between the lighthouses.

Explanation

Step 1: Determine the two-step communication links.

Calculate matrix C^2.

$$C^2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}^2 = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 3 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{bmatrix}$$

Step 2: Calculate matrix T.

Sum the one-step and two-step communication matrices.

$$T = C + C^2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 3 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{bmatrix}$$

Answer

$$T = \begin{array}{cc} & \begin{array}{cccc} A & B & C & D \end{array} \\ & \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 3 & 2 & 2 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 2 & 2 \end{bmatrix}\begin{array}{c} A \\ B \\ C \\ D \end{array} \end{array}$$

---

c. Using matrix T from part **b**, in how many ways can lighthouse B communicate with lighthouse D?

Explanation

Identify the element from matrix T that represents the communication between lighthouses B and D.

As this is an undirected matrix, either t_{24} or t_{42} are suitable.

$$T = \begin{array}{cc} & \begin{array}{cccc} A & B & C & D \end{array} \\ & \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 3 & 2 & 2 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 2 & 2 \end{bmatrix}\begin{array}{c} A \\ B \\ C \\ D \end{array} \end{array}$$

This means there are two ways in which lighthouse B can communicate with lighthouse D.

Answer

2

---

d. Using matrix T from part **b**, how many redundant communication links are there?

Explanation

Sum the elements in the leading diagonal.

$$T = \begin{array}{cc} & \begin{array}{cccc} A & B & C & D \end{array} \\ & \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 3 & 2 & 2 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 2 & 2 \end{bmatrix}\begin{array}{c} A \\ B \\ C \\ D \end{array} \end{array}$$

$1 + 3 + 2 + 2 = 8$

Answer

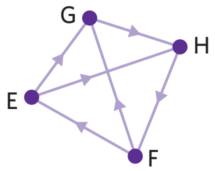
8

Interpreting and constructing dominance matrices

A **dominance matrix** is a square binary matrix where the 1's represent one-step dominances within the associated network. They are generally used for displaying and interpreting competitive matchups between two individuals or teams, where there is a winner and a loser.

Dominance matrices are similar to communication matrices in that they demonstrate links between two points. Unlike communication matrices however, dominance matrices are always directed.

The following network diagram displays the results of a tournament. An arrow from E to G indicates that team E defeated team G. This is represented in the dominance matrix, D, by the element d_{13}.



$$D = \begin{array}{c} \text{loser} \\ \begin{array}{cccc} E & F & G & H \end{array} \\ \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ \\ E \\ F \\ G \\ H \end{array} \text{winner}$$

One-step dominance refers to the direct dominance links from one point to another. Summing each row of the dominance matrix will give the one-step dominance scores.

Teams E and F both have dominance scores of 2. This means that both teams defeated 2 other teams in direct matchups.

$$D = \begin{array}{c} \text{loser} \\ \begin{array}{cccc} E & F & G & H \end{array} \\ \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ \\ E \\ F \\ G \\ H \end{array} \text{winner} \qquad \begin{array}{c} \\ \text{dominance} \\ 2 \\ 2 \\ 1 \\ 1 \end{array}$$

Two-step dominance occurs when one point has dominance over a second point, and the second point has dominance over a third point. In this case, the first point has two-step dominance over the third point. Squaring the dominance matrix will generate a matrix that indicates all two-step dominances.

In matrix D^2, element $(d^2)_{12}$ indicates that team E has two-step dominance over team F. Team E and F didn't play each other, however team E defeated team H, who in turn defeated team F.

$$D^2 = \begin{array}{c} \text{loser} \\ \begin{array}{cccc} E & F & G & H \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ \\ E \\ F \\ G \\ H \end{array} \text{winner} \qquad \begin{array}{c} \\ \text{dominance} \\ 2 \\ 3 \\ 1 \\ 2 \end{array}$$

Similar to combining communication matrices, summing one and two-step dominance matrices gives the total number of dominances one point has over each other point. Summing the rows of these total dominances will give a total dominance score, which allows each point to be ranked.

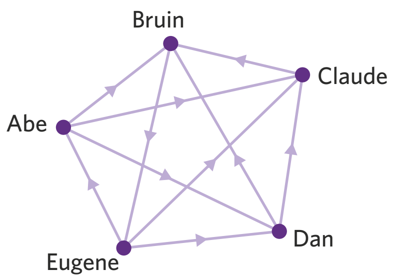
In the total dominance matrix T, team F has a total dominance score of 5. This makes them the top ranked team in the tournament.

$$T = D + D^2 = \begin{array}{c} \text{loser} \\ \begin{array}{cccc} E & F & G & H \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 2 \\ 1 & 0 & 2 & 2 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ \\ E \\ F \\ G \\ H \end{array} \text{winner} \qquad \begin{array}{c} \text{total} \\ \text{dominance} \\ 4 \\ 5 \\ 2 \\ 3 \end{array}$$

Back to contents

Worked example 2

Five players participated in a round-robin jousting tournament. The results are shown in the following network. An arrow from Abe to Bruin indicates that Abe defeated Bruin.



a. Construct a dominance matrix, D, that indicates all one-step dominances.

Explanation

Step 1: Set up an appropriate square matrix.

As there are five players, it will be a 5×5 matrix.

Label the rows and columns A to E to represent each player.

Label the rows *winner* and the columns *loser*.

$$D = \begin{array}{c} \text{loser} \\ \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \left[\begin{array}{ccccc} \ & \ & \ & \ & \ \\ \ & \ & \ & \ & \ \\ \ & \ & \ & \ & \ \\ \ & \ & \ & \ & \ \\ \ & \ & \ & \ & \ \end{array}\right] \end{array} \begin{array}{l} A \\ B \\ C \\ D \\ E \end{array} \text{ winner}$$

Step 2: Fill in the first row of the matrix.

Abe defeated Bruin, Claude and Dan.

$$D = \begin{array}{c} \text{loser} \\ \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \left[\begin{array}{ccccc} 0 & 1 & 1 & 1 & 0 \\ \ & \ & \ & \ & \ \\ \ & \ & \ & \ & \ \\ \ & \ & \ & \ & \ \\ \ & \ & \ & \ & \ \end{array}\right] \end{array} \begin{array}{l} A \\ B \\ C \\ D \\ E \end{array} \text{ winner}$$

Step 3: Repeat this process for the remaining rows.

Answer

$$D = \begin{array}{c} \text{loser} \\ \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \left[\begin{array}{ccccc} 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \end{array}\right] \end{array} \begin{array}{l} A \\ B \\ C \\ D \\ E \end{array} \text{ winner}$$

b. Rank the players from most to least dominant.

Explanation

Step 1: Determine the two-step dominances.

Calculate matrix D^2.

$$D^2 = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix}^2$$

$$= \begin{bmatrix} 0 & 2 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 3 & 2 & 1 & 0 \end{bmatrix}$$

Step 2: Calculate matrix T.

Sum the one-step and two-step communication matrices.

$$T = D + D^2 = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 2 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 3 & 2 & 1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 3 & 2 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 2 & 1 & 0 & 1 \\ 1 & 3 & 3 & 2 & 0 \end{bmatrix}$$

Continues →

Step 3: Sum the rows in matrix T.

$$T = \begin{array}{c} \\ \\ \\ \end{array}\begin{array}{c}\text{loser} \\ \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 0 & 3 & 2 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 2 & 1 & 0 & 1 \\ 1 & 3 & 3 & 2 & 0 \end{bmatrix}\end{array}\begin{array}{l} \\ \\ A \\ B \\ C \text{ winner} \\ D \\ E \end{array} \quad \begin{array}{c} \text{total} \\ \text{dominance} \\ 7 \\ 4 \\ 2 \\ 4 \\ 9 \end{array}$$

Step 4: Rank the players in order from most dominant to least dominant according to their total dominance score.

Answer

Eugene, Abe, Bruin/Dan (tied), Claude

Exam question breakdown

VCAA 2019 Exam 1 Matrices Q7

The following communication matrix shows the direct paths by which messages can be sent between two people in a group of six people, U to Z.

$$\begin{array}{c}\text{receiver} \\ \begin{array}{cccccc} U & V & W & X & Y & Z \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 \end{bmatrix}\end{array}\begin{array}{l} \\ \\ U \\ V \\ W \text{ sender} \\ X \\ Y \\ Z \end{array}$$

A ‘1’ 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 ‘1’ 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?

A. 0 **B.** 1 **C.** 2 **D.** 3 **E.** 4

Explanation

Step 1: Interpret the question.

The question asks for the number of ways Y can send a message to W via a third person. This is a two-step communication.

Step 2: Determine the two-step communication links for the entire matrix.

This is done by squaring the matrix.

$$\begin{bmatrix} 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 \end{bmatrix}^2 = \begin{array}{c}\text{receiver} \\ \begin{array}{cccccc} U & V & W & X & Y & Z \end{array} \\ \begin{bmatrix} 3 & 2 & 2 & 3 & 2 & 2 \\ 1 & 2 & 2 & 2 & 1 & 3 \\ 2 & 3 & 2 & 1 & 4 & 2 \\ 2 & 1 & 2 & 2 & 2 & 1 \\ 2 & 3 & 0 & 2 & 2 & 2 \\ 1 & 2 & 3 & 1 & 3 & 3 \end{bmatrix}\end{array}\begin{array}{l} \\ \\ U \\ V \\ W \text{ sender} \\ X \\ Y \\ Z \end{array}$$

Step 3: Identify the number of two-step communication links from Y to W.

This is represented by the element in row 5, column 3, which is 0.

Answer

A

54% of students answered this question correctly.

19% of students incorrectly answered B. This is the number of one-step communication links from Y to W. These students likely didn’t recognise that the question required them to find the number of communication links from Y to W via a third person.

7F Questions

Interpreting and constructing communication matrices

1. Which of the following can be classified as a communication matrix?

A. receiver

$$\begin{matrix} \begin{matrix} A & B & C & D \end{matrix} & \\ \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix} & \begin{matrix} A \\ B \\ C \\ D \end{matrix} \ \text{sender} \end{matrix}$$

B. receiver

$$\begin{matrix} \begin{matrix} A & B & C & D \end{matrix} & \\ \begin{bmatrix} 0 & 8 & 0 & 2 \\ 4 & 1 & 5 & 5 \\ 3 & 1 & 2 & 0 \\ 0 & 1 & 6 & 1 \end{bmatrix} & \begin{matrix} A \\ B \\ C \\ D \end{matrix} \ \text{sender} \end{matrix}$$

C. receiver

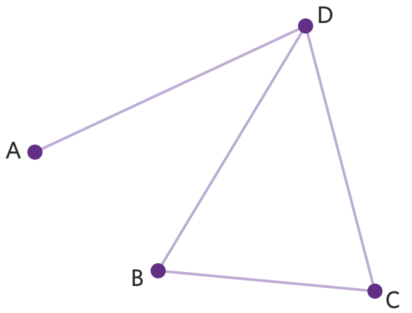
$$\begin{matrix} \begin{matrix} A & B & C & D \end{matrix} & \\ \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} & \begin{matrix} A \\ B \\ C \end{matrix} \ \text{sender} \end{matrix}$$

D. receiver

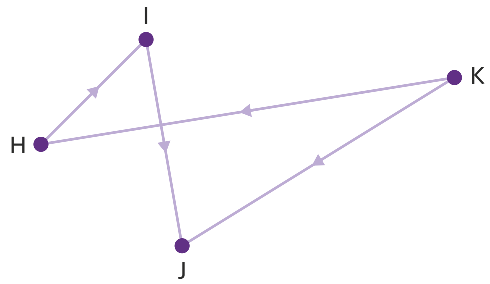
$$\begin{matrix} \begin{matrix} A & B & C & D \end{matrix} & \\ \begin{bmatrix} 2 & 4 & 1 & 3 \\ 0 & 1 & 4 & 5 \\ 1 & 0 & 8 & 0 \end{bmatrix} & \begin{matrix} A \\ B \\ C \end{matrix} \ \text{sender} \end{matrix}$$

2. Construct a communication matrix that shows all one-step communication links for each of the following diagrams.

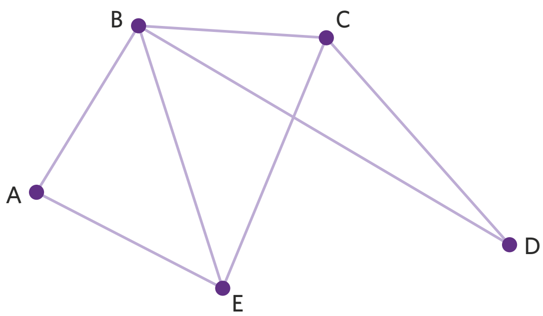
a.



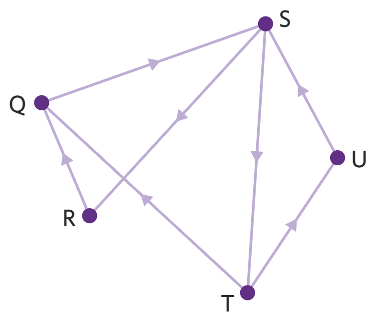
b.



c.



d.



3. The following communication matrix C, shows how four aeroplanes can communicate with each other.

$$C = \begin{matrix} \begin{matrix} P1 & P2 & P3 & P4 \end{matrix} & \\ \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} & \begin{matrix} P1 \\ P2 \\ P3 \\ P4 \end{matrix} \end{matrix}$$

a. Interpret c_{13}.

b. Which aeroplane(s) can P4 directly communicate with?

c. Is this a directed or undirected communication matrix?

4. Ariana, Beyonce, Chance, Drake and Eminem are all very careful about who they share their phone number with. The following communication matrix shows which celebrities can directly call each other.

$$\begin{array}{c} \begin{matrix} A & B & C & D & E \end{matrix} \\ \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}$$

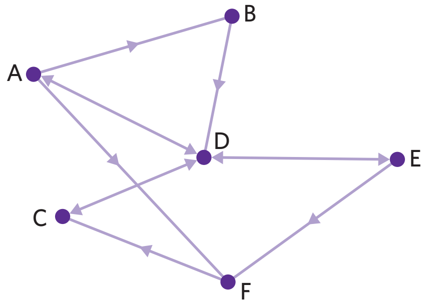
a. Who can Beyonce directly call?

b. Construct a matrix that will represent all the calls that can be made either directly or via a third person.

c. How many communication paths in the matrix found in part **b** are redundant?

---

5. The mayors of six towns, labelled A to F, decided to set up a communication network between the towns. They created a diagram to map out the existing communication network.



a. Construct a total communication matrix, T, that shows all one-step and two-step communication links between the towns.

The mayors decided it would be too expensive to set up a direct communication line between each town. Mayor Abrams, from town A, proposed that they add links to ensure each town had at least a two-step communication link to every other town.

b. Does the existing network already fulfil Mayor Abrams' proposal?

c. Determine the smallest number of one-way communication links that have to be added to fulfil Mayor Abrams' proposal and list each link.

Interpreting and constructing dominance matrices

6. Consider the following dominance matrix.

$$\begin{array}{c} \text{loser} \\ \begin{matrix} A & B & C & D \end{matrix} \\ \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ \\ A \\ B \\ C \\ D \end{array} \text{winner}$$

Which of the following statements is false?

A. Team A defeated team B.

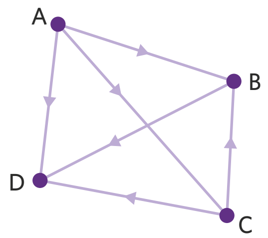
B. Team B lost to three teams.

C. Team C defeated two teams.

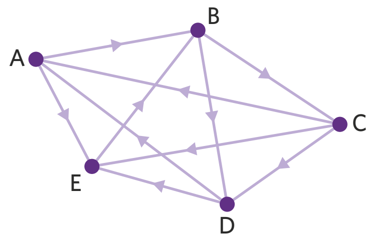
D. Team D lost to less teams than team A.

7. Construct a dominance matrix that shows all one-step dominances in the following networks.

a.



b.



8. Four players, ai.zhao999, rainbowmuffin, xthundercatx and felipe.sanchez. competed in a round-robin style League of the Ancients (LOTA) tournament. The following matrix shows the results of the tournament.

$$\begin{array}{c} \text{loser} \\ \begin{array}{cc} & \begin{array}{cccc} A & R & X & F \end{array} \\ D = & \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{array}{c} A \\ R \\ X \\ F \end{array} \text{ winner} \end{array} \end{array}$$

a. Interpret d_{23}.

b. Which player(s) did ai.zhao999 lose to?

c. Who is the top-ranking player according to one-step dominances?

9. Six teams competed in a round-robin beach volleyball tournament. After round 4, a tsunami warning was issued and the tournament was put on hold. A dominance matrix was constructed based on the four rounds played.

$$\begin{array}{c} \text{loser} \\ \begin{array}{cc} & \begin{array}{cccccc} A & B & C & D & E & F \end{array} \\ D = & \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{array}{c} A \\ B \\ C \\ D \\ E \\ F \end{array} \text{ winner} \end{array} \end{array}$$

a. Construct a dominance matrix that shows both one and two-step dominances.

b. Rank the teams from most to least dominant.

The tsunami warning turned out to be a false alarm and the tournament resumed the next day. In round 5, team C wins against team B, team E wins against team A, and team F wins against team D.

c. Taking into account both one and two-step dominances, rank the teams from most to least dominant after round 5.

10. Five members of a chess club played in a round-robin chess tournament, where each person played each of the other people once. None of the games ended in a stalemate.

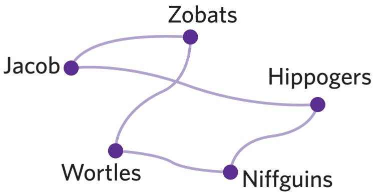
After all the matches were played, the following table of their one-step and two-step dominances was prepared to summarise the results.

member	one-step dominance	two-step dominance
Fred (F)	3	6
Georgia (G)	3	4
Harry (H)	2	2
Indiya (I)	1	3
Jax (J)	1	1

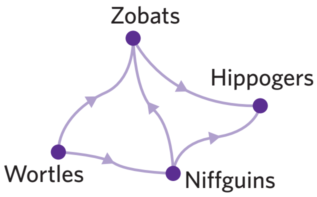
Use the information in the table to construct a one-step dominance matrix.

Joining it all together

11. Jacob stumbled upon a mysterious jungle and was curious to learn more about the inhabitants. After observing them for a couple of days, he realised that whilst he could understand some of them, not all inhabitants communicated to one another. He drew a diagram based on his observations. A line connecting Jacob and Zobats indicates that they directly communicated with one another.



He also discovered a hierarchy amongst predators and prey. The following network diagram illustrates that hierarchy. An arrow from Wortles to Zobats indicates that Wortles eat Zobats.



a. Complete the following matrix to represent all one-step and two-step communications links that Jacob observed.

$$\begin{array}{c c}
\begin{array}{ccccc} H & J & N & W & Z \end{array} & \\
\begin{bmatrix} \quad & \quad & \quad & \quad & \quad \\ & & & & \\ & & & & \\ & & & & \\ & & & & \end{bmatrix} & \begin{array}{c} H \\ J \\ N \\ W \\ Z \end{array}
\end{array}$$

b. Jacob believes that if some of the species translated for the others, Jacob will be able to communicate with all the species in the jungle. Is Jacob right? Justify your answer.

c. Jacob declares the leader of the most dominant species, based on the hierarchy amongst predators and prey, to be King of the Jungle. Taking into account one-step and two-step dominances, of which species does the King belong to?

Exam practice

12. 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 following matrix.

$$\begin{array}{c c c}
\text{loser} & & \\
\begin{array}{cccc} A & B & C & D \end{array} & & \\
\begin{bmatrix} 0 & 0 & f & 1 \\ 1 & 0 & 0 & 0 \\ 1 & g & 0 & 1 \\ 0 & 1 & 0 & h \end{bmatrix} & \begin{array}{c} A \\ B \\ C \\ D \end{array} & \text{winner}
\end{array}$$

A '1' in the matrix shows that the team named in that row defeated the team named in that column.

For example, the '1' in row 2 shows that team B defeated team A.

In this matrix, the values of f, g and h are

A. $f = 0, \quad g = 1, \quad h = 0$

B. $f = 0, \quad g = 1, \quad h = 1$

C. $f = 1, \quad g = 0, \quad h = 0$

D. $f = 1, \quad g = 1, \quad h = 0$

E. $f = 1, \quad g = 1, \quad h = 1$

VCAA 2017 Exam 1 Matrices Q5

74% of students answered this question correctly.

Back to contents

13. The main computer system in Elena's 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.

$$\begin{array}{cc} & \text{receiver} \\ & \begin{array}{ccccc} A & B & C & D & E \end{array} \\ M = & \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{array}{l} A \\ B \\ C \quad \text{sender} \\ D \\ E \end{array} \end{array}$$

In this matrix:

- the '1' in row A, column B indicates that Alex can send information to Brie.
- the '0' 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 is the square of matrix M and shows the number of two-step communication links between each pair of staff members.

$$\begin{array}{cc} & \text{receiver} \\ & \begin{array}{ccccc} A & B & C & D & E \end{array} \\ M^2 = & \begin{bmatrix} 0 & 0 & 1 & 2 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{array}{l} A \\ B \\ C \quad \text{sender} \\ D \\ E \end{array} \end{array}$$

Only one pair of individuals has two different two-step communication links.
List each two-step communication link for this pair. (1 MARK)

VCAA 2021 Exam 2 Matrices Q2

Part **a**: **92%** of students answered this question correctly.
Part **b**: **83%** of students answered this question correctly.
Part **c**: **21%** of students answered this question correctly.

Questions from multiple lessons

Data analysis

14. The amount of time, in minutes, that a population of Year 12 students spent watching MeTube over the weekend is approximately normally distributed with a mean of 310 minutes and a standard deviation of 40 minutes.

A student selected at random from this population has a standardised watch time of $z = 2.5$.

The actual amount of time that this student spent watching MeTube in minutes is

A. 160 **B.** 310 **C.** 330 **D.** 410 **E.** 775

Adapted from VCAA 2018 Exam 1 Data analysis Q3

Matrices *Year 11 content*

15. Consider the following matrix equation.

$$3 \times \begin{bmatrix} 2 & -5 \\ 0 & 1 \end{bmatrix} + A = \begin{bmatrix} 4 & -11 \\ 6 & 9 \end{bmatrix}$$

Matrix A is equal to

A. $\begin{bmatrix} 2 & -4 \\ 6 & 6 \end{bmatrix}$ **B.** $\begin{bmatrix} 10 & -26 \\ -6 & 12 \end{bmatrix}$ **C.** $\begin{bmatrix} 2 & -4 \\ -6 & -6 \end{bmatrix}$ **D.** $\begin{bmatrix} 10 & -26 \\ 6 & 12 \end{bmatrix}$ **E.** $\begin{bmatrix} -2 & 4 \\ 6 & 6 \end{bmatrix}$

Adapted from VCAA 2018NH Exam 1 Matrices Q2

Back to contents

Recursion and financial modelling

16. Pedro is a violin virtuoso and recently won $200 000 in a prestigious international chamber music competition. He decided to invest his entire winnings into a perpetuity.

The perpetuity earns interest at a rate of 6.6% per annum, with interest being calculated and paid monthly.

a. What is the value of the monthly payment that Pedro will receive? (1 MARK)

b. Pedro later decides to convert the perpetuity to an annuity investment. The annuity earns interest at a rate of 5.5% per annum, compounding monthly.

For the first three years of the investment, Pedro makes an extra payment of $600 every month to the annuity investment. This payment is made immediately after the interest has been added.

After three years of these payments, Pedro decides to increase his monthly payment into the investment. After two further years of this new monthly payment, Pedro's annuity has a balance of $320 000.

What is the value of Pedro's new monthly payment, correct to the nearest cent? (2 MARKS)

Adapted from VCAA 2017 Exam 2 Recursion and financial modelling Q7

Back to contents

7G Introduction to transition matrices

STUDY DESIGN DOT POINT

- use of the matrix recurrence relation: S_0 = initial state matrix, $S_{n+1} = TS_n$ or $S_{n+1} = LS_n$ where T is a transition matrix, L is a Leslie matrix, and S_n is a column state matrix, to generate a sequence of state matrices (assuming the next state only relies on the current state)

7A 7B 7C 7D 7E 7F 7G 7H 7I

KEY SKILLS

During this lesson, you will be:

- interpreting and constructing state matrices
- interpreting and constructing transition matrices
- calculating state matrices
- using Leslie matrices in applied scenarios.

KEY TERMS

- State matrix
- Initial state matrix
- Transition matrix
- Leslie matrix

A system consists of two or more interconnected parts working together. Examples of systems include the weather system, the ecosystem and computer systems. Systems are in a constant state of flux, with components of the system ebbing and flowing over time. Transition matrices help to define these changes and allow for the analysis of changes to the system. They can be designed to model real-life situations such as animal population mapping, election poll predictions, and even music composition.

Interpreting and constructing state matrices

A **state matrix** is a column matrix that is a snapshot of a system at a point in time. In a given scenario, there will be regular time intervals between these snapshots, such as days or weeks. The state matrix at time period n is denoted S_n.

For example, if state matrices are used to capture the populations of cities in different years, S_4 would show the populations after 4 years.

An **initial state matrix** is a state matrix that represents the initial, or starting, state of a system. It is most commonly denoted as S_0, but can also be represented as S_1.

When interpreting the elements of a state matrix, the context needs to be considered. For example, the following initial state matrix shows the number of people in a school who prefer apples, bananas, or pears.

See worked example 1

$$S_0 = \begin{array}{c} \begin{bmatrix} 130 \\ 180 \\ 145 \end{bmatrix} \end{array} \begin{array}{l} \text{apples} \\ \text{bananas} \\ \text{pears} \end{array}$$

It can be determined that, initially, 130 people prefer apples, 180 people prefer bananas, and 145 people prefer pears.

When constructing a state matrix, make sure to label the rows to make it clear what each element represents.

See worked example 2

Back to contents

Worked example 1

The number of Year 10, 11 and 12 students who chose to study maths this year is shown in the initial state matrix, S_0.

$$S_0 = \begin{array}{c} \begin{bmatrix} 137 \\ 112 \\ 82 \end{bmatrix} \end{array} \begin{array}{l} \text{Year 10} \\ \text{Year 11} \\ \text{Year 12} \end{array}$$

How many Year 11 students chose to study maths this year?

Explanation

Interpret the state matrix in the context of the question.

The second row represents Year 11 students.

Answer

112 students

Worked example 2

In a warehouse, employees are allocated to one of two possible work stations; the picking or packing station. Initially, there were 20 people at the picking station, and 35 people at the packing station. Construct an initial state matrix, S_0, for the information provided.

Explanation

Construct the initial state matrix, making sure to label the rows.

There were initially 20 people at the picking station.

There were initially 35 people at the packing station.

Answer

$$S_0 = \begin{bmatrix} 20 \\ 35 \end{bmatrix} \begin{array}{l} \text{picking} \\ \text{packing} \end{array}$$

Interpreting and constructing transition matrices

See worked example 3

A **transition matrix**, denoted T, is a square matrix that is used to represent the movement or changes of a system between states. It provides information about how much of the data from the system remains the same, or changes. Transition matrices follow the understanding that the value of one state is dependent on the value of a previous state. This helps predict the movement from one time period to another, and determine future and past states.

For example, the following transition matrix can be used to predict the number of workers allocated to each station (A or B) tomorrow based on their designated station today.

$$T = \begin{array}{c} \text{today} \\ \begin{array}{cc} \text{A} & \text{B} \end{array} \\ \begin{bmatrix} 0.2 & 0.6 \\ 0.8 & 0.4 \end{bmatrix} \end{array} \begin{array}{l} \\ \\ \text{A} \\ \text{B} \end{array} \text{ tomorrow}$$

To interpret the movement of workers between stations from one day to the next, it is important to understand that each element represents a proportion of the column that it is in. For example, element t_{11} of the transition matrix indicates that 20% of people working at station A on one day will continue to work at station A the following day. In contrast, element t_{21} indicates that the remaining 80% of people working at station A one day will instead be working at station B the following day.

See worked example 4

When constructing a transition matrix, it is critical to label the direction of movement with respect to columns and rows. Common words used to label the columns are ‘from’, ‘today’, and ‘start’, whilst common words used to label rows are ‘to’, ‘tomorrow’, and ‘end’.

All values must be represented as decimal proportions. As the movement of each individual in a system needs to be accounted for, each column will add up to 1.

Back to contents

Worked example 3

The transition matrix, T, is used to predict the number of people who drive, bus, or train to work on a day-to-day basis.

$$\begin{array}{cc} & \text{today} \\ & \begin{array}{ccc} \text{drive} & \text{bus} & \text{train} \end{array} \\ T = & \begin{bmatrix} 0.68 & 0.34 & 0.58 \\ 0.22 & 0.17 & 0.36 \\ 0.10 & 0.49 & 0.06 \end{bmatrix} \begin{array}{l} \text{drive} \\ \text{bus} \\ \text{train} \end{array} \quad \text{tomorrow} \end{array}$$

What percentage of people who caught the train today will choose to drive tomorrow?

Explanation

Step 1: Identify the relevant element.

The element that corresponds to the proportion of people who caught a train today and will drive tomorrow is located in column 3 (train today), and row 1 (drive tomorrow).

$$\begin{array}{cc} & \text{today} \\ & \begin{array}{ccc} \text{drive} & \text{bus} & \text{train} \end{array} \\ T = & \begin{bmatrix} 0.68 & 0.34 & 0.58 \\ 0.22 & 0.17 & 0.36 \\ 0.10 & 0.49 & 0.06 \end{bmatrix} \begin{array}{l} \text{drive} \\ \text{bus} \\ \text{train} \end{array} \quad \text{tomorrow} \end{array}$$

Step 2: Convert the element value to a percentage.

$0.58 = 58\%$

Answer

58%

Worked example 4

Johnny is an avid fan of fish, and has two different ponds (A and B) for his Koi fish.

The number of Koi fish in each pond can change on a daily basis.

- 10% of the Koi fish in pond A today will remain in pond A tomorrow.
- 15% of the Koi fish in pond B today will be moved to pond A tomorrow.

Construct a transition matrix, T, to represent the movement of Koi fish in each pond from one day to the next.

Explanation

Step 1: Set up a blank matrix, labelled T.

Since there are two possible states, pond A and pond B, a 2×2 matrix is required.

Label the rows and columns 'A' and 'B' to represent the two different ponds.

The columns represent today and the rows represent tomorrow.

$$\begin{array}{cc} & \text{today} \\ & \begin{array}{cc} \text{A} & \text{B} \end{array} \\ T = & \begin{bmatrix} & \\ & \end{bmatrix} \begin{array}{l} \text{A} \\ \text{B} \end{array} \quad \text{tomorrow} \end{array}$$

Step 2: Fill in the matrix with the information provided, converting percentages into decimals.

10% of Koi fish in pond A today will remain in pond A tomorrow.

$t_{11} = 0.10$

15% of Koi fish in pond B today will be moved to pond A tomorrow.

$t_{12} = 0.15$

$$\begin{array}{cc} & \text{today} \\ & \begin{array}{cc} \text{A} & \text{B} \end{array} \\ T = & \begin{bmatrix} 0.10 & 0.15 \\ & \end{bmatrix} \begin{array}{l} \text{A} \\ \text{B} \end{array} \quad \text{tomorrow} \end{array}$$

Continues →

Back to contents

Step 3: Calculate the remaining elements in the transition matrix.

Since each column must add up to 1, the unknown elements can be identified.

Column A: $1 - 0.10 = 0.90$

Column B: $1 - 0.15 = 0.85$

Answer

$$T = \begin{array}{c} \\ \\ \end{array}\begin{array}{cc} \text{today} & \\ \text{A} & \text{B} \end{array}$$

$$T = \left[\begin{array}{cc} 0.10 & 0.15 \\ 0.90 & 0.85 \end{array}\right]\begin{array}{l} \text{A} \\ \text{B} \end{array} \text{ tomorrow}$$

Calculating state matrices

The next state matrix, denoted S_{n+1}, can be calculated recursively by multiplying the transition matrix, T, with the current state matrix, S_n.

See worked example 5

This can be modelled using a recurrence relation of the form

$S_0 = \textit{initial state matrix}, \quad S_{n+1} = T \times S_n$, where

- S_n is the current state matrix
- S_{n+1} is the next state matrix
- T is the transition matrix.

It is also possible to calculate previous state matrices using inverse transition matrices.

See worked example 6

When the value of n is large, calculating a state matrix recursively is time-consuming. In these instances, the following rule can be used to calculate state matrices for any value of n.

See worked example 7

$S_n = T^n \times S_0$, where

- S_n is the current state matrix
- T is the transition matrix
- S_0 is the initial state matrix.

Worked example 5

Every night, a colony of seals can settle on either island A or island B. On Sunday, there were 130 seals on island A and 180 seals on island B.

The initial state and transition matrices are provided.

$$S_0 = \left[\begin{array}{c} 130 \\ 180 \end{array}\right]\begin{array}{l} \text{A} \\ \text{B} \end{array} \qquad T = \begin{array}{c} \text{today} \\ \begin{array}{cc} \text{A} & \text{B} \end{array} \\ \left[\begin{array}{cc} 0.2 & 0.6 \\ 0.8 & 0.4 \end{array}\right] \end{array}\begin{array}{l} \\ \\ \text{A} \\ \text{B} \end{array} \text{ tomorrow}$$

a. Determine the matrix recurrence relation that can be used to calculate the number of seals on each island every day.

Explanation

Substitute S_0 and T into the recurrence relation form.

$S_0 = \textit{initial state matrix}, \quad S_{n+1} = T \times S_n$

Answer

$$S_0 = \left[\begin{array}{c} 130 \\ 180 \end{array}\right], \quad S_{n+1} = \left[\begin{array}{cc} 0.2 & 0.6 \\ 0.8 & 0.4 \end{array}\right] \times S_n$$

Continues →

Back to contents

b. Use recursion to calculate the number of seals on each island on Tuesday, rounded to the nearest whole number.

Explanation

Step 1: Use the recurrence relation to calculate the number of seals on Monday, S_1.

$S_1 = T \times S_0$

$S_1 = \begin{bmatrix} 0.2 & 0.6 \\ 0.8 & 0.4 \end{bmatrix} \times \begin{bmatrix} 130 \\ 180 \end{bmatrix}$

$S_1 = \begin{bmatrix} 134 \\ 176 \end{bmatrix}$

Step 2: Use the recurrence relation to calculate the number of seals on Tuesday, S_2.

$S_2 = T \times S_1$

$S_2 = \begin{bmatrix} 0.2 & 0.6 \\ 0.8 & 0.4 \end{bmatrix} \times \begin{bmatrix} 134 \\ 176 \end{bmatrix}$

$S_2 = \begin{bmatrix} 132.4 \\ 177.6 \end{bmatrix}$

Answer

Island A: 132 seals

Island B: 178 seals

Worked example 6

Consider the following state matrix after 3 periods, S_3, and transition matrix, T.

$S_3 = \begin{bmatrix} 120 \\ 150 \end{bmatrix} \quad T = \begin{bmatrix} 0.84 & 0.21 \\ 0.16 & 0.79 \end{bmatrix}$

Determine S_2, rounding the matrix elements to the nearest whole number.

Explanation

Step 1: Substitute S_2 and S_3 into the equation $S_{n+1} = T \times S_n$.

$S_3 = T \times S_2$

Step 2: Solve for S_2.

Pre-multiply both sides by T^{-1}.

$T^{-1} \times S_3 = T^{-1} \times T \times S_2$

$T^{-1} \times S_3 = S_2$

$S_2 = \begin{bmatrix} 0.84 & 0.21 \\ 0.16 & 0.79 \end{bmatrix}^{-1} \times \begin{bmatrix} 120 \\ 150 \end{bmatrix}$

$S_2 = \begin{bmatrix} 100.47... \\ 169.52... \end{bmatrix}$

Answer

$\begin{bmatrix} 100 \\ 170 \end{bmatrix}$

Worked example 7

Consider the following initial state matrix and transition matrix.

$S_0 = \begin{bmatrix} 213 \\ 142 \end{bmatrix} \quad T = \begin{bmatrix} 0.3 & 0.9 \\ 0.7 & 0.1 \end{bmatrix}$

Determine S_{10}, rounding the matrix elements to the nearest whole number.

Explanation

Step 1: Substitute the known matrices into the rule $S_n = T^n \times S_0$.

$S_n = \begin{bmatrix} 0.3 & 0.9 \\ 0.7 & 0.1 \end{bmatrix}^n \times \begin{bmatrix} 213 \\ 142 \end{bmatrix}$

Step 2: Calculate S_{10}.

The n value required is $n = 10$.

$S_{10} = \begin{bmatrix} 0.3 & 0.9 \\ 0.7 & 0.1 \end{bmatrix}^{10} \times \begin{bmatrix} 213 \\ 142 \end{bmatrix}$

$S_{10} = \begin{bmatrix} 199.76... \\ 155.23... \end{bmatrix}$

Continues →

Back to contents

Answer

$$\begin{bmatrix} 200 \\ 155 \end{bmatrix}$$

Using Leslie matrices in applied scenarios

A **Leslie matrix**, denoted L, is a unique application of transition matrices that can be used to model the growth of a population and its age distribution over time. Generally, population growth is modelled year-to-year.

When applying a Leslie matrix to a population, only the females in the population are considered. In contrast to a standard transition matrix, the columns of a Leslie matrix do not add up to 1.

Leslie matrices are square matrices of size n and are presented in the following form.

$$\begin{array}{ccl} & \text{age} & \\ & \begin{array}{ccc} 0 & 1 & 2 \end{array} & \\ L = & \begin{bmatrix} F_0 & F_1 & F_2 \\ P_0 & 0 & 0 \\ 0 & P_1 & 0 \end{bmatrix} & \begin{array}{l} \text{fertility rate} \\ \text{survival rate from age 0 – age 1} \\ \text{survival rate from age 1 – age 2} \end{array} \end{array}$$

In a Leslie matrix, F represents fertility rates.

- F_0 is the average number of females born to each female that is less than 1 year old.
- F_1 is the average number of females born to each 1-year-old female.
- F_2 is the average number of females born to each 2-year-old female.

As such, F_n is the average number of females born to each female that is n years old.

In a Leslie matrix, P represents survival rates.

- P_0 is the average survival rate of females that are less than 1 year old.
- P_1 is the average survival rate of females that are 1 year old.
- The remaining elements contain zeros (0), allowing the Leslie matrix to calculate the number of females that survive at each age group, when multiplied with an initial state matrix.

As such, P_n is the average survival rate of females that are n years old.

As this is a 3×3 matrix, there is no P_2 value. This means that all females aged 2 years and over will not survive into the following year.

A state matrix S_n, is used to represent the breakdown of female age groups in the population. Future state matrices can be calculated by using the Leslie matrix as a transition matrix.

Worked example 8

Scientists have collected the yearly fertility and survival rates for siamese fighting fish. A small sample of the species is under observation. The following Leslie matrix for the survival rate of siamese fighting fish can be used to model the growth of the sample population.

$$\begin{array}{cc} & \text{age} \\ & \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ L = & \begin{bmatrix} 0.05 & 0.80 & 0.90 \\ 0.85 & 0 & 0 \\ 0 & 0.75 & 0 \end{bmatrix} \end{array}$$

The following initial state matrix represents the female population of the sample being observed at the start of the study.

$$S_0 = \begin{bmatrix} 50 \\ 40 \\ 30 \end{bmatrix} \begin{array}{l} 0 \\ 1 \quad \text{age} \\ 2 \end{array}$$

Continues →

Back to contents

a. Interpret the element in row 2 and column 1 of the Leslie matrix.

Explanation

Step 1: Locate the relevant element on the Leslie matrix.

$$L = \begin{array}{c} \text{age} \\ \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{bmatrix} 0.05 & 0.80 & 0.90 \\ 0.85 & 0 & 0 \\ 0 & 0.75 & 0 \end{bmatrix} \end{array}$$

Step 2: Interpret the element.

0.85 is the average survival rate of females that are less than 1 year old.

Answer

85% of the female population that is less than 1 year old will survive into the next year.

---

b. Interpret the element in row 1 and column 3 of the Leslie matrix.

Explanation

Step 1: Locate the relevant element on the matrix.

$$L = \begin{array}{c} \text{age} \\ \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{bmatrix} 0.05 & 0.80 & 0.90 \\ 0.85 & 0 & 0 \\ 0 & 0.75 & 0 \end{bmatrix} \end{array}$$

Step 2: Interpret the element.

0.90 is the average number of females born to each female that is 2 years old.

Answer

On average, there will be 0.90 females born to each 2-year-old female.

---

c. What is the estimated total number of female siamese fighting fish after two years?

Explanation

Step 1: Substitute S_0 and L into the recurrence relation.

$S_0 = \textit{initial state matrix}, \quad S_{n+1} = L \times S_n$

$$S_0 = \begin{bmatrix} 50 \\ 40 \\ 30 \end{bmatrix}, \quad S_{n+1} = \begin{bmatrix} 0.05 & 0.80 & 0.90 \\ 0.85 & 0 & 0 \\ 0 & 0.75 & 0 \end{bmatrix} \times S_n$$

Step 2: Use the recurrence relation to calculate S_1.

$S_1 = L \times S_0$

$$S_1 \begin{bmatrix} 0.05 & 0.80 & 0.90 \\ 0.85 & 0 & 0 \\ 0 & 0.75 & 0 \end{bmatrix} \times \begin{bmatrix} 50 \\ 40 \\ 30 \end{bmatrix}$$

$$S_1 = \begin{bmatrix} 61.50 \\ 42.50 \\ 30.00 \end{bmatrix}$$

Step 3: Use the recurrence relation to calculate S_2.

$S_2 = L \times S_1$

$$S_2 = \begin{bmatrix} 0.05 & 0.80 & 0.90 \\ 0.85 & 0 & 0 \\ 0 & 0.75 & 0 \end{bmatrix} \times \begin{bmatrix} 61.50 \\ 42.50 \\ 30.00 \end{bmatrix}$$

$$S_2 = \begin{bmatrix} 64.075 \\ 52.275 \\ 31.875 \end{bmatrix}$$

Step 4: Round the elements in S_2 to the nearest whole number.

There can only be a whole number of siamese fighting fish at each age.

$$S_2 \approx \begin{bmatrix} 64 \\ 52 \\ 32 \end{bmatrix}$$

Step 5: Sum the elements in S_2.

$64 + 52 + 32 = 148$

Answer

148 female siamese fighting fish

Continues →

d. Determine S_8, the expected female population in the study after 8 years. Round the matrix elements to the nearest whole number.

Explanation

Step 1: Substitute S_0 and L into the rule $S_n = L^n \times S_0$.

$$S_n = \begin{bmatrix} 0.05 & 0.80 & 0.90 \\ 0.85 & 0 & 0 \\ 0 & 0.75 & 0 \end{bmatrix}^n \times \begin{bmatrix} 50 \\ 40 \\ 30 \end{bmatrix}$$

Step 2: Calculate S_8.

The n value required is $n = 8$.

$$S_8 = \begin{bmatrix} 0.05 & 0.80 & 0.90 \\ 0.85 & 0 & 0 \\ 0 & 0.75 & 0 \end{bmatrix}^8 \times \begin{bmatrix} 50 \\ 40 \\ 30 \end{bmatrix}$$

$$S_8 = \begin{bmatrix} 127.95... \\ 97.23... \\ 65.64... \end{bmatrix}$$

Answer

$$\begin{array}{cc} \begin{bmatrix} 128 \\ 97 \\ 66 \end{bmatrix} & \begin{matrix} 0 \\ 1 \\ 2 \end{matrix} \;\text{age} \end{array}$$

Exam question breakdown

VCAA 2019 Exam 1 Matrices Q4

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.

$$T = \begin{array}{c} \text{this question} \\ \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} \end{array} \begin{matrix} \\ \\ A \\ B \\ C \\ D \\ E \end{matrix} \;\text{next question}$$

What answer did Stella choose for question number six?

A. A **B.** B **C.** C **D.** D **E.** E

Explanation

Step 1: Identify S_0, the initial state matrix.

S_0 would be a 5×1 matrix due to the five options (A, B, C, D, E). Since Stella chose E for the first question, the elements in the first 4 rows are 0, and the element in the fifth row is 1.

$$S_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Step 2: Substitute S_0 and T into the rule $S_n = T^n \times S_0$.

$$S_n = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}^n \times \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Step 3: Calculate S_5.

The n value required is $n = 5$. This is because the initial state matrix, with $n = 0$, represents Stella's answer to question 1.

$$S_5 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}^5 \times \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$$S_5 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Answer

E

56% of students answered this question correctly.

24% of students incorrectly answered C. This may be because some students raised their transition matrix to a power of 6 instead of 5 when using a rule to answer this question. Students likely overlooked the fact that their initial state matrix corresponds to Stella's answer to question 1, so only 5 more iterations of the matrix multiplication were required.

Back to contents

7G Questions

Interpreting and constructing state matrices

1. A transition model follows the recurrence relation.

$$S_0 = \begin{bmatrix} 27 \\ 45 \\ 15 \end{bmatrix}, \quad S_{n+1} = \begin{bmatrix} 0.5 & 0.3 & 0.8 \\ 0.3 & 0.3 & 0.1 \\ 0.2 & 0.4 & 0 \end{bmatrix} \times S_n$$

What is the initial state matrix?

A. $S_0 = \begin{bmatrix} 27 \\ 45 \\ 15 \end{bmatrix}$

B. $S_0 = \begin{bmatrix} 0.5 & 0.3 & 0.8 \\ 0.3 & 0.3 & 0.1 \\ 0.2 & 0.4 & 0 \end{bmatrix}$

C. $S_{n+1} = \begin{bmatrix} 0.5 & 0.3 & 0.8 \\ 0.3 & 0.3 & 0.1 \\ 0.2 & 0.4 & 0 \end{bmatrix} \times S_n$

D. $T = \begin{bmatrix} 0.5 & 0.3 & 0.8 \\ 0.3 & 0.3 & 0.1 \\ 0.2 & 0.4 & 0 \end{bmatrix}$

2. People attending a local carnival have the choice of enjoying 3 different rides: 'Timid Tiger' (T), 'Wild Warrior' (W) or 'The Devastator' (D), or can choose not to take a ride at all (N). After a ride, attendees can choose to take the same ride, take a different ride, or stop taking rides altogether.

The ride that attendees first chose, on a particular day, is shown in the initial state matrix, S_0. The way in which attendees are predicted to move between rides is modelled by the transition matrix, T.

$$S_0 = \begin{array}{c c} \begin{bmatrix} 32 \\ 85 \\ 57 \\ 24 \end{bmatrix} & \begin{array}{c} \text{T} \\ \text{W} \\ \text{D} \\ \text{N} \end{array} \end{array} \qquad T = \begin{array}{c c c} & \text{this ride} & \\ & \begin{array}{cccc} \text{T} & \text{W} & \text{D} & \text{N} \end{array} & \\ & \begin{bmatrix} 0.2 & 0.1 & 0.2 & 0 \\ 0.6 & 0.3 & 0.3 & 0 \\ 0.1 & 0.5 & 0.1 & 0 \\ 0.1 & 0.1 & 0.4 & 1 \end{bmatrix} & \begin{array}{c} \text{T} \\ \text{W} \\ \text{D} \\ \text{N} \end{array} \text{ next ride} \end{array}$$

a. How many people rode 'The Devastator' as their first ride of the day?

b. How many people attended the carnival on the given day?

c. What percentage of people chose not to take a ride at all when attending the carnival, rounded to the nearest percent?

3. The food company 'Barnets' releases 3 all-new flavours of their popular savoury biscuit brand: sweet-and-sour (S), hot-and-spicy (H), and American mustard (A).

a. When the flavours were first released, 47 270 boxes of sweet-and-sour, 39 231 boxes of hot-and-spicy, and 56 159 boxes of American mustard were sold in the first week.

Construct an initial state matrix, S_0, to represent this information.

b. At the end of the 10th week, Barnets noticed that their new flavours weren't selling as well as they hoped, with significant declines in sales. Only 21 452 boxes of sweet-and-sour, 23 527 boxes of hot-and-spicy, and 12 339 boxes of American mustard were sold in that week.

Construct a state matrix to represent this information.

Interpreting and constructing transition matrices

4. A transition model follows the recurrence relation

$$S_0 = \begin{bmatrix} 50 \\ 40 \end{bmatrix}, \quad S_{n+1} \begin{bmatrix} 0.2 & 0.7 \\ 0.8 & 0.3 \end{bmatrix} \times S_n$$

What is the transition matrix?

A. $S_0 = \begin{bmatrix} 50 \\ 40 \end{bmatrix}$

B. $S_0 = \begin{bmatrix} 0.2 & 0.7 \\ 0.8 & 0.3 \end{bmatrix}$

C. $S_{n+1} = \begin{bmatrix} 0.2 & 0.7 \\ 0.8 & 0.3 \end{bmatrix} \times S_n$

D. $T = \begin{bmatrix} 0.2 & 0.7 \\ 0.8 & 0.3 \end{bmatrix}$

5. When recycling, people either fold (F) or scrunch (S) their paper. A researcher finds that she can predict a person's technique each time they recycle paper using the following transition matrix.

$$\begin{array}{cc} & \text{this time} \\ & \begin{array}{cc} \text{F} & \text{S} \end{array} \\ T = & \begin{bmatrix} 0.91 & 0.23 \\ 0.09 & 0.77 \end{bmatrix} \begin{array}{c} \text{F} \\ \text{S} \end{array} \text{ next time} \end{array}$$

a. What percentage of people who fold their paper for recycling this time are expected to scrunch next time?

b. What percentage of people who scrunch their paper for recycling this time are expected to fold next time?

c. What percentage of people who fold their paper for recycling this time are expected to fold again next time?

d. What percentage of people who scrunch their paper for recycling this time are expected to scrunch again next time?

6. A teacher finds that 35% of students who buy lunch today (B) will buy lunch tomorrow, and 50% of students who do not buy lunch today (N) will buy lunch tomorrow.

The teacher also recorded that 36 students bought lunch today, whilst 15 students did not.

Use this information to construct a transition matrix, T.

Calculating state matrices

7. Consider the following initial state matrix and transition matrix.

$S_0 = \begin{bmatrix} 20 \\ 10 \end{bmatrix} \quad T = \begin{bmatrix} 0.2 & 0.5 \\ 0.8 & 0.5 \end{bmatrix}$

a. Which calculation can be used to determine S_1?

A. $S_1 = \begin{bmatrix} 20 \\ 10 \end{bmatrix} \times \begin{bmatrix} 0.2 & 0.5 \\ 0.8 & 0.5 \end{bmatrix}$

B. $S_1 = \begin{bmatrix} 20 \\ 10 \end{bmatrix}^{-1} \times \begin{bmatrix} 0.2 & 0.5 \\ 0.8 & 0.5 \end{bmatrix}$

C. $S_1 = \begin{bmatrix} 0.2 & 0.5 \\ 0.8 & 0.5 \end{bmatrix} \times \begin{bmatrix} 20 \\ 10 \end{bmatrix}$

D. $S_1 = \begin{bmatrix} 0.2 & 0.5 \\ 0.8 & 0.5 \end{bmatrix}^{-1} \times \begin{bmatrix} 20 \\ 10 \end{bmatrix}$

b. Calculate S_2, rounding the matrix elements to one decimal place.

c. Calculate S_3, rounding the matrix elements to one decimal place.

8. Consider the following initial state matrix and transition matrix.

$S_0 = \begin{bmatrix} 90 \\ 1968 \end{bmatrix} \quad T = \begin{bmatrix} 0.2 & 0.5 \\ 0.8 & 0.5 \end{bmatrix}$

a. Which calculation can be used to determine S_7?

A. $S_7 = \begin{bmatrix} 0.2 & 0.5 \\ 0.8 & 0.5 \end{bmatrix}^7 \times \begin{bmatrix} 90 \\ 1968 \end{bmatrix}$

B. $S_7 = \begin{bmatrix} 90 \\ 1968 \end{bmatrix}^7 \times \begin{bmatrix} 0.2 & 0.5 \\ 0.8 & 0.5 \end{bmatrix}$

C. $S_7 = \begin{bmatrix} 0.2 & 0.5 \\ 0.8 & 0.5 \end{bmatrix} \times \begin{bmatrix} 90 \\ 1968 \end{bmatrix}^7$

D. $S_7 = \begin{bmatrix} 90 \\ 1968 \end{bmatrix} \times \begin{bmatrix} 0.2 & 0.5 \\ 0.8 & 0.5 \end{bmatrix}^7$

b. Calculate S_5, rounding the matrix elements to two decimal places.

c. Calculate S_{40}, rounding the matrix elements to two decimal places.

9. Consider the following state matrix after one period, S_1, and transition matrix, T.

$S_1 = \begin{bmatrix} 169 \\ 116 \\ 125 \end{bmatrix} \quad T = \begin{bmatrix} 0.5 & 0.2 & 0.5 \\ 0.2 & 0.6 & 0.1 \\ 0.3 & 0.2 & 0.4 \end{bmatrix}$

Calculate S_0.

10. A psychology experiment is conducted where 50 participants decide to eat either a red or blue gummy every hour. A transition matrix, T, has been constructed to try and predict the number of participants who will eat each gummy each hour. The number of participants who decided to eat each of the gummies at the start of the experiment is shown in matrix S_0.

$$S_0 = \begin{array}{c} \begin{bmatrix} 40 \\ 10 \end{bmatrix} \begin{array}{l} \text{red} \\ \text{blue} \end{array} \end{array} \qquad T = \begin{array}{c} \text{this hour} \\ \begin{array}{cc} \text{red} & \text{blue} \end{array} \\ \begin{bmatrix} 0.3 & 0.6 \\ 0.7 & 0.4 \end{bmatrix} \end{array} \begin{array}{l} \text{red} \\ \text{blue} \end{array} \ \text{next hour}$$

a. How many participants are predicted to eat the red gummy in the third hour?

b. How many participants are predicted to eat the blue gummy in the sixth hour?

---

11. Superstar travel agent Tal has 300 regular customers who often go on trips to Mars. Based on previous years, Tal has come up with a transition matrix, T, to help him predict whether his customers will go on holiday or not. In 2050, 167 customers went to Mars and 133 customers did not.

$$T = \begin{array}{c} \text{this year} \\ \begin{array}{cc} \text{trip} & \text{no trip} \end{array} \\ \begin{bmatrix} 0.4 & 0.7 \\ 0.6 & 0.3 \end{bmatrix} \end{array} \begin{array}{l} \text{trip} \\ \text{no trip} \end{array} \ \text{next year} \qquad S_0 = \begin{bmatrix} 167 \\ 133 \end{bmatrix} \begin{array}{l} \text{trip} \\ \text{no trip} \end{array}$$

a. Calculate how many customers are predicted to go to Mars in 2060, rounded to the nearest whole number.

b. To make a profit, Tal must send at least 150 holiday goers to Mars every year. Determine if he will be profitable or not in 2063.

Using Leslie matrices in applied scenarios

12. The following Leslie matrix can be used to model the growth of a population of hedgehogs.

$$L = \begin{array}{c} \text{age} \\ \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0.2 & 0.3 & 0.5 & 0.8 \\ 0.9 & 0 & 0 & 0 \\ 0 & 0.7 & 0 & 0 \\ 0 & 0 & 0.6 & 0 \end{bmatrix} \end{array}$$

The female population of a sample of hedgehogs at the start of a study is shown in the initial state matrix, S_0.

$$S_0 = \begin{bmatrix} 120 \\ 100 \\ 85 \\ 65 \end{bmatrix} \begin{array}{l} 0 \\ 1 \\ 2 \\ 3 \end{array} \ \text{age}$$

a. Interpret the element in row 4 and column 3 of the Leslie matrix.

A. On average, 60% of females will be born to a 2-year-old female.

B. On average, there will be 0.6 females born to each 2-year-old female.

C. 0.6% of the female population that is 2 years old will survive into the next year.

D. 60% of the female population that is 2 years old will survive into the next year.

b. Interpret the element in row 1 and column 4 of the Leslie matrix.

A. On average, 80% of females will be born to a 3-year-old female.

B. On average, there will be 0.8 females born to each 3-year-old female.

C. 0.8% of the female population that is 3 years old will survive into the next year.

D. 80% of the female population that is 3 years old will survive into the next year.

13. The Leslie matrix for a critically endangered species is given.

$$L = \begin{array}{c} \text{age} \\ \begin{array}{cc} 0 & 1 \end{array} \\ \begin{bmatrix} 0.8 & 0.6 \\ 0.4 & 0 \end{bmatrix} \end{array}$$

The initial female population has also been provided.

$$S_0 = \begin{bmatrix} 45 \\ 20 \end{bmatrix} \begin{array}{c} 0 \\ 1 \end{array} \text{ age}$$

a. Using a matrix recurrence relation, determine the expected female population after 2 years.

b. Determine the state matrix that represents the expected female population after 10 years.

c. Has the female population of the critically endangered species increased after 10 years? If so, by how much?

---

14. Scientists have been monitoring the survival and reproductive rates of a rare species of chameleon on the island of Madagascar. They've collated their findings in the following Leslie matrix.

$$L = \begin{array}{c} \text{age} \\ \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0.4 & 0.7 & 0.8 & 0.6 \\ 0.8 & 0 & 0 & 0 \\ 0 & 0.6 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \end{bmatrix} \end{array}$$

The female population of the species at the start of the study is shown by the initial state matrix, S_0.

$$S_0 = \begin{bmatrix} 59 \\ 64 \\ 48 \\ 32 \end{bmatrix} \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} \text{ age}$$

If 45% of the population of the species of chameleons are female, what is the expected total population of the species after 15 years?

Joining it all together

15. In Circle Hill, every household has tacos for Mexican night once a week. They either have soft (S) or hard (H) shelled tacos; they cannot have both.

A survey of the neighbourhood found that 40% of families that eat soft shell tacos (S) for Mexican night this week will eat hard shell tacos next week. 65% of families that eat hard shell tacos (H) for Mexican night this week will eat hard shell tacos next week.

a. Use this information to construct a transition matrix, T.

b. What percentage of families that eat soft shell tacos this week are expected to eat soft shell tacos next week?

c. What percentage of families that eat hard shell tacos this week are expected to eat soft shell tacos next week?

There are 400 families in this neighbourhood.

d. If 200 families eat soft shell tacos this week, how many families will eat hard shell tacos next week?

e. In a particular week, 120 families eat soft shell tacos. How many families are expected to change taco shells the following week?

16. A new cafeteria is opening at Edrolo High and the principal wants to know how many students will be buying food at the cafeteria each day. Mr Barry is a matrices enthusiast and says he can predict the number of students that will buy food from the cafeteria each day. By surveying the students, he constructs a transition matrix, T, and an initial state matrix, S_0, based on the first Monday of school.

$$S_0 = \begin{array}{c} \begin{bmatrix} 122 \\ 78 \end{bmatrix} \end{array} \begin{array}{l} \text{buy} \\ \text{not buy} \end{array} \qquad T = \begin{array}{c} \text{today} \\ \begin{array}{cc} \text{buy} & \text{not buy} \end{array} \\ \begin{bmatrix} 0.5 & 0.3 \\ 0.5 & 0.7 \end{bmatrix} \end{array} \begin{array}{l} \\ \\ \text{buy} \\ \text{not buy} \end{array} \quad \text{tomorrow}$$

a. How many students will be expected to buy food on the first Tuesday?

b. How many students will be expected to buy food on the first Friday?

---

17. In the lead up to exams, teachers are offering study sessions in English (E), maths (M) and history (H). To ensure that there are enough teachers to help the students, they construct a transition matrix, T, which will help to predict the next study session a student will attend.

$$T = \begin{array}{c} \text{this period} \\ \begin{array}{ccc} \text{E} & \text{M} & \text{H} \end{array} \\ \begin{bmatrix} 0.29 & 0.32 & 0.29 \\ 0.32 & 0.62 & 0.12 \\ 0.39 & 0.06 & 0.59 \end{bmatrix} \end{array} \begin{array}{l} \\ \\ \text{E} \\ \text{M} \\ \text{H} \end{array} \quad \text{next period}$$

- 14 students attended the first English study session.
- 23 students attended the first maths study session.
- 13 students attended the first history study session.

a. Construct the initial state matrix, S_0.

b. How many students are predicted to attend the third maths study session?

c. There must be at least one teacher for every 5 students.
How many teachers will be required for the sixth history study session?

---

18. Scientists have been investigating the population changes of green tree frogs. They've collected their findings on the yearly survival and fertility rates in the following Leslie matrix.

$$L = \begin{array}{c} \text{age} \\ \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 1 & 1 & 2 & 3 \\ 0.9 & 0 & 0 & 0 \\ 0 & 0.7 & 0 & 0 \\ 0 & 0 & 0.7 & 0 \end{bmatrix} \end{array}$$

A study is being conducted on an isolated group of green tree frogs to verify if their findings are accurate. Initially, there were

- 10 females under 1 year old
- 12 female one-year-olds
- 14 female two-year-olds
- 20 female three-year-olds.

a. Construct the initial state matrix, S_0.

b. What is the average number of female babies that are expected to be born to each two-year-old female green tree frog?

c. What is the expected number of three-year-old female green tree frogs after three years of observations?

d. What is the expected number of female green tree frogs after 10 years of observations?

e. If 60% of the population are female, determine the total population of green tree frogs after 2 years of observations?

Back to contents

Exam practice

19. A travel company is studying the choice between air (A), land (L), and sea (S) or no (N) travel by some of its customers each year.

Matrix T contains the percentages of customers and their choice of travel from year to year.

$$\begin{array}{cc} & \text{this year} \\ & \begin{array}{cccc} \text{A} & \text{L} & \text{S} & \text{N} \end{array} \\ T = & \begin{bmatrix} 0.65 & 0.25 & 0.25 & 0.50 \\ 0.15 & 0.60 & 0.20 & 0.15 \\ 0.05 & 0.10 & 0.25 & 0.20 \\ 0.15 & 0.05 & 0.30 & 0.15 \end{bmatrix} \end{array} \begin{array}{c} \text{A} \\ \text{L} \\ \text{S} \\ \text{N} \end{array} \quad \text{next year}$$

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.

$$S_0 = \begin{bmatrix} 520 \\ 320 \\ 80 \\ 80 \end{bmatrix} \begin{array}{c} \text{A} \\ \text{L} \\ \text{S} \\ \text{N} \end{array}$$

Matrix S_1 shows the number of customers who chose each type of travel in 2015.

$$S_1 = TS_0 = \begin{bmatrix} 478 \\ d \\ e \\ f \end{bmatrix} \begin{array}{c} \text{A} \\ \text{L} \\ \text{S} \\ \text{N} \end{array}$$

Write the values missing from matrix S_1 (d, e, f) in the boxes provided. (1 MARK)

$d =$ ☐ $e =$ ☐ $f =$ ☐

VCAA 2016 Exam 2 Matrices Q3a

85% of students answered this question correctly.

---

20. Senior students at a high school must choose one elective activity in each of the four terms in 2018.

Their choices are communication (C), investigation (I), problem-solving (P) and services (S).

The transition matrix, T, shows the way in which senior students are expected to change their choice of elective activity from term to term.

$$\begin{array}{cc} & \text{this term} \\ & \begin{array}{cccc} \text{C} & \text{I} & \text{P} & \text{S} \end{array} \\ T = & \begin{bmatrix} 0.4 & 0.2 & 0.3 & 0.1 \\ 0.2 & 0.4 & 0.1 & 0.3 \\ 0.2 & 0.3 & 0.3 & 0.4 \\ 0.2 & 0.1 & 0.3 & 0.2 \end{bmatrix} \end{array} \begin{array}{c} \text{C} \\ \text{I} \\ \text{P} \\ \text{S} \end{array} \quad \text{next term}$$

Let S_n be the state matrix for the number of senior students expected to choose each elective activity in term n.

For the given matrix S_1, a matrix rule that can be used to predict the number of senior students in each elective activity in terms 2, 3 and 4 is

$$S_1 = \begin{bmatrix} 300 \\ 200 \\ 200 \\ 300 \end{bmatrix}, \quad S_{n+1} = TS_n$$

a. How many senior students will not change their elective activity from term 1 to term 2? (1 MARK)

b. Complete S_2, the state matrix for term 2. (1 MARK)

$$S_2 = \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix} \begin{array}{c} \text{C} \\ \text{I} \\ \text{P} \\ \text{S} \end{array}$$

c. Of the senior students expected to choose investigation (I) in term 3, what percentage choose service (S) in term 2? (2 MARKS)

VCAA 2017 Exam 2 Matrices Q3

Part **a**: **47%** of students answered this question correctly.

Part **b**: **73%** of students answered this question correctly.

Part **c**: The average mark on this question was **0.3**.

Back to contents

Questions from multiple lessons

Data analysis

21. The Year 12 cohort at Edrolo High just sat their end-of-year Psychology exam. Their exam scores were approximately normally distributed with a mean of 70.2% and a standard deviation of 10.1%. 278 students sat the exam. The number of students expected to have passed the exam (received a mark over 50%) is closest to

A. 7 **B.** 14 **C.** 236 **D.** 264 **E.** 271

Adapted from VCAA 2016 Exam 1 Data analysis Q4

Matrices *Year 11 content*

22. The elements in matrix M are determined by the rule $m_{ij} = 3i + 2j$.

Which of the following **cannot** be matrix M?

A. $\begin{bmatrix} 5 & 7 \\ 8 & 10 \end{bmatrix}$ **B.** $[5]$ **C.** $[5 \quad 8 \quad 11]$ **D.** $\begin{bmatrix} 5 \\ 8 \\ 11 \\ 14 \end{bmatrix}$ **E.** $\begin{bmatrix} 5 & 7 & 9 \\ 8 & 10 & 12 \\ 11 & 13 & 15 \end{bmatrix}$

Adapted from VCAA 2017NH Exam 1 Matrices Q4

Matrices *Year 11 content*

23. An op-shop sells tops (T), pants (P), and dresses (D).

The number of each sold on Monday, Tuesday and Wednesday is shown in matrix N.

$$N = \begin{array}{c} \begin{array}{ccc} \text{T} & \text{P} & \text{D} \end{array} \\ \begin{bmatrix} 14 & 10 & 21 \\ 23 & 16 & 27 \\ 28 & 15 & 22 \end{bmatrix} \end{array} \begin{array}{l} \\ \text{Monday} \\ \text{Tuesday} \\ \text{Wednesday} \end{array}$$

a. What was the total number of tops sold over the three days? (1 MARK)

b. Interpret element n_{32}. (1 MARK)

Consider the following matrix equation.

$$\begin{bmatrix} 14 & 10 & 21 \\ 23 & 16 & 27 \\ 28 & 15 & 22 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 225 \\ 322.50 \\ 305 \end{bmatrix}$$

- x is the cost of a top.
- y is the cost of one pair of pants.
- z is the cost of a dress.

c. What is the cost of a dress? (1 MARK)

d. The following matrix equation shows that the total value of all clothing sold over Monday and Tuesday is \$547.50.

$$M \times \begin{bmatrix} 225 \\ 322.50 \\ 305 \end{bmatrix} = [547.50]$$

Given that the order of matrix M is 1×3, write down matrix M. (1 MARK)

Adapted from VCAA 2017 Exam 2 Matrices Q1a-c

Back to contents

7H The equilibrium state matrix

STUDY DESIGN DOT POINT

- 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)

7A 7B 7C 7D 7E 7F 7G 7H 7I

KEY SKILLS

During this lesson, you will be:

- calculating the equilibrium state matrix
- interpreting the equilibrium state matrix.

KEY TERMS

- Equilibrium state matrix
- Steady state matrix

When state matrices are used to model data that transforms over time, the data may eventually settle to a point where there is no visible change between different states, even though the transition matrix is still functioning. This can provide useful insight into the long-term projections of data.

Calculating the equilibrium state matrix

The **equilibrium state matrix** (often called the **steady state matrix**) is the state matrix which has no difference compared to the matrix occurring after it.

From the recurrence relation

$S_0 = \textit{initial state matrix}, \quad S_{n+1} = T \times S_n,$

the equilibrium state matrix is the matrix in which $S_n = S_{n+1}$.

Although each data point is still affected by the transition matrix, the changes that occur ultimately cancel out, resulting in the previous state matrix. In other words, there is no net change of the elements within the state matrix. Once the equilibrium state has been reached, it will remain that way for all future states.

As large values of n are being used, the equilibrium state matrix should be calculated using a rule, $S_n = T^n \times S_0$, rather than a recurrence relation.

As T^n cannot be calculated for when n is infinite, an approximated value of the equilibrium can be found by calculating S_n for large values of n. Usually, the equilibrium state matrix will be found somewhere between $n = 15$ and $n = 30$.

Back to contents

Worked example 1

If $S_0 = \begin{bmatrix} 29 \\ 32 \end{bmatrix}$ and $T = \begin{bmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{bmatrix}$, calculate the equilibrium state matrix.

Explanation

Step 1: Calculate the state matrix for a large n.

$S_n = T^n \times S_0$

Usually the equilibrium state matrix will have occurred by the time $n = 30$.

$S_{30} = T^{30} \times S_0$

$$= \begin{bmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{bmatrix}^{30} \times \begin{bmatrix} 29 \\ 32 \end{bmatrix}$$

$$= \begin{bmatrix} 24.4 \\ 36.6 \end{bmatrix}$$

Step 2: Verify the equilibrium state matrix by comparing with S_{n+1}.

$S_{31} = T^{31} \times S_0$

$$= \begin{bmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{bmatrix}^{31} \times \begin{bmatrix} 29 \\ 32 \end{bmatrix}$$

$$= \begin{bmatrix} 24.4 \\ 36.6 \end{bmatrix}$$

Since S_{30} and S_{31} are equal, this is the equilibrium state matrix.

If S_{30} and S_{31} were not equal, steps 1 and 2 would have to be repeated by substituting in a higher value of n.

Answer

$$\begin{bmatrix} 24.4 \\ 36.6 \end{bmatrix}$$

Interpreting the equilibrium state matrix

An equilibrium state matrix contains information about the long term expectations for a given scenario. Using this, the equilibrium state matrix and its elements can be interpreted in relation to the context.

Worked example 2

The school canteen planned to offer two new menu options: spring rolls (S), and avo toast (A). All 100 students order one of the options each day. Their first ever orders are represented in matrix S_0. In order to predict how many of each option will be required, the canteen staff made a transition matrix, T.

$$S_0 = \begin{array}{c} \begin{bmatrix} 75 \\ 25 \end{bmatrix} \end{array}\begin{array}{l} S \\ A \end{array} \qquad T = \begin{array}{c} \text{today} \\ \begin{array}{cc} S & A \end{array} \\ \begin{bmatrix} 0.54 & 0.71 \\ 0.46 & 0.29 \end{bmatrix} \end{array}\begin{array}{l} \\ \\ S \\ A \end{array} \ \text{tomorrow}$$

From the matrix equation $S_n = T^n \times S_0$, the equilibrium state matrix is approximately:

$$\begin{bmatrix} 60.68 \\ 39.32 \end{bmatrix}\begin{array}{l} S \\ A \end{array}$$

Using the equilibrium state matrix values, rounded to the nearest whole number, what can the cafe expect in the long term?

Explanation

Interpret the equilibrium state matrix.

The elements in the equilibrium state matrix describe the expected daily sales in the long-term.

The first row corresponds to spring rolls, and the second row corresponds to avo toasts.

Answer

In the long term, the cafe can expect to sell 61 spring rolls and 39 avo toasts daily.

Exam question breakdown

VCAA 2016 Exam 1 Matrices Q7

Each week, 300 students at a primary school choose art (A), music (M) or sport (S) as an afternoon activity.

The following transition matrix shows how the students' choices change from week to week.

$$\begin{array}{cc} & \text{this week} \\ & \begin{array}{ccc} \text{A} & \text{M} & \text{S} \end{array} \\ T = & \begin{bmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.4 & 0.4 \\ 0.2 & 0.2 & 0.5 \end{bmatrix} \end{array} \begin{array}{l} \\ \\ \text{A} \\ \text{M} \quad \text{next week} \\ \text{S} \end{array}$$

Based on this 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.

Explanation

Step 1: Calculate the equilibrium state matrix.

As the question specifies 'in the long term', the equilibrium state matrix will be useful for verifying each option. Although the question doesn't specify S_0, it does state that there are 300 students. From this, an initial state matrix can be approximated, as the equilibrium state solution will be the same regardless.

$$S_0 = \begin{bmatrix} 100 \\ 100 \\ 100 \end{bmatrix} \begin{array}{l} \text{A} \\ \text{M} \\ \text{S} \end{array}$$

$$S_{30} = T^{30} \times S_0$$

$$= \begin{bmatrix} 104.76... \\ 109.52... \\ 85.71... \end{bmatrix} \begin{array}{l} \text{A} \\ \text{M} \\ \text{S} \end{array}$$

Step 2: Verify the equilibrium state matrix by comparing with S_{n+1}.

$$S_{31} = T^{31} \times S_0$$

$$= \begin{bmatrix} 104.76... \\ 109.52... \\ 85.71... \end{bmatrix} \begin{array}{l} \text{A} \\ \text{M} \\ \text{S} \end{array}$$

Step 3: Check whether each option is correct or incorrect.

A: This is incorrect because the equilibrium state matrix suggests that approximately 86 students will choose sport in the long term. ✗

B: This is incorrect because the transition matrix suggests that many students change their activity each week. ✗

C: This cannot be concluded from the information provided. While the transition matrix suggests that some students change their activity from week to week, there is no way of keeping track of which students change and which don't. This means there could be some students picking the same activity every week. ✗

D: This is correct, as $109.52 > 85.71$. ✓

E: This is incorrect, as $109.52 \neq 104.76$. ✗

Answer

D

52% of students answered this question correctly.

30% of students incorrectly answered option C, likely on the basis of reasoning rather than mathematics. While it is likely true that every student will change their activity at least once in the long term, this isn't guaranteed from the provided information.

7H Questions

Calculating the equilibrium state matrix

1. For the state matrix equation $S_n = T^n \times S_0$, which of the following expressions would be most likely to correctly calculate the equilibrium state matrix?

A. $T^5 \times S_0$ **B.** $T^{10} \times S_0$ **C.** $T^{20} \times S_0$ **D.** $T^{30} \times S_0$

2. If $T = \begin{bmatrix} 0.23 & 0.54 \\ 0.77 & 0.46 \end{bmatrix}$ and $S_0 = \begin{bmatrix} 98 \\ 42 \end{bmatrix}$:

a. Write down an expression, in terms of T and S_0, that can be used to calculate the equilibrium state matrix.

b. Calculate and verify the equilibrium state matrix. Round values to two decimal places.

3. Calculate the equilibrium state matrix for the following pairs of T and S_0. Round values to two decimal places.

a. $T = \begin{bmatrix} 0.5 & 0.9 \\ 0.5 & 0.1 \end{bmatrix}$, $S_0 = \begin{bmatrix} 50 \\ 70 \end{bmatrix}$

b. $T = \begin{bmatrix} 0.19 & 0.44 \\ 0.81 & 0.56 \end{bmatrix}$, $S_0 = \begin{bmatrix} 401 \\ 225 \end{bmatrix}$

c. $T = \begin{bmatrix} 0.7 & 0.5 & 0.2 \\ 0.2 & 0.5 & 0.4 \\ 0.1 & 0.0 & 0.4 \end{bmatrix}$, $S_0 = \begin{bmatrix} 90 \\ 47 \\ 211 \end{bmatrix}$

4. The transition matrix and initial state matrix used to model the changing preferences of 100 mice in a laboratory test, from hour to hour, are provided.

$$T = \begin{array}{c} \text{current hour} \\ \begin{array}{cc} \text{eat} & \text{stay} \end{array} \\ \begin{bmatrix} 0.6 & 0.9 \\ 0.4 & 0.1 \end{bmatrix} \end{array} \begin{array}{l} \text{eat} \\ \text{stay} \end{array} \text{ next hour} \qquad S_0 = \begin{bmatrix} 84 \\ 16 \end{bmatrix} \begin{array}{l} \text{eat} \\ \text{stay} \end{array}$$

Calculate the equilibrium state matrix, giving values as whole numbers.

Interpreting the equilibrium state matrix

5. Matrix F represents an initial state matrix for the daily selection of different fruits, and matrix E represents the corresponding equilibrium state matrix.

If matrix element f_{mn} represents the number of apples selected on the first day, then matrix element e_{mn} represents

A. the amount of fruit selected on the first day.

B. the daily number of apples selected in the long term.

C. the total number of apples selected in the long term.

D. the daily amount of fruit selected in the long term.

6. The musical instruments tried by 50 primary school students from day to day can be modelled using state matrices. From this, the equilibrium state matrix, M, has been calculated.

$$M = \begin{bmatrix} 9.2 \\ 15.0 \\ 13.9 \\ 7.6 \\ 4.3 \end{bmatrix} \begin{array}{l} \text{keyboard} \\ \text{percussion} \\ \text{guitar} \\ \text{boomwhackers} \\ \text{other} \end{array}$$

In the long term, which type of instrument will students pick most often?

7. The migration of penguins from two local regions, A and B, can be modelled by the following rule:

$S_n = T^n \times S_0$, where

$$S_0 = \begin{array}{c} \begin{bmatrix} 759 \\ 529 \end{bmatrix} \end{array}\begin{array}{l} A \\ B \end{array}, \quad T = \begin{array}{c} \begin{array}{cc} A & B \end{array} \\ \begin{bmatrix} 0.7 & 0.9 \\ 0.3 & 0.1 \end{bmatrix} \end{array}\begin{array}{l} \\ A \\ B \end{array}$$

The equilibrium state matrix for this situation is:

$$\begin{bmatrix} 966 \\ 322 \end{bmatrix}\begin{array}{l} A \\ B \end{array}$$

From this matrix, comment on the long-term projections of penguin migration occurring between the two regions.

Joining it all together

8. A study was conducted on the weekly assignment submissions for 60 university students. The transition matrix and initial state matrix are provided.

$$T = \begin{array}{c} \text{this week} \\ \begin{array}{cc} \text{on time} & \text{late} \end{array} \\ \begin{bmatrix} 0.8 & 0.5 \\ 0.2 & 0.5 \end{bmatrix} \end{array}\begin{array}{l} \\ \\ \text{on time} \\ \text{late} \end{array} \text{ next week} \qquad S_0 = \begin{bmatrix} 53 \\ 7 \end{bmatrix}\begin{array}{l} \text{on time} \\ \text{late} \end{array}$$

a. Write down an expression, in terms of the matrices provided, that can be used to calculate the equilibrium state matrix.

b. Calculate the equilibrium state matrix, rounding all values to whole numbers.

c. In the long term, how many students will hand in their weekly assignments on time each week?

---

9. The chances of a driver reaching a red light at an intersection depend on whether the driver came across a red or green light at the previous intersection. The following transition matrix demonstrates this.

$$T = \begin{array}{c} \text{current light} \\ \begin{array}{cc} \text{red} & \text{green} \end{array} \\ \begin{bmatrix} 0.15 & 0.25 \\ 0.85 & 0.75 \end{bmatrix} \end{array}\begin{array}{l} \\ \\ \text{red} \\ \text{green} \end{array} \text{ next light}$$

38 drivers start a journey with a red light at their first intersection, and 50 drivers start with a green light. In the long term, how many drivers would be expected to meet a red light at the intersection?

---

10. Scientists have discovered that turtles tend to migrate monthly between three islands: Amnio, Belix, and Chel. The number of turtles observed at the three islands in January is represented by matrix S_0, and the subsequent migration patterns are represented by the transition matrix, T.

$$T = \begin{array}{c} \text{this month} \\ \begin{array}{ccc} \text{Amnio} & \text{Belix} & \text{Chel} \end{array} \\ \begin{bmatrix} 0.55 & 0.27 & 0.23 \\ 0.20 & 0.49 & 0.10 \\ 0.25 & 0.24 & 0.67 \end{bmatrix} \end{array}\begin{array}{l} \\ \\ \text{Amnio} \\ \text{Belix} \\ \text{Chel} \end{array} \text{ next month} \qquad S_0 = \begin{bmatrix} 514 \\ 276 \\ 410 \end{bmatrix}\begin{array}{l} \text{Amnio} \\ \text{Belix} \\ \text{Chel} \end{array}$$

The scientists believe that, eventually, the turtle population will stabilise at all three islands. If the population at any one island exceeds 500 turtles in the long term, the scientists must then relocate some of the turtles to the less populated islands.

Will the scientists need to relocate any turtles? If so, from which island(s)?

Back to contents

11. Steve is a weatherman and wants to boost his ratings for 35 towns in regional Victoria. His aim is to improve his predictions for the upcoming autumn season. From previous years, he concluded that the chance that it will rain in a town on a given day will depend on whether it rained the previous day. The transition matrix, T, represents this information.

$$\begin{array}{ccc} & \text{today} & \\ & \begin{array}{cc}\text{rain} & \text{no rain}\end{array} & \\ T = & \begin{bmatrix} 0.65 & 0.87 \\ 0.35 & 0.13 \end{bmatrix} & \begin{array}{l}\text{rain} \\ \text{no rain}\end{array} \quad \text{tomorrow} \end{array}$$

Using the transition matrix, how many of the 35 towns can be expected to rain each day in the long term? Round the answer to the nearest whole number.

Exam practice

12. 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 provided, T, shows how the number of books borrowed per month by the members is expected to change from month to month.

$$\begin{array}{ccc} & \text{this month} & \\ & \begin{array}{ccccc}\text{J} & \text{K} & \text{L} & \text{M} & \text{N}\end{array} & \\ T = & \begin{bmatrix} 0.1 & 0.2 & 0.2 & 0 & 0 \\ 0.5 & 0.2 & 0.3 & 0.1 & 0 \\ 0.3 & 0.3 & 0.4 & 0.1 & 0.2 \\ 0.1 & 0.2 & 0.1 & 0.6 & 0.3 \\ 0 & 0.1 & 0 & 0.2 & 0.5 \end{bmatrix} & \begin{array}{l}\text{J} \\ \text{K} \\ \text{L} \\ \text{M} \\ \text{N}\end{array} \quad \text{next month} \end{array}$$

In the long term, which category is expected to have approximately 96 members each month?

A. J **B.** K **C.** L
D. M **E.** N

VCAA 2018 Exam 1 Matrices Q8

46% of students answered this question correctly.

13. 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.

$$S_0 = \begin{bmatrix} 700 \\ 400 \\ 200 \\ 1400 \end{bmatrix} \begin{array}{l}\text{G} \\ \text{R} \\ \text{S} \\ \text{N}\end{array}$$

The type of maintenance in sections of highway varies from year to year, as shown in the transition matrix, T.

$$\begin{array}{ccc} & \text{this year} & \\ & \begin{array}{cccc}\text{G} & \text{R} & \text{S} & \text{N}\end{array} & \\ T = & \begin{bmatrix} 0.2 & 0.1 & 0.0 & 0.2 \\ 0.1 & 0.1 & 0.0 & 0.2 \\ 0.2 & 0.1 & 0.2 & 0.1 \\ 0.5 & 0.7 & 0.8 & 0.5 \end{bmatrix} & \begin{array}{l}\text{G} \\ \text{R} \\ \text{S} \\ \text{N}\end{array} \quad \text{next year} \end{array}$$

In the long term, what percentage of highway each year is expected to have no maintenance (N)? Round the answer to one decimal place. (1 MARK)

VCAA 2018 Exam 2 Matrices Q3e

39% of students answered this question correctly.

Back to contents

14. 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.

One 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.

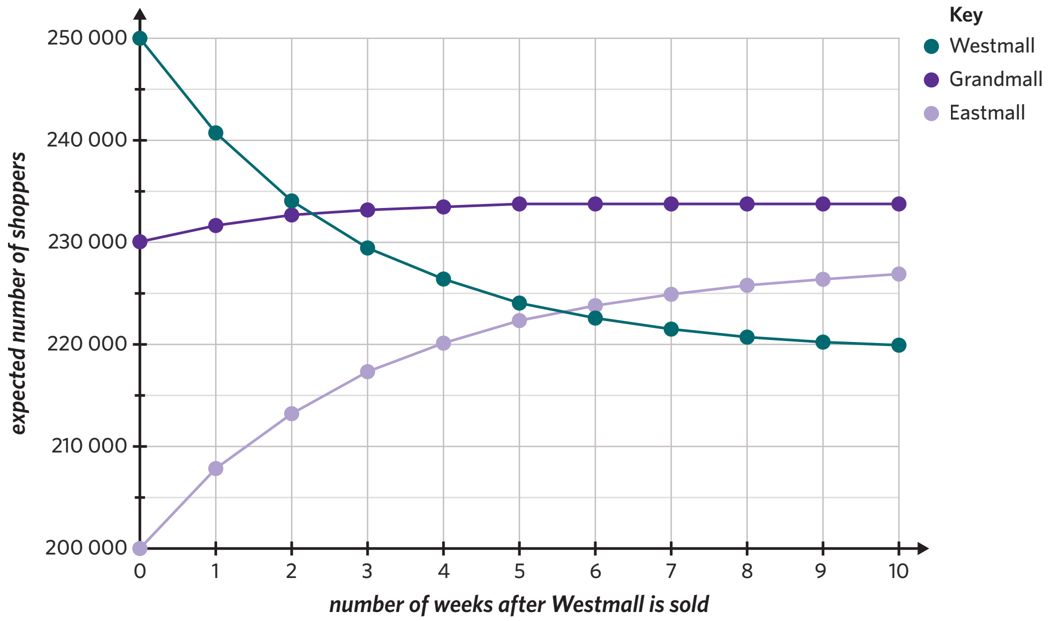
Let S_n be the state matrix that shows the expected number of shoppers at each of the three centres n weeks after Westmall is sold.

A matrix recurrence relation that generates values of S_n is

$S_{n+1} = T \times S_n$, where

$$T = \begin{array}{c} \text{this week} \\ \begin{array}{ccc} W & G & E \end{array} \\ \begin{bmatrix} 0.80 & 0.09 & 0.10 \\ 0.12 & 0.79 & 0.10 \\ 0.08 & 0.12 & 0.80 \end{bmatrix} \end{array} \begin{array}{l} \\ \\ W \\ G \\ E \end{array} \text{next week} \qquad S_0 = \begin{bmatrix} 250\,000 \\ 230\,000 \\ 200\,000 \end{bmatrix} \begin{array}{l} W \\ G \\ E \end{array}$$

Using values from the recurrence relation, the graph provided displays the expected number of shoppers at Westmall, Grandmall and Eastmall for each of the 10 weeks after Westmall is sold.



In the long term, what is the expected weekly number of shoppers at Westmall? Round your answer to the nearest whole number. (1 MARK)

VCAA 2020 Exam 2 Matrices Q3d

39% of students answered this question correctly.

Questions from multiple lessons

Matrices

15. Matrix P is a 3×3 permutation matrix.

Matrix Q is another matrix such that the matrix product $Q \times P$ is defined.

This matrix product results in the entire first and second columns of matrix Q being swapped.
The permutation matrix P is

A. $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ **B.** $\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ **C.** $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ **D.** $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ **E.** $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$

Adapted from VCAA 2018 Exam 1 Matrices Q4

Back to contents

Recursion and financial modelling

16. Nathan is looking to purchase a new Mercedes. He will take out a loan for \$80 000 with interest charged at a rate of 3.9% per annum, compounding fortnightly.

Each fortnight, Nathan pays back the exact amount of interest that is charged for that fortnight.

Let V_n be the value of Nathan's loan, in dollars, after n fortnights.

Which of the following recurrence relations correctly models the value of Nathan's loan?

A. $V_0 = 80\,000, \quad V_{n+1} = 1.0015V_n$

B. $V_0 = 80\,000, \quad V_{n+1} = 1.039V_n - 120$

C. $V_0 = 80\,000, \quad V_{n+1} = 1.0015V_n - 3120$

D. $V_0 = 80\,000, \quad V_{n+1} = 1.039V_n$

E. $V_0 = 80\,000, \quad V_{n+1} = 1.0015V_n - 120$

Adapted from VCAA 2017 Exam 1 Recursion and financial modelling Q20

Matrices

17. The matrix C represents the way in which five friends, Voula (V), Will (W), Xavier (X), Yasmin (Y), and Zoe (Z) interact on Instagram.

The matrix C^2 is also shown.

$$C = \begin{array}{c} \text{followed} \\ \begin{array}{ccccc} V & W & X & Y & Z \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix} \end{array} \begin{array}{l} \\ \\ V \\ W \\ X \\ Y \\ Z \end{array} \text{follower} \qquad C^2 = \begin{array}{c} \text{followed} \\ \begin{array}{ccccc} V & W & X & Y & Z \end{array} \\ \begin{bmatrix} 2 & 1 & 0 & 2 & 1 \\ 0 & 1 & 2 & 0 & 1 \\ 2 & 1 & 3 & 1 & 1 \\ 2 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 1 \end{bmatrix} \end{array} \begin{array}{l} \\ \\ V \\ W \\ X \\ Y \\ Z \end{array} \text{follower}$$

The '1' in row V, column W of matrix C indicates that Voula follows Will on Instagram.

The '0' in row Z, column Y of matrix C indicates that Zoe does not follow Yasmin on Instagram.

a. Who does Will follow? (1 MARK)

b. Yasmin wants to see a photo Voula posted, but cannot do this as she does not follow Voula. She plans to send a message over Instagram to the friend(s) that she follows, who follow Voula themselves, asking for a screenshot of Voula's post. Which friend(s) could she ask? (1 MARK)

Adapted from VCAA 2016 Exam 2 Matrices Q2

Back to contents

7I Applications of transition matrices

STUDY DESIGN DOT POINTS

- 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_0 = initial state matrix, $S_{n+1} = TS_n + B$ to extend modelling to populations that include culling and restocking

7A 7B 7C 7D 7E 7F 7G 7H 7I

KEY SKILLS

During this lesson, you will be:

- constructing and interpreting transition diagrams
- using transition matrices to model situations involving culling and restocking.

KEY TERMS

- Transition diagram
- Culling
- Restocking

Transition matrices can be used to model various practical situations such as growing and changing populations, storage levels, rotating activities, menu options and more. Transition diagrams are often used in applications to visually represent the information contained in transition matrices.

Constructing and interpreting transition diagrams

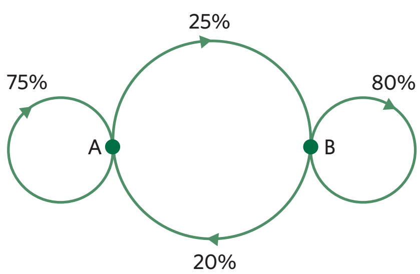
A **transition diagram** is a visual representation of how a transition matrix functions. Each state is represented by a point on the diagram, and the transitions between states are represented by lines with arrows, connecting all of the points together (and connecting each point to itself).

See worked example 1

For example, the transition matrix

$$T = \begin{array}{c} \text{today} \\ \begin{array}{cc} \text{A} & \text{B} \end{array} \\ \begin{bmatrix} 0.75 & 0.2 \\ 0.25 & 0.8 \end{bmatrix} \end{array} \begin{array}{l} \\ \\ \text{A} \\ \text{B} \end{array} \text{tomorrow}$$

can be represented by the following transition diagram.



In a transition diagram, the sum of the percentages moving away from a given point adds up to 100%. This includes the parts of the diagram where a point loops on itself.

A transition diagram can also be used to construct a transition matrix, and vice versa.

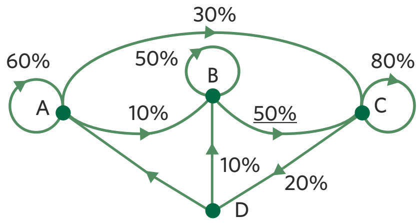
See worked example 2

Back to contents

Worked example 1

Trams operate between four depots, A, B, C, and D. The transition diagram provided represents where the trams end up at the end of each week.

a. What does the underlined value in the diagram represent?



Explanation

The 50% corresponds to the portion of the transition diagram that is flowing from point B to point C.

Answer

50% of trams operating at depot B end up at depot C at the end of each week.

b. Determine the missing percentage value on the transition diagram.

Explanation

Step 1: Identify the location of the missing percentage.

The missing percentage is between points D and A.

Step 2: Calculate the missing percentage.

The sum of the percentages moving away from point D should add up to 100%.

$100\% - 10\% = 90\%$

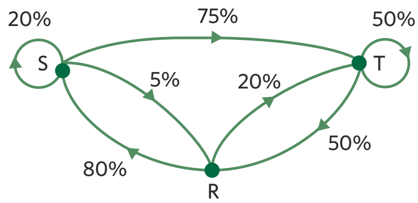
Answer

90%

Worked example 2

Complete the following conversions.

a. Convert the following transition diagram into a transition matrix.



Explanation

Step 1: Set up a square matrix.

As there are three points, the order of the matrix must be 3×3.

Since the points of the transition diagram are S, R and T, these will also be the row and column labels of the transition matrix.

$$\begin{array}{cc} \begin{array}{ccc} R & \quad S & \quad T \end{array} & \\ \left[\begin{array}{ccc} & & \\ & & \\ & & \end{array}\right] & \begin{array}{c} R \\ S \\ T \end{array} \end{array}$$

Step 2: Fill in the first column, R.

Column R will represent the transitions from R to each state.

There is no transition from R to itself (0%).

The transition from R to S is 80%.

The transition from R to T is 20%.

Converted to decimals, to two decimal places, these are 0.00, 0.80 and 0.20 respectively.

$$\begin{array}{cc} \begin{array}{ccc} R & \quad S & \quad T \end{array} & \\ \left[\begin{array}{ccc} 0.00 & & \\ 0.80 & & \\ 0.20 & & \end{array}\right] & \begin{array}{c} R \\ S \\ T \end{array} \end{array}$$

Continues →

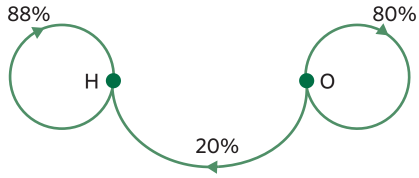
Step 3: Repeat for every other column.

Answer

$$\begin{array}{c c} \begin{array}{ccc} \mathrm{R} & \mathrm{S} & \mathrm{T} \end{array} & \\ \begin{bmatrix} 0.00 & 0.05 & 0.50 \\ 0.80 & 0.20 & 0.00 \\ 0.20 & 0.75 & 0.50 \end{bmatrix} & \begin{array}{c} \mathrm{R} \\ \mathrm{S} \\ \mathrm{T} \end{array} \end{array}$$

b. Use the information in the transition matrix T to complete the corresponding transition diagram.

$$T = \begin{array}{c c c} \text{this month} & & \\ \begin{array}{cc} \mathrm{H} & \mathrm{O} \end{array} & & \\ \begin{bmatrix} 0.88 & 0.20 \\ 0.12 & 0.80 \end{bmatrix} & \begin{array}{c} \mathrm{H} \\ \mathrm{O} \end{array} & \text{next month} \end{array}$$



Explanation

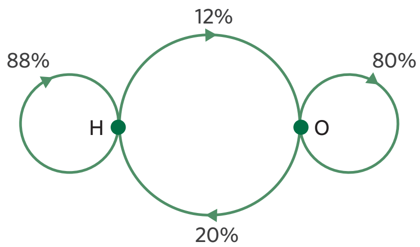
Step 1: Identify which value(s) are missing.

All values except 0.12 (12%) are included in the diagram.

Step 2: Annotate the transition diagram.

Since 0.12 corresponds to the movement from H to O, the line should be drawn between those points, with the arrowhead pointing towards O. Label the line with '12%'.

Answer



Using transition matrices to model situations involving culling and restocking

Situations that can be modelled using the recurrence relation $S_{n+1} = TS_n$ generally assume that all factors are accounted for from one state to the next, represented in the transition matrix, T.

In other situations, this assumption is not valid. In scenarios where the total population is regularly increased or decreased by a set quantity, for a reason that is not already accounted for, a new model is required.

When these types of changes are made to the total population from one state to the next, a more suitable model is

$S_{n+1} = TS_n + B,$

where B is a matrix of the same order as S_n and S_{n+1}.

This model is particularly useful for modelling the population of animals, where humans may choose to add or remove a set number of animals at a regular interval.

Culling is the reduction of an animal population by slaughter. In a matrix recurrence relation scenario, culling represents a subtraction to the population from one state to the next.

Restocking typically refers to replacing stock with a new supply. In a matrix recurrence relation scenario, restocking represents an addition to the population from one state to the next.

For example, the recurrence relation $S_{n+1} = TS_n + B$ can be used to model the number of cows, goats and sheep, respectively, on a farm each month. Matrix $B = \begin{bmatrix} -5 \\ 0 \\ 10 \end{bmatrix}$ indicates that 5 cows are removed (culled), no goats are added (restocked) or removed, and 10 sheep are added each month. This is separate to any information provided in the transition matrix, T.

Back to contents

Worked example 3

Eddortsford High School runs an optional extra-curricular arts program. Each year, students can choose from Music (M), Drama (D), Visual Arts (V), or to opt out (O) of the program. The movement of students from one year to the next is summarised in the transition matrix, T.

$$T = \begin{array}{c} \text{this year} \\ \begin{array}{cccc} \text{M} & \text{D} & \text{V} & \text{O} \end{array} \\ \begin{bmatrix} 0.73 & 0.07 & 0.11 & 0 \\ 0.10 & 0.71 & 0.04 & 0 \\ 0.06 & 0.08 & 0.72 & 0 \\ 0.11 & 0.14 & 0.13 & 1 \end{bmatrix} \end{array} \begin{array}{l} \text{M} \\ \text{D} \\ \text{V} \\ \text{O} \end{array} \text{next year}$$

The matrix S_0 represents the number of students enrolled in each course at the beginning of 2023.

$$S_0 = \begin{bmatrix} 73 \\ 36 \\ 52 \\ 0 \end{bmatrix} \begin{array}{l} \text{M} \\ \text{D} \\ \text{V} \\ \text{O} \end{array}$$

As the program expands, 30 new students are added to the program each year. 13 of these students are expected to be added to the music course, 6 to the drama course and 11 to the visual arts course.

a. Construct matrix B to represent the predicted new enrolments and their preferences.

Explanation

Step 1: Set up a blank matrix.

Since matrix B is the same order as S_0, its order will be 4×1.

$$B = \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix} \begin{array}{l} \text{M} \\ \text{D} \\ \text{V} \\ \text{O} \end{array}$$

Step 2: Fill in the missing values.

The missing values are given by the number of new enrolments for each course.

Answer

$$B = \begin{bmatrix} 13 \\ 6 \\ 11 \\ 0 \end{bmatrix} \begin{array}{l} \text{M} \\ \text{D} \\ \text{V} \\ \text{O} \end{array}$$

b. How many students are expected to be enrolled in the music course at the beginning of 2025? Round to the nearest whole number.

Explanation

Step 1: Calculate S_1, the state matrix at the start of 2024.

$S_1 = TS_0 + B$

$$= \begin{bmatrix} 0.73 & 0.07 & 0.11 & 0 \\ 0.10 & 0.71 & 0.04 & 0 \\ 0.06 & 0.08 & 0.72 & 0 \\ 0.11 & 0.14 & 0.13 & 1 \end{bmatrix} \begin{bmatrix} 73 \\ 36 \\ 52 \\ 0 \end{bmatrix} + \begin{bmatrix} 13 \\ 6 \\ 11 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 74.53 \\ 40.94 \\ 55.70 \\ 19.83 \end{bmatrix} \begin{array}{l} \text{M} \\ \text{D} \\ \text{V} \\ \text{O} \end{array}$$

Step 2: Calculate S_2, the state matrix at the start of 2025.

$S_2 = TS_1 + B$

$$= \begin{bmatrix} 0.73 & 0.07 & 0.11 & 0 \\ 0.10 & 0.71 & 0.04 & 0 \\ 0.06 & 0.08 & 0.72 & 0 \\ 0.11 & 0.14 & 0.13 & 1 \end{bmatrix} \begin{bmatrix} 74.53 \\ 40.94 \\ 55.70 \\ 19.83 \end{bmatrix} + \begin{bmatrix} 13 \\ 6 \\ 11 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 76.3997 \\ 44.7484 \\ 58.8510 \\ 41.0009 \end{bmatrix} \begin{array}{l} \text{M} \\ \text{D} \\ \text{V} \\ \text{O} \end{array}$$

Step 3: Identify the element that corresponds with music in matrix S_2.

The value for music in S_2 is 76.3997.

Answer

76 students

Back to contents

Exam question breakdown

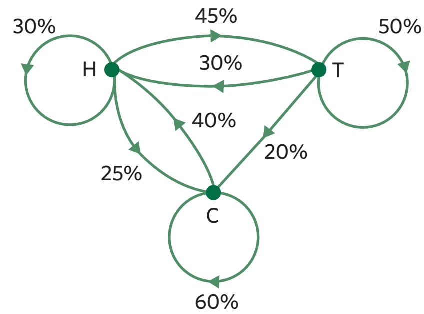
VCAA 2016 Exam 1 Matrices Q6

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 following 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.
$$\begin{array}{cc} \begin{array}{ccc} & \text{from} & \\ H & T & C \end{array} & \\ \begin{bmatrix} 0.30 & 0.75 & 0.65 \\ 0.75 & 0.50 & 0.20 \\ 0.65 & 0.20 & 0.60 \end{bmatrix} & \begin{array}{l} H \\ T \ \text{to} \\ C \end{array} \end{array}$$

B.
$$\begin{array}{cc} \begin{array}{ccc} & \text{from} & \\ H & T & C \end{array} & \\ \begin{bmatrix} 0.30 & 0.30 & 0.40 \\ 0.45 & 0.50 & 0 \\ 0.25 & 0.20 & 0.60 \end{bmatrix} & \begin{array}{l} H \\ T \ \text{to} \\ C \end{array} \end{array}$$

C.
$$\begin{array}{cc} \begin{array}{ccc} & \text{from} & \\ H & T & C \end{array} & \\ \begin{bmatrix} 0.30 & 0.30 & 0.40 \\ 0.45 & 0.50 & 0.20 \\ 0.25 & 0.20 & 0.60 \end{bmatrix} & \begin{array}{l} H \\ T \ \text{to} \\ C \end{array} \end{array}$$

D.
$$\begin{array}{cc} \begin{array}{ccc} & \text{from} & \\ H & T & C \end{array} & \\ \begin{bmatrix} 0.30 & 0.30 & 0.40 \\ 0.45 & 0.50 & 0.20 \\ 0.25 & 0.20 & 0.40 \end{bmatrix} & \begin{array}{l} H \\ T \ \text{to} \\ C \end{array} \end{array}$$

E.
$$\begin{array}{cc} \begin{array}{ccc} & \text{from} & \\ H & T & C \end{array} & \\ \begin{bmatrix} 0.30 & 0.45 & 0.25 \\ 0.30 & 0.50 & 0.20 \\ 0.40 & 0 & 0.60 \end{bmatrix} & \begin{array}{l} H \\ T \ \text{to} \\ C \end{array} \end{array}$$

Explanation

Step 1: Set up a square matrix.

As there are three points, the matrix size needs to be 3×3.

$$\begin{array}{cc} \begin{array}{ccc} & \text{from} & \\ H & T & C \end{array} & \\ \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix} & \begin{array}{l} H \\ T \ \text{to} \\ C \end{array} \end{array}$$

Step 2: Fill in the first column, H.

Column H will represent the transitions from H to each state.

The transition from H to itself is 30%.

The transition from H to T is 45%.

The transition from H to C is 25%.

As decimals, these are 0.30, 0.45 and 0.25 respectively.

$$\begin{array}{cc} \begin{array}{ccc} & \text{from} & \\ H & T & C \end{array} & \\ \begin{bmatrix} 0.30 & & \\ 0.45 & & \\ 0.25 & & \end{bmatrix} & \begin{array}{l} H \\ T \ \text{to} \\ C \end{array} \end{array}$$

Step 3: Repeat for every other column.

$$\begin{array}{cc} \begin{array}{ccc} & \text{from} & \\ H & T & C \end{array} & \\ \begin{bmatrix} 0.30 & 0.30 & 0.40 \\ 0.45 & 0.50 & 0 \\ 0.25 & 0.20 & 0.60 \end{bmatrix} & \begin{array}{l} H \\ T \ \text{to} \\ C \end{array} \end{array}$$

Answer

B

86% of students answered this question correctly.

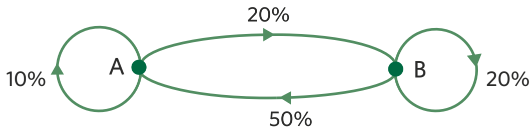
For this question, the key is to understand how the missing connection between T and C on the transition diagram translates to a transition matrix. B is the only answer which displays this feature with a correctly placed 0.

7I Questions

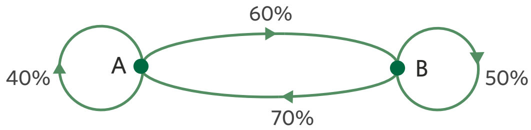
Constructing and interpreting transition diagrams

1. Which of the following is a valid transition diagram?

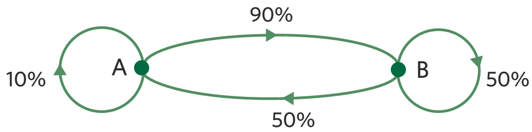
A.



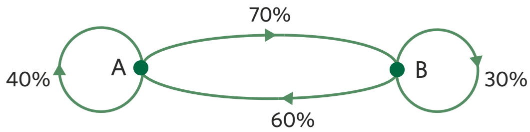
B.



C.

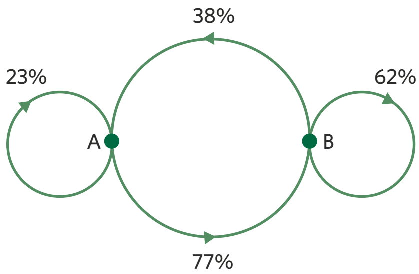


D.

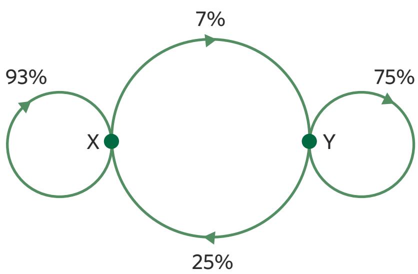


2. Construct a transition matrix from each of the following transition diagrams. Convert all percentages to decimals.

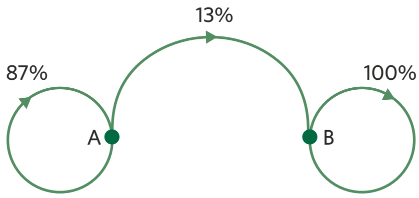
a.



b.

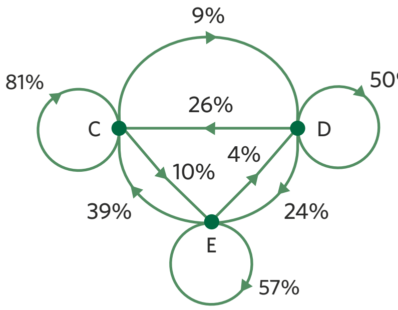


c.

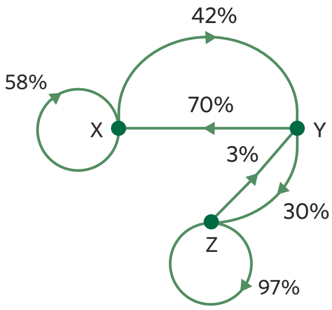


3. Construct a transition matrix from each of the following transition diagrams. Convert all percentages to decimals.

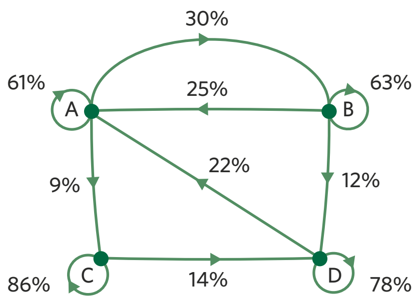
a.



b.



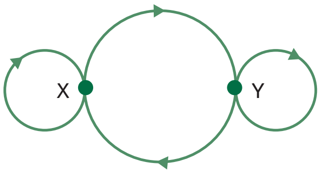
c.



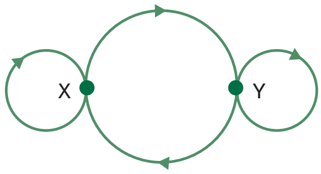
Back to contents

4. Construct a transition diagram from each of the following transition matrices. Convert all decimals to percentages.

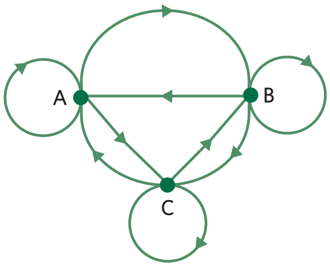
a.
$$\begin{array}{ccc} & \text{from} & \\ & \begin{matrix} X & \;\; Y \end{matrix} & \\ & \begin{bmatrix} 0.7 & 0.55 \\ 0.3 & 0.45 \end{bmatrix} & \begin{matrix} X \\ Y \end{matrix}\ \text{to} \end{array}$$



b.
$$\begin{array}{ccc} & \text{from} & \\ & \begin{matrix} X & \;\; Y \end{matrix} & \\ & \begin{bmatrix} 0.35 & 0.21 \\ 0.65 & 0.79 \end{bmatrix} & \begin{matrix} X \\ Y \end{matrix}\ \text{to} \end{array}$$

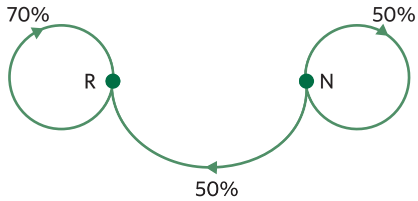


c.
$$\begin{array}{ccc} & \text{from} & \\ & \begin{matrix} A & \;\; B & \;\; C \end{matrix} & \\ & \begin{bmatrix} 0.33 & 0.4 & 0.05 \\ 0.34 & 0.42 & 0.67 \\ 0.33 & 0.18 & 0.28 \end{bmatrix} & \begin{matrix} A \\ B \\ C \end{matrix}\ \text{to} \end{array}$$



5. An Indian food appreciation group likes to order takeaway each week. Members can either order roti (R) or naan bread (N). It is discovered that the transition matrix, T, can be used to predict each member's choice from week to week.

$$\begin{array}{ccc} & \text{this week} & \\ & \begin{matrix} R & \;\; N \end{matrix} & \\ & \begin{bmatrix} 0.7 & 0.5 \\ 0.3 & 0.5 \end{bmatrix} & \begin{matrix} R \\ N \end{matrix}\ \text{next week} \end{array}$$



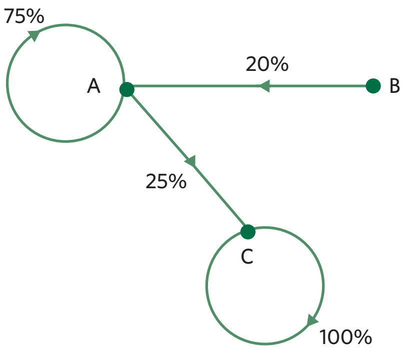
a. Use the information in the transition matrix, T, to complete the transition diagram.

b. What percentage of members are expected to change their order from naan bread to roti each week?

c. If there are 120 members in this club and 80 decide to order roti this week, how many members are expected to order roti next week?

6. Aaron, Brett and Charlie are brothers who like to collect rocks. At the end of each week, Aaron and Brett always give the worst 25% of their rocks to Charlie. Charlie keeps all his rocks. Brett is very picky about his collection and also gives 20% of his rocks to Aaron each week.

a. Using this information, complete the transition diagram shown, which describes the movement of rocks week-to-week between the brothers.



b. Use the answer from part **a** to construct a transition matrix.

c. At the beginning of this week, Aaron has 40 rocks. How many of these rocks will he still have next week?

d. If Aaron and Brett each currently have 24 rocks, how many rocks will be given to Charlie at the end of the week?

e. Charlie received 90 rocks this week. If each of the brothers had an equal number of rocks last week, how many rocks did the brothers have in total?

Back to contents

Using transition matrices to model situations involving culling and restocking

7. Consider the matrix recurrence relation $S_{n+1} = TS_n + B$

If T is a 3×3 matrix, then

A. S_n and B will be 1×1 matrices.

B. S_n and B will be 1×3 matrices.

C. S_n and B will be 3×1 matrices.

D. S_n and B will be 3×3 matrices.

---

8. Consider the following matrix recurrence relation.

$$S_0 = \begin{bmatrix} 38 \\ 64 \end{bmatrix} \quad S_{n+1} = \begin{bmatrix} 0.7 & 0.46 \\ 0.3 & 0.54 \end{bmatrix} S_n + B$$

For each value of B provided, determine S_3. Round to 3 significant figures.

a. $B = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$ **b.** $B = \begin{bmatrix} -13 \\ 5 \end{bmatrix}$ **c.** $B = \begin{bmatrix} 17 \\ -12 \end{bmatrix}$

---

9. Consider the following recurrence relation.

$$S_0 = \begin{bmatrix} 200 \\ 300 \end{bmatrix} \quad S_{n+1} = \begin{bmatrix} 0.6 & 0.5 \\ 0.4 & 0.5 \end{bmatrix} S_n + B$$

a. Which of the following must be true in order to keep the sum of the elements in each state matrix, S_n, constant?

A. Matrix B can only be the zero matrix.

B. The sum of elements in matrix B must equal 0.

C. The sum of elements in matrix B must be equal to the sum of elements in the initial state matrix S_0.

D. Matrix B must be equal to the initial state matrix S_0.

b. Determine matrix B for which $S_n = S_{n+1}$.

---

10. Two cosmetics companies, Edphora and L'Órolo, are in direct competition. Each month, 7% of those who bought cosmetics from Edphora last month are expected to buy from L'Órolo this month, and 12% of those who bought cosmetics from L'Órolo last month are expected to buy from Edphora this month.

a. Complete the following transition matrix using the provided information.

$$\begin{array}{c} \text{last month} \\ \begin{array}{cc} \text{Edphora} & \text{L' Órolo} \end{array} \end{array}$$

$$T = \begin{bmatrix} \quad & \quad \\ \quad & \quad \end{bmatrix} \begin{array}{l} \text{Edphora} \\ \text{L' Órolo} \end{array} \quad \text{this month}$$

L'Órolo suddenly realise that they are losing their customers to Edphora and run a targeted advertising campaign. They project that the campaign will result in 32 000 current Edphora customers swapping to L'Órolo next month.

b. Create a matrix which represents this change. It should be of an appropriate order so that it can be used in a matrix equation with the transition matrix, T.

c. Last month, 49 000 people bought cosmetics from Edphora and 56 000 people bought cosmetics from L'Órolo. Including changes due to the successful advertising campaign, how many people will buy from L'Órolo this month?

11. John, Leila, and Lo all have way too much money and decide to start their own business. They decide to buy a fleet of 20 Teslolo self-driving cars and start an automated taxi service. Customers around Greater Melbourne will be able to book trips via an app and are then driven by a Teslolo car to their destinations.

John, Leila, and Lo also establish 4 depots at which the cars can refuel. They are located in Doncaster (D), Frankston (F), Reservoir (R), and Werribee (W).

At 3:30 am every morning, the Teslolo cars will drive themselves to the nearest depot to refuel.

The transition matrix, T, describes the proportion of cars at each depot based on the distribution of cars the previous morning.

$$T = \begin{array}{c} \text{this morning} \\ \begin{array}{cccc} \text{D} & \text{F} & \text{R} & \text{W} \end{array} \\ \begin{bmatrix} 0.23 & 0.42 & 0.18 & 0.10 \\ 0.36 & 0.25 & 0.10 & 0.05 \\ 0.34 & 0.28 & 0.50 & 0.23 \\ 0.07 & 0.05 & 0.22 & 0.62 \end{bmatrix} \end{array} \begin{array}{l} \text{D} \\ \text{F} \\ \text{R} \\ \text{W} \end{array} \text{ tomorrow morning}$$

a. Given that, initially, all the cars are evenly distributed across the four depots, find the state matrix that describes the distribution of the Teslolo cars in 2 mornings' time. Round to the nearest whole number.

After the two days, John, Leila, and Lo decide that there are too many Teslolo cars at the Reservoir depot. Each morning they redirect one car from the Reservoir depot to the Frankston depot.

b. Write down a matrix representing this redistribution.

c. Find the state matrix that describes the distribution of the Teslolo cars two days later. Round to the nearest whole number. Use the answer from part **a** for the initial state matrix.

After these 2 days, in a sneaky way to make some more cash, John decides to start selling one car a day for 4 days from the reservoir depot, hoping that his colleagues won't notice.

d. Find the state matrix that describes the distribution of the Teslolo cars 4 days later. Round to the nearest whole number. Use the answer from part **c** for the initial state matrix.

Joining it all together

12. On a crowded bus, people will either be standing or sitting on a seat. On one bus journey, the number of sitting and standing passengers after n stops can be modelled using the following recurrence relation.

$$B_0 = \begin{bmatrix} 3 \\ 15 \end{bmatrix} \begin{array}{l} \text{stand} \\ \text{sit} \end{array} \qquad B_{n+1} = \begin{bmatrix} 0.85 & 0.05 \\ 0.15 & 0.95 \end{bmatrix} B_n + \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

a. What does the matrix $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ represent?

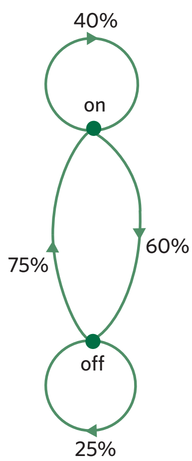
b. Calculate B_3, the matrix displaying the standing and sitting passengers on the bus after 3 stops. Round values to the nearest whole number.

After a while, the bus reaches its maximum capacity, and cannot take any more passengers.

c. Represent the transitioning of sitting and standing passengers, after the bus stops taking passengers, in a transition diagram.

---

13. In basketball, each team always has 5 players on the court, and other players off court on standby. Each time a break is called, some players are taken off the court, and some players on standby are put onto the court to play. The transition diagram provided demonstrates the transitioning pattern for a particular team within a game.



a. Each break, 3 players are taken off the court, and 3 players are put back on the court. How many players are on this team in total?

b. Convert the transition diagram into a transition matrix, T, and construct matrix P_0, which shows the initial number of players on and off court.

c. Let P_n be the state matrix representing the players on and off court after n breaks. Construct a matrix recurrence relation in terms of P_n, P_{n+1}, P_0 and T.

d. How does P_2 and P_5 compare to P_0?

Back to contents

During a basketball game, the audience members are either in the stands (S), standing near the court (C), or in the lobby (L). Additionally, some people may arrive or leave during each break. This information is summarised by the following recurrence relation.

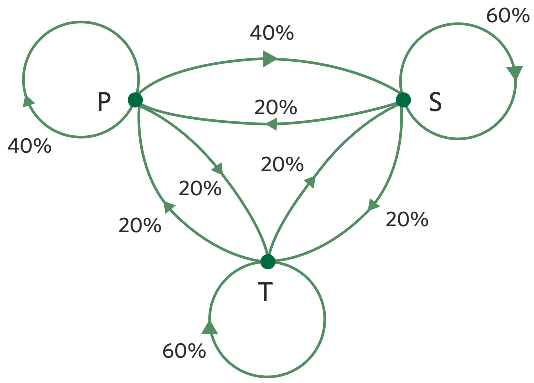
$$A_0 = \begin{array}{c} \begin{bmatrix} 56 \\ 24 \\ 21 \end{bmatrix} \end{array}\begin{array}{l} S \\ C \\ L \end{array} \qquad A_{n+1} = \begin{bmatrix} 0.85 & 0.30 & 0.25 \\ 0.05 & 0.50 & 0.25 \\ 0.10 & 0.20 & 0.5 \end{bmatrix} A_n + \begin{bmatrix} -2 \\ -3 \\ 2 \end{bmatrix}$$

e. Use the recurrence relation to determine the expected number of people in the stands after 1 break, to the nearest whole number.

f. Use the recurrence relation to determine the expected number of people in the lobby after 2 breaks, to the nearest whole number.

Exam practice

14. Junior students at this 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 provided 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)

b. Matrix J_1 lists the number of junior students who will be in each elective activity in Term 1.

$$J_1 = \begin{bmatrix} 300 \\ 240 \\ 210 \end{bmatrix}\begin{array}{l} P \\ S \\ T \end{array}$$

306 junior students are expected to choose sport (S) in Term 2.

Complete the following calculation to show this. (1 MARK)

$300 \times$ ☐ $+ 240 \times$ ☐ $+ 210 \times$ ☐ $= 306$

VCAA 2017 Exam 2 Matrices Q2a,b

Part **a**: **95%** of students answered this question correctly.

Part **b**: **69%** of students answered this question correctly.

---

15. 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.

$$F_0 = \begin{bmatrix} 50\,000 \\ 10\,000 \\ 7\,000 \\ 0 \end{bmatrix}\begin{array}{l} Y \\ J \\ A \\ D \end{array}$$

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 relation

$$F_{n+1} = \begin{bmatrix} 0.65 & 0 & 0 & 0 \\ 0.25 & 0.75 & 0 & 0 \\ 0 & 0.20 & 0.95 & 0 \\ 0.10 & 0.05 & 0.05 & 1 \end{bmatrix} F_n + B$$

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.

Back to contents

Each month, the fish farm will

A. sell 1650 adult fish.

B. buy 1750 adult fish.

C. sell 17 500 young fish.

D. buy 50 000 young fish.

E. buy 10 000 juvenile fish.

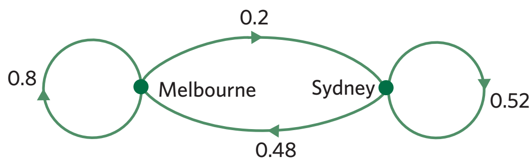
VCAA 2017 Exam 1 Matrices Q7

36% of students answered this question correctly.

---

16. An airline parks all of its planes at Sydney airport or Melbourne airport overnight.

The transition diagram provided shows the change in the location of the planes from night to night.



There are always m planes parked at Melbourne airport.

There are always s planes parked at Sydney airport.

Of the planes parked at Melbourne airport on Tuesday night, 12 had been parked at Sydney airport on Monday night.

How many planes does the airline have?

A. 25 **B.** 37 **C.** 62

D. 65 **E.** 85

VCAA 2019 Exam 1 Matrices Q8

15% of students answered this question correctly.

Questions from multiple lessons

Data analysis

17. The heights of a population of high school students are approximately normally distributed with a mean of 172 cm and a standard deviation of 6 cm.

From a random sample of 400 students, how many students are between 166 cm and 184 cm tall?

A. 68 **B.** 82 **C.** 272 **D.** 326 **E.** 380

Adapted from VCAA 2018 Exam 1 Data analysis Q5

Matrices

18. A local gym offers five different classes: Boxing (B), Pilates (P), Spin (S), Yoga (Y), and Zumba (Z). The gym has 150 members who attend one of the five classes each week.

A transition matrix, T, shows how the class attended by members is expected to change from week to week.

$$\begin{array}{ccc} & \text{this week} & \\ & \begin{array}{ccccc} B & P & S & Y & Z \end{array} & \\ T = & \begin{bmatrix} 0.3 & 0 & 0.2 & 0.1 & 0.4 \\ 0.1 & 0.2 & 0.3 & 0.5 & 0 \\ 0.1 & 0.2 & 0 & 0.1 & 0.2 \\ 0.1 & 0.5 & 0.3 & 0.2 & 0.3 \\ 0.4 & 0.1 & 0.2 & 0.1 & 0.1 \end{bmatrix} & \begin{array}{l} B \\ P \\ S \quad \text{next week} \\ Y \\ Z \end{array} \end{array}$$

Which gym class is expected to have approximately 43 attendees each week in the long run?

A. Boxing **B.** Pilates **C.** Spin **D.** Yoga **E.** Zumba

Adapted from VCAA 2018 Exam 1 Matrices Q8

Back to contents

Matrices

19. The population, in millions, of three East Asian countries, in 2019, is shown in matrix P_{2019}.

$$P_{2019} = \begin{array}{c} \begin{bmatrix} 51 \\ 1419 \\ 127 \end{bmatrix} \end{array} \begin{array}{l} \text{South Korea} \\ \text{China} \\ \text{Japan} \end{array}$$

The expected percentage annual growth of each of the populations is shown in the following table.

Country	**South Korea**	**China**	**Japan**
Annual Change	0.30% increase	0.35% increase	0.25% decrease

a. Find matrix P_{2020}, which shows the expected population, in millions, to two decimal places, of each country in 2020. (1 MARK)

b. The expected population of each of the countries in 2020 can be determined by the matrix product

$P_{2020} = G \times P_{2019}$

where G is a diagonal matrix.

Find matrix G. (1 MARK)

Adapted from VCAA 2018 Exam 2 Matrices Q2

Back to contents

AOS 2
CHAPTER 8
Networks and decision mathematics

Back to contents

8

KEY KNOWLEDGE

- 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
- 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 spanning trees
- minimum spanning trees in a weighted connected graph and their determination by inspection or by Prim's algorithm
- use of minimal spanning trees to solve minimal connector 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
- 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
- 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
- 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.

Image: ChristianChan/Shutterstock.com

Back to contents

8A Introduction to graphs and networks

STUDY DESIGN DOT POINT

- the concepts, conventions and terminology of graphs including planar graphs and Euler's rule, and directed (digraphs) and networks

8A 8B 8C 8D 8E 8F 8G 8H 8I 8J

KEY SKILLS

During this lesson, you will be:

- identifying properties of a graph
- identifying and constructing graphs
- using Euler's rule.

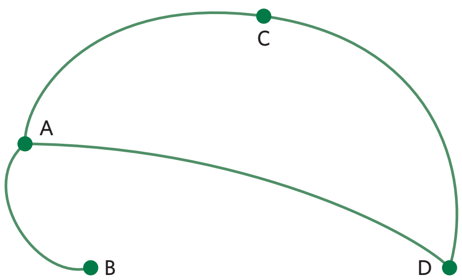
KEY TERMS

- Graph
- Vertex
- Edge
- Degree
- Loop
- Face
- Bridge
- Isolated vertex
- Isomorphic graphs
- Planar graph
- Connected graph
- Simple graph
- Complete graph
- Degenerate graph
- Euler's rule

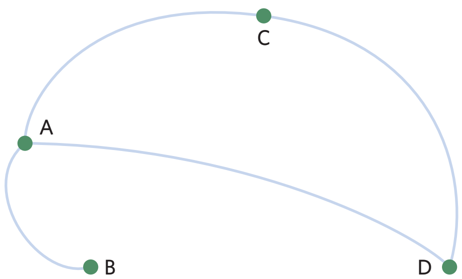
There are many occurences in everyday life that involve connections. In such situations, it can be useful to visually display these connections. This allows them to be analysed mathematically in order to solve problems.

Identifying properties of a graph

A **graph**, also known as a network, is a diagram that is used to show the connections between a group of common elements, such as objects, locations, people, or activities.

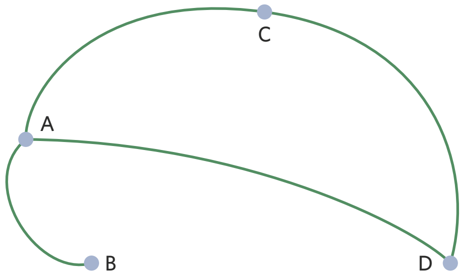


A **vertex** is a point on a graph. Points A, B, C and D are all vertices.

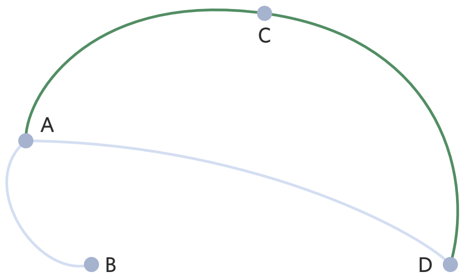


Back to contents

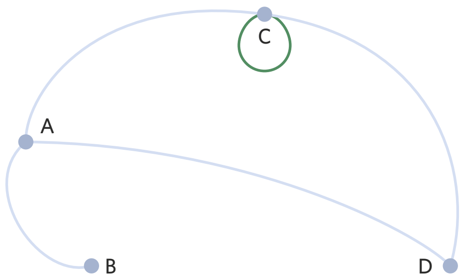
An **edge** is a line connecting one vertex to either another vertex or itself. They represent connections between vertices. Edges can overlap each other.



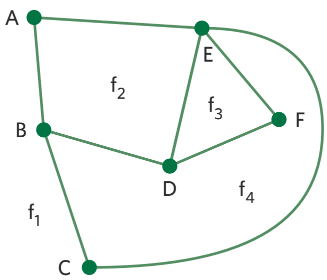
A **degree** of a vertex is the number of times an edge attaches to it. The degree of vertex C is two as there are two edges attached to it.



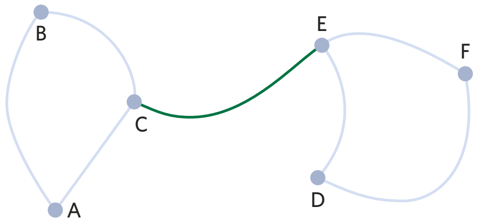
A **loop** is an edge that connects a vertex back to itself without passing through any other vertices. A loop adds two degrees to a vertex as there are two connections, so the degree of vertex C is four.



A **face** is a separate area on a graph that is bordered by edges. For the number of faces to be counted directly from a graph, there must be no overlapping edges. The space outside the graph is counted as an additional face. If there are overlapping edges, the graph must be redrawn so that the number of faces can be counted. The following graph has four faces.

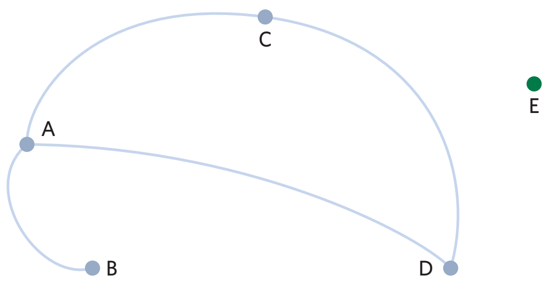


In some graphs, all vertices are connected to each other, either directly or indirectly. Every vertex is able to be reached by another vertex. A **bridge** is an edge that, when removed, will prevent this. In the following graph, the edge connecting vertices C and E is a bridge. If it was removed, vertices A, B and C would no longer be able to reach vertices D, E and F and vice versa.



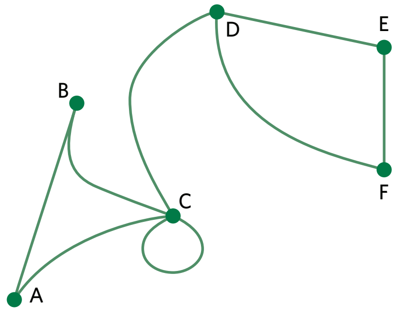
Back to contents

An **isolated vertex** is not connected by an edge to any other vertices on the graph. Vertex E is isolated.



Worked example 1

Consider the graph.



a. How many edges are in this graph?

Explanation

There are eight edges in the graph.

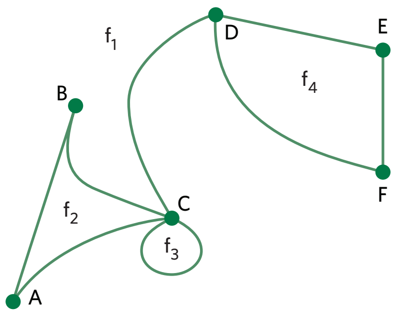
Answer

8

b. How many faces are in this graph?

Explanation

This graph has no overlapping edges so the number of faces can be counted directly from the graph. The space outside the graph is counted as an additional face.



Answer

4

Continues →

Back to contents

c. How many vertices are in this graph?

Explanation

There are six points that are connected by edges.

Answer

6

---

d. Which vertex has a loop?

Explanation

A loop can be found at vertex C.

Answer

C

---

e. Determine the degree of vertex C.

Explanation

There are four edges connected to vertex C, however a loop counts as two degrees as it connects to the vertex twice.

Answer

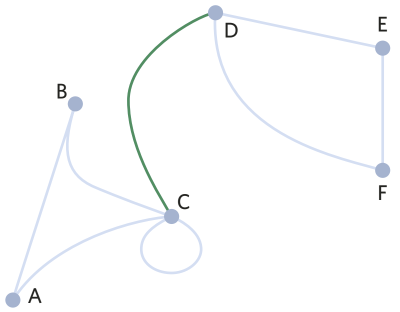
5

---

f. Between which vertices can a bridge be found?

Explanation

If the edge between vertices C and D were removed, vertices A, B and C would be unable to reach vertices D, E and F and vice versa.



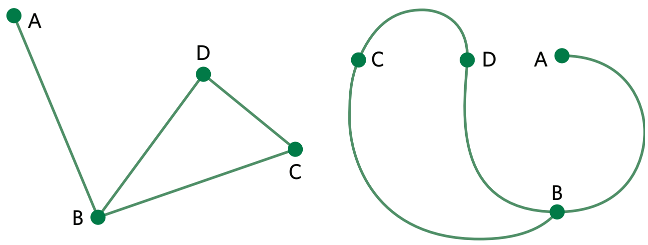
Answer

C and D

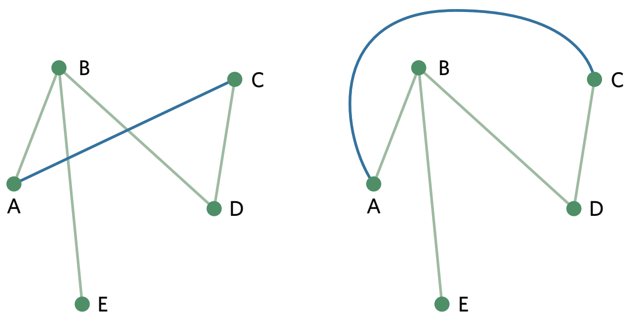
Back to contents

Identifying and constructing graphs

Isomorphic graphs, or equivalent graphs, are graphs that display the same information as each other. They have the same vertices and connections, but they are drawn differently. The following graphs are isomorphic.

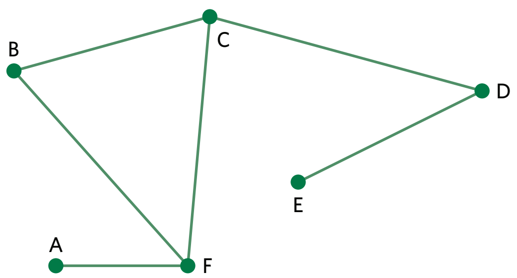


A **planar graph** is a graph that can be drawn with no overlapping edges. If a graph cannot be redrawn this way, it is considered non-planar. The following is an example of a graph being redrawn in planar form with no overlapping edges.

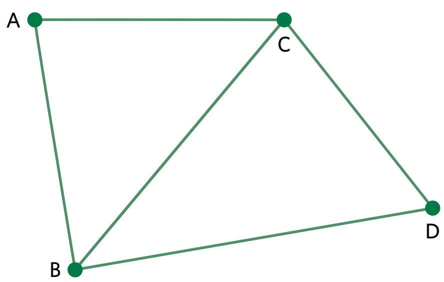


In addition to redrawing edges, vertices can also be moved around to make it easier to redraw in planar form. However, a non-planar graph will always remain non-planar, regardless of how the vertices are rearranged.

A **connected graph** is a graph in which every vertex is connected to each other, either directly or indirectly. Every vertex is able to be reached by any other vertex by travelling along the edges.

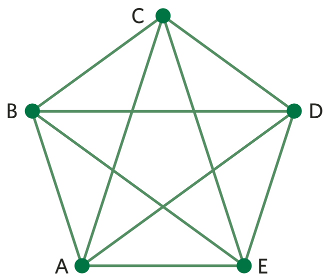


A **simple graph** does not contain any loops or duplicate edges.

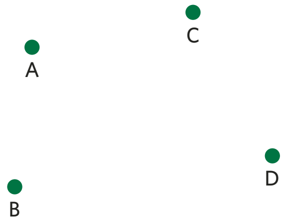


Back to contents

A **complete graph** directly connects every vertex to every other vertex.

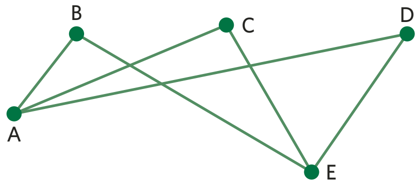


A **degenerate graph** is a graph with no edges. All of the vertices are isolated.



Worked example 2

Consider the following graph.



a. Redraw the graph in planar form.

Explanation

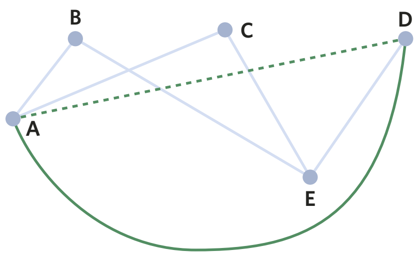
Step 1: Identify instances of overlapping edges.

The A-C edge is overlapping with the B-E edge.

The A-D edge is overlapping with the B-E edge and the C-E edge.

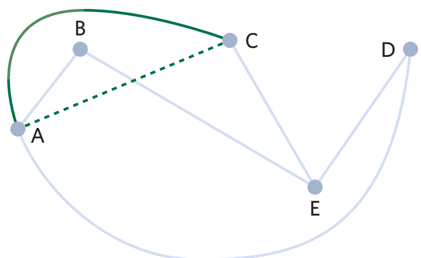
Step 2: Remove an overlapping edge and redraw it.

The A-D edge can be redrawn on the outside so it is not crossing any other edges.



Step 3: Repeat for the remaining overlapping edges.

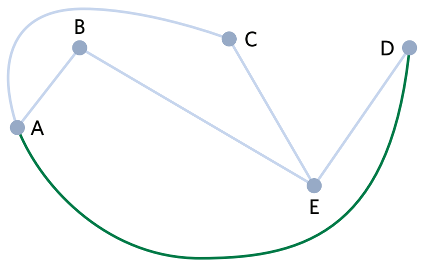
The A-C edge can be redrawn on the outside so it is not crossing any other edges.



Continues →

Back to contents

Answer



Note: This is just one example of the graph in planar form. There are other possible solutions.

b. Is this a connected graph?

Explanation

This graph is considered connected as vertices A, B, C, D and E can all reach each other using the edges.

Answer

Yes

c. Is this a complete graph?

Explanation

In this graph, there are multiple instances of vertices not being directly connected to all other vertices. For example, vertex A is not directly connected to vertices C or E.

Answer

No

Using Euler's rule

Euler's rule describes the relationship between the number of vertices, edges and faces for all connected planar graphs. This relationship can be written as

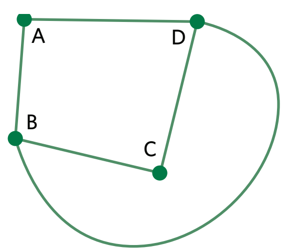
$v - e + f = 2$, where

- v is the number of vertices
- e is the number of edges
- f is the number of faces

Worked example 3

Use Euler's rule to solve the following questions.

a. Show that the graph is a connected planar graph.



Continues →

Explanation

Step 1: Identify the number of vertices, edges and faces.

There are four vertices, so $v = 4$.

There are five edges, so $e = 5$.

There are three faces, so $f = 3$.

Step 2: Apply these values to Euler's rule, ensuring to show calculations in the final answer.

Answer

$$v - e + f = 4 - 5 + 3$$
$$= 2$$

Therefore, the graph is planar.

b. If a connected planar graph has five vertices and five faces, how many edges does the graph have?

Explanation

Step 1: Identify the known values.

There are five vertices, so $v = 5$.

There are five faces, so $f = 5$.

Step 2: Substitute into Euler's rule.

$$v - e + f = 2$$
$$5 - e + 5 = 2$$
$$10 - e = 2$$
$$e = 8$$

Answer

8 edges

Exam question breakdown

VCAA 2019 Exam 1 Networks and decision mathematics Q4

Two graphs, labelled Graph 1 and Graph 2, are shown.

Which of the following statements is **not** true?

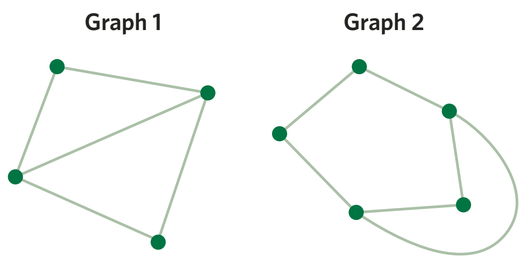
A. Graph 1 and Graph 2 are isomorphic.

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.



Explanation

To solve this question, determine whether each is true or not.

A: This is not true. Graph 1 and Graph 2 are not isomorphic, as they have a different number of vertices. Graph 1 has four vertices and five edges whereas Graph 2 has five vertices and six edges. ✓

B: This is true. Graph 1 has five edges and Graph 2 has six edges. ✗

C: This is true. All vertices in both graphs are able to be reached by all other vertices. ✗

D: This is true. Both graphs contain three faces. ✗

E: This is true. All vertices in a graph must be directly connected to every other vertex in a graph for it to be considered complete. ✗

Answer

A

71% of students answered this question correctly.

18% of students incorrectly chose option E. It is likely that students were confused between the definitions of connected and complete graphs, leading them to answer incorrectly.

8A Questions

Identifying properties of a graph

1. The following is a connected graph with four vertices.

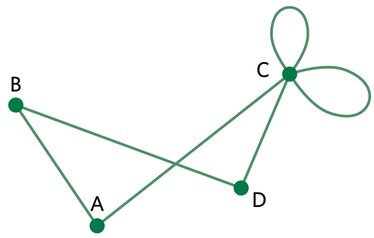
What is the degree of vertex C?

A. 3

B. 4

C. 5

D. 6

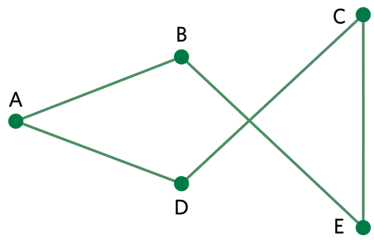


2. Consider the following graph.

a. How many vertices does it contain?

b. All vertices in this graph have the same degree.
What is the degree of each vertex?

c. How many faces are there in this graph?



3. Consider the following graph.

a. How many vertices does it contain?

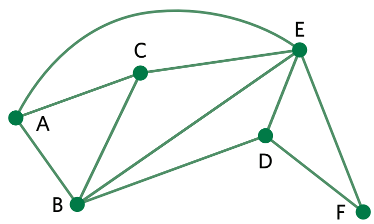
b. What is the degree of vertex C?

c. What is the degree of vertex D?

d. What is the degree of vertex E?

e. How many faces are there in this graph?

f. Is there a bridge in this graph? If so, which vertices does it connect to?



4. The following graph shows roads connecting countries in Europe.

a. How many vertices are present?

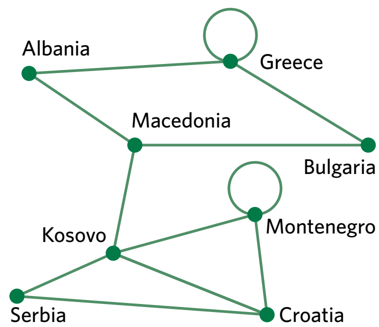
b. What is the degree of Bulgaria?

c. What is the degree of Greece?

d. The degree of Croatia is three. What does this mean in this context?

e. How many faces are there in this graph?

f. Is there a bridge in this graph? If so, which vertices does it connect to?



Identifying and constructing graphs

5. Consider the following graph.

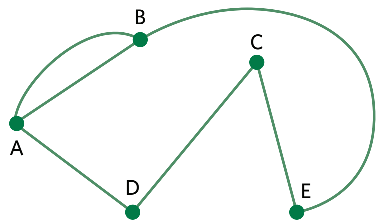
What term can be used to describe this graph?

A. Simple

B. Complete

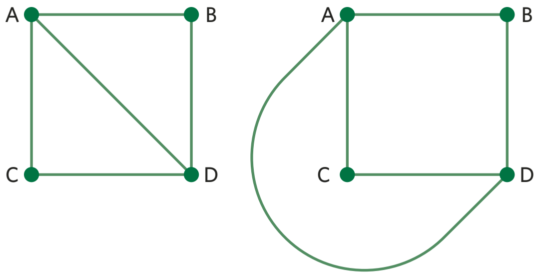
C. Planar

D. Degenerate

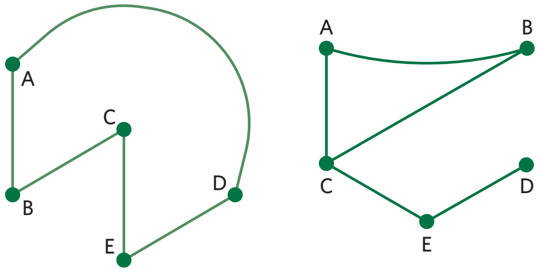


6. Which of the following graphs are isomorphic?

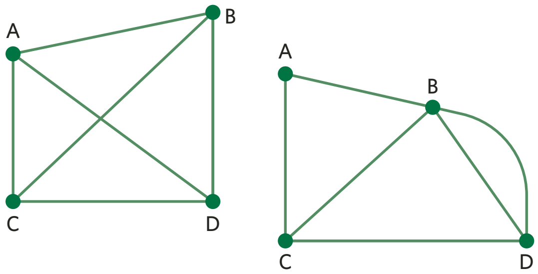
A.



B.



C.



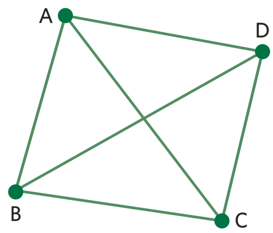
D. None of these.

---

7. Consider the graph.
Is this a connected graph or a degenerate graph? Explain briefly.

---

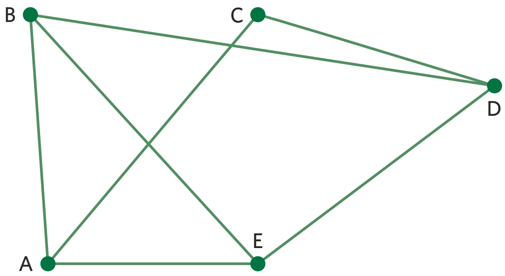
8. Consider the graph.
What type of graph is this? There may be more than one type of graph.



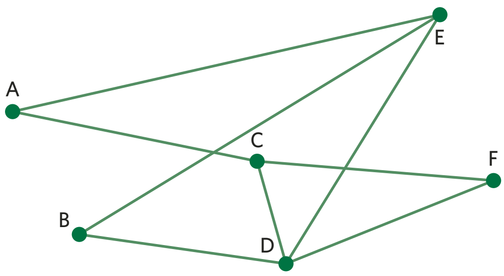
---

9. Redraw the following graphs in planar form.

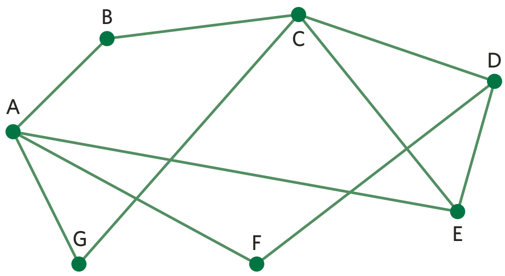
a.



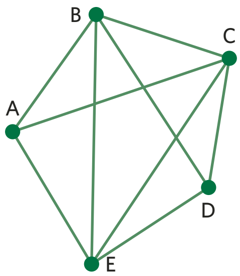
b.



c.



d.



10. Create a graph that represents the following information.

- There are five towns connected by roads: Melbourne, Ballarat, Geelong, Bendigo and Mildura.
- There is a road between Melbourne and Ballarat.
- There is a road between Ballarat and Mildura.
- Bendigo is directly connected to Geelong and Mildura.
- There are two roads connecting Melbourne and Geelong.
- None of the roads cross over each other.

Using Euler's rule

11. Consider a graph that has five vertices and eight edges. If the graph was connected and planar, how many faces would it have?

A. 3

B. 4

C. 5

D. 6

---

12. For a connected planar graph,

a. find v if $e = 6$ and $f = 5$.

b. find e if $v = 4$ and $f = 6$.

c. find f if $v = 7$ and $e = 9$.

d. find e if $v = 5$ and $f = 8$.

e. find f if $v = 10$ and $e = 13$.

f. find v if $e = 14$ and $f = 7$.

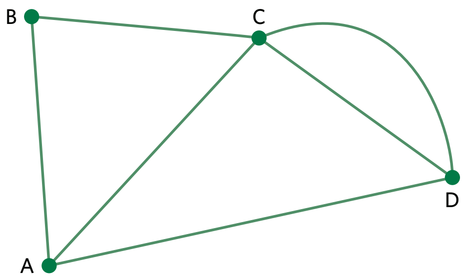
---

13. The city council has decided they want to plant grass in the areas between and around some roads. There are seven roads in the area that connect four landmarks and none of them intersect. It is only possible to travel along these seven roads to all of the landmarks. How many grassy areas will there be?

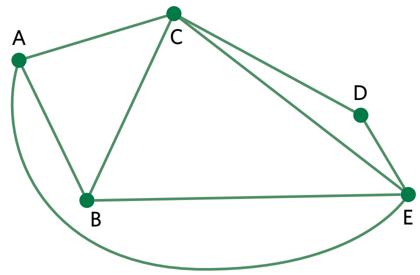
---

14. Determine the values of v, e and f for the following graphs and use Euler's rule to show that these graphs are connected and planar.

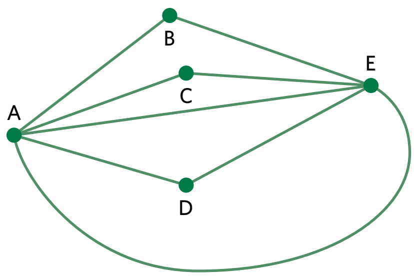
a.



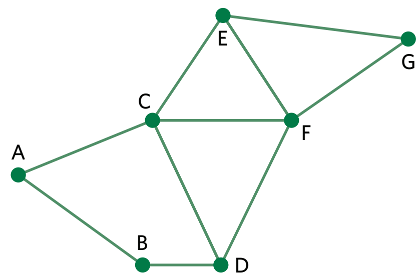
b.



c.



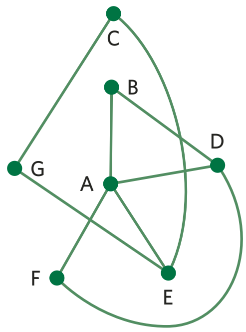
d.



Back to contents

Joining it all together

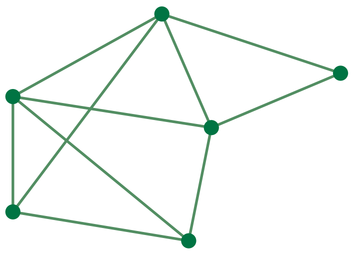
15. Consider the following graph.



a. Between which vertices does a bridge exist?

b. Redraw the graph in planar form.

16. Consider the following graph.



a. What type of graph is this? There may be more than one type of graph.

b. Redraw this graph without any edges overlapping.

c. How many faces are in this graph?

d. Calculate the sum of the degrees in this graph.

e. How many vertices have an odd degree?

17. A group of seven people all give handshakes to each other.

a. Create a graph that represents this.

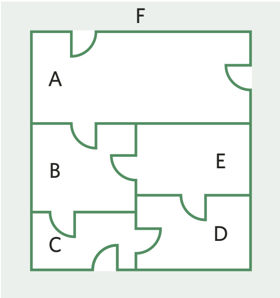
b. What type of graph is this? There may be more than one type of graph.

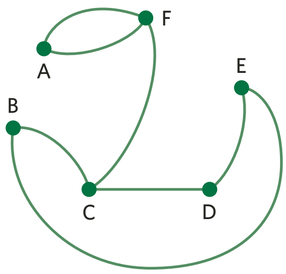
c. All vertices have the same degree. What is the degree of each vertex?

Exam practice

18. Maggie's house has five rooms, A, B, C, D and E, and eight doors. The floor plan of these rooms and doors is shown. The outside area, F, is shown shaded on the floor plan.

The floor plan is represented by the following 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. Which vertices should the edge be drawn between? (1 MARK)

b. What is the degree of vertex E? (1 MARK)

VCAA 2021 Exam 2 Networks and decision mathematics Q1a,b

Part **a**: **93%** of students answered this question correctly.

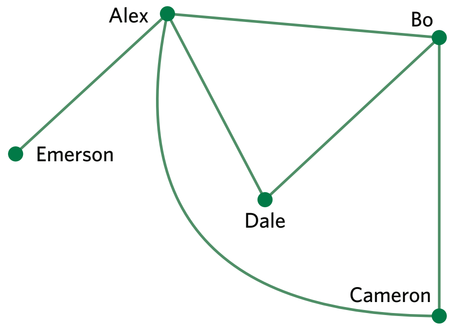
Part **b**: **88%** of students answered this question correctly.

Back to contents

19. The Sunny Coast Cricket Club has five new players join its team: Alex, Bo, Cameron, Dale and Emerson. The following 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.

Who had played cricket with both Alex and Bo before joining the team? (1 MARK)

VCAA 2020 Exam 2 Networks and decision mathematics Q1b



58% of students answered this question correctly.

---

20. Consider the following graph.

The number of faces is

A. 2

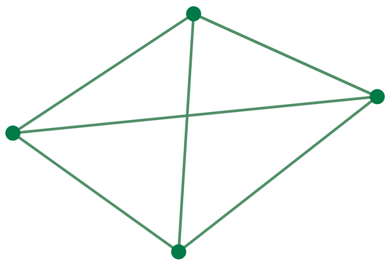
B. 3

C. 4

D. 5

E. 6

VCAA 2021 Exam 1 Networks and decision mathematics Q3



34% of students answered this question correctly.

Questions from multiple lessons

Data analysis

21. The following scatterplot shows *income*, in thousands of dollars, plotted against *happiness*, on a scale from one to ten. A least squares line has been fitted to the data and the resulting residual plot is also shown.

The equation of this least squares line is

happiness $= 1.00 + 0.139 \times$ *income*

The coefficient of determination is $r^2 = 0.768$.

Which of the following statements is **not** true?

A. Using the least squares line to predict the *happiness* of a person with an *income* of \$50 000 is an example of interpolation.

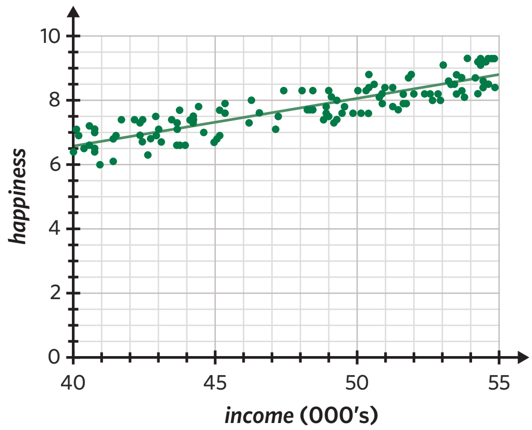
B. The value of the correlation coefficient is close to 0.88.

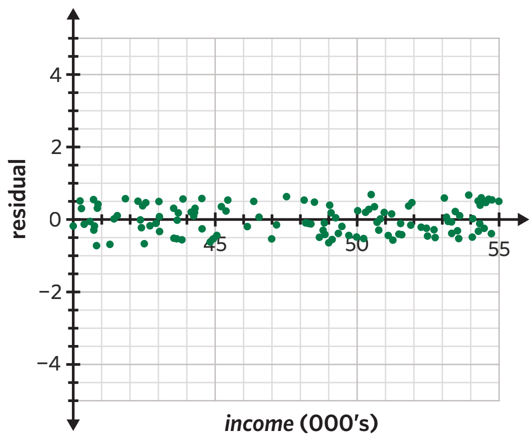
C. 23.2% of the variation in *happiness* is not explained by variation in *income*.

D. On average, *happiness* increases by 1.00 for every \$10 000 increase in *income*.

E. Ignoring any outliers, the association between *happiness* and *income* can be described as strong, linear and positive.

Adapted from VCAA 2016 Exam 1 Data analysis Q9





Back to contents

Recursion and financial modelling

22. Five 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
8	75.00	9.46	84.46	1661.77
9	75.00	9.97	84.97	1746.74
10	75.00	10.48	85.48	1832.22
11	75.00	10.99	85.99	1918.22
12				2000.00

The balance of the investment after payment number 12 is $2000.

The value of payment number 12 is closest to

A. $70 **B.** $75 **C.** $82 **D.** $87 **E.** $90

Adapted from VCAA 2017 Exam 1 Recursion and financial modelling Q23

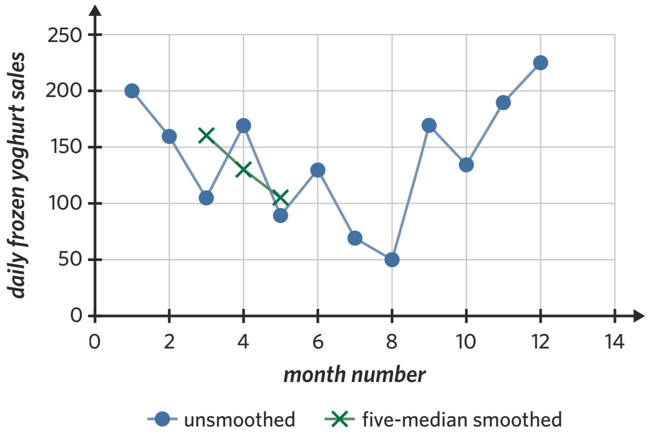
Data analysis

23. A local frozen yoghurt establishment decided to take note of their average daily sales every month of the year. The *daily frozen yoghurt sales* were recorded each month and were plotted against the *month number* (1 = January, 2 = February, and so on) in the following graph.

The data was collected over a period of one year.

a. Five-median smoothing has been used to smooth the time series plot.

The first three smoothed points are shown as crosses (X). Complete the five-median smoothing by marking smoothed values with crosses. (2 MARKS)



The *daily ice cream* sales each month of the store next door were also recorded in the following table.

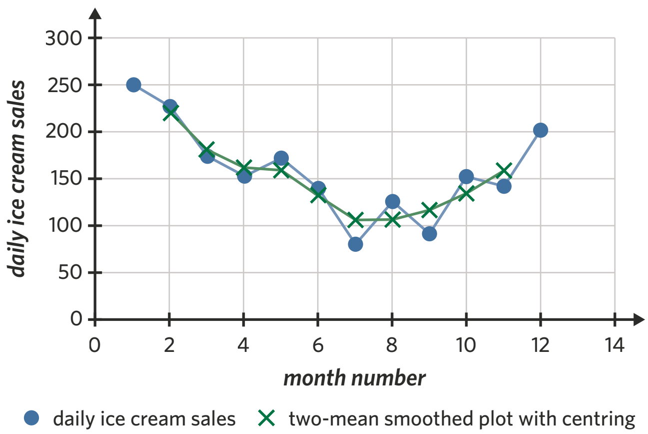
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
month number	1	2	3	4	5	6	7	8	9	10	11	12
daily ice cream sales	251	227	174	153	172	139	81	126	93	152	143	201

The data in the table has been used to plot *daily ice cream sales* against *month number* in a time series plot.

b. Two-mean smoothing with centring has been used to smooth the time series plot.

The smoothed values are marked with crosses (X). Using the data given in the table, show that the two-mean smoothed ice cream sales centred on July is 106.75. (2 MARKS)

Adapted from VCAA 2016 Exam 2 Data analysis Q4



Back to contents

8B Graphs, networks and matrices

STUDY DESIGN DOT POINT

- use of matrices to represent graphs, digraphs and networks and their application

8A 8B 8C 8D 8E 8F 8G 8H 8I 8J

KEY SKILLS

During this lesson, you will be:

- constructing an adjacency matrix from a graph
- constructing a graph from an adjacency matrix
- representing directed graphs.

KEY TERMS

- Adjacency matrix
- Directed graph (digraph)

Matrices can be used to represent the links between the vertices in a graph, by displaying the number of edges connecting each vertex. These matrices can be used to construct both undirected and directed graphs. Conversely, these matrices can be constructed if a graph is provided.

Constructing an adjacency matrix from a graph

An **adjacency matrix** represents the number of connections between vertices. It is a square matrix that has a row and column for each vertex and contains the number of edges between each vertex.

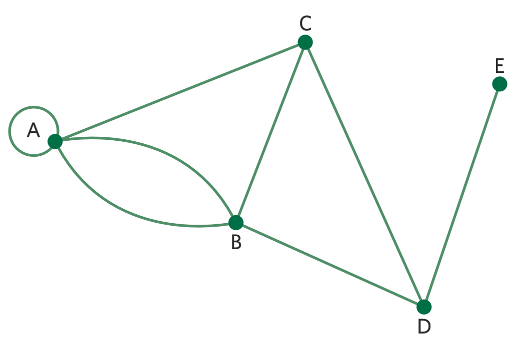
The order of the matrix will represent the number of vertices in the graph since it has a row and column for each vertex. The number of edges between vertices is represented as an element within the matrix. For example, an element of 3 indicates there are three edges connecting two different vertices, or a vertex with itself. A loop counts as one edge.

Columns and rows should add up to the degree of that particular vertex since the degree shows the total number of connections at a vertex. However, if a loop is involved, the columns and rows will add up to one less than the degree of the vertex for each loop involved, as a loop is counted as two degrees but only one edge.

Undirected graphs produce adjacency matrices that are symmetric about the leading diagonal. This is because the number of edges between vertices A and B is the same as the number of edges between vertices B and A.

Worked example 1

Construct an adjacency matrix from the following graph.



Continues →

Back to contents

Explanation

Step 1: Determine the order of the matrix.

All adjacency matrices are square.

There are five vertices.

The order will be 5×5.

Step 2: Label the matrix.

Each vertex will have its own row and column.

$$\begin{array}{c c} \begin{array}{ccccc} A & B & C & D & E \end{array} & \\ \begin{bmatrix} & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \end{bmatrix} & \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} \end{array}$$

Step 3: Fill in row A of the matrix.

There is one edge from vertex A back to vertex A. Element m_{AA} is 1.

There are two edges between vertex A and vertex B. Element m_{AB} is 2.

There is one edge between vertex A and vertex C. Element m_{AC} is 1.

There are no edges between vertex A and vertex D. Element m_{AD} is 0.

There are no edges between vertex A and vertex E. Element m_{AE} is 0.

$$\begin{array}{c c} \begin{array}{ccccc} A & B & C & D & E \end{array} & \\ \begin{bmatrix} 1 & 2 & 1 & 0 & 0 \\ & & & & \\ & & & & \\ & & & & \\ & & & & \end{bmatrix} & \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} \end{array}$$

Step 4: Fill in row B of the matrix.

There are two edges between vertex B and vertex A. Element m_{BA} is 2.

There are no loops at vertex B. Element m_{BB} is 0.

There is one edge between vertex B and vertex C. Element m_{BC} is 1.

There is one edge between vertex B and vertex D. Element m_{BD} is 1.

There are no edges between vertex B and vertex E. Element m_{BE} is 0.

$$\begin{array}{c c} \begin{array}{ccccc} A & B & C & D & E \end{array} & \\ \begin{bmatrix} 1 & 2 & 1 & 0 & 0 \\ 2 & 0 & 1 & 1 & 0 \\ & & & & \\ & & & & \\ & & & & \end{bmatrix} & \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} \end{array}$$

Step 5: Repeat for the remaining rows.

Answer

$$\begin{array}{c c} \begin{array}{ccccc} A & B & C & D & E \end{array} & \\ \begin{bmatrix} 1 & 2 & 1 & 0 & 0 \\ 2 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} & \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} \end{array}$$

Back to contents

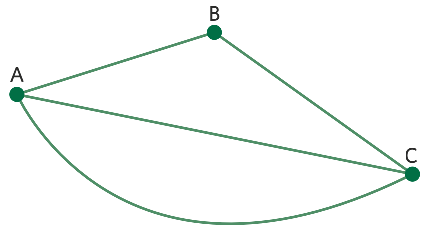
Constructing a graph from an adjacency matrix

Adjacency matrices contain the information needed to construct a graph.

Each row and column in the matrix must be represented as a vertex. The value of each element can then be used to determine the number of edges connecting the two corresponding vertices. If the elements in the leading diagonal are all zeros, then the graph will not contain any loops.

For example, element m_{AC} in the following matrix is 2. This indicates that there are two edges between vertex A and vertex C.

$$\begin{array}{c c} \begin{matrix} A & B & C \end{matrix} & \\ \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix} & \begin{matrix} A \\ B \\ C \end{matrix} \end{array}$$



Worked example 2

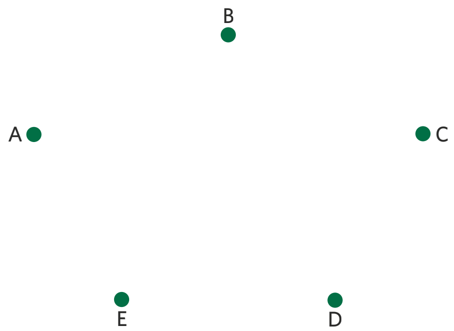
Construct a graph from the following adjacency matrix.

$$\begin{array}{c c} \begin{matrix} A & B & C & D & E \end{matrix} & \\ \begin{bmatrix} 0 & 1 & 2 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 2 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 2 \\ 1 & 0 & 0 & 2 & 0 \end{bmatrix} & \begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix} \end{array}$$

Explanation

Step 1: Draw and label the vertices.

The order of the matrix is 5×5. There will be five vertices.



Step 2: Draw the edges connecting to vertex A.

Note that these can be straight or curved lines.

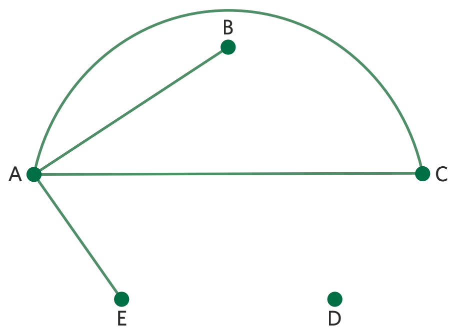
$m_{AA} = 0$, so there is no loop at vertex A.

$m_{AB} = 1$, so there is one edge between vertices A and B.

$m_{AC} = 2$, so there are two edges between vertices A and C.

$m_{AD} = 0$, so there are no edges between vertices A and D.

$m_{AE} = 1$, so there is one edge between vertices A and E.



Continues →

Back to contents

Step 3: Draw the edges connecting to vertex B.

$m_{BA} = 1$, so there is one edge between vertices B and A.

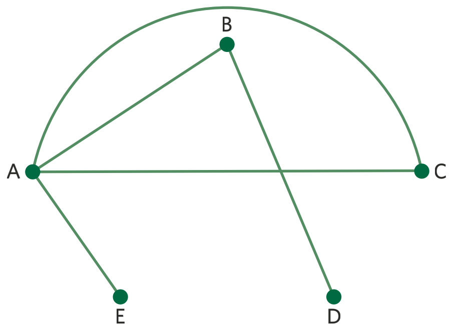
$m_{BB} = 0$, so there is no loop at vertex B.

$m_{BC} = 0$, so there are no edges between vertices B and C.

$m_{BD} = 1$, so there is one edge between vertices B and D.

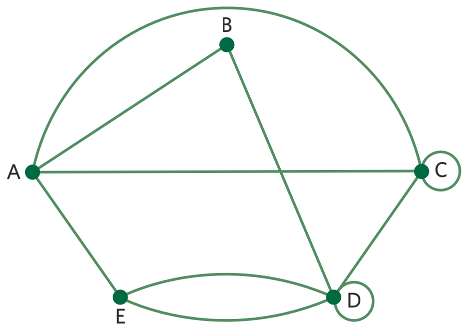
$m_{BE} = 0$, so there are no edges between vertices B and E.

Note that the edge between vertices A and B is already drawn so only the edge from B to D needs to be added.



Step 4: Repeat for the remaining vertices.

Answer

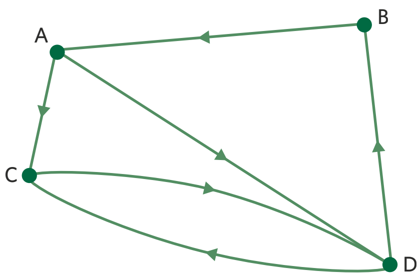


Note: Graphs may look different. A correct graph must simply include all vertices and edges leading to and from them.

Representing directed graphs

A **directed graph** is a network containing arrows on each edge that show the way in which one can travel between two vertices.

For example, the following directed graph shows the roads connecting four towns.



The edge from A to D shows a one-way road from town A to town D.

The edges between C and D show that traffic is allowed in both directions between towns C and D.

A directed graph can be represented in an adjacency matrix by labelling the rows as 'from' and the columns as 'to' in order to show the directional flow of each edge. Each element in the matrix represents the number of edges going from the row vertex to the column vertex. Adjacency matrices for directed graphs do not have to be symmetric.

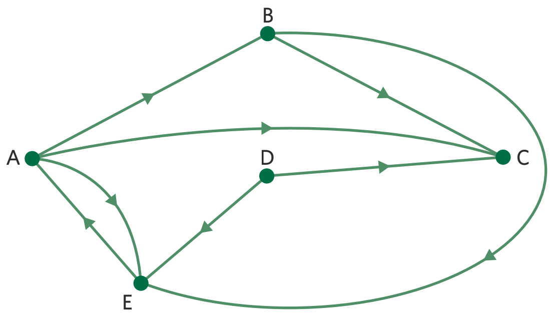
See worked example 3

Directional information contained in an adjacency matrix can be used to construct a digraph.

See worked example 4

Worked example 3

Represent the following directed graph in an adjacency matrix.



Explanation

Step 1: Determine the order of the matrix.

All adjacency matrices are square.

There are five vertices.

The order will be 5 × 5.

Step 2: Label the matrix.

Each vertex will have its own row and column.

Label the rows with 'from' and the columns with 'to'.

$$\begin{array}{c} \text{to} \\ \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \left[\begin{array}{ccccc} & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \end{array}\right] \end{array} \begin{array}{l} \\ \\ A \\ B \\ C \quad \text{from} \\ D \\ E \end{array}$$

Step 3: Fill in row A of the matrix.

There are no loops at vertex A. Element m_{AA} is 0.

There is one edge from vertex A to vertex B. Element m_{AB} is 1.

There is one edge from vertex A to vertex C. Element m_{AC} is 1.

There are no edges from vertex A to vertex D. Element m_{AD} is 0.

There is one edge from vertex A to vertex E. Element m_{AE} is 1.

$$\begin{array}{c} \text{to} \\ \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \left[\begin{array}{ccccc} 0 & 1 & 1 & 0 & 1 \\ & & & & \\ & & & & \\ & & & & \\ & & & & \end{array}\right] \end{array} \begin{array}{l} \\ \\ A \\ B \\ C \quad \text{from} \\ D \\ E \end{array}$$

Step 4: Fill in row B of the matrix.

There are no edges from vertex B to vertex A. Element m_{BA} is 0.

There are no loops at vertex B. Element m_{BB} is 0.

There is one edge from vertex B to vertex C. Element m_{BC} is 1.

There are no edges from vertex B to vertex D. Element m_{BD} is 0.

There is one edge from vertex B to vertex E. Element m_{BE} is 1.

$$\begin{array}{c} \text{to} \\ \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \left[\begin{array}{ccccc} 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ & & & & \\ & & & & \\ & & & & \end{array}\right] \end{array} \begin{array}{l} \\ \\ A \\ B \\ C \quad \text{from} \\ D \\ E \end{array}$$

Step 5: Repeat for the remaining rows.

Continues →

Back to contents

Answer

$$\begin{array}{cl} \begin{array}{ccccc} & & \text{to} & & \\ A & B & C & D & E \end{array} & \\ \begin{bmatrix} 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} & \begin{array}{l} A \\ B \\ C \ \ \text{from} \\ D \\ E \end{array} \end{array}$$

Worked example 4

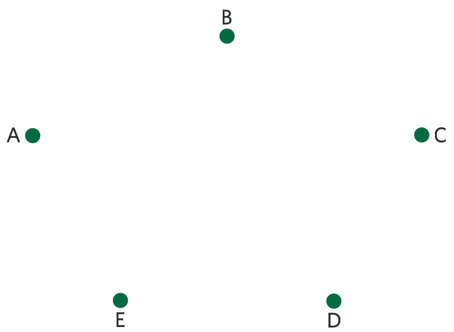
Construct a digraph from the adjacency matrix.

$$\begin{array}{cl} \begin{array}{ccccc} & & \text{to} & & \\ A & B & C & D & E \end{array} & \\ \begin{bmatrix} 0 & 2 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 2 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} & \begin{array}{l} A \\ B \\ C \ \ \text{from} \\ D \\ E \end{array} \end{array}$$

Explanation

Step 1: Draw and label the vertices.

The order of the matrix is 5 × 5. There will be five vertices.



Step 2: Draw the edges connecting from vertex A.

Note that these can be straight or curved lines.

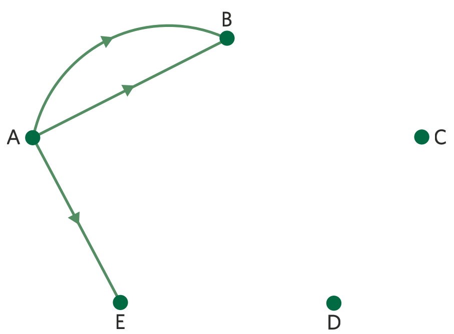
m_{AA} = 0. There is no loop at vertex A.

m_{AB} = 2. There are two edges from vertex A to B.

m_{AC} = 0. There are no edges from vertex A to C.

m_{AD} = 0. There are no edges from vertex A to D.

m_{AE} = 1. There is one edge from vertex A to E.



Step 3: Draw the edges connecting from vertex B.

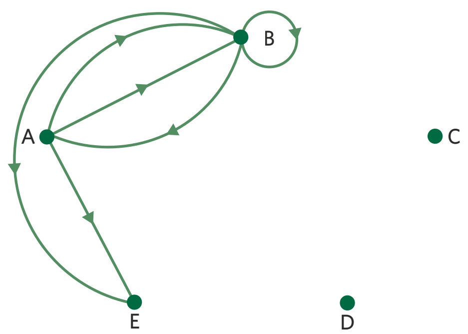
m_{BA} = 1. There is one edge from vertex B to A.

m_{BB} = 1. There is a loop at vertex B.

m_{BC} = 0. There are no edges from vertex B to C.

m_{BD} = 0. There are no edges from vertex B to D.

m_{BE} = 1. There are no edges from vertex B to E.

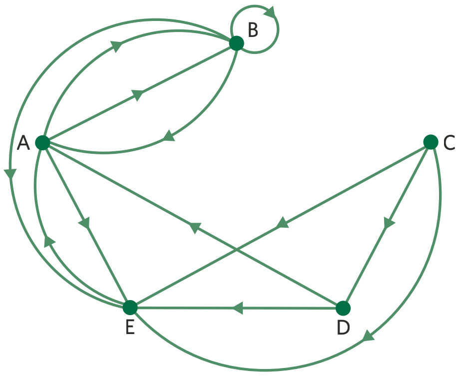


Step 4: Repeat for the remaining vertices.

Continues →

Back to contents

Answer

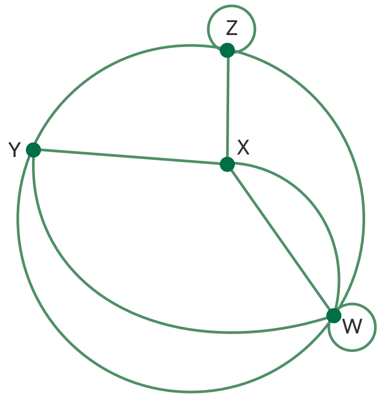


Note: Graphs may look different. A correct graph must simply include all vertices and edges leading to and from them, with the correct directional information.

Exam question breakdown

VCAA 2017 Exam 1 Networks and decision mathematics Q3

Consider the following graph.



The adjacency matrix for this graph, with some elements missing, is shown.

$$\begin{array}{cc} \begin{matrix} W & X & Y & Z \end{matrix} & \\ \begin{bmatrix} 1 & & & \\ & 0 & & \\ & & 0 & \\ & & & 1 \end{bmatrix} & \begin{matrix} W \\ X \\ Y \\ Z \end{matrix} \end{array}$$

This adjacency matrix contains 16 elements when complete.

Of the 12 missing elements

A. eight are '1' and four are '2'.

B. four are '1' and eight are '2'.

C. six are '1' and six are '2'.

D. two are '0', six are '1' and four are '2'.

E. four are '0', four are '1' and four are '2'.

Explanation

Step 1: Determine the missing elements in row W of the matrix.

There are two edges between vertex W and vertex X. Element m_{WX} is 2.

There are two edges between vertex W and vertex Y. Element m_{WY} is 2.

There is one edge between vertex W and vertex Z. Element m_{WZ} is 1.

Step 2: Determine the missing elements in row X of the matrix.

There are two edges between vertex X and vertex W. Element m_{XW} is 2.

There is one edge between vertex X and vertex Y. Element m_{XY} is 1.

There is one edge between vertex X and vertex Z. Element m_{XZ} is 1.

Step 3: Determine the missing elements in row Y of the matrix.

There are two edges between vertex Y and vertex W. Element m_{YW} is 2.

There is one edge between vertex Y and vertex X. Element m_{YX} is 1.

There is one edge between vertex Y and vertex Z. Element m_{YZ} is 1.

Step 4: Determine the missing elements in row Z of the matrix.

There is one edge between vertex Z and vertex W. Element m_{ZW} is 1.

There is one edge between vertex Z and vertex X. Element m_{ZX} is 1.

There is one edge between vertex Z and vertex Y. Element m_{ZY} is 1.

Continues →

Back to contents

Step 5: Determine the total number of each missing element.

Eight of the missing elements are '1'.

Four of the missing elements are '2'.

Answer

A

92% of students answered this question correctly.

5% of students incorrectly chose either D or E. It was important to be aware that the graph was a complete graph which means each vertex is directly connected to every other vertex by at least one edge, so none of the missing elements should be '0'.

8B Questions

Constructing an adjacency matrix from a graph

1. Which of the following is not an adjacency matrix for an undirected graph?

A. $\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$

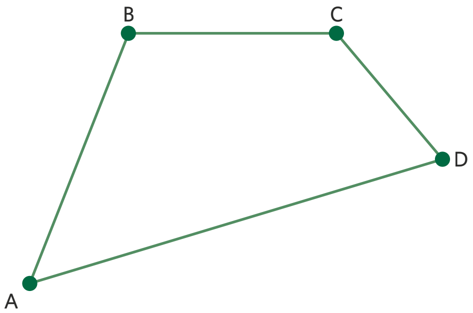
B. $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 2 \\ 0 & 2 & 0 \end{bmatrix}$

C. $\begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$

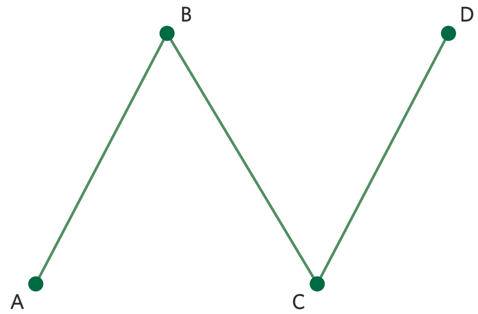
D. $\begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$

2. Construct an adjacency matrix from the following graphs.

a.

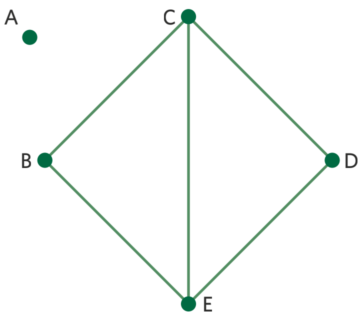


b.

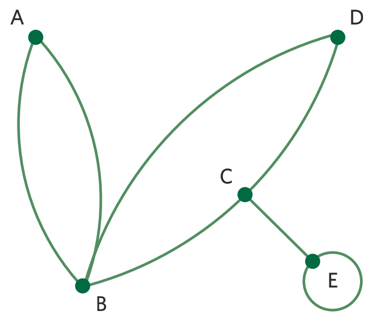


3. Construct an adjacency matrix from the following graphs.

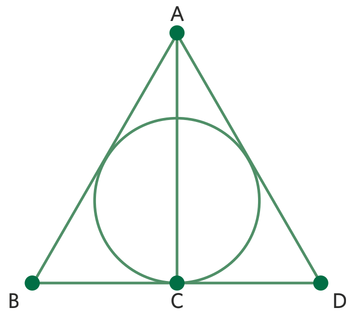
a.



b.



4. Harry Potter can't seem to remember how to draw the symbol of the Deathly Hallows. Create an adjacency matrix that will help Harry remember which vertices are connected.

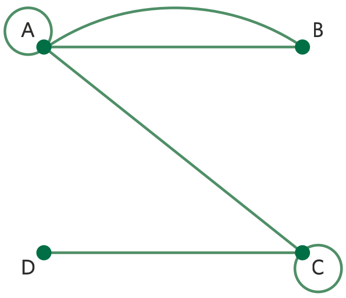


Constructing a graph from an adjacency matrix

5. Consider the following adjacency matrix.

$$\begin{array}{cc} \begin{array}{cccc} A & B & C & D \end{array} & \\ \begin{bmatrix} 1 & 2 & 1 & 0 \\ 2 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} & \begin{array}{c} A \\ B \\ C \\ D \end{array} \end{array}$$

The matrix is used to construct the following incorrect graph.



To make the graph correct,

A. an extra edge needs to be added between vertices A and B.

B. an edge needs to be added between vertices B and D.

C. an edge needs to be removed between vertices A and B.

D. a loop needs to be added on vertex D.

6. Construct a graph from the following adjacency matrices.

a.
$$\begin{array}{cc} \begin{array}{cccc} A & B & C & D \end{array} & \\ \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} & \begin{array}{c} A \\ B \\ C \\ D \end{array} \end{array}$$

b.
$$\begin{array}{cc} \begin{array}{cccc} A & B & C & D \end{array} & \\ \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} & \begin{array}{c} A \\ B \\ C \\ D \end{array} \end{array}$$

7. Construct a graph from the following adjacency matrices.

a.
$$\begin{array}{cc} \begin{array}{ccccc} A & B & C & D & E \end{array} & \\ \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 2 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 2 & 0 & 1 & 0 \end{bmatrix} & \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} \end{array}$$

b.
$$\begin{array}{cc} \begin{array}{cccccc} A & B & C & D & E & F \end{array} & \\ \begin{bmatrix} 1 & 2 & 1 & 2 & 1 & 0 \\ 2 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} & \begin{array}{c} A \\ B \\ C \\ D \\ E \\ F \end{array} \end{array}$$

Back to contents

8. The following adjacency matrix shows the railway connections between Melbourne CBD and seven regional towns in Victoria.

Construct a graph to show Victoria's regional train services.

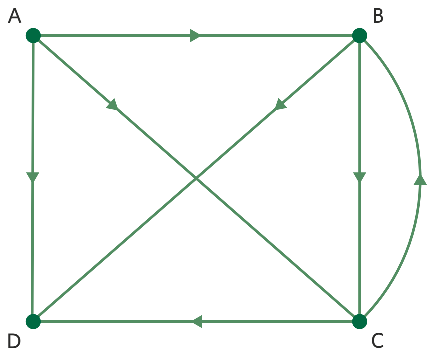
$$\begin{array}{c}
\begin{array}{cccccccc} \text{Ararat} & \text{Ballarat} & \text{Bendigo} & \text{Echuca} & \text{Geelong} & \text{Maryborough} & \text{Melbourne} & \text{Warrnambool} \end{array} \\
\left[\begin{array}{cccccccc}
0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0
\end{array}\right]
\begin{array}{l} \text{Ararat} \\ \text{Ballarat} \\ \text{Bendigo} \\ \text{Echuca} \\ \text{Geelong} \\ \text{Maryborough} \\ \text{Melbourne} \\ \text{Warrnambool} \end{array}
\end{array}$$

Representing directed graphs

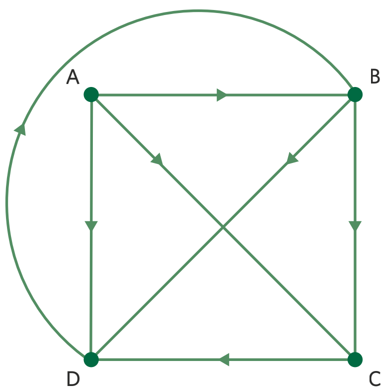
9. Which is the correct graph to represent the following adjacency matrix?

$$\begin{array}{c}
\text{to} \\
\begin{array}{cccc} A & B & C & D \end{array} \\
\left[\begin{array}{cccc}
0 & 1 & 1 & 1 \\
0 & 0 & 1 & 1 \\
0 & 1 & 0 & 1 \\
0 & 1 & 0 & 0
\end{array}\right]
\begin{array}{l} A \\ B \\ C \\ D \end{array}
\end{array} \text{ from}$$

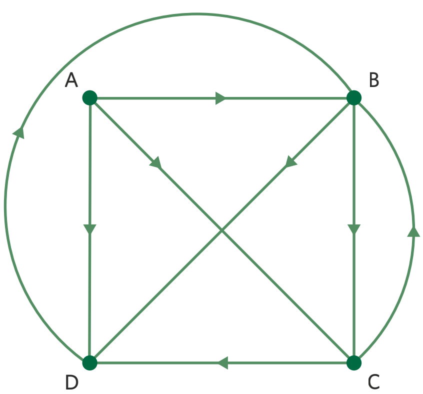
A.



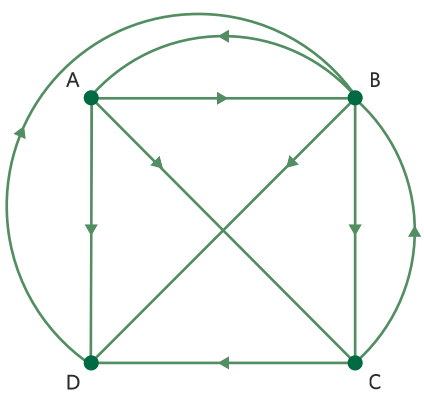
B.



C.

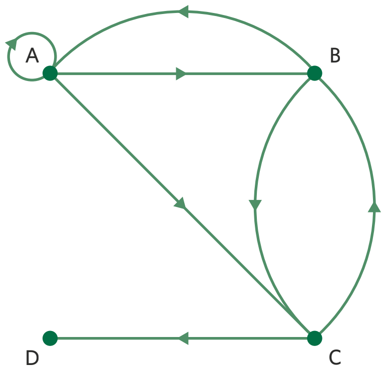


D.

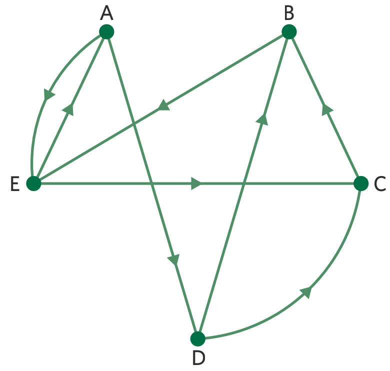


10. Construct an adjacency matrix to represent each of the following directed graphs.

a.



b.



11. Use the following adjacency matrices to construct directed graphs.

a.

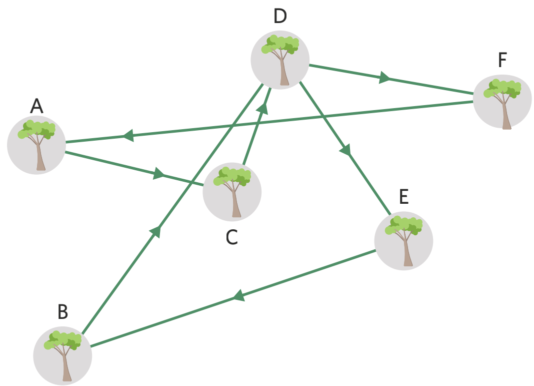
$$\begin{array}{c} \text{to} \\ \begin{array}{cc} \begin{array}{cccc} A & B & C & D \end{array} & \\ \begin{bmatrix} 0 & 2 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} & \begin{array}{c} A \\ B \\ C \\ D \end{array} \end{array} \text{ from} \end{array}$$

b.

$$\begin{array}{c} \text{to} \\ \begin{array}{cc} \begin{array}{ccccc} A & B & C & D & E \end{array} & \\ \begin{bmatrix} 0 & 1 & 0 & 0 & 2 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix} & \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} \end{array} \text{ from} \end{array}$$

12. Kayla decided to take her dogs for a walk in the local park. She starts and ends at tree A. The following graph shows the route she took through the park.

Kayla's dogs really enjoyed the walk so she wants to record the route in an adjacency matrix to remember it for the future. Construct an adjacency matrix to show Kayla's walk.



Joining it all together

13. A museum has several exhibits on display. The exhibits are connected by various different pathways. These pathways are summarised in the following adjacency matrix.

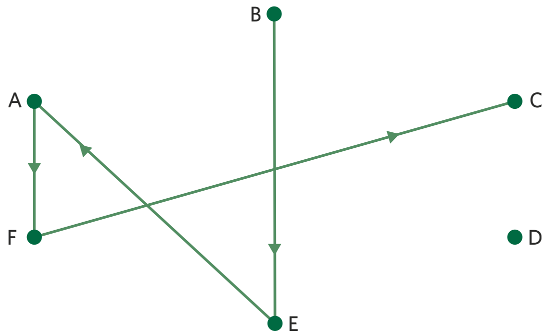
$$\begin{array}{cc} \begin{array}{cccccc} A & B & C & D & E & F \end{array} & \\ \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \end{bmatrix} & \begin{array}{c} A \\ B \\ C \\ D \\ E \\ F \end{array} \end{array}$$

a. Construct a graph showing the pathways between the exhibits.

b. The museum is undergoing renovations and, as a result, the pathway between exhibit A and exhibit B is no longer accessible. They decided to establish a new pathway between exhibits A and E. Construct a new adjacency matrix showing the museum's pathways.

Back to contents

c. Sarah is visiting the museum. She only has time to see five of the exhibits. The graph shows the exhibits she decides to see and the route she takes.



Show this information in an adjacency matrix.

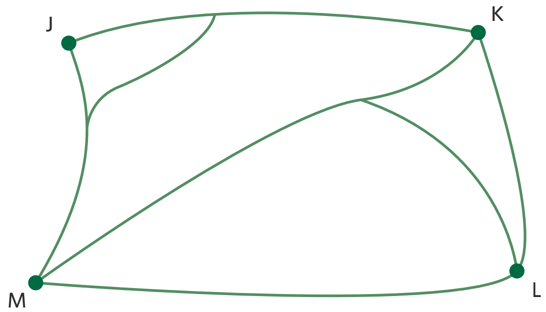
Exam practice

14. The adjacency matrix shows the number of pathway connections between four landmarks: J, K, L and M.

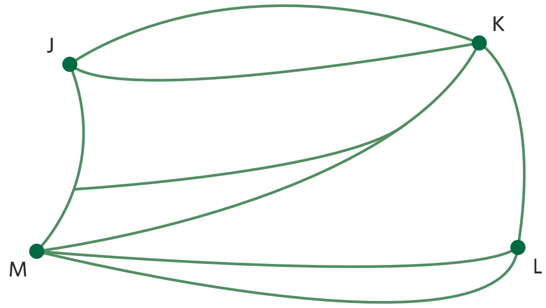
$$\begin{array}{c} \begin{array}{cccc} J & K & L & M \end{array} \\ \begin{bmatrix} 1 & 3 & 0 & 2 \\ 3 & 0 & 1 & 2 \\ 0 & 1 & 0 & 2 \\ 2 & 2 & 2 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ J \\ K \\ L \\ M \end{array}$$

A network of pathways that could be represented by the adjacency matrix is

A.

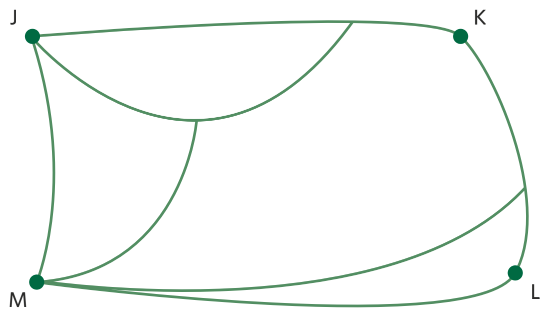


B.

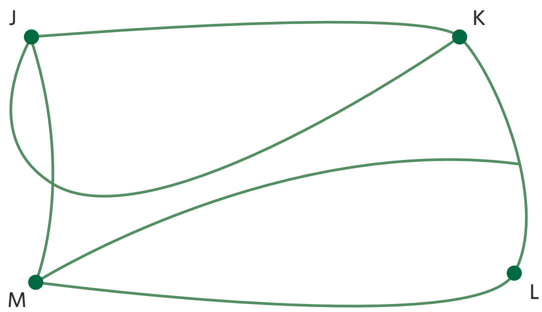


C.

D.



E.

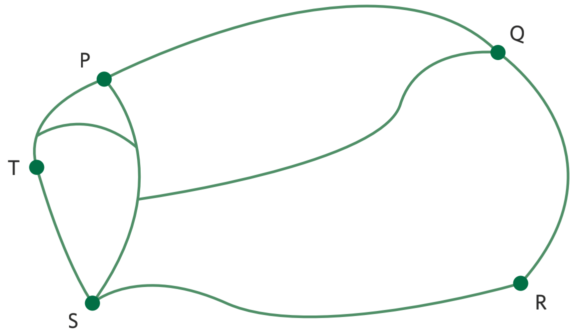


VCAA 2020 Exam 1 Networks and decision mathematics Q8

75% of students answered this question correctly.

Back to contents

15. The map shows all the road connections between five towns, P, Q, R, S and T.



The road connections could be represented by the adjacency matrix

A.
$$\begin{array}{cc} \begin{matrix} P & Q & R & S & T \end{matrix} & \\ \begin{bmatrix} 1 & 3 & 0 & 2 & 2 \\ 3 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 2 & 1 & 1 & 0 & 2 \\ 2 & 1 & 0 & 2 & 0 \end{bmatrix} & \begin{matrix} P \\ Q \\ R \\ S \\ T \end{matrix} \end{array}$$

B.
$$\begin{array}{cc} \begin{matrix} P & Q & R & S & T \end{matrix} & \\ \begin{bmatrix} 1 & 2 & 0 & 2 & 2 \\ 2 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 2 & 1 & 1 & 0 & 2 \\ 2 & 1 & 0 & 2 & 0 \end{bmatrix} & \begin{matrix} P \\ Q \\ R \\ S \\ T \end{matrix} \end{array}$$

C.
$$\begin{array}{cc} \begin{matrix} P & Q & R & S & T \end{matrix} & \\ \begin{bmatrix} 0 & 3 & 0 & 2 & 2 \\ 3 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 2 & 1 & 1 & 0 & 2 \\ 2 & 1 & 0 & 2 & 0 \end{bmatrix} & \begin{matrix} P \\ Q \\ R \\ S \\ T \end{matrix} \end{array}$$

D.
$$\begin{array}{cc} \begin{matrix} P & Q & R & S & T \end{matrix} & \\ \begin{bmatrix} 0 & 2 & 0 & 2 & 2 \\ 2 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 2 & 1 & 1 & 1 & 2 \\ 2 & 1 & 0 & 2 & 0 \end{bmatrix} & \begin{matrix} P \\ Q \\ R \\ S \\ T \end{matrix} \end{array}$$

E.
$$\begin{array}{cc} \begin{matrix} P & Q & R & S & T \end{matrix} & \\ \begin{bmatrix} 1 & 2 & 0 & 2 & 2 \\ 2 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 2 & 1 & 1 & 1 & 1 \\ 2 & 1 & 0 & 1 & 0 \end{bmatrix} & \begin{matrix} P \\ Q \\ R \\ S \\ T \end{matrix} \end{array}$$

VCAA 2019 Exam 1 Networks and decision mathematics Q6

39% of students answered this question correctly.

---

16. The network shows the pathways between five buildings: J, K, L, M and N.

An adjacency matrix for this network is formed.

The number of zeros in this matrix is

A. 8

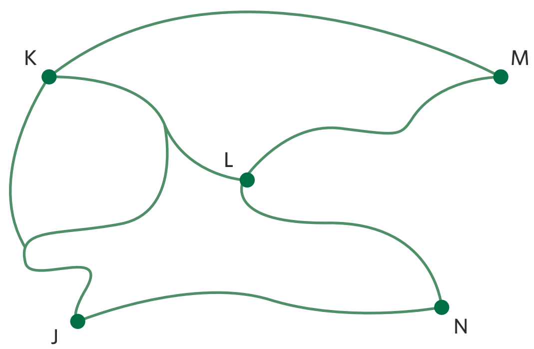
B. 9

C. 10

D. 11

E. 12

VCAA 2021 Exam 1 Networks and decision mathematics Q7



33% of students answered this question correctly.

Questions from multiple lessons

Data analysis

17. The seasonal index for a shop's sales of umbrellas in March is 0.8.

To correct the March sales of umbrellas for seasonality, March's actual number of sales should be

A. increased by 25%

B. increased by 80%

C. reduced by 20%

D. reduced by 25%

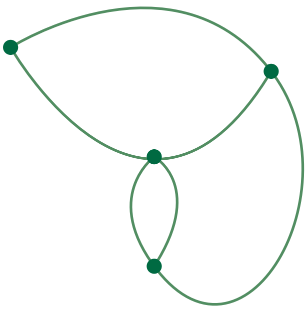
E. reduced by 80%

Adapted from VCAA 2017 Exam 1 Data analysis Q16

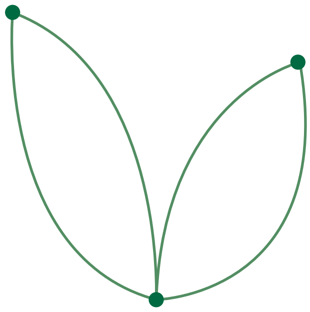
Networks and decision mathematics

18. Which of the following graphs contains a loop?

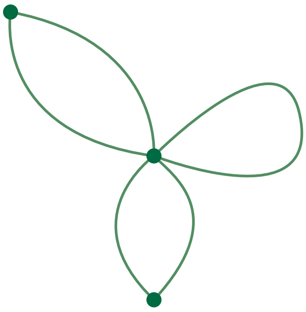
A.



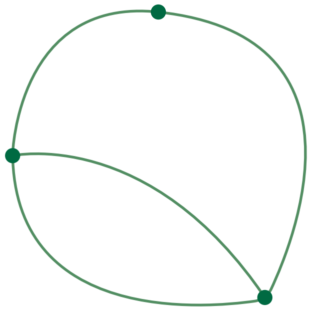
B.



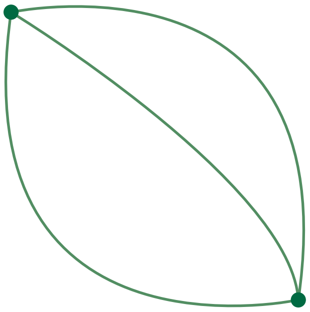
C.



D.



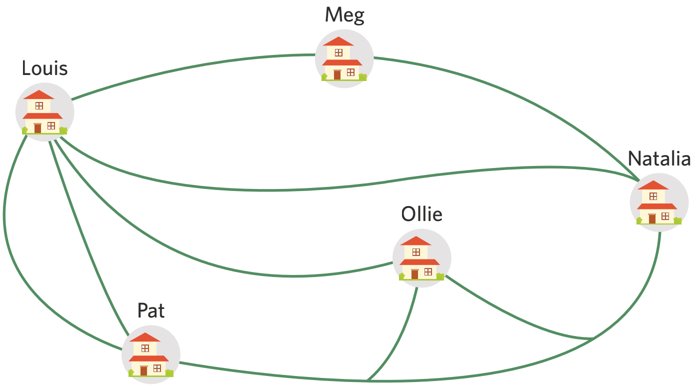
E.



Adapted from VCAA 2017 Exam 1 Network and decision mathematics Q1

Networks and decision mathematics *Year 11 content*

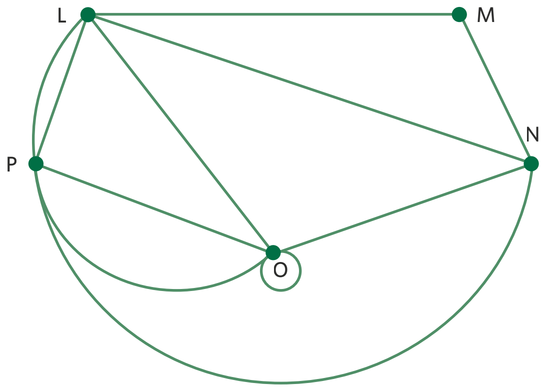
19. Shown is a map of the roads between five friends' houses; Louis (L), Meg (M), Natalia (N), Ollie (O), and Pat (P).



a. Meg is currently at her house. Which two friends can she visit, travelling along one road only? (1 MARK)

Back to contents

b. The map has been converted into the following graph. However, the graph is missing an edge.



i. Which two vertices is the missing edge between? (1 MARK)

ii. What does the loop at O represent in the context of travel from Ollie's house? (1 MARK)

Adapted from VCAA 2016 Exam 2 Networks and decision mathematics Q1

Back to contents

8C Exploring and travelling problems

STUDY DESIGN DOT POINTS

- 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

8A 8B **8C** 8D 8E 8F 8G 8H 8I 8J

KEY SKILLS

During this lesson, you will be:

- identifying types of walks
- identifying Eulerian trails and circuits
- identifying Hamiltonian paths and cycles.

KEY TERMS

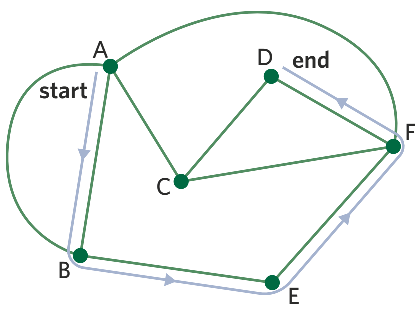
- Route
- Walk
- Trail
- Path
- Circuit
- Cycle
- Eulerian trail
- Eulerian circuit
- Hamiltonian path
- Hamiltonian cycle

Graphs can be used to represent physical locations similar to maps, with locations shown as vertices and roads/paths shown as edges. These can be used to model exploring and travelling problems, with different ways of travelling through graphs described with different terms.

Identifying types of walks

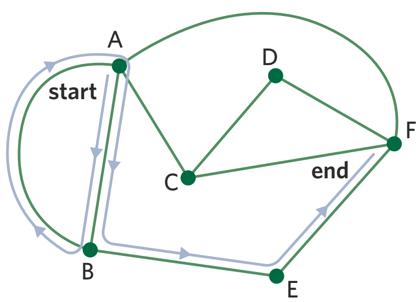
A **route** is a list of the vertices travelled through, in order, when moving from one vertex to another. It shows the pathway travelled in a graph.

For example, the route in the following graph can be written as A–B–E–F–D.



A **walk** is a continuous sequence of edges that passes through any number of vertices, in any order, starting and finishing at any vertex.

For example, the walk A–B–A–B–E–F is shown in the following graph.

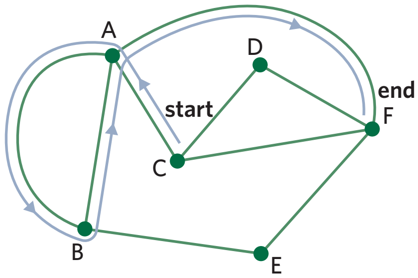


Back to contents

There are different types of walks that exhibit varying properties. These include trails, paths, circuits and cycles.

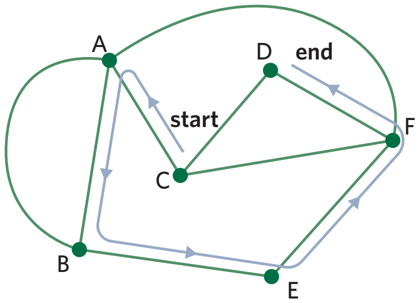
A **trail** is a walk in which no edges are repeated. However, they may pass through the same vertex multiple times.

For example, the walk C–A–B–A–F in the following graph is a trail since there are no repeated edges.



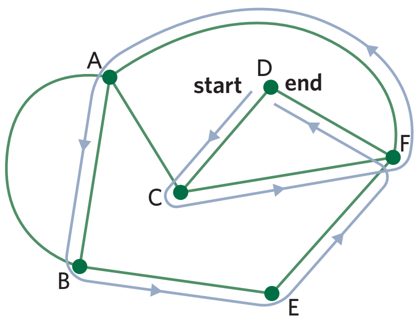
A **path** is a walk in which no edges or vertices are repeated.

For example, the walk C–A–B–E–F–D in the following graph is a path since there are no repeated edges or vertices.



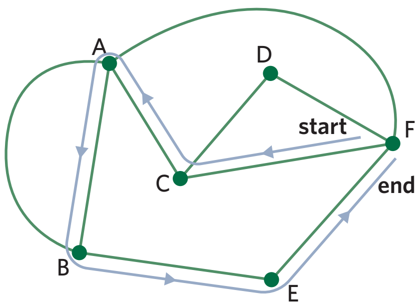
A **circuit** is a trail beginning and ending at the same vertex. Since it is a type of trail, no edges are repeated. Vertices may be repeated.

For example, the walk D–C–F–A–B–E–F–D in the following graph is a circuit because it has no repeated edges, and starts and ends at the same vertex.



A **cycle** is a path beginning and ending at the same vertex. Since it is a type of path, no edges or vertices are repeated, except the start and ending vertex.

For example, the walk F–C–A–B–E–F in the following graph is a cycle because it has no repeated edges or vertices, and starts and ends at the same vertex.

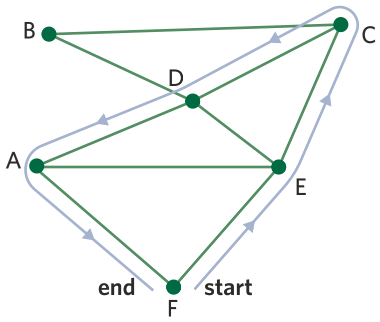


Back to contents

Worked example 1

Determine if the walk in each graph is a trail, path, circuit, cycle, or just a walk.

a.



Explanation

Step 1: Determine whether the walk starts and ends at the same vertex.

The walk starts and ends at vertex F.

Since it starts and ends at the same vertex, consider whether it is a circuit or a cycle.

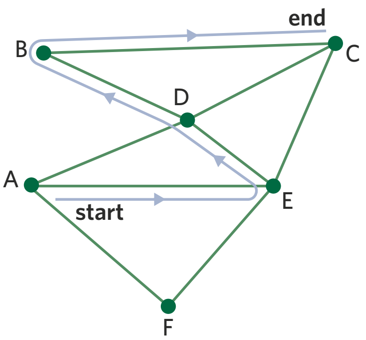
Step 2: Determine whether any edges or vertices are repeated.

There are no repeated edges and no repeated vertices, except the start and ending vertex.

Answer

Cycle

b.



Explanation

Step 1: Determine whether the walk starts and ends at the same vertex.

The walk starts at vertex A and ends at vertex C.

Since it starts and ends at different vertices, consider whether it is a trail or a path.

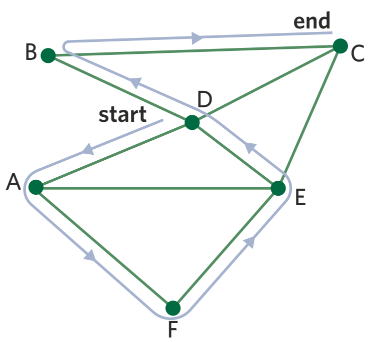
Step 2: Determine whether any edges or vertices are repeated.

There are no repeated edges or vertices.

Answer

Path

c.



Continues →

Explanation

Step 1: Determine whether the walk starts and ends at the same vertex.

The walk starts at vertex D and ends at vertex C.

Since it starts and ends at different vertices, consider whether it is a trail or a path.

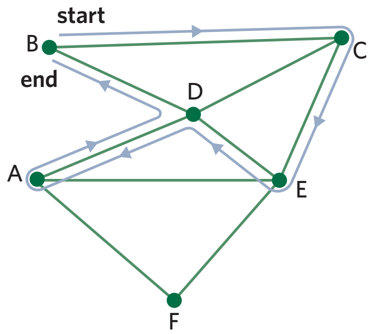
Step 2: Determine whether any edges or vertices are repeated.

There are no repeated edges. Vertex D is repeated.

Answer

Trail

d.



Explanation

Step 1: Determine whether the walk starts and ends at the same vertex.

The walk starts and ends at vertex B.

Since it starts and ends at the same vertex, consider whether it is a circuit or a cycle.

Step 2: Determine whether any edges or vertices are repeated.

The edge between vertex A and vertex D is repeated.

Vertex D is repeated.

Answer

Walk

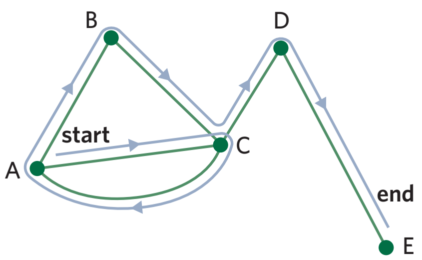
Identifying Eulerian trails and circuits

An **Eulerian trail** is a walk that includes every edge in a graph exactly once. For an Eulerian trail to exist, the graph must:

- be connected
- have exactly two vertices of an odd degree, with all other vertices of an even degree

The Eulerian trail will always start and finish at vertices with an odd degree.

For example, an Eulerian trail A–C–A–B–C–D–E is possible in the following connected graph because vertices A and E have an odd degree and all other vertices have an even degree. The Eulerian trail will start and end at these two vertices since they have an odd degree.

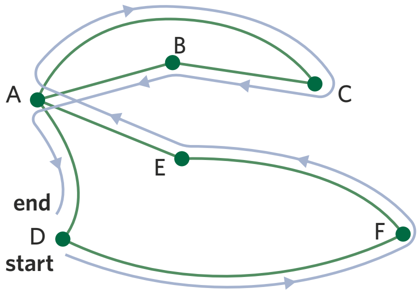


An **Eulerian circuit** is an Eulerian trail that starts and ends at the same vertex. For an Eulerian circuit to exist, the graph must:

- be connected
- have all vertices of even degree.

Back to contents

For example, an Eulerian circuit D–F–E–A–C–B–A–D is possible in the following connected graph because all vertices have an even degree.

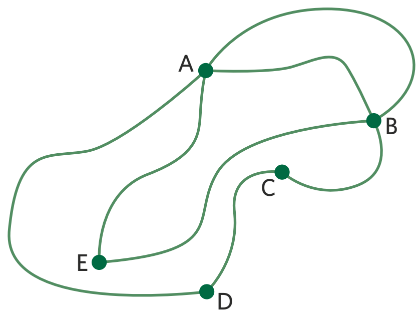


It is possible for multiple different Eulerian trails and circuits to exist in a graph.

Worked example 2

The following graph represents a map of a walking track in the Dandenong Ranges.

The vertices represent lookout points.



a. Is an Eulerian trail possible in this graph?

Explanation

Step 1: Check if the graph is connected.

All vertices in the graph are connected.

Step 2: Count the degrees of all of the vertices.

In order for an Eulerian trail to be possible, exactly two vertices must be of an odd degree.

Vertices A and B are of degree 4.

Vertices C, D and E are of degree 2.

All vertices are of an even degree.

Answer

Not possible

b. Is an Eulerian circuit possible in this graph?

Explanation

Step 1: Check if the graph is connected.

All vertices in the graph are connected.

Step 2: Count the degrees of all of the vertices.

In order for an Eulerian circuit to be possible, all vertices must be of an even degree.

Vertices A and B are of degree 4.

Vertices C, D and E are of degree 2.

All vertices are of an even degree.

Answer

Possible

Continues →

Back to contents

c. Write down an Eulerian circuit starting from lookout B.

Explanation

Step 1: Find the starting vertex.

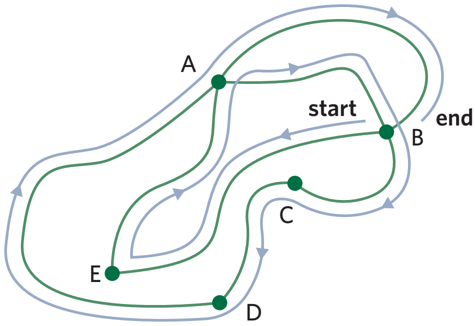
The starting vertex is lookout B.

Step 2: Follow the edges to create an Eulerian circuit.

Starting from lookout B, follow the edges until all have been included exactly once.

Vertices can be repeated.

The walk must end at lookout B.



Answer

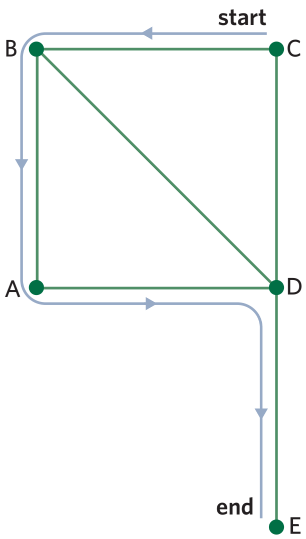
B–E–A–B–C–D–A–B

Note: There are other possible Eulerian circuits.

Identifying Hamiltonian paths and cycles

A **Hamiltonian path** is a walk that includes every vertex in a graph exactly once, with no repeated edges. Although Hamiltonian paths do not repeat edges, they do not necessarily have to include every edge.

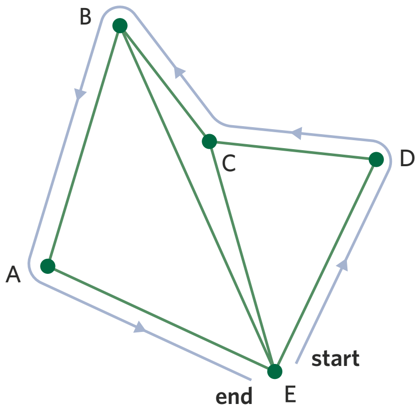
For example, the walk C–B–A–D–E in the following graph is a Hamiltonian path since it includes all vertices and does not repeat any edges.



A **Hamiltonian cycle** is a Hamiltonian path that starts and ends at the same vertex. As with a Hamiltonian path, a Hamiltonian cycle does not repeat edges but does not have to include every edge.

Back to contents

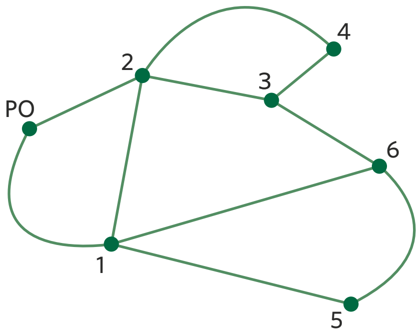
For example, the walk E–D–C–B–A–E in the following graph is a Hamiltonian cycle since it includes all vertices, without repeating edges, and starts and ends at vertex E.



It is possible for multiple Hamiltonian paths and cycles to exist in a graph.

Worked example 3

Max the mailman is doing his morning delivery run and has to deliver packages to six houses. These houses and the post office where the packages are stored are shown in the graph.



a. Write down a Hamiltonian path that allows Max to deliver all of his packages.

Explanation

Step 1: Determine the starting vertex.

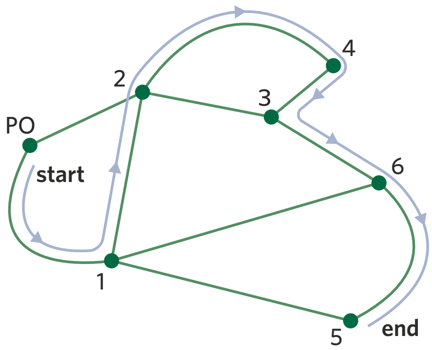
Max will start at the post office because that is where he must pick up his packages.

Step 2: Follow the edges from the starting vertex to create a Hamiltonian path.

Max can travel from the post office to either house 1 or house 2.

He must then travel to every other house exactly once.

He can end at any house.



Answer

PO–1–2–4–3–6–5

Note: There are other possible Hamiltonian paths.

Continues →

b. Write down a Hamiltonian cycle that allows Max to deliver all of his packages before returning to the post office.

Explanation

Step 1: Determine the starting vertex.

Max must start at the post office because that is where he must pick up his packages.

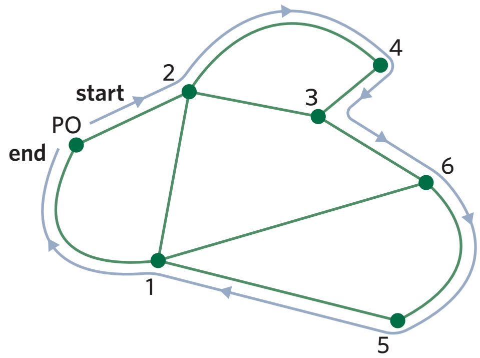
As this is a Hamiltonian cycle, the post office is also the ending vertex.

Step 2: Follow the edges from the starting vertex to create a Hamiltonian cycle.

Max can travel from the post office to either house 1 or house 2.

He must then travel to every other house exactly once.

He must end at the post office.



Answer

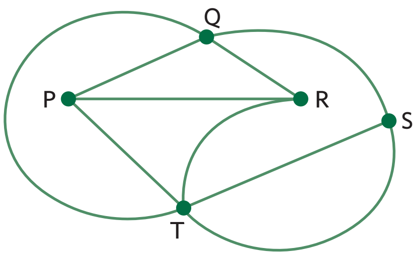
PO–2–4–3–6–5–1–PO

Note: There are other possible Hamiltonian cycles.

Exam question breakdown

VCAA 2018 Exam 1 Networks and decision mathematics Q4

Consider the graph.



Which one of the following is **not** a path for this graph?

A. P–R–Q–T–S **B.** P–Q–R–T–S **C.** P–R–T–S–Q **D.** P–T–Q–S–R **E.** P–T–R–Q–S

Explanation

Determine whether any edges or vertices are repeated in each option.

A: P–R–Q–T–S has no repeated edges or vertices. It is a path. ✗

B: P–Q–R–T–S has no repeated edges or vertices. It is a path. ✗

C: P–R–T–S–Q has no repeated edges or vertices. It is a path. ✗

D: P–T–Q–S–R is not possible unless a vertex is repeated since S and R are not connected. It is not a path. ✓

E: P–T–R–Q–S has no repeated edges or vertices. It is a path. ✗

88% of students answered this question correctly.

7% of students incorrectly chose option A. Students needed to check each route and ensure it had no repeated edges or vertices in order for it to meet the definition of a path.

Answer

C

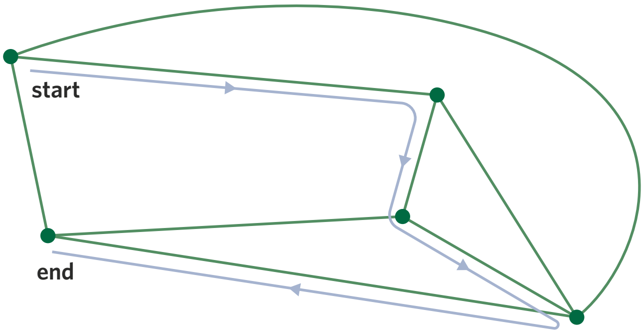
Back to contents

8C Questions

Identifying types of walks

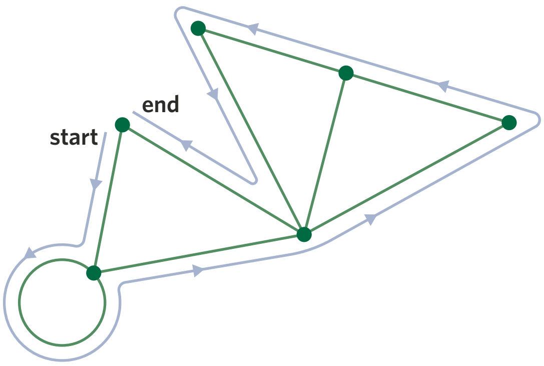
1. What type of walk is shown in the following graph?

A. Trail
B. Path
C. Circuit
D. Cycle

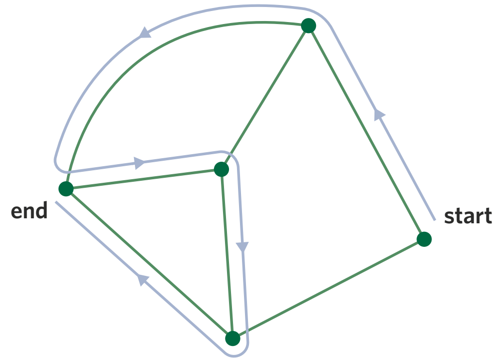


2. Determine if the walk in each of the following graphs is a trail, path, circuit, cycle, or just a walk.

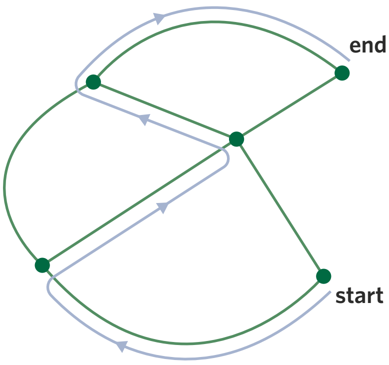
a.



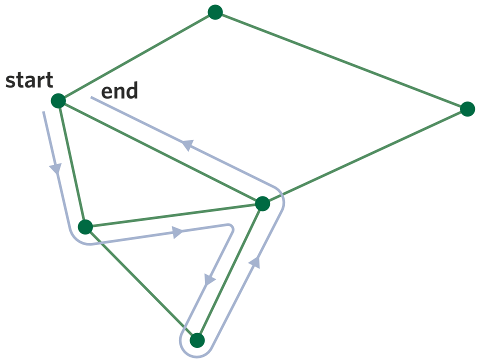
b.



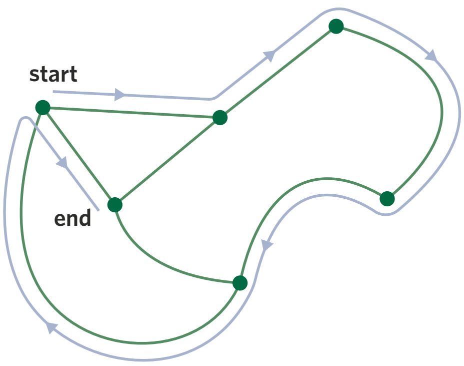
c.



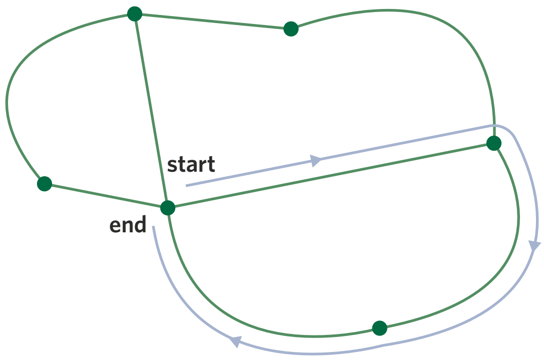
d.



e.



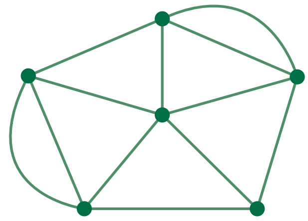
f.



3. Chelsea is going for a walk around her neighbourhood. She wants to make sure she starts and ends at her house and does not walk along any of the roads more than once. She likes her walks to be interesting, so she also does not want to visit any of the intersections more than once. What type of walk is Chelsea hoping to take?

Identifying Eulerian trails and circuits

4. How many edges need to be removed from the following graph for an Eulerian circuit to be possible?



A. 0

B. 1

C. 2

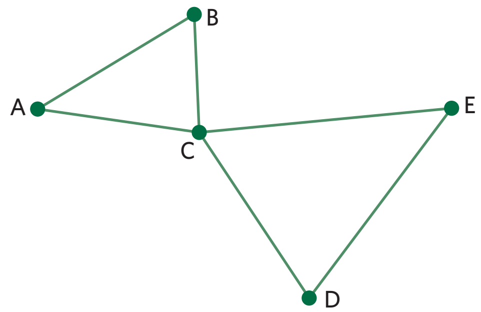
D. 3

---

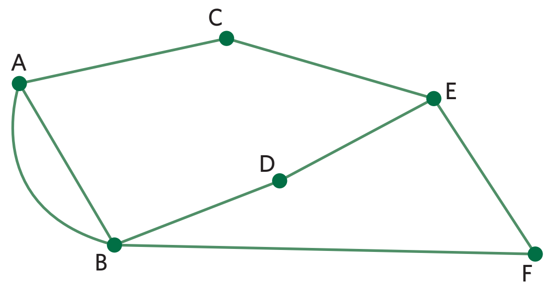
5. For each of the following graphs:

- Identify whether an Eulerian trail or Eulerian circuit is possible.
- If possible, write down an Eulerian trail or Eulerian circuit starting from vertex A.

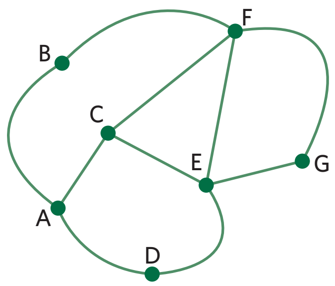
a.



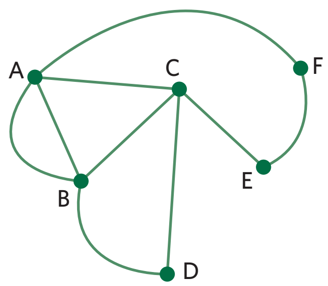
b.



c.



d.

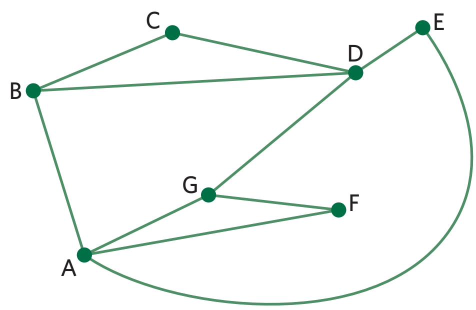


---

6. Consider the following graph.

a. Add one edge to the graph to make an Eulerian circuit possible.

b. Write down a possible Eulerian circuit.



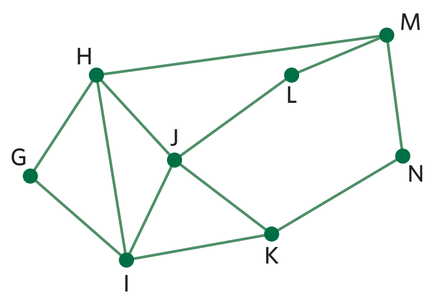
---

7. Jesse is skating at the local skatepark. The skatepark has eight ramps which are represented as vertices on the following graph.

a. Jesse is currently at ramp K and decides to follow an Eulerian trail. At which ramp will Jesse finish?

b. Write down a possible Eulerian trail that Jesse could take, starting from ramp K.

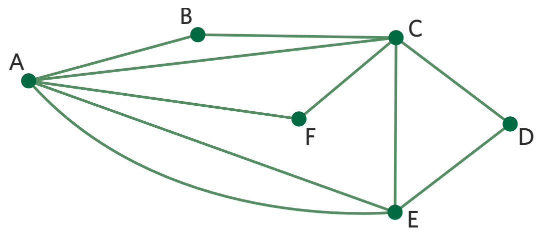
c. If Jesse wanted to start and end at the same ramp, but still skate along every path, which two ramps would need to be connected?



Back to contents

Identifying Hamiltonian paths and cycles

8. Consider the following graph.



Which of the following is **not** a Hamiltonian path?

A. F–A–B–C–E–D

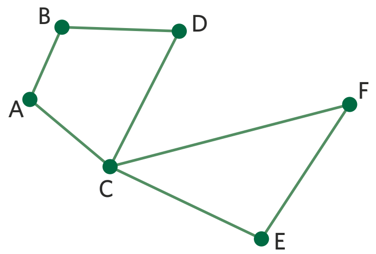
B. F–C–B–A–E–D

C. D–C–E–A–F–B

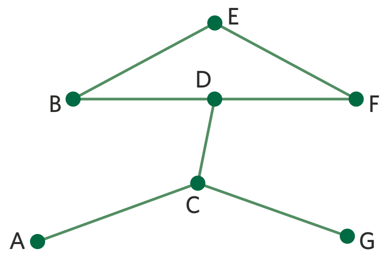
D. B–C–F–A–E–D

9. Determine if a Hamiltonian path is possible in each of the following graphs. If so, write one down.

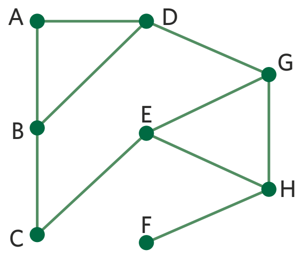
a.



b.

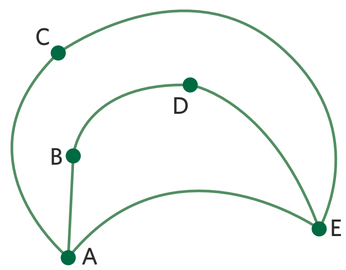


c.

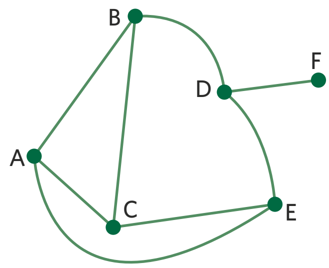


10. Determine if a Hamiltonian cycle is possible in each of the following graphs. If so, write one down.

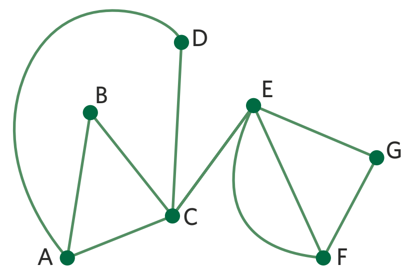
a.



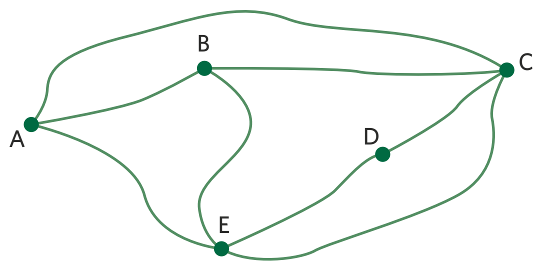
b.



c.



11. Five friends each live in different towns. The following graph shows the roads that link the five towns.



a. Natasha lives in town A and wants to visit each of her friends without travelling along the same road more than once. If she returns home after visiting her friends, what is the name of the walk she would have taken?

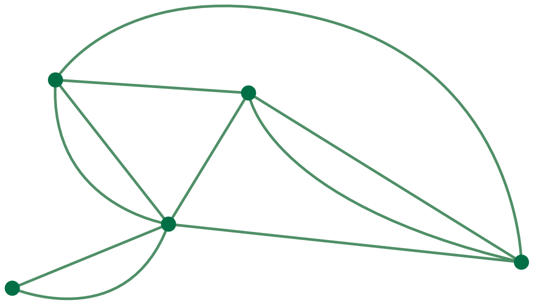
b. Write down a possible Hamiltonian path that Natasha could take in order to visit each of her friends.

Back to contents

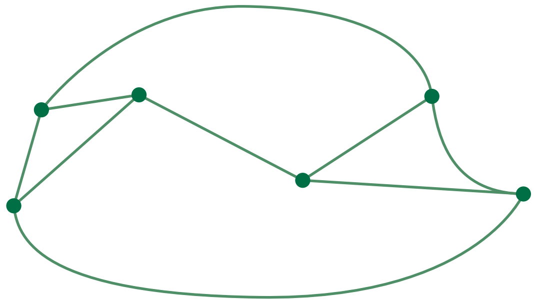
Joining it all together

12. Consider the following five graphs.

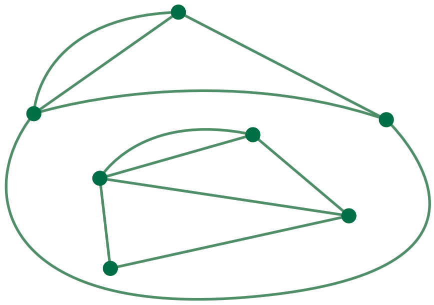
I.



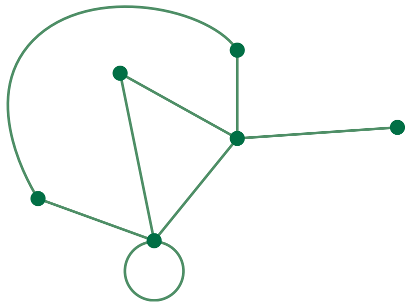
II.



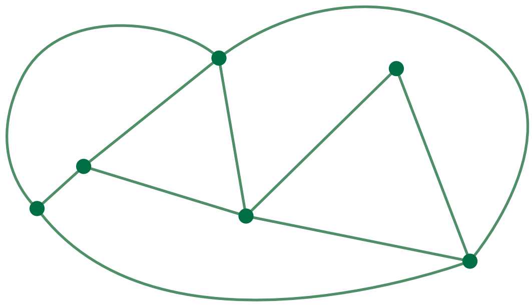
III.



IV.



V.



Which of the graphs

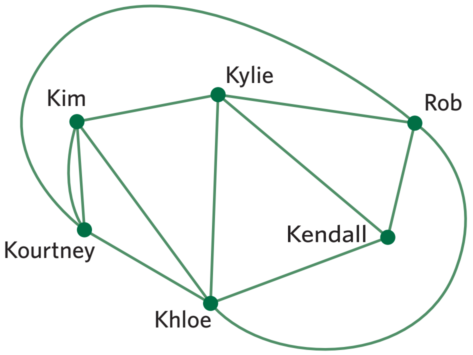
a. does not contain a Hamiltonian path?

b. contain a Hamiltonian cycle?

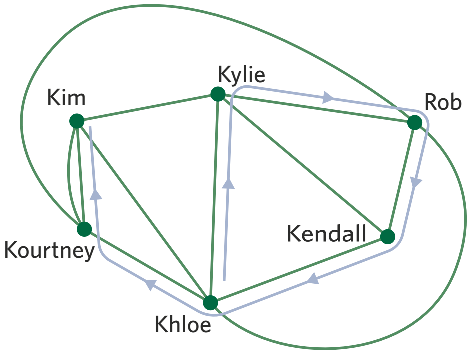
c. contain an Eulerian trail?

d. contain an Eulerian circuit?

13. The following graph displays the location of the Los Angeles houses of the six Kardashian siblings. The edges represent roads and the vertices represent houses where each of the siblings live.



a. Khloe is bored on a Sunday afternoon and decides to visit each of her siblings at their houses. She takes the route shown.



What type of walk did she take?

b. Kendall is back in Los Angeles after a long time in Europe and wants to drive around the neighbourhood. She decides to drive along every road once, starting from her house.

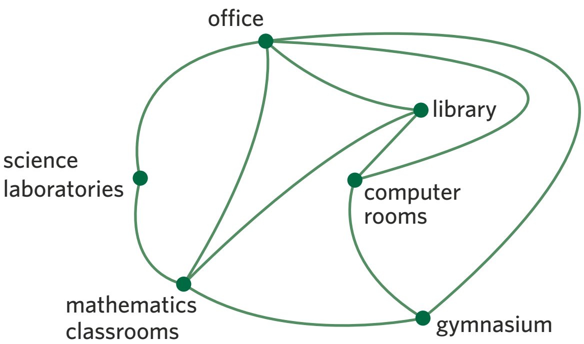
Which sibling's house will she end up at?

c. A new road is being built which will add an extra edge connecting Rob and Khloe's houses. Starting from her house, Kendall still wants to drive along every road once. Based on the new graph, which house will she now finish her trip at?

d. Kim decides she also wants to visit everyone. However, she decides to travel more efficiently than Kendall and does not want to travel along every road. She plans to visit each person once, starting and ending at her house. What type of walk is Kim planning on taking? Write down a possible route she could take.

Exam practice

14. Fencedale High School has six buildings. The network shows these buildings represented by vertices. The edges represent the paths between the buildings.



A school tour is to start and finish at the office, visiting each building only once.

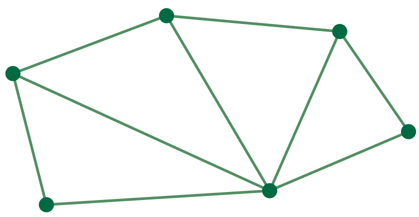
What is the mathematical term for this route? (1 MARK)

VCAA 2019 Exam 2 Networks and decision mathematics Q1bi

70% of students answered this question correctly.

---

15. Consider the following graph.



The minimum number of extra edges that are required so that an Eulerian circuit is possible in this graph is

A. 0

B. 1

C. 2

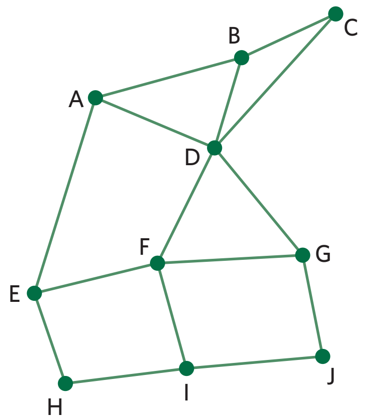
D. 3

E. 4

VCAA 2019 Exam 1 Networks and decision mathematics Q2

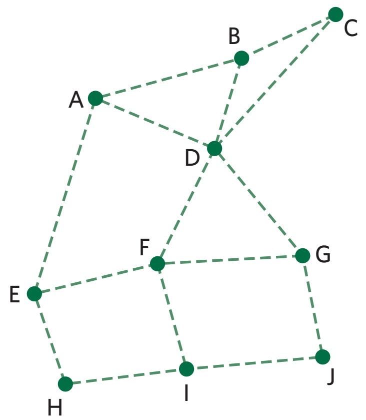
64% of students answered this question correctly.

16. In one area of the town of Zenith, a postal worker delivers mail to 10 houses labelled as vertices A to J on the following graph.



The postal worker has delivered the mail at F and will continue her deliveries by following a Hamiltonian path from F.

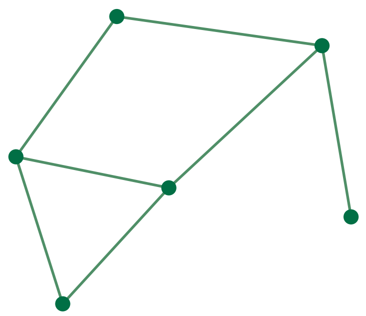
Draw in a possible Hamiltonian path for the postal worker on the following diagram. (1 MARK)



VCAA 2018 Exam 2 Networks and decision mathematics Q2c

62% of students answered this question correctly.

17.



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?

A. 1 **B.** 2 **C.** 3
D. 4 **E.** 5

VCAA 2021 Exam 1 Networks and decision mathematics Q5

17% of students answered this question correctly.

Back to contents

Questions from multiple lessons

Recursion and financial modelling

18. Which of the following recurrence relations could model the value of a perpetuity investment, P_n, after n months?

A. $P_0 = 340\,000, \quad P_{n+1} = 1.0031 \times P_n - 1024$

B. $P_0 = 245\,000, \quad P_{n+1} = 1.0056 \times P_n - 1382$

C. $P_0 = 375\,000, \quad P_{n+1} = 1.0018 \times P_n - 670$

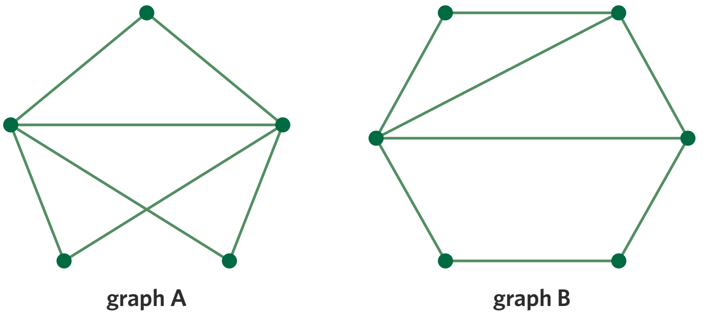
D. $P_0 = 310\,000, \quad P_{n+1} = 1.0046 \times P_n - 1426$

E. $P_0 = 285\,000, \quad P_{n+1} = 1.0064 \times P_n - 1844$

Adapted from VCAA 2018 Exam 1 Recursion and financial modelling Q21

Network and decision mathematics

19. Consider graphs A and B shown.



The sum of the degrees of the vertices of graph B is

A. two less than the sum of the degrees of the vertices of graph A.

B. one less than the sum of the degrees of the vertices of graph A.

C. equal to the sum of the degrees of the vertices of graph A.

D. one more than the sum of the degrees of the vertices of graph A.

E. two more than the sum of the degrees of the vertices of graph A.

Adapted from VCAA 2017 Exam 1 Networks and decision mathematics Q2

Recursion and financial modelling

20. Chloe has sold her investment property and has received \$865 200 after taxes.

She can invest this money in two ways.

Option 1

Chloe could put the \$865 200 into a perpetuity investment. She would then receive \$3287.76 per month for the rest of her life.

Option 2

Chloe could invest the \$865 200 in an annuity, instead of a perpetuity.

The annuity earns interest at 4.55% per annum, compounded fortnightly.

The balance of Chloe's annuity at the end of the first year would be \$850 160.19.

a. What is the interest rate, per annum, of the **Option 1** investment? (1 MARK)

b. For the **Option 2** investment,

i. what fortnightly payment, to the nearest dollar, would Chloe receive? (1 MARK)

ii. how much interest would Chloe's annuity earn in the second year of the investment? Round to the nearest cent. (2 MARKS)

Adapted from VCAA 2018 Exam 2 Recursion and financial modelling Q6

Back to contents

8D Minimum connector problems

STUDY DESIGN DOT POINTS

- trees and spanning trees
- minimum spanning trees in a weighted connected graph and their determination by inspection or by Prim's algorithm
- use of minimal spanning trees to solve minimal connector problems

8A 8B 8C 8D 8E 8F 8G 8H 8I 8J

KEY SKILLS

During this lesson, you will be:

- identifying and finding the weight of a spanning tree
- finding the minimum spanning tree.

KEY TERMS

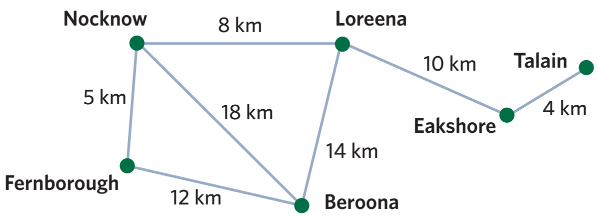
- Weighted graph
- Tree
- Spanning tree
- Minimum spanning tree
- Prim's algorithm

Graphs can provide additional information by including numeric values, known as weights, at each edge. A useful application of these weights is to minimise the total weight of a graph, while still keeping all vertices connected. For example, determining the minimum length of power lines required to connect different towns. This is known as a connector problem and can be solved using minimum spanning trees.

Identifying and finding the weight of a spanning tree

A **weighted graph** is a graph that has numeric values attached to each edge, which provides further information about the connections between the vertices. These weights represent physical quantities such as time, distance or cost.

For example, in the following graph the weights represent the lengths of train tracks between different stations.

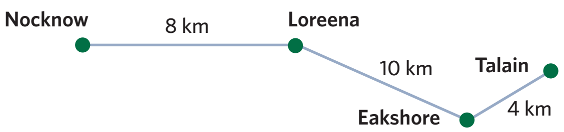


A **tree** is a type of graph that has no loops, duplicate edges or cycles. That is, it uses the least number of edges to connect the vertices.

The number of edges in a tree is always one less than the number of vertices, $e = v - 1$.

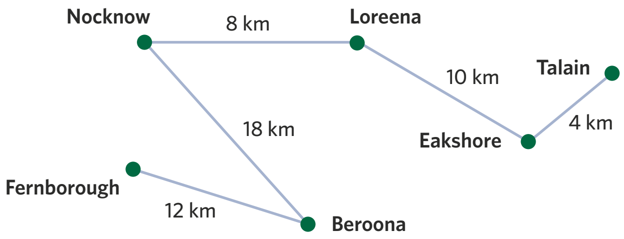
A tree can be a subgraph of a larger graph, so not all vertices in the larger graph need to be included.

For example, the following graph is a tree of the previous graph.



A **spanning tree** is a tree which connects all vertices in the original graph. There may be multiple spanning trees in a single graph.

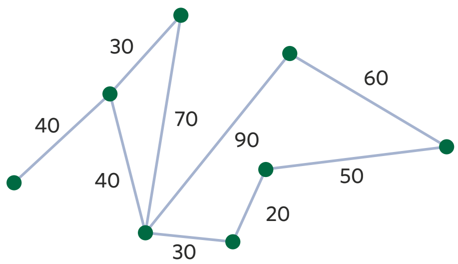
For example, the following graph is a spanning tree of the original graph.



The weight of a graph or tree is the sum of the weights of all its edges. For example, this spanning tree has a weight of 52 km.

Worked example 1

Consider the following graph.



a. Construct a spanning tree for the graph

Explanation

Step 1: Determine the number of vertices and edges required.

A spanning tree connects all vertices. Therefore the spanning tree will use all 8 vertices.

$e = v - 1$

$= 8 - 1$

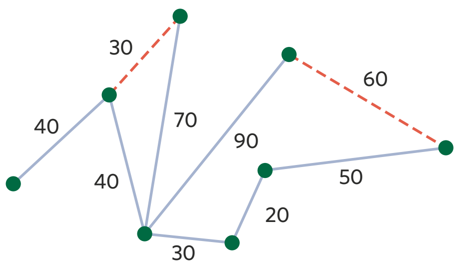
$= 7$

The spanning tree will have 7 edges.

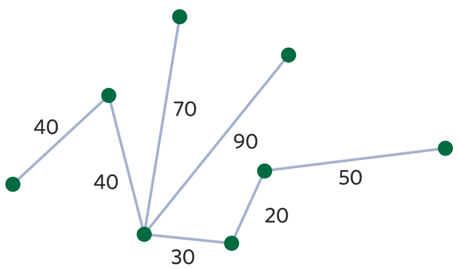
Step 2: Remove excess edges.

The original graph has 9 edges so 2 edges can be removed.

Ensure all vertices remain connected.



Answer



Note: There are other possible spanning trees.

Continues →

Back to contents

b. Calculate the weight of the spanning tree.

Explanation

Sum the weights of the edges of the spanning tree constructed in part **a**.

$weight = 40 + 40 + 70 + 90 + 30 + 20 + 50$

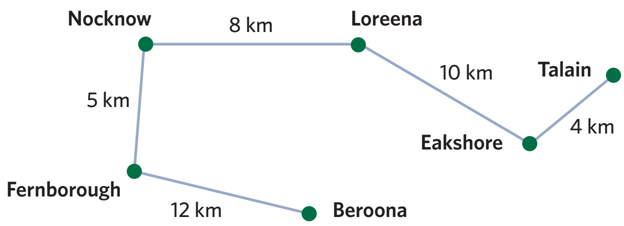
Answer

340

Finding the minimum spanning tree

The **minimum spanning tree** of a graph is the spanning tree with the lowest total weight. A key application of minimum spanning trees involves using them to solve connector problems. These are problems that involve finding the minimum distance, cost, weight or time to connect all vertices in a graph.

For example, the following graph is the minimum spanning tree of the original graph.



The minimum spanning tree has a weight of 39 km.

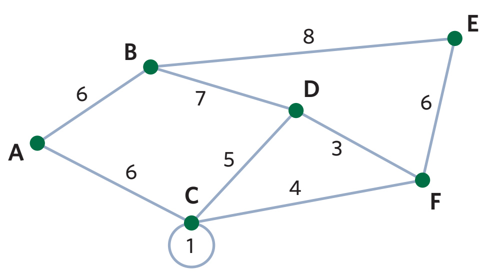
The minimum spanning tree of a graph can be found using a method known as **Prim's algorithm**, which has the following steps.

1. Select any vertex.
2. Select the edge with the lowest weight connected to that vertex.
3. There are now two connected vertices. Inspect all edges connected to either vertex and select the edge with the lowest weight that will connect a new vertex to the tree.
4. Ignore any edges which would create duplicate edges, cycles or loops.
5. Continue the process until all vertices are connected.

Note: If two edges have the same weight, it does not matter which is selected.

Worked example 2

For the following graph:



Continues →

Back to contents

a. Use Prim's algorithm to find the minimum spanning tree.

Explanation

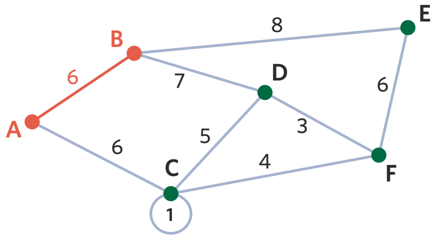
Step 1: Select any vertex.

For this example, start with vertex A.

Step 2: Select the edge with the lowest weight connected to that vertex.

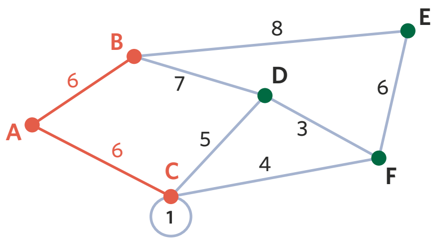
Both edges connected to vertex A are weighted 6 so it does not matter which is chosen.

For this example, choose the edge connected to vertex B.



Step 3: Inspect all edges connected to either of the two connected vertices. Select the edge with the lowest weight that will connect a new vertex to the tree.

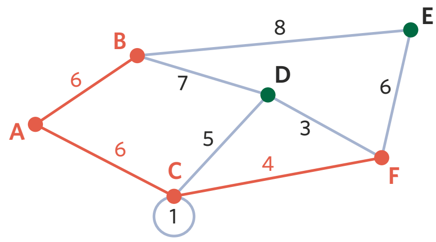
The edge connecting vertex A to C has the lowest weight.



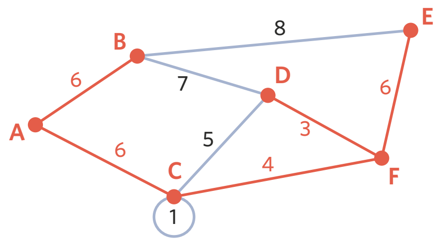
Step 4: Inspect all edges connected to the three connected vertices. Select the edge with the lowest weight that will connect a new vertex to the tree.

The loop at vertex C is the lowest weighted edge, however it should not be used as it does not connect a new vertex to the tree.

The edge connecting vertex C to F has the lowest weight.

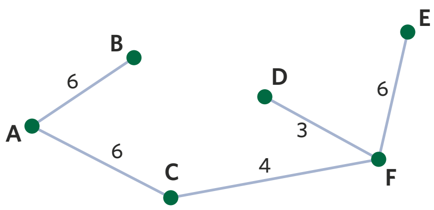


Step 5: Continue this process until all vertices are connected.



Step 6: Redraw the minimum spanning tree on its own.

Answer



b. Calculate the weight of the minimum spanning tree.

Explanation

Sum the weights of the edges of the minimum spanning tree.

$weight = 6 + 6 + 4 + 3 + 6$

Answer

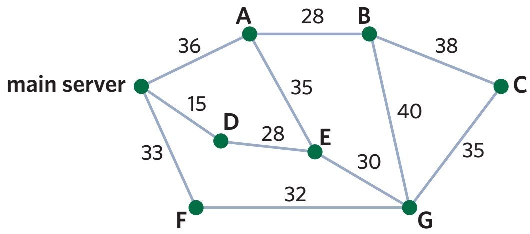
25

Exam question breakdown

VCAA 2019 Exam 1 Networks and decision mathematics Q5

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

Explanation

Step 1: Identify the connector problem.

The question requires the minimum length of cable needed to connect all points to be determined. This is the minimum spanning tree of the graph.

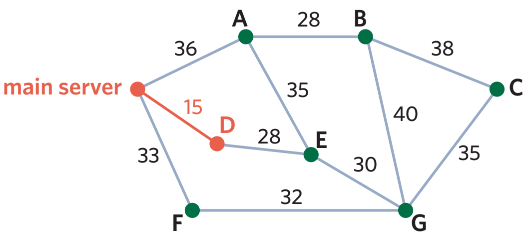
Prim's algorithm can be used to find the minimum spanning tree.

Step 2: Select any vertex.

Start with the main server.

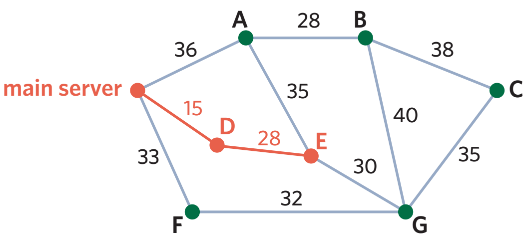
Step 3: Select the edge with the lowest weight connected to that vertex.

The edge connecting the main server to point D has the lowest weight.



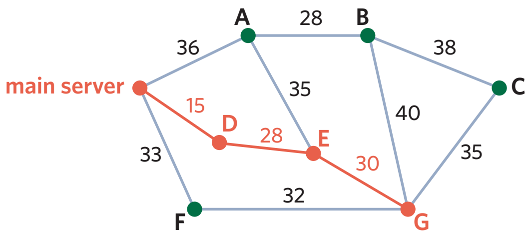
Step 4: Inspect all edges connected to either of the two connected vertices. Select the edge with the lowest weight that will connect a new vertex to the tree.

The edge connecting vertex D to E has the lowest weight.

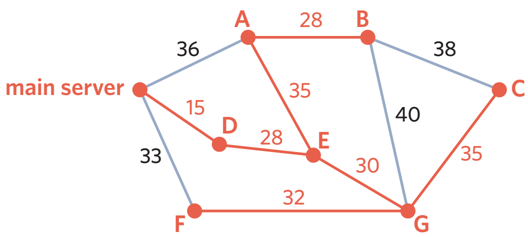


Step 5: Inspect all edges connected to the three connected vertices. Select the edge with the lowest weight that will connect a new vertex to the tree.

The edge connecting vertex E to G has the lowest weight.



Step 6: Continue this process until all vertices are connected.



Step 7: Calculate the weight of the minimum spanning tree.

Sum the weights of the edges.

$15 + 28 + 35 + 28 + 30 + 32 + 35 = 203$

Answer

B

75% of students answered this question correctly.

Students needed to be careful not to select any edges which would create duplicate edges, cycles, or loops. Edges needed to be selected that connected new vertices to the tree.

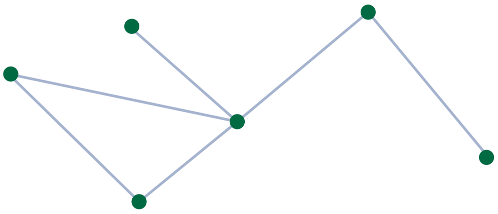
Back to contents

8D Questions

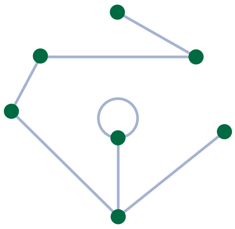
Identifying and finding the weight of a spanning tree

1. Which of the following graphs is a tree?

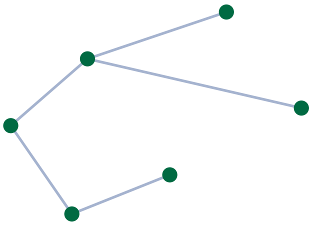
A.



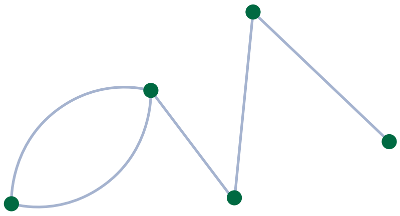
B.



C.

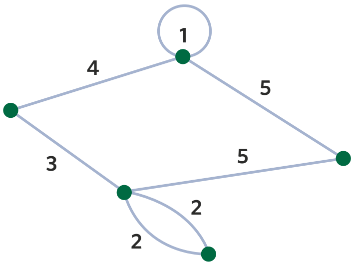


D.



---

2. Consider the following weighted graph.



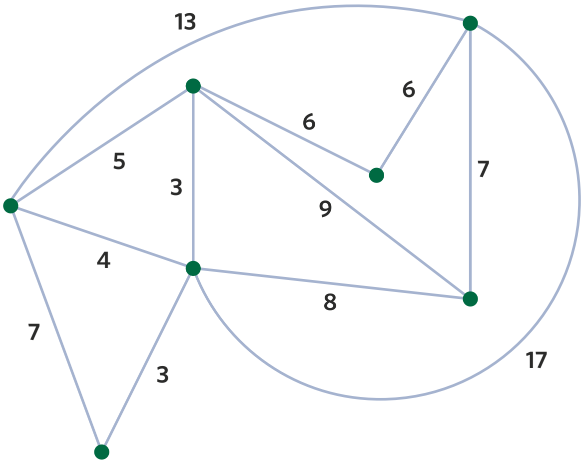
a. How many vertices and edges will a spanning tree of the weighted graph contain?

b. Find a possible spanning tree within the weighted graph.

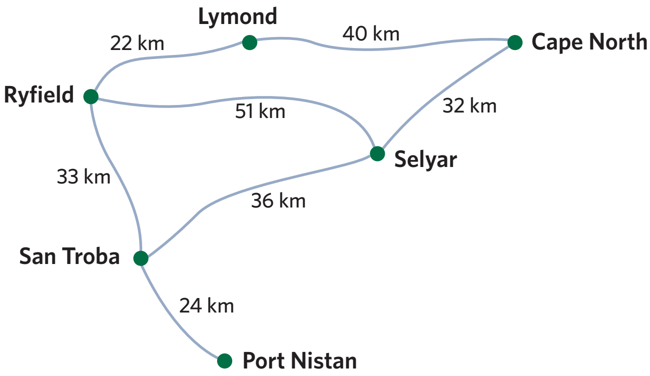
c. Calculate the weight of the spanning tree found in part **a**.

---

3. How many edges need to be removed from the following graph in order to create a spanning tree?



4. The following graph shows the roads connecting six towns.

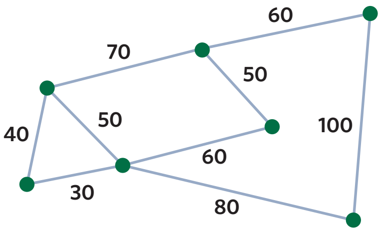


The roads require urgent maintenance and repairs. However, while roadworks are underway, all towns need to remain accessible.

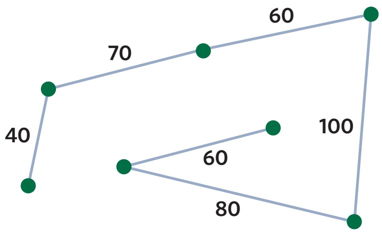
Find three possible spanning trees that ensure each town can still be accessed.

Finding the minimum spanning tree

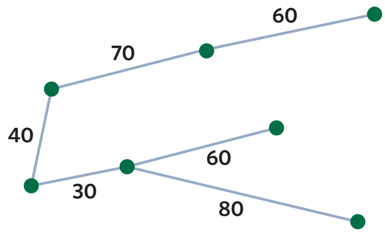
5. Which of the following graphs is the minimum spanning tree of the graph provided?



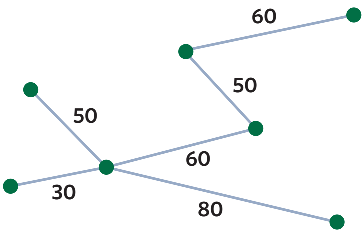
A.



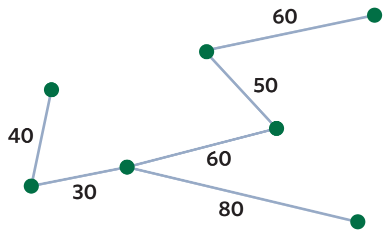
B.



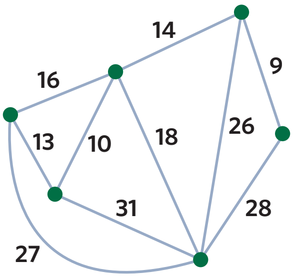
C.



D.



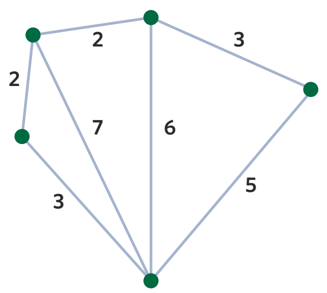
6. Draw the minimum spanning tree of the following weighted graph.



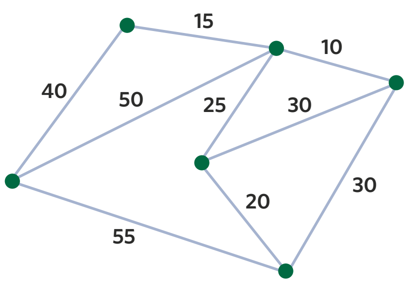
Back to contents

7. Calculate the weight of the minimum spanning tree for each of the following graphs.

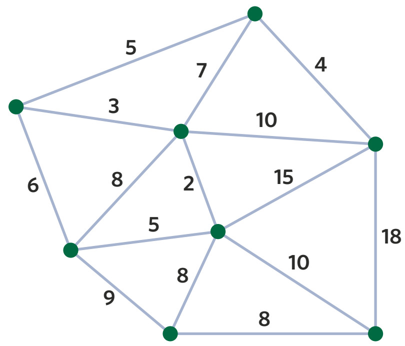
a.



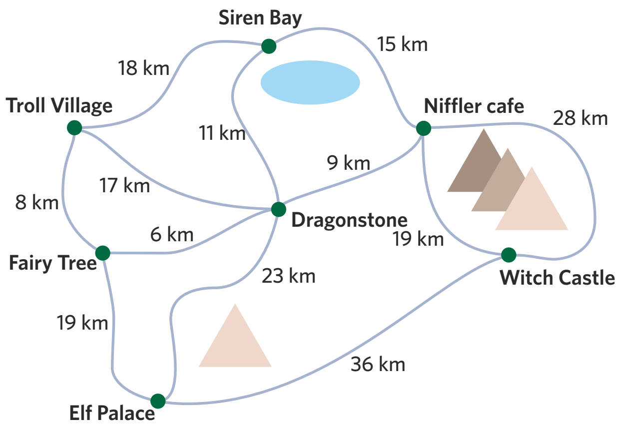
b.



c.



8. Since transportation dust has been discovered, Dragonride Services have been going out of business. In an effort to save some gold, they have decided to cut down on some of their flight paths.



a. Draw the flight map with the shortest possible flight distance that still accesses all of the landmarks.

b. What is the total distance of the flight map?

Joining it all together

9. The total weight of the minimum spanning tree of the following graph is 150.

a. How many edges will be in the minimum spanning tree?

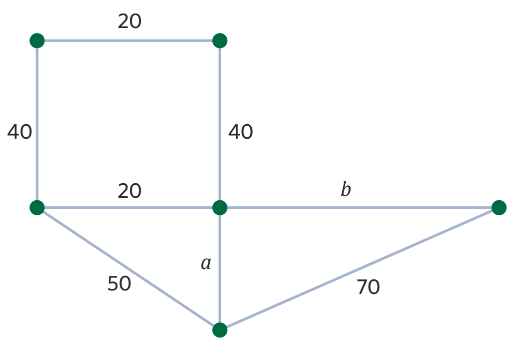
b. The values of a and b could be:

A. $a = 10, \quad b = 30$

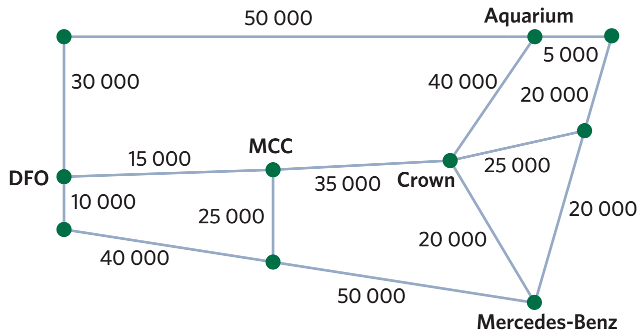
B. $a = 20, \quad b = 60$

C. $a = 10, \quad b = 60$

D. $a = 30, \quad b = 50$



10. Global leaders from every country are scheduled to meet at the Melbourne Convention Centre (MCC) for a conference on climate change. The government is worried about the security of the event and decides to block off as many roads surrounding the event as possible.



Only major landmarks (DFO, Mercedes-Benz, Crown, Aquarium, and the MCC) still need to be accessible.

The numbers represent the operational cost in dollars of keeping each road open.

a. Create an appropriate street map which will minimise the cost for the government.

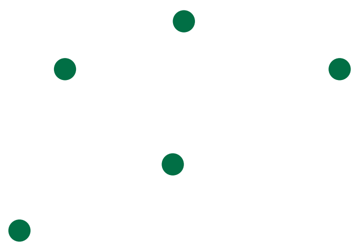
Due to security purposes, the roads connecting DFO to Mercedes-Benz ($10 000, $40 000 and $50 000) must stay open.

b. Draw a new appropriate street map which is still as cheap as possible, while including these streets.

c. Suppose the government decides that traffic congestion will increase by too much if only the landmarks are accessible. They decide that all points on the original map need to be accessible. What is the minimum cost required to keep the necessary roads open?

Exam practice

11. Consider the following graph with five isolated vertices.



To form a tree, the minimum number of edges that must be added to the graph is

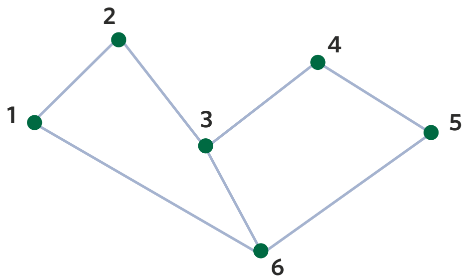
A. 1
B. 4
C. 5
D. 6
E. 10

VCAA 2018 Exam 1 Networks and decision mathematics Q1

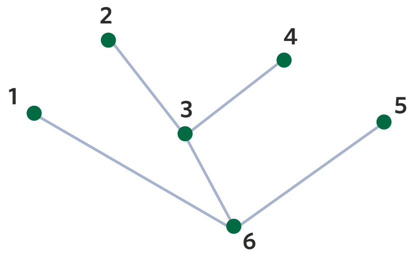
92% of students answered this question correctly.

Back to contents

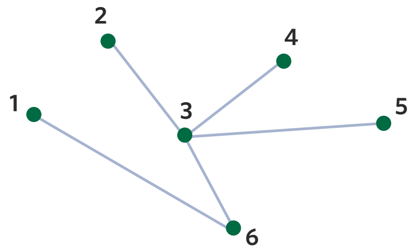
12. Which graph is **not** a spanning tree for the following network?



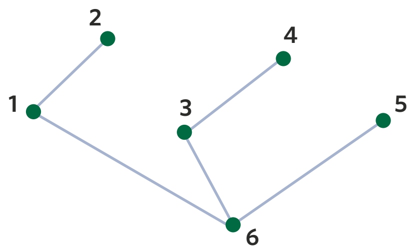
A.



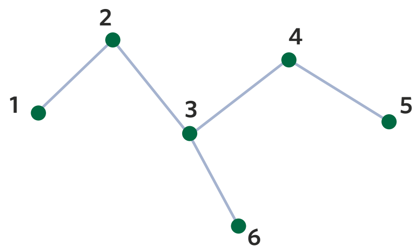
B.



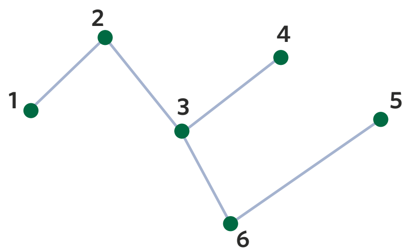
C.



D.



E.



VCAA 2020 Exam 1 Networks and decision mathematics Q3

81% of students answered this question correctly.

---

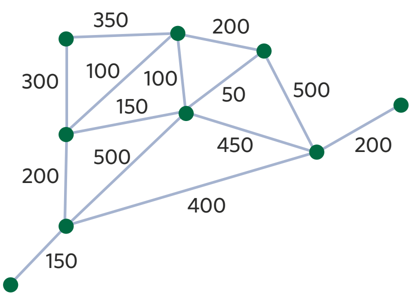
13. While on holiday, four friends visit a theme park where there are nine rides. On the following graph, the positions of the rides are indicated by the vertices.

The number on the edges represent the distances, in metres, between rides.

Electrical cables are required to power the rides.

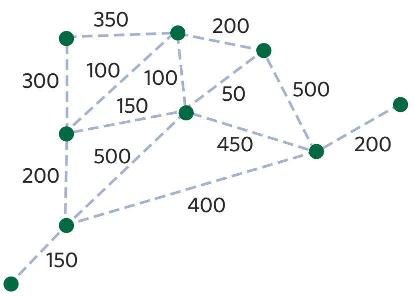
These cables will form a connected graph.

The shortest total length of cable will be used.



a. Give a mathematical term to describe a graph that represents these cables. (1 MARK)

b. Draw the graph that represents these cables using the following diagram (1 MARK)



VCAA 2017 Exam 2 Networks and decision mathematics Q3ai,ii

Part **a**: **63%** of students answered this question correctly.

Part **b**: **53%** of students answered this question correctly.

Back to contents

Questions from multiple lessons

Data analysis *Year 11 content*

14. The following table shows the average maximum daily *temperature* (°C) in Melbourne for each month in the first half of the year.

	Jan	Feb	Mar	Apr	May	Jun
***temperature* (°C)**	26	27	24	20	17	15

Using four-mean smoothing with centring, what is the smoothed value for April?

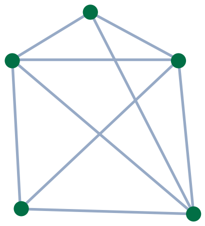
A. 19 °C **B.** 20.5 °C **C.** 22 °C **D.** 23.5 °C **E.** 24.25 °C

Adapted from VCAA 2018 Exam 1 Data analysis Q15

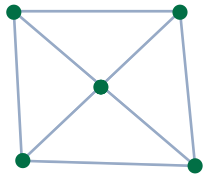
Networks and decision mathematics *Year 11 content*

15. Which one of the following graphs is **not** a planar graph?

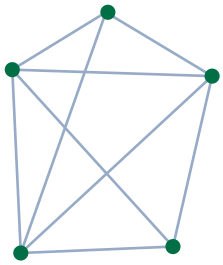
A.



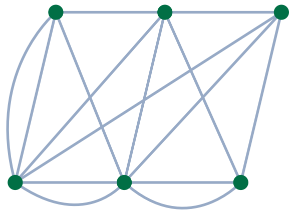
B.



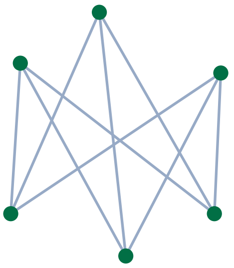
C.



D.



E.



Adapted from VCAA 2018 Exam 1 Networks and decision mathematics Q6

Data analysis *Year 11 content*

16. The equation of the least squares regression line relating *pond size* (m^2) and *number of fish* is:

number of fish = 2 + 0.85 × *pond size*

a. Interpret the slope of the regression line in terms of the variables *pond size* and *number of fish*. (1 MARK)

b. There are 23 fish in the pond when the pond has an area of 20 m^2. What is the residual value when the *pond size* is 20 m^2? (1 MARK)

c. The correlation coefficient is 0.73. What percentage of the variation in *number of fish* can be explained by the variation in *pond size*? Round your answer to two decimal places if required. (1 MARK)

Adapted from VCAA 2017 Exam 2 Data analysis Q3

Back to contents

8E Flow problems

STUDY DESIGN DOT POINTS

- 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

8A 8B 8C 8D 8E 8F 8G 8H 8I 8J

KEY SKILLS

During this lesson, you will be:

- navigating directed graphs
- calculating cut capacities
- determining the maximum flow.

KEY TERMS

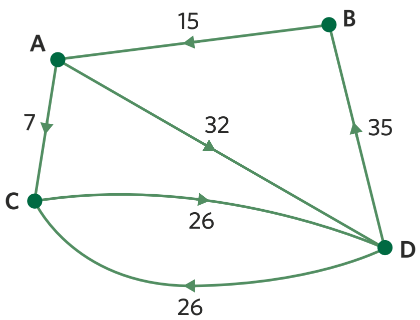
- Flow
- Source
- Sink
- Cut
- Cut capacity
- Maximum flow
- Minimum cut

Directed graphs can be used to solve flow problems, where the maximum capacity of a network, when moving from one side to another, needs to be calculated. A practical application of a flow problem is the movement of water through a system of pipes. The amount of water that can flow through the system is dictated by the capacity of the smallest pipe.

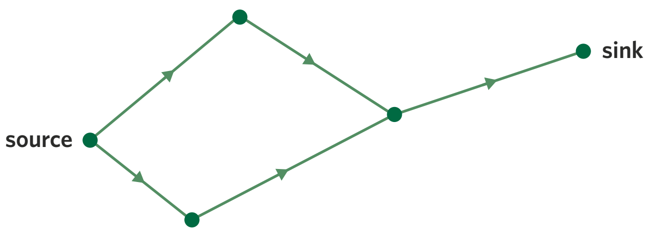
Navigating directed graphs

A directed graph is a network containing arrows on each edge that show the way in which one can travel between two vertices. Weights can be added to the edges to represent physical quantities such as time, distance or cost.

For example, the following directed graph shows the roads connecting four towns, where the weights represent the length, in km, of each road.



Flow occurs when networks only allow for movement in one direction, starting at one vertex (the **source**) and ending at another vertex (the **sink**).

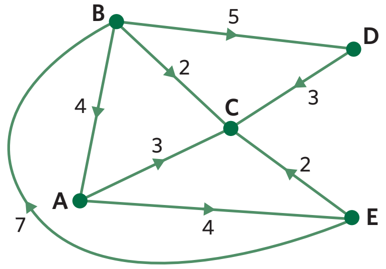


Back to contents

Worked example 1

For the following directed graphs, determine if it is possible to travel from vertex A to vertex D. If so, identify a route and calculate the weight of the route.

a.



Explanation

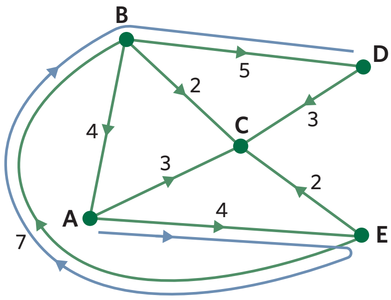
Step 1: Follow the edges starting from A.

From A, travel is possible either to C or E.

Travelling to C results in a dead end. Travelling to E is the only option.

Step 2: Continue this pattern until D is reached.

The route is A–E–B–D



Step 3: Calculate the weight of the route.

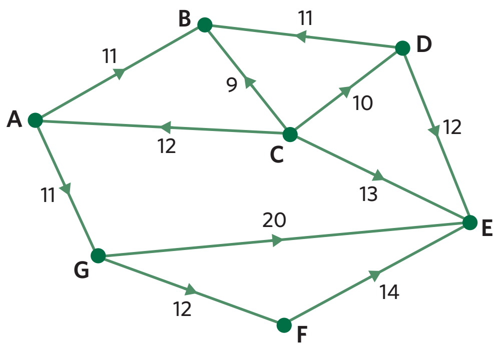
Sum the weights of the edges (A to E, E to B and B to D) in the route.

$weight = 4 + 7 + 5$

Answer

Yes, A–E–B–D, 16

b.



Explanation

From A, you can travel either to B or G.

Travelling to B results in a dead end. Travelling to G is the only option.

After travelling to G, you will end up at E, which is a dead end.

Answer

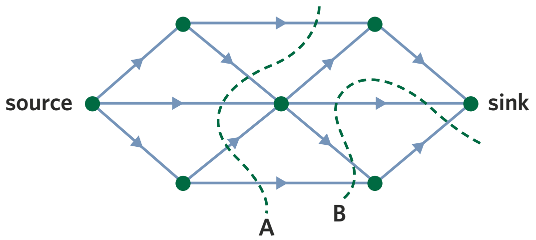
No

Back to contents

Calculating cut capacities

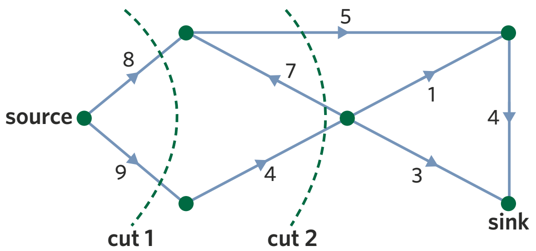
A **cut** is a line passing through a network that cuts the network in two. A valid cut must completely stop any network flow from the source to the sink.

For example, line A is a cut because it completely prevents any flow from the source to the sink. Line B is not a cut because it does not prevent all flow from source to sink.



The **cut capacity** is the total capacity, or weight, of all edges that a cut passes through. It is calculated by summing the weights of all edges that flow from source to sink included in the cut. If an edge flows from the sink side of the cut to the source side of the cut, it is not included in the cut capacity.

For example, consider the following network.



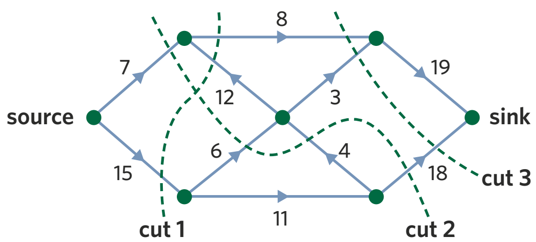
The capacity of cut 1 is $8 + 9 = 17$.

The capacity of cut 2 is $5 + 4 = 9$.

The edge with a weight of 7 is not included in the calculation for cut 2 as it flows from the sink side of the cut to the source side of the cut.

Worked example 2

Calculate the capacity of each cut in the following network.



Explanation

Step 1: Determine if any edges will not be counted.

If an edge flows from the sink side of the cut to the source side of the cut, it is not counted in the cut capacity. Hence, the edge with a weight of 12 in cut 1 will not be counted.

All edges in cuts 2 and 3 will be counted.

Step 2: Calculate the sum of the capacities of the edges.

Cut 1: $8 + 15 = 23$

Cut 2: $7 + 6 + 4 + 18 = 35$

Cut 3: $8 + 3 + 18 = 29$

Answer

Cut 1: 23

Cut 2: 35

Cut 3: 29

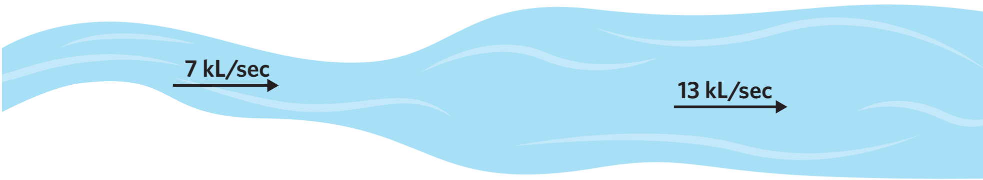
Back to contents

Determining the maximum flow

The **maximum flow** is equal to the minimum capacity through a network.

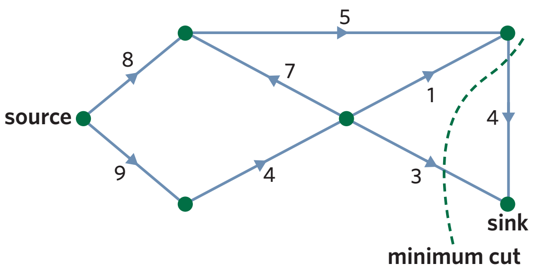
The maximum flow can best be explained using an example.

The start of a river system has the capacity to allow 7 kilolitres (kL) of water to flow through it each second. The river system then widens and the capacity of the next section is 13 kL per second. However, since the first section of river is smaller, only 7 kL will be allowed into the river system each second. Therefore, the full capacity of the second section will not be used. The maximum flow of the river system is 7 kL/sec.



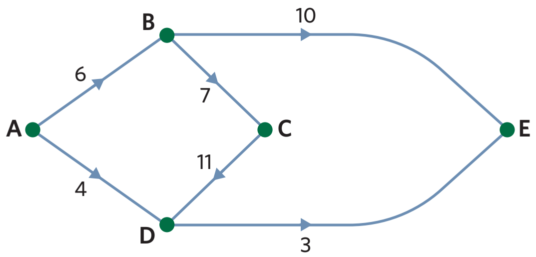
The **minimum cut** of a network is the cut with the smallest cut capacity. The capacity of the minimum cut is the maximum flow through the network.

For example, the minimum cut capacity of the following network is 7. Therefore, the maximum flow of the network is also 7.



Worked example 3

The following network shows the number of cars that can travel between five different towns each second.



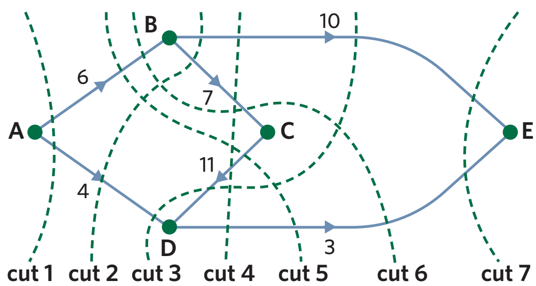
What is the maximum number of cars that can travel from town A to town E each second?

Explanation

Step 1: Identify all possible cuts through the network.

A cut must separate the source from the sink and completely stop the network flow.

There are seven possible cuts for this network.



Continues →

Back to contents

Step 2: Calculate the capacity of each cut.

The cut capacity is the sum of the capacities of all edges (from source to sink) that the cut passes through.

Cut 1: $6 + 4 = 10$

Cut 2: $10 + 7 + 4 = 21$

Cut 3: $10 + 11 + 4 = 25$

Cut 4: $10 + 7 + 3 = 20$

Note: In cut 4, the edge from vertex C to vertex D is not counted.

Cut 5: $6 + 3 = 9$

Note: In cut 5, the edge from vertex C to vertex D is not counted.

Cut 6: $6 + 3 = 9$

Note: In cut 6, the edge from vertex B to vertex C is not counted.

Cut 7: $10 + 3 = 13$

Step 3: Identify the minimum cut(s).

The minimum cuts are cut 4 and cut 5, with a capacity of 9 cars per second.

The maximum flow is equal to the capacity of the minimum cuts.

Answer

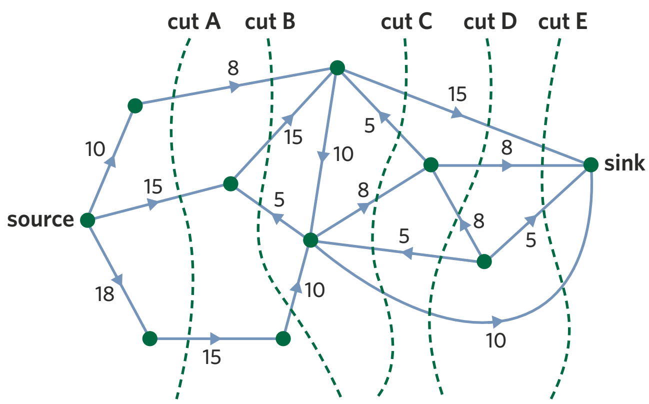
9 cars

Exam question breakdown

VCAA 2020 Exam 1 Networks and decision mathematics Q9

The flow of liquid through a series of pipelines, in litres per minute, is shown in the following directed network.

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



A. 1

B. 2

C. 3

D. 4

E. 5

Explanation

Step 1: Calculate the capacity of each cut.

The cut capacity is the sum of the capacities of all edges (from source to sink) that the cut passes through.

Cut A: $8 + 15 + 15 = 38$

Cut B: $8 + 15 + 10 = 33$

Cut C: $15 + 8 + 10 = 33$

Cut D: $15 + 8 + 10 = 33$

Cut E: $15 + 8 + 5 + 10 = 38$

Step 2: Determine the maximum flow.

The maximum flow is equal to the capacity of the minimum cut.

The capacity of the minimum cut is 33 L/min.

Step 3: Count the number of cuts with the minimum capacity.

Cuts B, C, and D all have a capacity of 33 L/min.

60% of students answered this question correctly.

18% of students incorrectly chose option B. This is likely because they included edges in their cut capacity calculations that were flowing in the wrong direction. It is important that only edges flowing from the source side of the cut to the sink side of the cut are included in the cut capacity.

Answer

C

Back to contents

8E Questions

Navigating directed graphs

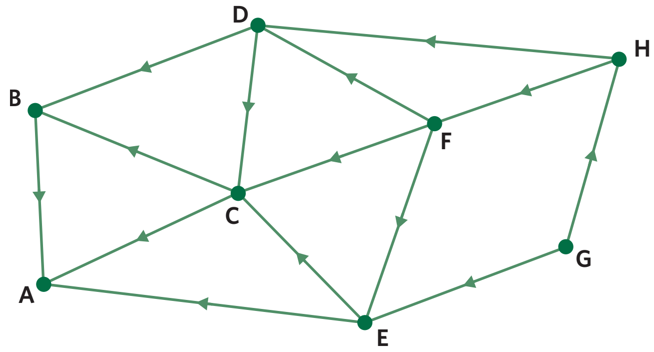
1. For the following directed graph, the source and sink, respectively, are

A. vertex A and vertex H

B. vertex G and vertex A

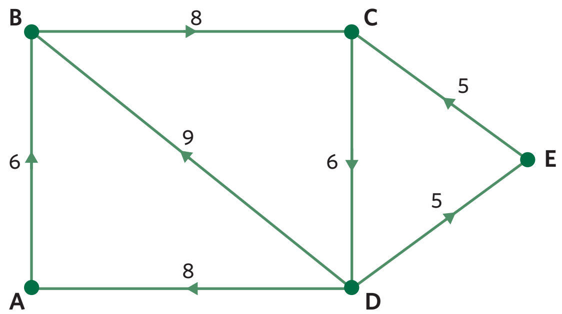
C. vertex H and vertex B

D. vertex H and vertex C

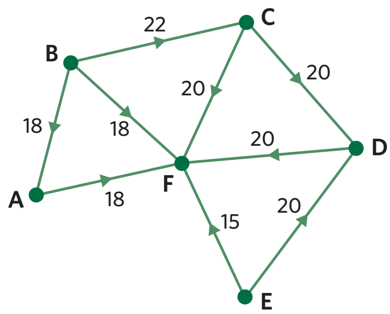


2. For each of the following directed graphs, is it possible to travel from vertex B to vertex E? If so, write down a route and calculate the weight of the route.

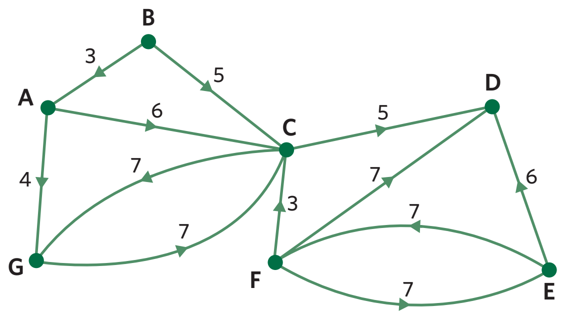
a.



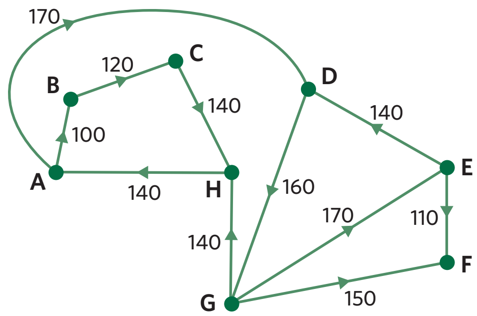
b.



c.



d.



3. The following network represents a number of one-way roads, with intersections represented as vertices. The numerical values above each edge represent the lengths of the roads, in km.

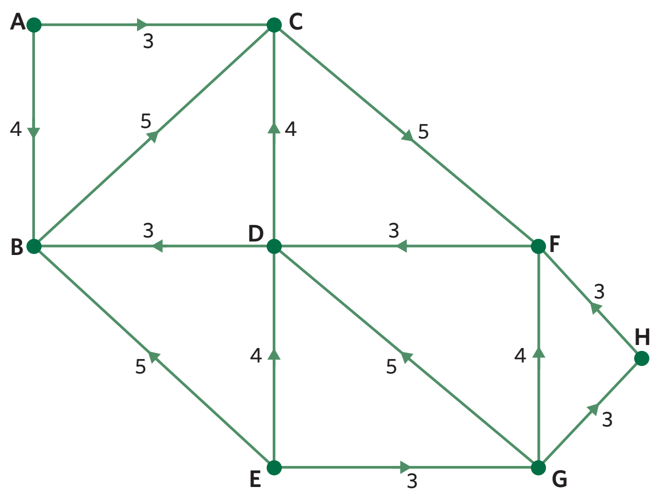
Which route is possible and has a weight of 13?

A. intersection A to intersection F.

B. intersection C to intersection G.

C. intersection G to intersection A.

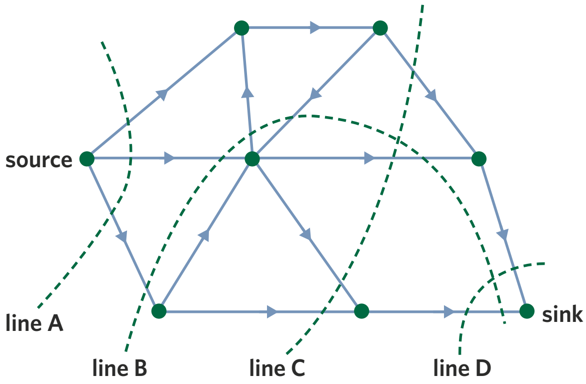
D. intersection B to intersection D.



Back to contents

Calculating cut capacities

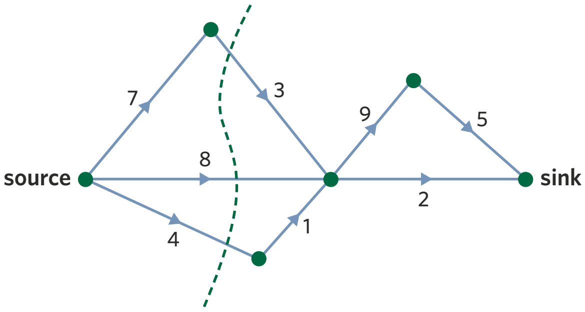
4. On the following network, which of the lines is **not** a cut?



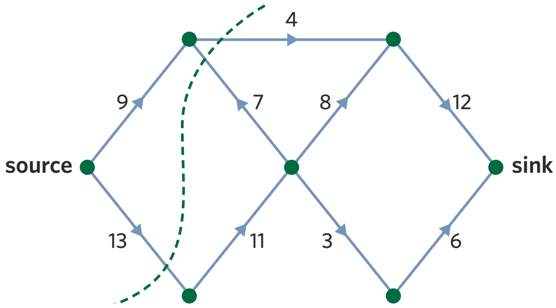
A. line A **B.** line B **C.** line C **D.** line D

5. Calculate the capacity of each of the following cuts.

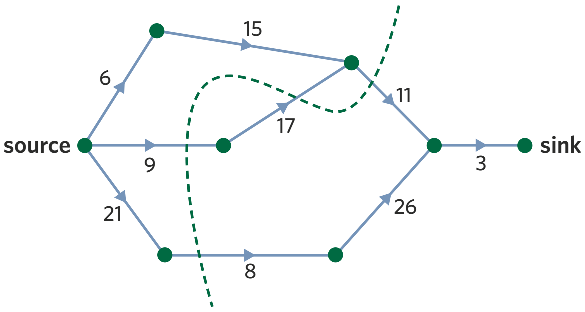
a.



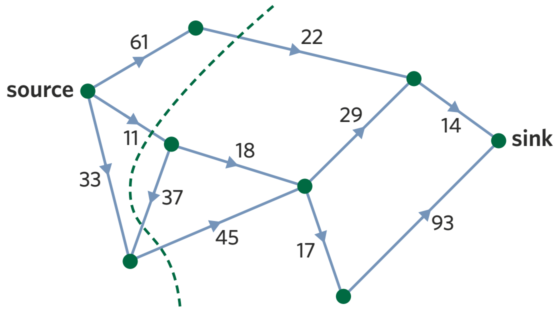
b.



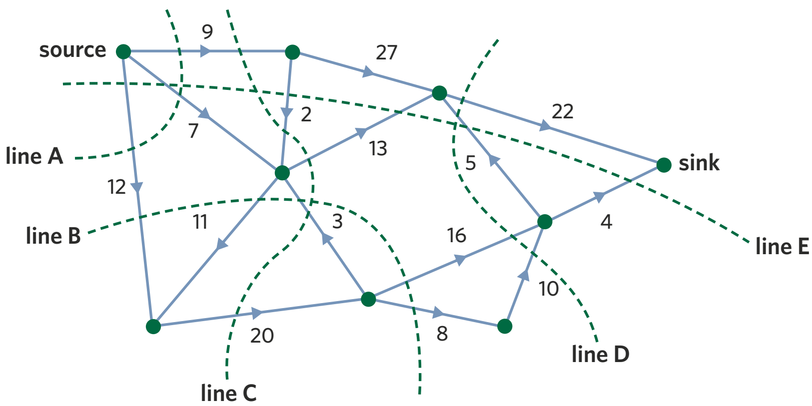
c.



d.



6. Consider the following network.



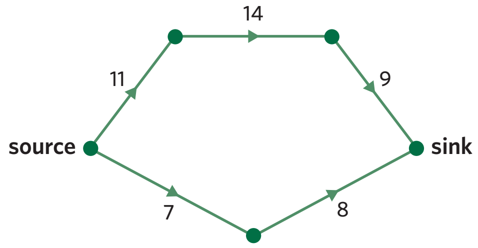
a. Which of the lines are valid cuts that stop flow from the source to the sink?

b. Determine the capacity of each cut identified in part **a**.

Back to contents

Determining the maximum flow

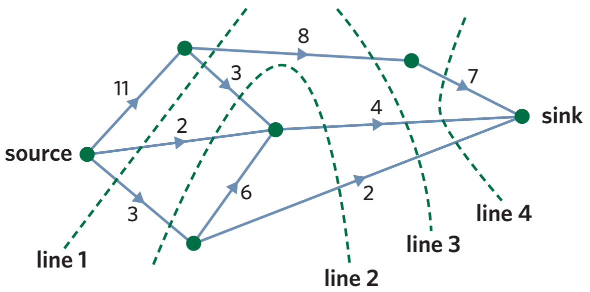
7. The maximum flow from source to sink through the following network is



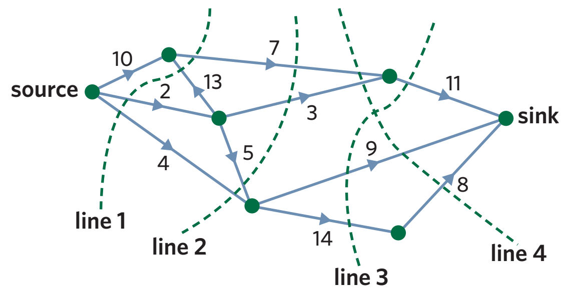
A. 16 **B.** 17 **C.** 18 **D.** 22

8. Four lines have been drawn on each of the following networks. One of these four lines is the minimum cut. For each network, determine which line is the minimum cut.

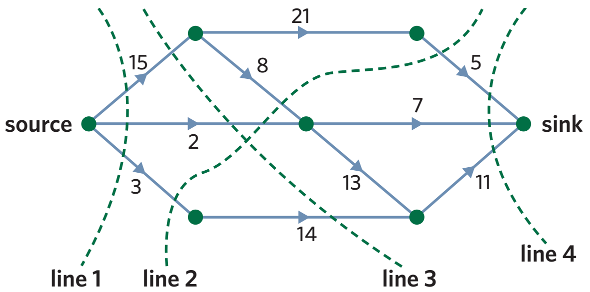
a.



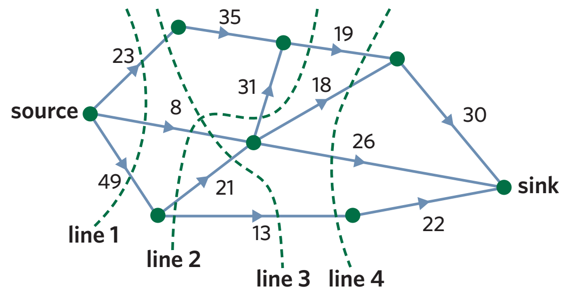
b.



c.

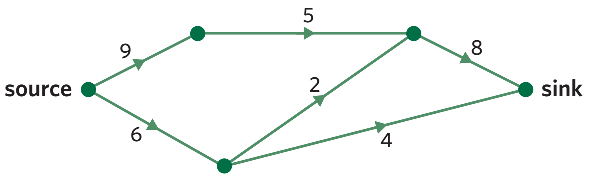


d.

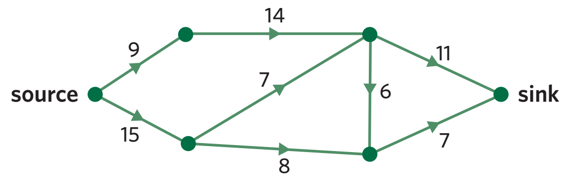


9. Find the maximum flow from source to sink for each of the following networks.

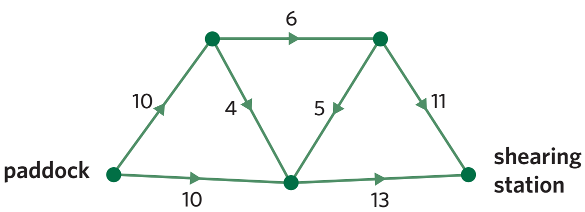
a.



b.



10. The following network displays the route from a paddock in which sheep are currently herded to a shearing station. The number of sheep per second that can travel between different gates on the route is shown.



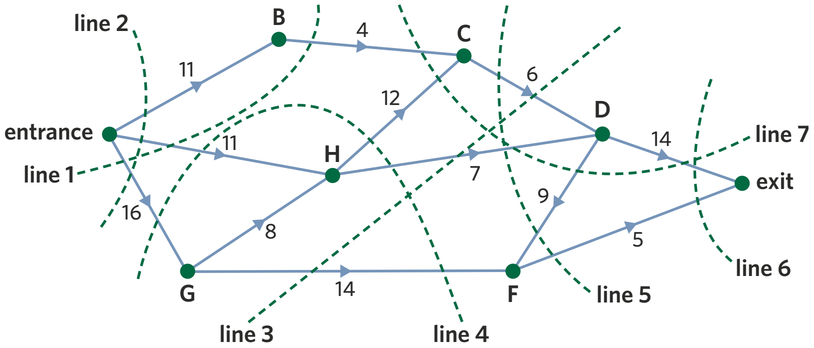
Find the maximum number of sheep that can travel each second from the paddock to the shearing station.

Back to contents

Joining it all together

11. The following network is a map of the paths between stalls and rides at the community fair. It shows the number of people that can travel along each of the paths every minute, as well as which direction they are able to walk.

There are seven lines on the map. One of these lines is the minimum cut.



a. It is possible to travel from

A. attraction D to attraction G.

B. attraction C to attraction G.

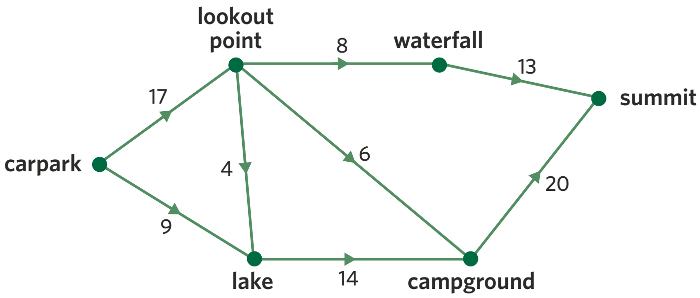
C. attraction F to attraction C.

D. attraction B to attraction F.

b. Which of the lines are not valid cuts that stop flow from the entrance to the exit?

c. What is the maximum number of people that can move from the entrance to the exit each minute?

12. The following network shows the paths that connect landmarks on a hiking trail, and the number of people per hour that can travel along each path.

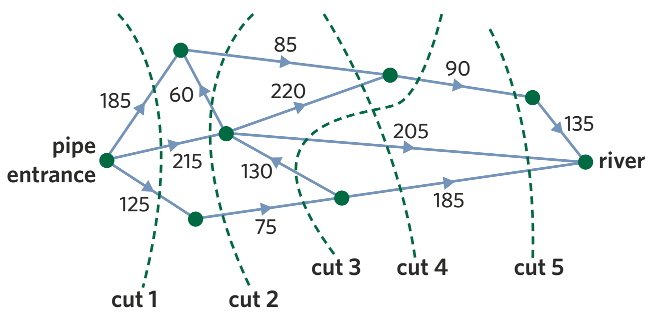


a. What is the maximum number of people that can travel from the carpark to the campground each hour?

b. What is the maximum number of people that can travel from the carpark to the summit each hour?

c. This week, park rangers are doing some work on the path from the waterfall to the summit, so now only 5 people can travel along this path each hour. How many less people can now travel from the carpark to the summit each hour?

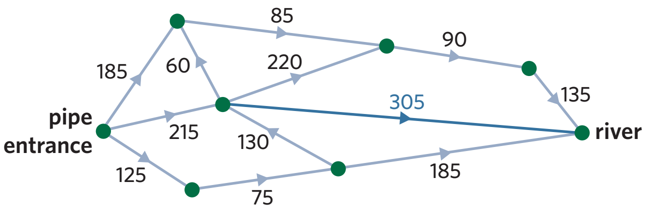
13. The following network shows the layout of a pipe system that transports water into a river. The edges represent pipes while the vertices are intersections of pipes. All values are in L/sec. There are five cuts on the network, one of them being the minimum cut.



a. What is the maximum amount of water that can be transported from the pipe entrance to the river each second?

Back to contents

In an effort to maximise transportation rate of the pipe system, the council increased the size of one pipe. They increased the capacity of the highlighted pipe from 205 L/sec of water to 305 L/sec.



b. What is now the maximum amount of water that can be transported from the pipe entrance to the river each second?

Exam practice

14. The following 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

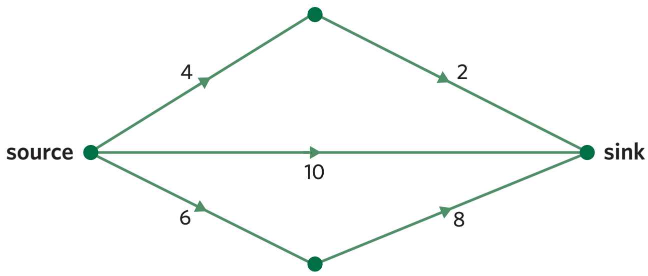
B. 14

C. 18

D. 20

E. 22

VCAA 2016 Exam 1 Networks and decision mathematics Q2



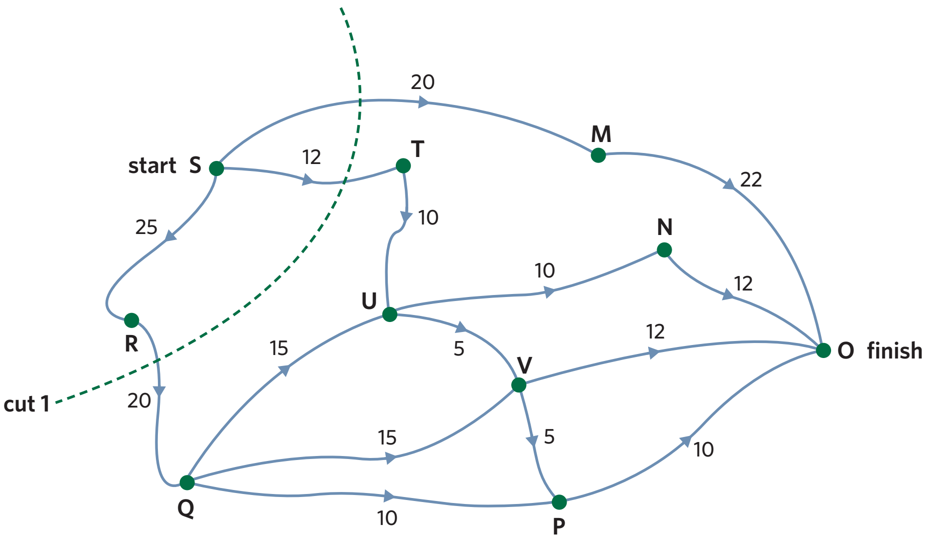
54% of students answered this question correctly.

---

15. A training program has a cricket team 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.



When considering the possible flow of people through this network, many different cuts can be made.

a. Determine the capacity of cut 1. (1 MARK)

b. What is the maximum flow from S to O, in number of people per minute? (1 MARK)

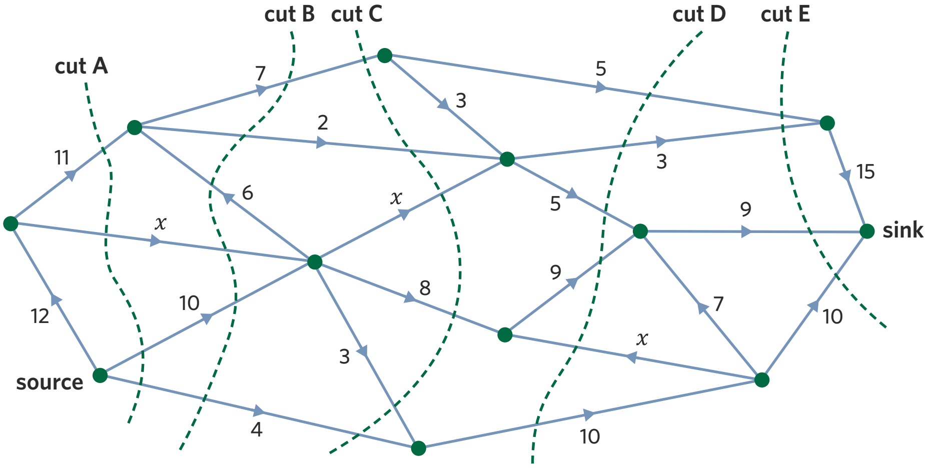
VCAA 2020 Exam 2 Networks and decision mathematics Q4b,c

Part **a**: **69%** of students answered this question correctly.

Part **b**: **32%** of students answered this question correctly.

Back to contents

16. The flow of oil through a series of pipelines, in litres per minute, is shown in the following 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 $x = 3$

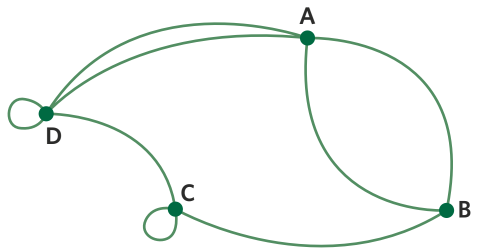
VCAA 2017 Exam 1 Networks and decision mathematics Q8

44% of students answered this question correctly.

Questions from multiple lessons

Networks and decision mathematics

17. Consider the following graph.



The adjacency matrix for this graph, with some elements missing, is shown.

$$\begin{array}{cc} \begin{matrix} A & B & C & D \end{matrix} & \\ \begin{bmatrix} 0 & & & \\ & 0 & & \\ & & 1 & \\ & & & 1 \end{bmatrix} & \begin{matrix} A \\ B \\ C \\ D \end{matrix} \end{array}$$

The adjacency matrix contains 16 elements when complete.

Of the 12 missing elements

A. four are ‘1’ and eight are ‘2’.

B. eight are ‘1’ and four are ‘2’.

C. two are ‘0’, six are ‘1’ and four are ‘2’.

D. six are ‘1’ and six are ‘2’.

E. four are ‘0’, four are ‘1’ and four are ‘2’.

Adapted from VCAA 2017 Exam 1 Networks and decision mathematics Q3

Back to contents

Recursion and financial modelling

18. Five lines of an amortisation table for a reducing balance loan with fortnightly repayments are shown.

repayment number	repayment	interest	principal reduction	balance of loan
17	$1050	$348.68	$701.32	$173 639.60
18	$1050	$347.28	$702.72	$172 936.88
19	$1050	$345.87	$704.13	$172 232.75
20	$1050	$310.02	$739.98	$171 492.77
21	$1050	$308.69	$741.31	$170 751.46

The interest rate for this loan changed at the start of the period that applies to repayment 20.

This change in interest rate is best described as

A. an increase of 0.52% per annum.

B. a decrease of 0.0002% per annum.

C. an increase of 0.02% per annum.

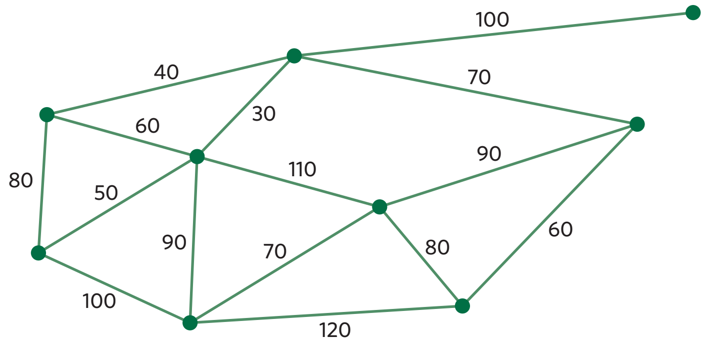
D. a decrease of 0.52% per annum.

E. a decrease of 0.02% per annum.

Adapted from VCAA 2018 Exam 1 Recursion and financial modelling Q23

Networks and decision mathematics *Year 11 content*

19. A village has nine homes. They are represented as vertices in the following graph. The edges represent possible power lines that could be built between the houses, with the numbers representing the distances, in metres, between them.



Power lines are required so that the houses can use electricity.

These power lines will create a connected graph.

The shortest length of power line will be used.

a. Give a mathematical term to describe a graph that represents these power lines. (1 MARK)

b. Draw the graph that represents these power lines. (1 MARK)

Adapted from VCAA 2017 Exam 2 Networks and decision mathematics Q3

Back to contents

8F Shortest path problems

STUDY DESIGN DOT POINTS

- 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

8A 8B 8C 8D 8E **8F** 8G 8H 8I 8J

KEY SKILLS

During this lesson, you will be:

- determining the shortest path by inspection
- applying Dijkstra's algorithm.

KEY TERMS

- Shortest path
- Dijkstra's algorithm

When looking at weighted graphs, it is important to be able to find the shortest path to maximise efficiency when travelling between vertices. For practical examples of weighted graphs, the shortest path can represent the cheapest option to get from point A to point B, or the shortest distance between points.

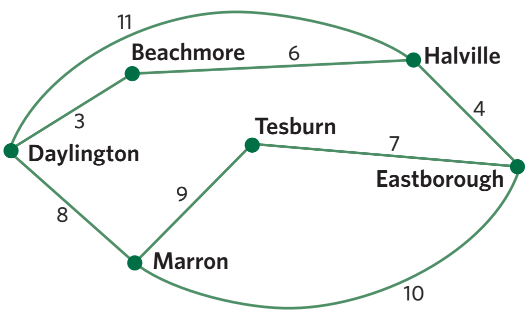
Determining the shortest path by inspection

The **shortest path** between two vertices on a weighted graph is the path with the minimum sum of weights. For simple weighted graphs, the shortest path can often be found by inspection.

Worked example 1

The weighted graph shown displays the distance, in kilometres, between six train stations.

Find the length and route of the shortest path from Daylington to Eastborough by eye.



Explanation

Step 1: List all the possible paths to get from Daylington (D) to Eastborough (E).

D–H–E, D–M–E, D–M–T–E, D–B–H–E

Step 2: Find the sum of the weights for each of these paths, and determine the minimum value.

D–H–E: $11 + 4 = 15$

D–M–E: $8 + 10 = 18$

D–M–T–E: $8 + 9 + 7 = 24$

D–B–H–E: $3 + 6 + 4 = 13$

Answer

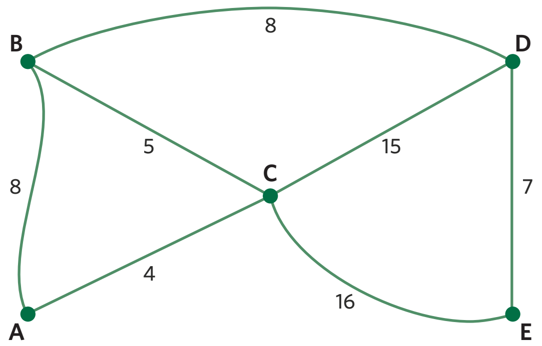
13 km, Daylington–Beachmore–Halville–Eastborough

Back to contents

Applying Dijkstra's algorithm

For more complex weighted graphs, determining the shortest path by inspection can be tedious. **Dijkstra's algorithm** can be used to determine the length and route of the shortest path between vertices on any weighted network.

The following steps detail the use of Dijkstra's algorithm to find the shortest path between vertex A and D in the given weighted network.



1. Construct a table.
   - The starting vertex should be the first row vertex.
   - Each of the other vertices should be listed as column vertices in any order.

	B	C	D	E
A				

2. Complete the first row.
   Write down the distance of direct connections from the starting vertex to each of the other vertices.
   - If there is no edge between the starting vertex and the column vertex, mark it with a cross.

	B	C	D	E
A	8	4	X	X

3. Identify the smallest number in this row.
   - If there are multiple equally small numbers, choose any one of them.
   - Put a square around the chosen number.
   - Copy the number into the next row, and put a square around it.
   - The column vertex of this number will become the new row vertex.

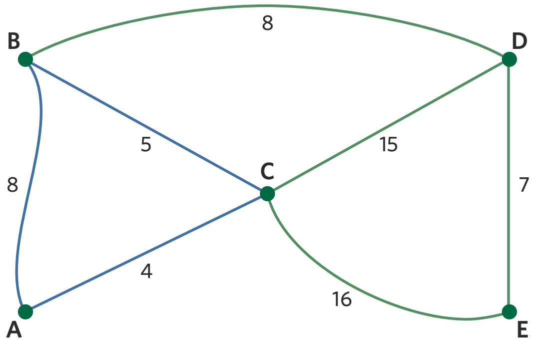
	B	C	D	E
A	8	$\boxed{4}$	X	X
C		$\boxed{4}$		

4. Complete the next row.
   For the rest of the columns, add the number that recently had a square put around it to the distance from the row vertex to the column vertex.
   - If this number is less than or equal to the number above it in the column, write it down.
   - If the number is greater than the number above it in the column, ignore it and copy the value from above.
   - If there is no connection from the starting vertex to the column vertex, via the row vertex, copy the number from above. If there is no number above, mark it with a cross.

Back to contents

	B	C	D	E
A	8	$\boxed{4}$	X	X
C	8	$\boxed{4}$	19	20

Although there is a connection from the starting vertex (A) to vertex B, via the row vertex (C), the direct connection from vertex A to B has a shorter length of 8.



5. Identify the smallest number in this row that is not already in a square.
   - If there are multiple equally small numbers, choose any one of them.
   - Put a square around the chosen number.
   - The column vertex of this number will become the new row vertex.
   - Copy down all numbers in squares into the next row.

	B	C	D	E
A	8	$\boxed{4}$	X	X
C	$\boxed{8}$	$\boxed{4}$	19	20
B	$\boxed{8}$	$\boxed{4}$		

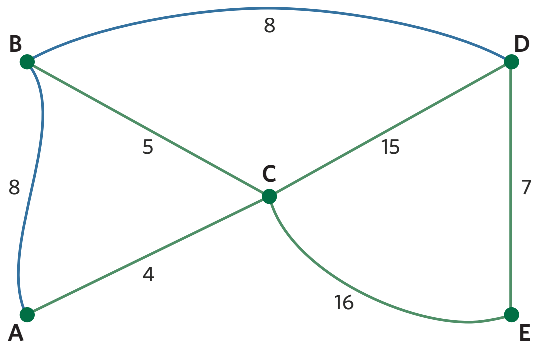
6. Complete the table by repeating steps 4 and 5.
   - Continue until the value in the destination vertex column has a square around it.

	B	C	D	E
A	8	$\boxed{4}$	X	X
C	$\boxed{8}$	$\boxed{4}$	19	20
B	$\boxed{8}$	$\boxed{4}$	$\boxed{16}$	20

7. Identify the length of the shortest path.
   The number in a square in the column of the destination vertex is the length of the shortest path.
   The length of the shortest path is 16.
8. Identify the shortest path.
   - Start from the number in the square in the destination vertex column.
   - Draw a line up the column from this final value until you reach the highest row with this same value. This value does not need to be in a square.
   - Locate the row vertex for the row you are now at and draw a horizontal line to this vertex's column.
   - Repeat until you reach the row of the starting vertex.
   - The horizontal lines show the route of the shortest path.

	B	C	D	E
A	8	$\boxed{4}$	X	X
C	$\boxed{8}$	$\boxed{4}$	19	20
B	$\boxed{8}$	$\boxed{4}$	$\boxed{16}$	20

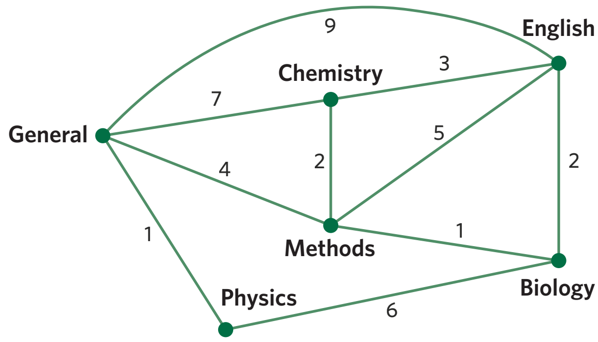
The shortest path is A–B–D.



Worked example 2

Reuben is struggling to find his way around his new school and is constantly running late. Luckily his friend in General Maths drew him a weighted graph of the school, representing the time it takes, in minutes, to walk to each class.

Find the time and route of the shortest path from General (G) to English (E).



Explanation

Step 1: Construct the table.

For efficiency purposes, label each vertex as their initial.

The first row is the starting vertex (G).

	P	B	M	C	E
G					

Step 2: Complete the first row.

There are edges directly connecting vertex G to vertices P, M, C and E. Write the weights of these edges in the table.

There is no edge directly connecting vertex G to vertex B. Mark column B with a cross.

	P	B	M	C	E
G	1	X	4	7	9

Step 3: Identify the smallest number in this row.

1 is the smallest number in the row.

Put a square around the '1' and copy it to the next row.

The '1' is located in column P, so vertex P will become the next row vertex.

	P	B	M	C	E
G	$\boxed{1}$	X	4	7	9
P	$\boxed{1}$				

Step 4: Complete the second row.

For the rest of the columns, add the number that recently had a square put around it (1) to the distance from the row vertex (P) to each column vertex.

Vertex B: $1 + 6 = 7$

There is no number above, so write it down.

Vertex P is not directly connected to vertices M, C and E. As there is a number above each, copy them down.

	P	B	M	C	E
G	$\boxed{1}$	X	4	7	9
P	$\boxed{1}$	7	4	7	9

Continues →

Back to contents

Step 5: Identify the smallest number in this row that is not already in a square.

4 is the smallest number in the row that is not already in a square.

Put a square around the '4' and copy it to the next row.

The '4' is located in column M, so vertex M will become the next row vertex.

	P	B	M	C	E
G	$\boxed{1}$	X	4	7	9
P	$\boxed{1}$	7	$\boxed{4}$	7	9
M	$\boxed{1}$		$\boxed{4}$		

Step 6: Complete the table by repeating steps 4 and 5.

Continue until the value in the destination vertex column has a square around it.

	P	B	M	C	E
G	$\boxed{1}$	X	4	7	9
P	$\boxed{1}$	7	$\boxed{4}$	7	9
M	$\boxed{1}$	$\boxed{5}$	$\boxed{4}$	6	9
B	$\boxed{1}$	$\boxed{5}$	$\boxed{4}$	$\boxed{6}$	$\boxed{7}$
C	$\boxed{1}$	$\boxed{5}$	$\boxed{4}$	$\boxed{6}$	$\boxed{7}$

Step 7: Identify the length of the shortest path.

The number in a square in column E is the length of the shortest path.

The length of the shortest path is 7.

Step 8: Identify the route of the shortest path.

Start from the square in column E.

Draw a line up the column from this final value until you reach the highest row with this same value. This is row B.

Draw a horizontal line to this vertex's column (B).

Repeat until you reach the row of the starting vertex.

The horizontal lines show the route of the shortest path.

	P	B	M	C	E
G	$\boxed{1}$	X	4	7	9
P	$\boxed{1}$	7	$\boxed{4}$	7	9
M	$\boxed{1}$	$\boxed{5}$	$\boxed{4}$	6	9
B	$\boxed{1}$	$\boxed{5}$	$\boxed{4}$	$\boxed{6}$	$\boxed{7}$
C	$\boxed{1}$	$\boxed{5}$	$\boxed{4}$	$\boxed{6}$	$\boxed{7}$

The shortest path is G–M–B–E.

Answer

7 minutes, General–Methods–Biology–English

Exam question breakdown

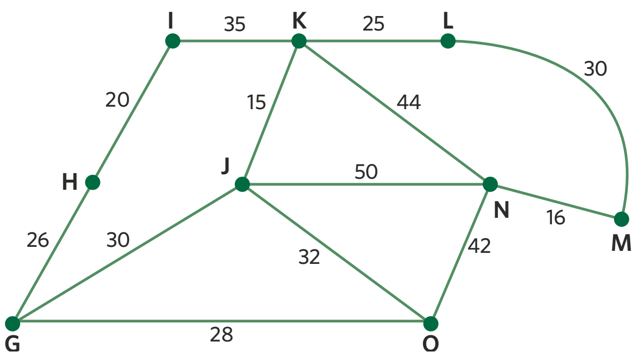
VCAA 2021 Exam 2 Networks and decision mathematics Q2a

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, I, J, K, L, M, N and O represent towns.

The edges represent the main roads. The numbers on the edges indicate the distances, in kilometres, between adjacent towns.



What is the shortest distance, in kilometres, between town G and town M? (1 MARK)

Continues →

Back to contents

Explanation

Step 1: Apply Dijkstra's algorithm to the weighted network.

Town G is the starting vertex and town M is the destination vertex.

	H	I	J	K	L	M	N	O
G	$\boxed{26}$	X	30	X	X	X	X	28
H	$\boxed{26}$	46	30	X	X	X	X	$\boxed{28}$
O	$\boxed{26}$	46	$\boxed{30}$	X	X	X	70	$\boxed{28}$
J	$\boxed{26}$	46	$\boxed{30}$	$\boxed{45}$	X	X	70	$\boxed{28}$
K	$\boxed{26}$	$\boxed{46}$	$\boxed{30}$	$\boxed{45}$	70	X	70	$\boxed{28}$
I	$\boxed{26}$	$\boxed{46}$	$\boxed{30}$	$\boxed{45}$	70	X	$\boxed{70}$	$\boxed{28}$
N	$\boxed{26}$	$\boxed{46}$	$\boxed{30}$	$\boxed{45}$	$\boxed{70}$	86	$\boxed{70}$	$\boxed{28}$
L	$\boxed{26}$	$\boxed{46}$	$\boxed{30}$	$\boxed{45}$	$\boxed{70}$	$\boxed{86}$	$\boxed{70}$	$\boxed{28}$

Step 2: Identify the length of the shortest path.

The number in a square in column M is the length of the shortest path.

The shortest path from town G to town M is 86 km.

Answer

86 km

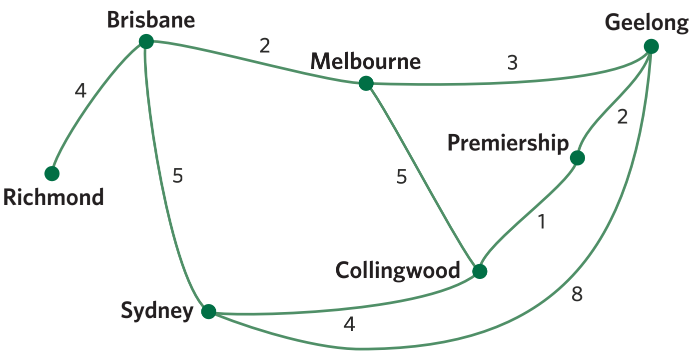
85% of students answered this question correctly.

Some students correctly found the route but did not find the distance of the path, as asked for in the question.

8F Questions

Determining the shortest path by inspection

1. Seven fictitious towns, connected by roads, are represented in the following weighted network. The number on each edge is the distance, in kilometres, of the road.



Which option outlines the shortest path between the towns Richmond and Premiership?

A. Richmond–Brisbane–Melbourne–Geelong–Premiership

B. Richmond–Brisbane–Sydney–Collingwood–Premiership

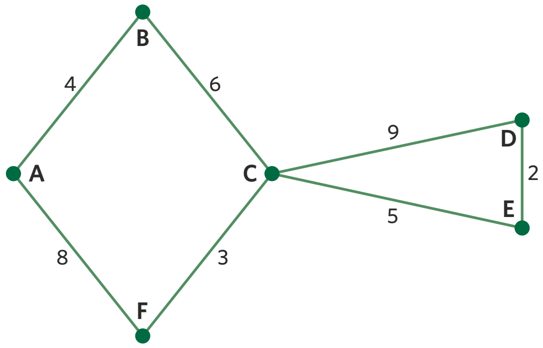
C. Richmond–Brisbane–Melbourne–Collingwood–Premiership

D. Richmond–Brisbane–Sydney–Geelong–Premiership

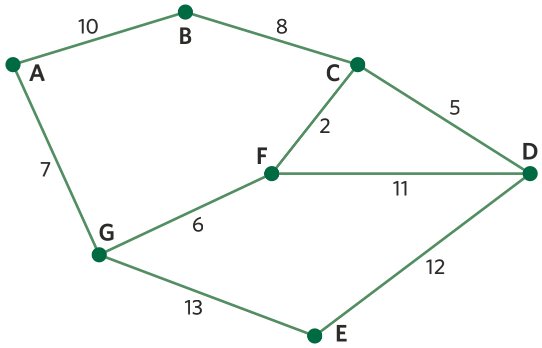
Back to contents

2. Find, by inspection, the length and route of the shortest path from vertex A to vertex D for each of the following weighted networks.

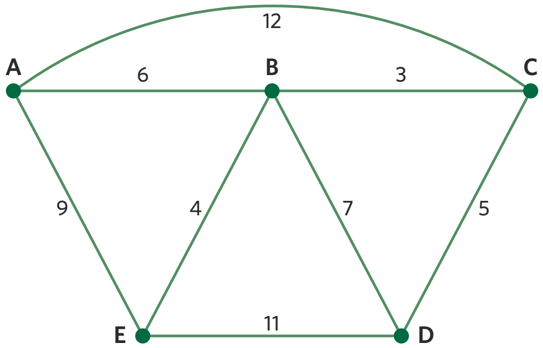
a.



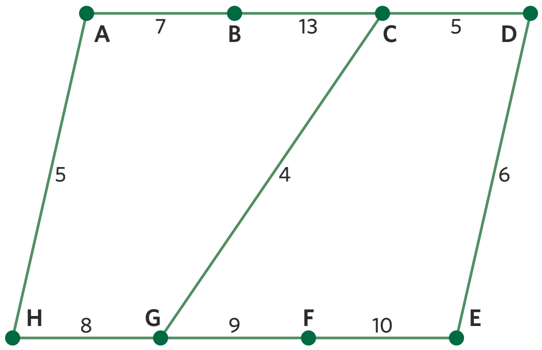
b.



c.



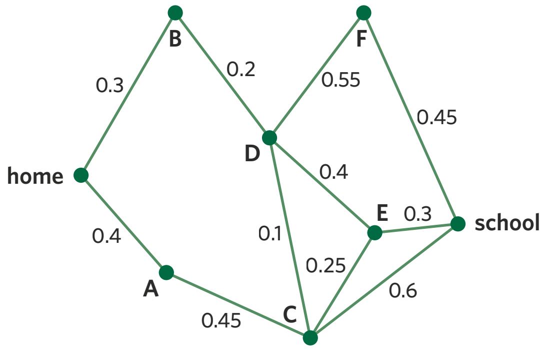
d.



3. Fiona is preparing for her first day at her new school, and is planning on walking to school. The given weighted network shows the distance, in kilometres, of the roads between her house and her school. Intersections are labelled A–F.

a. Fiona decides to walk to school via the route home–A–C–D–E–school. What distance will she walk?

b. Find, by inspection, the shortest distance that Fiona can walk to school.



Applying Dijkstra's algorithm

4. The following table is a complete Dijkstra's algorithm solution to find the shortest path through a particular weighted network.

	B	C	D	E	F
A	5	$\boxed{4}$	X	X	X
C	$\boxed{5}$	$\boxed{4}$	7	10	12
B	$\boxed{5}$	$\boxed{4}$	$\boxed{7}$	10	12
D	$\boxed{5}$	$\boxed{4}$	$\boxed{7}$	$\boxed{10}$	11
E	$\boxed{5}$	$\boxed{4}$	$\boxed{7}$	$\boxed{10}$	$\boxed{11}$

a. What is the length of the shortest path from A to D?

A. 4 **B.** 7 **C.** 11 **D.** 14

b. What is the length of the shortest path from A to F?

c. What is the route of the shortest path from A to F?

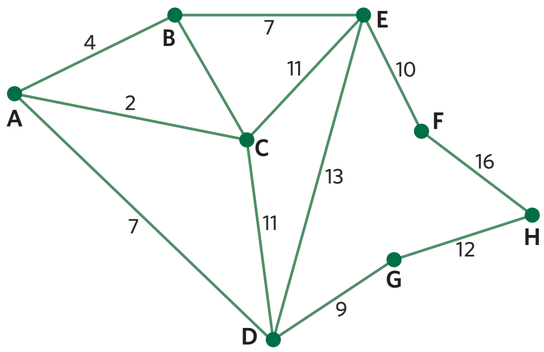
Back to contents

5. Consider the table containing a complete Dijkstra's algorithm solution.

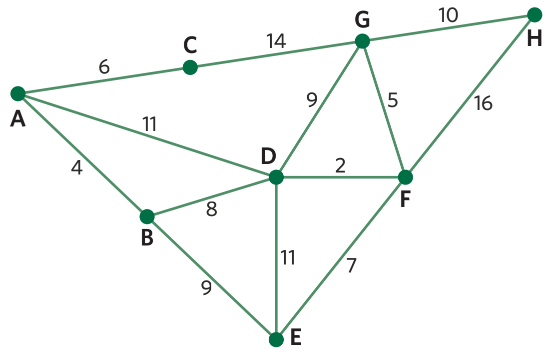
	B	C	D	E	F	G	H
A	$\boxed{4}$	6	11	X	X	X	X
B	$\boxed{4}$	$\boxed{6}$	11	13	X	X	X
C	$\boxed{4}$	$\boxed{6}$	$\boxed{11}$	13	X	20	X
D	$\boxed{4}$	$\boxed{6}$	$\boxed{11}$	$\boxed{13}$	13	20	X
E	$\boxed{4}$	$\boxed{6}$	$\boxed{11}$	$\boxed{13}$	$\boxed{13}$	20	X
F	$\boxed{4}$	$\boxed{6}$	$\boxed{11}$	$\boxed{13}$	$\boxed{13}$	$\boxed{18}$	29
G	$\boxed{4}$	$\boxed{6}$	$\boxed{11}$	$\boxed{13}$	$\boxed{13}$	$\boxed{18}$	$\boxed{28}$

a. Which of the following weighted networks could this Dijkstra's algorithm be used for?

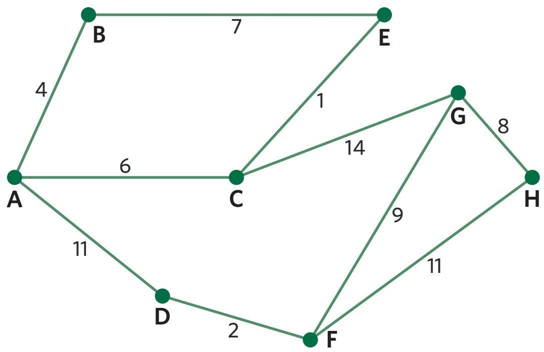
A.



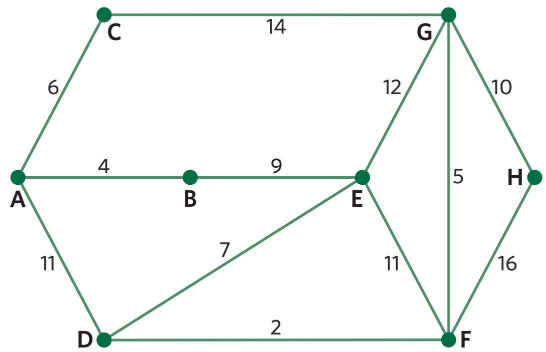
B.



C.



D.



b. What is the length of the shortest path from A to F?

c. What is the route of the shortest path from A to G?

6. The first three rows of a Dijkstra's algorithm solution have been filled out for a particular weighted network. Which of the following statements is not true?

	B	C	D	E	F	G	H
A	8	X	$\boxed{7}$	X	X	13	X
D	$\boxed{8}$	16	$\boxed{7}$	X	9	12	X
B	$\boxed{8}$	14	$\boxed{7}$	X	$\boxed{9}$	11	15

A. The shortest path between vertices A and F has a length of 9.

B. The edge connecting vertices D and G has a length of 5.

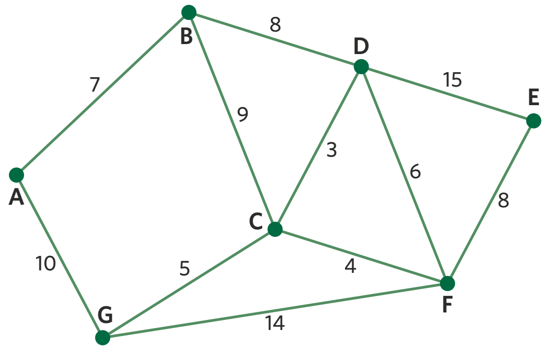
C. There must be an edge connecting vertices B and F.

D. Vertex E is not connected directly to vertices A, B or D.

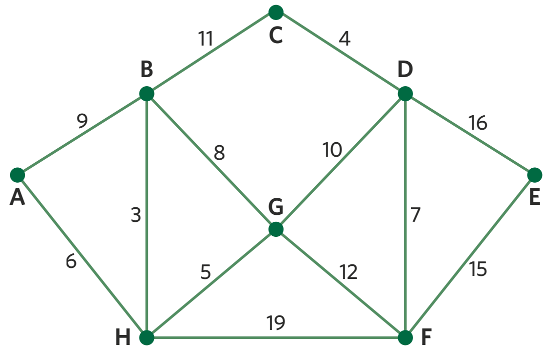
Back to contents

7. Use Dijkstra's algorithm to find, for each of the following weighted networks:
   - the length of the shortest path from vertex A to vertex E.
   - the route of the shortest path from vertex A to vertex E.

a.



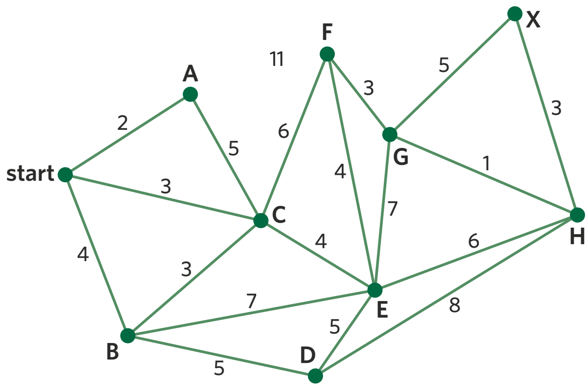
b.



8. During an orienteering task at a school camp, students are given a map and are tasked with getting to point X in the shortest time possible. One group of students converted the map into the following weighted network diagram. Each weight represents the time, in minutes, it would take to travel down the path.

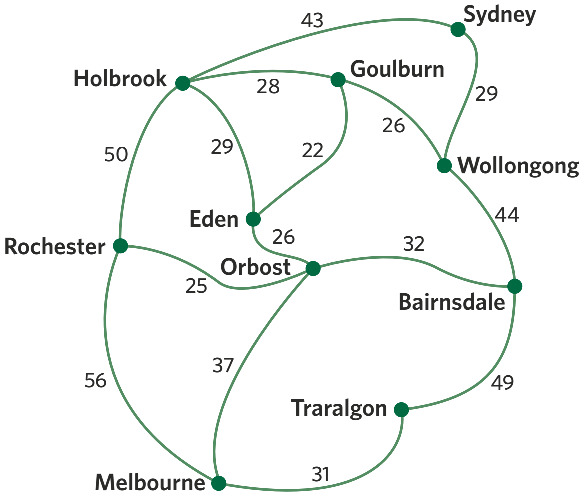
   a. What is the shortest possible time it would take the students to get from the start to point X?

   b. What is the route of the shortest path?



Joining it all together

9. Tom is a truck driver who frequently travels between Melbourne and Sydney. He has calculated his petrol costs for different routes passing through small towns on the way. All values are correct to the nearest dollar.



   a. By inspection, what is the minimum cost of petrol from Melbourne to Bairnsdale?

   A. \$63 **B.** \$69 **C.** \$75 **D.** \$80

   b. When driving from Melbourne to Sydney, what is the minimum amount that Tom has to spend on petrol?

c. While following this route that allows for the minimum cost, which one of the following towns will Tom pass through?

A. Rochester B. Wollongong C. Goulburn D. Eden

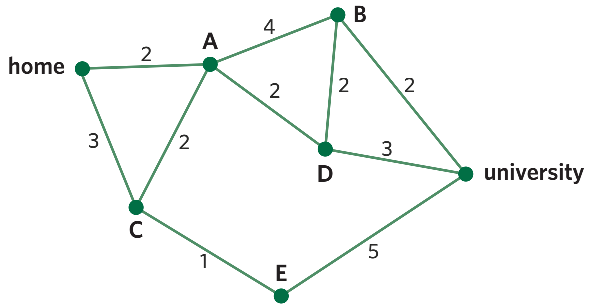
d. Tom has remembered that his favourite bakery is in Goulburn, and decides that he is willing to pay the extra money in order to visit it on his way to Sydney. What is his new minimum cost for petrol?

Exam practice

10. Niko drives from his home to university.

The following network shows the distances, in kilometres, along a series of streets connecting Niko's 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. Prim's algorithm.

C. Dijkstra's algorithm.

D. a minimum spanning tree.

E. an Eulerian trail.

Adapted from VCAA 2018 Exam 1 Networks and decision mathematics Q2

68% of students answered this type of question correctly.

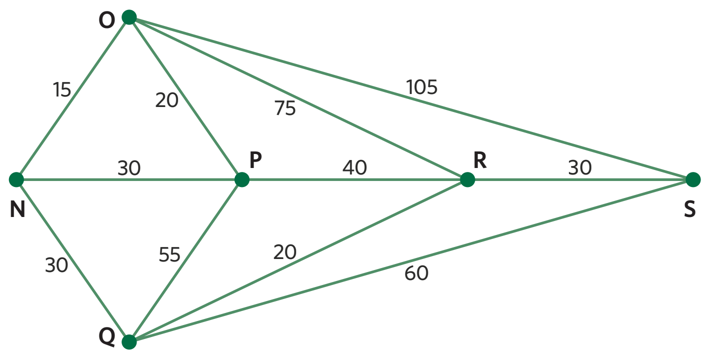
---

11. Bus routes connect six towns.

The towns are Northend (N), Opera (O), Palmer (P), Quigley (Q), Rosebush (R) and Seatown (S).

The following 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).



If Bai takes the cheapest route from Northend (N) to Seatown (S), which other town(s) will he pass through? (1 MARK)

VCAA 2017 Exam 2 Networks and decision mathematics Q1b

79% of students answered this question correctly.

Back to contents

Questions from multiple lessons

Data analysis

12. You are completing some research on how students at your school use public transport. You have decided to focus on collecting information on two variables:

- *trips per week* (less than 5, 5–10, more than 10)
- *public transport type* (bus, tram, train)

These variables are

A. numerical and categorical respectively.

B. ordinal and nominal respectively.

C. nominal and ordinal respectively.

D. both ordinal.

E. both nominal.

Adapted from VCAA 2017 Exam 1 Data analysis Q7

Networks and decision mathematics

13. A connected planar graph has 7 faces.

This graph could have

A. eight vertices and nine edges.

B. seven vertices and twelve edges.

C. twelve vertices and seven edges.

D. ten vertices and thirteen edges.

E. nine vertices and eight edges.

Adapted from VCAA 2018 Exam 1 Networks and decision mathematics Q3

Recursion and financial modelling

14. Daniel has opened a savings account to buy a new motorbike.

He puts the $5000 that he has saved up in a savings account which pays a 3% interest rate, compounding annually.

a. Complete the recurrence relation to model the amount of money, in dollars, in the savings account after n years, V_n. (1 MARK)

$V_0 = \boxed{}, \quad V_{n+1} = \boxed{} \times V_n$

b. The amount of money, in dollars, in the account after n years, V_n, can be determined using a rule. Complete the rule by writing the appropriate numbers in the boxes provided. (1 MARK)

$V_n = \boxed{}^{\,n} \times \boxed{}$

c. How much money will Daniel have in his savings account after 5 years? Round to the nearest cent. (1 MARK)

Adapted from VCAA 2016 Exam 2 Data analysis Q5

Back to contents

8G Matching problems

STUDY DESIGN DOT POINTS

- 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

8A 8B 8C 8D 8E 8F 8G 8H 8I 8J

KEY SKILLS

During this lesson, you will be:

- representing matching problems using a bipartite graph
- solving matching problems.

KEY TERMS

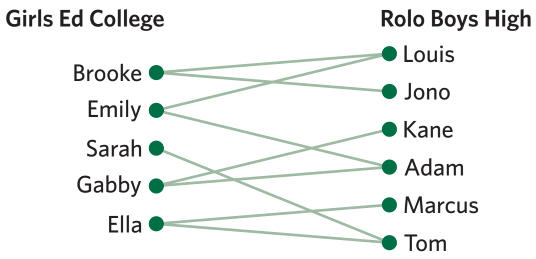
- Bipartite graph
- Hungarian algorithm

There are often situations in which activities or tasks need to be allocated or matched in a way to minimise costs or time, or to maximise scores. An example of this would be the allocation of swimmers to different strokes in a medley relay race to achieve the fastest time. These allocations can be represented and solved using bipartite graphs and the Hungarian algorithm.

Representing matching problems using a bipartite graph

A **bipartite graph** can be used to show connections between vertices which fall into two separate groups. A vertex cannot be directly connected to another vertex in the same group.

For example, the following bipartite graph shows students from two different schools that have met at a party.



Bipartite graphs are sometimes shown with arrows instead of lines. This is to show the direction of the connection from one vertex to another.

Worked example 1

A random survey of individuals' sports preferences was conducted. The results of 4 random individuals from the survey are shown in the following table. Use the information from the table to construct a bipartite graph.

individual	preferences
James	basketball, soccer, tennis
Alice	football
Harry	soccer, tennis
Nathan	soccer, tennis, football

Continues →

Back to contents

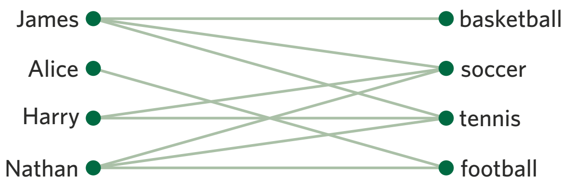
Explanation

Step 1: Draw two sets of dots representing the two sets of groups.

There are 4 different sports, so there will be 4 dots, each representing a sport.

Step 2: Draw lines to connect individuals to the sport they like based on the information from the table.

Answer



Solving matching problems

Matching problems are situations in which multiple allocations of tasks or activities can be made. The optimal allocation is the allocation that minimises costs, such as time or money, or maximises a desirable outcome such as scores.

For smaller scale matching problems, the optimal allocation can be solved using inspection. However, to solve large scale matching problems, the Hungarian algorithm must be used.

See worked example 2

The **Hungarian algorithm** uses cost matrices to find the optimal allocation, or the allocation which will give the minimum cost.

See worked example 3

The following steps can be used to find the optimal allocation for any cost matrix. The table shown gives information relating to the amount of time a group of students would each take to prepare a presentation on a subject in days.

	art	maths	science	drama
Alice	3	6	8	10
Bonnie	7	3	9	11
Carl	8	6	2	4
Danny	4	7	8	7

1. Locate the lowest value in each row. Subtract it from each element in the same row.
2. Find the minimum number of lines required to cover all of the zeros in the table.

   If this number is the same as the number of allocations needed to be made, jump to step 7.

	art	maths	science	drama
Alice	0	3	5	7
Bonnie	4	0	6	8
Carl	6	4	0	2
Danny	0	3	4	3

3. Locate the lowest value in each column. This may be a zero. Subtract it from each element in the same column.

Back to contents

4. Find the minimum number of lines required to cover all of the zeros in the table.

   If this number is the same as the number of allocations needed to be made, jump to step 7.

	art	maths	science	drama
Alice	0	3	5	5
Bonnie	4	0	6	6
Carl	6	4	0	0
Danny	0	3	4	1

5. Locate the lowest value that is not covered by a line. Subtract it from all uncovered values and add it to all values covered by two lines.
6. Find the minimum number of lines required to cover all of the zeros in the table.

   If it is not the same as the number of allocations to be made, repeat step 5.

	art	maths	science	drama
Alice	0	3	4	4
Bonnie	4	0	5	5
Carl	7	5	0	0
Danny	0	3	3	0

7. Allocate the activities based on the location of zeros.

	art	maths	science	drama
Alice	0	3	4	4
Bonnie	4	0	5	5
Carl	7	5	0	0
Danny	0	3	3	0

Alice: art, Bonnie: maths, Carl: science, Danny: drama

Worked example 2

A university group assignment details a writing task on a research topic. The members in the group decide to split up the work based on which member was the most efficient at each task. Using inspection, find the minimum cost of the following cost matrix. The cost values are in hours.

	planning	researching	writing	editing
Bryan	1	4	7	5
Mason	4	5	3	7
Lily	2	2	3	7
Carmen	8	7	6	4

Explanation

Step 1: Determine the most efficient individual at each stage.

For the planning stage, Bryan is the most efficient with a time of 1 hour.

For the researching stage, Lily is the most efficient with a time of 2 hours.

For the writing stage, Mason and Lily are the most efficient with a time of 3 hours.

For the editing stage, Carmen is the most efficient with a time of 4 hours.

Step 2: Find the optimal allocation.

Bryan: planning, Lily: researching, Mason: writing, Carmen: editing

Both Mason and Lily are the most efficient for the writing stage, however, Lily is the most efficient for the researching stage. This means Lily will be allocated to researching, and Mason will be allocated to writing.

Continues →

Back to contents

Step 3: Calculate the minimum cost.

$1 + 2 + 3 + 4 = 10$

Answer

10 hours

Worked example 3

Using the Hungarian algorithm, find the minimum cost of the following cost matrix. The cost values are in hours.

	activity A	activity B	activity C	activity D
Amy	4	4	2	4
Bryce	4	3	4	5
Cherry	2	5	6	3
Dinnith	8	7	4	6

Explanation

Step 1: Locate the lowest value in each row. Subtract it from each element in the same row.

	activity A	activity B	activity C	activity D	number subtracted
Amy	2	2	0	2	2
Bryce	1	0	1	2	3
Cherry	0	3	4	1	2
Dinnith	4	3	0	2	4

Step 2: Find the minimum number of lines required to cover all of the zeros in the table.

	activity A	activity B	activity C	activity D
Amy	2	2	0	2
Bryce	1	0	1	2
Cherry	0	3	4	1
Dinnith	4	3	0	2

Only three lines are required, but there are four allocations to be made.

Allocation is not yet possible.

Step 3: Locate the lowest value in each column. This may be a zero. Subtract it from each element in the same column.

	activity A	activity B	activity C	activity D
Amy	2	2	0	1
Bryce	1	0	1	1
Cherry	0	3	4	0
Dinnith	4	3	0	1
number subtracted	0	0	0	1

Continues →

Back to contents

Step 4: Find the minimum number of lines required to cover all of the zeros in the table.

	activity A	activity B	activity C	activity D
Amy	2	2	0	1
Bryce	1	0	1	1
Cherry	0	3	4	0
Dinnith	4	3	0	1

Only three lines are required, but there are four allocations to be made.

Allocation is not yet possible.

Step 5: Locate the lowest value that is not covered by a line. Subtract it from all uncovered values and add it to all values covered by two lines.

	activity A	activity B	activity C	activity D
Amy	1	1	0	0
Bryce	1	0	2	1
Cherry	0	3	5	0
Dinnith	3	2	0	0

Step 6: Find the minimum number of lines required to cover all of the zeros in the table.

	activity A	activity B	activity C	activity D
Amy	1	1	0	0
Bryce	1	0	2	1
Cherry	0	3	5	0
Dinnith	3	2	0	0

Four lines are required.

Allocation is now possible.

Step 7: Allocate the activities based on the location of zeros.

	activity A	activity B	activity C	activity D
Amy	1	1	0	0
Bryce	1	0	2	1
Cherry	0	3	5	0
Dinnith	3	3	0	0

Cherry will complete activity A.

Bryce will complete activity B.

Amy and Dinnith may both complete either activity C or D.

Step 8: Calculate the minimum cost using the cost matrix.

	activity A	activity B	activity C	activity D
Amy	4	4	2	4
Bryce	4	3	4	5
Cherry	2	5	6	3
Dinnith	8	7	4	6

If Amy completes activity C and Dinnith completes activity D: $2 + 3 + 2 + 6 = 13$

If Amy completes activity D and Dinnith completes activity C: $2 + 3 + 4 + 4 = 13$

Answer

13 hours

Exam question breakdown

VCAA 2016 Exam 1 Networks and decision mathematics Q8

Five children, Alan, Brianna, Chamath, Deidre and Ewen, are each to be assigned a different job by their teacher. The following table shows the time, in minutes, that each child would take to complete each of the five jobs.

	Alan	Brianna	Chamth	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

Continues →

Back to contents

The teacher wants to allocate the jobs so as to minimise the total time taken to complete 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. Chamth or Ewen

E. Deidre or Ewen

Explanation

Step 1: Apply the Hungarian algorithm.

Subtract the lowest value in each row from all elements in the same row and check the number of lines required to cover the zeros. There are 3 lines required to cover all the zeros.

As each column contains a zero, subtracting the lowest value in each column from all elements in the same column has no effect.

Subtract the smallest uncovered number ('1') from all uncovered elements and add it to elements that are covered twice.

	Alan	Brianna	Chamth	Deidre	Ewen
job 1	0	3	0	3	3
job 2	0	2	1	2	0
job 3	4	0	2	0	5
job 4	1	1	3	2	0
job 5	0	0	0	0	0

Step 2: Allocate the activities based on the location of zeros.

Job 4 must be allocated to Ewen.

As Ewen is allocated to job 4, job 2 must be allocated to Alan.

As Alan is allocated to job 2, job 1 must be allocated to Chamth.

Brianna or Deidre can be allocated to either job 3 or job 5.

Step 3: Compare and interpret the two possible allocations.

The two possible allocations are shown in the following table.

		Alan	Brianna	Chamath	Deidre	Ewen
allocation 1	job	2	3	1	5	4
	completion time	5 min	5 min	5 min	4 min	5 min
allocation 2	job	2	5	1	3	4
	completion time	5 min	4 min	5 min	5 min	5 min

The child that finishes their allocated job the fastest could be either Deidre or Brianna

30% of students answered this question correctly.

Many students applied the Hungarian algorithm incorrectly. In addition, some students did not recognise that three of the children had a single optimal allocation.

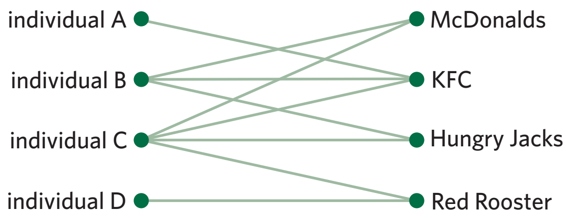
Answer

B

8G Questions

Representing matching problems using a bipartite graph

1. The fast-food preferences of four random individuals are displayed in the following bipartite graph.



Which of the following tables correctly shows the fast-food preferences of the four individuals?

A.

individual	preferences
A	KFC
B	McDonalds, KFC, Hungry Jacks
C	McDonalds, KFC, Hungry Jacks, Red Rooster
D	Red Rooster

B.

individual	preferences
A	KFC
B	KFC, Hungry Jacks
C	McDonalds, KFC, Hungry Jacks
D	Red Rooster

C.

individual	preferences
A	KFC
B	McDonalds, KFC, Hungry Jacks, Red Rooster
C	McDonalds, KFC, Hungry Jacks
D	Red Rooster

D.

individual	preferences
A	Red Rooster
B	McDonalds, KFC, Hungry Jacks, Red Rooster
C	McDonalds, KFC, Hungry Jacks
D	KFC

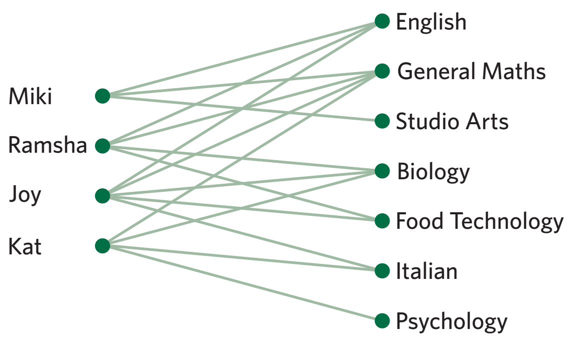
2. Miki, Ramsha, Joy, and Kat are trying to set up a study group for each of their Year 12 subjects. The bipartite graph shows which subjects the students have in common.

a. How many of the students study English?

b. Which subject has the most students studying it?

c. The students want to set up a study group only for subjects that three or more of them are studying. How many study groups will they have?

d. Which study groups will Miki and Ramsha be in together?



3. Priya, Rhys, Ted, Georgia, Natalia, and Andy were all asked to compete in one event at the school athletics carnival. Each student was asked to write down which events they would be happy to participate in. This is summarised in the following table.

Construct a bipartite graph to represent the following information.

student	events
Priya	100 m, 200 m, high jump
Rhys	100 m, 200 m, high jump
Ted	800 m
Georgia	800 m, javelin
Natalia	100 m, 200 m, triple jump, high jump
Andy	high jump, javelin

Back to contents

Solving matching problems

4. The following table summarises the cost, in dollars, of producing three types of clothing at three different factories.

	shirts	pants	hoodies
factory A	2	3	5
factory B	5	1	5
factory C	6	7	4

Using inspection, which of the following allocations will minimise the total cost of producing the different types of clothing?

A. factory A: hoodies, factory B: shirts, factory C: pants

B. factory A: pants, factory B: hoodies, factory C: shirts

C. factory A: shirts, factory B: hoodies, factory C: pants

D. factory A: shirts, factory B: pants, factory C: hoodies

5. Khan, Henry, Jamie and Mya have been assigned to work with each other on a group project. The project involves four different sections: research, writing, editing, and presentation. Each member of the group will complete one section.

Table 1 shows the time, in minutes, it would take each member of the group to complete each section.

Table 1

	research	writing	editing	presentation
Khan	60	70	65	65
Henry	50	60	40	70
Jamie	45	60	50	70
Mya	60	55	60	75

The sections will be allocated based on the minimum total time to complete the entire project.

The Hungarian algorithm will be used to make the allocation.

The first step of the Hungarian algorithm is to locate the lowest value in each row and subtract that value from each element in the row. The first step is shown in Table 2, however the values for Mya are missing.

Table 2

	research	writing	editing	presentation
Khan	0	10	5	5
Henry	10	20	0	30
Jamie	0	15	5	25
Mya				

a. Complete Table 2.

b. Explain why it is not yet possible to allocate sections of the project.

The steps of the Hungarian algorithm have continued to be applied until allocation is possible.

Table 3

	research	writing	editing	presentation
Khan	0	10	5	0
Henry	10	20	0	25
Jamie	0	15	5	20
Mya	5	0	5	15

c. Explain why Jamie must complete the research section instead of Khan.

d. Determine the final allocation for each section. If there is more than one possible allocation, provide them all.

e. How long will it take the group to complete the project?

6. Anna, Margot, Tina, and Svetlana have all been asked to perform a solo piece at their end of year dance concert. To keep the concert exciting, they will each perform a different discipline of dance; ballet, jazz, hip hop and contemporary. They have been ranked from 1 to 4 on their recent performances in each discipline. The best dancer achieves a ranking of 1 for that discipline.

	ballet	jazz	hip hop	contemporary
Anna	3	1	2	2
Margot	4	3	1	4
Tina	2	2	3	3
Svetlana	1	4	4	1

Use the Hungarian algorithm to determine which discipline each dancer should perform. If there is more than one possible allocation, provide them all.

7. The Question Creation Team (QCT) at Edrolo wanted to minimise mistakes in their textbook. They decided to have each QCT member be responsible for checking for one of the following mistake types: clarity, grammar, mathematical errors, and other. To decide who was best placed to look for each mistake type, they recorded the number of each type of mistake each QCT member made themselves in one lesson.

	Sophie	Zephyr	Anna	Cassie
clarity	10	8	21	9
grammar	17	17	11	9
mathematical errors	2	1	4	4
other	9	4	5	8

Use the Hungarian algorithm to allocate the QCT members to a mistake type. The allocation should be based on the minimum number of mistakes possible. If there is more than one possible allocation, provide them all.

Joining it all together

8. There are four builders available to complete four tasks, A, B, C and D, for a small construction project. All builders are qualified to complete all tasks.

a. Create a bipartite graph showing each of the builders and the tasks they are able to complete.

The building company wishes to complete the project for the lowest total price possible. In order to complete the project on time, each builder must complete a different task. The table below shows the cost, in dollars, for each worker to complete each task.

	task A	task B	task C	task D
builder 1	100	80	120	90
builder 2	130	110	100	110
builder 3	120	110	120	80
builder 4	90	80	130	100

The first step of the Hungarian algorithm involves locating the lowest value in each row and subtracting that value from each element in the row.

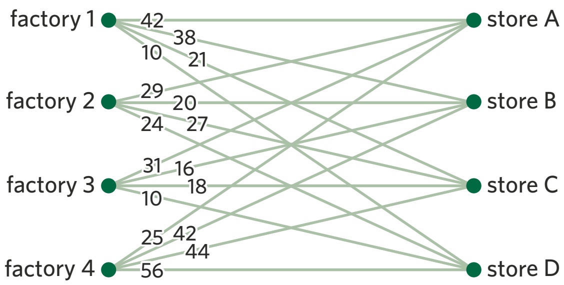
b. Complete the first step of the Hungarian algorithm.

c. Allocate each of the builders to a task which will provide the lowest total cost for the building company.

d. Calculate the cost of this allocation.

Back to contents

9. Bedazzled Bathers has four different production factories. Four retail stores wish to stock bathers from Bedazzled Bathers. The distance, in kilometres, from each production factory to each retail store is shown in the following bipartite graph.



Each factory only makes enough bathers to stock one retail store.

Calculate the lowest possible total distance for all retail stores to receive bathers from a different factory.

Exam practice

10. Lee, Mandy, Nola, Oscar and Pieter are each to be allocated one particular task at work. The following bipartite graph 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

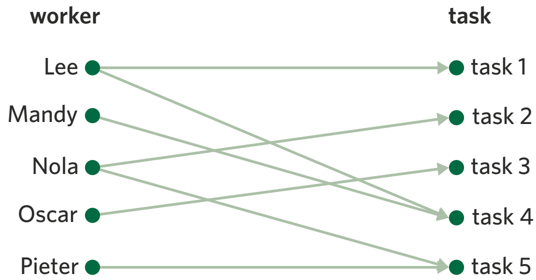
B. Mandy

C. Nola

D. Oscar

E. Pieter

VCAA 2016 Exam 1 Networks and decision mathematics Q1



97% of students answered this question correctly.

11. A cricket team has 11 players who are each assigned to a batting position.

Three of the new players, Alex, Bo and Cameron, can bat in position 1, 2, or 3.

The following table shows the average scores, in runs, for each player for the batting positions 1, 2 and 3.

		batting position		
		1	**2**	**3**
player	**Alex**	22	24	24
	Bo	25	25	21
	Cameron	24	25	19

Each player will be assigned to one batting position.

To which position should each player be assigned to maximise the team's score? (1 MARK)

VCAA 2020 Exam 2 Networks and decision mathematics Q2

57% of students answered this question correctly.

Back to contents

12. Bai joins his friend Agatha, Colin and Diane when he arrives for a 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	13	7	13	12
tour 2	14	9	8	7
tour 3	19	25	21	18
tour 4	10	7	11	10

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

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)

VCAA 2017 Exam 2 Networks and decision mathematics Q2a

37% of students answered this question correctly.

---

13. 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 following table.

Buddhi takes 3 minutes for task 3.

He takes k minutes for each other task.

	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

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

A. 0 **B.** 1 **C.** 2
D. 3 **E.** 4

VCAA 2018 Exam 1 Networks and decision mathematics Q8

24% of students answered this question correctly.

Back to contents

Questions from multiple lessons

Recursion and financial modelling

14. Four lines of an amortisation table for an annuity investment are shown in the following table.

The interest rate for this investment remains constant, but the payment value may vary.

payment number	payment	interest	principal addition	balance of investment
27	250	114.64	364.64	14 695.14
28	250	117.56	367.56	15 062.71
29	250	120.50	370.50	15 433.21
30				16 000.00

The balance of the investment after payment number 30 is $16 000.

The value of payment number 30 is closest to

A. $123 **B.** $250 **C.** $373 **D.** $443 **E.** $567

Adapted from VCAA 2017 Exam 1 Data analysis Q23

Networks and decision mathematics

15. The following directed graph shows the flow of traffic, in people per minute, in a system of paths connecting a car park to an arena.

The maximum flow, in people per minute, from the car park to the arena is

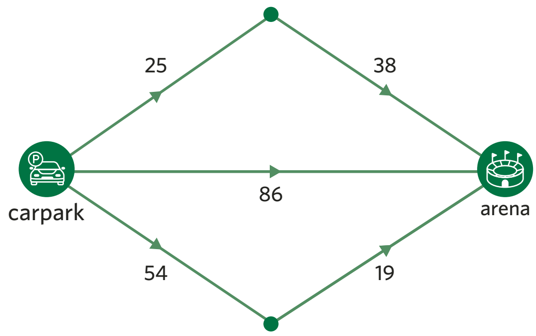
A. 73

B. 86

C. 130

D. 143

E. 178



Adapted from VCAA 2016 Exam 1 Networks and decision mathematics Q2

Networks and decision mathematics

16. In Venice, Italy, a gondola rider takes tourists to nine destinations along canals labelled as vertices A to I on the following graph.

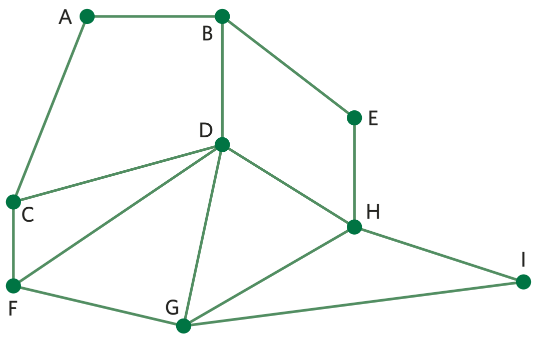
a. Which one of the vertices on the graph has a degree of 5? (1 MARK)

b. For this graph, an Eulerian trail does not currently exist.

For an Eulerian trail to exist, what is the minimum number of extra edges that the graph would require? (1 MARK)

c. The gondola rider has taken tourists to D and continues the tour by following a Hamiltonian path from D.

Draw a possible Hamiltonian path for the gondola rider. (1 MARK)



Adapted from VCAA 2018 Exam 2 Networks and decision mathematics Q2

Back to contents

8H Activity networks and precedence tables

STUDY DESIGN DOT POINT

- construction of an activity network from a precedence table (or equivalent) including the use of dummy activities where necessary

8A 8B 8C 8D 8E 8F 8G **8H** 8I 8J

KEY SKILLS

During this lesson, you will be:

- interpreting precedence tables
- constructing activity networks
- constructing activity networks with dummy activities.

KEY TERMS

- Immediate predecessor
- Precedence table
- Activity network
- Dummy activity

Networks can be used to visualise the workflow of large projects when the individual activities within it are known, as well as the order in which activities need to be completed. Modelling projects using networks is useful for assigning tasks with the goal of increasing efficiency.

Interpreting precedence tables

In projects that require multiple activities to be completed, an **immediate predecessor** is an activity that must be completed directly before another activity can begin. For example, when getting dressed, putting on socks is an immediate predecessor to putting on shoes.

A **precedence table** displays all of the activities within a project, and each of their immediate predecessors.

Worked example 1

Consider the following precedence table.

activity	immediate predecessor(s)
A	–
B	A
C	A
D	B
E	A, D
F	E
G	F
H	C

What activities need to be completed before activity F can begin?

Explanation

Step 1: Determine the immediate predecessors of activity F.

The immediate predecessor of activity F is activity E.

Continues →

Back to contents

Step 2: Continue to determine the immediate predecessors of activities that occur before *F*.

The immediate predecessors of activity *E* are activities *A* and *D*.

The immediate predecessor of activity *D* is activity *B*.

The immediate predecessor of activity *B* is activity *A*.

Activity *A* has no immediate predecessors.

Answer

A, *B*, *D*, *E*

Constructing activity networks

See worked example 2

An **activity network** is a visual display of the information in a precedence table. Each activity is represented by a directed edge. The vertices remain unlabelled, except for the vertices that represent the start and finish of the project.

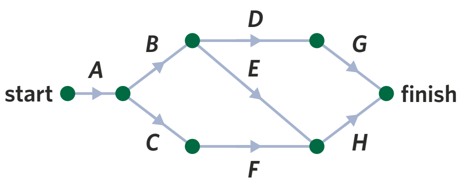
Activities with no immediate predecessors are represented as edges leading from the start vertex. Activities that are not immediate predecessors to any other activities are represented by edges leading to the finish vertex.

See worked example 3

A precedence table can also be created from an activity network.

For example, the following precedence table and activity network display the same project.

activity	immediate predecessor(s)
A	–
B	*A*
C	*A*
D	*B*
E	*B*
F	*C*
G	*D*
H	*E*, *F*



Worked example 2

Construct an activity network from the following precedence table.

activity	immediate predecessor(s)
A	–
B	–
C	*A*
D	*A*
E	*B*
F	*B*
G	*D*, *E*
H	*C*, *G*
I	*F*, *H*

Continues →

Back to contents

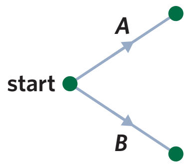
Explanation

Step 1: Determine the starting activities.

The starting activities are those with no immediate predecessors.

Activities A and B have no immediate predecessors, so they are the starting activities.

Draw these activities as edges connected to the start vertex.

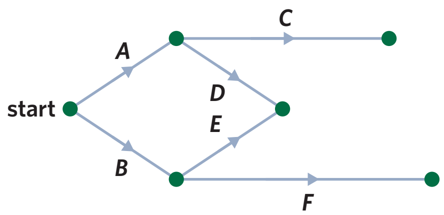


Step 2: Draw all activities with A or B as their immediate predecessor.

Activities C and D have activity A as their immediate predecessor.

Activities E and F have activity B as their immediate predecessor.

Activity G has activities D and E as its immediate predecessors, so activities D and E must meet at a single vertex.



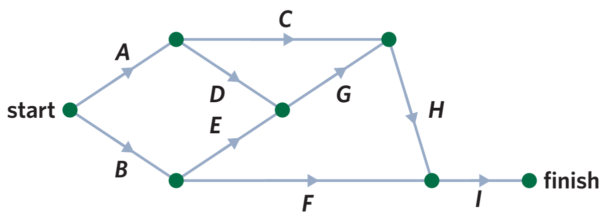
Step 3: Continue to draw activities based on their immediate predecessor(s).

Activity G has activities D and E as its immediate predecessors. Activity H has activities C and G as its immediate predecessors, so activity G must start at the end of activities D and E and reach the same vertex as activity C, which will lead to activity H.

Activity I has activities F and H as its immediate predecessors, so activity H must start at the end of activities C and G and reach the same vertex as activity F.

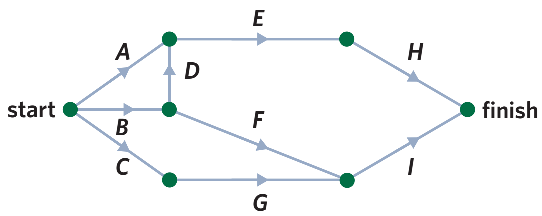
Activity I is not an immediate predecessor for any other activity, so it will connect to the finish vertex.

Answer



Worked example 3

Construct a precedence table from the following activity network.



Continues →

Back to contents

Explanation

Step 1: Construct a table.

The table will need 2 columns and a row for each activity.

Label each of the columns 'activity' and 'immediate predecessor(s)' respectively. Fill in the activity column with the names of each activity.

activity	immediate predecessor(s)
A	
B	
C	
D	
E	
F	
G	
H	
I	

Step 2: Enter a dash for each activity with no immediate predecessors.

Activities A, B and C start at the start vertex, so they have no immediate predecessors.

Step 3: Fill in the rest of the table.

The immediate predecessor of activity D is activity B.

The immediate predecessors of activity E are activities A and D.

The immediate predecessor of activity F is activity B.

The immediate predecessor of activity G is activity C.

The immediate predecessor of activity H is activity E.

The immediate predecessors of activity I are activities F and G.

Answer

activity	immediate predecessor(s)
A	–
B	–
C	–
D	B
E	A, D
F	B
G	C
H	E
I	F, G

Constructing activity networks with dummy activities

When activities in a project share some, but not all, of the same immediate predecessors, it becomes impossible to draw an activity network correctly. In these circumstances, a **dummy activity** can be used to show the additional predecessors.

A dummy activity is drawn from the end of the shared immediate predecessor to the start of the activity that has an additional immediate predecessor.

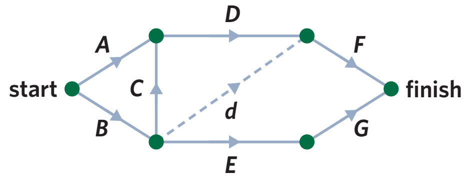
Dummy activities are drawn using dotted lines, and are often denoted 'dummy' or '*d*'.

Note: A dummy activity does not actually exist as a real activity. It is merely a directed edge to properly show the immediate predecessors in an activity network.

Back to contents

For example, the following precedence table and activity network display the same project. The dummy activity runs from the end of activity *B* (shared immediate predecessor) to the start of activity *F* (activity with an additional immediate predecessor).

activity	immediate predecessor(s)
A	–
B	–
C	*B*
D	*A*, *C*
E	*B*
F	*B*, *D*
G	*E*



Worked example 4

Consider the following precedence table.

activity	immediate predecessor(s)
A	–
B	*A*
C	*A*
D	*A*
E	*B*
F	*C*
G	*D*
H	*E*, *F*
I	*C*, *G*

a. Complete the following sentence.

A dummy activity needs to be drawn from the end of activity __________ to the start of activity __________.

Explanation

Step 1: Find activities that share some, but not all, of the same immediate predecessors.

Activity *F* has activity *C* as its immediate predecessor.

Activity *I* has activites *C* and *G* as its immediate predecessors.

Step 2: Find the activity that the dummy activity should start at.

A dummy activity is drawn from the end of the shared immediate predecessor.

The shared immediate predecessor is activity *C*.

Step 3: Find the activity that the dummy activity should end at.

A dummy activity will be drawn to the start of the activity with an additional immediate predecessor.

The activity with an additional shared immediate predecessor is activity *I*.

Answer

A dummy activity needs to be drawn from the end of activity *C* to the start of activity *I*.

Continues →

Back to contents

b. Construct an activity network from the precedence table.

Explanation

Step 1: Determine the starting activities.

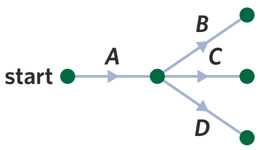
The starting activities are those with no immediate predecessors.

Activity *A* has no immediate predecessor and so is the starting activity.

Draw this as an edge connected to the start vertex.

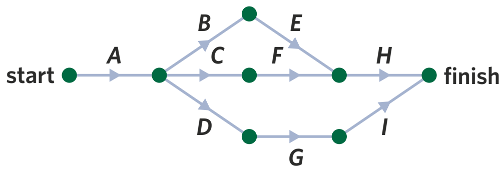
Step 2: Draw all activities with *A* as their immediate predecessor.

Activities *B*, *C* and *D* all have activity *A* as their immediate predecessor.



Step 3: Continue to draw activities based on their immediate predecessor(s).

Ignore activity *C* as an immediate predecessor of activity *I* as this will be shown as a dummy activity.

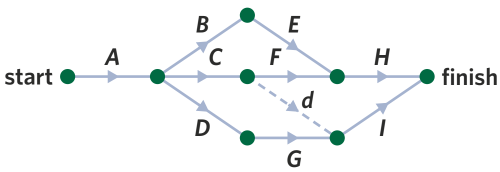


Step 4: Draw the dummy activity.

The dummy activity needs to be drawn from the end of activity *C* to the start of activity *I*.

It should be a dotted line denoted '*d*'.

Answer



Exam question breakdown

VCAA 2019 Exam 1 Networks and decision mathematics Q7

A project involves nine activities, *A* to *I*.

The immediate predecessor(s) of each activity is shown in the following 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 predecessor(s)
A	–
B	*A*
C	*A*
D	*B*
E	*B*, *C*
F	*D*
G	*D*
H	*E*, *F*
I	*G*, *H*

Explanation

Step 1: Find activities that share some, but not all, of the same immediate predecessors.

Activity *D* has activity *B* as its immediate predecessor.

Activity *E* has activites *B* and *C* as its immediate predecessors.

Step 2: Find the activity that the dummy activity should start at.

A dummy activity is drawn from the end of the shared immediate predecessor.

The shared immediate predecessor is activity *B*.

Continues →

Back to contents

Step 3: Find the activity that the dummy activity should end at.

A dummy activity will be drawn to the start of the activity with an additional immediate predecessor.

The activity with an additional shared immediate predecessor is activity *E*.

Answer

B

42% of students answered this question correctly.

21% of students incorrectly chose option A. These students correctly determined that activity *B* was a shared predecessor, and that the dummy should start at activity *B*, but could not determine where the dummy activity ends.

8H Questions

Interpreting precedence tables

1. Consider the following precedence table.

What activities need to be completed before activity *E* can begin?

A. *C*

B. *A*, *C*

C. *A*, *B*, *C*

D. *A*, *B*, *C*, *D*, *F*

activity	immediate predecessor(s)
A	–
B	*A*
C	*A*
D	*B*
E	*C*
F	*D*
G	*F*, *E*

2. Hannah created a precedence table for a DIY project she wanted to complete.

Which activities must be completed before activity *H* can commence?

activity	immediate predecessor(s)
A	–
B	*A*
C	*A*
D	*B*
E	*C*
F	*C*
G	*E*
H	*D*, *E*
I	*F*, *G*

Constructing activity networks

3. Consider the following activity network.

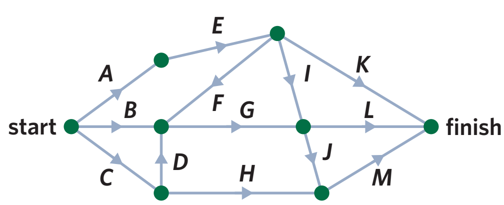
Which activities must be completed before activity G can commence?

A. *B*

B. *B*, *C*, *D*

C. *A*, *B*, *C*, *D*

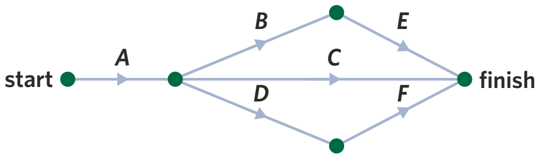
D. *A*, *B*, *C*, *D*, *E*, *F*



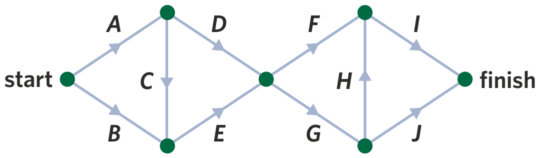
Back to contents

4. Construct a precedence table for each of the following activity networks.

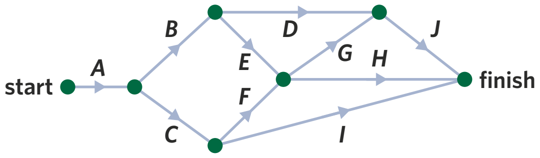
a.



b.



c.



5. Construct an activity network for each of the following precedence tables.

a.

activity	immediate predecessor(s)
A	–
B	–
C	–
D	A
E	B
F	C
G	D, E

b.

activity	immediate predecessor(s)
A	–
B	–
C	A
D	C
E	B, D
F	B, D
G	B, D
H	E
I	F
J	G

c.

activity	immediate predecessor(s)
A	–
B	–
C	–
D	B
E	A
F	B
G	C, D
H	C, D
I	E, F
J	I, G
K	I, G
L	H
M	J

Back to contents

6. Dustin Martin has multiple strength and conditioning tests that he needs to complete before being considered fit to play in the AFL finals. The tests are shown as activities in the following precedence table.

Construct an activity network for Dustin Martin's strength and conditioning tests.

activity	immediate predecessor(s)
A	–
B	–
C	A
D	A
E	D
F	D
G	B
H	C
I	F, G
J	H
K	E, I

Constructing activity networks with dummy activities

7. Consider the project displayed in the following precedence table.

A corresponding activity network will include a dummy activity drawn from

A. the end of activity B to the start of activity E.

B. the end of activity B to the start of activity H.

C. the end of activity D to the start of activity G.

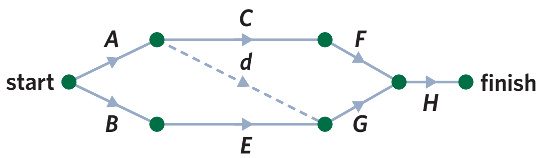
D. the end of activity D to the start of activity H.

activity	immediate predecessor(s)
A	–
B	A
C	A
D	A
E	B
F	D
G	E
H	B, C, F

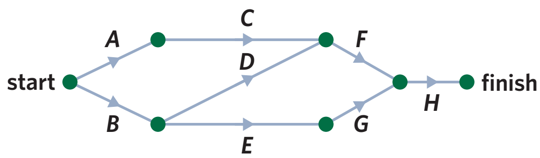
8. Which activity network corresponds to the following precedence table?

activity	immediate predecessor(s)
A	–
B	–
C	A
D	C
E	B
F	C
G	A, E
H	F, G

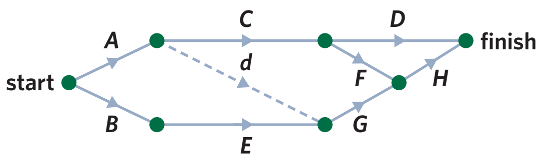
A.



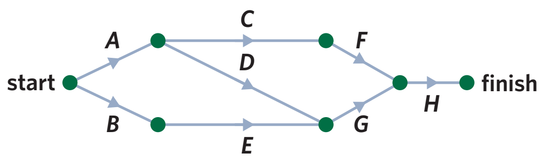
B.



C.



D.



Back to contents

9. For each of the following precedence tables, if a corresponding activity network was constructed, how many dummy activities would be drawn?

a.

activity	immediate predecessor(s)
A	–
B	–
C	–
D	*A*
E	*A*, *B*
F	*C*
G	*D*
H	*E*
I	*E*, *F*

b.

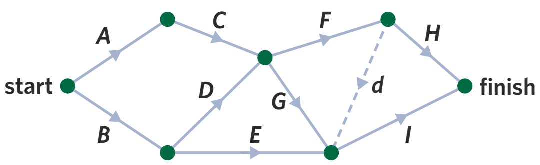
activity	immediate predecessor(s)
A	–
B	–
C	*B*
D	*B*
E	*A*, *C*
F	*D*, *E*
G	*F*
H	*F*

c.

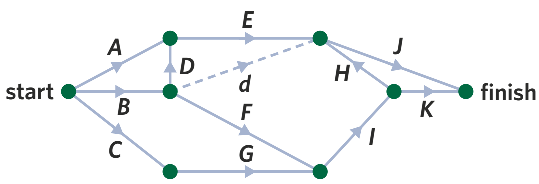
activity	immediate predecessor(s)
A	–
B	*A*
C	*A*
D	*B*
E	*B*
F	*C*, *D*
G	*C*, *D*
H	*E*, *F*
I	*F*
J	*F*, *G*

10. Construct a precedence table for each of the following activity networks.

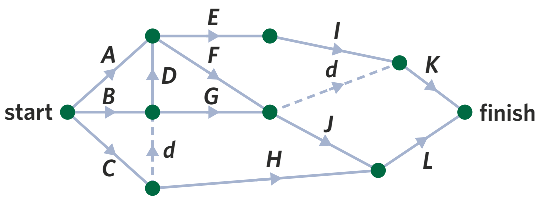
a.



b.



c.



Back to contents

11. Construct an activity network for each of the following precedence tables.

a.

activity	immediate predecessor(s)
A	–
B	–
C	–
D	A
E	C
F	B, D
G	B, D, E

b.

activity	immediate predecessor(s)
A	–
B	A
C	A
D	B
E	A, D
F	C
G	B
H	E
I	E, F

c.

activity	immediate predecessor(s)
A	–
B	–
C	A
D	A
E	B, C, D
F	B, C
G	B, C
H	F
I	F
J	G
K	E, H
L	I, J

12. Nova is creating a Piñata for her daughter's 4th birthday party. The steps required to create the Piñata are shown in the following precedence table.

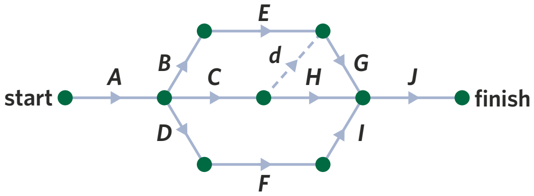
Construct an activity network for the project.

activity	immediate predecessor(s)
A	–
B	–
C	–
D	A, B
E	B
F	B, C
G	D, E, F
H	G
I	E, F
J	B, C

Back to contents

Joining it all together

13. Christine is planning on sewing an outfit for her birthday party. She has decided to make a jumpsuit with a matching jacket. The activity network shows all the activities she must complete for the project.



a. Construct a precedence table for the project.

Descriptions of activity are shown in the following table.

activity	description
A	cut out pattern pieces
B	sew the top of the jumpsuit
C	sew the pant legs
D	sew the body of the jacket
E	sew the jumpsuit sleeves
F	sew the jacket sleeves
G	sew the top and bottom of the jumpsuit together
H	hem the pant legs
I	sew a zip onto the jacket
J	iron

b. What must Christine do immediately before she can sew the top and bottom of the jumpsuit together?

c. Which of the following activities can Christine **not** complete before sewing the pant legs?

A. Sew a zip onto the jacket

B. Sew the top of the jumpsuit

C. Sew the jacket sleeves

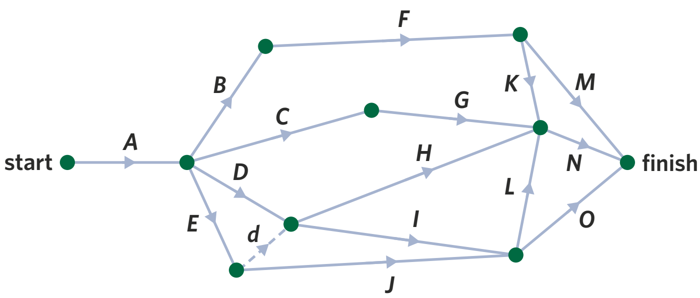
D. Sew the top and bottom of the jumpsuit together

Exam practice

14. The rides at a theme park are set up at the beginning of each holiday season.

This project involves activities A to O.

The following directed network shows these activities.



Write down the two immediate predecessors of activity I. (1 MARK)

Adapted from VCAA 2017 Exam 2 Networks and decision mathematics Q4a

67% of students answered this type of question correctly.

Back to contents

15. The Sunny Coast cricket clubroom is undergoing a major works project.

This project involves nine activities: A to I.

The following table shows the immediate predecessor(s) of each activity.

The information in the table can be used to complete a directed network.

This network will require a dummy activity.

Complete the following sentence by filling in the boxes provided.

This dummy activity could be drawn as a directed edge from the end of

activity ☐ to the start of activity ☐. (1 MARK)

Adapted from VCAA 2020 Exam 2 Networks and decision mathematics Q5a

activity	immediate predecessor(s)
A	–
B	–
C	A, B
D	C
E	C
F	B
G	D
H	F
I	E, G, H

26% of students answered this type of question correctly.

Questions from multiple lessons

Data analysis

16. The heights of students at Edrolo High are approximately normally distributed with a mean of 170 cm and a standard deviation of 6 cm.

If there are 600 students at the school, the number of students expected to be taller than 164 cm is closest to

A. 84 **B.** 408 **C.** 442 **D.** 504 **E.** 507

Adapted from VCAA 2017NH Exam 1 Data analysis Q9

Networks and decision mathematics

17. Anton, Belinda, Christian and Daniella all swim together. They compete together in the 4×100 mixed medley relay. In the race, each swimmer will complete one of four strokes: backstroke, breaststroke, butterfly and freestyle.

The usual completion times for Anton, Belinda, Christian and Daniella are shown in the following table.

	backstroke	breaststroke	butterfly	freestyle
Anton	61	78	71	51
Belinda	t	72	t	t
Christian	57	73	82	56
Daniella	64	79	61	62

Belinda takes 72 seconds to complete the breaststroke.

She takes t seconds for each other stroke.

Their coach allocated the strokes for the race as follows:

- Backstroke to Christian
- Breaststroke to Belinda
- Butterfly to Daniella
- Freestyle to Anton

This allocation will achieve the minimum total race completion time if the value of t is at least

A. 50 **B.** 51 **C.** 56 **D.** 72 **E.** 73

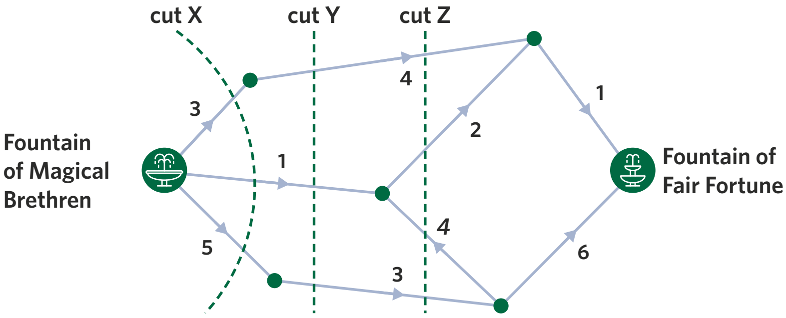
Adapted from VCAA 2018 Exam 1 Networks and decision mathematics Q8

Back to contents

Networks and decision mathematics

18. The following graph shows the amount of water flowing through pipes in a series of connected fountains.

The weighting of each edge represents the maximum amount of water (in litres) that can flow through each pipe every second.



a. Cut X, shown on the graph, has a capacity of 9 L/s.

Two other cuts are labelled as cut Y and cut Z.

i. Write down the capacity of cut Y. (1 MARK)

ii. Write down the capacity of cut Z. (1 MARK)

b. Determine the maximum volume of water that can be delivered each second from the 'Fountain of Magical Brethren' to the 'Fountain of Fair Fortune'. (1 MARK)

Adapted from VCAA 2018 Exam 2 Networks and decision mathematics Q1

Back to contents

8I Critical path analysis

STUDY DESIGN DOT POINTS

- 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

8A 8B 8C 8D 8E 8F 8G 8H 8I 8J

KEY SKILLS

During this lesson, you will be:

- interpreting weighted precedence tables
- forward scanning to determine the EST and minimum completion time
- backward scanning to determine the LST
- determining the critical path and float times.

KEY TERMS

- Scheduling
- Minimum completion time
- Earliest start time (EST)
- Latest start time (LST)
- Critical path
- Float time

Each activity within a project can be assigned a time weighting, or an estimated completion time. The activity network can then be used to find the minimum completion time of the project, which activities need to be prioritised, and where breaks can be taken without reducing the project completion time.

Interpreting weighted precedence tables

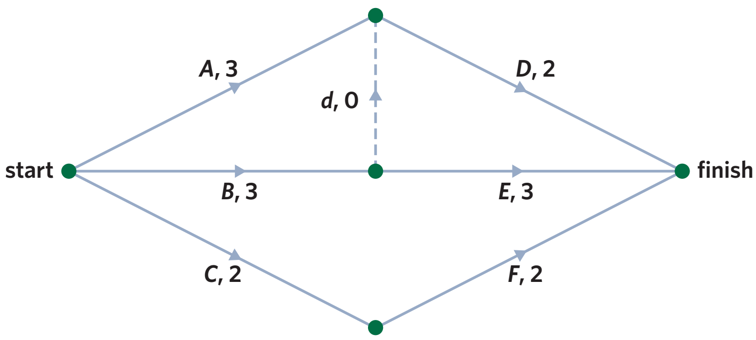
In addition to information about the order of activities, a time weighting, or estimated completion time, for each of the activities may also be known. The allocation of an estimated completion time for each activity is known as **scheduling**.

Scheduling information can be shown in a weighted precedence table by including an additional column with the estimated completion time of each activity. This information can then be used to calculate the minimum time for the entire project to be completed.

Note: As dummy activities are not real activities, they have an estimated completion time of 0.

For example, the following weighted precedence table and activity network represent the same project.

activity	estimated completion time	immediate predecessor(s)
A	3	–
B	3	–
C	2	–
D	2	*A*, *B*
E	3	*B*
F	2	*C*



Back to contents

Worked example 1

Construct an activity network based on the following weighted precedence table.

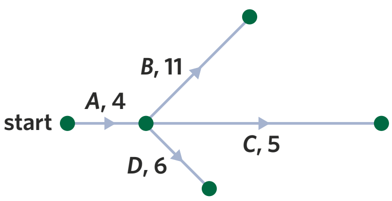
activity	duration (hours)	immediate predecessor(s)
A	4	–
B	11	A
C	5	A
D	6	A
E	2	D
F	3	E
G	4	B
H	9	B, C, F

Explanation

Step 1: Draw the start vertex along with the starting activities.

Activity A has no immediate predecessor so it is the starting activity. Draw this as an edge connected to the starting vertex, and label its estimated completion time.

Step 2: Draw the activities with A as their immediate predecessor, and label their estimated completion times.



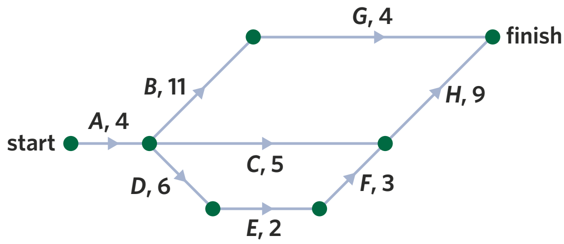
Step 3: Continue to draw activities based on their immediate predecessor, and label their estimated completion times.

The predecessor of activity G is activity B, so draw activity G from activity B.

The predecessor of activity E is activity D, so draw activity E from activity D.

Activity H has shared predecessors of F and C, so draw activity F from the end of activity E to the end of activity C, with H following from that vertex.

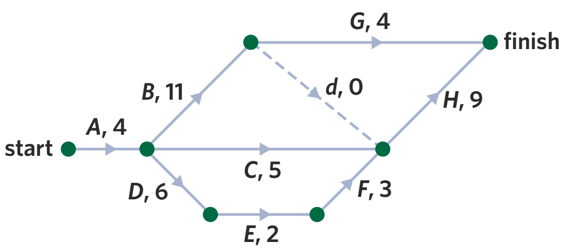
Activities G and H are not immediate predecessors for any other activity so they will connect to the finish vertex.



Step 4: Draw a dummy activity.

Activity B is also an immediate predecessor of activity H. To complete the activity network, a dummy activity must be drawn from activity B to activity H. Dummy activities have an estimated completion time of 0.

Answer



Back to contents

Forward scanning to determine the EST and minimum completion time

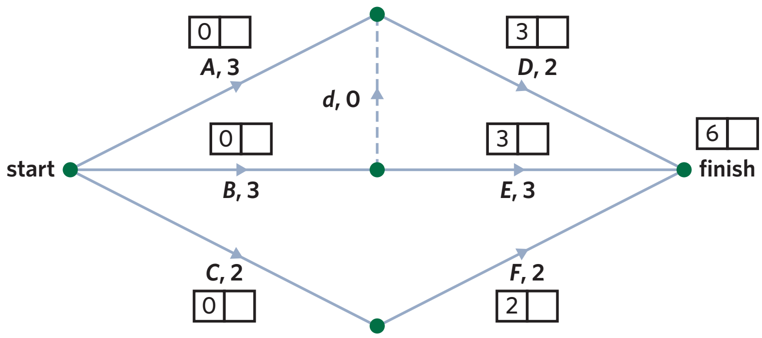
The **minimum completion time** of a weighted activity network is the shortest time in which the project, which is all the activities within the network, can be completed. The **earliest start time (EST)** is the earliest time, relative to the beginning of the project, that an individual activity can start at. To minimise the total completion time for a project, each activity should begin at its EST.

The process for finding the EST of each activity within a project is called forward scanning.

1. Draw a box alongside each activity of the network, as well as the finish vertex. The box should be split into two separate cells.
2. Fill in the EST as zero in the left-hand cell of each box for the starting activities (activities that connect to the start vertex).
3. Continuing filling in the EST in the left-hand cell for all other activities, as well as the finish vertex, working forwards from start to finish.
   - If an activity has one immediate predecessor, its EST can be found by adding the EST and duration of the immediate predecessor.
     EST = *EST of predecessor* + *duration of predecessor*
   - If an activity has two or more immediate predecessors, its EST will be the largest value found after calculating the EST using each of the predecessors.

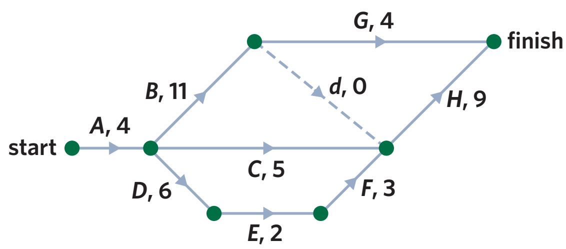
The EST of the finish vertex is the minimum possible completion time for the project.

For example, the following project has a minimum completion time of 6 hours.



Worked example 2

Consider the following activity network. The estimated completion times are in hours.

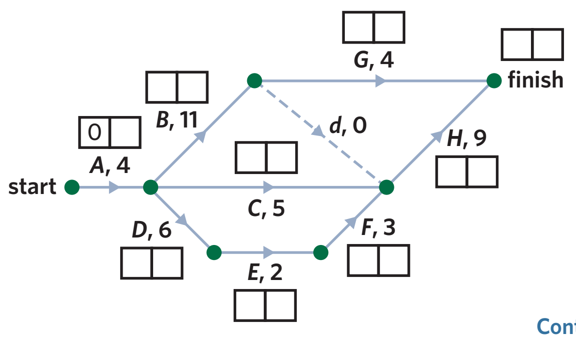


a. Find the earliest start time of activity *H*.

Explanation

Step 1: Draw a box, split into two cells, at each activity, excluding the dummy activity.

Fill in the EST as 0 in the left-hand cell of the starting activity (*A*).



Continues →

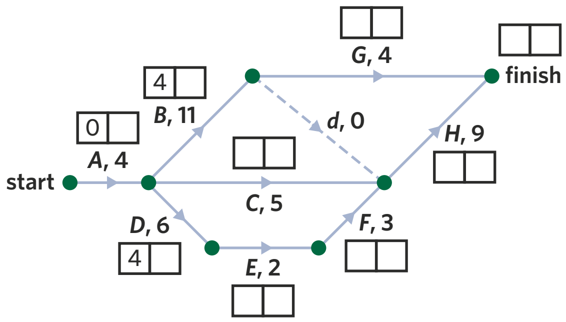
Step 2: Fill in the EST of activities B, C and D.

If an activity has one immediate predecessor, its EST can be found by adding the EST and duration of the previous activity.

$EST = EST\ of\ predecessor + duration\ of\ predecessor$

The immediate predecessor of activities B, C and D is activity A.

$EST = 0 + 4 = 4$

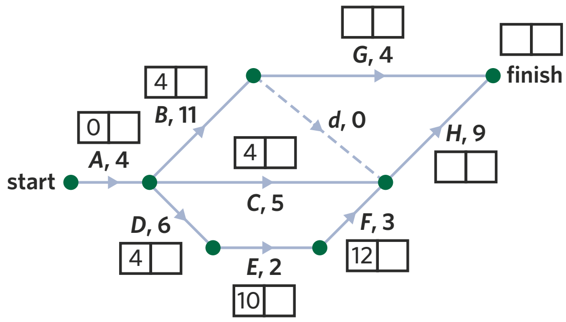


Step 3: Find the EST of activities E and F.

Activities E and F have only one immediate predecessor.

Activity E: $EST = 4 + 6 = 10$

Activity F: $EST = 10 + 2 = 12$



Step 4: Find the EST of activity H.

If an activity has two or more immediate predecessors, its EST will be the largest value found when looking at the predecessors individually.

Using the dummy activity as the predecessor:

$EST = 4 + 11 + 0 = 15$

Using activity C as the predecessor:

$EST = 4 + 5 = 9$

Using activity F as the predecessor:

$EST = 12 + 3 = 15$

Answer

15 hours

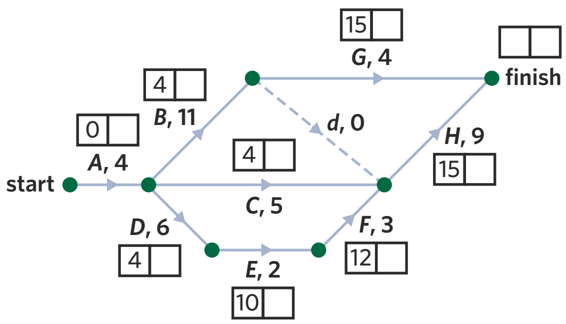
---

b. Find the minimum completion time of the project.

Explanation

Step 1: Continuing on from part **a**, find the EST of activity G.

$EST = 4 + 11 = 15$



Step 2: Find the EST of the finish vertex.

The EST will be the largest value found when looking at the predecessors individually.

Using activity G as the predecessor:

$EST = 15 + 4 = 19$

Using activity H as the predecessor:

$EST = 15 + 9 = 24$

Step 3: Determine the minimum completion time of the project.

The minimum completion time is the EST of the finish vertex.

Answer

24 hours

Backward scanning to determine the LST

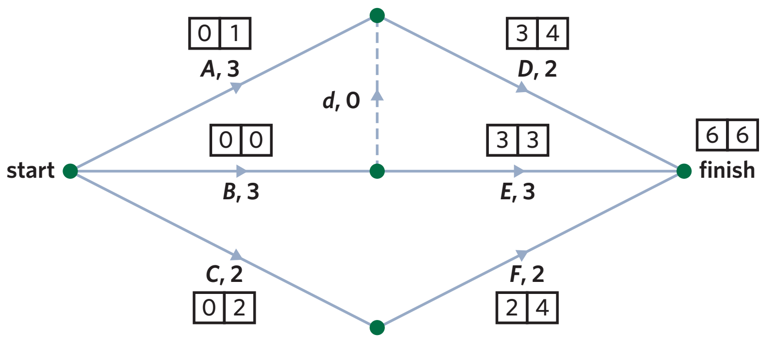
The **latest start time (LST)** is the latest time, relative to the beginning of the project, that an individual activity can start at without increasing the completion time of the entire project. The LST can be greater than the EST because some activities may be able to start after the earliest possible start time and not impact the total minimum completion time for the project.

The process for finding the LST for each activity within a project is called backward scanning.

1. Complete forward scanning.
   - All of the cells on the left should be full.
   - LSTs will be written in the right-hand cell.
2. Fill in the LST of the finish vertex as the minimum possible completion time, which is the same as the EST for the finish vertex.
3. Continue filling in the LST of all other activities, working backwards from finish to start.
   - If an activity has only one activity following it, its LST can be found by subtracting the duration of the activity from the LST of the activity following it.
     LST = LST of following activity − duration of activity
   - If an activity has two or more activities following it, its LST will be the smallest value found when looking at each of the activities immediately following it.

Note: The LST of at least one of the starting activities should be zero.

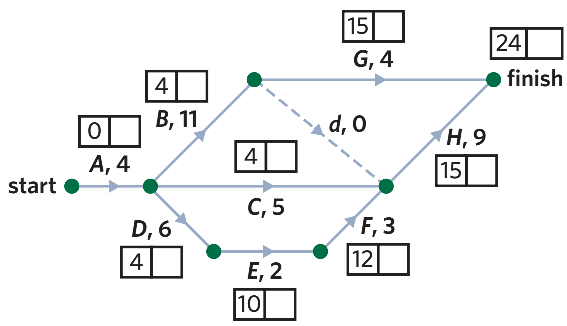
For example, activity *C* in the following project has an LST of 2.



Worked example 3

Consider the following activity network with the earliest start time for each activity filled in. The estimated completion times are in hours.

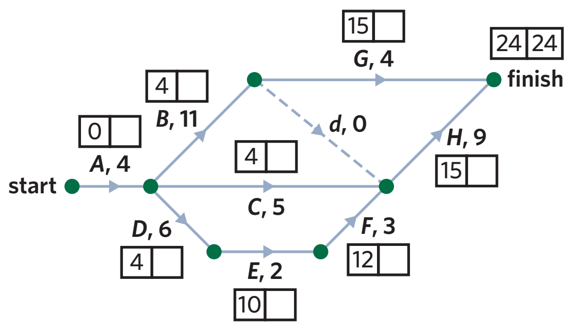
Find the latest start time of activity *B*.



Explanation

Step 1: Fill in the LST of the finish vertex as the minimum possible completion time.

The EST for the finish vertex is 24 hours. This will also be the LST.



Continues →

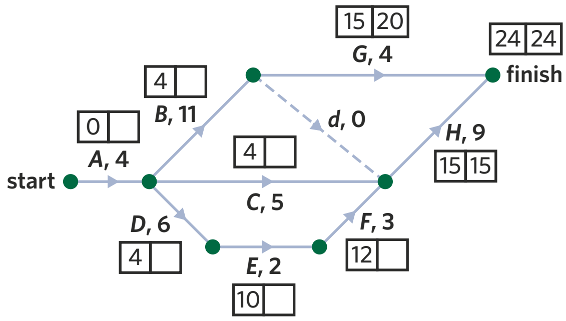
Step 2: Fill in the LST of activities G and H.

Only the finish vertex follows activities G and H.

$LST = LST\ of\ following\ activity - duration\ of\ activity$

Activity G: $LST = 24 - 4 = 20$

Activity H: $LST = 24 - 9 = 15$



Step 3: Find the LST of activity B.

If an activity has two or more activities following it, its LST will be the smallest value found when looking at each of the activities immediately following it.

Using activity H as the following activity:

$LST = 15 - 0 - 11 = 4$

Using activity G as the following activity:

$LST = 20 - 11 = 9$

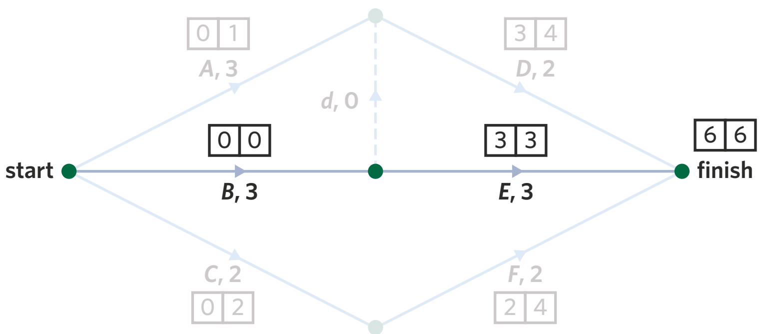
Answer

4 hours

Determining the critical path and float times

An activity is critical when its EST is equal to its LST. A **critical path** is a sequence of critical activities, each of which cannot be delayed without increasing the minimum completion time of the project. It is possible to have multiple critical paths within a project.

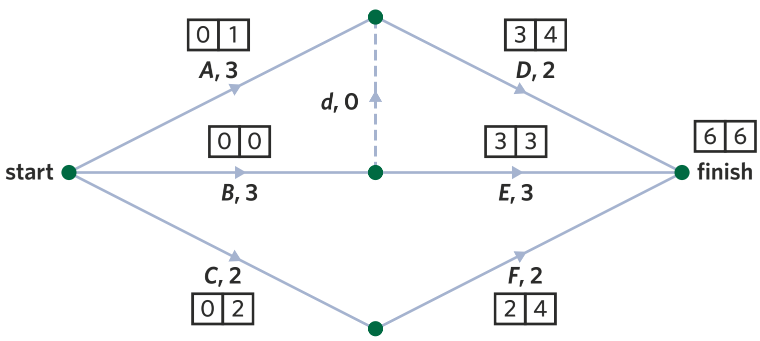
For example, the critical path for the following project is B–E, as the ESTs and LSTs of each activity are equal. This means that neither activity B or E can be delayed without increasing the completion time of the project.



The **float time** of an activity is the maximum amount of time that it can be delayed without changing the completion time of the project. It can be calculated by finding the difference between the EST and LST. The float time of a critical activity is 0.

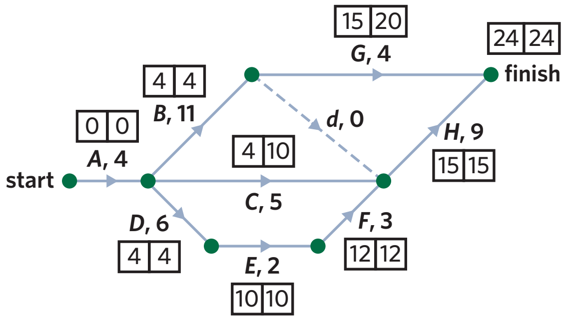
$float\ time = LST - EST$

For example, activity F has a float time of $4 - 2 = 2$ in the following project. This means that activity F can start up to 2 hours late without impacting the project's minimum completion time.



Worked example 4

Consider the following activity network with the earliest and latest starting time for each activity shown. The estimated completion times are in hours.

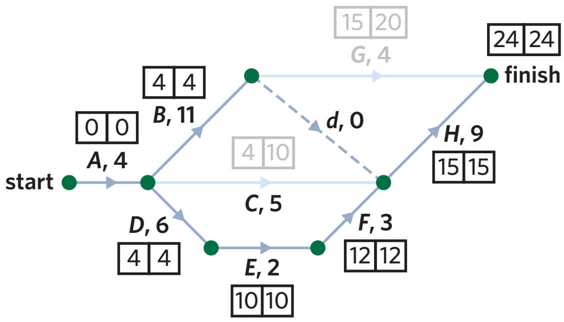


a. Determine the critical path(s) of the project.

Explanation

Step 1: Highlight all activities that have an EST and LST that are equal.

Step 2: Find all the paths that connect the start vertex to the finish vertex using only critical activities.



Answer

A–*B*–*H* and *A*–*D*–*E*–*F*–*H*

b. Determine the float time of activity *C*.

Explanation

Step 1: Determine the EST and LST of activity *C*.

The LST of activity *C* is 10.

The EST of activity *C* is 4.

Step 2: Calculate the float time.

$$\begin{aligned} \textit{float time} &= LST - EST \\ &= 10 - 4 \\ &= 6 \end{aligned}$$

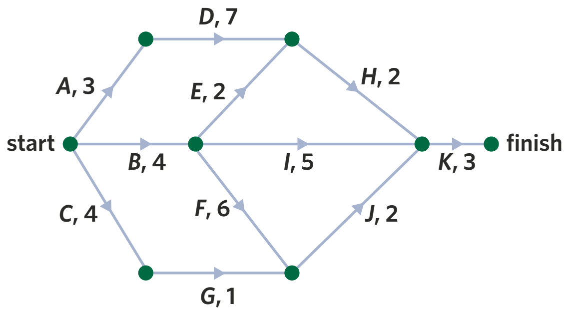
Answer

6 hours

Exam question breakdown

VCAA 2018 Exam 1 Networks and decision mathematics Q5

The directed network below 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.



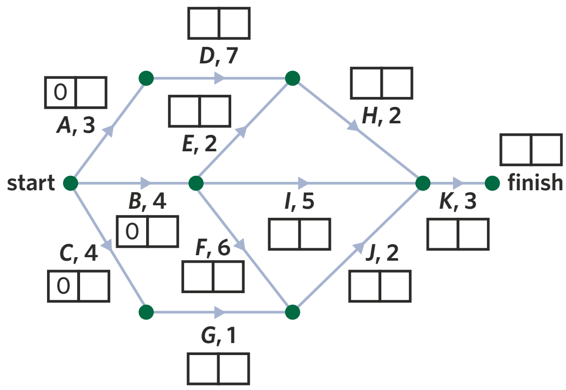
How many of these activities could be delayed without affecting the minimum completion time of the project?

A. 3 **B.** 4 **C.** 5 **D.** 6 **E.** 7

Explanation

Step 1: Draw a box, split into two cells, at each activity and the finish vertex.

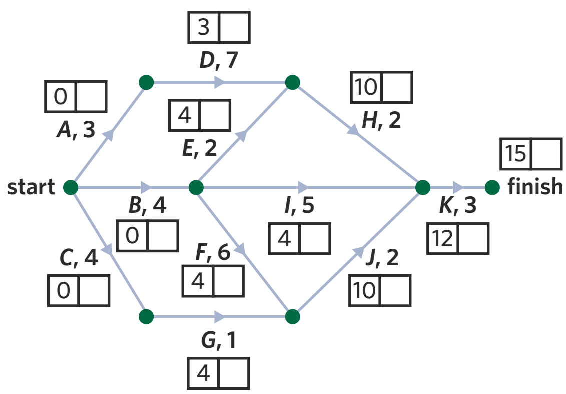
Fill in the EST as 0 in the left-hand cell of each of the starting activities (*A*, *B* and *C*).



Step 2: Fill in the EST in the left-hand cell for all remaining activities and the finish vertex.

EST = EST of predecessor + duration of predecessor

If an activity has two or more immediate predecessors, its EST will be the largest value found when looking at the predecessors individually.

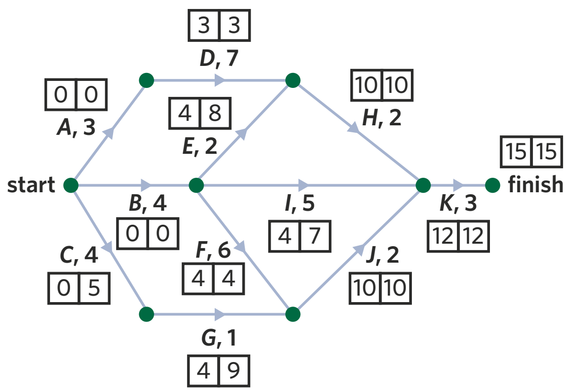


Step 3: Fill in the LST in the right-hand cell for all activities and the finish vertex.

LST = LST of following activity − duration of activity

The LST of the finish vertex is the same as the EST.

If an activity has two or more activities following it, its LST will be the smallest value found when looking at each of the activities immediately following it.



Step 4: Determine the number of non-critical activities.

Non-critical activities have an LST that is larger than its EST.

The non-critical activities are *C*, *E*, *G* and *I*.

Answer

B

43% of students answered this question correctly.

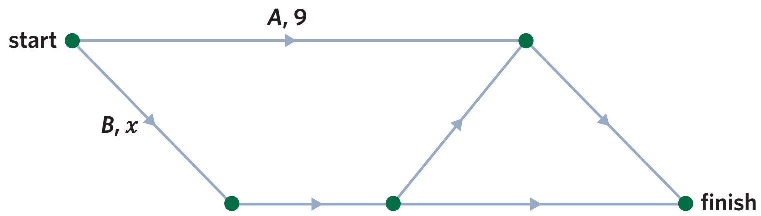
27% of students incorrectly chose option A. These students likely found the 2 critical paths, which each had 4 activities, and subtracted 8 from 11 total activities to reach 3 non-critical activities. However, activity *K* is in both critical paths, meaning that there are actually 4 non-critical activities.

8I Questions

Interpreting weighted precedence tables

1. A weighted precedence table and its corresponding activity network are shown. Some labels on the activity network are missing.

activity	duration (days)	immediate predecessor(s)
A	9	–
B	5	–
C	4	*B*
D	6	*C*
E	2	*A*, *D*
F	3	*C*



a. What is the value of x?

A. 2

B. 4

C. 5

D. 6

b. Copy the activity network and fill in the missing labels.

2. Consider the following weighted precedence table.

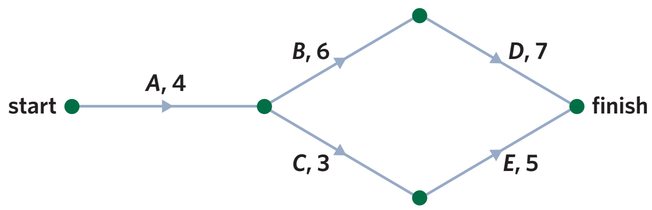
a. Draw an activity network based on the table.

b. The precedence table has been altered such that activity *F* cannot start until activity *D* has been completed. Show this change on the activity network.

activity	duration (minutes)	immediate predecessor(s)
A	4	–
B	13	*A*
C	12	*A*
D	11	*A*
E	4	*B*
F	6	*C*, *E*
G	13	*D*
H	18	*B*
I	15	*F*, *G*

Forward scanning to determine the EST and minimum completion time

3. Consider the following activity network.



a. The EST of activity *B* is

A. 3

B. 4

C. 6

D. 7

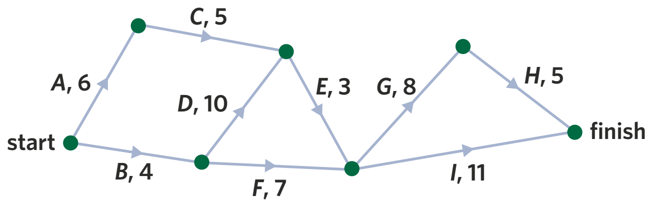
b. Determine the EST of activity *E*.

c. Determine the minimum completion time of the project.

4. The following activity network displays the time in minutes taken for each of the steps of a DIY craft project.

a. What is the earliest start time after the beginning of the project that activity *E* can start?

b. What is the minimum amount of time that it would take to finish the DIY craft project?



Backward scanning to determine the LST

5. Consider the following activity network.

The earliest start time of each activity is filled in.

a. The LST of activity *C* is

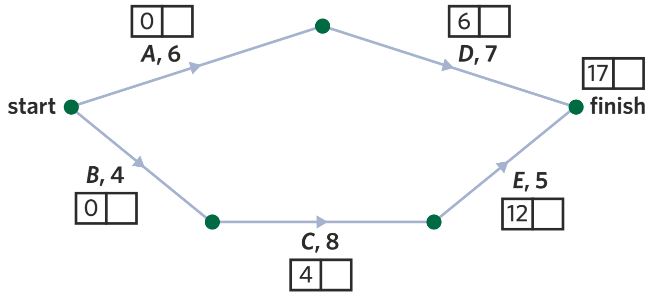
A. 4

B. 7

C. 10

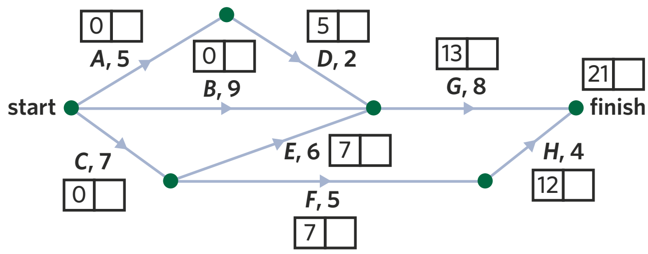
D. 12

b. Determine the LST of activity *A*.



6. A few students are doing a small group project in class. The activity network shows the time in minutes required for all of the tasks within the project, as well as the earliest start time for each.

Danny, who has been assigned to activity *D*, decides to go to the bathroom just as the group begins the project. How long can Danny spend on his bathroom break before he holds up the rest of the project?



Determining the critical path and float times

7. Which one of the following statements about critical paths is true?

A. Each project has only one critical path.

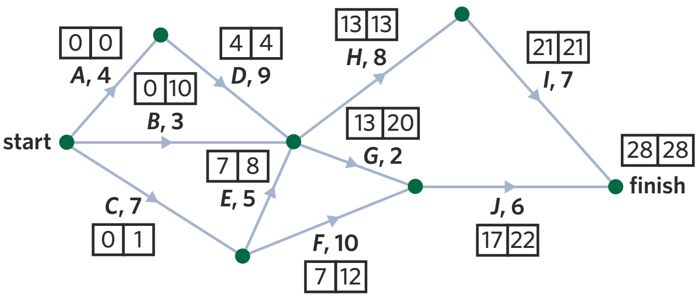
B. A critical path must include at least three activities.

C. A critical path must include the activity that takes the longest time to complete.

D. Reducing the estimated completion time of any activity on a critical path for a project will always reduce the minimum completion time for the entire project.

E. Increasing the estimated completion time of any activity on a critical path for a project will always increase the minimum completion time for the entire project.

8. Consider the following activity network.

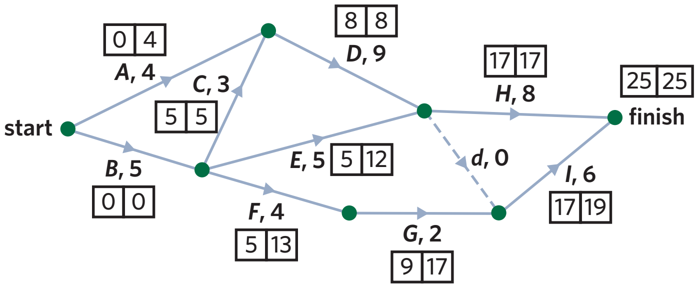


a. Find the critical path(s) of the activity network.

b. What is the float time of activity *J*?

Back to contents

9. A group of young film students are editing a short film. The activity network below shows the activities required to complete the film and their duration in hours.



a. Find the critical path(s) of the project.

b. Would the critical path change if there was no dummy activity?

c. Determine the float time of activity *G*.

d. Would the completion time of the short film change if activity *D* was delayed 1 hour?

e. Would the completion time of the short film change if activity *E* was delayed 4 hours?

Joining it all together

10. Consider the following activity network.

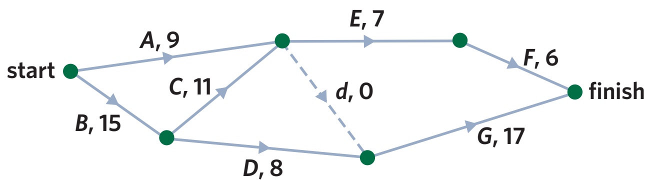
The critical path is

A. *A–E–F*

B. *B–C–G*

C. *B–C–E–F*

D. *B–D–G*



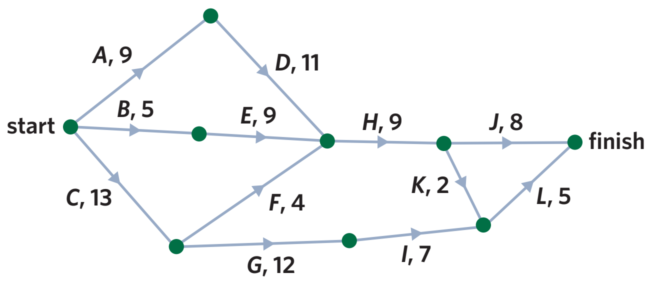
11. The following network has been used to show the activities required to fix a broken laptop. All values are in minutes.

a. What is the least amount of time required to fix the laptop?

b. Determine the float time of activity *G*.

c. Determine the float time of activity *B*.

d. Find the critical path(s) of the project.



12. Jamie Oliver has come up with a new recipe for beef stroganoff. The time (in minutes) needed for each step of the recipe can be represented in the following activity network.

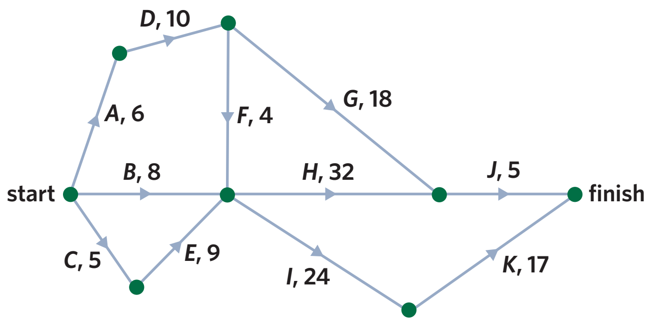
Which of the following statements is true?

A. Activity *J* is critical.

B. Activity *E* has an EST of 11 minutes.

C. Activity *K* has a LST of 44 minutes.

D. Activity *G* has a float time of 4 minutes.



Back to contents

13. Amanda is setting up for her 16th birthday party and has created the following precedence table for the project.

a. Draw an activity network to represent this information.

b. What is the least amount of time that Amanda will need to set up for her party?

c. Find the critical path(s) of the project.

Amanda's mum has bought her a surprise cake for her birthday. This alters the activity network slightly because it means activity H cannot start until activity G has been completed.

d. Show this change on the activity network.

e. How long will Amanda now need to set up for her party?

f. What is the new critical path of the project?

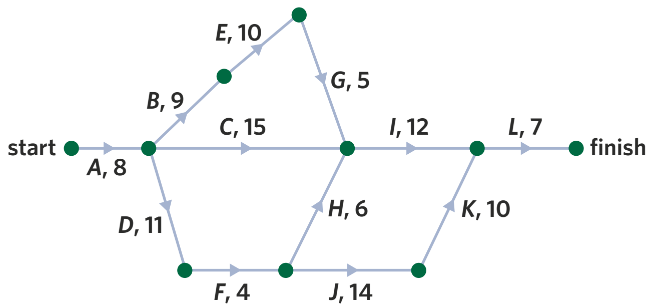
activity	duration (hours)	immediate predecessor(s)
A	8	–
B	10	–
C	5	–
D	11	A
E	12	C
F	6	C
G	9	B, D, E
H	7	F
I	3	G
J	8	G
K	9	H, I
L	4	J, K

14. Beyoncé is mid-way through recording her new album 'Ginger Ale'. Her personal assistant has created an activity network for the activities left in the project. All values are in days.

a. Find the critical path(s) of the project.

b. How long will it take for the album to be finished?

Beyoncé decides at the last minute that she wants to include a bonus track in the album. This requires an extra activity, M, to be added to the project. Activity M can begin after the completion of activity E and it must be finished before activity L can begin. Activity M will take 3 weeks.



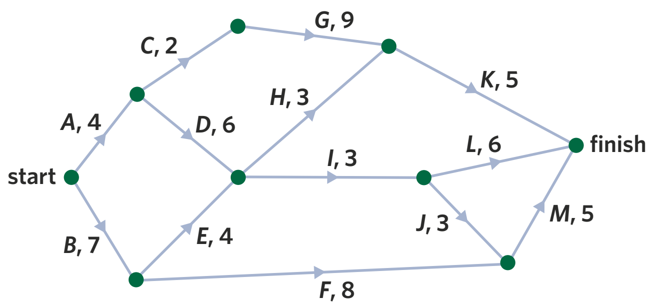
c. Draw activity M on the activity network.

d. Find the critical path(s) of the project with activity M.

e. How much longer will the album take to record due to the addition of the bonus track?

Exam practice

15. The following 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.



The minimum completion time for this project, in days, is

A. 18

B. 19

C. 20

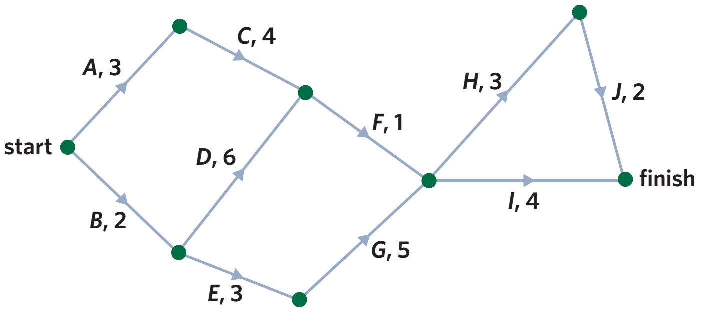
D. 21

E. 22

VCAA 2020 Exam 1 Networks and decision mathematics Q6

61% of students answered this question correctly.

16. At the Zenith Post Office all computer systems are to be upgraded. This project involves 10 activities, A to J. The following directed network 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. Complete the following sentence by filling in the boxes provided.

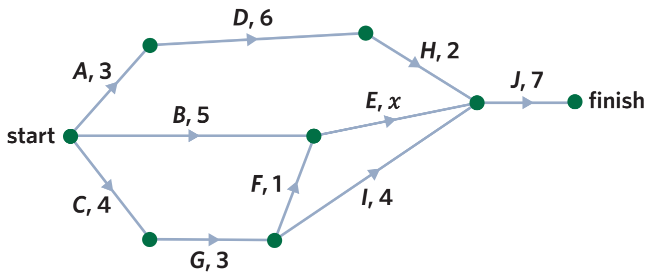
The extra activity could be represented on the network by a directed edge from the end of activity ________ to the start of activity ________. (1 MARK)

VCAA 2018 Exam 2 Networks and decision mathematics Q3a,d

Part **a**: **64%** of students answered this question correctly.
Part **b**: **76%** of students answered this question correctly.
Part **c**: **45%** of students answered this question correctly.
Part **d**: **45%** of students answered this question correctly.

---

17. The directed graph below shows the sequence of activities required to complete a project. The time taken to complete each activity, in hours, is also shown.



The minimum completion time for this project is 18 hours.

The time taken to complete activity E is labelled x.

The maximum value of x is

A. 2

B. 3

C. 4

D. 5

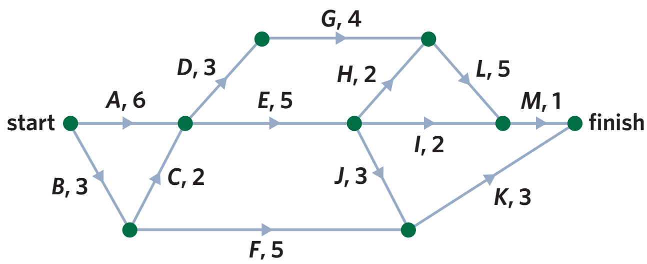
E. 6

VCAA 2021 Exam 1 Networks and decision mathematics Q6

46% of students answered this question correctly.

---

18. Roadworks planned by the local council require 13 activities to be completed. The following 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)

VCAA 2021 Exam 2 Networks and decision mathematics Q4a,b

Part **a**: **45%** of students answered this question correctly.
Part **b**: **22%** of students answered this question correctly.

Back to contents

Questions from multiple lessons

Recursion and financial modelling

19. Anna is planning on taking out a $4000 loan to fund a semester exchange in Copenhagen. She plans to repay it in full after one year.

She has been offered five different loans. Their annual interest rates and compounding periods are provided.

- Loan A: 11.0% p.a, compounding quarterly
- Loan B: 11.8% p.a, compounding quarterly
- Loan C: 11.4% p.a, compounding quarterly
- Loan D: 11.2% p.a, compounding weekly
- Loan E: 10.9% p.a, compounding weekly

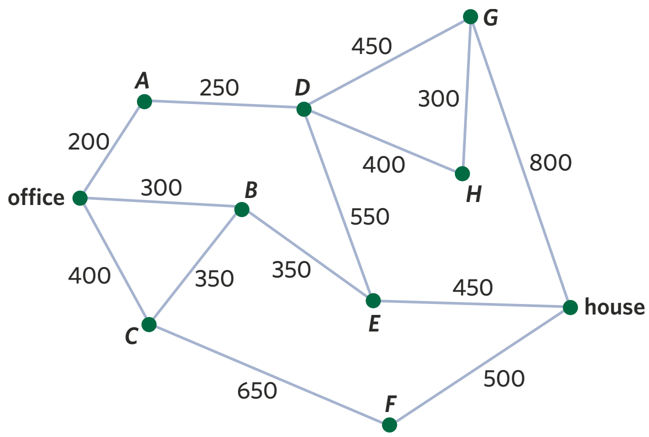
Which loan should Anna choose if she wants to minimise the amount of interest paid?

A. Loan A **B.** Loan B **C.** Loan C **D.** Loan D **E.** Loan E

Adapted from VCAA 2018 Exam 1 Recursion and financial modelling Q19

Networks and decision mathematics

20. The following network shows the distance, in metres, of several streets between Yolanda's office and her house. The intersections of the streets have been labelled from *A* to *H*.



Yolanda wants to work out the shortest path from her office to her house. She could do this using

A. Dijkstra's algorithm.
B. the Hungarian algorithm.
C. Prim's algorithm.
D. a minimum cut.
E. critical path analysis.

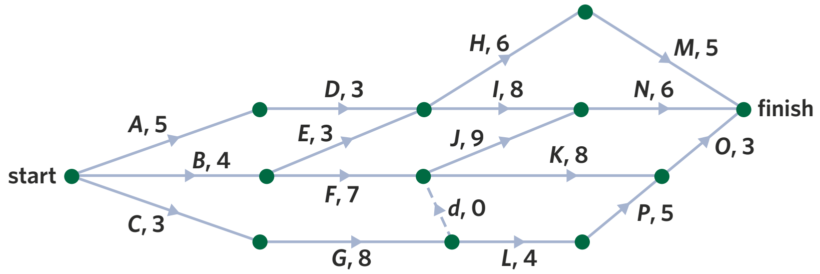
Adapted from VCAA 2018 Exam 1 Networks and decision mathematics Q2

Networks and decision mathematics

21. The tailors at Dior's atelier are constructing an haute couture jumpsuit for Lady Gaga to wear to the 2023 annual MET Gala.

The construction of the jumpsuit requires the completion of 16 activities, labelled from *A* to *P*.

The network to the right shows these activities and each of their completion times, in days.



a. What are the two immediate predecessors of activity *J*? (1 MARK)

b. The minimum completion time for the project is 26 days.

i. This activity network has two critical paths. One of them is *B*–*F*–*J*–*N*. What is the other critical path? (1 MARK)

ii. Determine the float time for activity *K*. (1 MARK)

Adapted from VCAA 2017 Exam 2 Networks and decision mathematics Q4

Back to contents

8J Crashing

STUDY DESIGN DOT POINT

- use of crashing to reduce the completion time of the project or task being modelled

8A 8B 8C 8D 8E 8F 8G 8H 8I 8J

KEY SKILLS

During this lesson, you will be:

- crashing an activity network
- crashing an activity network with restrictions.

KEY TERMS

- Crashing

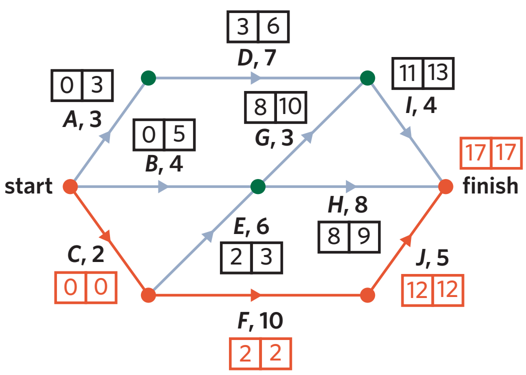
It is often advantageous to complete a project as quickly as possible. Sometimes it is so important that a manager of a project is willing to pay to reduce the completion time. Real life examples of this include hiring more workers, purchasing better equipment or improving technology. However, it is crucial to determine which activities to reduce the duration of, since not all will actually have an impact on a project's completion time.

Crashing an activity network

It is sometimes possible to reduce the time required to complete certain activities, and thereby lower the total completion time of the project.

If the activity that is being altered is on all critical paths of the project, the minimum completion time of the project will also change. This process of reducing the completion time of a project is known as **crashing**. Crashing can cause the critical path to change if activities on the critical path are altered by more than the minimum float time outside of the critical path.

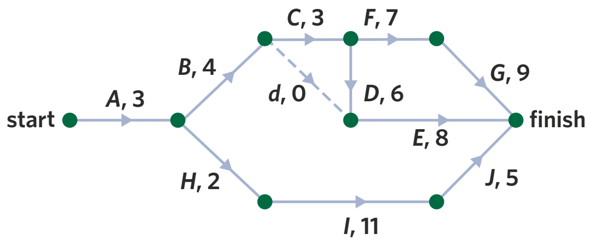
For example, the critical path of the following activity network is currently *C*–*F*–*J*, and the minimum float time outside of the critical path is 1, for activities *E* and *H*.



- If the duration of either activity *F* or *J* is reduced by 1 hour, the minimum completion time of the project will be reduced by 1 hour and there will be a second critical path *C*–*E*–*H*.
- If the duration of either activity *F* or *J* is reduced by 2 hours, the minimum completion time of the project will be reduced by only 1 hour, and *C*–*E*–*H* will be the only critical path.
- Reducing the duration of activity *C* by 2 hours will decrease the minimum completion time of the project by 2 hours, but the critical path will remain the same as the activity is on both *C*–*F*–*J* and *C*–*E*–*H*.

Worked example 1

Claire is part of a study group that is planning to create revision notes on chapter 8 of this textbook. The activity network shows all the lessons within chapter 8. All values are in hours.



Claire asks her friend Klare to help the study group. With Klare's assistance, the study group can reduce the time it will take to complete revision notes for lesson *C* by 2 hours and lessons *B*, *G* and *I* by 3 hours.

a. What is the minimum amount of time it will take the group to complete revision notes for chapter 8 with Klare's help?

Explanation

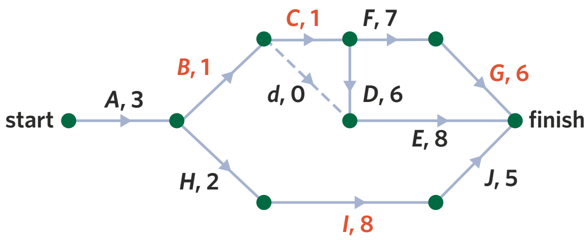
Step 1: Reduce the duration of each of the activities by the maximum number of hours.

Activity *B*: $4 - 3 = 1$

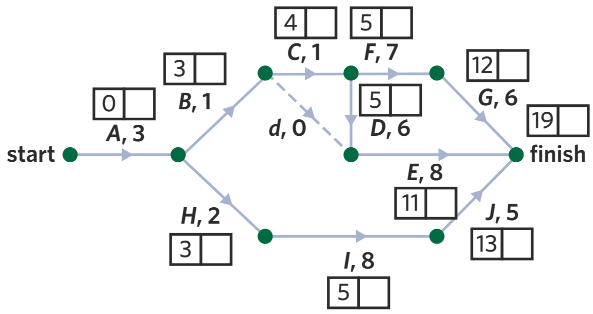
Activity *C*: $3 - 2 = 1$

Activity *G*: $9 - 3 = 6$

Activity *I*: $11 - 3 = 8$



Step 2: Use forward scanning to find the new minimum completion time.



The minimum completion time is the EST of the finish vertex.

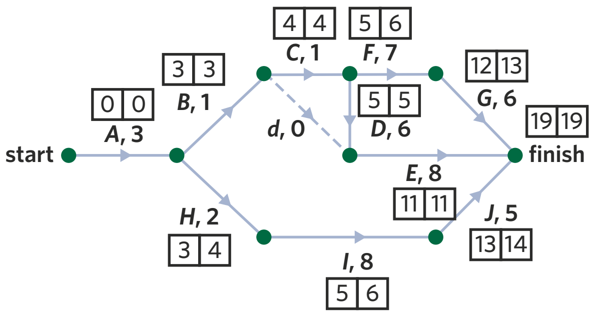
Answer

19 hours

b. What is the critical path of the project?

Explanation

Step 1: Use backward scanning to find the LSTs of each lesson.



Step 2: Find the critical path(s).

Critical activities have an EST and LST that are equal.

Answer

A–*B*–*C*–*D*–*E*

Back to contents

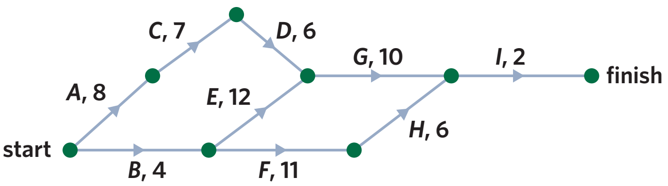
Crashing an activity network with restrictions

Often a cost may be incurred in order to reduce the completion time of an activity. The amount of money spent to reduce the completion time of a project is balanced against the amount of time it will save, and the most efficient strategy is chosen.

For example, if the manager of a project can choose to reduce the length of two different activities, and both reduce the completion time of a project by an equal amount, it is more efficient to reduce the length of the activity that is cheaper.

Worked example 2

A landscaping company is preparing to renovate and landscape a garden. The activities that must be completed for the project are represented in the following activity network. All values are in days.



The manager is willing to employ more staff in order to speed up the completion of the project.

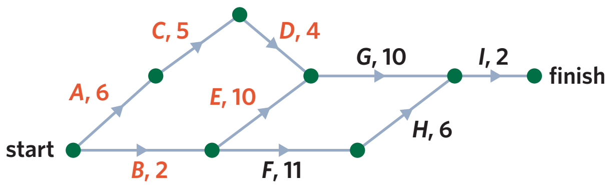
The completion times of activities *A*, *B*, *C*, *D* and *E* can each be reduced by $150 per day, for up to two days.

---

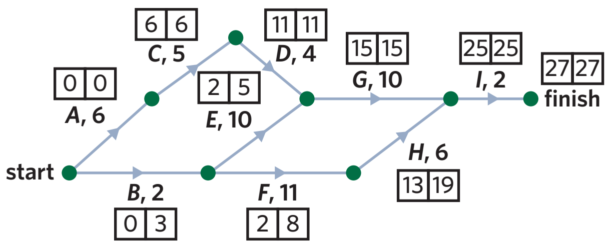
a. What is the minimum cost of reducing the minimum completion time by as much as possible?

Explanation

Step 1: Reduce the duration of each of the activities by the maximum number of days.



Step 2: Complete forward and backward scanning.



Activities *A*, C and *D* have float times of 0.

Activities *B* and *E* have float times of 3.

Step 3: Determine any redundant reduction in the duration of activities.

It is inefficient to pay for the reduction in the duration of activities when there will be float time.

The minimum completion time of the project would not change if the duration of activities *B* and *E* were a total of 3 days longer. Therefore, the duration of one of activity *B* and *E* does not need to be reduced at all, and the duration of the other only needs to be reduced by 1 day.

Step 4: Identify the cost of reducing the minimum completion time by as much as possible.

The duration of activities *A*, *C* and *D* should each be reduced by 2 days and the duration of activity *B* (or *E*) should be reduced by 1 day. This is a total of 7 days.

Each day of reduced time costs $150.

$150 \times 7 = 1050$

Answer

$1050

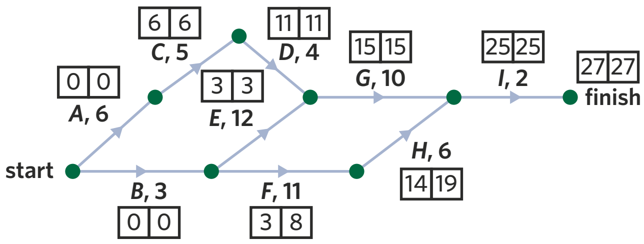
Continues →

Back to contents

b. On top of the current reductions, the manager may also be able to further reduce either the duration of activity *A* by 1 day for \$140, the duration of activity *G* by 1 day for \$175 or the duration of activity *I* by 1 day for \$225. Which activity should the manager reduce the duration of?

Explanation

Step 1: Find the critical path(s) of the project after the initial crashing.



Critical activities have an EST and LST that are equal.

The critical paths are *A*–*C*–*D*–*G*–*I* and *B*–*E*–*G*–*I*.

Step 2: Identify whether activities *A*, *G* and *I* are on all of the critical paths.

Activities *G* and *I* are on both critical paths while activity *A* is not. Hence, reducing the duration of activity *A* would not alter the completion time of the project.

Step 3: Identify the cheaper option.

It costs \$175 to reduce the duration of activity *G* by 1 day and \$225 to reduce the duration of activity *I* by 1 day. Both reductions will have the same effect on the minimum completion time of the project. The cheaper option is activity *G*.

Answer

Activity *G*

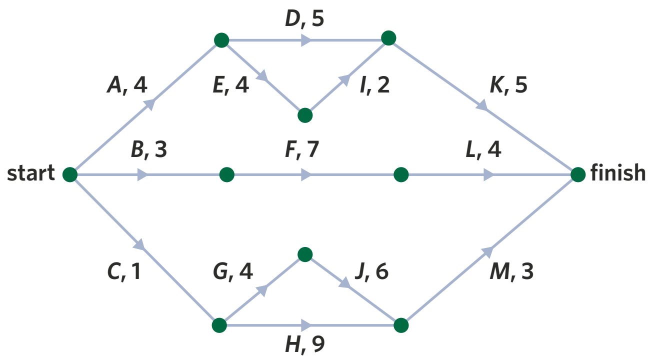
Exam question breakdown

VCAA 2016 Exam 2 Networks and decision mathematics Q3d

A new skateboard park is to be built in Beachton.

This project involves 13 activities, *A* to *M*.

The directed network shows these activities and their completion times in days.



The completion times for activities *E*, *F*, *G*, *I* and *J* can each be reduced by one day.

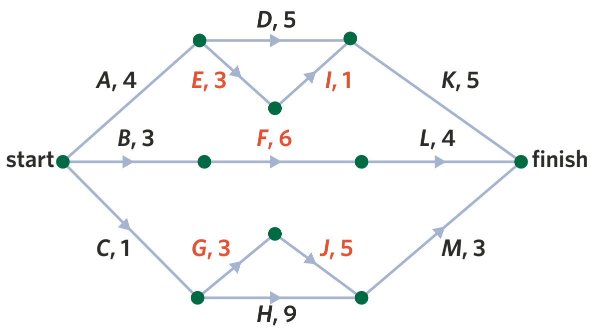
The cost of reducing the completion time by one day for these activities is shown in the following table.

activity	*E*	*F*	*G*	*I*	*J*
cost (\$)	3000	1000	5000	2000	4000

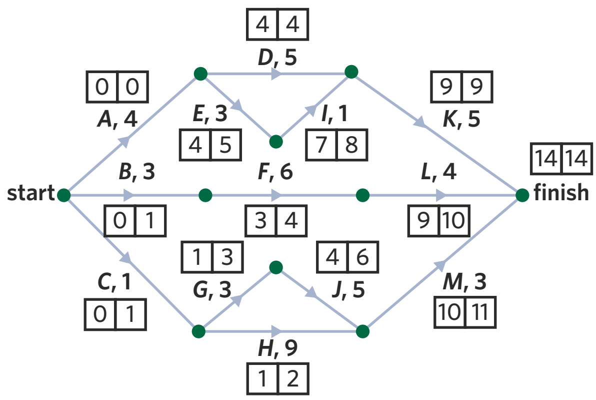
What is the minimum cost to complete the project in the shortest time possible? (1 MARK)

Explanation

Step 1: Reduce the duration of each of the activities by the maximum number of days.



Step 2: Complete forward and backward scanning.



Activities *E*, *F* and *I* have a float time of 1.

Activities *G* and *J* have a float time of 2.

Continues →

Back to contents

Step 3: Determine any redundant reduction in the duration of activities.

The minimum completion time of the project would not change if the duration of activity F was 1 day longer.

The minimum completion time of the project would not change if the duration of activities G and J were a total of 2 days longer.

Therefore, the durations of each of activity F, G and J do not need to be reduced at all.

The minimum completion time of the project would not change if the duration of activities E and I were a total of 1 day longer.

Therefore, the duration of only one of activity E and I needs to be reduced.

Step 4: Identify the cheaper option.

It costs \$3000 to reduce the duration of activity E and \$2000 to reduce the duration of activity I. The cheaper option is activity I.

Answer

\$2000

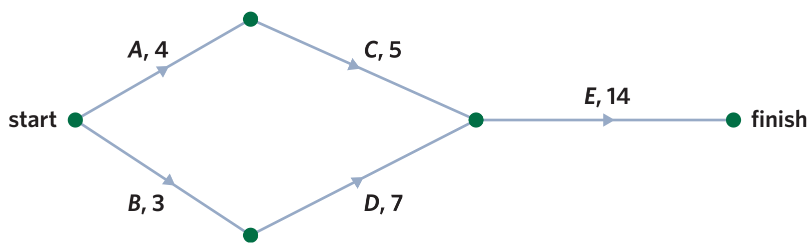
21% of students answered this question correctly.

Many students likely did not recognise that A-D-K becomes a critical path after reducing the duration of activity I. This led to answers of \$10 000 as the three new critical paths could be reduced in duration by a further day. Other students reduced the duration of activity E instead of I, failing to make the most efficient choice.

8J Questions

Crashing an activity network

1. Consider the following activity network. The minimum completion time of the project is 24 days.

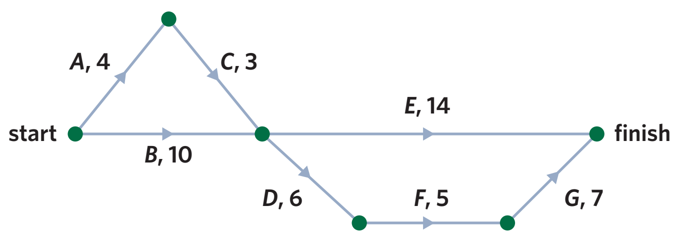


The duration of each activity within the project can be reduced by one day.

To complete this project in 21 days, the minimum number of activities that must have their durations reduced by one day is

A. 2 **B.** 3 **C.** 4 **D.** 5

2. The activity network for a project is shown. All values are in days.



a. Which one of the following activities **cannot** be reduced by 1 day without altering the completion time of the project?

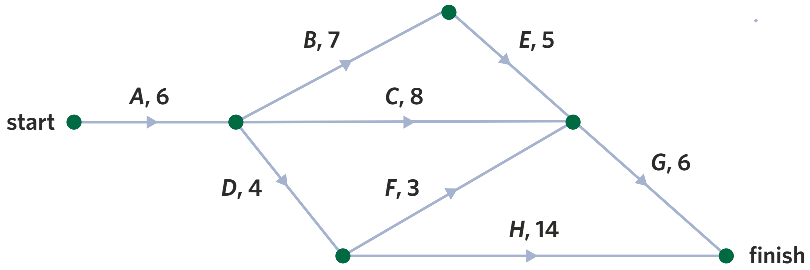
Back to contents

The durations of activities *B* and *G* can be reduced by up to 5 days and the durations of activities *C* and *E* can be reduced by up to 2 days.

b. If all are reduced as much as possible, what would be the new minimum completion time?

c. If all are reduced as much as possible, what would be the new critical path(s)?

3. In order to be more efficient, Alyssa, a Year 12 student, has timed all of her activities as she gets ready for school in the morning. She is able to multitask by doing numerous activities at once. The activity network for her morning activities is shown. All values are in minutes.



a. Find the critical path(s).

Alyssa is able to cut down the time it takes to complete activity *B*. She can reduce its duration by 2 minutes.

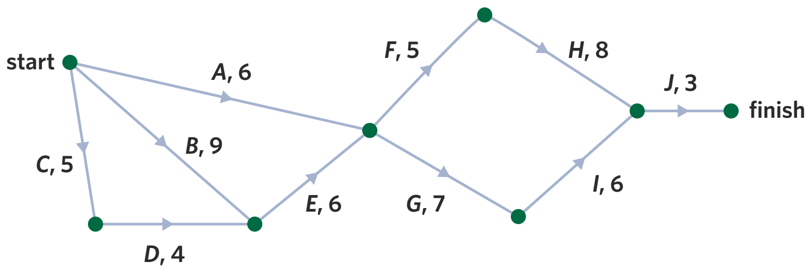
b. Will Alyssa be able to get ready for school faster? If not, explain why. If so, how much time has she saved?

c. What is the new float time of activity *B*?

d. Find the new critical path(s).

Crashing an activity network with restrictions

4. Consider the following activity network. All values are in minutes.



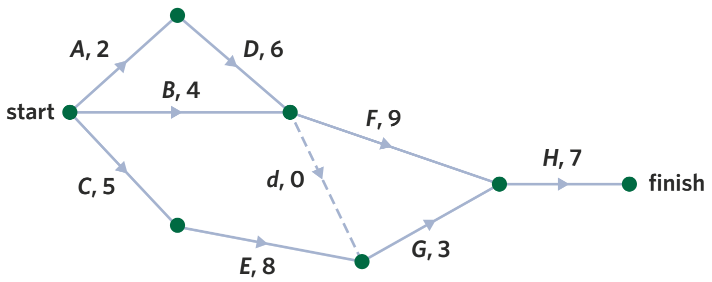
The durations of activities *E* and *I* can each be reduced by one minute for \$200.

The durations of activities *B*, *D* and *F* can each be reduced by one minute for \$250.

If the durations of only three activities can be reduced, it would be best to reduce the durations of activities

A. *B*, *D* and *E*. **B.** *B*, *D* and *F*. **C.** *E*, *D* and *I*. **D.** *E*, *F* and *I*.

5. The organisers of Falls Music Festival have drawn an activity network for the setup of their festival. All values are in days.



Back to contents

a. Find the critical path(s).

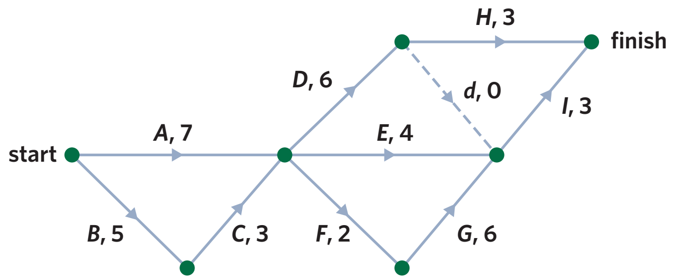
The organisers are able to employ more workers to reduce the overall setup time. They can reduce the duration of activity D for \$1800 per day and activity F for \$1300 per day. The duration of the activities can be reduced by a maximum of four days each.

b. What is the minimum cost of reducing the setup time as much as possible?

c. How long would it take the organisers of Falls Music Festival to set up the festival after employing the extra workers?

d. Find the new critical path(s).

6. The following activity network shows the steps required to build an automatic farm on Minecraft.



The current minimum completion time is 19 minutes.

Activities A and B can each be reduced for 9 emeralds per minute.

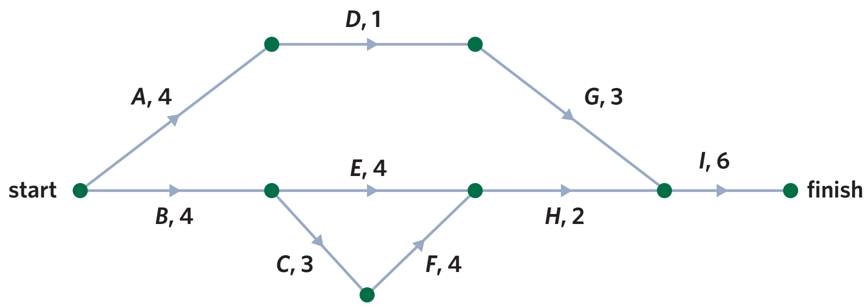
Activities C and D can each be reduced for 7 emeralds per minute.

Each activity can be reduced by a maximum of 2 minutes.

What is the minimum additional cost of completing the automatic farm in 16 minutes?

Joining it all together

7. There are plans for a new basketball court to be built in West Melbourne. Nine activities have been identified for this project. The following activity network shows the activities and their completion times in weeks.



The builders are able to speed up the project. Activities A, C, E, F and G can be reduced in completion time at an additional cost. The maximum reduction in time for each of the five activities is 2 weeks.

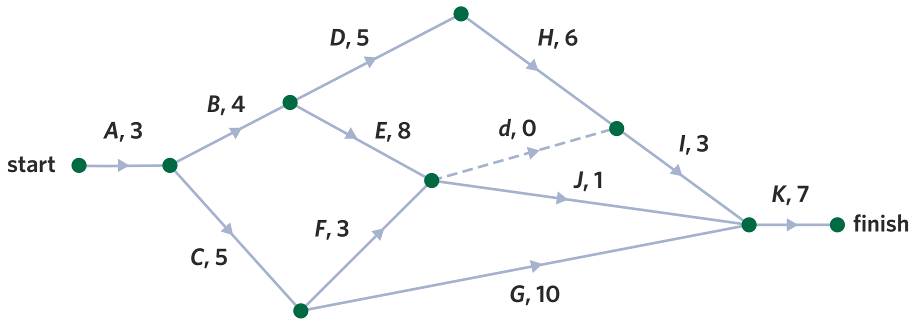
a. Which of these activities, if their durations were reduced individually, would **not** result in an earlier completion of the project?

b. Determine the minimum time, in weeks, that the project can be completed now that certain activities can be reduced in duration.

c. The cost of reducing the time of each activity is \$5000 per week. Determine the minimum additional cost for completing the project in this reduced time.

Back to contents

8. Hermione is planning on brewing a batch of Polyjuice Potion and has created an activity network for the project. Each activity is in hours.



a. What is the minimum brewing time for the Polyjuice Potion?

b. Find the critical path(s).

Hermione is able to buy special ingredients from Diagon Alley that speed up the time required for certain activities. Each special ingredient can reduce the duration of one of the activities *D*, *E*, *F*, *G*, *H* and *I* by exactly 2 hours.

c. If Hermione purchased all the special ingredients, what would be the new minimum brewing time for the Polyjuice Potion?

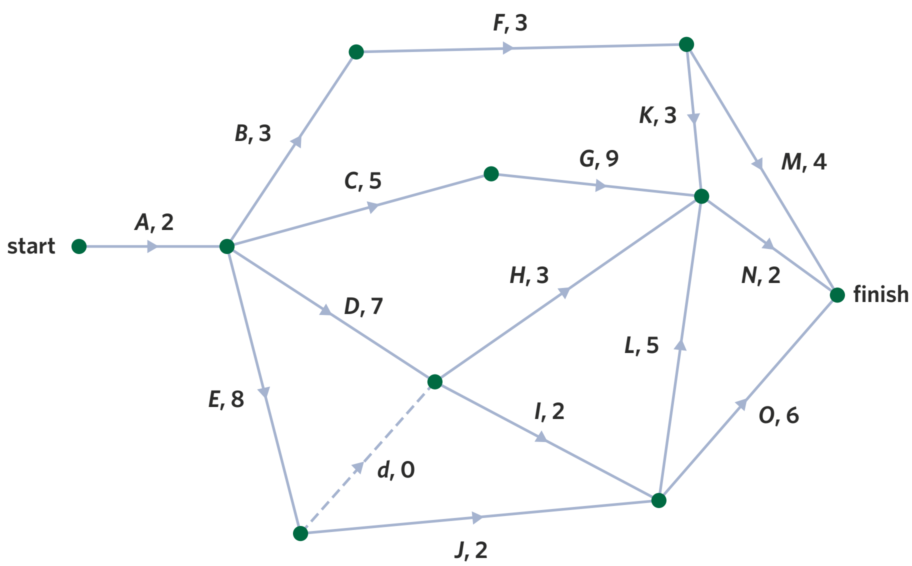
d. The cost of each special ingredient is 30 galleons. What is the least Hermione can spend while reducing the brewing time as much as possible?

Exam practice

9. The rides at a theme park are set up at the beginning of each holiday season.

This project involves activities *A* to *O*.

The directed network shows these activities and their completion times in days.



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.

a. What is the minimum number of days in which the project could now be completed? (1 MARK)

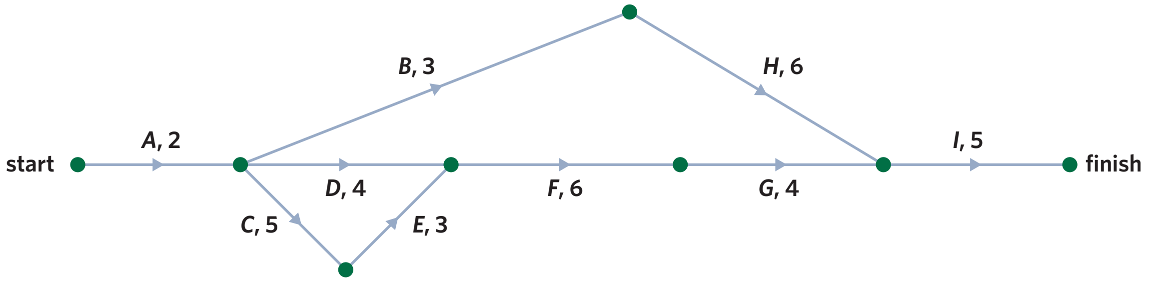
b. What is the minimum cost of completing the project in this time? (1 MARK)

VCAA 2017 Exam 2 Networks and decision mathematics Q4ci,ii

Part **a**: **35%** of students answered this question correctly.
Part **b**: **15%** of students answered this question correctly.

10. The following directed network 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.

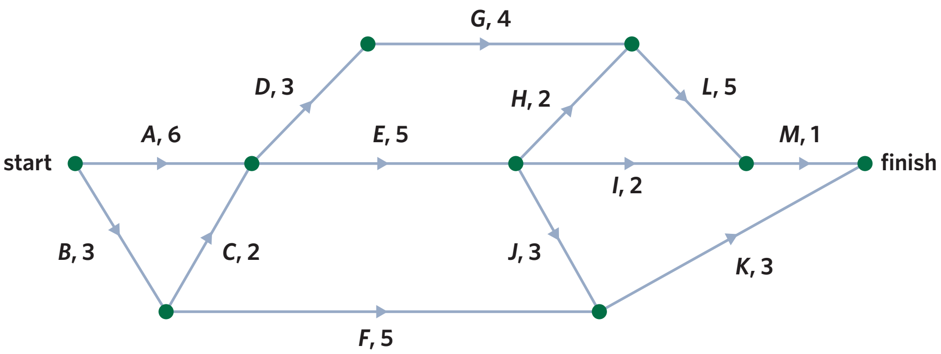
VCAA 2020 Exam 1 Networks and decision mathematics Q10

24% of students answered this question correctly.

---

11. Roadworks planned by the local council require 13 activities to be completed.

The following network shows these 13 activities and their completion times in weeks.



It is possible to reduce the completion time for activities A, E, F, L and K.

The reduction in completion time for each of these five activities will incur an additional cost.

The table shows the five activities that can have their completion time reduced and the associated weekly cost, in dollars.

activity	A	E	F	L	K
weekly cost ($)	140 000	100 000	100 000	120 000	80 000

Back to contents

The completion time for each of these five activities can be reduced by a maximum of two weeks.

The overall completion time for the roadworks can be reduced to 16 weeks.

What is the minimum cost, in dollars, of this change in completion time? (1 MARK)

VCAA 2021 Exam 2 Networks and decision mathematics Q4c

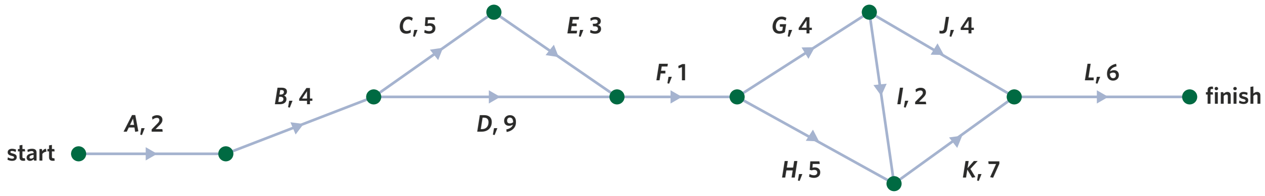
9% of students answered this question correctly.

---

12. Fencedale High School is planning to renovate its gymnasium.

This project involves 12 activities, A to L.

The directed network shows these activities and their completion times, in weeks.



The minimum completion time for the project is 35 weeks.

It is possible to reduce the completion time for activities C, D, G, H and K by employing more workers.

The reduction in completion time for each of these five activities will incur an additional cost to the school.

The table shows the five activities that can have their completion times reduced and the associated weekly cost, in dollars.

activity	C	D	G	H	K
weekly cost ($)	3000	2000	2500	1000	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.

Complete the following table, showing the reductions in individual activity completion times that would achieve this. (2 MARKS)

activity	C	D	G	H	K
reduction in completion time (0, 1 or 2 weeks)					

VCAA 2019 Exam 2 Networks and decision mathematics Q3e

The average mark on this question was **0.2**.

Questions from multiple lessons

Recursion and financial modelling

13. Seth purchased a TI-Nspire CX CAS Graphing Calculator for $286.

After three years, he sold the calculator for its estimated value of $58 on Melbourne Trade.

On average, Seth pressed the ‘enter’ button 643 times per school term.

The value of the calculator was depreciated using a unit-cost method of depreciation.

The depreciation in the value of the TI-Nspire CX CAS Graphing Calculator, per press of the ‘enter’ button, is closest to

A. 2 cents. **B.** 3 cents. **C.** 4 cents. **D.** 10 cents. **E.** 12 cents.

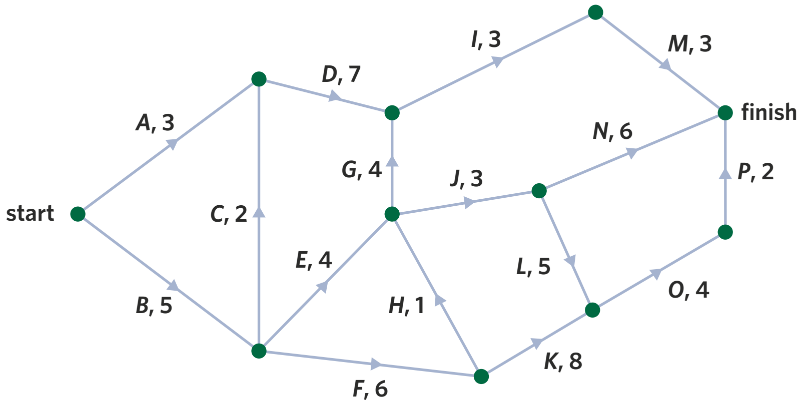
Adapted from VCAA 2017 Exam 1 Recursion and financial modelling Q21

Back to contents

Networks and decision mathematics

14. Sophie is making her renowned sticky date pudding.

The completion times, in minutes, for the activities required to make the sticky date pudding are shown in the following activity network.



What is the earliest starting time for activity *O*?

A. 16 minutes

B. 17 minutes

C. 18 minutes

D. 19 minutes

E. 20 minutes

Adapted from VCAA 2017 Exam 1 Networks and decision mathematics Q4

Data analysis

15. In a large population of high school schools, the number of students per school is approximately normally distributed with a mean of 360 students and a standard deviation of 43 students.

a. Using the 68–95–99.7% rule, determine the percentage of schools expected to have less than 446 students. (1 MARK)

b. Using the 68–95–99.7% rule, determine the number of schools expected to have less than 317 students in a sample of 2000 schools. (1 MARK)

c. The standardised number of students in a school is $z = -3.4$. Determine the actual number of students in this school. Round to the nearest whole number. (1 MARK)

Adapted from VCAA 2017 Exam 2 Data analysis Q1b,c

Back to contents

ANSWERS

CONTENTS

Image: ChristianChan/Shutterstock.com

Back to contents

1A Types of data

Classifying data as categorical or numerical

1. D

2. A

3. **a.** Numerical **b.** Categorical

Classifying categorical data as nominal or ordinal

4. C

5. B

6. **a.** Nominal **b.** Ordinal

Classifying numerical data as discrete or continuous

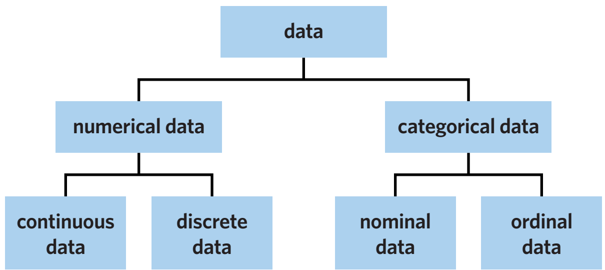
7. C

8. D

9. **a.** Discrete **b.** Continuous

Joining it all together

10.



11. **a.** Nominal **b.** Discrete

c. Continuous **d.** Continuous

e. Nominal **f.** Ordinal

g. Nominal **h.** Discrete

i. Nominal **j.** Ordinal

12. **a.** *serve speed* (km/h)

b. *number of tennis racquets used*

13. B

Exam practice

14. **Explanation**

Step 1: Identify the variables in the question.

The table has presented two variables, *day number* and *minimum temperature*.

Step 2: Classify the variable *day number*.

The variable *day number* acts as a unique identifier and each number represents a different day. As such, *day number* is a categorical variable.

Since each day comes after the previous one, it can also be sorted into ascending or descending order. Therefore, *day number* can be further classified as an ordinal variable.

Step 3: Classify the variable *minimum temperature* (°C).

The variable temperature is measured in degrees Celsius. As such, this is a numerical variable.

Since numerical variables cannot be ordinal, *minimum temperature* is not an ordinal variable.

Answer

day number

A number of students answered 'neither' because they did not recognise that a variable can consist of numeric responses, without necessarily being a numerical variable. In this case, *day number* is not a variable that is counted or measured.

15. **Explanation**

Step 1: Classify the variable *colour*.

The variable *colour* is a qualitative property that can have a variety of categories, such as blue, teal, or grey. As such, it is a categorical variable.

Since categorical variables cannot be discrete, *colour* is not a discrete variable.

Step 2: Classify the variable *name of species*.

The variable *name of species* is a qualitative property that can have a variety of categories, such as ladybug, butterfly, or beetle. As such, it is a categorical variable.

Since categorical variables cannot be discrete, *name of species* is not a discrete variable.

Step 3: Classify the variable *number of wings*.

The variable *number of wings* is a quantitative property that is counted and can only consist of whole number values. As such, it is a numerical variable that is discrete.

Step 4: Classify the variable *body length* (in millimetres).

The variable *body length* (in millimetres) is a quantitative property that is measured and can consist of any value on a continuous scale. As such, it is a numerical variable that is continuous, not discrete.

Step 5: Classify the variable *body weight* (in milligrams).

The variable *body weight* (in milligrams) is a quantitative property that is measured and can consist of any value on a continuous scale. As such, it is a numerical variable that is continuous, not discrete.

Answer

A

A number of students incorrectly chose options B and C. This was likely because they had classified the variables *body length* (in millimetres) and *body weight* (in milligrams) as discrete, rather than continuous. However, since both variables are measured in millimetres and milligrams, they can consist of any value on a continuous scale.

16. **Explanation**

Step 1: Identify the variables in the question.

The table has four variables, *athlete number*, *high jump* (metres), *shot-put* (metres) and *javelin* (metres).

Step 2: Classify the variable *athlete number*.

The variable *athlete number* acts as a unique identifier and each number represents a different athlete. As such, this is a categorical variable.

Step 3: Classify the variable *high jump* (metres).

The variable *high jump* (metres) is a distance that is measured. This is a quantitative property. As such, this is a numerical variable.

Step 4: Classify the variable *shot-put* (metres).

The variable *high jump* (metres) is a distance that is measured. This is a quantitative property. As such, this is a numerical variable.

Back to contents

Step 5: Classify the variable *javelin* (metres).

The variable *javelin* (metres) is a distance that is measured. This is a quantitative property. As such, this is a numerical variable.

Answer

3 numerical variables

A number of students answered 45 by counting the number of values instead of the number of unique numerical variables.

17. Explanation

Step 1: Classify the variable *number of moths* (less than 250, 250–500, more than 500).

The variable *number of moths* has three categories, 'less than 250', '250–500' and 'more than 500'. As such, this is a categorical variable.

These categories can be sorted into ascending or descending order. Therefore, the variable *number of moths* (less than 250, 250–500, more than 500) can be further classified as an ordinal variable.

Step 2: Classify the variable *trap type* (sugar, scent, light).

The variable *trap type* has three categories, 'sugar', 'scent' and 'light'. As such, this is a categorical variable.

These categories cannot be sorted into a logical order. Therefore, *trap type* (sugar, scent, light) can be further classified as a nominal variable.

Answer

E

36% of students incorrectly chose option C. Many assumed that the variable *number of moths* (less than 250, 250–500, more than 500) was numerical because the categories included numbers. However, a variable can consist of numeric responses, without necessarily being a numerical variable.

Questions from multiple lessons

18. D **19.** C **20. a.** 55 cars **b.** 57.1%

1B Displaying and describing categorical data

Constructing frequency tables

1. C

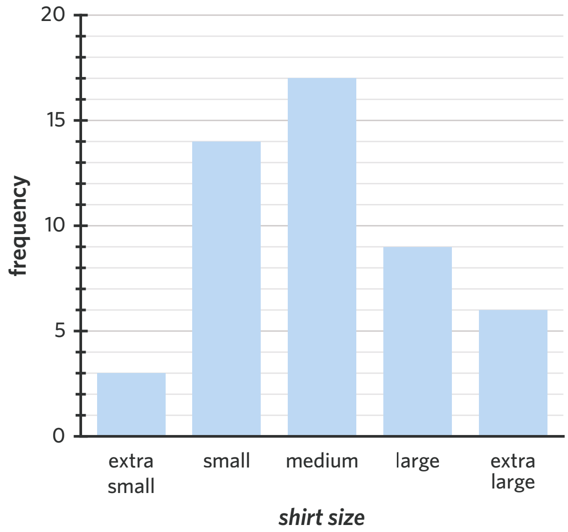
2.

favourite type of pasta	frequency	
	number	**%**
penne	7	35
spaghetti	6	30
fettuccine	5	25
macaroni	2	10
total	20	100

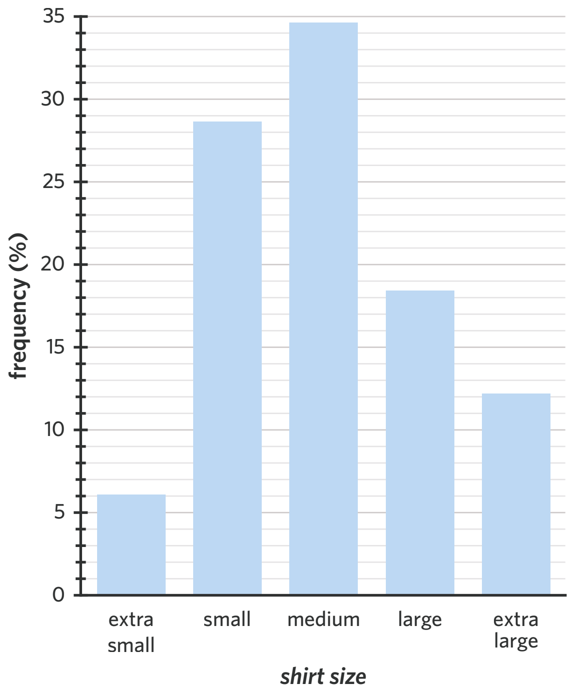
Constructing bar charts

3. D

4. a.



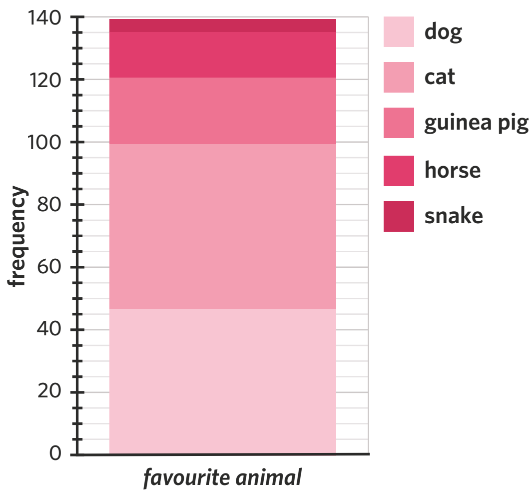
b.



Constructing segmented bar charts

5. B

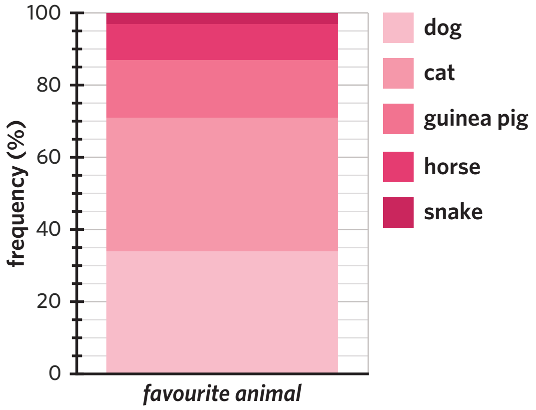
6. a.



1B ANSWERS

Back to contents

b.



Describing the distribution of categorical data

7. B

8. The brands of 50 cars were recorded as they entered a car park. All the cars were either 'Holden', 'Ford', or 'Toyota'. The most commonly occurring brand of car was Toyota, accounting for 60% of all cars. The next most commonly occurring brand was Holden, representing 30%. Finally, the last 10% of the cars were Fords.

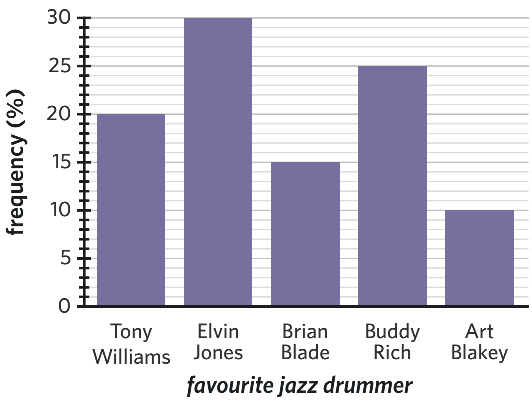
Joining it all together

9. A

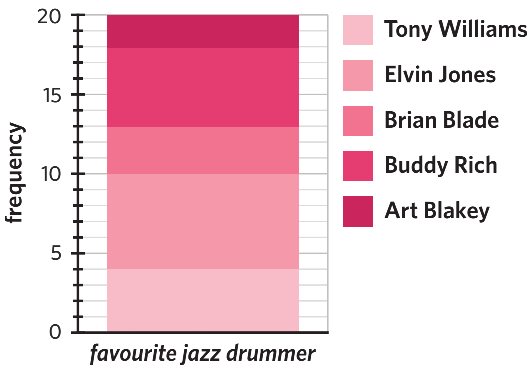
10. a.

favourite jazz drummer	frequency	
	number	%
Tony Williams	4	20
Elvin Jones	6	30
Brian Blade	3	15
Buddy Rich	5	25
Art Blakey	2	10
total	20	100

b.

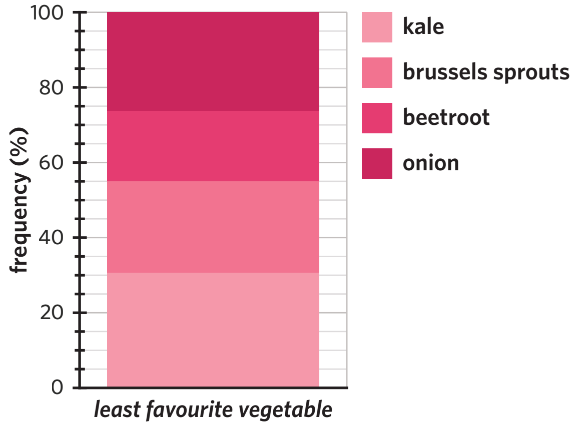


c.



d. 20 musicians were asked who their *favourite jazz drummer* was. The most common response was Elvin Jones, accounting for 30% of people surveyed. Buddy Rich also accounted for 25% of the votes. The least popular drummer of the five was Art Blakey, who accounted for 10% of the votes.

11.



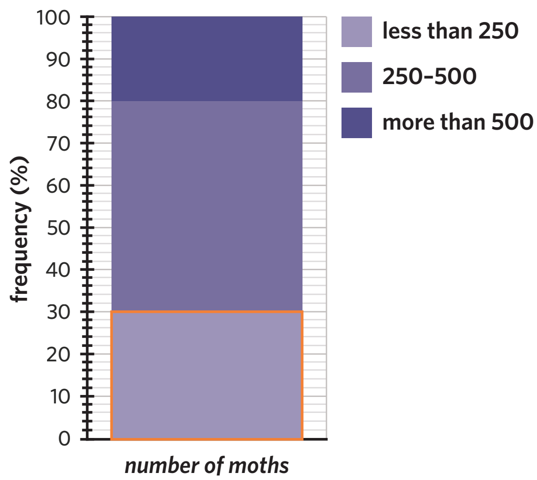
12.

favourite TV show	frequency	
	number	%
The Bachelor	10	25
Squid Game	24	60
Peaky Blinders	6	15
total	40	100

Exam practice

13. Explanation

Step 1: Identify the relevant segment.



Step 2: Interpret the relevant segment.

30% of sugar traps caught less than 250 moths.

Step 3: Calculate the number of sugar traps.

There were 300 sugar traps. Find 30% of 300.

$300 \times 0.3 = 90$

Answer

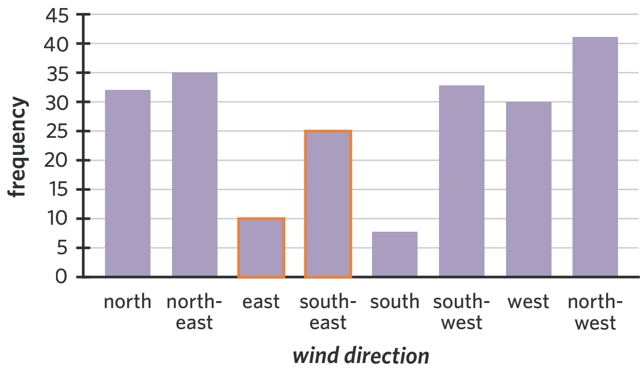
B

21% of students incorrectly chose option A. These students did not convert the percentage of sugar traps to the number of sugar traps.

Back to contents

14. Explanation

Step 1: Identify the relevant columns.



Step 2: Interpret the relevant columns.

Looking at the vertical axis, the *wind direction* was east on 10 days and south-east on 25 days.

This means there were 10 + 25 = 35 days on which the *wind direction* was east or south-east.

Step 3: Represent this as a percentage.

Data was collected over 214 days, and the *wind direction* was east or south-east on 35 of these days.

$\frac{35}{214} = 0.163... = 16.3...\%$

Answer

B

19% of students incorrectly chose option E. These students did not convert the number of days to a percentage.

Questions from multiple lessons

15. D **16.** C

17. a. $M_0 = 1000, \quad M_{n+1} = 1.04 \times M_n$

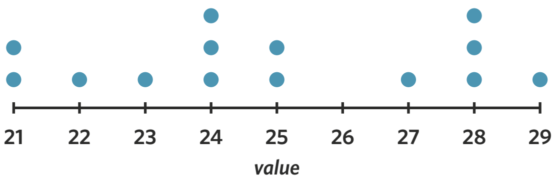
b. 11.81%

1C Displaying numerical data

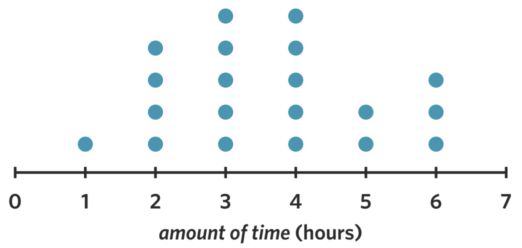
Displaying data using dot plots

1. B

2.



3. a.



b. 4 students **c.** 10 students **d.** 25%

Displaying data using stem plots

4. D

5. Key: 4 | 0 = 40

4	0 1 4 6 7
5	3 4 4 8
6	3
7	1 4 5
8	2 6 7 9 9
9	1 9

6. a. 7 students **b.** 13 students **c.** 38 minutes

7. a. Key: 17 | 0 = 170 cm

17	0 1 4
17	5 6 7 8 9
18	0 1 2 3 4 4
18	5 7 8 9
19	0
19	5

b. 5 times **c.** 14 times **d.** 40%

Constructing grouped frequency tables

8. a. C **b.** 87.5%

c. 1.0–<1.1 kg

9. a.

maximum flying height (cm)	frequency	
	number	%
14.5–<15.0	4	25.00
15.0–<15.5	5	31.25
15.5–<16.0	3	18.75
16.0–<16.5	4	25.00
total	16	100.00

b. 25%

Displaying data using histograms

10. B

11. a. i. 2 students

ii. 1 student

iii. 9 students

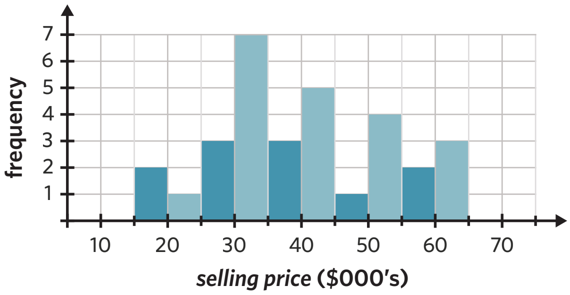
b. i. 1 student

ii. 10 students

ii. 6 students

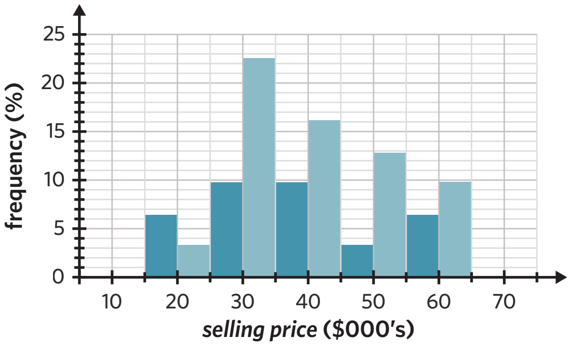
c. 50–<55 seconds

12. a.



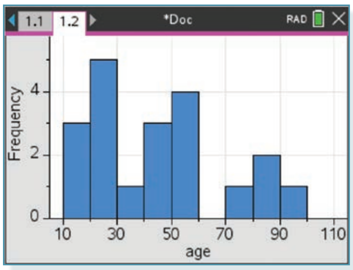
b. 13 NFTs

c.

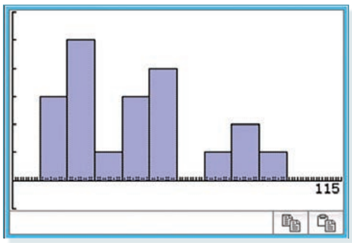


d. 48%

13. a. TI-Nspire



Casio ClassPad



b. 10 is the largest round number less than the minimum age in the dataset.

c. 50 years **d.** 20–<30 years

Joining it all together

14. a. Stem plot

Note: A histogram can also display this data, but with a sample size of 14, a stem plot is more appropriate.

b. **Key:** 0 | 8 = 8 °C

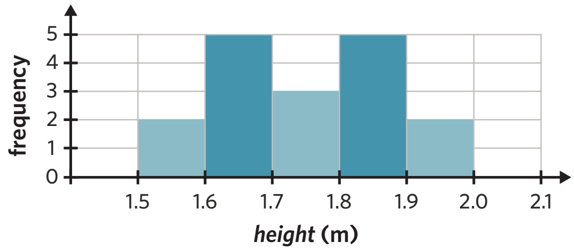
0	8
1	0 0 2 2 5 6 8
2	1 8
3	1 2 5
4	2

c. 3 days **d.** 50%

15. a.

height (m)	frequency	
	number	%
1.5–<1.6	2	11.8
1.6–<1.7	5	29.4
1.7–<1.8	3	17.6
1.8–<1.9	5	29.4
1.9–<2.0	2	11.8
total	17	100.00

b.



c. 10 members **d.** 1.51 m

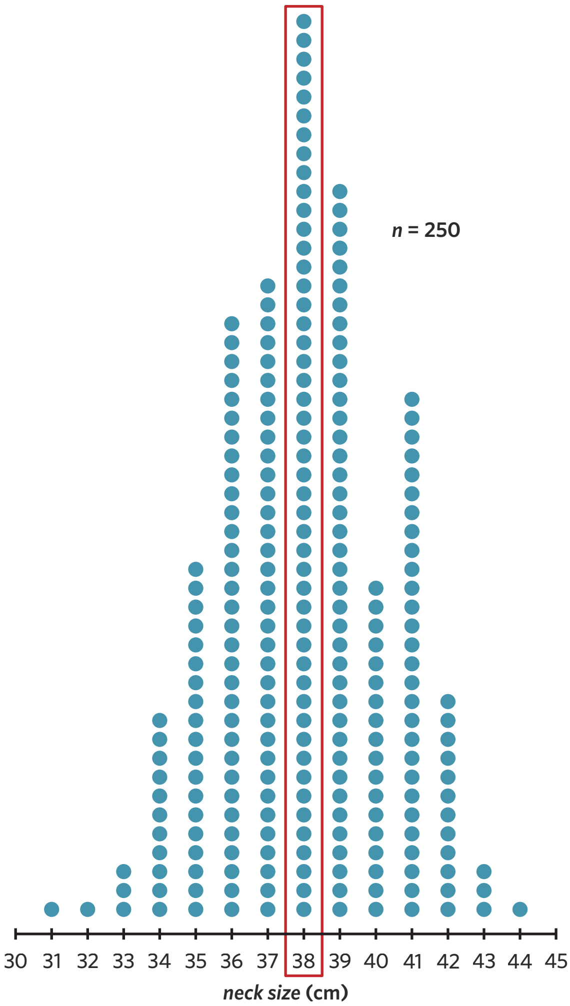
e. 1.6–<1.7 m, 1.8–<1.9 m

f. No. The mode of the original data is not a good measure of centre as it is also the minimum value of the data set.

Exam practice

16. Explanation

The modal *neck size* will be the value on the axis with the most dots.



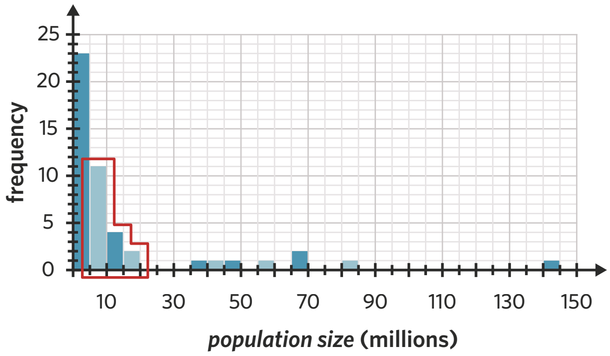
Back to contents

Answer

38 cm

17. Explanation

Sum the frequencies of the intervals from a *population size* of 5 million to a *population size* of 20 million.



11 + 4 + 2 = 17

Answer

B

18. Explanation

Count the most frequently occurring leaf.

Note: Leaves in different stems are counted separately.

temperature **Key:** 2 | 2 = 2.2 °C

2	2	2	4	4									
2	5	7	8	8	8	8	8	8	9	9	9	9	
3	1	2	3	3	4	4	4						
3	5	6	7	7	7	7							
4	1												

The leaf '8' on the stem '2' represents 2.8 °C.

Answer

A

19. Explanation

Step 1: Use the data in the dot plot to construct a grouped frequency table.

The percentage frequency is not necessary.

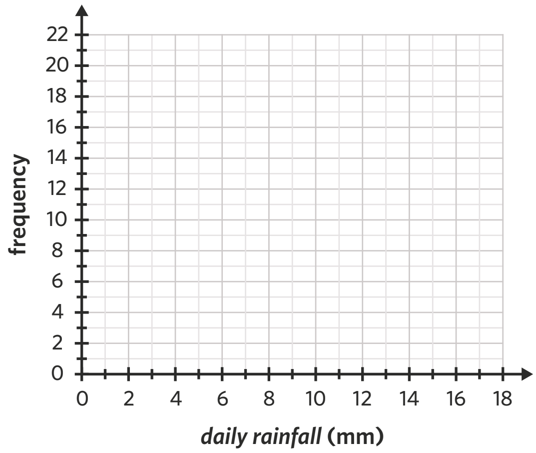
The intervals should have a width of 2 and start at 0.

daily rainfall (mm)	frequency
0–<2	21
2–<4	3
4–<6	0
6–<8	3
8–<10	0
10–<12	0
12–<14	1
14–<16	0
16–<18	2
total	30

Step 2: Construct a set of axes for the histogram.

Label the horizontal axis '*daily rainfall* (mm)'.
The horizontal axis needs to range from 0 to at least 18.

Label the vertical axis 'frequency'. The vertical axis needs to range from 0 to at least 21.

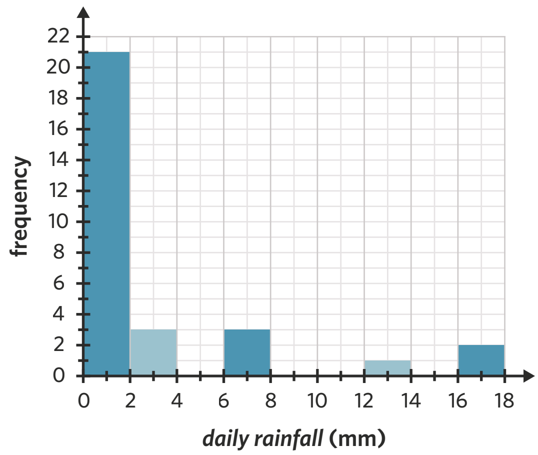


Step 3: Draw a column for each interval.

The height of the column represents the frequency of data values within the corresponding interval.

Histograms do not have spaces between columns.

Answer



Many students incorrectly drew the histogram with interval widths of 1. The question specifically asked for interval widths of 2.

Questions from multiple lessons

20. D **21.** C

22. a. 3

b. 1

c. aardvark, dodo

1C ANSWERS

Back to contents

1D Log scales and graphs

Calculating logarithmic values

1. C

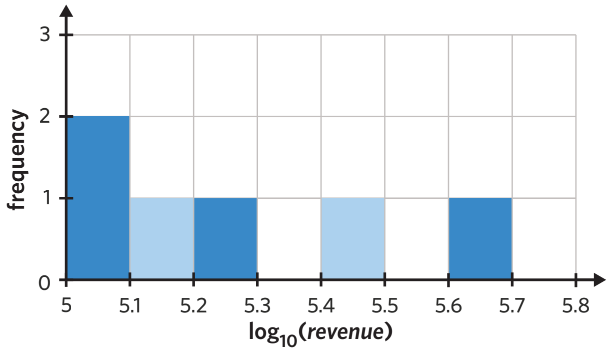
2. a. −1.30 b. −0.30 c. 0.70 d. 1.70
   e. 3.58 f. 2.58 g. 1.58 h. 0.58

3. a. 0.32 b. 3.16 c. 31.62

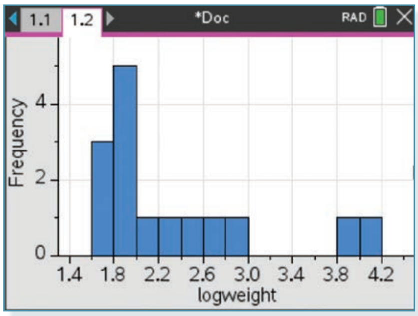
Displaying data using a logarithmic scale

4. D

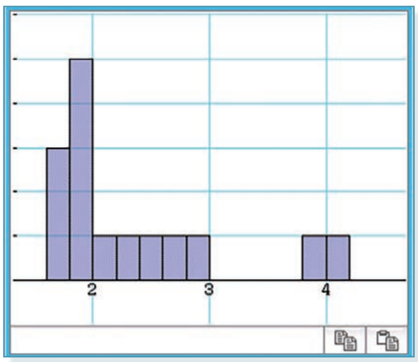
5.



6. TI-Nspire



Casio ClassPad



Interpreting data displayed using a logarithmic scale

7. a. C b. C

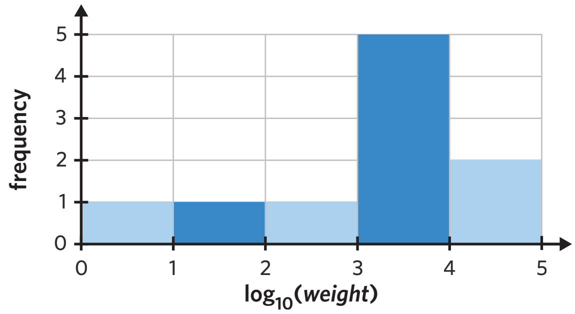
8. a. 12 trees b. 42.86%

Joining it all together

9. a. 3981 kg b. 3.4–<3.6

10. a. 0.18 b. 19 aliens
    c. 39.61 m

11. a.



b. 5 vehicles weigh between 1000 kg and 10 000 kg.

Exam practice

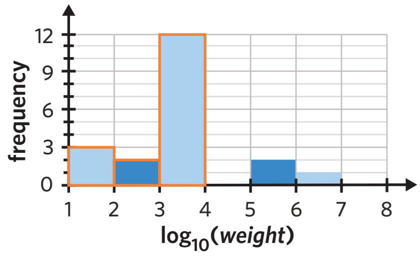
12. **Explanation**

Step 1: Calculate the appropriate log value.

$\log_{10}(10\,000) = 4$

This means that on a $\log_{10}$ scale, 10 000 is plotted as 4.

Step 2: Sum the columns less than 4.



$3 + 2 + 12 = 17$

Step 3: Represent this as a percentage of the total.

$\frac{17}{20} = 0.85$

$= 85\%$

Answer

E

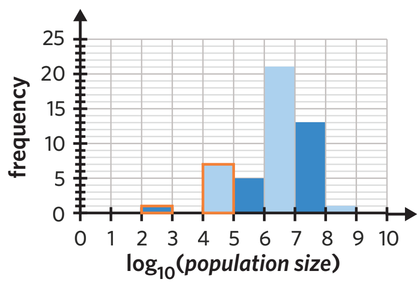
13. **Explanation**

Step 1: Calculate the appropriate log value.

$\log_{10}(100\,000) = 5$

This means that on a log scale, 100 000 is plotted as 5.

Step 2: Sum the columns less than 5.



$7 + 1 = 8$

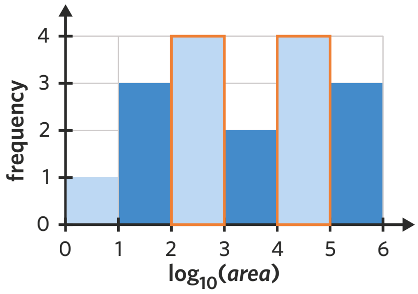
Answer

D

Back to contents

14. **Explanation**

Step 1: Identify the columns with the highest frequency.



This occurs between log values of 2 and 3, as well as 4 and 5.

Step 2: Use the $\log_{10}$ values to calculate the *area*.

$\log_{10}(a) = 2$

$a = 10^2 = 100$

$\log_{10}(b) = 3$

$b = 10^3 = 1000$

The first modal category is between 100 km^2 and 1000 km^2.

$\log_{10}(c) = 4$

$c = 10^4 = 10\ 000$

$\log_{10}(d) = 5$

$d = 10^5 = 100\ 000$

The second modal category is between 10 000 km^2 and 100 000 km^2.

Answer

E

A common error with this type of question was ignoring the $\log_{10}$ scale and not converting the log values to actual values. This would have led to an incorrect answer of B.

Questions from multiple lessons

15. E **16.** E

17. **a.** $87 500

b. $90 475

c. 4 years

1E The five-number summary and boxplots

Calculating the five-number summary

1. C **2.** 6, 14.5, 24, 31.5, 35

3. 1, 2, 3, 4, 6 **4.** 7, 12, 19, 25, 41

Calculating the range and interquartile range

5. D

6. **a.** $7 **b.** $3

7. 1.3 seconds

Identifying outliers

8. A **9.** B, C, D, E

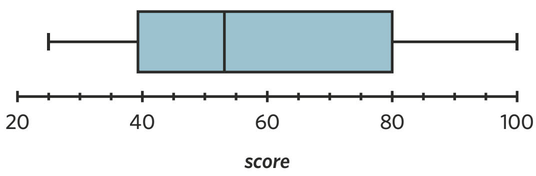
10. **a.** 15

b. 19, 20, 30

Constructing and interpreting boxplots

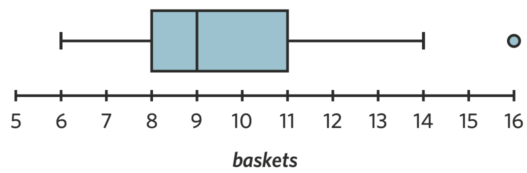
11. A

12. **a.**



b. A **c.** B

13. **a.**



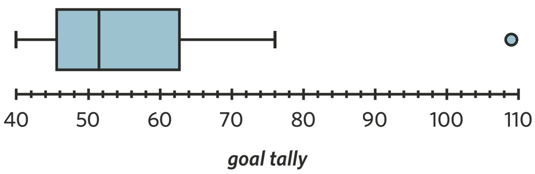
b. B **c.** C

Joining it all together

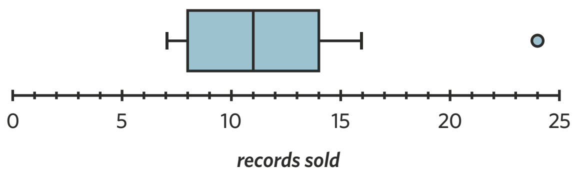
14. **a.** 40, 45.5, 51.5, 62.5, 109

b. 17 **c.** 109

d.



15. **a.**

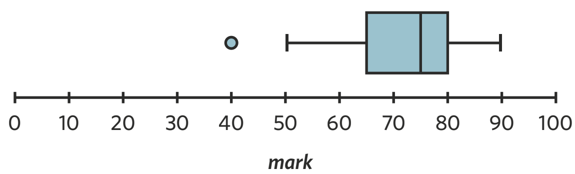


b. 24 **c.** 7, 8, 11, 14, 24

16. **a.** **i.** 25% **ii.** 50%

iii. 50% **iv.** 75%

b. The minimum value of the original boxplot, 40, should be an outlier.



Back to contents

Exam practice

17. Explanation

Step 1: Determine the lower half and upper half of the data set.

$n = 23$

Position of median: $\frac{23 + 1}{2} = 12$

Key: 4 | 2 = 42 $\quad n = 23$

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

Lower half:

40 41 44 44 52 57 59 59 59 65 66

Upper half:

68 69 69 70 70 75 76 77 78 85 89

Step 2: Determine the value of Q_1 and Q_3.

Q_1 is the median of the lower half of the data set.

40 41 44 44 52 57 59 59 59 65 66

$Q_1 = 57$

Q_3 is the median of the upper half of the data set.

68 69 69 70 70 75 76 77 78 85 89

$Q_3 = 75$

Step 3: Calculate the IQR.

$IQR = Q_3 - Q_1$

$= 75 - 57$

$= 18$

Answer

C

18. Explanation

Step 1: Determine the IQR.

The value of the left border of the box is Q_1 and the value of the right border of the box is Q_3.

$Q_1 = 8$

$Q_3 = 11$

$IQR = Q_3 - Q_1$

$= 11 - 8$

$= 3$

Step 2: Calculate the upper fence.

$upper\ fence = Q_3 + (1.5 \times IQR)$

$= 11 + (1.5 \times 3)$

$= 15.5$

Answer

15.5 °C

Some students read Q_3 off the boxplot as 12 °C. This led to an incorrect answer of 18 °C.

19. Explanation

Step 1: Check for any outliers.

Calculate the IQR.

$IQR = Q_3 - Q_1$

$= 39 - 36$

$= 3$

Calculate the lower fence.

$lower\ fence = Q_1 - (1.5 \times IQR)$

$= 36 - (1.5 \times 3)$

$= 31.5$

Calculate the upper fence.

$upper\ fence = Q_3 + (1.5 \times IQR)$

$= 39 + (1.5 \times 3)$

$= 43.5$

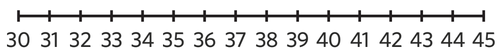
Locate any outliers from the boxplot.

31 is less than the lower fence.

44 is greater than the upper fence.

Step 2: Construct an axis with an appropriate scale.

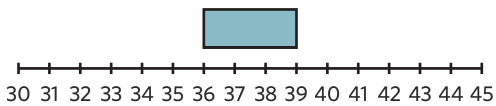
The same scale as seen in the dot plot can be used.



Step 3: Draw the border of the box.

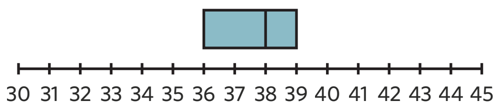
The value of the left border is Q_1 and the value of the right border is Q_3. As given by the five-number summary, Q_1 is 36 and Q_3 is 39.

The height of the box is not important.



Step 4: Mark in the vertical line.

The value of the vertical line in the middle of the box is the median. As given by the five-number summary, the median value is 38.



Step 5: Draw the whiskers.

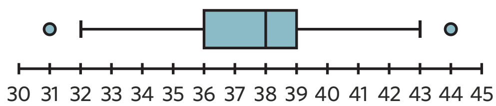
As there are outliers, the left whisker ends at the minimum value that is greater than the lower fence and the right whisker ends at the maximum value that is less than the upper fence.

Draw a whisker ranging from 32 to the leftmost point of the box.

Draw another whisker ranging from the rightmost point of the box to 43.

Remember to include the outliers as dots.

Answer



Many students drew a boxplot from the five-number summary without consideration of outliers. Students who found the two outliers often extended the whiskers to the fence values rather than to the smallest and largest values within the fences.

Back to contents

20. Explanation

The ordered data set is:

1.68 1.73 1.76 1.78 1.79 1.80 1.82 1.82 1.83 1.83
1.83 1.87 1.87 1.87 1.87

maximum = 1.87

Upper half: 1.83 1.83 1.83 1.87 1.87 1.87 1.87

$Q_3 = 1.87$

Answer

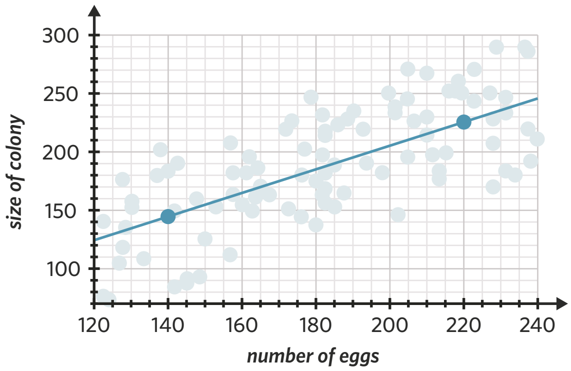
The value of Q_3 is equal to the maximum, so the whisker is not shown.

Many responses mentioned that the maximum was 1.87 without indicating this was also Q_3.

Questions from multiple lessons

21. B **22.** D

23. a.



b. 101 termites **c.** Extrapolation

1F Describing numerical data

Describing the distribution of histograms

1. C

2. a. Positively skewed with one potential outlier

b. Negatively skewed

c. Approximately symmetric bimodal distribution

d. Approximately symmetric

3. a. 41 °F

b. Positively skewed with a range of 24 °F and one potential outlier

Describing the distribution of dot plots and stem plots

4. D

5. a. Approximately symmetric with one potential outlier

b. Positively skewed

c. Negatively skewed with two potential outliers

d. Positively skewed with one potential outlier

6. a. Negatively skewed

b. Approximately symmetric with one potential outlier

c. Positively skewed with one potential outlier

d. Negatively skewed with two potential outliers

7. a. 26 years

b. Positively skewed with a range of 15 years and two potential outliers

8. a. 211 edits

b. Negatively skewed with a range of 72 edits and one potential outlier

Identifying the best measure of centre

9. C

10. a. Mean **b.** Median **c.** Median **d.** Median

11. Sally is correct. While the distribution is approximately symmetric, the accuracy of the mean would be affected by the potential outlier more than the median. Therefore, the median is the best measure of centre.

Describing the distribution of boxplots

12. B

13. a. Symmetric with a median of 10, a range of 8 and no outliers

b. Negatively skewed with a median of 8, a range of 9 and no outliers

c. Positively skewed with a median of 10, a range of 9 and one outlier

d. Symmetric with a median of 9, a range of 6 and no outliers

Joining it all together

14. C **15.** D

16. a. Prue's choir: Approximately symmetric with an approximate median of 16.5 years, a range of 30 years and no potential outliers.

Gil's choir: Approximately symmetric with a median of 26 years, a range of 65 years, and one potential outlier.

b. Prue's choir: Mean is the best measure of centre because the distribution is approximately symmetric with no outliers.

Gil's choir: Median is the best measure of centre because the distribution contains a potential outlier.

Exam practice

17. Explanation

Step 1: Determine the shape of the histogram.

The bulk of the data is towards the smaller values in the distribution.

The distribution trails off in the positive direction.

This means the histogram is positively skewed.

Step 2: Identify any potential outliers.

Several values on the upper end of the distribution appear to be outside what looks reasonable. This means there are potential outliers.

Answer

B

Back to contents

18. a. **Explanation**

Step 1: Determine the shape of the stem plot.

The bulk of the data is towards the smaller values in the distribution.

The distribution trails off in the positive direction.

This means the stem plot is positively skewed.

Step 2: Identify any potential outliers.

There are no data values that fall outside of what looks reasonable.

Answer

Positively skewed

b. **Explanation**

Step 1: Determine the number of data points.

The question specifies that the data was collected for a sample of 32 dalmatians.

Step 2: Determine the location of the median.

$\frac{32 + 1}{2} = \frac{33}{2} = 16.5$

The median will be the average of the 16th and 17th data points.

Step 3: Locate the data points and calculate the median.

Key: 21 | 6 = 21.6 kg $n = 32$

21	6 9 9
22	1 2 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

$\frac{24.5 + 24.6}{2} = \frac{49.1}{2} = 24.55$

Answer

24.55 kg

Many students with this type of question might answer 24.5 or 24.6 kg instead of finding the average of the two values.

19. **Explanation**

Step 1: Identify the location of the median.

The median is 150, which is larger than Q_1 by 80 and smaller than Q_3 by 150. This indicates that the median is closer to Q_1.

Step 2: Determine the shape of the data set.

As the median is closer to Q_1, more data exists towards the negative side of the data set. This indicates a positive skew to the data set.

Step 3: Identify any outliers.

The maximum value of 1380 is a potential outlier. Calculating the upper fence will confirm if this is the case.

$$\begin{aligned} \textit{upper fence} &= Q_3 + (1.5 \times IQR) \\ &= 300 + (1.5 \times 230) \\ &= 645 \end{aligned}$$

As 1380 is larger than the upper fence of 645, it is an outlier.

Answer

C

27% of students selected option E. This is likely due to the assumption that when the majority of data is towards the negative end of the distribution, the distribution is negatively skewed.

Questions from multiple lessons

20. C **21.** D

22. a. $122.64

b. $T_0 = 246, \quad T_{n+1} = 1.022 \times T_n$

c. $28.28

1G Introduction to standard deviation

Calculating the sample mean

1. C **2.** 148.5 **3.** 1.2

Calculating the sample mean and standard deviation using technology

4. B

5. $\bar{x} = 158.4$
$s_x = 13.3$

6. $\bar{x} = 1.7$
$s_x = 1.3$

Joining it all together

7. a. $\Sigma x = 77$
$n = 19$

b. 4 minutes c. 1.61 minutes

8. a. $\Sigma x = 916$
$n = 14$

b. 65 sales c. 15.43 sales

9. a. 79 colleagues b. 79 coffee cups

Exam practice

10. **Explanation**

Step 1: Calculate Σx and determine n.

$$\begin{aligned} \Sigma x &= 1.76 + 1.79 + 1.83 + 1.82 + 1.87 + 1.73 \\ &\quad + 1.68 + 1.82 + 1.83 + 1.87 + 1.87 + 1.80 \\ &\quad + 1.83 + 1.87 + 1.78 \\ &= 27.15 \end{aligned}$$

$n = 15$

Step 2: Calculate the mean.

$$\bar{x} = \frac{\Sigma x}{n} = \frac{27.15}{15} = 1.81$$

Answer

statistic	*high jump* (metres)	*shot-put* (metres)
mean	1.81	13.74
standard deviation	0.06	1.43

11. **Explanation**

Step 1: Calculate Σx and determine n.

$$\Sigma x = 1.07 + 1.07 + 1.08 + 1.08 + 1.03 + 1.05 + 1.07 + 1.06 + 1.07 + 1.09 + 1.02 + 1.09 = 12.78$$

$n = 12$

Step 2: Calculate the mean.

$$\bar{x} = \frac{\Sigma x}{n} = \frac{12.78}{12} = 1.065$$

Answer

1.065 kg/litre

Students rounded the answer, when a rounded answer was unnecessary. As the value has a terminating decimal, rounding is not required and makes the answer incorrect.

Questions from multiple lessons

12. A **13.** E

14. **a.** **i.** 28 goals

ii. 24 goals

b. 28 goals

1H The normal distribution

Calculating proportions in normal distributions

1. D

2. **a.** 16% **b.** 5%

3. 60 turkeys

Calculating values in normal distributions

4. C

5. **a.** 301 kg **b.** 322 kg

Determining the mean and standard deviation of a normal distribution

6. D

7. **a.** 0.04 g/cm^3 **b.** 1.04 g/cm^3

Joining it all together

8. **a.** **i.** 34% **ii.** 2.35%

b. 3 plants **c.** 116 cm

9. **a.** 84% **b.** 3 students

c. 25–65 minutes

10. **a.** 400 engines **b.** 500 hours

c. 64 engines **d.** 24 500 and 25 500 hours

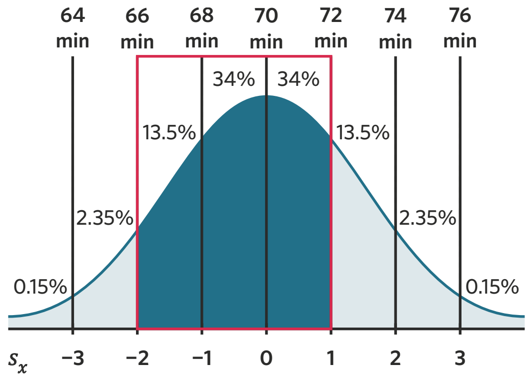
Exam practice

11. **Explanation**

Step 1: Identify and label the boundaries indicated in the question.

The lower boundary is 66 minutes, which is 2 standard deviations below the mean.

The upper boundary is 72 minutes, which is 1 standard deviation above the mean.



Step 2: Add the percentages within the boundaries.

$13.5 + 34 + 34 = 81.5\%$

Answer

D

12. **Explanation**

Step 1: Identify the percentage of days that are expected to have a *temperature difference* greater than 9.4 °C.

As 9.4 °C is the mean of the data set, 50% of the days in November will have a *temperature difference* greater than 9.4 °C.

Step 2: Calculate the number of days.

November has 30 days.

Calculate 50% of 30.

$0.5 \times 30 = 15$

Answer

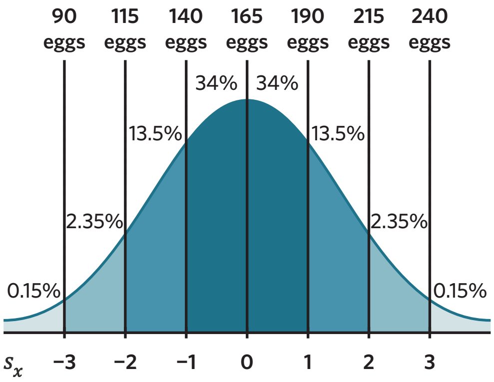
15 days

Many students who answered this question incorrectly gave the percentage of 50% rather than taking the further step to calculate the number of days this equates to.

13. **a.** **Explanation**

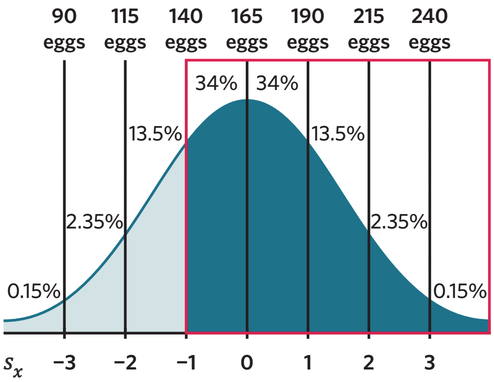
Step 1: Identify and label the boundaries indicated in the question.

The mean is 165 eggs, with a standard deviation of 25 eggs.



Step 2: Identify and label the boundaries indicated in the question.

The lower boundary is 140 eggs, with no upper boundary.



Step 3: Add the percentages within the boundaries.

$34 + 34 + 13.5 + 2.35 + 0.15 = 84\%$

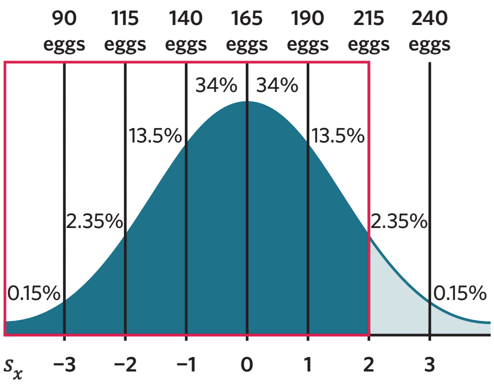
Answer

84%

b. Explanation

Step 1: Use the labelled normal distribution graph from part **a** to identify and label the boundaries indicated in the question.

The upper boundary is 215 eggs, with no lower boundary.



Step 2: Add the percentages within the boundaries.

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

Step 3: Calculate the expected number of clusters.

97.5% of 1000 clusters

$0.975 \times 1000 = 975$

Answer

975 clusters

> Many students who answered this question incorrectly gave only a percentage value of 97.5% rather than take the answer a step further and calculate the number of clusters, as asked in the question.

Questions from multiple lessons

14. C **15.** B

16. a. *blood pressure*

b. On average, *blood pressure* decreases by 10.47 mmHg for every additional hour of *sleep*.

11 z-scores

Calculating standardised (z) scores

1. D

2. a. 2.00 **b.** 1.50 **c.** −0.25 **d.** −1.38

3. a. −2.00 **b.** 2.00 **c.** 0.667 **d.** −2.33

e. −3.67

Calculating actual scores from standardised (z) scores

4. C

5. a. 4.40 kg **b.** 3.02 kg **c.** 2.00 kg

6. a. \$69 633 **b.** \$51 493 **c.** \$80 517 **d.** \$75 002

e. \$41 843

Using standardised (z) scores to interpret data

7. B

8. In Australia, Matt's standardised score was $z = 1.00$, placing him in the top 16% of male heights. In the Netherlands, Matt's standardised score was $z = -0.23$, placing him in the bottom 50% of male heights. Therefore, Matt is considered taller in Australia, relative to the male population, than in the Netherlands.

9. In 2017, Kamil's standardised score was $z = 1.33$, placing him in the slowest 16% of athletes. In 2018, Kamil's standardised score was $z = -0.22$, placing him in the fastest 50% of athletes. In 2019, Kamil's standardised score was $z = 0.60$, placing him in the slowest 50% of athletes. Therefore Kamil performed his best in 2018 when compared to other athletes.

Joining it all together

10. a. $z = -1.9$ **b.** 167.3 cm

c. i. 174.2 cm **ii.** 97.5%

11. a. In Chemistry, Isaac's standardised score was $z = 2.6$. In Physics, Isaac's standardised score was $z = 1$. Therefore, Isaac performed better in Chemistry, relative to his class, than in Physics.

Back to contents

b. In class A, Mohit's standardised score was $z = -0.57$.
In class B, Sarika's standardised score was $z = 1.33$.
Therefore, Sarika performed better in Physics, relative to her class, than Mohit, relative to his class.

c. 2.5% **d.** 7 marks **e.** 16% **f.** 34%

Exam practice

12. Explanation

Step 1: Identify the mean, standard deviation and standardised score.

$\overline{x} = 69$

$s_x = 4$

$z = -2.5$

Step 2: Calculate the actual score.

Substitute the values into the formula $x = \overline{x} + (z \times s_x)$.

$x = 69 + (-2.5 \times 4)$
$= 69 - 10$
$= 59$

Answer

A

13. Explanation

Step 1: Identify the mean, standard deviation and actual score.

$\overline{x} = 13.74$

$s_x = 1.43$

$x = 14.50$

Step 2: Calculate the z-score.

Substitute the values into the formula $z = \frac{x - \overline{x}}{s_x}$.

$z = \frac{14.50 - 13.74}{1.43}$
$= \frac{0.76}{1.43}$
$= 0.531...$

Answer

$z = 0.5$

Most incorrect responses were from students who did not round as required.

14. Explanation

Step 1: Identify the mean, standard deviation and standardised score.

$\overline{x} = 165$

$s_x = 25$

$z = -2.4$

Step 2: Calculate the actual score.

Substitute the values into the formula $x = \overline{x} + (z \times s_x)$.

$x = 165 + (-2.4 \times 25)$
$= 165 - 60$
$= 105$

Answer

105 eggs

15. Explanation

Step 1: Identify the mean, standard deviation and actual score.

$\overline{x} = 3.5$

$s_x = 3$

$x = -4.3$

Step 2: Calculate the z-score.

Substitute the values into the formula $z = \frac{x - \overline{x}}{s_x}$.

$z = \frac{-4.3 - 3.5}{3}$
$= \frac{-7.8}{3}$
$= -2.6$

Answer

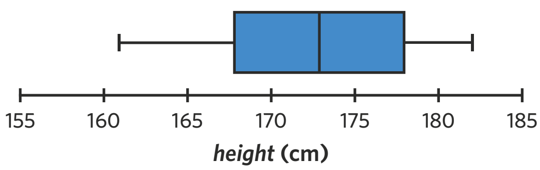
B

Many students misread this type of question and assumed that the actual value given in the question was the standardised score. This would have led to an incorrect answer of A.

Questions from multiple lessons

16. A **17.** C

18. a.



b. 75%

11 ANSWERS

Back to contents

2A Associations between two categorical variables

Displaying bivariate data using two-way frequency tables

1. a. D b. C

2. a.

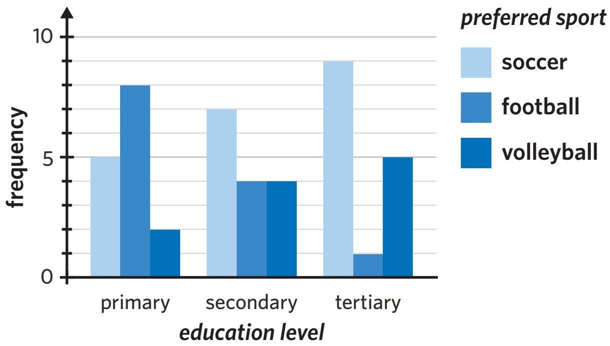
	coffee consumption	
reusable cup	often	rarely
frank green	41%	17%
KeepCup	40%	11%
other	19%	72%
total	**100%**	**100%**

b. 40%

Displaying bivariate data using grouped bar charts

3. D

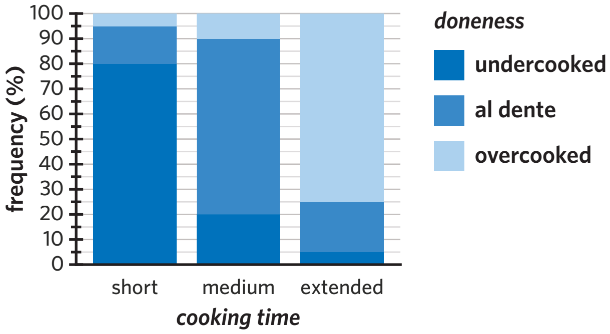
4.



Displaying bivariate data using percentage segmented bar charts

5. D

6.



Describing the association between two categorical variables

7. C

8. Yes, there is an association between *cryptocurrency investment* and *age*. The percentage of those with a cryptocurrency investment tends to increase with age, from 15% of participants aged under 45 to 75% of participants aged 45 and over having an investment.

Answers may vary.

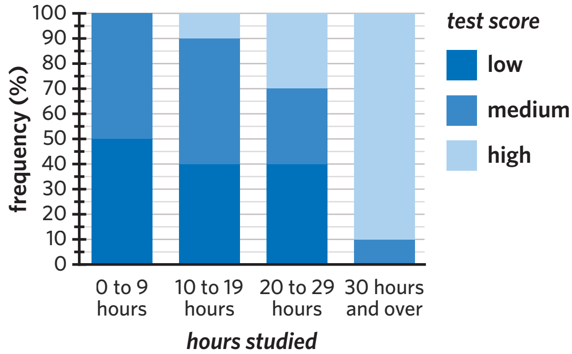
Joining it all together

9. a. *hours studied*

b.

	hours studied			
test score	0 to 9 hours	10 to 19 hours	20 to 29 hours	30 or more hours
low	80%	30%	20%	0%
medium	20%	60%	50%	10%
high	0%	10%	30%	90%
total	**100%**	**100%**	**100%**	**100%**

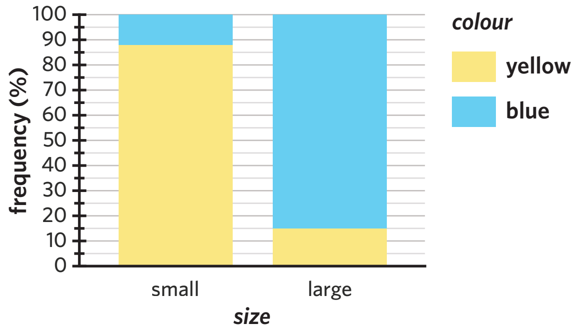
c.



d. Yes, the proportion of students with a high score tends to increase with the number of hours studied. For the students who obtained a 'high' test score, 0% studied for 0 to 9 hours, while 10% studied for 10 to 19 hours, 30% studied for 20–29 hours and 90% studied for 30 hours or over.

Answers may vary.

10. a.



b. Yes, there is an association between a star's *size* and its *colour*. The percentage of blue stars tends to increase with size. 12% of small stars are blue, while 85% of large stars are blue.

Answers may vary.

Exam practice

11. **Explanation**

Step 1: Fill in the column for 'small' cities by counting from the table.

	size	
congestion level	small	large
high	4	
medium	4	
low	8	
total	**16**	

Back to contents

Step 2: Ensure that the total for that column matches the overall number of ‘small’ cities.

$4 + 4 + 8 = 16$

There are a total of 16 ‘small’ cities.

Step 3: Fill in the column for ‘large’ cities by counting from the table.

	size	
congestion level	small	large
high	4	2
medium	4	2
low	8	3
total	**16**	

Step 4: Ensure that the total for that column matches the overall number of ‘large’ cities.

$2 + 2 + 3 = 7$

There are a total of 7 ‘large’ cities.

Answer

	size	
congestion level	small	large
high	4	2
medium	4	2
low	8	3
total	**16**	**7**

12. Explanation

To solve this question, check whether each option is correct or incorrect.

A: This is incorrect. It only mentions one category of each variable. ✗

B: This is incorrect. It only mentions one category of each variable. ✗

C: This is incorrect. It compares two separate categories of the *number of moths* for two categories of *trap types* and does not quote specific percentages. ✗

D: This is incorrect. It compares two separate categories of the *number of moths* for two categories of *trap type* and while it does quote specific percentages that support the statement, this is comparing the incorrect percentages. ✗

E: This is correct. It compares two different trap types for the same number of moths and also quotes specific percentages that support the statement. ✓

Answer

E

12% of students incorrectly chose option C. This was an insufficient response because it compares two different categories for both variables and also does not quote any significant percentages that support the statement.

13. Explanation

Step 1: Determine the two-way percentage frequency table for the percentage segmented bar chart.

- 65% of visitors under 55 years prefer domestic travel.
- 35% of visitors under 55 years prefer international travel.
- 45% of visitors 55 years and over prefer domestic travel.
- 55% of visitors 55 years and over prefer international travel.

	age	
preferred travel destination	under 55 years	55 years and over
domestic	65%	45%
international	35%	55%
total	**100%**	**100%**

Step 2: Calculate the percentage frequency for each option and compare it with the table.

A: This **matches** the percentages in the given segmented bar chart. ✓

	age			
preferred travel destination	under 55 years		55 years and over	
domestic	91	$\frac{91}{140} \times 100 = 65\%$	90	$\frac{90}{200} \times 100 = 45\%$
international	49	$\frac{49}{140} \times 100 = 35\%$	110	$\frac{110}{200} \times 100 = 55\%$
total	**140**	**100%**	**200**	**100%**

B: This **does not** match the percentages in the given segmented bar chart. ✗

	age			
preferred travel destination	under 55 years		55 years and over	
domestic	65	$\frac{65}{110} \times 100 \approx 59\%$	35	$\frac{35}{90} \times 100 \approx 39\%$
international	45	$\frac{45}{110} \times 100 \approx 41\%$	55	$\frac{55}{90} \times 100 \approx 61\%$
total	**110**	**100%**	**90**	**100%**

C: This **does not** match the percentages in the given segmented bar chart. ✗

	age			
preferred travel destination	under 55 years		55 years and over	
domestic	35	$\frac{35}{140} \times 100 = 35\%$	55	$\frac{55}{100} \times 100 = 55\%$
international	65	$\frac{65}{140} \times 100 = 65\%$	45	$\frac{45}{100} \times 100 = 45\%$
total	**100**	**100%**	**100**	**100%**

Back to contents

D: This **does not** match the percentages in the given segmented bar chart. ×

preferred travel destination	*age*			
	under 55 years		55 years and over	
domestic	50	$\frac{50}{150} \times 100 \approx 33\%$	70	$\frac{70}{120} \times 100 \approx 58\%$
international	100	$\frac{100}{150} \times 100 \approx 67\%$	50	$\frac{50}{120} \times 100 \approx 42\%$
total	**150**	**100%**	**120**	**100%**

E: This **does not** match the percentages in the given segmented bar chart. ×

preferred travel destination	*age*			
	under 55 years		55 years and over	
domestic	71	$\frac{71}{100} \times 100 = 71\%$	39	$\frac{39}{100} \times 100 = 39\%$
international	29	$\frac{29}{100} \times 100 = 29\%$	61	$\frac{61}{100} \times 100 = 61\%$
total	**100**	**100%**	**100**	**100%**

Answer

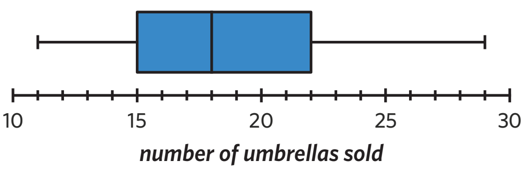
A

28% of students incorrectly chose option C. This was likely because they confused the percentage frequency with the actual frequency. It is likely they also confused the two categories 'domestic' and 'international' for visitors under 55 years old.

Questions from multiple lessons

14. C **15.** E

16. a.



b. The median number of umbrellas sold differs from month to month. The median in April is 28 umbrellas, the median in May is 37 umbrellas, and the median in June is 49 umbrellas.

2B Associations between numerical and categorical variables

Displaying data using back-to-back stem plots

1. The median number of 'scones' sold per day is 29 while the median number of 'muffins' sold per day is 38. In order to maximise revenue, the bakery should remove scones from their selection.

2. a. **Key:** 0 | 2 = 2

NSW Blues		Queensland Maroons
8 7 6 6 6 5 2	0	2 6
8 8 2 2 0	1	0 0 1 2 2 4 4 6 6 6 6 6 8
6 4 2 2 2 1 0 0	2	0 0 4 9
	3	
	4	3

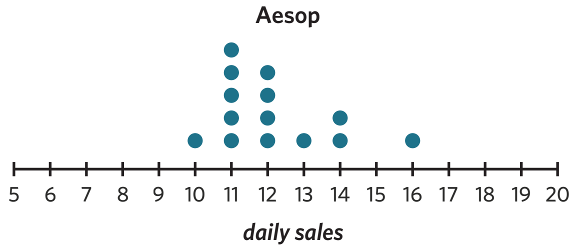
b. Queensland Maroons **c.** NSW Blues

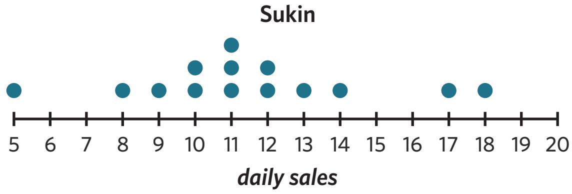
d. Most of the scoring actions are worth an even number of points.

Displaying data using parallel dot plots

3. a. D **b.** D

4. a.



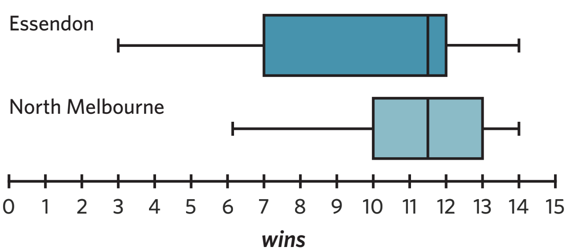


b. Aesop **c.** Sukin

Displaying data using parallel boxplots

5. C

6. a.



b. Essendon **c.** Essendon

Describing the association between numerical and categorical variables

7. a. B

b. Yes, there is an association between *average pay rate* and *year*. The *average pay rate* increases over time. The median *average pay rate* was $8/hour in 1980, $13/hour in 1990 and $14/hour in 2000.

8. a. Yes, the training program improved Charlotte's running time. Before the training program, Charlotte's median running time was 48.5 minutes. Her median running time decreased to 43 minutes after completing the training program.

b. Yes, the training program improved Charlotte's consistency. The range decreased from 9 minutes to 6 minutes after completing the training program, and the *IQR* decreased from 4 minutes to 1.5 minutes.

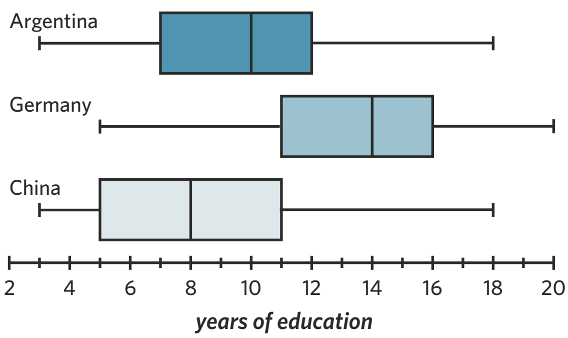
9. a. No, there is not an association between *maximum temperature* (°C) and *location*. The median temperature was 15 °C in both locations.

b. Yes, there is an association between the variability of *maximum temperature* (°C) and *location*. The variability of *maximum temperature* (°C) is greater in Launceston than Hobart. The range in Hobart is 14 °C and the *IQR* is 4 °C. The range in Launceston is 31 °C and the *IQR* is 6 °C.

c. No, there is not an association between the distribution of *maximum temperature* (°C) and *location*. The shape of the distribution in both locations is approximately symmetrical.

Joining it all together

10. a.



b. 75% of adults in Germany have more *years of education* than 75% of adults in China. The *IQR* of the *years of education* of adults in Argentina is equal to the *IQR* of adults in Germany.

c. Yes, there is an association between *years of education* and *country*. The number of *years of education* differs in each *country*. The median *years of education* is 14 in Germany, 10 in Argentina, and 8 in China.

d. No, there is not an association between the variability of *years of education* and *country* of residence. The range of *years of education* is 15 in each country, whilst the *IQR* is 5 years in Germany and Argentina, and 6 years in China.

Exam practice

11. **Explanation**

Step 1: Determine which pairing of variable types can be displayed in parallel boxplots.

Parallel boxplots can depict the relationship between a numerical variable and a categorical variable.

Step 2: Find which option can be the second variable.

Median rainfall, in millimetres, is a numerical variable. Therefore, the other variable must be a categorical variable.

The only categorical variable is option C.

Answer

C

46% of students incorrectly chose options B or E. This was likely because these variables are likely to be associated with median rainfall, in millimetres. However, the question was asking which variable could be displayed graphically on a parallel boxplot.

12. **Explanation**

Step 1: Find the appropriate statistic.

The appropriate statistic is the median as the question is asking us to investigate the relationship between the value of the *minimum daily temperature* and the *month*.

Step 2: Determine the median.

The median for each month can be found in the table.

February: 10.9 °C

May: 7.5 °C

July: 5.0 °C

Answer

There is a difference/change in the median of the *minimum daily temperature* for each of the months. The median *minimum daily temperature* is 10.9 °C in February, 7.5 °C in May and 5.0 °C in July. This implies that *minimum daily temperature* is associated with the *month*.

Students were required to state that a decrease or change in the median or *IQR* signals an association. Some students incorrectly estimated the median or *IQR* by reading off the boxplot, instead of using the more accurate figures in the table. Other incorrect responses involved the words 'averages' or 'means', comments on the minimums and maximums, or comments on the shape of the boxplots.

13. **Explanation**

Step 1: Find the appropriate statistic.

The appropriate statistic is the median as the question is asking to investigate the relationship between the value of the *wingspan* and the *place of capture*.

Step 2: Determine the median.

The median for each *place of capture* can be found using the back-to-back stem plot.

Forest: 21 mm

Grassland: 30 mm

Answer

There is a difference in the median of the *wingspan* for each *place of capture*. The median *wingspan* is 21 mm in the forest and 30 mm in the grassland. This implies that the *wingspan* of moths is associated with the *place of capture*.

Students were required to state that a difference in medians signals an association. Some students calculated the mean as the appropriate statistic, however as there is an outlier, the median is more appropriate. Other incorrect responses involved quoting comparisons of other irrelevant statistics.

Back to contents

Questions from multiple lessons

14. E

15. D

16. **a.** −1.8

b. 16%

c. 693 people

2C Associations between two numerical variables

Identifying the response and explanatory variables

1. D

2. **a.** *RV*: *number of umbrellas sold*
EV: *amount of rainfall*

b. *RV*: *selling price*
EV: *age*

c. *RV*: *price*
EV: *season*

d. *RV*: *salary*
EV: *years of experience*

3. **a.** *RV*: *high score*
EV: *time spent playing*

b. *RV*: *money spent*
EV: *distance*

c. *RV*: *amount of rainfall*
EV: *month*

d. *RV*: *number of homes owned*
EV : *age*

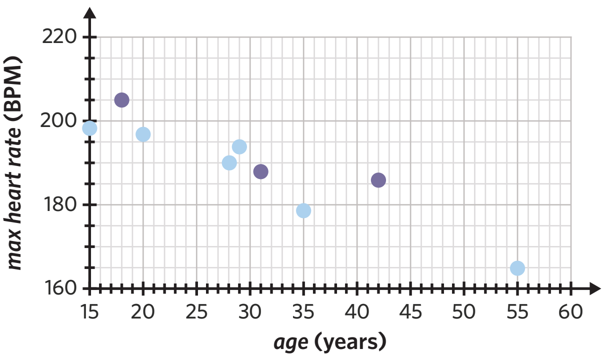
4. *RV*: *weight*
EV: *height*

5. *time spent watching Grey's Anatomy*

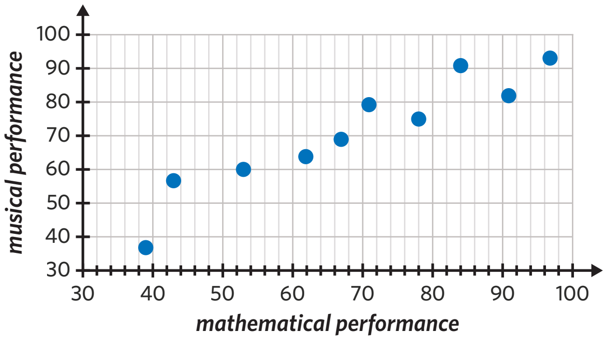
Using technology to construct scatterplots

6. B

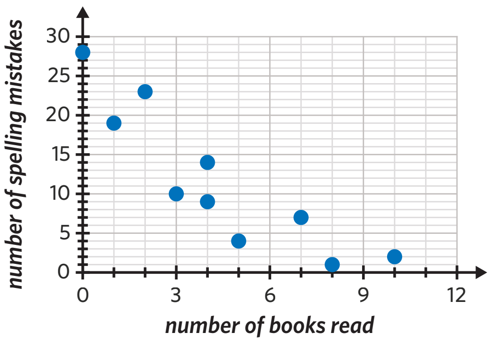
7.



8.



9.



Describing the relationship between two numerical variables

10. B

11. **a.** Strong, positive, and linear with no visible outliers

b. Strong, negative, and linear with one visible outlier at (30, 60)

c. Weak, positive, and linear with no visible outliers

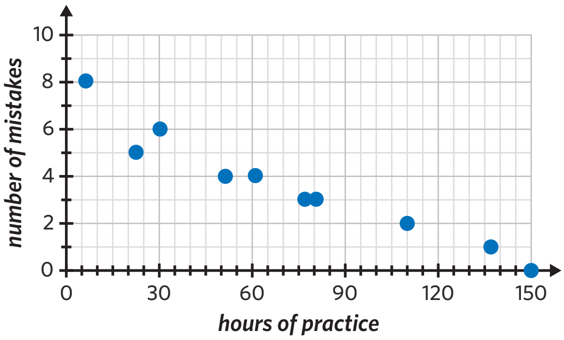
d. Moderate, positive, and linear with no visible outliers

Joining it all together

12. A

13. **a.** *RV*: *number of mistakes*
EV: *hours of practice*

b.

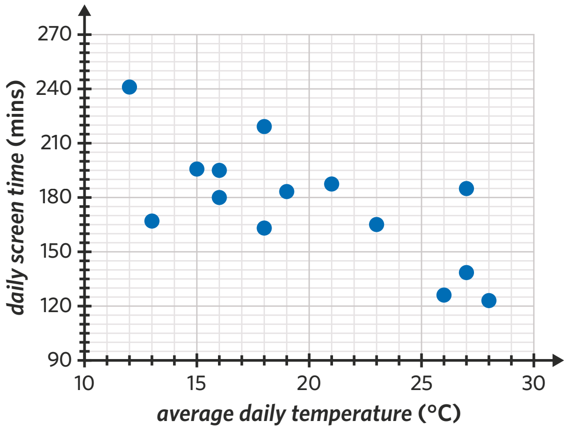


c. Strong, negative, and linear

14. **a.** Francis has incorrectly put the response variable on the horizontal axis and the explanatory variable on the vertical axis. The variable *average daily temperature* (°C) should be on the horizontal axis and *daily screen time* (mins) should be on the vertical axis.

Back to contents

b.



c. Moderate, negative, and linear

Exam practice

15. Explanation

Step 1: Assess whether each variable is predicting an outcome or being predicted.

evening congestion level is being predicted from changes in *morning congestion level*.

Step 2: Determine the response variable.

The *RV* is predicted from changes in the *EV*.

Answer

evening congestion level

16. Explanation

Step 1: Determine the strength.

All data points appear to be very close to the underlying trend. The association is strong.

Step 2: Determine the direction.

As *actual temperature*, the *EV*, increases, the *apparent temperature*, the *RV*, also increases. The association is positive.

Step 3: Determine the form.

The distribution of data best resembles a straight line. The form is linear.

Answer

Strong, positive, and linear

17. Explanation

Step 1: Determine the strength.

All data points appear to be very close to the underlying trend. The association is strong.

Step 2: Determine the direction.

As *mean age*, the *EV*, increases, *mean height*, the *RV*, decreases. The association is negative.

Answer

Strong and negative

It is important to note that the question asks for the association 'in terms of strength and direction'. Many students provided additional information such as form that resulted in 0 marks.

Questions from multiple lessons

18. D **19.** E

20. a. $7999.55

b. $7973.90

2D Correlation and causation

Calculating and interpreting the Pearson correlation coefficient

1. A **2.** B

3. Penelope is correct. An r value between -0.25 and 0.25 is too weak for any association to be determined.

4. a. Yes **b.** 0.95

c. Strong, positive, and linear

5. a. -0.06 **b.** No association

Distinguishing between correlation and causation

6. D **7.** C

8. No. The *total profit* could also be positively correlated with the sales of other products, for example, the number of *croissants sold*.

9. *population*

10. a. Potential variables include, but are not limited to:

amount of free time, *time spent at work*, *weather*

b. Robbie could try to eliminate all external variables by recording on the same day each week, as his schedule would not change significantly, and over a shorter period of time, where variables like *weather* are less likely to change drastically.

Joining it all together

11. a. Yes **b.** -0.76

c. Strong, negative, and linear

d. No. This association is likely due to a coincidence.

12. a. Yes. There is a linear relationship between the variables *time spent on recovery* and *goals*.

b. 0.63

c. Moderate, positive, and linear

d. Potential variables include, but are not limited to:

hydration, *position played*, *hours of sleep*, *weather*, *opposition team*

e. No. In order for the experiment to determine causation, Mikayla needs to eliminate any changes in external variables.

Back to contents

Exam practice

13. **Explanation**

To solve this question, check whether each option is true or false.

A: This is false. It isn't confirmed whether there is a causal relationship between the number of stray cats and stray dogs. ✗

B: This is false. Causal relationships can exist even if the correlation coefficient is not exactly $+1$ or -1. ✗

C: This is false. The positive correlation between population and both the number of stray cats and dogs show there is a logical explanation for the association. ✗

D: This is true. Both the number of stray cats and dogs are positively associated with the population. An increasing population is a reasonable explanation for increasing numbers of stray cats and dogs. ✓

E: This is false. The correlation coefficient doesn't tell us whether or not people were careful about keeping their cats properly contained. ✗

Answer

D

14. **Explanation**

Method 1: TI-Nspire

Step 1: From the home screen, select '1: New' → '4: Add Lists & Spreadsheet'.

Step 2: Name column A 'am' and column B 'pm'.

Enter the *humidity* 9 *am* values into column A, starting from row 1.

Enter the *humidity* 3 *pm* values into column B, starting from row 1.

Step 3: Determine the response and explanatory variables.

The variable *humidity* 3 *pm* is being predicted from the *humidity* 9 *am*, so *humidity* 3 *pm* is the response variable.

RV: *humidity* 3 *pm*

EV: *humidity* 9 *am*

Step 4: Press menu and select '4: Statistics' → '1: Stat Calculations' → '4: Linear Regression (a+bx)'.

Select 'am' in 'X List:' and 'pm' in 'Y List:'.

Select 'OK'.

Step 5: Read the r value from the screen and round to the number of decimal places specified in the question.

Method 2: Casio ClassPad

Step 1: From the main menu, tap Statistics.

Step 2: Name the first list 'am' and the second list 'pm'.

Enter the *humidity* 9 *am* values into list 'am', starting from row 1.

Enter the *humidity* 3 *pm* values into list 'pm', starting from row 1.

Step 3: Determine the response and explanatory variables.

The variable *humidity* 3 *pm* is being predicted from the *humidity* 9 *am*, so *humidity* 3 *pm* is the response variable.

RV: *humidity* 3 *pm*

EV: *humidity* 9 *am*

Step 4: Tap 'Calc' → 'Regression' → 'Linear Reg'.

Select 'main\am' in 'XList:' and 'main\pm' in 'YList:'.

Tap 'OK' to confirm.

Step 5: Read the r value from the screen and round to the number of decimal places specified in the question.

Answer - Method 1 and 2

$r = 0.871$

Questions from multiple lessons

15. D

16. C

17. a. *size*

b. 11.8 kg

c. $a = 12.5$

$b = 11.8$

2D ANSWERS

Back to contents

3A Fitting a least squares regression line

Using technology to determine the least squares regression equation

1. D **2.** C

3. *result* $= 41.250 + 5.151 \times$ *number of hours slept*

Determining the least squares regression equation from a graph

4. B

5. a. $y = 2 + 1.1x$

b. $y = 8 - 0.5x$

c. *daily sales* $= 0.75 \times$ *daily customers*

d. *daily sales* $= 32 - 0.26 \times$ *price*

Calculating the least squares regression equation from summary statistics

6. C

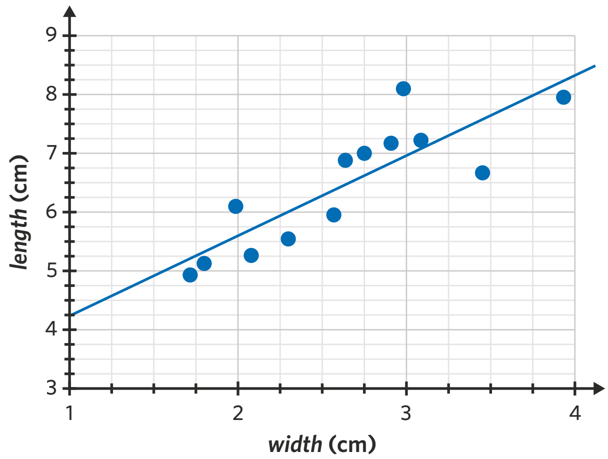
7. $s_y = 16.3, \bar{y} = 57.1$

8. *height* $= -3.4 + 1.2 \times$ *time*

Sketching a least squares regression line from its equation

9. B

10.



Joining it all together

11. D

12. a. *force* $= 1.0 + 9.7 \times$ *mass*

b. *force* $= 0.96 + 9.69 \times$ *mass*

c.

$$a = \bar{y} - b\bar{x}$$
$$0.96 = \bar{y} - 9.69 \times 1.482$$
$$\bar{y} = 0.96 + 9.69 \times 1.482$$
$$= 15.32$$

Exam practice

13. Explanation

Step 1: Identify the relevant formula for this question.

The value of r can be found using the formula $b = r \times \frac{s_y}{s_x}$.

Step 2: Substitute the values for b, s_y and s_x into the formula.

$$1.31 = r \times \frac{3.24}{2.33}$$

Step 3: Solve to find the value of r.

$$1.31 = 1.39r$$
$$r = \frac{1.31}{1.39}$$
$$= 0.942...$$

Answer

D

14. Explanation

Method 1: TI-Nspire

Step 1: From the home screen, select '1: New' → '4: Add Lists & Spreadsheet'.

Step 2: Name a list 'height' and another list 'weight' and enter the data as shown.

Step 3: Identify the explanatory and response variables.

As *height* is being used to predict *weight*, *height* is the explanatory variable.

EV: *height*

RV: *weight*

Step 4: Press ctrl + doc and select '5: Add Data & Statistics'.

Step 5: Add the variables on each axis using the 'Click to add variable' function.

The *RV* will be positioned on the vertical axis and the *EV* will be positioned on the horizontal axis.

Step 6: Press menu. Select → '4: Analyse' → '6: Regression' → '2: Show Linear (a+bx)' to plot the least squares regression line.

The least squares regression line and its equation in the form $y = a + bx$ will appear.

Step 7: Count the number of dots below the least squares line.

The line represents predicted values, and the dots represent actual values.

There are 6 dots below the least squares line.

Method 2: Casio ClassPad

Step 1: From the main menu, tap Statistics.

Step 2: Name a list 'height' and another list 'weight' and enter the data as shown.

Step 3: Identify the explanatory and response variables.

As *height* is being used to predict *weight*, *height* is the explanatory variable.

EV: *height*

RV: *weight*

Step 4: Configure the settings of the graph by tapping in the icon bar.

Create a scatterplot by selecting 'Type' as 'Scatter'.

Specify the data set by changing 'XList:' to 'main\height' and 'YList:' to 'main\weight.

Step 5: Tap 'Set' to confirm and then to plot the scatterplot.

Back to contents

Step 6: Fit a least squares regression line to the scatterplot by tapping 'Calc' → 'Regression' → 'Linear Reg'. Specify the data set by changing 'XList:' to 'main\height' and 'YList:' to 'main\weight'.

Tap 'OK' to confirm.

Step 7: The Stat Calculation window shows the values of a and b which will be used to write the least squares regression line.

Note: If $y = a \bullet x + b$ is selected, the a and b values need to be switched when writing the equation in $y = a + bx$ form.

Step 8: To visualise the data, tap 'OK', and the regression line will be generated on the scatterplot.

Step 9: Count the number of dots below the least squares line.

The line represents predicted values, and the dots represent actual values.

There are 6 dots below the least squares line.

Answer – Method 1 and 2

D

> 29% of students chose B, due to the incorrect assumption that the dots above the line needed to be counted.

15. a. Explanation

Method 1: TI-Nspire

Step 1: From the home screen, select '1: New' → '4: Add Lists & Spreadsheet'.

Step 2: Name a list 'moths' and another list 'density' and enter the data as shown.

Step 3: Identify the explanatory and response variables.

As *number of male moths* is being used to predict *egg density*, *number of male moths* is the explanatory variable.

EV: *number of male moths*

RV: *egg density*

Step 4: Press ctrl + doc▾ and select '5: Add Data & Statistics'.

Step 5: Add the variables on each axis using the 'Click to add variable' function.

The *RV* will be positioned on the vertical axis and the *EV* will be positioned on the horizontal axis.

Step 6: Press menu. Select → '4: Analyse' → '6: Regression' → '2: Show Linear (a+bx)' to plot the least squares regression line.

The least squares regression line and its equation in the form $y = a + bx$ will appear.

Step 7: Fill in the boxes corresponding to a and b, making sure to round as specified.

Method 2: Casio ClassPad

Step 1: From the main menu, tap Statistics.

Step 2: Name a list 'moths' and another list 'density' and enter the data as shown.

Step 3: Identify the explanatory and response variables.

As *number of male moths* is being used to predict *egg density*, *number of male moths* is the explanatory variable.

EV: *number of male moths*

RV: *egg density*

Step 4: Configure the settings of the graph by tapping the graph settings icon in the icon bar.

Create a scatterplot by selecting 'Type' as 'Scatter'.

Specify the data set by changing 'XList:' to 'main\moths' and 'YList:' to 'main\density.

Step 5: Tap 'Set' to confirm and then the graph icon to plot the scatterplot.

Step 6: Fit a least squares regression line to the scatterplot by tapping 'Calc' → 'Regression' → 'Linear Reg'. Specify the data set by changing 'XList:' to 'main\moths' and 'YList:' to 'main\density'.

Tap 'OK' to confirm.

Step 7: The Stat Calculation window shows the values of a and b which will be used to write the least squares regression line.

Note: If $y = a \bullet x + b$ is selected, the a and b values need to be switched when writing the equation in $y = a + bx$ form.

Step 8: Fill in the boxes corresponding to a and b, making sure to round as specified.

Answer – Method 1 and 2

egg density $= -46.8 + 18.9 \times$ *number of male moths*

> Occasionally students would incorrectly round the calculated values For example, writing -46.9 instead of -46.8 for the first number.

b. Explanation

Step 1: Decide on two horizontal axis values to substitute into the equation.

For accuracy, it is best to use two values for *number of female moths* that are on either end of the visible plane.

The best values to use are 10 and 60.

Step 2: Substitute both values for *number of female moths* into the equation to find their corresponding *egg density* values.

number of female moths $= 10$

egg density $= 191 + 31.3 \times 10 = 504$

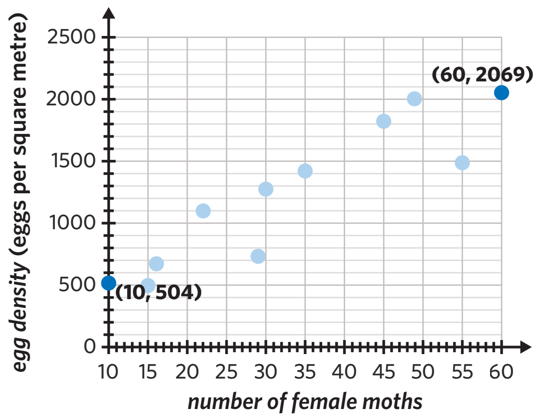
number of female moths $= 60$

egg density $= 191 + 31.3 \times 60 = 2069$

Step 3: Determine the points to plot on the graph.

Point 1: (10, 504)

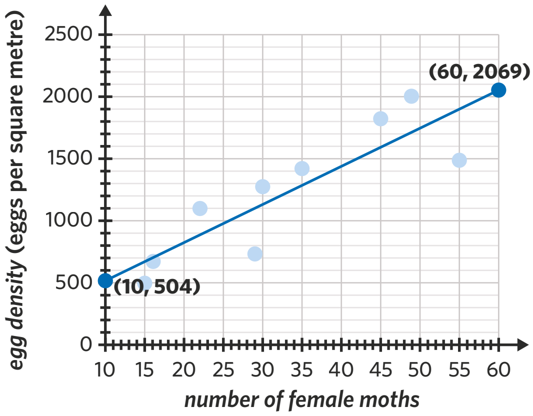
Point 2: (60, 2069)



Back to contents

Step 4: Draw a straight line passing through the two points.

Answer



Many students did not realise that the horizontal axis starts at 10, not 0.

Questions from multiple lessons

16. C **17.** E **18.** E

3B Interpreting a least squares regression line

Interpreting a least squares regression line

1. C

2. **a.** \$24 000 **b.** \$3500

3. **a.** 5.61 minutes

b. The y-intercept shows that on average it takes 0.32 minutes to run 0 km. This interpretation does not make sense.

Making predictions using a least squares regression line

4. C

5. **a.** 44.7 **b.** 207.1

6. **a.** 9 mL **b.** 13 cm **c.** Extrapolation

7. **a.** 88 people

b. Yes, the prediction is reliable because it is an example of interpolation.

c. 3 people

d. No, it may not be reliable because the prediction is an example of extrapolation.

Joining it all together

8. **a.** 10.89 cm **b.** 3.23 cm **c.** 28.66 cm

9. **a.** On average, for each unit increase in year, the price of the dress decreases by \$3.03.

b. \$156 **c.** \$74

d. Yes, the prediction is reliable because it is an example of interpolation.

10. **a.** On average, the *precipitation* level decreases by 0.43 mm for every 1 °C increase in the *maximum daily temperature*.

b. 22.72 °C

c. −0.98 mm. This prediction is not reliable because it is an example of extrapolation.

d. D

Exam practice

11. **Explanation**

Determine whether the value is within the range of data.

The range of the data of the pressure at 9 am is between 1013 hPa and 1027.25 hPa. 1025 hPa is within the range of the data set.

Answer

Interpolation

12. **Explanation**

Step 1: Substitute *morning congestion level* = 60 into the regression equation.

$\textit{evening congestion level} = 8.48 + 0.922 \times 60$

Step 2: Solve the equation.

$\textit{evening congestion level} = 8.48 + 55.32$

$= 63.8$

Answer

63.8%

An answer of 64% was not accepted since the question did not specify to round to the nearest whole number.

13. **Explanation**

Step 1: Identify the slope.

In the rule $y = a + bx$, the slope is represented by b.

$b = 31.3$

Step 2: Interpret the slope.

The slope is the average change in *egg density* for every one-unit increase in the *number of female moths* caught.

Answer

On average, *egg density* increases by 31.3 eggs/m^2 for each additional female moth caught.

The main problem with this question is that many students did not reference the one-unit increase in female moths. A response stating that *egg density* increases by 31.3 for each increase in the *number of female moths* caught is not acceptable as it suggests that the same increase in *egg density* applies to any increase in the *number of female moths* regardless of the size of the increase.

14. **Explanation**

Step 1: Calculate the 2008 congestion level for Sydney.

Substitute *year* = 2008 into the equation for Sydney and solve for *congestion level*.

$\textit{congestion level} = -2280 + 1.15 \times 2008$

$= -2280 + 2309.2$

$= 29.2$

Back to contents

Step 2: Calculate 2008 congestion level for Melbourne.

Substitute $year = 2008$ into the equation for Melbourne and solve for *congestion level*.

$$\begin{aligned} congestion\ level &= -1515 + 0.7667 \times 2008 \\ &= -1515 + 1539.53... \\ &\approx 24.5 \end{aligned}$$

Step 3: Compare the starting values for Sydney and Melbourne.

$29.2 > 24.5$

Step 4: Compare the slopes.

The slope of the least squares line for Sydney's congestion level is 1.15.

The slope of the least squares line for Melbourne's congestion level is 0.7667.

$1.15 > 0.7667$

Answer

The 2008 Sydney congestion level is higher, $29.2 > 24.5$, and the Sydney line slope is greater, $1.15 > 0.7667$. This means that Sydney's congestion levels start higher and increase by a greater amount, on average, than Melbourne's congestion levels each year.

Students needed to refer to both the starting values and the slopes for Sydney and Melbourne.A common mistake was students focusing on the starting values from the table instead of calculating the starting values from the least squares line. Students also needed to clearly state that Sydney's slope was greater, as simply stating the gradient figures for Sydney and Melbourne was not sufficient.

Questions from multiple lessons

15. E **16.** D

17. **a.** $a = 15\ 000$, $b = 1.000041$

b. $n = 2190$

3C Performing a regression analysis

Calculating and interpreting the coefficient of determination

1. A

2. **a.** 0.99

b. 99% of the variation in *hours slept* can be explained by the variation in the amount of *hours worked*.

Performing residual calculations and constructing a residual plot

3. C

4. **a.** $predicted\ test\ score = -85.67 + 1.57 \times IQ$

b. $predicted\ test\ score = 57.26$, $residual = -3.26$

5.

predicted *width* (cm)	25.48	37.31	41.86	21.84	32.76	28.21	47.32
residual value (cm)	0.52	−2.31	8.14	−1.84	−7.76	6.79	−3.32

Performing a residual analysis

6. **a.** The residual plot does not support the assumption of linearity.

b. The residual plot supports the assumption of linearity.

c. The residual plot does not support the assumption of linearity.

7. The assumption of linearity is not supported.

Joining it all together

8. **a.**

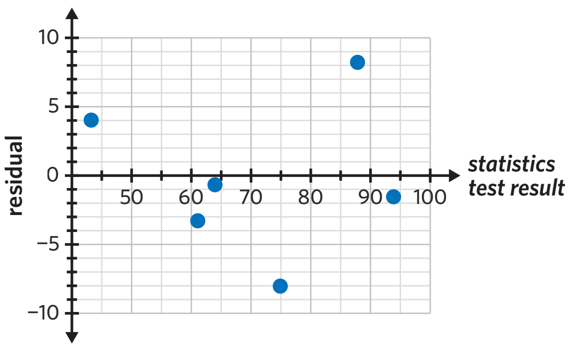
student	Sophie	Isaac	Mayu	Deanne	Selby	Emily
predicted *calculus test result* (%)	72.1	56.3	86.8	35.9	93.6	59.7

b.

student	Sophie	Isaac	Mayu	Deanne	Selby	Emily
residual value (%)	−8.1	−3.3	8.2	4.1	−1.6	−0.7

c. Sophie, Isaac, Selby and Emily

d.



The assumption of linearity is not supported.

9. **a.** $r^2 = 0.82$

b. 82% of the variation in *reaction time* can be explained by the variation in *age*.

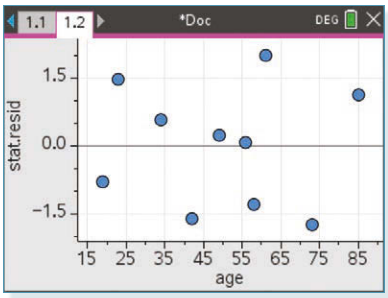
c. $reaction\ time = 1.40 + 0.14 \times age$

d.

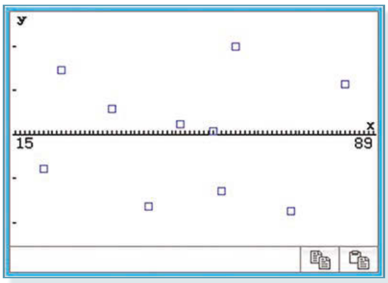
age (years)	19	42	73	56	23
reaction time (ms)	3.2	5.5	9.6	9.1	6.0
residual value (ms)	−0.86	−1.78	−2.02	−0.14	1.38

age (years)	85	61	49	34	58
reaction time (ms)	14.1	11.7	8.3	6.6	8.0
residual value (ms)	0.8	1.76	0.04	0.44	−1.52

e. TI-Nspire

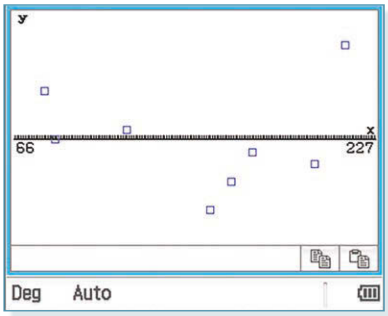


Casio ClassPad



The assumption of linearity is supported.

Casio ClassPad



The assumption of linearity is supported.

10. a. $r^2 = 0.952$

b. 95.2% of the variation in *sleep* can be explained by the variation in *time spent on phone*.

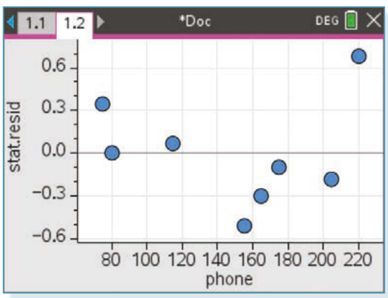
c. $sleep = 11.45 - 0.03058 \times time\ spent\ on\ phone$

d.

time spent on phone **(min)**	175	115	75	205
sleep **(hours)**	6.0	8.0	9.5	5.0
residual value (hours)	−0.10	0.07	0.34	−0.18

time spent on phone **(min)**	155	220	80	165
sleep **(hours)**	6.2	5.4	9.0	6.1
residual value (hours)	−0.51	0.68	0.00	−0.30

e. TI-Nspire



Exam practice

11. Explanation

Step 1: Calculate the coefficient of determination.

$r^2 = (0.92)^2$

$= 0.8464$

Step 2: Convert the coefficient of determination into a percentage.

$0.8464 \times 100 = 84.64$

Answer

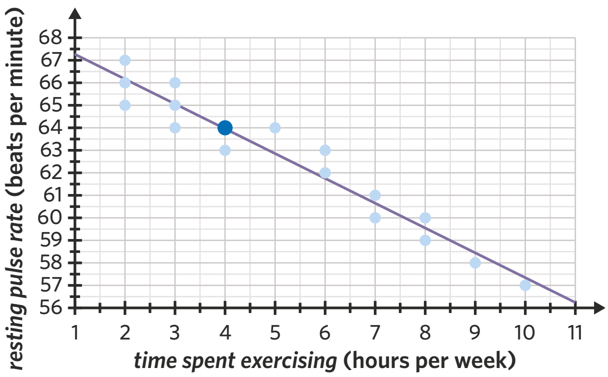
85%

The main problem that the students found with this question was understanding what the question was asking for. A number of students confused the correlation coefficient with the coefficient of determination, which led to an answer of 92% instead of 85%.

12. Explanation

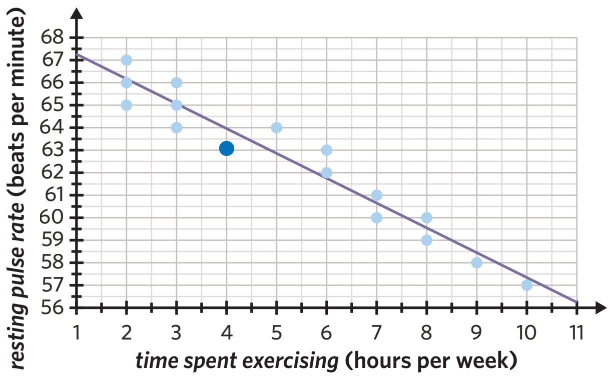
Step 1: Determine the predicted data value from the regression line on the graph.

From the regression line, it can be seen that 4 hours of *time spent exercising* per week corresponds to a predicted *resting pulse rate* of 64 beats per minute.



Step 2: Determine the actual data value from the graph.

$actual\ data\ value = 63$



Step 3: Calculate the residual value.

$$residual = actual\ data\ value - predicted\ data\ value$$
$$= 63 - 64$$
$$= -1$$

Answer

B

13. a. Explanation

Residuals are used to determine if a linear relationship between two numerical variables is the most appropriate fit for the data. Hence, it is testing the assumption of linearity.

Answer

The assumption of linearity.

A number of students struggled with understanding what the question was asking for. They did not need to prove or disprove linearity, but instead were required to explain what residual plots are actually testing.

b. Explanation

Check the residual plot for the conditions that support the assumption of linearity.

The residual plot shows an unequal number of points above and below the horizontal axis. There also appears to be a clear curved pattern. As such, this suggests that the assumption of linearity is not supported.

Answer

The residual plot displays a pattern and has an unequal distribution of points about the horizontal axis.

A large number of students struggled to answer this question, mainly because they had ignored the given statement that the residual plot does not support the assumption of linearity and instead tried to prove that linearity existed. A few others attempted to use the correlation coefficient to disprove linearity or textbook definitions that had no links to the question.

14. Explanation

Step 1: Calculate the correlation coefficient.

If the coefficient of determination is r^2, then the correlation coefficient can be calculated as the square root of the coefficient of determination.

$$r = \sqrt{0.8794}$$
$$r = 0.93776...$$

Step 2: Determine if the correlation coefficient is positive or negative.

The direction of the association between the variables on the scatterplot will dictate whether or not the correlation coefficient is positive or negative.

The scatterplot data and linear regression line are both sloping downwards, displaying a negative direction. As such, the correlation coefficient should be negative.

Answer

-0.938

A number of students incorrectly assumed that the correlation coefficient was a positive number and did not consider the direction of the association between the two variables. Students were required to look at the scatterplot to determine that the direction of the association was negative. As such, this led to students ignoring the negative option.

Questions from multiple lessons

15. C **16.** C

17. a. 25%

b. 105 students

c. All three schools have approximately the same percentage of students playing soccer: 28% of students at School A, 30% of students at School B, and 29% of students at School C. Answers may vary.

3D Data transformations

Choosing an appropriate data transformation

1. C

2. a. x-squared transformation

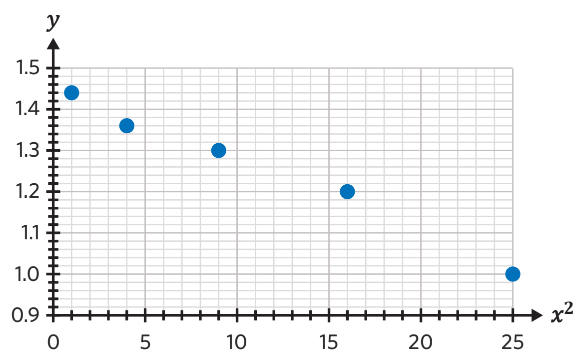
b. x-squared and y-squared transformation

3. a. D **b.** B **c.** D

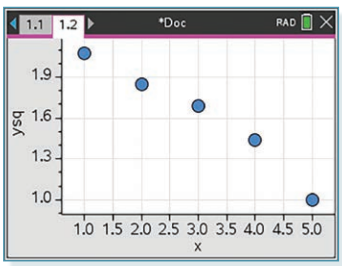
Applying a squared transformation

4. C

5. a.

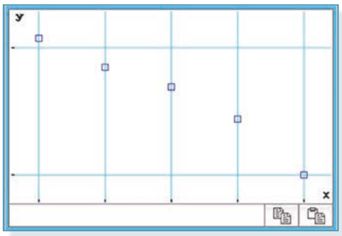


b. TI-Nspire

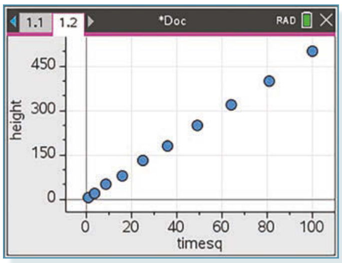


Back to contents

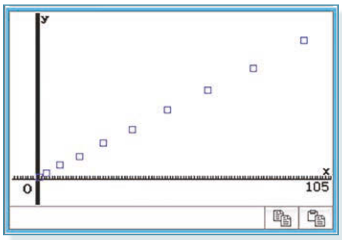
Casio ClassPad



6. TI-Nspire



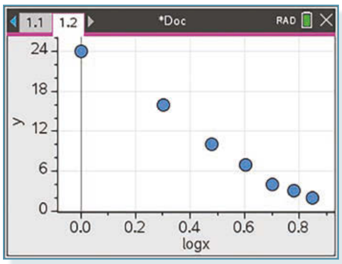
Casio ClassPad



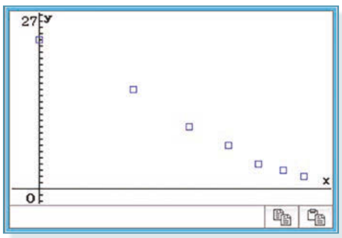
Applying a log transformation

7. A

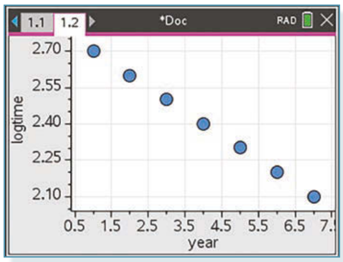
8. TI-Nspire



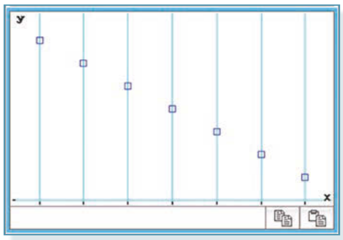
Casio ClassPad



9. TI-Nspire



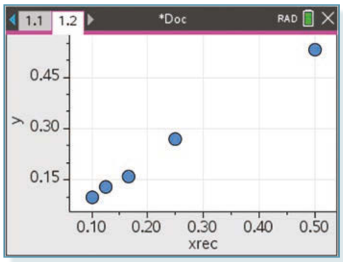
Casio ClassPad



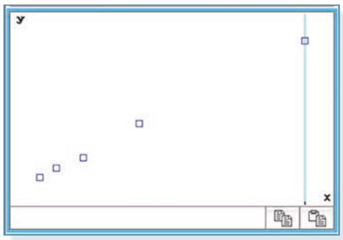
Applying a reciprocal transformation

10. D

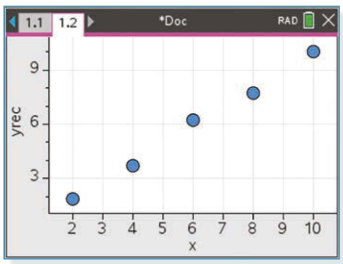
11. a. TI-Nspire



Casio ClassPad

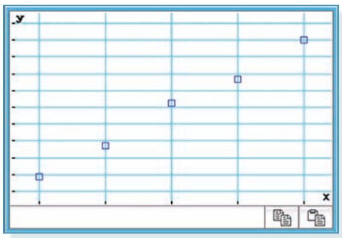


b. TI-Nspire

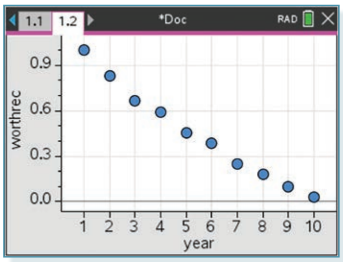


Back to contents

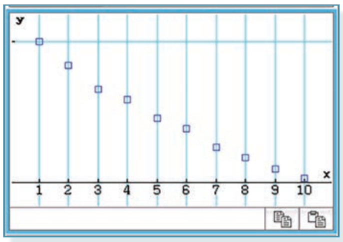
Casio ClassPad



12. **TI-Nspire**



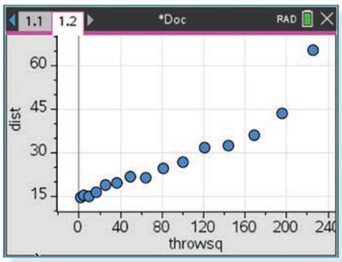
Casio ClassPad



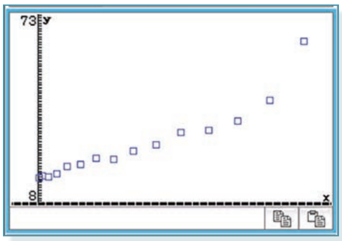
Joining it all together

13. a. A

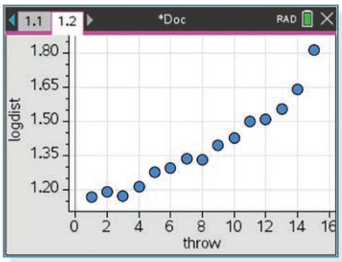
b. **TI-Nspire**



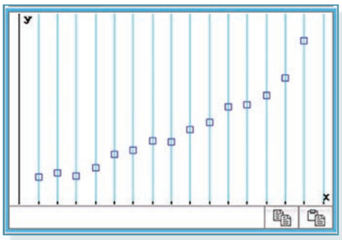
Casio ClassPad



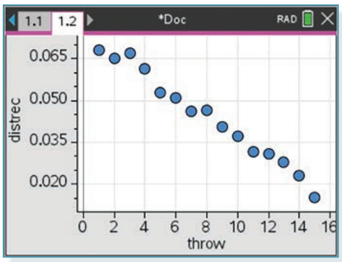
c. **TI-Nspire**



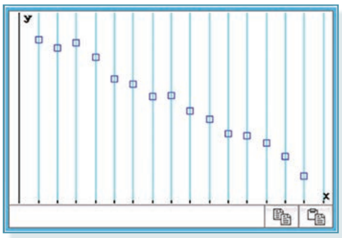
Casio ClassPad



d. **TI-Nspire**



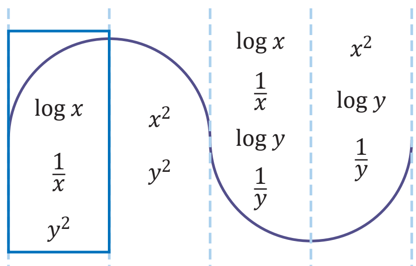
Casio ClassPad



Exam practice

14. **Explanation**

Step 1: Identify which segment of the transformation wave this scatterplot resembles.



The scatterplot most closely resembles the first segment of the transformation wave.

Step 2: Identify which transformations may be applied to this segment.

The $\log x$, $\frac{1}{x}$ or y^2 transformations may be applied.

Back to contents

Step 3: Express the transformations in terms of the variables given.

The x variable is *internet use* and the y variable is *mathematics achievement* in this instance.

Therefore, the possible ways to linearise the data are:

- *mathematics achievement* against $\log_{10}(internet\ use)$
- *mathematics achievement* against $\frac{1}{internet\ use}$
- $(mathematics\ achievement)^2$ against *internet use*

Answer

C

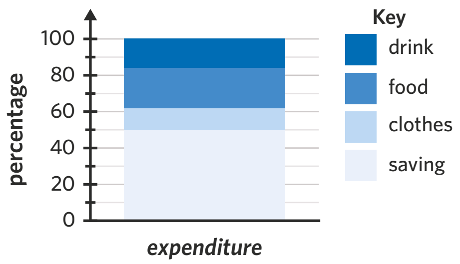
32% of students incorrectly answered either option B, D or E. These options are incorrect because the transformation has been applied to the wrong variable.

Questions from multiple lessons

15. E **16.** C

17. **a.** $275

b.



c. Yes, there is an association between the *season* and the type of *expenditure*. In summer, Eddie only saves 24% of his money, while in winter, he is able to save 50%.

3E Data transformations – applications

Calculating the equation of the least squares regression line for transformed data

1. B

2. $\log_{10}(y) = 1.379 + 0.045x$

3. $\frac{1}{views} = -0.0047 + 0.0077 \times day$

4. **a.** $time = 30.3 + 32.3 \times \log_{10}(lap)$

b. $(time)^2 = 820.7 + 367.5 \times lap$

c. $time = 30.3 + 32.3 \times \log_{10}(lap)$

Making predictions using the regression equation of transformed data

5. C **6.** D

7. **a.** 5 sprays

b. Yes. The estimate is reliable because it is an interpolation.

8. **a.**

year	11	12	13	14	15
estimated sea level rise **(cm)**	0.32	0.38	0.46	0.59	0.81

b. No. The predictions may be of limited reliability because they are extrapolations.

Joining it all together

9. **a.** $value = 21.938 + 0.252 \times (trading\ day)^2$

$\log_{10}(value) = 1.296 + 0.037 \times trading\ day$

$\frac{1}{value} = 0.044 - 0.002 \times trading\ day$

b. $\frac{1}{value} = 0.044 - 0.002 \times trading\ day$

c. **i.** $250 000

ii. Day 20

iii. $20 000

d. None of the predictions are completely reliable as they are all extrapolations.

Exam practice

10. **Explanation**

Step 1: Substitute the known value into the regression equation.

Let $time = 3$.

Step 2: Solve for the unknown value.

$\log_{10}(number\ of\ tasks) = 1.160 + 0.03617 \times 3$

$\log_{10}(number\ of\ tasks) = 1.26851$

$number\ of\ tasks = 10^{1.26851}$

$number\ of\ tasks = 18.557...$

Answer

C

11. **Explanation**

Method 1: TI-Nspire

Step 1: From the home screen, select '1: New' → '4: Add Lists & Spreadsheet'.

Step 2: Name column A 'x' and column B 'y'.

Enter the x values into column A, starting from row 1.

Enter the y values into column B, starting from row 1.

Step 3: Name column C 'yrec' (short for y-reciprocal).

Enter '=1/y' into the cell below the 'yrec' heading.

Select 'Variable Reference' → 'OK'.

Step 4: Press menu and select '4: Statistics' → '1: Stat Calculations' → '4: Linear Regression (a+bx)'.

Select 'x' in 'X List:' and 'yrec' in 'Y List:'

Select 'OK'.

Step 5: Write the equation in terms of the variables in the question.

$\frac{1}{y} = -0.0039 + 0.01179x$

Method 2: Casio ClassPad

Step 1: From the main menu, tap Statistics.

Step 2: Name the first list 'x' and the second list 'y'.

Enter the x values into list 'x', starting from row 1.

Enter the y values into list 'y', starting from row 1.

Step 3: Name the third list 'yrec' (short for y-reciprocal).

In the third list, go down to the calculation cell Cal▶ and enter '1/y'.

Step 4: Tap 'Calc' → 'Regression' → 'Linear Reg'.

Specify the data set by changing 'XList:' to 'main\x' and 'YList:' to 'main\yrec'.

Tap 'OK' to confirm.

Step 5: Write the equation in terms of the variables in the question.

$$\frac{1}{y} = -0.00395 + 0.0117973x$$

Answer – Method 1 and 2

A

> 32% of students incorrectly chose options B or C. This was likely because students recognised that the reciprocal transformation needed to be applied the the variable y, but were unable to complete the calculation of the intercept and slope.

12. a. Explanation

Method 1: TI-Nspire

Step 1: From the home screen, select '1: New' → '4: Add Lists & Spreadsheet'.

Step 2: Name column A 'year' and column B 'area'.

Enter the *year* values into column A, starting from row 1.

Enter the *area* values into column B, starting from row 1.

Step 3: Name column C 'logarea' (short for log(*area*)).

Enter '=log(area)' into the cell below the 'logarea' heading.

Step 4: Press menu and select '4: Statistics' → '1: Stat Calculations' → '4: Linear Regression (a+bx)'.

Select 'year' in 'X List:' and 'logarea' in 'Y List:'

Select 'OK'.

Step 5: Round the values of a and b to three significant figures.

$a = -14.4$

$b = 0.00854$

Step 6: Write the equation in terms of the variables in the question.

The explanatory variable is *year*.

The response variable is $\log_{10}(area)$.

Method 2: Casio ClassPad

Step 1: From the main menu, tap Statistics.

Step 2: Name the first list 'year' and the second list 'area'.

Enter the *year* values into list 'year', starting from row 1.

Enter the *area* values into list 'area', starting from row 1.

Step 3: Name the third list 'logarea' (short for log(*area*)).

In the third list, go down to the calculation cell Cal▶ and enter 'log(area)'.

Step 4: Tap 'Calc' → 'Regression' → 'Linear Reg'.

Specify the data set by changing 'XList:' to 'main\year' and 'YList:' to 'main\logarea'.

Tap 'OK' to confirm.

Step 5: Round the values of a and b to three significant figures.

$a = -14.4$

$b = 0.00854$

Step 6: Write the equation in terms of the variables in the question.

The explanatory variable is *year*.

The response variable is $\log_{10}(area)$.

Answer – Method 1 and 2

$\log_{10}(area) = -14.4 + 0.00854 \times year$

> Some students renumbered the values of the variable year as 1, 2, 3, etc. This was not asked for in the question, and resulted in incorrect values of the intercept and slope.

b. Explanation

Solve for *area*.

$\log_{10}(area) = 2.85$

$area = 10^{2.85}$

$area = 707.945...$

Answer

708 hectares

> Some students predicted the year in which $\log_{10}(area) = 2.85$, instead of solving for the *area* value. Other common incorrect answers were a result of inaccurate rounding.

c. Explanation

Identify whether the prediction is an interpolation or extrapolation.

The variable *year* ranges between 1900 and 1980.

The prediction year of 2020 falls outside the range of the data set.

Therefore, it is an extrapolation.

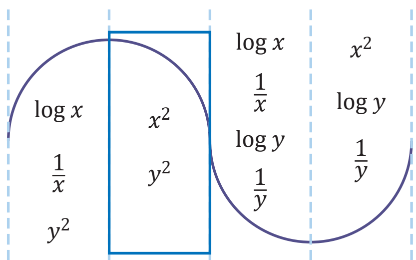
Answer

The prediction may have limited reliability because it is an extrapolation.

13. Explanation

Method 1: TI-Nspire

Step 1: Determine the appropriate transformation.



The scatterplot most closely resembles the second segment of the transformation wave. Hence, the x^2 or y^2 transformations may be applied.

As the question specifies that the transformation must be applied to the variable *mean age*, the appropriate transformation is $(mean\ age)^2$.

Step 2: From the home screen, select '1: New' → '4: Add Lists & Spreadsheet'.

Back to contents

Step 3: Name column A 'age' and column B 'height'.

Enter the *mean age* values into column A, starting from row 1.

Enter the *mean height* values into column B, starting from row 1.

Step 4: Name column C 'agesq' (short for *mean age* squared).

Enter '=age^2' into the cell below the 'agesq' heading.

Step 5: Press menu and select '4: Statistics' → '1: Stat Calculations' → '4: Linear Regression (a+bx)'.

Select 'agesq' in 'X List:' and 'height' in 'Y List:'

Select 'OK'.

Step 6: Round the values of a and b to four significant figures.

$a = 167.9$

$b = 0.001621$

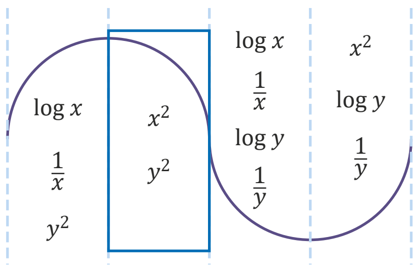
Step 7: Write the equation in terms of the variables in the question.

The explanatory variable is $(mean\ age)^2$.

The response variable is *mean height*.

Method 2: Casio ClassPad

Step 1: Determine the appropriate transformation.



The scatterplot most closely resembles the second segment of the transformation wave. Hence, the x^2 or y^2 transformations may be applied.

As the question specifies that the transformation must be applied to the variable *mean age*, the appropriate transformation is $(mean\ age)^2$.

Step 2: From the main menu, tap Statistics.

Step 3: Name the first list 'age' and the second list 'height'.

Enter the *mean age* values into list 'age', starting from row 1.

Enter the *mean height* values into list 'height', starting from row 1.

Step 4: Name the third list 'agesq' (short for *mean age* squared).

In the third list, go down to the calculation cell Cal▶ and enter 'age^2'.

Step 5: Tap 'Calc' → 'Regression' → 'Linear Reg'.

Specify the data set by changing 'XList:' to 'main\agesq' and 'YList:' to 'main\height'.

Tap 'OK' to confirm.

Step 6: Round the values of a and b to four significant figures.

$a = 167.9$

$b = 0.001621$

Step 7: Write the equation in terms of the variables in the question.

The explanatory variable is $(mean\ age)^2$.

The response variable is *mean height*.

Answer – Method 1 and 2

$mean\ height = 167.9 - 0.001621 \times (mean\ age)^2$

A majority of students who attempted this question chose the squared transformation correctly. A small number tried to work with a reciprocal transformation or a logarithm transformation. The rounding to significant figures was correctly done by a small proportion of students.

Questions from multiple lessons

14. D

15. D

16. **a.** 159.1

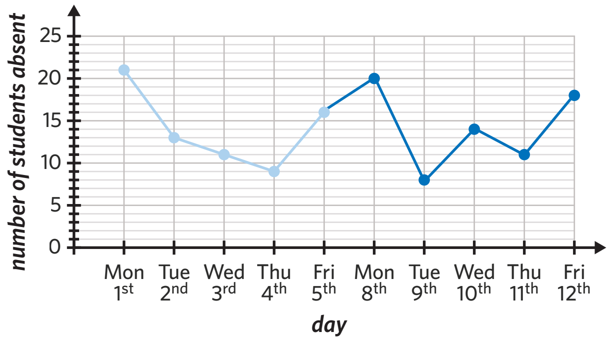
b. $words\ written = 159.1 + 6.074 \times time\ spent$

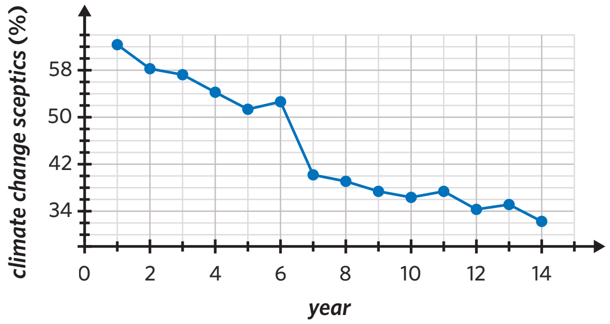
Back to contents

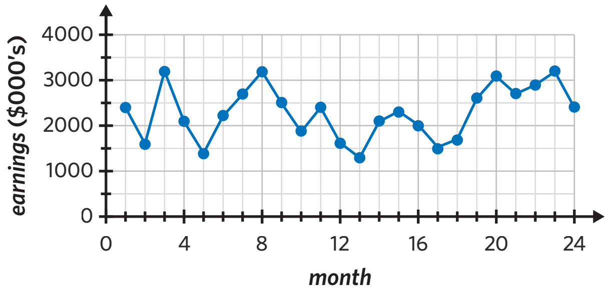
4A Time series data and their graphs

Constructing time series plots

1. C

2. 

3. 

4. 

Identifying characteristics of time series data

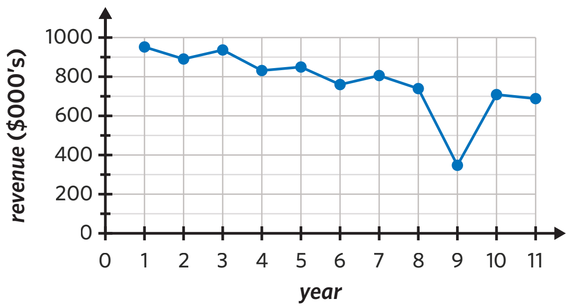
5. A

6. a. Increasing trend including an outlier
   b. Seasonality and structural change

7. Cyclical variation, with cycles of approximately 100 000 years

8. Irregular fluctuations until approximately 1900, then a structural change

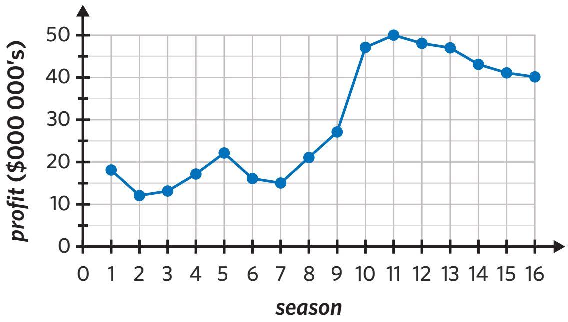
Joining it all together

9. a. 

   b. Decreasing trend with an outlier in the year 2020.

10. a.

season	*coded season*	*profit* ($000 000's)
Summer 2018	1	18
Autumn 2018	2	12
Winter 2018	3	13
Spring 2018	4	17
Summer 2019	5	22
Autumn 2019	6	16
Winter 2019	7	15
Spring 2019	8	21
Summer 2020	9	27
Autumn 2020	10	47
Winter 2020	11	50
Spring 2020	12	48
Summer 2021	13	47
Autumn 2021	14	43
Winter 2021	15	41
Spring 2021	16	40

   b. 

   c. Increasing trend with seasonality, with peaks occurring in summer each year.

   d. Structural change

   e. Decreasing trend

Exam practice

11. **Explanation**

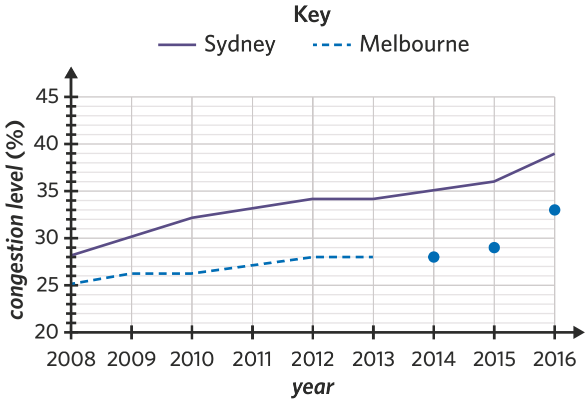
Step 1: Identify the correct values in the table.

The values for Melbourne are in the middle row.
The years that we need to plot the values for are 2014, 2015, and 2016.

	congestion level (%)								
year (2000's)	'08	'09	'10	'11	'12	'13	'14	'15	'16
Melb.	25	26	26	27	28	28	28	29	33
Syd.	28	30	32	33	34	34	35	36	39

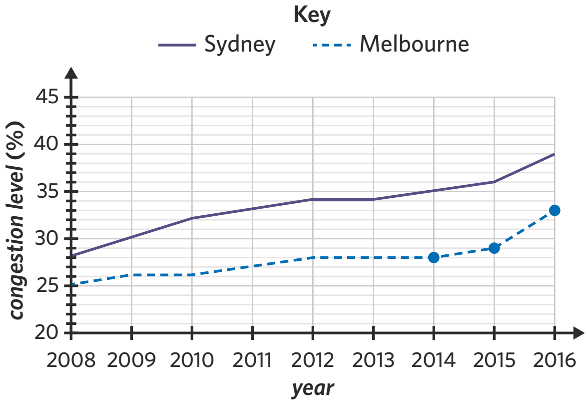
Back to contents

Step 2: Plot each data point on the graph.



Step 3: Connect the data points with straight dotted lines.

Answer



12. Explanation

Step 1: Look for a general trend in the data.

There is no general trend in this time series plot. The peaks and troughs in the data show no clear pattern of increase or decrease.

Step 2: Look for any seasonality.

The peaks seem to occur roughly every 12 months, indicating seasonality.

Answer

C

15% of students incorrectly chose option D. This is because they thought there was also an increasing trend, however there is not enough of a clear pattern in the peaks and troughs to suggest this.

13. Explanation

Step 1: Look for a general trend in the data.

The peaks and troughs in the data show that the data is decreasing slightly over time.

Step 2: Look for any seasonality.

The peaks do not occur at regular intervals. Some of the peaks are 1.25 years apart, while some are 0.75 years apart. The peaks and troughs must occur at regular intervals for seasonality to be present.

Answer

D

41% of students incorrectly chose option E. This is likely because they thought the fluctuations represented seasonality, however the peaks and troughs must be more regular in order to indicate seasonality.

Questions from multiple lessons

14. C **15.** E

16. a. $\textit{yearly salary} = -116\,275.9 + 128\,883.4 \times \log_{10}(\textit{age})$

b. $79 435

4B Smoothing – moving means

Smoothing over an odd number of data points using moving means

1. A

2. a. 38 minutes **b.** 39.4 minutes

3. a.

-	15.0	17.7	19.3	17.0	16.0	15.3	15.3	16.0	-

b.

-	-	16.8	16.4	17.8	17.0	15.0	15.8	-	-

4. a. 19

b.

-	6.0	9.0	9.7	13.0	15.0	19.0	22.0	24.3	27.7	30.7	-

c.

-	-	8.0	10.8	12.8	15.8	18.6	21.6	24.6	28.0	-	-

5. a. 0.92 **b.** 0.95

c.

-	1.00	1.01	0.99	0.94	0.89	0.88	-

d.

-	-	0.99	0.97	0.94	0.91	-	-

Smoothing over an even number of data points using moving means

6. A

7. a. 104 points **b.** 104.5 points

c. 100.5 points

8. a. 1.2 **b.** 1.3

c.

-	2.0	1.9	1.7	1.4	1.2	1.2	1.4	1.6	1.7	1.8	-

d.

-	-	1.8	1.6	1.4	1.3	1.3	1.4	1.6	1.7	-	-

9. a. 22.85 kg

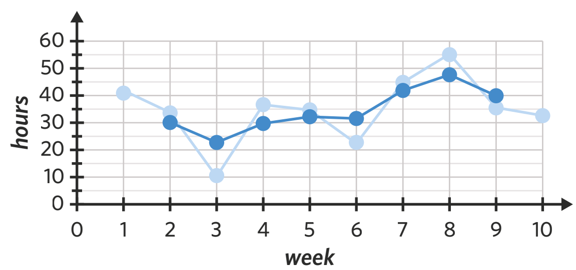
b.

-	-	22.2	23.5	24.3	24.8	25.0	24.7	24.1	-	-

Plotting and interpreting a mean smoothed time series

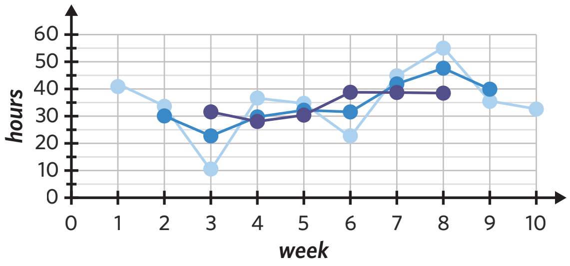
10. B

11. a.

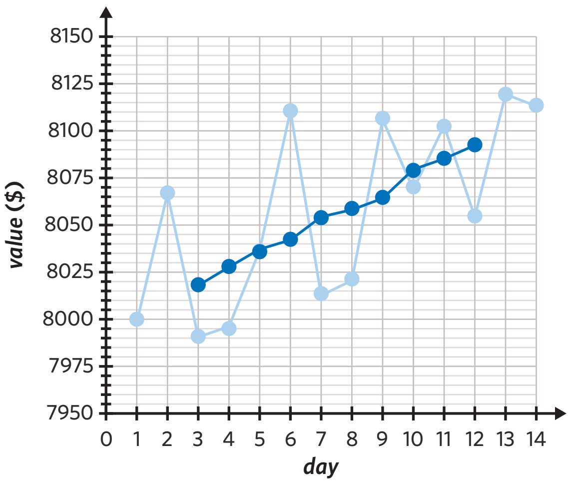


Back to contents

b.



12. a.



b. Increasing trend.

Over the first two weeks, Amira's crypto portfolio has increased in value.

Joining it all together

13. a. 250 **b.** 136 **c.** 155 **d.** 402.5

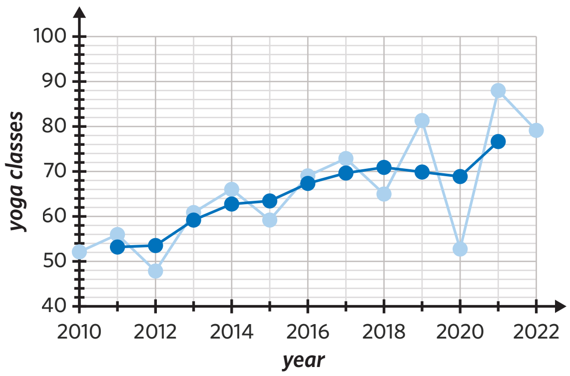
e. 190 **f.** 201.7

14. a. 2014 **b.** 61 yoga classes

c.

-	53	53	59	63	63	68	70	71	70	69	77	-

d.



e. Increasing trend.

The number of *yoga classes* that Alfredo has attended has increased from 2010 to 2022.

Exam practice

15. Explanation

There are 12 data points. Seven-mean smoothing requires a data point to have at least 3 values on either side of it, so the first 3 and last 3 values cannot be smoothed using seven-mean smoothing. Hence, there would be 6 smoothed data points.

Answer

C

16. Explanation

Step 1: Find the *maximum wind speed* for day 11 as well as the two values on either side and write them in the order they appear in the time series.

22 19 22 43 37

Step 2: Find the mean of the first four values.

$$mean_1 = \frac{22 + 19 + 22 + 43}{4} = 26.5$$

Step 3: Find the mean of the last four values.

$$mean_2 = \frac{19 + 22 + 43 + 37}{4} = 30.25$$

Step 4: Calculate the centred mean.

$$mean_{centred} = \frac{30.25 + 26.5}{2}$$
$$= 28.375$$

Answer

D

17. Explanation

Step 1: Write an equation for Thursday's five-mean smoothed *quantity*.

Let x be the total *quantity* sold over Wednesday, Thursday, and Friday.

The five-mean smoothed value is the sum of the *quantity* sold from Tuesday to Saturday, divided by 5.

$$\frac{186 + x + 346}{5} = 206$$

Step 2: Solve the equation for x.

$$186 + x + 346 = 5 \times 206$$
$$x + 532 = 1030$$
$$x = 1030 - 532$$
$$x = 498$$

Step 3: Calculate Thursday's three-mean smoothed value.

The three-mean smoothed value is the sum of the *quantity* sold from Wednesday to Friday, divided by 3.

x is *quantity* sold from Wednesday to Friday

$$\frac{x}{3} = \frac{498}{3}$$
$$= 166$$

Answer

B

40% of students incorrectly selected options C or D. This was because they either thought the three-mean value would be the same as the five-mean value, or they assumed it would fall roughly halfway between the Tuesday and Saturday values.

18. Explanation

Step 1: Find the *maximum daily rainfall* for October as well as its adjacent values and write them in the order they appear in the time series.

124 140 225

Step 2: Find the mean of the first two values.

$$mean_1 = \frac{124 + 140}{2} = 132$$

Step 3: Find the mean of the last two values.

$$mean_2 = \frac{140 + 225}{2} = 182.5$$

Back to contents

Step 4: Calculate the centred mean.

$$mean_{centred} = \frac{132 + 182.5}{2} = 157.25$$

Answer

$$mean_{centred} = \frac{\frac{124 + 140}{2} + \frac{140 + 225}{2}}{2}$$

$$= \frac{132 + 182.5}{2}$$

$$= 157.25 \text{ mm}$$

49% of students got 0 marks on this question. This was generally because they attempted to use values from the graph instead of the table, or they didn't understand the question and didn't attempt it.

Questions from multiple lessons

19. C **20.** B

21. **a.** Strong, linear, positive correlation.

b. $a = 15.67$

$b = 0.92$

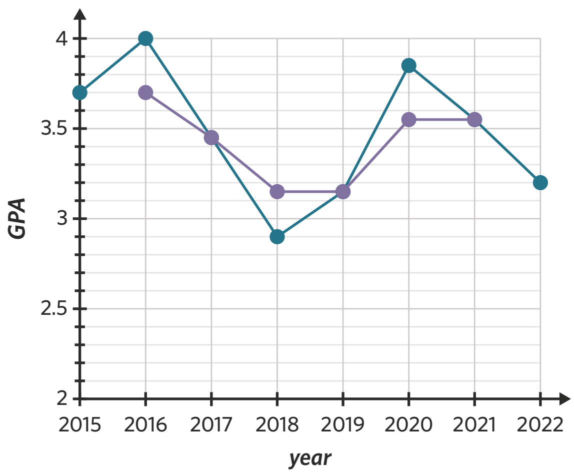
4C Smoothing – moving medians

Smoothing using three-moving medians

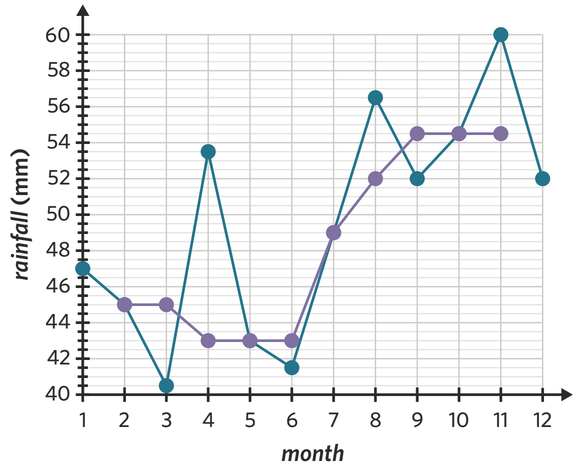
1. B

2. **a.** 26 °C **b.** 21 °C **c.** 25 °C **d.** 27 °C

3. **a.**

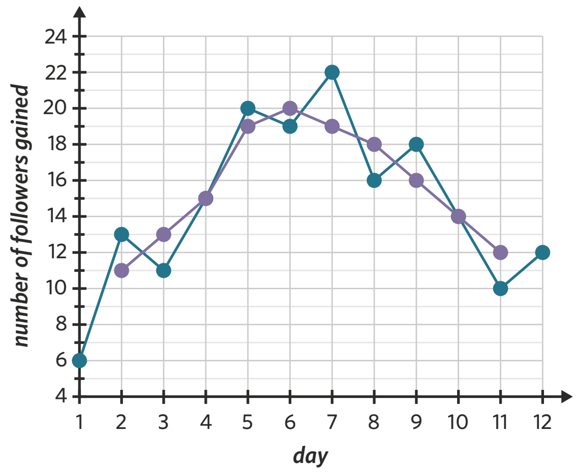


b.



4. **a.** 16 followers

b.



c. There are two key trends in the smoothed time series. The *number of followers gained* is generally increasing until day 6. After day 6, *number of followers gained* decreases.

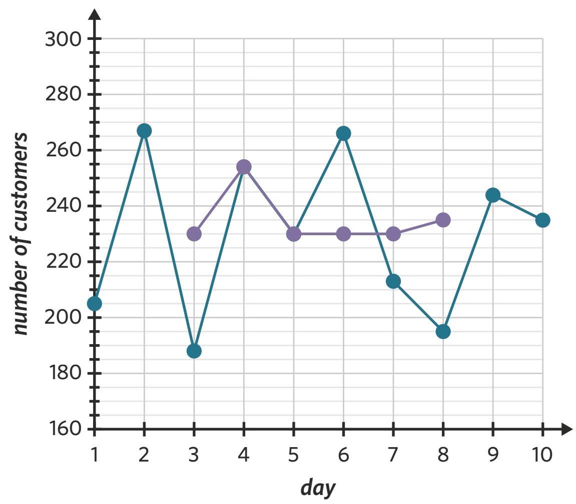
Smoothing using five (or more)-moving medians

5. B

6. **a.** 65 potatoes **b.** 43 potatoes

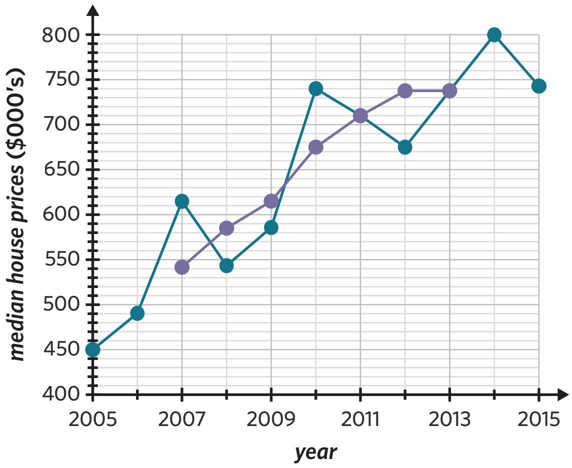
c. 59 potatoes **d.** 65 potatoes

7. **a.**



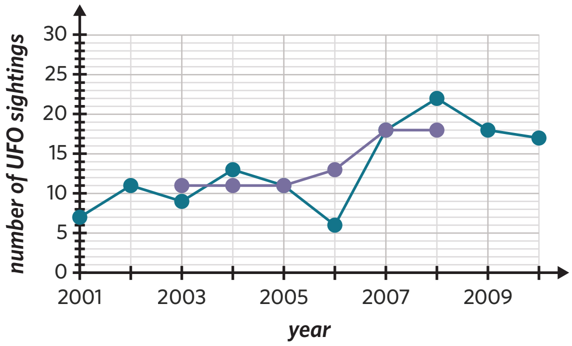
Back to contents

b.



8. **a.** 13 UFO sightings

b.



c. Increasing trend

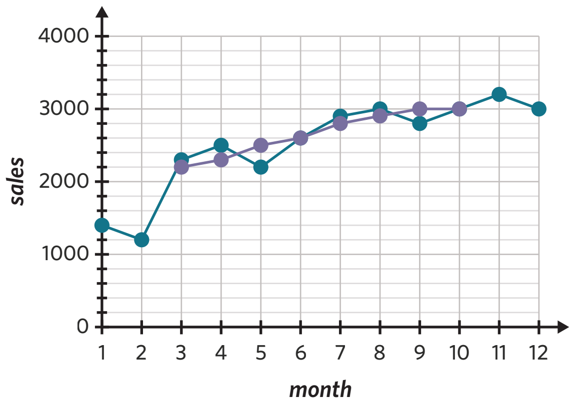
Joining it all together

9. C

10. **a.** 22 mm **b.** 18 mm **c.** 20 mm **d.** 21 mm

11. **a.** 1400 sales

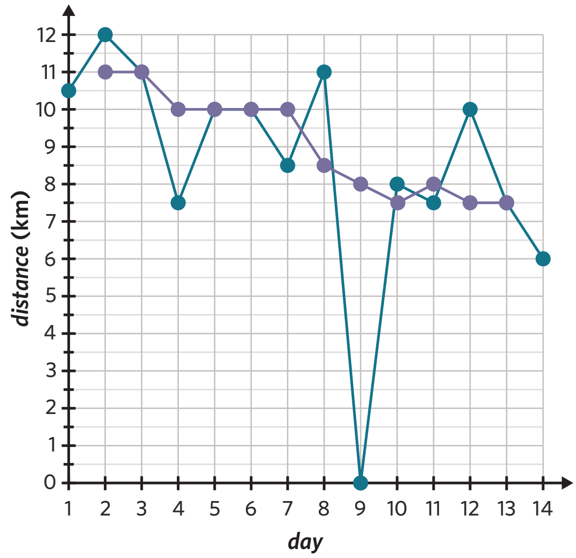
b.



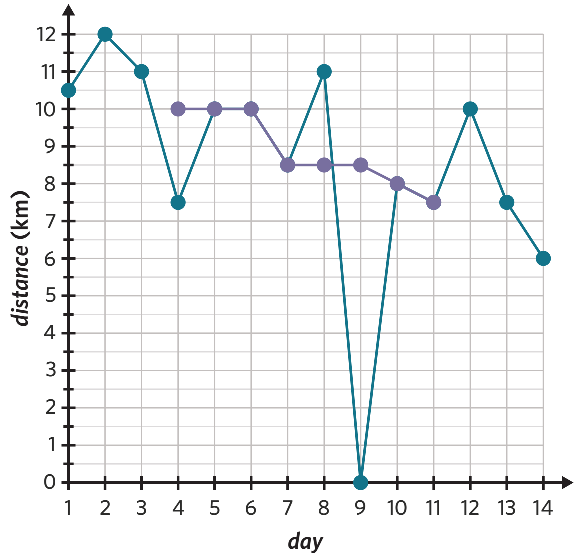
c. Increasing trend

12. **a.** 8 km

b.



c.



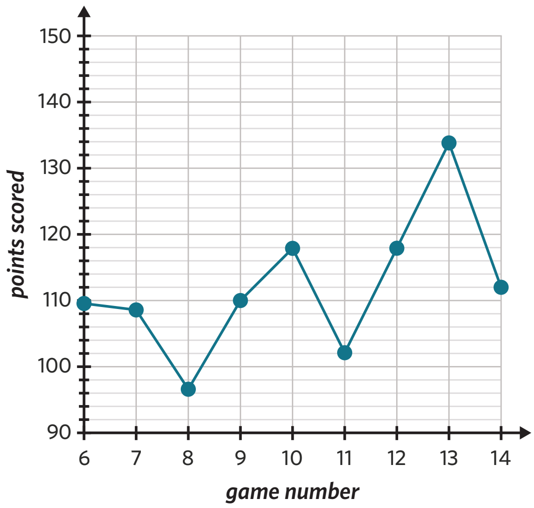
d. Decreasing trend

Exam practice

13. **Explanation**

Step 1: Identify the necessary data points.

The nine-median smoothed value for game number 10 is the median *points scored* from game 6 to game 14.

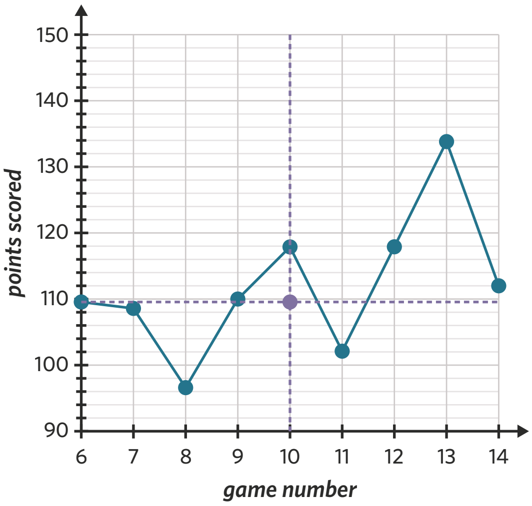


Back to contents

Step 2: Identify the median *points scored* from game 6 to game 14.

The variable *points scored* is represented by the vertical height of the points.

As there are nine values, the median is the fifth value from bottom to top.



The median *points scored* is closest to 110.

Answer

C

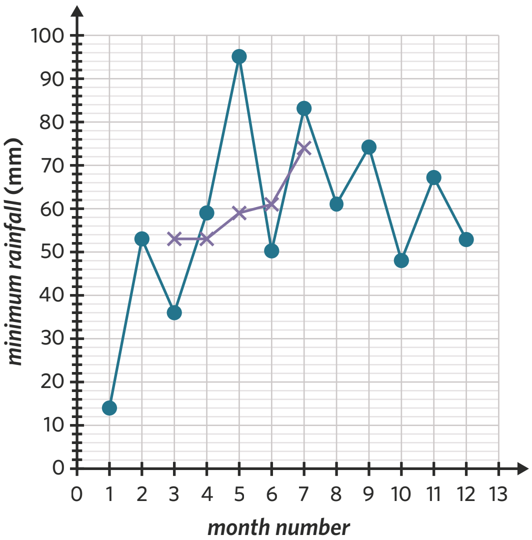
14. Explanation

Step 1: Determine the five-median smoothed value for month 7.

The smoothed value for month 7 is the median *minimum rainfall* for months 5 to 9.

The *minimum rainfall* values for months 5 to 9 are approximately 95, 50, 83, 61, and 74 mm.

The median *minimum rainfall* is approximately 74 mm.

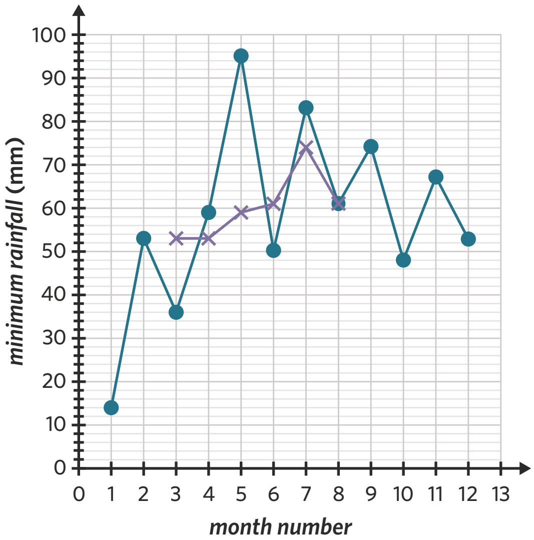


Step 2: Determine the five-median smoothed value for month 8.

The smoothed value for month 8 is the median *minimum rainfall* for months 6 to 10.

The *minimum rainfall* values for months 6 to 10 are approximately 50, 83, 61, 74, and 48 mm.

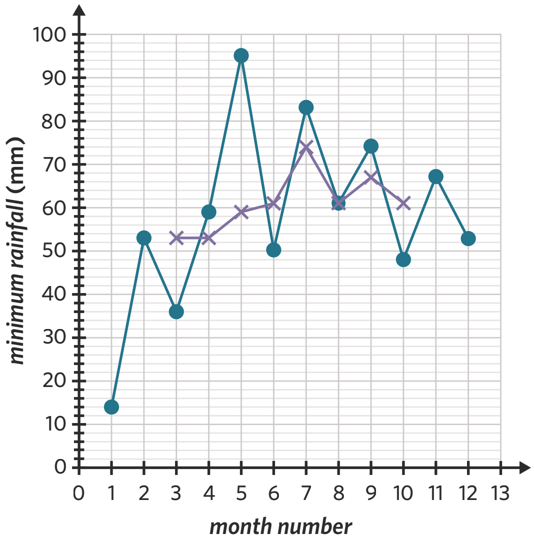
The median *minimum rainfall* is approximately 61 mm.



Step 3: Apply this technique for the months 9 and 10.

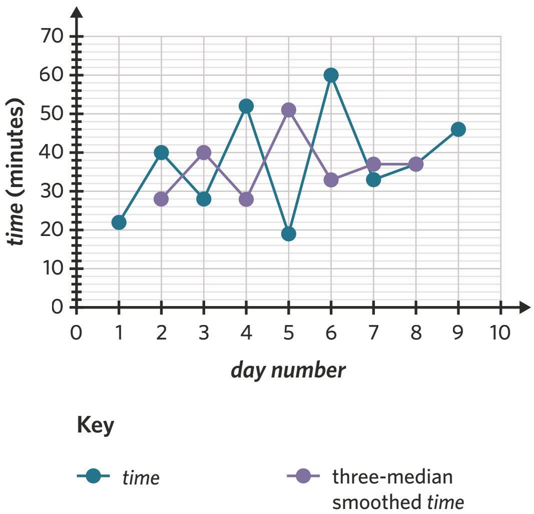
The smoothing will stop at month 10, as this is the last month with two data values on each side.

Answer



15. Explanation

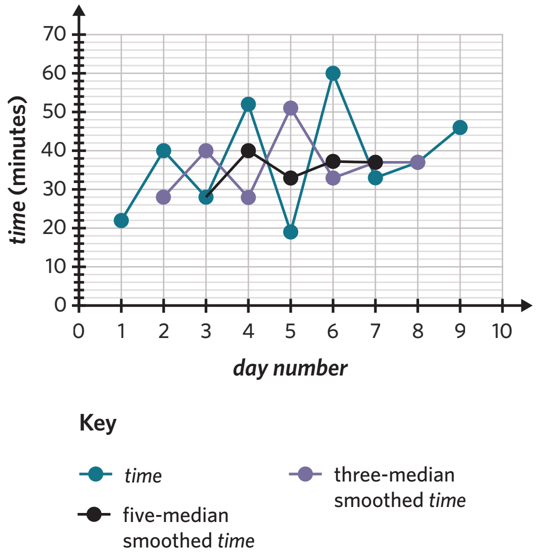
Step 1: Smooth the time series using three-median smoothing.



4C ANSWERS

Back to contents

Step 2: Smooth the time series using five-median smoothing.



Step 3: Identify on which day both of the smoothed values are the same.

On day 7, the three-median and five-median smoothed values are both 37 minutes.

Answer

E

Questions from multiple lessons

16. E **17.** C

18. a. 8 **b.** No outlier

4D Seasonal adjustments

Calculating and interpreting seasonal indices

1. B

2. a. 0.70

b. On average, sales in quarter 1 are 70% of, or 30% less than, the average quarter.

3. D

4. a. summer: 0.72, autumn: 1.03, winter: 1.26, spring: 0.99

b. Mon: 0.55, Tue: 1.23, Wed: 1.28, Thu: 1.62, Fri: 1.20, Sat: 0.76, Sun: 0.37

c. quarter 1: 0.68, quarter 2: 1.04, quarter 3: 1.30, quarter 4: 0.98

Deseasonalising a time series

5. D **6.** C

7. a. summer: 30.8, autumn: 40.0, winter: 28.7, spring: 31.7

b. Mon: 44.0, Tue: 32.5, Wed: 38.2, Thu: 36.1, Fri: 37.2, Sat: 43.4, Sun: 21.7

8. B

9. a. 60% **b.** 11 360 holidays booked

Reseasonalising a time series

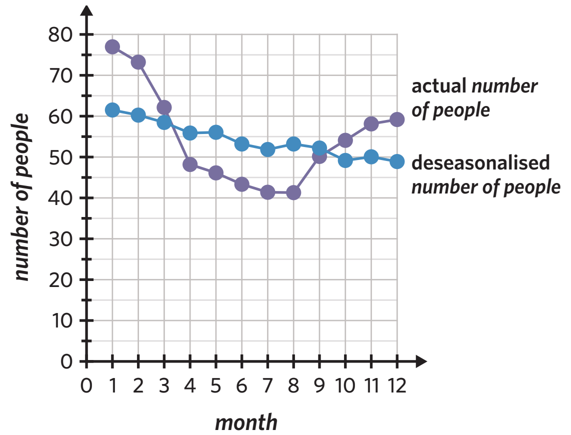
10. A **11.** D

12. Mon: 42, Tue: 49, Wed: 29, Thu: 56, Fri: 84, Sat: 154, Sun: 126

Plotting and interpreting a deseasonalised time series

13. A

14. a.



b. The number of people wearing Birkenstocks is decreasing over time.

Joining it all together

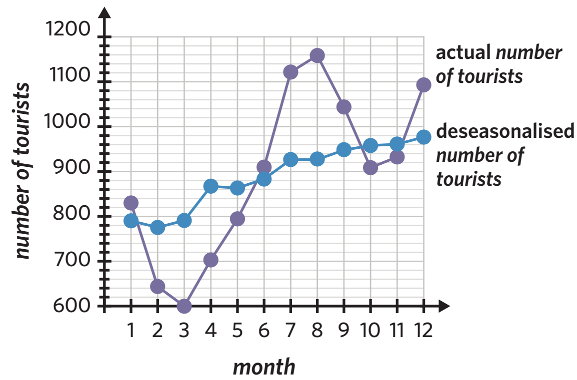
15. a. summer: 0.59, spring: 1.03

b. On average, sales in autumn are 6% greater than the average season.

c. To correct for seasonality, the amount of sales in winter should be decreased by 24%.

16. a. 0.95 **b.** 601 **c.** 977.7

d.



e. The number of tourists is increasing over time.

Exam practice

17. Explanation

Step 1: Identify the actual value and seasonal index for May 2019.

$\textit{actual value} = 92.6$

$\textit{seasonal index} = 1.222$

Step 2: Substitute the values into the formula.

$$\textit{deseasonalised value} = \frac{92.6}{1.222} = 75.77\ldots$$

Answer

B

Back to contents

18. a. **Explanation**

Step 1: Calculate the seasonal indices for the summer, autumn, and spring in 2015.

$$\begin{aligned} \textit{mean of all seasons} &= \frac{142 + 156 + 222 + 120}{4} \\ &= 160 \end{aligned}$$

Summer 2015: $\textit{seasonal index} = \frac{142}{160} = 0.8875$

Autumn 2015: $\textit{seasonal index} = \frac{156}{160} = 0.975$

Spring 2015: $\textit{seasonal index} = \frac{120}{160} = 0.75$

Step 2: Calculate the seasonal indices for the summer, autumn, and spring in 2016.

$$\begin{aligned} \textit{mean of all seasons} &= \frac{135 + 153 + 216 + 96}{4} \\ &= 150 \end{aligned}$$

Summer 2016: $\textit{seasonal index} = \frac{135}{150} = 0.9$

Autumn 2016: $\textit{seasonal index} = \frac{153}{150} = 1.02$

Spring 2016: $\textit{seasonal index} = \frac{96}{150} = 0.64$

Step 3: Find the average seasonal indices by calculating the mean for each season.

Summer: $\textit{seasonal index} = \frac{0.8875 + 0.9}{2} \approx 0.89$

Autumn: $\textit{seasonal index} = \frac{0.975 + 1.02}{2} \approx 1.00$

Spring: $\textit{seasonal index} = \frac{0.75 + 0.64}{2} \approx 0.70$

Answer

summer: 0.89, autumn: 1.00, spring: 0.70

> The main problem students found with this question was rounding to 2 decimal places, in particular rounding the autumn value of 0.9975 to the nearest whole number (1) instead of 2 decimal places (1.00).

b. **Explanation**

Step 1: Identify the actual value for winter 2017 from Table 2 and the seasonal index.

$\textit{actual value} = 262$

$\textit{seasonal index} = 1.41$

Step 2: Substitute the values into the formula.

$$\begin{aligned} \textit{deseasonalised value} &= \frac{262}{1.41} \\ &= 185.816 \end{aligned}$$

Answer

186

19. **Explanation**

Step 1: Find the reciprocal of the seasonal index.

Divide 1 by the seasonal index.

$1 \div 1.6 = 0.625$

Step 2: Interpret the value of the reciprocal.

The reciprocal value of 0.625 suggests that deseasonalised values are 62.5% of the actual value. This means that values for the season need to be decreased by 37.5%.

Answer

A

> 49% of students incorrectly selected options C or D. Many would have assumed that the value needed to be reduced by 60%. Option C is closest to this conclusion and option D is the only response with a 60% adjustment.

Questions from multiple lessons

20. E

21. E

22. a. 22.0 cm
b. 26.9%
c. The association is linear.

4E Time series data and least squares regression modelling

Modelling time series data

1. B

2. a. 413 customers
b. 510 customers
c. 608 customers
d. 705 customers

3. a. $\textit{number of steps} = 6764.3 + 485.7 \times \textit{day}$
13 564 steps
b. $\textit{number of steps} = 8492.9 - 414.3 \times \textit{day}$
2693 steps

4. a. 1132 alien sightings
b. The prediction is an example of extrapolation since it is outside the range of recorded data.
c. On average, the number of *alien sightings* in the northern hemisphere increases by 41.2 each year.

5. a. $\textit{bank balance} = 9588.27 - 573.46 \times \textit{week}$
b. $412.91
c. 12 weeks
d. $573.46
e. $9588.27

Modelling seasonal data

6. C

7. $81 253.25

8. $\textit{deseasonalised average price} = 2.76 + 0.21 \times \textit{coded season}$

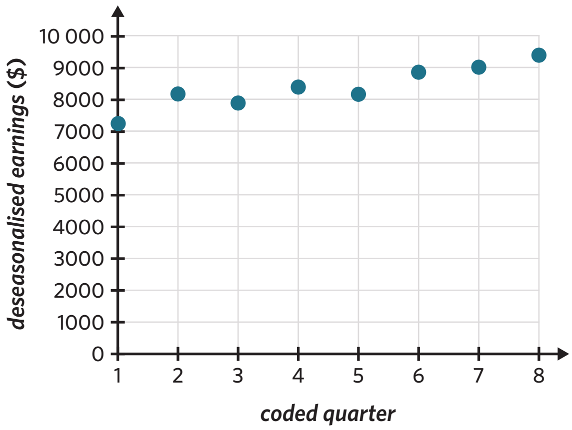
9. a. $\textit{deseasonalised profit} = 7923.9 - 194.9 \times \textit{coded month}$
b. $4791
c. No, the prediction may be of limited reliability since it is an example of extrapolation.

Back to contents

10. a.

coded quarter	*deseasonalised earnings* ($)
1	7237
2	8168
3	7888
4	8387
5	8157
6	8853
7	9011
8	9387

b.



There is a positive linear trend. A least squares regression line would be suitable.

c. *deseasonalised earnings* = 7211 + 261 × *coded quarter*

d. $10 956.28

e. Q1 2024

Joining it all together

11. a. *number of students* = 62.78 + 1.19 × *coded day*

b. 160 students

c.

coded day	*deseasonalised number of students*
1	57.02
2	68.22
3	58.93
4	66.98
5	70.37
6	63.64
7	67.96
8	71.05
9	71.03
10	90.18
11	77.36
12	75.31
13	84.42
14	84.47

d. *deseasonalised number of students* = 57.43 + 1.93 × *coded day*

e. 175 students

f. On average, the *deseasonalised number of students* in the library increases by 1.93 each day.

Exam practice

12. a. i. **Explanation**

Step 1: Identify the slope.

The slope of the regression line is 1.15.

Step 2: Interpret the slope.

The slope indicates the average change in the response variable for every one-unit increase in the explanatory variable. In this case, the slope indicates that, on average, the *congestion level* increases by 1.15% each year.

Answer

1.15% per year

64% of students received no marks for this question. Many students did not identify the slope of the regression line as a measure of the average rate of increase per year. Some students answered 15% due to confusion with the *R* value used in financial mathematics.

ii. **Explanation**

Substitute *congestion level* = 43 into the regression equation and solve for *year*.

congestion level = −2280 + 1.15 × *year*

43 = −2280 + 1.15 × *year*

year = 2020

Answer

2020

b. **Explanation**

Method 1: TI-Nspire

Step 1: From the home screen, select '1: New' → '4: Add Lists & Spreadsheet'.

Step 2: Name column A 'year' and column B 'congestion'.

Enter Melbourne's *year* values into column A, starting from row 1.

Enter Melbourne's *congestion level* values into column B, starting from row 1.

Step 3: Press menu and select '4: Statistics' → '1: Stat Calculations' → '4: Linear Regression (a+bx)'. Select 'year' in 'X List:' and 'congestion' in 'Y List:'

Step 4: Write down the equation for the least squares regression line, and round the values of *a* and *b* to four significant figures.

$y = -1515 + 0.7667 \times x$

Step 5: Rewrite the equation in terms of the variables in the question.

The explanatory variable is *year*.

The response variable is *congestion level*.

Method 2: Casio ClassPad

Step 1: From the main menu, tap Statistics.

Step 2: Name the first list 'year' and the second list 'congest'.

Enter Melbourne's *year* values into list 'year', starting from row 1.

Enter Melbourne's *congestion level* values into list 'congest', starting from row 1.

Step 3: Tap 'Calc' → 'Regression' → 'Linear Reg'.

Specify the data set by changing 'XList:' to 'main\year' and 'YList:' to 'main\congest'. Tap 'OK'.

Back to contents

Step 4: Change the form of the equation in the drop down box to 'y=a+b·x'. Write down the equation for the least squares regression line, and round the values of a and b to four significant figures.

$y = -1515 + 0.7667 \times x$

Step 5: Rewrite the equation in terms of the variables in the question.

The explanatory variable is *year*.

The response variable is *congestion level*.

Answer – Method 1 and 2

$\textit{congestion level} = -1515 + 0.7667 \times \textit{year}$

Questions from multiple lessons

13. A

14. B

15. **a.** \$5508.70 **b.** 62.7%

Back to contents

5A Recurrence relations and their graphs

Interpreting a recurrence relation

1. C **2.** D

3. a. Initial value is 4; geometric decay

b. Initial value is 3.2; geometric growth and linear decay

c. Initial value is 14; linear growth

Constructing a recurrence relation

4. B

5. a. Common ratio of 2 **b.** Common difference of 7

c. Common ratio of 0.9 **d.** Common ratio of −1.2

6. a. $u_0 = 22, \quad u_{n+1} = u_n + 4$

Note: Any letter could be used to denote the terms of the sequence.

b. $u_0 = 100, \quad u_{n+1} = 0.2 \times u_n$

Note: Any letter could be used to denote the terms of the sequence.

c. $u_0 = 14, \quad u_{n+1} = 2 \times u_n - 3$

Note: Any letter could be used to denote the terms of the sequence.

d. $u_0 = 3, \quad u_{n+1} = 0.4 \times u_n + 6$

Note: Any letter could be used to denote the terms of the sequence.

7. a. $u_0 = 4, \quad u_{n+1} = u_n + 2.5$

b. $u_0 = 768, \quad u_{n+1} = 0.25 \times u_n$

c. $u_0 = \frac{1}{4}, \quad u_{n+1} = 5 \times u_n$

d. $u_0 = 3, \quad u_{n+1} = 3 \times u_n + 2$

8. a. 96 people

b. $P_1 = 3, \quad P_{n+1} = 2 \times P_n$

Generating a sequence of terms using a recurrence relation

9. C

10. a. $U_2 = 17, \quad U_4 = 23, \quad U_7 = 32$

b. $V_2 = \frac{1}{3}, \quad V_4 = 3, \quad V_7 = 81$

c. $W_2 = 0, \quad W_4 = -12, \quad W_7 = -124$

d. $X_2 = 13.84, \quad X_4 = 9.258, \quad X_7 = -3.060$

11. a. 4.5, 4.25, 4, 3.75, 3.5

b. 10, 11, 12.1, 13.31, 14.641

c. 17, 29, 65, 173, 497

d. 12.5, 15, 17, 18.6, 19.88

12. $G_{15} = 293$ **13.** $M_0 = 11$

Joining it all together

14. a. 64 **b.** 9.75

15. a. 5 years

b. $V_0 = 3500, \quad V_{n+1} = V_n - 550$

16. 2 cm

Exam practice

17. Explanation

To solve this question, check whether each option is true or false.

This can be done either by analysing the sequence or by determining the value of each term in the sequence using the recurrence relation.

A: This is false. It is not an arithmetic sequence as it does not have a common difference. ✗

B: This is false. The recurrence relation only generates the first term after the initial value. ✗

C: This is false. It is not a geometric sequence as it does not have a common ratio. ✗

D: This is true. The recurrence relation generates all five terms. ✓

E: This is false. The recurrence relation only generates the first term after the initial value. ✗

Answer

D

Students need to ensure that the recurrence relation generates all the terms in the sequence and not just the first term.

18. Explanation

Substitute the initial value T_0 into the recurrence relation.

$$\begin{aligned} T_1 &= T_0 + 3 \\ &= 10 + 3 \\ &= 13 \end{aligned}$$

The number 13 appears in the sequence as T_1.

Answer

A

The subscript denotes the term's position rather than its value. Students may have chosen E if they incorrectly interpreted the subscript as the value of the term.

19. Explanation

Step 1: Calculate L_1.

$$\begin{aligned} L_1 &= L_0 + C \\ &= 37 + C \end{aligned}$$

Step 2: Substitute L_1 into the recurrence relation.

$$\begin{aligned} L_2 &= L_1 + C \\ &= (37 + C) + C \\ &= 37 + 2C \end{aligned}$$

Step 3: Substitute $L_2 = 25$ and solve for C.

$$\begin{aligned} 25 &= 37 + 2C \\ -12 &= 2C \\ -6 &= C \end{aligned}$$

Answer

A

Back to contents

20. Explanation

Step 1: Identify the value needed to generate A_3.

A_3 can be found from A_2.

$A_3 = 2A_2 + 4$

Step 2: Calculate A_2.

$A_1 = 2 \times A_0 + 4$
$= 2 \times 3 + 4$
$= 10$

$A_2 = 2 \times A_1 + 4$
$= 2 \times 10 + 4$
$= 24$

Step 3: Substitute $A_2 = 24$ into the equation from step 1 to find A_3.

$A_3 = 2 \times 24 + 4$

Answer

D

16% of students answered A. Students needed to first calculate the value of A_2 to substitute into the recurrence relation. A common mistake was substituting the initial value, A_0, into the equation.

Questions from multiple lessons

21. C **22.** A

23. **a.** 13.52

b. *intercept* = −36.77 and *slope* = 13.52

5B Flat rate and unit cost depreciation – recurrence relations

Modelling flat rate and unit cost depreciation using recurrence relations

1. Courtney's yacht was worth $100 000 when brand new. The value of her yacht decreases by $1500 every time she uses it. This amount can be expressed as 1.5% of the principal.

2. **a.** $V_0 = 18\,000, \quad V_{n+1} = V_n - 2000$

b. $V_0 = 800, \quad V_{n+1} = V_n - 128$

c. $V_0 = 575, \quad V_{n+1} = V_n - 57.5$

3. **a.** $V_0 = 300, \quad V_{n+1} = V_n - 2$

b. $V_0 = 20\,000, \quad V_{n+1} = V_n - 0.05$

c. $V_0 = 500, \quad V_{n+1} = V_n - 0.01$

4. $V_0 = 14\,000, \quad V_{n+1} = V_n - 2000$

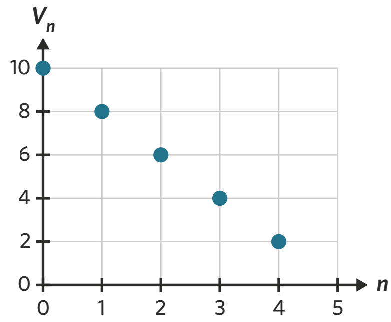
5. $V_0 = 1200, \quad V_{n+1} = V_n - 80$

Using an arithmetic recurrence relation to determine the value of an asset after *n* depreciation periods

6. B

7. **a.** $V_1 = 8, \quad V_2 = 6, \quad V_3 = 4, \quad V_4 = 2$

b.



8. **a.** $184 **b.** $128

c. 6 uses

9. **a.** $664.28 **b.** 70 000 times

Joining it all together

10. **a.** $V_0 = 600, \quad V_{n+1} = V_n - 33$

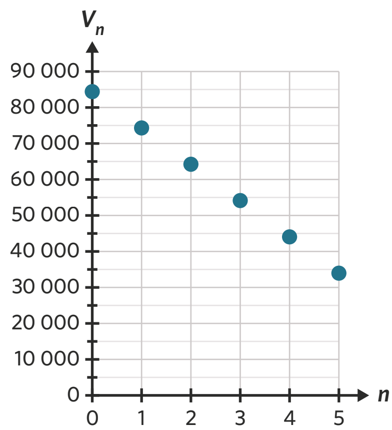
b. $435

11. **a.** $780 **b.** 3 times

c. 17.5%

12. **a.** $V_0 = 84\,000, \quad V_{n+1} = V_n - 10\,000$

b.



c. 3 years

13. **a.** $V_0 = 280, \quad V_{n+1} = V_n - 22$

b. $192 **c.** 1000 days

Exam practice

14. Explanation

Step 1: Determine the initial value, V_0.

The computer was purchased for $4500.

$V_0 = 4500$

Step 2: Calculate the depreciation amount, d.

$d = \frac{r}{100} \times V_0$

The rate of depreciation is 10%, so $r = 10$.

$d = \frac{10}{100} \times 4500$
$= 450$

Step 3: Substitute these values into the recurrence relation.

$V_0 = principal, \quad V_{n+1} = V_n - d$

Back to contents

Answer

A

13% of students answered C. These students did not recognise that flat rate depreciation is modelled using an arithmetic recurrence relation. Instead they misinterpreted it as a geometric recurrence relation, where the value of the computer would be multiplied by 0.9 each recursion.

15. Explanation

Step 1: Determine the initial value, V_0.

$V_0 = 120\,000$

Step 2: Calculate V_1, the value after one year.

$$V_1 = V_0 - 15\,000$$
$$= 120\,000 - 15\,000$$
$$= 105\,000$$

Step 3: Calculate V_2, the value after two years.

$$V_2 = V_1 - 15\,000$$
$$= 105\,000 - 15\,000$$
$$= 90\,000$$

Answer

$V_1 = 120\,000 - 15\,000 = 105\,000$

$V_2 = 105\,000 - 15\,000 = 90\,000$

The main problem that students found with this question was understanding what was required. Several students did not show recursive calculations, as the question specified. The final answer of 90 000 needed to be obtained by recursion and therefore workings were necessary.

16. Explanation

Step 1: Determine the initial value, V_0.

The initial value of the tools was $60 000.

$V_0 = 60\,000$

Step 2: Calculate the depreciation amount, d.

$$d = \frac{r}{100} \times V_0$$

The rate of depreciation is 8%, so $r = 8$.

$$d = \frac{8}{100} \times 60\,000$$
$$= 4800$$

Step 3: Substitute these values into the recurrence relation.

$V_0 = principal, \quad V_{n+1} = V_n - d$

Answer

$V_0 = 60\,000, \quad V_{n+1} = V_n - 4800$

Students needed to ensure that they wrote a recurrence relation for this question. Several students wrote a rule with V_n in terms of n which was not accepted. Additionally, a few students did not use a recurrence relation of the form $V_0 = principal, \quad V_{n+1} = V_n - d$.

17. Explanation

Step 1: Determine the initial value, V_0.

The caravan was purchased for $38 000.

$V_0 = 38\,000$

Step 2: Calculate the depreciation per year.

After 8 years, the value was $16 000.

$$\textit{depreciation over eight years} = 38\,000 - 16\,000$$
$$= 22\,000$$

$$\textit{depreciation per year} = \frac{22\,000}{8}$$
$$= 2750$$

Step 3: Calculate the depreciation per kilometre travelled.

On average, 5000 km were travelled each year.

$$\textit{depreciation per km travelled} = \frac{2750}{5000}$$
$$= 0.55$$

Answer

$0.55 per km

A common incorrect response was $4.40. This is likely because the students divided the depreciation over 8 years by the annual kilometres travelled (5000 km).

Questions from multiple lessons

18. E

19. D

20. a. Negatively skewed **b.** −14.5

5C Reducing balance depreciation – recurrence relations

Modelling reducing balance depreciation using recurrence relations

1. D

2. a. $2000 **b.** 6%

3. a. $O_0 = 600, \quad O_{n+1} = 0.92 \times O_n$

b. $A_0 = 500, \quad A_{n+1} = 0.86 \times A_n$

4. $V_0 = 36\,000, \quad V_{n+1} = 0.8 \times V_n$

Using a geometric recurrence relation to determine the value of an asset after n depreciation periods

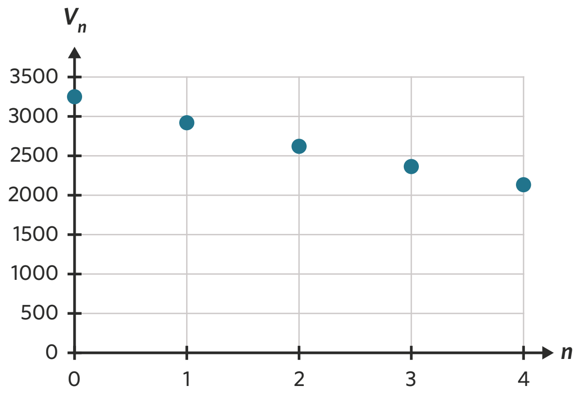
5. C

6. a. $8892.14 **b.** $2907.86

7. a. $23 981.29 **b.** 5 months **c.** $400

Back to contents

8. a. $V_1 = 2925, \quad V_2 = 2632.5, \quad V_3 = 2369.25, \quad V_4 = 2132.33$

b. 

c. 3 years

9. a. $30 876 b. $24 457

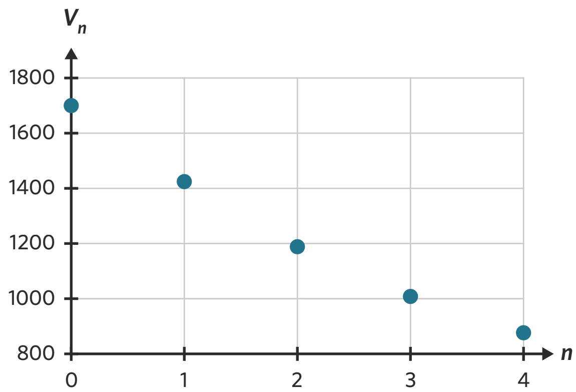
Joining it all together

10. a. 2% b. $47.05 c. 2.5%

d. $D_0 = 900, \quad D_{n+1} = 0.975 \times D_n$

11. a. $S_0 = 1700, \quad S_{n+1} = 0.84 \times S_n$

b. $846.38

c. 

d. 2 years

12. a. $C_0 = 3250, \quad C_{n+1} = 0.9 \times C_n$

b. 7 years c. $1403

Exam practice

13. **Explanation**

Step 1: Determine the initial value, V_0.

$V_0 = 60\,000$

Step 2: Calculate the value after 1 year.

$$V_1 = 0.9 \times V_0$$
$$= 0.9 \times 60\,000$$
$$= 54\,000$$

Step 3: Calculate the value after 2 years.

$$V_2 = 0.9 \times V_1$$
$$= 0.9 \times 54\,000$$
$$= 48\,600$$

Note: These calculations can also be performed using a calculator as shown in 5A Worked example 5.

Answer

$V_1 = 0.9 \times 60\,000 = 54\,000$

$V_2 = 0.9 \times 54\,000 = 48\,600$

Students needed to ensure that they included their workings as the question specified to show that the value after two years was $48 600.

14. **Explanation**

To solve this question, check whether each option is true or false.

A: The item does not depreciate by a constant amount each period. ✗

B: The item depreciates by a constant ratio each period and is therefore depreciated according to the reducing balance method. The value of the asset decreases by $420 in the first period, which is 6% of the initial value. The asset continues to depreciate by 6% each year. ✓

C: The graph shows a decreasing balance. An investment that earns interest increases in value. ✗

D: The item depreciates by a constant ratio each period and is therefore depreciated according to the reducing balance method. However, the reducing balance rate seen in the graph is 6%. ✗

E: The item does not depreciate by a constant amount each period. ✗

Answer

B

15% of students incorrectly chose option A. Students most likely calculated the depreciation in year 1 as 6%, but then incorrectly assumed that the item was depreciated at a constant amount each period instead of recognising that it depreciated on a reducing balance basis.

15. a. **Explanation**

Step 1: Recall the formula that relates the common ratio and the rate of depreciation.

$$R = 1 - \frac{r}{100}$$

Step 2: Substitute $R = 0.85$ and solve for the value of r.

$$0.85 = 1 - \frac{r}{100}$$
$$\frac{r}{100} = 0.15$$
$$r = 15$$

Answer

15%

A number of students gave the R value of 0.85 due to incorrectly interpreting the value of R in the recurrence relation as the rate of depreciation. Students needed to recognise that R represents the common ratio for the recurrence relation.

b. **Explanation**

Step 1: Using the recurrence relation and value of R, calculate the value after three years.

Julie's car is depreciated at a rate of 15% for the first three years.

$C_0 = 14\,000$

$C_1 = 0.85 \times C_0 = 0.85 \times 14\,000 = 11\,900$

$C_2 = 0.85 \times C_1 = 0.85 \times 11\,900 = 10\,115$

$C_3 = 0.85 \times C_2 = 0.85 \times 10\,115 = 8597.75$

Back to contents

Step 2: Calculate the value after eight years.

The new rate of depreciation is 8.6% so $R = 0.914$.

$C_4 = 0.914 \times C_3 = 0.914 \times 8597.75 = 7858.34...$

$C_5 = 0.914 \times C_4 = 0.914 \times 7858.34... = 7182.53...$

$C_6 = 0.914 \times C_5 = 0.914 \times 7182.53... = 6564.83...$

$C_7 = 0.914 \times C_6 = 0.914 \times 6564.83... = 6000.25...$

$C_8 = 0.914 \times C_7 = 0.914 \times 6000.25... = 5484.23...$

Note: These calculations can also be performed using a calculator as shown in 5A Worked example 5.

Answer

$5484.23

Many students incorrectly depreciated the car by 8.6% for the entire eight years instead of depreciating by 15% for the first three. Some students obtained one mark for correctly finding the value after three years.

Questions from multiple lessons

16. C **17.** B

18. a. Positively skewed with two possible outliers

b. 90% **c.** 24 handbags

5D Depreciation – finding the rule for the *n*th term

Creating and using a rule for flat rate depreciation

1. C **2.** A

3. a. $V_n = 1\,000\,000 - 60\,000n$
$V_9 = \$460\,000$

b. $V_n = 550 - 16.5n$
$V_9 = \$401.50$

c. $V_n = 380 - 24.7n$
$V_9 = \$157.70$

d. $V_n = 1545 - 61.8n$
$V_9 = \$988.80$

4. a. $S_n = 550 - 55n$ **b.** $385

c. 10 years

5. a. 11% **b.** $C_n = 1225 - 134.75n$

c. $281.75

6. 10%

Creating and using a rule for unit cost depreciation

7. C

8. a. $V_n = 65 - 0.08n$
$V_5 = \$64.60$

b. $V_n = 315 - 0.05n$
$V_5 = \$314.75$

c. $V_n = 20\,000 - 245n$
$V_5 = \$18\,775$

d. $V_n = 825 - 7.5n$
$V_5 = \$787.50$

9. a. $P_n = 400 - 8n$ **b.** $240

c. 25 jumps

10. a. $3.40 **b.** $5 718 200

11. $7418.75

Creating and using a rule for reducing balance depreciation

12. D

13. a. $V_n = 13\,000 \times 0.94^n$
$V_3 = \$10\,797.59$

b. $V_n = 3100 \times 0.9^n$
$V_7 = \$1482.72$

c. $V_n = 2000 \times 0.965^n$
$V_8 = \$1504.00$

d. 6500×0.98^n
$V_{26} = \$3844.07$

14. One year and nine months

Joining it all together

15. a. Flat rate: $H_n = 170 - 4.76n$
Unit cost: $H_n = 170 - 0.22n$
Reducing balance: $H_n = 170 \times 0.97^n$

b. Flat rate
$160.48

c. Reducing balance
$141.61

16. a. 3.75% **b.** 3.82%

c. Flat rate: $750
Reducing balance: $812.88
The reducing balance method will allow Mildred to sell the washer for the highest amount.

Exam practice

17. Explanation

Step 1: Determine the value of the van after three years.

In the fourth year, the van depreciated by 15 cents per kilometre and travelled 30 000 km.

This means it depreciated by $0.15 \times 30\,000 = \$4500$ in the fourth year.

$V_3 = 26\,166.24 + 4500$
$= 30\,666.24$

Step 2: Determine the rule that models V_n, the value of the van after n years.

$V_n = V_0 \times \left(1 - \frac{r}{100}\right)^n$

Substitute $V_n = 30\,666.24$, $V_0 = 45\,000$, $n = 3$ and $r = k$ into the rule.

$30\,666.24 = 45\,000 \times \left(1 - \frac{k}{100}\right)^3$

Step 3: Solve for k.

Answer

B

18. Explanation

Step 1: Interpret the question.

The question asks for the value that the boat will depreciate by 10% of in the third year.

In the third year, the boat will depreciate by 10% of the value in the second year.

Step 2: Determine the rule that models the value of the boat after n years.

$V_n = V_0 \times \left(1 - \frac{r}{100}\right)^n$

Substitute $V_0 = 72\,000$ and $r = 10$ into the rule.

$V_n = 72\,000 \times 0.9^n$

Step 3: Calculate V_2, the value of the boat after 2 years.

Substitute $n = 2$ into the rule for V_n.

$V_2 = 72\,000 \times 0.9^2$

$= 58\,320$

Answer

C

40% of students incorrectly selected option B. These students calculated the value of the boat at the end of the third year, rather than at the end of the second year.

19. a. Explanation

Step 1: Determine the initial value, V_0.

The coffee machine was purchased for \$12 000.

$V_0 = 12\,000$

Step 2: Determine the depreciation amount, d.

From the recurrence relation, $d = 1440$.

Step 3: Calculate the flat rate depreciation rate, r.

$d = \frac{r}{100} \times V_0$

Substitute $d = 1440$ and $V_0 = 12\,000$ into the rule.

$1440 = \frac{r}{100} \times 12\,000$

$r = \frac{1400 \times 100}{12\,000}$

$r = 12$

Answer

12%

51% of students answered this question incorrectly. Students likely did not understand that the recurrence relation showed the annual depreciation, and instead tried to calculate the flat rate percentage using the unit cost depreciation of \$0.05 per cup of coffee.

b. Explanation

Step 1: Determine the initial value, V_0.

$V_0 = 12\,000$

Step 2: Determine the depreciation per unit of use, d.

The coffee machine depreciates by \$0.05 per cup of coffee made.

$d = -0.05$

Note: The question provides a '+' sign between the two values. As the coffee machine is depreciating, d needs to be negative.

Answer

$M_n = 12\,000 + -0.05 \times n$

Many students failed to recognise that the value was in terms of cups of coffee made, and instead provided a percentage answer based on the previous question. Many students also left d as a positive value.

Questions from multiple lessons

20. E

21. D

22. a. \$243.00 **b.** May

c. \$1576.25

5E Simple interest

Modelling simple interest using recurrence relations

1. D

2. a. $T_0 = 1500, \quad T_{n+1} = T_n + 30$

b. $A_0 = 500, \quad A_{n+1} = A_n + 40$

3. a. \$775.00 **b.** \$1500.00

c. \$1035.00

4. a. \$1600.00 **b.** 2.69%

Modelling simple interest using a rule

5. D

6. D

7. \$344.00

8. a. \$1200.00 **b.** \$50.00

c. 4.17%

9. a. \$112.50 **b.** \$3675.00

10. a. Bank of Edrolo **b.** \$459.00

11. a. $J_n = 400 + 12n$ **b.** \$472.00

c. 9 months

12. a. \$500.00 **b.** \$650.00

Joining it all together

13. a. 2.27% **b.** $T_n = 3650 + 83n$

c. \$4148.00

Back to contents

14. a. \$550.00
 b. $C_0 = 550, \quad C_{n+1} = C_n + 22$
 c. 7 years

15. a. \$3600.00 b. $H_n = 3600 + 90n$

Exam practice

16. **Explanation**

Step 1: Identify the principal and interest rate.

$V_0 = 3000$

$r = 6.5\%$ per annum

Step 2: Calculate the interest, d.

$$d = \frac{r}{100} \times V_0$$
$$= \frac{6.5}{100} \times 3000$$
$$= 195$$

This means that \$195 is earned in interest each year.

Step 3: Calculate the interest earned in three years.

$$\text{interest} = d \times 3$$
$$= 195 \times 3$$
$$= 585$$

Answer

C

17. **Explanation**

Step 1: Calculate the yearly interest amount.

The interest paid over 6 years is \$27 000.

$$d = \frac{27\,000}{6}$$
$$= 4500$$

This means that \$4500 is earned in interest each year.

Step 2: Determine the interest rate.

$r = 8\%$ per annum

Step 3: Calculate the principal using $d = \frac{r}{100} \times V_0$.

$$4500 = \frac{8}{100} \times V_0$$
$$= 0.08 \times V_0$$
$$V_0 = \frac{4500}{0.08}$$
$$= 56\,250$$

Answer

C

17% of students incorrectly answered A. These students calculated the amount of interest earned after 6 years with \$27 000 as the principal, instead of calculating the principal given \$27 000 as the total interest earned.

Questions from multiple lessons

18. C 19. B

20. a. 64.9%
 b. 1.3
 c. 62%

5F Compound interest

Modelling compound interest using recurrence relations

1. C

2. a. \$3000 b. 3%

3. a. \$1823.26
 b. After 6 months

4. a. $L_0 = 2800, \quad L_{n+1} = 1.02 \times L_n$
 b. $S_0 = 800, \quad S_{n+1} = 1.04 \times S_n$

5. a. $H_0 = 29\,000, \quad H_{n+1} = 1.008 \times H_n$
 b. 5 months

6. \$5248.80

7. a. \$112.13
 b. $B_0 = 1100, \quad B_{n+1} = 1.02 \times B_n$

Modelling compound interest using a rule

8. D

9. a. \$138.92
 b. \$13 346 851.03
 c. \$363 152.91

10. a. \$3423.50 b. \$3608.12 c. \$3372.21

11. a. \$42 185 b. \$1353 c. \$694

12. Eliza will have \$3.90 more.

Joining it all together

13. B 14. D

15. a. $V_0 = 1000, \quad V_{n+1} = 1.018 \times V_n$
 $V_n = 1000 \times 1.018^n$
 b. $V_0 = 3500, \quad V_{n+1} = 1.002 \times V_n$
 $V_n = 3500 \times 1.002^n$
 c. $V_0 = 5000, \quad V_{n+1} = 1.015 \times V_n$
 $V_n = 5000 \times 1.015^n$

16. Account A: $V_3 = 5000 \times \left(1 + \frac{6.8}{100}\right)^3$
$= 6090.93...$

Account B: $V_{36} = 5000 \times \left(1 + \frac{6.6}{100 \times 12}\right)^{36}$
$= 6091.50...$

Account B is more profitable.

Back to contents

Exam practice

17. Explanation

Step 1: Calculate B_1, the value of the investment after 1 month.

$B_1 = 1.0048 \times B_0$
$= 1.0048 \times 3000$
$= 3014.40$

Step 2: Calculate B_2, the value of the investment after 2 months.

$B_2 = 1.0048 \times B_1$
$= 1.0048 \times 3014.40$
$= 3028.869...$

Step 3: Calculate B_3, the value of the investment after 3 months.

$B_3 = 1.0048 \times B_2$
$= 1.0048 \times 3028.869...$
$= 3043.407...$

Step 4: Calculate B_4, the value of the investment after 4 months.

$B_4 = 1.0048 \times B_3$
$= 1.0048 \times 3043.407...$
$= 3058.016...$

Step 5: Calculate B_5, the value of the investment after 5 months.

$B_5 = 1.0048 \times B_4$
$= 1.0048 \times 3058.016...$
$= 3072.694...$

Step 6: Calculate the interest earned after 5 months.

This is the difference between B_5 and B_0.

$interest = 3072.694... - 3000$
$= 72.694...$

Answer

D

18. a. Explanation

Step 1: Identify V_0 and R.

$V_0 = 200$

$R = 1 + \frac{r}{100}$
$= 1 + \frac{1.5}{100}$
$= 1.015$

Step 2: Construct a rule, and use it to find V_1.

$V_n = 200 \times 1.015^n$

$V_1 = 200 \times 1.015^1$

Answer

$203

Some students misread the 1.5% interest rate as being per annum, leading to a calculated monthly interest rate of 0.125%, and an incorrect answer of $200.25.

b. Explanation

Step 1: Identify the principal and interest rate.

$A_0 = 428$

The interest rate is 1.5% per month, compounding monthly.

$r = 1.5\%$ per compounding period

Step 2: Calculate R.

$R = 1 + \frac{r}{100}$
$= 1 + \frac{1.5}{100}$
$= 1.015$

Step 3: Construct the recurrence relation.

Answer

$A_0 = 428, \quad A_{n+1} = 1.015 \times A_n$

A number of students did not write a recurrence relation in the correct form. The initial value needed to be written first, and the recurrence relation needed to use the appropriate terms A_0, A_n and A_{n+1}. Some students wrote a rule instead of a recurrence relation, which was not accepted.

c. Explanation

Step 1: Calculate A_1, the value of the account after 1 quarter.

$A_1 = 1.015 \times A_0$
$= 1.015 \times 428$
$= 434.42$

Step 2: Calculate A_2, the value of the account after 2 quarters.

$A_2 = 1.015 \times A_1$
$= 1.015 \times 434.42$
$= 440.936..$

Step 3: Calculate A_3, the value of the account after 3 quarters.

$A_3 = 1.015 \times A_2$
$= 1.015 \times 440.936..$
$= 447.550...$

Step 4: Calculate A_4, the value of the account after 4 quarters.

$A_4 = 1.015 \times A_3$
$= 1.015 \times 447.550...$
$= 454.263...$

Step 5: Calculate the interest charged after 4 quarters.

This is the difference between V_4 and V_0.

$interest = 454.263... - 428$
$= 26.263...$

Answer

$26.26

From the students who answered parts **a** and **b** correctly, some gave the total amount paid, $454.26, as the answer, which was not what the question asked for.

Questions from multiple lessons

19. C

20. B

21. a. 2

b. Arithmetic

c. $b = 2, \quad c = 3$

5F ANSWERS

Back to contents

5G Nominal and effective interest rates

Calculating effective interest rates

1. C
2. a. 10.47% b. 8.32% c. 9.64%
   d. 7.32% e. 2.03%
3. 12.4%

Comparing nominal and effective interest rates and returns

4. D
5. a. 7.9% per annum, compounding monthly
   b. Effective interest rate of 11.3% per annum
   c. 8.7% per annum, compounding weekly
6. SuperAccount
7. Westbank

Joining it all together

8. C
9. AussieBank
10. No, because her current bank offers a higher effective interest rate.

Exam practice

11. **Explanation**

To solve this question, calculate the effective interest of each multiple choice option, and check whether each one corresponds to the required rate.

A: $r = 8\%$ p.a. and $n = 365$ compounding periods per year

$r_{effective} = 8.3277...\%$

$\approx 8.33\%$ ✗

B: $r = 8\%$ p.a. and $n = 52$ compounding periods per year

$r_{effective} = 8.3220...\%$

$\approx 8.32\%$ ✗

C: $r = 8\%$ p.a. and $n = 26$ compounding periods per year

$r_{effective} = 8.3220...\%$

$\approx 8.32\% \neq 8.31\%$ ✓

D: $r = 8\%$ p.a. and $n = 12$ compounding periods per year

$r_{effective} = 8.2999...\%$

$\approx 8.30\%$ ✗

E: $r = 8\%$ p.a. and $n = 4$ compounding periods per year

$r_{effective} = 8.2432...\%$

$\approx 8.24\%$ ✗

Answer

C

12. **Explanation**

Step 1: Determine the loan options that have the lowest nominal rate for their respective compounding frequency.

Loans I, II and III all compound weekly. As loan I has the lowest nominal interest rate, it will also have the lowest effective interest rate of all weekly-compunding loans.

Loans IV and V both compound quarterly. As loan IV has the lowest nominal interest rate, it will also have the lowest effective interest rate of all quarterly-compounding loans.

The cheapest option will either be loan I or IV.

Step 2: Calculate the effective interest rate for options I and IV.

Loan I: $r = 12.6\%$ p.a. and $n = 52$ compounding periods per year

$r_{effective} = 13.4109...\%$

Loan IV: $r = 12.7\%$ p.a. and $n = 4$ compounding periods per year

$r_{effective} = 13.3177...\%$

Step 3: Determine the best option.

The question involves a loan so the option with the lowest effective interest rate will be the preferred choice.

Loan IV has the lowest effective interest rate.

Answer

D

28% of students incorrectly chose A. This is likely because they selected the option with the lowest interest rate without considering the number of compounding periods.

Questions from multiple lessons

13. E
14. A
15. a. \$300
    b. $V_0 = 4000, \quad V_{n+1} = V_n - 300$
    c. \$0.15 per page

Back to contents

6A Introducing financial applications

Using an amortisation table to model financial problems

1. a. A b. B

2. Hannes takes out money from the safe every week and doesn't add back to it.

   $balance = previous\ balance - principal\ reduction$

   $= 500.00 - 50.00 = 450.00$

   Answers may vary.

3. a. $principal\ reduction = payment - interest$

   $= 50.00 - 15.00 = 35.00$

   b. $balance\ of\ loan = previous\ loan\ balance - principal\ reduction$

   $= 500.00 - 35.00 = 465.00$

   c.

payment number	payment	interest	principal reduction	balance of loan
0	0.00	0.00	0.00	500.00
1	50.00	15.00	35.00	465.00
2	50.00	15.00	35.00	430.00
3	50.00	15.00	35.00	395.00
4	50.00	15.00	35.00	360.00

Using financial applications of technology

4. C

5. a. 1.5 b. −215 c. 0

6. a. 2 b. −4000 c. 365

Joining it all together

7. a.

payment number	payment	interest	principal reduction	balance of loan
0	0.00	0.00	0.00	500.00
1	15.00	2.08	12.92	487.08
2	15.00	2.03	12.97	474.11
3	15.00	1.98	13.02	461.09
4	15.00	1.92	13.08	448.01

   b. i. The establishment of a loan gives the borrower a lump sum of money. This amount is going towards them, hence the positive sign.
   Answers may vary.

   ii. The payments made by the borrower go towards paying off their debt. This amount is going away from them, hence the negative sign.
   Answers may vary.

8. a. $344.28

   b. i. Negative, because the investor puts away a sum of money, so it's going away from them.
   Answers may vary.

   ii. Positive, because the investor receives payments, the money is going towards them.
   Answers may vary.

Questions from multiple lessons

9. C 10. C

11. a. Increasing

    b. $number\ of\ ice\ creams\ sold = 21.38 + 1.185 \times month\ number$

6B Reducing balance loans

Using recurrence relations to model reducing balance loans

1. A

2. a. $10 000 b. $600 c. 4.8% p.a.

3. a. $V_0 = 8500, \quad V_{n+1} = 1.0018 \times V_n - 250$

   b. $V_0 = 985\,000, \quad V_{n+1} = 1.0006 \times V_n - 1200$

4. a. 6.72% p.a. b. $22 137.69 c. 5 months

5. 3 fortnights

Using amortisation tables to solve problems involving reducing balance loans

6. a. D b. C c. D

7.

payment number	payment	interest	principal reduction	balance of loan
0	0.00	0.00	0.00	18 000.00
1	1627.00	90.00	1537.00	16 463.00

8.

payment number	payment	interest	principal reduction	balance of loan
0	0.00	0.00	0.00	1200.00
1	242.42	4.02	238.40	961.60
2	242.42	3.22	239.20	722.40
3	242.42	2.42	240.00	482.40
4	242.42	1.62	240.80	241.60
5	242.41	0.81	241.60	0.00

9. 2.4% p.a.

Back to contents

10.

payment number	payment	interest	principal reduction	balance of loan
0	0.00	0.00	0.00	3000.00
1	500.00	10.00	490.00	2510.00
2	500.00	8.37	491.63	2018.37
3	500.00	6.73	493.27	1525.10

Using financial applications of technology to solve problems involving reducing balance loans

11. B

12. **a.** 6 quarters **b.** 40% p.a. **c.** $42 000
d. $141.73

13. 5 weeks

14. **a.** $15 984.69 **b.** 95 months **c.** $28 563.48
d. $9563.48

Joining it all together

15. **a.** 7.2% p.a.
b. $V_0 = 5000, \quad V_{n+1} = 1.006 \times V_n - 1200$
c. $279.38

16. $R = 1.0195$

17. **a.** $162.61 **b.** 42 weeks **c.** $231.40

Exam practice

18. **Explanation**

Use recursion to calculate V_1, V_2, V_3, V_4 and V_5.

$V_0 = 26\,000$
$V_1 = 25\,678.00$
$V_2 = 25\,355.034...$
$V_3 = 25\,031.099...$
$V_4 = 24\,706.192...$
$V_5 = 24\,380.310...$

Answer

A

19. **Explanation**

Step 1: Calculate the interest for payment number 2.

$$\begin{aligned} \textit{interest} &= \frac{r}{100} \times \textit{previous loan balance} \\ &= \frac{4.8}{12 \times 100} \times 249\,500.00 \\ &= \frac{4.8}{1200} \times 249\,500.00 \\ &= 998.00 \end{aligned}$$

Step 2: Calculate the principal reduction for payment number 2.

$$\begin{aligned} \textit{principal reduction} &= \textit{payment} - \textit{interest} \\ &= 1500.00 - 998.00 \\ &= 502.00 \end{aligned}$$

Answer

B

19% of students incorrectly answered option C. This is likely because they understood that the principal reduction increases with each payment, but didn't know how to calculate it.

20. **Explanation**

Step 1: Determine the annual interest rate.

N	240	(there are 240 months in 20 years)
I(%)		
PV	450 000	(this is positive because Bimal receives it from the lender)
PMT	−2633	(this is negative since Bimal pays the lender)
FV	0	(the loan is to be fully repaid)
PpY	12	(payments made monthly)
CpY	12	(interest compounds monthly)

I(%)	3.599...

The annual interest rate is 3.599...% p.a.

Step 2: Calculate R.

The interest compounds monthly, so $r = \frac{3.599...}{12}$% per compounding period.

$$\begin{aligned} R &= 1 + \frac{3.599...}{12 \times 100} \\ &= 1 + \frac{3.599...}{1200} \\ &= 1.00299... \\ &\approx 1.003 \end{aligned}$$

Answer

A

33% of students incorrectly answered option D. This is likely because they did not factor in the compounding period before calculating R.

21. **Explanation**

Step 1: Identify R from the recurrence relation.

$R = 1.001$

Step 2: Calculate r, the interest rate per compounding period.

$$\begin{aligned} R &= 1 + \frac{r}{100} \\ 1.001 &= 1 + \frac{r}{100} \\ r &= 0.001 \times 100 \\ &= 0.1\% \text{ per compounding period} \end{aligned}$$

Step 3: Calculate the annual interest rate.

Interest compounds on a fortnightly basis.

$$\begin{aligned} \textit{annual interest rate} &= 0.1\% \times 26 \\ &= 2.6\% \text{ p.a.} \end{aligned}$$

Back to contents

Answer

$1.001 = 1 + \frac{r}{100}$

$r = 0.001 \times 100 = 0.1\%$ per compounding period

annual interest rate $= 0.1\% \times 26 = 2.6\%$ p.a.

A significant number of students incorrectly responded with calculator commands. Any response involving calculator syntax or notation such as writing an equation involving *r* and 'solve' does not warrant full marks.

22. Explanation

Step 1: Determine the balance remaining after the initial three years.

N	36	(there are 36 months in 3 years)
I(%)	6.9	(annual interest rate)
PV	70 000	(this is positive because Ken receives it from the lender)
PMT	−800	(this is negative since Ken pays the lender)
FV		
PpY	12	(payments made monthly)
CpY	12	(interest compounds monthly)

FV	−54 151.599...

This means that after 3 years of repayments, Ken will still owe $54 151.60.

At this point in time, Ken will make a lump sum payment, $L to reduce the balance of the loan.

Step 2: Determine the required balance at the start of the next three years.

The next three years will see that the loan is paid off in full. The balance at the start of this period is unknown.

N	36	(there are 36 months in 3 years)
I(%)	6.9	(annual interest rate)
PV		
PMT	−800	(this is negative since Ken pays the lender)
FV	0	(the loan is to be fully repaid)
PpY	12	(payments made monthly)
CpY	12	(interest compounds monthly)

PV	25 947.576...

This means that Ken's loan balance needs to be $25 947.58 for him to fully pay off the loan in the remaining 3 years.

The difference between the balance at the end of the first three years and the required beginning balance of the next three years will be equal to the lump sum payment $L that Ken makes.

Step 3: Calculate the difference between the two balances.

$54\,151.60 - 25\,947.58 = 28\,204.02$

$\approx 28\,204$

Answer

$28 204

A significant number of students only calculated the loan balance after three years.

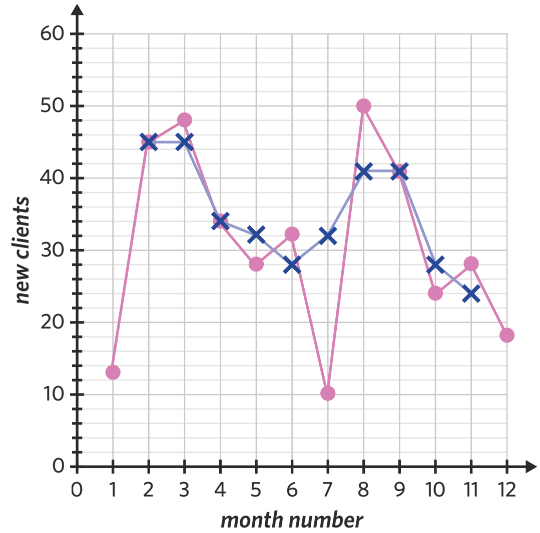
Questions from multiple lessons

23. D **24.** B

25. a. 40 clients

b. 28 clients

c.



6C Interest-only loans

Using recurrence relations to model interest-only loans

1. B

2. a. $8000 **b.** $240 **c.** 36% p.a.

d. $V_0 = 8000, \quad V_{n+1} = 1.03 \times V_n - 240$

3. a. $336.00 **b.** $3080.00 **c.** $75.60

d. $4.85

4. a. $23 250

b. $V_0 = 23\,250, \quad V_{n+1} = 1.02 \times V_n - 465$

5. a. 18.0% p.a. **b.** 2.2% p.a. **c.** 42.9% p.a.

6.

payment number	payment	interest	principal reduction	balance of loan
11	3125.00	3125.00	0.00	750 000.00
12	3125.00	3125.00	0.00	750 000.00
13	3125.00	3125.00	0.00	750 000.00
14	3125.00	3125.00	0.00	750 000.00

Back to contents

Using financial applications of technology to solve problems involving interest-only loans

7. D

8. **a.** 10.4% p.a. **b.** 52.0% p.a. **c.** 15.9% p.a.

9. **a.** \$912.00 **b.** \$730.77 **c.** \$2429.17

10. **a.** \$383 488.70 **b.** 3.0% p.a. **c.** \$34 513.44

Joining it all together

11. **a.** \$138 000 **b.** 7.5% p.a.

c. $V_0 = 500\,000, \quad V_{n+1} = 1.00625 \times V_n - 3125$

d. Melbourne Bank. They offer a lower interest rate of 6.9% p.a., while Sydney Bank offers 7.5% p.a.

12. **a.** \$1 000 000

b. $V_0 = 900\,000, \quad V_{n+1} = 1.0002 \times V_n - 180$

c. \$197 100 **d.** \$900 000 **e.** 3.65% p.a.

Exam practice

13. **Explanation**

Step 1: Calculate the scheduled monthly payment.

This represents the original schedule that was planned, prior to Bob deciding to make interest-only payments.

N	240	(there are 240 months in 20 years)
I(%)	3.14%	(annual interest rate)
PV	400 000	(this is positive because Bob receives it from the lender)
PMT		
FV	0	(the loan is to be fully repaid)
PpY	12	(payments made monthly)
CpY	12	(interest compounds monthly)

PMT	−2246.528...

The PMT is negative because Bob pays it to the lender.

Step 2: Calculate the new interest rate.

The payment remains the same as the scheduled monthly payment.

The PV of the loan remains constant because only the interest was paid over the first two years.

N	216	(there are 216 months in 18 years)
I(%)		
PV	400 000	(this is positive because Bob receives it from the lender)
PMT	−2246.528...	(this is negative because Bob pays it to the lender)
FV	0	(the loan is to be fully repaid)
PpY	12	(payments made monthly)
CpY	12	(interest compounds monthly)

I(%)	2.211...

Answer

B

25% of students incorrectly chose C. It is likely that these students did not realise that the balance of the loan remained the same in the first two years of the loan. In addition to this, it is possible that a number of students did not recognise that they were required to calculate the planned scheduled monthly repayment amount before using it to model the remaining 18 years of the loan.

Questions from multiple lessons

14. D **15.** C

16. **a.** $4324 = \left(1 - \frac{r}{100}\right) \times 4600$

$r = \left(1 - \frac{4324}{4600}\right) \times 100$

$r = 6\%$ p.a.

b. $V_0 = 4600, \quad V_{n+1} = 0.94 \times V_n$

6D Amortising annuities

Using recurrence relations to model amortising annuities

1. D

2. **a.** \$25 000 **b.** \$1500

c. 14.4%

3. **a.** $V_0 = 50\,000.00$

$V_1 = 1.150 \times 50\,000.00 - 10\,000 = 47\,500.00$

$V_2 = 1.150 \times 47\,500.00 - 10\,000 = 44\,625.00$

$V_3 = 1.150 \times 44\,625.00 - 10\,000 = 41\,318.75$

$V_4 = 1.150 \times 41\,318.75 - 10\,000 \approx 37\,516.56$

b. \$44 791.25

4. **a.** $V_0 = 2500, \quad V_{n+1} = 1.0058 \times V_n - 100$

b. $V_0 = 6200, \quad V_{n+1} = 1.0103 \times V_n - 250$

c. $V_0 = 1500, \quad V_{n+1} = 1.0019 \times V_n - 60$

Back to contents

5. $1402.08

Using amortisation tables to solve problems involving amortising annuities

6. A

7. a. $r = \frac{476.95}{136\,270.19} \times 100 = 0.350...\%$ per compounding period
   annual interest rate $= 0.350...\% \times 12 \approx 4.2\%$ p.a.
   Note: Answers may vary.
   b. $455.51   c. $1544.49
   d. $128 601.43

8. a. $x = 0.00, \quad y = 40.00, \quad z = 379.00$
   b. $x = 86.00, \quad y = 1514.00, \quad z = 60.56$
   c. $x = 74.88, \quad y = 1223.56, \quad z = 1194.88$

Using financial applications of technology to solve problems involving amortising annuities

9. B

10. a. $8962.24   b. $356.29
    c. $4870.66   d. $41 899.33

11. $149 019.29

12. a. $23 144.64   b. $431 856.49
    c. 378 fortnights

Joining it all together

13. a. $1 250 000   b. 10.4% p.a.
    c.

payment number	payment	interest	principal reduction	balance of investment
10	8500.00	4871.97	3628.03	1 214 363.24
11	8500.00	4857.45	3642.55	1 210 720.69

    d. $54 220.69

14. a. $2000
    b. $V_0 = 759\,231, \quad V_{n+1} = 1.0017 \times V_n - 2000$
    c. 465 weeks

15. 1.0003

16. a. $V_0 = 100\,000, \quad V_{n+1} = 1.0018 \times V_n - 1250$
    b. $10 483.34
    c. 3 years and 12 fortnights

Exam practice

17. **Explanation**
    Use recursion to calculate V_1, V_2, V_3, V_4 and V_5.
    $V_0 = 100\,000$
    $V_1 = 99\,695.00$
    $V_2 = 99\,389.237...$
    $V_3 = 99\,082.710...$
    $V_4 = 98\,775.417...$
    $V_5 = 98\,467.355...$

    Answer
    C

18. **Explanation**
    Step 1: Determine the financial solver inputs.

N	12	(there are 12 months in 1 year)
I(%)	4.32	(annual interest rate)
PV	−492 800	(this is negative because Julie invests it with the bank)
PMT		
FV	480 242.25	(this is positive because it represents future payments that will be made to Julie)
PpY	12	(payments made monthly)
CpY	12	(interest compounds monthly)

    Step 2: Use the financial solver to solve for PMT.

PMT	2800.00...

    This is positive because Julie receives it from the bank.

    Answer
    $2800

19. **Explanation**
    Step 1: Identify the principal and payment amount.
    principal = $500 000.00
    payment = $44 970.55
    Step 2: Calculate the annual interest rate using the amortisation table.
    This is calculated using the interest and previous investment balance. Any line in the amortisation table can be used.

payment number	payment	interest	principal reduction	balance of investment
0	0.00	0.00	0.00	500 000.00
1	44 970.55	20 000.00	24 970.55	475 029.45
2	44 970.55	19 001.18	25 969.37	449 060.08

    $\textit{interest} = \frac{r}{100} \times \textit{previous investment balance}$
    $20\,000.00 = \frac{r}{100} \times 500\,000.00$
    Solve for r, the interest rate per compounding period.
    $r = \frac{20\,000.00}{500\,000.00} \times 100$
    $= 4\%$ per compounding period
    The interest compounds annually, so this is the annual interest rate.
    annual interest rate = 4% p.a.

Step 3: Determine the financial solver inputs.

N		
I(%)	4	(annual interest rate)
PV	−500 000	(this is negative because Deepa invests it with the bank)
PMT	44 970.55	(this is positive because Deepa receives it from the bank)
FV	0	(annuity to be fully paid out)
PpY	1	(payments made annually)
CpY	1	(interest compounds annually)

Step 4: Use the financial solver to solve for N.

N	15.000...

This means that there will be 15 payments.

Since payments are made annually, this means that Deepa will receive payments of $44 970.55 for 15 consecutive years.

Answer

B

20. a. Explanation

Step 1: Determine the principal, interest rate and payment amount.

$V_0 = 152\,431$

$r = \frac{5.1}{12}\%$ per compounding period

$d = 900$

Step 2: Calculate R.

$$R = 1 + \frac{5.1}{12 \times 100}$$
$$= 1 + \frac{5.1}{1200}$$
$$= 1.00425$$

Step 3: Construct the recurrence relation.

$V_0 = 152\,431, \quad V_{n+1} = 1.00425 \times V_n - 900$

Step 4: Calculate V_1, the value of the annuity after 1 month.

$$V_1 = 1.00425 \times V_0 - 900$$
$$= 1.00425 \times 152\,431 - 900$$
$$= 152\,178.831...$$

Step 5: Calculate V_2, the value of the annuity after 2 months.

$$V_2 = 1.00425 \times V_1 - 900$$
$$= 1.00425 \times 152\,178.831... - 900$$
$$\approx 151\,925.591...$$

Answer

$V_0 = 152\,431$

$V_1 = 1.00425 \times 152\,431 - 900 = 152\,178.831...$

$V_2 = 1.00425 \times 152\,178.831... - 900 \approx 151\,925.59$

A number of students did not construct the recurrence relation correctly. Some had added the $900 rather than subtracted it and others did not show the complete recursive calculations. All three lines of working are required to obtain full marks.

b. Explanation

Step 1: Determine the balance remaining after the initial 2 years.

N	24	(there are 24 months in 2 years)
I(%)	5.1	(annual interest rate)
PV	−152 431	(this is negative because Sienna invests it with the bank)
PMT	900	(this is positive because Sienna receives it from the bank)
FV		
PpY	12	(payments made monthly)
CpY	12	(interest compounds monthly)

FV	146 073.740...

This means that after 2 years, the annuity still has $146 073.74 left.

At this point in time, Sienna will make a lump sum payment to increase the balance of the annuity investment.

Step 2: Determine the required starting balance for the remaining 23 years.

The remaining 23 years will see the annuity paid out in full at an interest rate of 4.6% per annum. The balance at the start of this period is unknown.

N	276	(there are 276 months in 23 years)
I(%)	4.6	(annual interest rate)
PV		
PMT	900	(this is positive because Sienna receives it from the bank)
FV	0	(annuity to be fully paid out)
PpY	12	(payments made monthly)
CpY	12	(interest compounds monthly)

Back to contents

PV	−153 112.939...

This means that the annuity balance must be $153 122.94 if the annuity is to be fully paid out in the remaining 23 years.

The difference between the balance at the end of the first 2 years and the beginning balance of the remaining 23 years will be equal to the lump sum payment.

Step 3: Calculate the difference between the two balances.

$153\,112.94 - 146\,073.74 = 7039.20$

Answer

$7039.20

A significant number of students likely only calculated the future value after two years.

Questions from multiple lessons

21. B **22.** C

23. a. 25% **b.** $7360.94

6E Perpetuities

Using recurrence relations to model perpetuities

1. D

2. a. $90 000 **b.** $270 **c.** 3.6% p.a.

d. $V_0 = 90\,000, \quad V_{n+1} = 1.003 \times V_n - 270$

3. a. $842.15 **b.** 6% p.a. **c.** $168 429

4. a. $233.33 **b.** $1750 **c.** $79 411.76

d. 7.17% p.a.

5. a. $1 400 002.50

b. $V_0 = 1\,400\,002.50, \quad V_{n+1} = 1.0015 \times V_n - 2153.85$

6. a. $754 270

b.

payment number	payment	interest	principal reduction	balance of investment
20	4525.62	4525.62	0.00	754 270.00
21	4525.62	4525.62	0.00	754 270.00

Using financial applications of technology to solve problems involving perpetuities

7. C **8.** $75.05 **9.** 2.93% p.a.

10. $2407.73

11. a. $151 080.96 **b.** $429.32 **c.** $12 879.60

Joining it all together

12. a. $2941.26

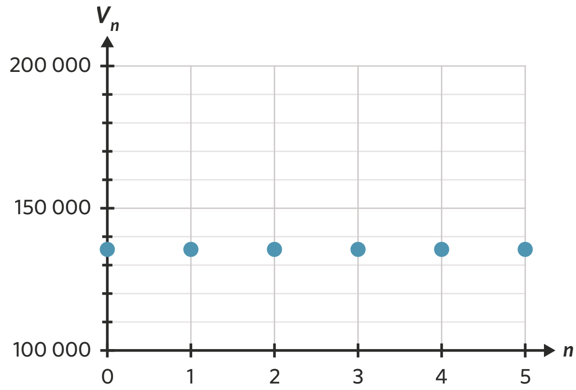
b. $V_0 = 245\,105, \quad V_{n+1} = 1.012 \times V_n - 2941.26$

c. $245 105

13. a. 0.2%

b. $V_0 = 135\,560, \quad V_{n+1} = 1.002 \times V_n - 271.12$

c.



14. a. $455 891.43 **b.** $263.01 **c.** $294 051.42

Exam practice

15. Explanation

Step 1: Interpret the graph.

The value of the annuity investment, A_n, remains constant each time period. This means it is modelling a perpetuity.

Step 2: To solve this question, check if $A_1 = 200\,000$ for each option.

A: $A_1 = 1.015 \times 200\,000 - 2500$
$= 200\,500$ ✗

B: $A_1 = 1.025 \times 200\,000 - 5000$
$= 200\,000$ ✓

C: $A_1 = 1.03 \times 200\,000 - 5500$
$= 200\,500$ ✗

D: $A_1 = 1.04 \times 200\,000 - 6000$
$= 202\,000$ ✗

E: $A_1 = 1.05 \times 200\,000 - 8000$
$= 202\,000$ ✗

Answer

B

16. Explanation

Step 1: Determine the principal of the perpetuity.

$80 000 was invested.

$V_0 = 80\,000$

Step 2: Identify the correct graph.

The *value* ($) of the perpetuity needs to remain constant at $80 000.

Answer

B

43% of students incorrectly answered D or E. These students likely did not realise the value of a perpetuity remains the same over all compounding periods.

17. a. Explanation

Calculate the total amount received over 12 payments.

$\textit{total amount} = 1890 \times 12$
$= 22\,680$

Back to contents

Answer

$22 680

b. **Explanation**

Step 1: Determine the principal of the perpetuity.

$420 000 was invested.

$V_0 = 420\ 000$

Step 2: Determine the balance after one year.

The balance of a perpetuity remains the same each period.

Answer

$420 000

A significant number of students did not realise that the balance of a perpetuity remains the same for all compounding periods.

c. **Explanation**

Step 1: Determine the principal and interest rate.

$S_0 = 420\ 000$

$r = \frac{5.4}{12}\%$ per compounding period

Step 2: Calculate R.

$$R = 1 + \frac{5.4}{12 \times 100}$$
$$= 1 + \frac{5.4}{1200}$$
$$= 1.0045$$

Step 3: Construct the recurrence relation.

Answer

$S_0 = 420\ 000, \quad S_{n+1} = 1.0045 \times S_n - 1890$

A number of students incorrectly calculated R as 1.054. This is likely because they used the annual interest rate rather than the rate per compounding period.

Questions from multiple lessons

18. E **19.** C

20. a. $28 733.34 b. $1747.94 c. $12 682

6F Annuity investments

Using recurrence relations to model annuity investments

1. B

2. a. $6500 b. $180 c. 4.2% p.a.

3. a. $V_0 = 2000, \quad V_{n+1} = 1.001 \times V_n + 10$

b. $V_0 = 500, \quad V_{n+1} = 1.0035 \times V_n + 40$

4. a. 3% p.a. b. $8500.63 c. 4 months

Using tables to solve problems involving annuity investments

5. C

6. a. $2500.00 b. $2859.34 c. $369.07

d. $114 296.89 e. 4.08% p.a.

7.

payment number	payment	interest	principal addition	balance of investment
0	0.00	0.00	0.00	5600.00
1	200.00	19.13	219.13	5819.13
2	200.00	19.88	219.88	6039.01

8. a.

payment number	payment	interest	principal addition	balance of investment
0	0.00	0.00	0.00	104 123.00
1	1000.00	90.24	1090.24	105 213.24
2	1000.00	91.18	1091.18	106 304.42
3	1000.00	92.13	1092.13	107 396.55

b. $273.55

c. The interest amount is calculated off the increasing previous investment balance.

d. The principal addition is calculated using the increasing interest calculation.

Using financial applications of technology to solve problems involving annuity investments

9. C **10.** $729.87 **11.** 2.1% p.a.

12. a. $130 000 b. $3090.58 c. 14 quarters

13. a. $40 837.31 b. $30 064.79 c. 2.8% p.a.

Joining it all together

14. a. 4.316% p.a.

b. $V_0 = 7000, \quad V_{n+1} = 1.00083 \times V_n + 30$

c. 6 weeks

15. a. $184.96 b. $281.51

Exam practice

16. a. **Explanation**

Identify d, the payment made per compounding period.

A recurrence relation for an annuity investment is of the form $V_0 = principal, \quad V_{n+1} = R \times V_n + d$.

In this case, $d = 500$.

This means that $500 is added to the investment annually.

Answer

D

11% of students incorrectly answered E. This may be because they considered the interest calculated as a part of the regular payment the annuity receives.

Back to contents

b. Explanation

Step 1: Use recursion to calculate V_1, V_2 and V_3.

$V_0 = 46\,000$

$V_1 = 46\,656.40$

$V_2 = 47\,315.031...$

$V_3 = 47\,975.902...$

Step 2: Calculate the difference between V_3 and V_2.

$$\begin{aligned} \textit{difference} &= 47\,975.902... - 47\,315.031... \\ &= 660.871... \\ &\approx 661 \end{aligned}$$

Answer

C

17% of students who chose B calculated the difference between V_1 and V_0 to obtain \$658. 11% of students who answered D calculated the difference between V_0 and V_2 to obtain \$1315.

17. Explanation

Step 1: Determine the financial solver inputs

N	60	(there are 60 months in 5 years)
I(%)	2.8	(annual interest rate)
PV	−12 000	(this is negative because Joanna invests it with the bank)
PMT		
FV	25 000	(this is positive because it represents a future payment that will be made to Joanna once she closes the annuity)
PpY	12	(payments made monthly)
CpY	12	(interest compounds monthly)

Step 2: Use the financial solver to solve for PMT.

PMT	−174.106...

This is negative because Joanna deposits it into the account.

Answer

B

23% of students incorrectly answered D. This is likely because they had either entered both PV and FV as positive values or both as negative. It is important to note the direction of cash flow in questions that use the financial solver so that accurate figures can be obtained.

18. Explanation

Step 1: Calculate the principal addition after payment 20.

It is the difference between the current and previous investment balance.

$$\begin{aligned} \textit{principal addition} &= 7500 - 7233.83 \\ &= 266.17 \end{aligned}$$

Step 2: Calculate the interest rate per compounding period.

The interest rate can be calculated using any two consecutive payment numbers.

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

$$27.91 = \frac{r}{100} \times 6977.50$$

$$r = \frac{27.91}{6977.50} \times 100$$

$r = 0.4\%$ per compounding period.

Step 3: Calculate the interest earned in payment 20.

$$\begin{aligned} \textit{interest} &= \frac{0.4}{100} \times 7233.83 \\ &= 28.94 \end{aligned}$$

Step 4: Calculate the payment amount.

$$266.17 = \textit{payment} + 28.94$$

$$\begin{aligned} \textit{payment} &= 266.17 - 28.94 \\ &= 237.23 \\ &\approx 237 \end{aligned}$$

Answer

D

35% of students incorrectly answered B. This is likely because they had assumed that the payment amount is always constant, despite the question stating that the payment might vary from one compounding period to the next.

19. Explanation

Step 1: Determine the value of the investment after four years.

N	48	(there are 48 months in 4 years)
I(%)	3.8	(annual interest rate)
PV	−360 000	(this is negative because Alex invests it with the bank)
PMT	−500	(this is negative because Alex deposits it into the account)
FV		
PpY	12	(payments made monthly)
CpY	12	(interest compounds monthly)

FV	444 872.944...

This is positive because it represents a future payment that will be made to Alex once he closes the annuity.

Back to contents

Step 2: Determine the required payment for the next two years.

The balance at the start of this period is \$444 872.944..., so this is treated as the present value, PV.

N	24	(there are 24 months in 2 years)
I(%)	3.8	(annual interest rate)
PV	−444 872.944...	(this is negative because Alex invests it with the bank)
PMT		
FV	500 000	(this is positive because it represents a future payment that will be made to Alex once he closes the annuity)
PpY	12	(payments made monthly)
CpY	12	(interest compounds monthly)

PMT	−805.650...

This is negative because Alex deposits it into the account.

Answer

\$805.65

A number of students incorrectly answered \$393 121.15 as the value of the investment after four years. This is likely because they had entered a positive PMT value into their financial solver in the first part of this question, rather than a negative.

Questions from multiple lessons

20. B

21. D

22. a. \$1000

b. $V_0 = 1000$
$V_1 = 1.05 \times 1000 = 1050$
$V_2 = 1.05 \times 1050 = 1102.50$
$V_3 = 1.05 \times 1102.50 \approx 1157.63$

c. 5%

d. i. $V_n = 1.05^n \times 1000$

ii. \$1407.10

Back to contents

7A Introduction to matrices

Identifying matrix properties and types

1. A

2. a. 4 rows and 2 columns. **b.** 4×2

c. 8

3. a. 3×1 **b.** 2×3 **c.** 3×3 **d.** 2×4

4. a. 46 **b.** 47 **c.** 111 **d.** 53

5. a. M, N and O **b.** M and O

c. O **d.** M, O and P

Constructing and interpreting matrices

6. C

7. a. The number of residents in household 2, which is 3.

b. The number of cats in household 3, which is 9.

c. The total number of residents across all 3 households, which is 9.

d. The total number of cats across all 3 households, which is 12.

8. a.

$$\begin{array}{cl} \begin{array}{ccc} F & P & R \end{array} & \\ \begin{bmatrix} 15 & 8 & 9 \\ 13 & 14 & 5 \end{bmatrix} & \begin{array}{l} \text{boys} \\ \text{girls} \end{array} \end{array}$$

b. The number of girls whose favourite animal is a porcupine, which is 14.

c. The total number of boys surveyed, which is 32.

9. a. $\begin{bmatrix} 4 & 6 \\ 6 & 8 \end{bmatrix}$ **b.** $\begin{bmatrix} 2 & 4 \\ 4 & 8 \\ 6 & 12 \end{bmatrix}$

c. $\begin{bmatrix} 4 & 9 & 16 & 25 \\ 9 & 16 & 25 & 36 \end{bmatrix}$

Joining it all together

10. a. $\begin{bmatrix} 23 & 19 & 28 \\ 13 & 20 & 9 \\ 14 & 11 & 26 \end{bmatrix}$

b. The total number of ice cream scoops sold over three days, which is 163.

c. $\begin{bmatrix} 19 \\ 20 \\ 11 \end{bmatrix}$

d. The total number of ice cream scoops sold on day 2, which is 50.

11. a. 1×4

b. Row matrix

c. The total amount of water stored on Wednesday, which is 900 L.

Exam practice

12. a. Explanation

Identify the element that corresponds to G.

Each row corresponds to a particular section of the toll road.

$$\begin{array}{cl} \begin{bmatrix} 3.58 \\ 2.22 \\ 2.87 \end{bmatrix} & \begin{array}{l} E \\ F \\ G \end{array} \end{array}$$

The element in row 3 refers to section G of the toll road.

Answer

\$2.87

b. Explanation

Step 1: Count the number of rows.

The matrix has 3 rows.

$$\begin{array}{cl} \begin{bmatrix} 3.58 \\ 2.22 \\ 2.87 \end{bmatrix} & \begin{array}{l} \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \end{array} \end{array}$$

Step 2: Count the number of columns.

The matrix has 1 column.

$$\begin{array}{c} \mathbf{1} \\ \begin{bmatrix} 3.58 \\ 2.22 \\ 2.87 \end{bmatrix} \end{array}$$

Step 3: Determine the order of matrix C.

Answer

3×1

c. Explanation

Step 1: Set up the matrix.

A row matrix has one row and any number of columns. As there are three separate pieces of information, there should be three columns, one for each section. As such, the order of the matrix should be 3×1.

$$\begin{array}{c} \begin{array}{ccc} E & F & G \end{array} \\ \begin{bmatrix} \quad & \quad & \quad \end{bmatrix} \end{array}$$

Step 2: Fill in the elements.

The first column corresponds to the number of times Kim travelled on section E, which is 2.

The second column corresponds to the number of times Kim travelled on section F, which is 0.

The third column corresponds to the number of times Kim travelled on section G, which is 1.

Answer

$\begin{bmatrix} 1 & 0 & 2 \end{bmatrix}$

Questions from multiple lessons

13. B **14.** B

15. a. 79.1% **b.** 12.2% **c.** 55%

7B Operations with matrices

Adding and subtracting matrices

1. C

2. a. defined
$\begin{bmatrix} 7 & -3 & 4 & 10 \end{bmatrix}$

b. undefined

c. defined
$\begin{bmatrix} 1 & 1 & 4 \\ 2 & 1 & 9 \end{bmatrix}$

d. defined
$\begin{bmatrix} -1 & -6 \\ -17 & 10 \\ 3 & -11 \\ 2 & 0 \end{bmatrix}$

3. a. $S = \begin{bmatrix} 10 \\ 8 \\ 15 \end{bmatrix}$, $D = \begin{bmatrix} 9 \\ 8 \\ 12 \end{bmatrix}$

b. $\begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}$

The difference between the average number of passengers driven by Simon and Darren each day.

c. $T = \begin{bmatrix} 14 \\ 11 \\ 16 \\ 20 \end{bmatrix}$

d. No, it is not possible to evaluate $S + D + T$ because they are not all of the same order.

Multiplying matrices by a scalar

4. D

5. a. $\begin{bmatrix} 6 & 12 & 2 & 18 & 10 \end{bmatrix}$

b. $\begin{bmatrix} 8 & 20 & -24 \\ -12 & 4 & 16 \end{bmatrix}$

c. $\begin{bmatrix} -36 \\ 21 \\ 12 \\ -27 \\ -24 \\ -33 \end{bmatrix}$

d. $\begin{bmatrix} 9 & 15 & 7.5 \\ 6 & -4.5 & -30 \\ -1.5 & 12 & 3 \end{bmatrix}$

6. a. 0.8

b. $\begin{bmatrix} 32 & 36 \\ 64 & 72 \\ 80 & 96 \end{bmatrix}$

c. \$72

Determining the transpose of a matrix

7. A

8. a. $\begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix}$

b. $\begin{bmatrix} 1 & 9 \\ 5 & 12 \end{bmatrix}$

c. $\begin{bmatrix} -7 & 13 & -1 \\ 6 & 5 & 15 \end{bmatrix}$

d. $\begin{bmatrix} 8 & -11 & 4 \\ 9 & 1 & 0 \\ 12 & 5 & -2 \end{bmatrix}$

9.

$$C^T = \begin{array}{c} \begin{array}{ccc} \text{Mon} & \text{Tue} & \text{Wed} \end{array} \\ \begin{bmatrix} 51 & 38 & 50 \\ 33 & 21 & 26 \\ 60 & 30 & 62 \end{bmatrix} \end{array} \begin{array}{l} \\ \text{almond croissant} \\ \text{brownie} \\ \text{croissant} \end{array}$$

Joining it all together

10. a. undefined

b. undefined

c. defined
$\begin{bmatrix} 16 \\ 2 \\ 15 \\ -7 \\ 5 \end{bmatrix}$

d. defined
$\begin{bmatrix} -1 & -1 \\ 19 & -1 \\ -26 & 0 \end{bmatrix}$

11. a. $x = -2$
$y = 6$

b. $x = -8$
$y = 6$

c. $x = -1.5$
$y = -6$

d. $x = -18$
$y = 1$

12. a.

$$O_{2021} = \begin{array}{c} \begin{array}{cccc} \text{medicine} & \text{literature} & \text{law} & \text{philosophy} \end{array} \\ \begin{bmatrix} 276 & 288 & 336 & 180 \\ 264 & 180 & 312 & 144 \end{bmatrix} \end{array} \begin{array}{l} \\ \text{local} \\ \text{international} \end{array}$$

b. 572 students

c.

$$C_{2021}{}^T = \begin{array}{c} \begin{array}{cccc} \text{medicine} & \text{literature} & \text{law} & \text{philosophy} \end{array} \\ \begin{bmatrix} 305 & 372 & 320 & 266 \\ 350 & 231 & 318 & 175 \end{bmatrix} \end{array} \begin{array}{l} \\ \text{local} \\ \text{international} \end{array}$$

d. $\begin{bmatrix} 29 & 84 & -16 & 86 \\ 86 & 51 & 6 & 31 \end{bmatrix}$

e. 86 students

Exam practice

13. Explanation

A matrix addition is defined if the matrices involved in the addition are of the same order.

In the first expression, both matrices are of the order 2×1. The expression is defined.

In the second expression, one matrix is of the order 2×1 and one matrix is of the order 2×2. The expression is undefined.

In the third expression, one matrix is of the order 2×2 and one matrix is of the order 2×1. The expression is undefined.

In the fourth expression, both matrices are of the order 2×2. The expression is defined.

Answer

C

14. Explanation

Method 1: By hand

Step 1: Determine the order of the transpose.

The order of the original matrix is 2×3.

Hence, the order of the transpose will be 3×2.

Step 2: Transpose the matrix by swapping its rows and columns.

$\begin{bmatrix} 2 & 13 \\ 7 & 19 \\ 10 & 8 \end{bmatrix}$

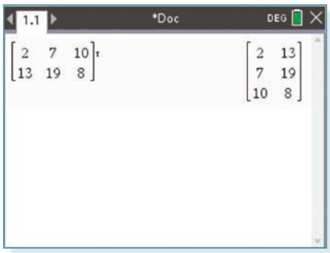
Method 2: TI-Nspire

Step 1: From the home screen, select '1: New' → '1: Add Calculator'.

Back to contents

Step 2: Press and select . On the settings window, set 'Number of rows' as 2 and 'Number of columns' as 3. Select 'OK'. Enter the values for the original matrix.

Step 3: Press menu and select '7: Matrices' → '2: Transpose'. Press enter.

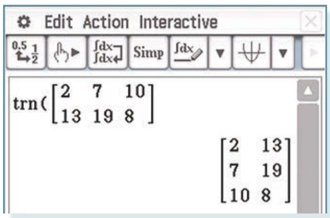


Method 3: Casio ClassPad

Step 1: From the main menu, tap √α Main.

Step 2: Tap 'Action' → 'Matrix' → 'Create' → 'trn'.

Step 3: Press keyboard and tap Math2. Tap to create a matrix and to add an extra column. Enter the values for the original matrix. Press EXE.



Answer – Method 1, 2 and 3

C

15. Explanation

Step 1: Add the rules for matrices A and B to find a singular rule for matrix $A + B$.

$$(a+b)_{ij} = 2i + j + i - j$$
$$= 3i$$

Step 2: Interpret the rule to construct matrix $A + B$.

The rule for each element is $3i$.

All elements in the first row will be $3 \times 1 = 3$.

All elements in the second row will be $3 \times 2 = 6$.

All elements in the third row will be $3 \times 3 = 9$.

$$\begin{bmatrix} 3 & 3 & 3 \\ 6 & 6 & 6 \\ 9 & 9 & 9 \end{bmatrix}$$

Answer

D

13% of students incorrectly chose option C. This is likely because they mixed up the definitions of i and j and believed that i referred to the matrix's columns instead of its rows.

Questions from multiple lessons

16. B **17.** B

18. a. 4×1

b. i. [2146]

ii. The total revenue from booking fees collected last month.

7C Advanced operations with matrices

Defining matrix products

1. C

2. a. Not defined **b.** Defined; 2×3

c. Not defined **d.** Defined; 2×1

e. Defined; 1×2 **f.** Not defined

3. a. B **b.** B

Calculating a matrix product

4. D

5. a. $[7]$ **b.** $[-2 \quad 4]$

c. $\begin{bmatrix} 2 \\ 4 \end{bmatrix}$ **d.** $\begin{bmatrix} 2 & 0 \\ 3 & -4 \end{bmatrix}$

e. $\begin{bmatrix} 25 & 17 \\ 0 & -31 \end{bmatrix}$ **f.** $\begin{bmatrix} -32 & -27 & 29 \\ 94 & 11 & 43 \\ 71 & 5 & 75 \end{bmatrix}$

6. a. $\begin{bmatrix} 26 \\ 19 \end{bmatrix}$

b. The total number of points received in the tournament by each team.

c. Soccerolo; 26 points

7. a. The number of likes for post 3.

b. $[255 \quad 345 \quad 390 \quad 270]$

c. $12.60

d. Post 3

8. a. $M = \begin{bmatrix} 5 \\ 2 \\ 3 \end{bmatrix} \quad LM = \begin{bmatrix} 31.35 \\ 29.8 \end{bmatrix}$

b. Cheesetopia

c. No, Cheeseworld is still more expensive, charging $29.25 in total compared to $28.80 at Cheesetopia.

Using a summing matrix

9. A

10. a. $[1 \quad 1 \quad 1 \quad 1]$ **b.** $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$

11. a. $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

b. Store 1

$$\begin{bmatrix} 20 & 24 & 24 \\ 16 & 30 & 21 \\ 18 & 15 & 28 \\ 22 & 24 & 19 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 68 \\ 67 \\ 61 \\ 65 \end{bmatrix}$$

c. $[1 \quad 1 \quad 1 \quad 1]$

Back to contents

d. Macaroons

$$[1 \quad 1 \quad 1 \quad 1] \times \begin{bmatrix} 20 & 24 & 24 \\ 16 & 30 & 21 \\ 18 & 15 & 28 \\ 22 & 24 & 19 \end{bmatrix} = [76 \quad 93 \quad 92]$$

Calculating a matrix power

12. C

13. a. Defined **b.** Not defined

c. Not defined **d.** Defined

e. Defined

14. a. $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ **b.** $\begin{bmatrix} 3 & -6 \\ 5 & -2 \end{bmatrix}$

c. $\begin{bmatrix} -3 & -6 \\ -1 & -2 \end{bmatrix}$

Joining it all together

15. a. Defined; 2×2 **b.** Matrix J

c. Defined; $\begin{bmatrix} -3450 & -3511 \\ -6242 & -6403 \end{bmatrix}$

d. $\begin{bmatrix} 3999 & 3762 \\ 7227 & 6870 \end{bmatrix}$

16. a. $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ **b.** $\begin{bmatrix} 18.4 & 9.6 & 5.6 \\ 23.4 & 18.2 & 10.4 \end{bmatrix}$

Area 1: 42 000 birds
Area 2: 40 000 birds

c. 85 600 birds **d.** Area 1: 20% decrease
Area 2: 30% increase

Exam practice

17. Explanation

Step 1: Construct a summing matrix that sums the columns of matrix M.

To sum the columns of a matrix, a row summing matrix is required.

The number of columns in the summing matrix should equal the number of rows in matrix M.

Matrix M has 3 rows. The summing matrix must be a 1×3 matrix.

$[1 \quad 1 \quad 1]$

Step 2: Pre-multiply matrix M with the summing matrix.

$$[1 \quad 1 \quad 1] \times \begin{bmatrix} 35 & 24 & 60 \\ 28 & 32 & 43 \\ 32 & 30 & 56 \end{bmatrix} = [95 \quad 86 \quad 159]$$

Step 3: Determine how many sandwiches were sold in the three weeks.

The third element represents the sum of the sandwiches sold each week.

Answer

159

Students needed to make sure to find the sum of the columns and not the rows. Also, a few students obtained an answer of 56 which was the number of sandwiches sold only in week 3.

18. Explanation

Step 1: Evaluate the matrix product.

$$[4 \quad 2 \quad 0] \times \begin{bmatrix} 4 \\ 12 \\ 8 \end{bmatrix} = [4 \times 4 + 2 \times 12 + 0 \times 8]$$
$$= [40]$$

Step 2: Evaluate each option and compare it with the matrix product.

A: $[144] \neq [40]$ ✗

B: $\begin{bmatrix} 16 \\ 24 \\ 0 \end{bmatrix} \neq [40]$ ✗

C: $4 \times [1 \quad 2 \quad 0] \times \begin{bmatrix} 1 \\ 12 \\ 8 \end{bmatrix} = [100] \neq [40]$ ✗

D: $2 \times [2 \quad 1 \quad 0] \times \begin{bmatrix} 2 \\ 6 \\ 4 \end{bmatrix} = [20] \neq [40]$ ✗

E: $4 \times [2 \quad 1 \quad 0] \times \begin{bmatrix} 2 \\ 6 \\ 4 \end{bmatrix} = [40]$ ✓

Answer

E

15% of students incorrectly chose option B. This resulted from incorrectly applying matrix multiplication whereby each corresponding element was multiplied, but the sum of these values was not calculated.

Questions from multiple lessons

19. B **20.** E **21. a.** 23% **b.** $7675

7D Inverse matrices

Calculating the determinant of a matrix

1. A

2. a. 0 **b.** 2 **c.** $-\frac{19}{8}$

3. a. -49 **b.** -6 **c.** 37

4. $\frac{3}{2}$

Calculating the inverse of a matrix

5. B

6. a. $A^{-1} = \begin{bmatrix} \frac{3}{22} & -\frac{1}{22} \\ -\frac{1}{11} & \frac{4}{11} \end{bmatrix}$ **b.** $B^{-1} = \begin{bmatrix} \frac{3}{4} & -\frac{1}{2} \\ -\frac{1}{6} & 0 \end{bmatrix}$

c. $C^{-1} = \begin{bmatrix} \frac{17}{65} & \frac{12}{65} & -\frac{14}{65} \\ \frac{24}{65} & -\frac{6}{65} & \frac{7}{65} \\ -\frac{4}{13} & \frac{1}{13} & \frac{1}{13} \end{bmatrix}$ **d.** Undefined

7. a. Yes **b.** No **c.** No

8. 6

Back to contents

Solving simultaneous equations using matrix equations

9. B

10. a. $\begin{bmatrix}1 & 2\\2 & 3\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}2\\6\end{bmatrix}$ **b.** $\begin{bmatrix}3 & -2\\1 & -3\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}5\\5\end{bmatrix}$

c. $\begin{bmatrix}-3 & 1\\2 & 1\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}-3\\1\end{bmatrix}$ **d.** $\begin{bmatrix}2 & -4 & 6\\-1 & 3 & 5\\-3 & 2 & 5\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix} = \begin{bmatrix}12\\2\\5\end{bmatrix}$

11. a. $x = 2, \quad y = 6$ **b.** $x = -8, \quad y = 5$

c. $x = 1, \quad y = 5$

12. a. $\begin{bmatrix}3 & 5\\3 & 6\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}-2\\-2\end{bmatrix}$

The determinant is 3, so a unique solution exists.

b. $\begin{bmatrix}-2 & 3\\-3 & 4.5\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}$

The determinant is 0, so no unique solution exists.

c. $\begin{bmatrix}1 & -3\\2 & 4\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}6\\3\end{bmatrix}$

The determinant is 10, so a unique solution exists.

d. $\begin{bmatrix}2 & -3\\-3 & 2\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}6\\6\end{bmatrix}$

The determinant is -5, so a unique solution exists.

Joining it all together

13. a. $\begin{bmatrix}7 & 6\\10 & 4\end{bmatrix}\begin{bmatrix}b\\m\end{bmatrix} = \begin{bmatrix}120\\120\end{bmatrix}$

b. -32

c. $p = -32$, and $q = -10$

d. Body wash: \$7.50
Moisturiser: \$11.25

14. a. D

b. $\begin{bmatrix}350\\150\end{bmatrix}$

c. $\begin{bmatrix}78 & 19\\37 & 5\end{bmatrix}\begin{bmatrix}r\\l\end{bmatrix} = \begin{bmatrix}350\\150\end{bmatrix}$

d. Regular = \$3.50, Large = \$4.00

Exam practice

15. Explanation

To solve this question, calculate the determinant for each option.

A: $\det\left(\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}\right) = 0 \times 0 - 1 \times 1$
$= -1$ ✗

B: $\det\left(\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right) = 1 \times 1 - 0 \times 0$
$= 1$ ✗

C: $\det\left(\begin{bmatrix}1 & 2\\-3 & 6\end{bmatrix}\right) = 1 \times 6 - 2 \times (-3)$
$= 12$ ✗

D: $\det\left(\begin{bmatrix}3 & 6\\2 & 4\end{bmatrix}\right) = 3 \times 4 - 6 \times 2$
$= 0$ ✓

E: $\det\left(\begin{bmatrix}4 & 0\\0 & -2\end{bmatrix}\right) = 4 \times (-2) - 0 \times 0$
$= -8$ ✗

Answer

D

16. a. Explanation

Identify the feature of a matrix that enables it to be an inverse matrix.

The determinant of a matrix must not be equal to zero. The determinant of this matrix is 1, therefore it has an inverse.

Answer

The inverse of a matrix is defined if the matrix's determinant is not equal to zero.

Many students were unable to realise that for an inverse matrix to exist, a non-zero determinant is required. Some students stated that the determinant had to be greater than 0, and did not realise the determinant can be negative for the inverse to exist.

b. Explanation

Step 1: Identify a, b, c and d in the matrix.

In the matrix $\begin{bmatrix}5 & -9\\4 & -7\end{bmatrix}$:

$a = 5$

$b = -9$

$c = 4$

$d = -7$

Step 2: Substitute the values into the formula to find the inverse matrix.

$= \frac{1}{ad - bc}\begin{bmatrix}d & -b\\-c & a\end{bmatrix}$

$= \frac{1}{(5 \times -7) - (-9 \times 4)}\begin{bmatrix}-7 & 9\\-4 & 5\end{bmatrix}$

$= \frac{1}{1}\begin{bmatrix}-7 & 9\\-4 & 5\end{bmatrix}$

Answer

$\begin{bmatrix}-7 & 9\\-4 & 5\end{bmatrix}$

When substituting the values of a, b, c and d into the inverse matrix formula, some students did not include the negative signs. Meanwhile, other students did not correctly interchange the position of 5 and -7.

c. Explanation

Step 1: Use the inverse matrix found in part **b** to solve the matrix equation.

$\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}-7 & 9\\-4 & 5\end{bmatrix} \times \begin{bmatrix}7\\6\end{bmatrix}$

$= \begin{bmatrix}5\\2\end{bmatrix}$

Step 2: Interpret the column matrix.

The second row shows the number of preferred sandwich bars in Grandmall's food court.

Answer

2

Back to contents

17. Explanation

Step 1: Find the determinant of the coefficient matrix.

$$\det\left(\begin{bmatrix} 12 & 9 \\ m & 3 \end{bmatrix}\right) = 12 \times 3 - 9 \times m$$
$$= 36 - 9m$$

Step 2: Let the determinant equal 0 and solve for m.

For there to be no unique solution, the determinant must be equal to 0.

$$36 - 9m = 0$$
$$9m = 36$$
$$m = 4$$

Answer

E

15% of students incorrectly chose option C. This is most likely because students failed to recognise that the determinant must equal 0 in order for there to be no unique solution. Students needed to be careful not to make arithmetic mistakes when solving for m, and needed to ensure the determinant equals 0.

Questions from multiple lessons

18. B

19. C

20. a. i. 16%

ii. 1997 days

b. \$238

7E Binary and permutation matrices

Applying permutations to matrices

1. a. B, C, D and E

b. C and E

2. a. $MP = \begin{bmatrix} 7 & 4 \\ 9 & 3 \end{bmatrix}$

b. $MP = \begin{bmatrix} 0 & -3 & 4 \end{bmatrix}$

c. $MP = \begin{bmatrix} f & v & m \\ c & p & q \end{bmatrix}$

d. $MP = \begin{bmatrix} 19 & 15 & 12 & 27 \\ 9 & 7 & 11 & 15 \\ 17 & 14 & 15 & 8 \end{bmatrix}$

3. a. $PQ = \begin{bmatrix} 3 & 8 \\ 4 & 9 \\ 6 & 7 \end{bmatrix}$

b. $PQ = \begin{bmatrix} 4 \\ 1 \\ -2 \\ 0 \end{bmatrix}$

c. $PQ = \begin{bmatrix} h & j & k & l \\ a & s & d & f \end{bmatrix}$

d. $PQ = \begin{bmatrix} 22 & 31 \\ 25 & 19 \\ 14 & 17 \end{bmatrix}$

Constructing permutation matrices

4. C

5. a. $P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$

b. Pre-multiplication

6. $$\begin{bmatrix} H & S & M & A \end{bmatrix} \times \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} M & A & S & H \end{bmatrix}$$

Joining it all together

7. a. $$\begin{bmatrix} i & r & l & s \\ b & u & h & d \\ n & a & e & m \\ t & c & a & t \end{bmatrix} \times \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} s & l & r & i \\ d & h & u & b \\ m & e & a & n \\ t & a & c & t \end{bmatrix}$$

Note: There are other possible solutions.

b. $$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} i & r & l & s \\ b & u & h & d \\ n & a & e & m \\ t & c & a & t \end{bmatrix} = \begin{bmatrix} b & u & h & d \\ n & a & e & m \\ t & c & a & t \\ i & r & l & s \end{bmatrix}$$

Note: There are other possible solutions.

Exam practice

8. Explanation

Complete the matrix multiplication using a calculator.

$$\begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} L \\ E \\ A \\ P \\ S \end{bmatrix} = \begin{bmatrix} P \\ A \\ L \\ E \\ S \end{bmatrix}$$

Answer

D

9. Explanation

Step 1: Determine the type of permutation.

As the permutation matrix is pre-multiplying another matrix, this is a row permutation.

Step 2: Determine the location of each '1' in the permutation matrix from the information given.

Row 1 becomes row 3 – p_{31}

Row 2 stays as row 2 – p_{22}

Row 3 becomes row 1 – p_{13}

Row 4 stays as row 4 – p_{44}

Step 3: Define the permutation matrix.

$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Answer

B

The most common incorrect response was option D. This was a combination of not placing the 1's in the correct location in rows 1 and 3 of the permutation matrix, but also not realising that there still needed to be a 1 placed in each column and row for it to be a valid permutation matrix.

10. Explanation

Step 1: Input and store matrices P and W using your calculator.

Step 2: Substitute each possible value for n to determine the correct response.

For $n = 1$

$$P^1 \times W = \begin{bmatrix} T \\ R \\ S \\ O \\ A \end{bmatrix} ✗$$

Back to contents

For $n = 2$

$$P^2 \times W = \begin{bmatrix} S \\ A \\ R \\ O \\ T \end{bmatrix} ✗$$

For $n = 3$

$$P^3 \times W = \begin{bmatrix} R \\ T \\ A \\ O \\ S \end{bmatrix} ✗$$

For $n = 4$

$$P^4 \times W = \begin{bmatrix} A \\ S \\ T \\ O \\ R \end{bmatrix} ✓$$

For $n = 5$

$$P^5 \times W = \begin{bmatrix} T \\ R \\ S \\ O \\ A \end{bmatrix} ✗$$

Answer

D

19% of students incorrectly chose option A. This was most likely due to the incorrect assumption that raising the permutation matrix to a power of 1 will result in no change. These students most likely did not understand that they could input each possible response into their calculator to find the correct response.

11. Explanation

Step 1: Determine the type of permutation.

The first two rows are being interchanged. Therefore a row permutation is being applied.

Step 2: Determine the location of each '1' in the permutation matrix from the information given.

Row 1 becomes row 2 – m_{21}

Row 2 becomes row 1 – m_{12}

Row 3 stays as row 3 – m_{33}

Row 4 stays as row 4 – m_{44}

Step 3: Construct the permutation matrix.

Answer

$$M = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Many students might not have recalled that a permutation matrix is a specific type of binary matrix, and did not understand that a permutation matrix was required.
A number of students incorrectly gave the identity matrix as the solution.

Questions from multiple lessons

12. C

13. C

14. a. \$10

b. 3×1

c. $S = [3 \quad 2 \quad 0]$

7F Communication and dominance matrices

Interpreting and constructing communication matrices

1. A

2. a.

$$\begin{array}{c@{}c} \begin{matrix} \mathrm{A} & \mathrm{B} & \mathrm{C} & \mathrm{D} \end{matrix} & \\ \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} & \begin{matrix} \mathrm{A} \\ \mathrm{B} \\ \mathrm{C} \\ \mathrm{D} \end{matrix} \end{array}$$

b.

$$\begin{array}{c@{}cc} \text{receiver} & & \\ \begin{matrix} \mathrm{H} & \mathrm{I} & \mathrm{J} & \mathrm{K} \end{matrix} & & \\ \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix} & \begin{matrix} \mathrm{H} \\ \mathrm{I} \\ \mathrm{J} \\ \mathrm{K} \end{matrix} & \text{sender} \end{array}$$

c.

$$\begin{array}{c@{}c} \begin{matrix} \mathrm{A} & \mathrm{B} & \mathrm{C} & \mathrm{D} & \mathrm{E} \end{matrix} & \\ \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{bmatrix} & \begin{matrix} \mathrm{A} \\ \mathrm{B} \\ \mathrm{C} \\ \mathrm{D} \\ \mathrm{E} \end{matrix} \end{array}$$

d.

$$\begin{array}{c@{}cc} \text{receiver} & & \\ \begin{matrix} \mathrm{Q} & \mathrm{R} & \mathrm{S} & \mathrm{T} & \mathrm{U} \end{matrix} & & \\ \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} & \begin{matrix} \mathrm{Q} \\ \mathrm{R} \\ \mathrm{S} \\ \mathrm{T} \\ \mathrm{U} \end{matrix} & \text{sender} \end{array}$$

3. a. P1 and P3 cannot communicate with each other.

b. P2 and P3

c. Undirected

4. a. Ariana and Chance

b.

$$\begin{array}{c@{}c} \begin{matrix} \mathrm{A} & \mathrm{B} & \mathrm{C} & \mathrm{D} & \mathrm{E} \end{matrix} & \\ \begin{bmatrix} 2 & 1 & 2 & 1 & 1 \\ 1 & 2 & 1 & 2 & 1 \\ 2 & 1 & 3 & 2 & 2 \\ 1 & 2 & 2 & 3 & 2 \\ 1 & 1 & 2 & 2 & 2 \end{bmatrix} & \begin{matrix} \mathrm{A} \\ \mathrm{B} \\ \mathrm{C} \\ \mathrm{D} \\ \mathrm{E} \end{matrix} \end{array}$$

c. 12

5. a.

$$T = \begin{array}{c@{}cc} \text{receiver} & & \\ \begin{matrix} \mathrm{A} & \mathrm{B} & \mathrm{C} & \mathrm{D} & \mathrm{E} & \mathrm{F} \end{matrix} & & \\ \begin{bmatrix} 1 & 1 & 2 & 2 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 3 & 1 & 2 \\ 1 & 0 & 2 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix} & \begin{matrix} \mathrm{A} \\ \mathrm{B} \\ \mathrm{C} \\ \mathrm{D} \\ \mathrm{E} \\ \mathrm{F} \end{matrix} & \text{sender} \end{array}$$

b. No

c. 3 links – F to D, D to B and D to F

Back to contents

Interpreting and constructing dominance matrices

6. B

7. a.

$$\begin{array}{cc} \text{loser} & \\ \begin{matrix} A & B & C & D \end{matrix} & \\ \begin{bmatrix} 0&1&1&1\\0&0&0&1\\0&1&0&1\\0&0&0&0 \end{bmatrix} & \begin{matrix} A\\B\\C\\D \end{matrix}\ \text{winner} \end{array}$$

b.

$$\begin{array}{cc} \text{loser} & \\ \begin{matrix} A & B & C & D & E \end{matrix} & \\ \begin{bmatrix} 0&1&0&0&1\\0&0&1&1&0\\1&0&0&1&1\\1&0&0&0&1\\0&1&0&0&0 \end{bmatrix} & \begin{matrix} A\\B\\C\\D\\E \end{matrix}\ \text{winner} \end{array}$$

8. a. rainbowmuffin defeated xthundercatx

b. rainbowmuffin and felipe.sanchez

c. rainbowmuffin

9. a.

$$\begin{array}{cc} \text{loser} & \\ \begin{matrix} A & B & C & D & E & F \end{matrix} & \\ \begin{bmatrix} 0&1&2&1&0&1\\1&0&1&3&1&2\\1&0&0&2&1&2\\1&0&1&0&1&0\\0&1&2&1&0&1\\2&1&1&1&2&0 \end{bmatrix} & \begin{matrix} A\\B\\C\\D\\E\\F \end{matrix}\ \text{winner} \end{array}$$

b. B, F, C, A/E (tied), D **c.** C, F, B/E (tied), A, D

10.

$$\begin{array}{cc} \text{loser} & \\ \begin{matrix} F & G & H & I & J \end{matrix} & \\ \begin{bmatrix} 0&1&1&0&1\\0&0&1&1&1\\0&0&0&1&1\\1&0&0&0&0\\0&0&0&1&0 \end{bmatrix} & \begin{matrix} F\\G\\H\\I\\J \end{matrix}\ \text{winner} \end{array}$$

Joining it all together

11. a.

$$\begin{array}{cc} \begin{matrix} H & J & N & W & Z \end{matrix} & \\ \begin{bmatrix} 2&1&1&1&1\\1&2&1&1&1\\1&1&2&1&1\\1&1&1&2&1\\1&1&1&1&2 \end{bmatrix} & \begin{matrix} H\\J\\N\\W\\Z \end{matrix} \end{array}$$

b. Yes. Jacob's row of the one and two-step communication matrix has no zeros.

c. Wortle

Exam practice

12. Explanation

Step 1: Determine f.

$c_{31} = 1$, meaning team C defeated team A.

This means $f = 0$, as team A did not defeat team C.

Step 2: Determine g.

$c_{23} = 0$, meaning team B lost to team C.

This means $g = 1$, as team C defeated team B.

Step 3: Determine h.

As team D cannot have played itself, $h = 0$.

Answer

A

13. a. Explanation

Step 1: Determine Alex's communication links.

The 1's in elements m_{12} and m_{15} indicate that Alex can send information to Brie and Elena.

However, looking at column A, Alex can only receive information from Chai.

Step 2: Determine Brie's communication links.

The 1's in elements m_{23} and m_{24} indicate that Brie can send information to Chai and Dex.

Looking at column B, Brie cannot receive information from Chai, but the 1 in element m_{42} indicates that Brie can receive information from Dex. This means they can send information directly to each other.

$$\begin{array}{ccc} & \text{receiver} & \\ & \begin{matrix} A & B & C & D & E \end{matrix} & \\ M = & \begin{bmatrix} 0&1&0&0&1\\0&0&1&1&0\\1&0&0&1&0\\0&1&0&0&0\\0&0&0&1&0 \end{bmatrix} & \begin{matrix} A\\B\\C\\D\\E \end{matrix}\ \text{sender} \end{array}$$

Answer

Brie and Dex

b. Explanation

Step 1: Determine the people that Elena can send information to.

Looking at row E, Elena can only send information to Dex. This will be the second person in the sequence.

$$\begin{array}{ccc} & \text{receiver} & \\ & \begin{matrix} A & B & C & D & E \end{matrix} & \\ M = & \begin{bmatrix} 0&1&0&0&1\\0&0&1&1&0\\1&0&0&1&0\\0&1&0&0&0\\0&0&0&1&0 \end{bmatrix} & \begin{matrix} A\\B\\C\\D\\E \end{matrix}\ \text{sender} \end{array}$$

Step 2: Determine the people that Dex can send information to.

Looking at row D, Dex can only send information to Brie. This will be the third person in the sequence.

$$\begin{array}{ccc} & \text{receiver} & \\ & \begin{matrix} A & B & C & D & E \end{matrix} & \\ M = & \begin{bmatrix} 0&1&0&0&1\\0&0&1&1&0\\1&0&0&1&0\\0&1&0&0&0\\0&0&0&1&0 \end{bmatrix} & \begin{matrix} A\\B\\C\\D\\E \end{matrix}\ \text{sender} \end{array}$$

Step 3: Determine the people that Brie can send information to.

Looking at row B, Brie can send information to Chai and Dex. As Elena needs to send documents to Chai, this will be the last person in the sequence.

$$\begin{array}{ccc} & \text{receiver} & \\ & \begin{matrix} A & B & C & D & E \end{matrix} & \\ M = & \begin{bmatrix} 0&1&0&0&1\\0&0&1&1&0\\1&0&0&1&0\\0&1&0&0&0\\0&0&0&1&0 \end{bmatrix} & \begin{matrix} A\\B\\C\\D\\E \end{matrix}\ \text{sender} \end{array}$$

Answer

Elena, Dex, Brie, Chai

c. **Explanation**

Step 1: Interpret the question.

The question asks for a pair of individuals (a sender and a receiver) that have 2 two-step communication links. This pair can be found by locating a '2' in matrix M^2.

Step 2: Identify the pair of individuals.

A '2' can be located in element $(m^2)_{14}$. The pair of individuals are Alex and Dex.

$$M^2 = \begin{array}{c} \text{receiver} \\ \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 0 & 0 & 1 & 2 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ \\ A \\ B \\ C \\ D \\ E \end{array} \quad \text{sender}$$

Step 3: Identify the two-step communication links.

This can be done by referring to the original matrix M. As Alex is the sender and Dex is the receiver, the third party in each two-step communication link will be able to receive information from Alex and send information to Dex.

$$M = \begin{array}{c} \text{receiver} \\ \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ \\ A \\ B \\ C \\ D \\ E \end{array} \quad \text{sender}$$

Brie and Elena can receive information from Alex

Brie and Elena can send information to Dex

The two individuals that Alex can send information through to Dex are Brie and Elena.

Answer

Alex, Brie, Dex

Alex, Elena, Dex

Many students failed to interpret this question correctly. It is possible they didn't realise that both the one and two-step communication matrices needed to be used to solve the question. It is also possible they thought the answer needed to be a pair of individuals instead of 2 two-step communication paths for a pair of individuals.

Questions from multiple lessons

14. D

15. E

16. a. \$1100 **b.** \$1213.66

7G Introduction to transition matrices

Interpreting and constructing state matrices

1. A

2. a. 57 people **b.** 198 people

c. 12%

3. a. $S_0 = \begin{bmatrix} 47\,270 \\ 39\,231 \\ 56\,159 \end{bmatrix} \begin{array}{c} S \\ H \\ A \end{array}$ **b.** $S_9 = \begin{bmatrix} 21\,452 \\ 23\,527 \\ 12\,339 \end{bmatrix} \begin{array}{c} S \\ H \\ A \end{array}$

Interpreting and constructing transition matrices

4. D

5. a. 9% **b.** 23% **c.** 91% **d.** 77%

6.

$$T = \begin{array}{c} \text{today} \\ \begin{array}{cc} B & N \end{array} \\ \begin{bmatrix} 0.35 & 0.50 \\ 0.65 & 0.50 \end{bmatrix} \end{array} \begin{array}{c} \\ \\ B \\ N \end{array} \quad \text{tomorrow}$$

Calculating state matrices

7. a. C **b.** $\begin{bmatrix} 12.3 \\ 17.7 \end{bmatrix}$ **c.** $\begin{bmatrix} 11.3 \\ 18.7 \end{bmatrix}$

8. a. A **b.** $\begin{bmatrix} 793.24 \\ 1264.76 \end{bmatrix}$ **c.** $\begin{bmatrix} 791.54 \\ 1266.46 \end{bmatrix}$

9. $\begin{bmatrix} 150 \\ 120 \\ 140 \end{bmatrix}$

10. a. 23 participants **b.** 27 participants

11. a. 162 customers

b. Yes, Tal will be profitable in 2063.

Using Leslie matrices in applied scenarios

12. a. D **b.** B

13. a. 68 females **b.** $\begin{bmatrix} 64 \\ 25 \end{bmatrix} \begin{array}{c} 0 \\ 1 \end{array}$ age

c. Yes; 24 females

14. 7229 chameleons

Joining it all together

15. a.

$$T = \begin{array}{c} \text{this week} \\ \begin{array}{cc} S & H \end{array} \\ \begin{bmatrix} 0.60 & 0.35 \\ 0.40 & 0.65 \end{bmatrix} \end{array} \begin{array}{c} \\ \\ S \\ H \end{array} \quad \text{next week}$$

b. 60% **c.** 35%

d. 210 families **e.** 146 families

16. a. 84 students **b.** 75 students

17. a. $S_0 = \begin{bmatrix} 14 \\ 23 \\ 13 \end{bmatrix} \begin{array}{c} E \\ M \\ H \end{array}$ **b.** 19 students

c. 4 teachers

18. a. $S_0 = \begin{bmatrix} 10 \\ 12 \\ 14 \\ 20 \end{bmatrix} \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array}$ age **b.** 2 females

c. 4 females **d.** 58 171 females

e. 460

Back to contents

Exam practice

19. Explanation

Complete the calculation provided.

$S_1 = T \times S_0$

$$S_1 = \begin{bmatrix} 0.65 & 0.25 & 0.25 & 0.50 \\ 0.15 & 0.60 & 0.20 & 0.15 \\ 0.05 & 0.10 & 0.25 & 0.20 \\ 0.15 & 0.05 & 0.30 & 0.15 \end{bmatrix} \times \begin{bmatrix} 520 \\ 320 \\ 80 \\ 80 \end{bmatrix}$$

$$S_1 = \begin{bmatrix} 478 \\ 298 \\ 94 \\ 130 \end{bmatrix}$$

Answer

$d = 298$; $e = 94$; $f = 130$

20. a. Explanation

Step 1: Analyse the transition matrix, T.

The elements on the diagonal of the transition matrix are the proportion of senior students that are expected to remain in the elective activity in the next term.

Step 2: Analyse the state matrix, S_1.

300 students chose communications in term 1.

200 students chose investigation in term 1.

200 students chose problem-solving in term 1.

200 students chose services in term 1.

Step 3: Multiply the number of students that chose each elective activity in term 1 by the proportion of students that are expected to remain in that elective activity.

$300 \times 0.4 + 200 \times 0.4 + 200 \times 0.3 + 300 \times 0.2$

$= 320$

Answer

320 senior students

b. Explanation

Substitute the known matrices into the equation $S_{n+1} = TS_n$.

S_2 is required, so substitute S_1 for S_n and S_2 for S_{n+1}.

$$S_2 = \begin{bmatrix} 0.4 & 0.2 & 0.3 & 0.1 \\ 0.2 & 0.4 & 0.1 & 0.3 \\ 0.2 & 0.3 & 0.3 & 0.4 \\ 0.2 & 0.1 & 0.3 & 0.2 \end{bmatrix} \times \begin{bmatrix} 300 \\ 200 \\ 200 \\ 300 \end{bmatrix}$$

Answer

$$S_2 = \begin{array}{c} \begin{bmatrix} 250 \\ 250 \\ 300 \\ 200 \end{bmatrix} \end{array} \begin{array}{l} \text{C} \\ \text{I} \\ \text{P} \\ \text{S} \end{array}$$

Some students assumed that the given S_1 matrix was S_0, and therefore calculated S_3, giving information on term 3.

c. Explanation

Step 1: Identify the number of senior students that choose services in term 2.

$$S_2 = \begin{bmatrix} 250 \\ 250 \\ 300 \\ 200 \end{bmatrix}$$

200 senior students choose services in term 2.

Step 2: Calculate the number of senior students that move from services to investigation in term 3.

The element in row 2 and column 4 represents the proportion of senior students that are in service that will move to investigation between any term. Therefore, 30% of students that choose services in term 2 are expected to choose investigation in term 3.

$200 \times 0.3 = 60$

Step 3: Determine the total number of students that choose investigation in term 3.

Calculate S_3.

$S_3 = TS_2$

$$S_3 = \begin{bmatrix} 0.4 & 0.2 & 0.3 & 0.1 \\ 0.2 & 0.4 & 0.1 & 0.3 \\ 0.2 & 0.3 & 0.3 & 0.4 \\ 0.2 & 0.1 & 0.3 & 0.2 \end{bmatrix} \begin{bmatrix} 250 \\ 250 \\ 300 \\ 200 \end{bmatrix}$$

$$S_3 = \begin{bmatrix} 260 \\ 240 \\ 295 \\ 205 \end{bmatrix}$$

240 senior students choose investigation in term 3.

Step 4: Find the percentage of students that choose investigation in term 3 that chose services in term 2.

$\frac{60}{240} = 0.25 = 25\%$

Answer

25%

Some students were able to find either the number of senior students that choose investigation in term 3 or the number of senior students that move from services to investigation in term 3, yet were unable to relate them to each other.

Questions from multiple lessons

21. E **22.** C

23. a. 65 tops

b. 15 pairs of pants were sold on Wednesday.

c. $6.00

d. $M = [1 \quad 1 \quad 0]$

7H The equilibrium state matrix

Calculating the equilibrium state matrix

1. D

2. a. $T^{30} \times S_0$

Answers may vary, but should use a high value for n.

b. $S_{30} = \begin{bmatrix} 57.71 \\ 82.29 \end{bmatrix}$

$= S_{31}$

Answers may vary.

3. a. $\begin{bmatrix} 77.14 \\ 42.86 \end{bmatrix}$ **b.** $\begin{bmatrix} 220.35 \\ 405.65 \end{bmatrix}$

c. $\begin{bmatrix} 204.71 \\ 109.18 \\ 34.12 \end{bmatrix}$

Back to contents

4. $\begin{array}{c|l} \begin{bmatrix} 69 \\ 31 \end{bmatrix} & \begin{array}{l} \text{eat} \\ \text{stay} \end{array} \end{array}$

Interpreting the equilibrium state matrix

5. B

6. Percussion

7. In the long term, 966 penguins will be living in region A and 322 will be living in region B.

Joining it all together

8. **a.** $\begin{bmatrix} 0.8 & 0.5 \\ 0.2 & 0.5 \end{bmatrix}^{30} \times \begin{bmatrix} 53 \\ 7 \end{bmatrix}$

b. $\begin{bmatrix} 43 \\ 17 \end{bmatrix} \begin{array}{l} \text{on time} \\ \text{late} \end{array}$

c. 43 students

9. 20 drivers

10. Yes, from Chel island.

11. 25 towns

Exam practice

12. Explanation

Step 1: Calculate the state matrix for a large n.

$S_n = T^n \times S_0$

Usually the equilibrium state matrix will have occurred by the time $n = 30$.

$S_{30} = T^{30} \times S_0$

Note: As long as the sum of the elements in the initial state matrix is 500, any 5×1 initial state matrix can be used, as the equilibrium state will be the same regardless.

$$= \begin{bmatrix} 0.1 & 0.2 & 0.2 & 0 & 0 \\ 0.5 & 0.2 & 0.3 & 0.1 & 0 \\ 0.3 & 0.3 & 0.4 & 0.1 & 0.2 \\ 0.1 & 0.2 & 0.1 & 0.6 & 0.3 \\ 0 & 0.1 & 0 & 0.2 & 0.5 \end{bmatrix}^{30} \times \begin{bmatrix} 100 \\ 100 \\ 100 \\ 100 \\ 100 \end{bmatrix}$$

$$= \begin{bmatrix} 48.96... \\ 96.07... \\ 124.24... \\ 151.07... \\ 79.64... \end{bmatrix} \begin{array}{l} J \\ K \\ L \\ M \\ N \end{array}$$

Step 2: Verify the equilibrium state matrix by comparing with S_{n+1}.

$S_{31} = T^{31} \times S_0$

$$= \begin{bmatrix} 0.1 & 0.2 & 0.2 & 0 & 0 \\ 0.5 & 0.2 & 0.3 & 0.1 & 0 \\ 0.3 & 0.3 & 0.4 & 0.1 & 0.2 \\ 0.1 & 0.2 & 0.1 & 0.6 & 0.3 \\ 0 & 0.1 & 0 & 0.2 & 0.5 \end{bmatrix}^{31} \times \begin{bmatrix} 100 \\ 100 \\ 100 \\ 100 \\ 100 \end{bmatrix}$$

$$= \begin{bmatrix} 48.96... \\ 96.07... \\ 124.24... \\ 151.07... \\ 79.64... \end{bmatrix} \begin{array}{l} J \\ K \\ L \\ M \\ N \end{array}$$

Since S_{30} and S_{31} are equal, this is the equilibrium state matrix.

Step 3: Interpret the equilibrium state matrix.

The value 96.07 corresponds to category K.

Answer

B

13. Explanation

Step 1: Calculate the state matrix for a large n.

$S_n = T^n \times S_0$

Usually the equilibrium state matrix will have occurred by the time $n = 30$.

$S_{30} = T^{30} \times S_0$

$$= \begin{bmatrix} 0.2 & 0.1 & 0.0 & 0.2 \\ 0.1 & 0.1 & 0.0 & 0.2 \\ 0.2 & 0.1 & 0.2 & 0.1 \\ 0.5 & 0.7 & 0.8 & 0.5 \end{bmatrix}^{30} \times \begin{bmatrix} 700 \\ 400 \\ 200 \\ 1400 \end{bmatrix}$$

$$= \begin{bmatrix} 431.56... \\ 388.41... \\ 347.95... \\ 1532.06... \end{bmatrix} \begin{array}{l} G \\ R \\ S \\ N \end{array}$$

Step 2: Verify the equilibrium state matrix by comparing with S_{n+1}.

$S_{31} = T^{31} \times S_0$

$$= \begin{bmatrix} 0.2 & 0.1 & 0.0 & 0.2 \\ 0.1 & 0.1 & 0.0 & 0.2 \\ 0.2 & 0.1 & 0.2 & 0.1 \\ 0.5 & 0.7 & 0.8 & 0.5 \end{bmatrix}^{31} \times \begin{bmatrix} 700 \\ 400 \\ 200 \\ 1400 \end{bmatrix}$$

$$= \begin{bmatrix} 431.56... \\ 388.41... \\ 347.95... \\ 1532.06... \end{bmatrix} \begin{array}{l} G \\ R \\ S \\ N \end{array}$$

Since S_{30} and S_{31} are equal, this is the equilibrium state matrix.

Step 3: Calculate the percentage of highway falling under category N.

$$\frac{1532.06...}{2700} = 0.5674... = 56.74...\%$$

Answer

56.7%

Many students who correctly obtained a long term value for N left the answer as 1532.1, instead of calculating the percentage of the total road.

14. Explanation

Step 1: Calculate the state matrix for a large n.

$S_n = T^n \times S_0$

In this case, S_{30} and S_{31} differ significantly, so a higher value of n must be used.

$S_{60} = T^{60} \times S_0$

$$= \begin{bmatrix} 0.80 & 0.09 & 0.10 \\ 0.12 & 0.79 & 0.10 \\ 0.08 & 0.12 & 0.80 \end{bmatrix}^{60} \times \begin{bmatrix} 250\,000 \\ 230\,000 \\ 200\,000 \end{bmatrix}$$

$$= \begin{bmatrix} 218\,884.12... \\ 233\,476.39... \\ 227\,639.48... \end{bmatrix} \begin{array}{l} W \\ G \\ E \end{array}$$

Step 2: Verify the equilibrium state matrix by comparing with S_{n+1}.

$S_{61} = T^{61} \times S_0$

$$= \begin{bmatrix} 0.80 & 0.09 & 0.10 \\ 0.12 & 0.79 & 0.10 \\ 0.08 & 0.12 & 0.80 \end{bmatrix}^{61} \times \begin{bmatrix} 250\,000 \\ 230\,000 \\ 200\,000 \end{bmatrix}$$

$$= \begin{bmatrix} 218\,884.12... \\ 233\,476.39... \\ 227\,639.48... \end{bmatrix} \begin{array}{l} W \\ G \\ E \end{array}$$

Back to contents

Since S_{60} and S_{61} are equal, this is the equilibrium state matrix.

Step 3: Interpret the equilibrium state matrix.

The expected long term weekly number of shoppers at Westmall is given by the element in the top row.

Answer

218 884

Some students tried to read the value off the graph, giving an answer of approximately 220 000. As the question specifies to round to the nearest whole number, the provided transition matrix and initial state matrix must be used in order to obtain a more specific value.

Questions from multiple lessons

15. C

16. E

17. a. Voula and Yasmin

b. Xavier and Zoe

7I Applications of transition matrices

Constructing and interpreting transition diagrams

1. C

2. a.

$$\begin{array}{c} \text{from} \\ \begin{matrix} A & B \end{matrix} \\ \begin{bmatrix} 0.23 & 0.38 \\ 0.77 & 0.62 \end{bmatrix} \end{array} \begin{matrix} A \\ B \end{matrix} \ \text{to}$$

b.

$$\begin{array}{c} \text{from} \\ \begin{matrix} X & Y \end{matrix} \\ \begin{bmatrix} 0.93 & 0.25 \\ 0.07 & 0.75 \end{bmatrix} \end{array} \begin{matrix} X \\ Y \end{matrix} \ \text{to}$$

c.

$$\begin{array}{c} \text{from} \\ \begin{matrix} A & B \end{matrix} \\ \begin{bmatrix} 0.87 & 0 \\ 0.13 & 1 \end{bmatrix} \end{array} \begin{matrix} A \\ B \end{matrix} \ \text{to}$$

3. a.

$$\begin{array}{c} \text{from} \\ \begin{matrix} C & D & E \end{matrix} \\ \begin{bmatrix} 0.81 & 0.26 & 0.39 \\ 0.09 & 0.5 & 0.04 \\ 0.1 & 0.24 & 0.57 \end{bmatrix} \end{array} \begin{matrix} C \\ D \\ E \end{matrix} \ \text{to}$$

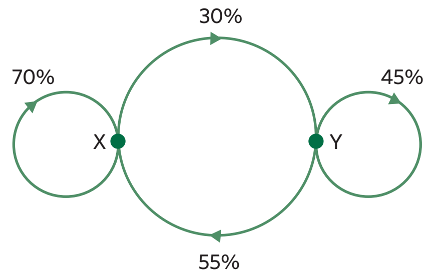
b.

$$\begin{array}{c} \text{from} \\ \begin{matrix} X & Y & Z \end{matrix} \\ \begin{bmatrix} 0.58 & 0.7 & 0 \\ 0.42 & 0 & 0.03 \\ 0 & 0.3 & 0.97 \end{bmatrix} \end{array} \begin{matrix} X \\ Y \\ Z \end{matrix} \ \text{to}$$

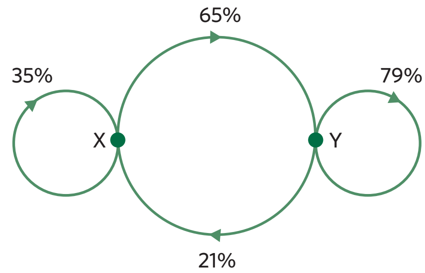
c.

$$\begin{array}{c} \text{from} \\ \begin{matrix} A & B & C & D \end{matrix} \\ \begin{bmatrix} 0.61 & 0.25 & 0 & 0.22 \\ 0.3 & 0.63 & 0 & 0 \\ 0.09 & 0 & 0.86 & 0 \\ 0 & 0.12 & 0.14 & 0.78 \end{bmatrix} \end{array} \begin{matrix} A \\ B \\ C \\ D \end{matrix} \ \text{to}$$

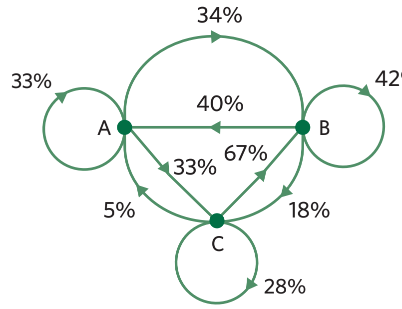
4. a.



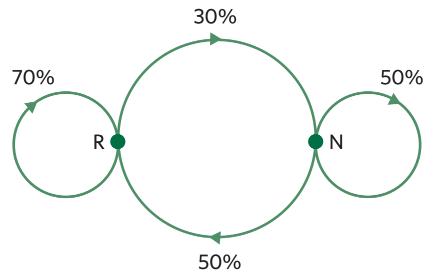
b.



c.



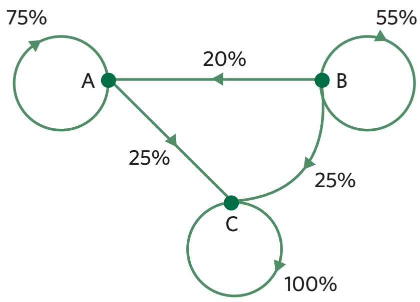
5. a.



b. 50%

c. 76 members

6. a.



b.

$$\begin{array}{c} \text{from} \\ \begin{matrix} A & B & C \end{matrix} \\ \begin{bmatrix} 0.75 & 0.2 & 0 \\ 0 & 0.55 & 0 \\ 0.25 & 0.25 & 1 \end{bmatrix} \end{array} \begin{matrix} A \\ B \\ C \end{matrix} \ \text{to}$$

c. 30 rocks

d. 12 rocks

e. 180 rocks

Back to contents

Using transition matrices to model situations involving culling and restocking

7. C

8. **a.** $\begin{bmatrix} 72.8 \\ 47.2 \end{bmatrix}$ **b.** $\begin{bmatrix} 36.3 \\ 41.7 \end{bmatrix}$ **c.** $\begin{bmatrix} 88.6 \\ 28.4 \end{bmatrix}$

9. **a.** B **b.** $\begin{bmatrix} -70 \\ 70 \end{bmatrix}$

10. **a.** $\begin{bmatrix} 0.93 & 0.12 \\ 0.07 & 0.88 \end{bmatrix}$ **b.** $\begin{bmatrix} -32\,000 \\ 32\,000 \end{bmatrix}$

c. 84 710 people

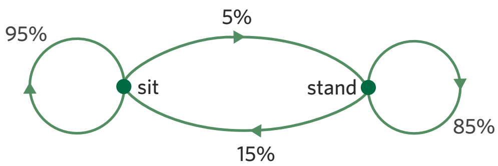
11. **a.** $\begin{bmatrix} 4 \\ 4 \\ 7 \\ 5 \end{bmatrix}$ **b.** $\begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}$ **c.** $\begin{bmatrix} 4 \\ 5 \\ 6 \\ 5 \end{bmatrix}$ **d.** $\begin{bmatrix} 4 \\ 4 \\ 4 \\ 4 \end{bmatrix}$

Joining it all together

12. **a.** The people getting on and off the bus at each stop.

b. $B_3 = \begin{array}{c} \begin{bmatrix} 9 \\ 18 \end{bmatrix} \end{array} \begin{array}{l} \text{stand} \\ \text{sit} \end{array}$

c.



13. **a.** 9 players

b. $T = \begin{array}{c} \begin{array}{cc} \text{on} & \text{off} \end{array} \\ \begin{bmatrix} 0.40 & 0.75 \\ 0.60 & 0.25 \end{bmatrix} \end{array} \begin{array}{l} \\ \text{on} \\ \text{off} \end{array}$ $\quad P_0 = \begin{bmatrix} 5 \\ 4 \end{bmatrix} \begin{array}{l} \text{on} \\ \text{off} \end{array}$

c. $P_0 = \begin{bmatrix} 5 \\ 4 \end{bmatrix} \begin{array}{l} \text{on} \\ \text{off} \end{array}$ $\quad P_{n+1} = \begin{bmatrix} 0.40 & 0.75 \\ 0.60 & 0.25 \end{bmatrix} \times P_n$

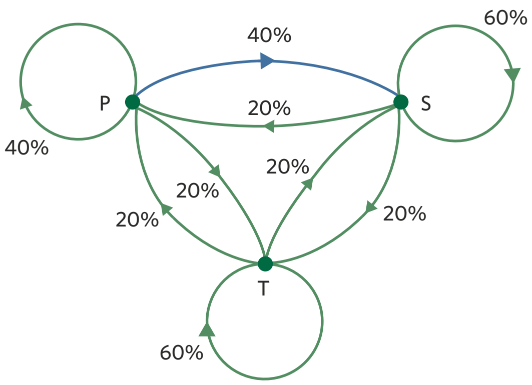
d. They are identical. **e.** 58 people

f. 23 people

Exam practice

14. **a.** **Explanation**

Determine the percentage moving from P to S on the transition diagram.



Answer

40%

b. **Explanation**

Step 1: Determine the percentage of each state that transitions into S.

40% of state P transitions into S.

60% of state S transitions into itself.

20% of state T transitions into S.

Step 2: Convert the percentages to decimals and fill in the boxes.

Answer

$300 \times 0.4 + 240 \times 0.6 + 210 \times 0.2 = 306$

15. **Explanation**

Step 1: Determine F_1, the fish population after 1 month.

$$F_1 = \begin{bmatrix} 0.65 & 0 & 0 & 0 \\ 0.25 & 0.75 & 0 & 0 \\ 0 & 0.20 & 0.95 & 0 \\ 0.10 & 0.05 & 0.05 & 1 \end{bmatrix} F_0 + B$$

$$= \begin{bmatrix} 0.65 & 0 & 0 & 0 \\ 0.25 & 0.75 & 0 & 0 \\ 0 & 0.20 & 0.95 & 0 \\ 0.10 & 0.05 & 0.05 & 1 \end{bmatrix} \times \begin{bmatrix} 50\,000 \\ 10\,000 \\ 7\,000 \\ 0 \end{bmatrix} + B$$

$$= \begin{bmatrix} 32\,500 \\ 20\,000 \\ 8\,650 \\ 5\,850 \end{bmatrix} + B$$

Step 2: Determine matrix B.

Since the population remains constant, $F_1 = F_0$.

$$\begin{bmatrix} 50\,000 \\ 10\,000 \\ 7\,000 \\ 0 \end{bmatrix} = \begin{bmatrix} 32\,500 \\ 20\,000 \\ 8\,650 \\ 5\,850 \end{bmatrix} + B$$

$$B = \begin{bmatrix} 50\,000 \\ 10\,000 \\ 7\,000 \\ 0 \end{bmatrix} + \begin{bmatrix} 32\,500 \\ 20\,000 \\ 8\,650 \\ 5\,850 \end{bmatrix}$$

$$= \begin{bmatrix} 17\,500 \\ -10\,000 \\ -1\,650 \\ -5\,850 \end{bmatrix} \begin{array}{l} \text{Y} \\ \text{J} \\ \text{A} \\ \text{D} \end{array}$$

Step 3: Check whether each option is correct or incorrect.

Note: A positive number in matrix B indicates fish being bought, and a negative number indicates fish being sold.

A: This is correct as $-1650 = -1650$. ✓

B: This is incorrect as $1750 \neq -1650$. ✗

C: This is incorrect as $-17\,500 \neq 17\,500$. ✗

D: This is incorrect as $50\,000 \neq 17\,500$. ✗

E: This is incorrect as $10\,000 \neq -10\,000$. ✗

Answer

A

> Many students did not consider the fact that the population of each type of fish remained constant, which was the key to solving the question.

16. **Explanation**

Step 1: Calculate s, the number of planes always parked at Sydney airport.

12 of Melbourne airport's planes on Tuesday had been parked at Sydney airport on Monday. This means that there are 12 planes transitioning from Sydney to Melbourne. This fact can be used to calculate s.

Back to contents

$$12 = 0.48 \times s$$
$$s = \frac{12}{0.48}$$
$$= 25$$

Step 2: Calculate m, the number of planes always parked at Melbourne airport.

Since the number of planes parked at each airport remains constant, the number of planes moving from Melbourne to Sydney must be the same as the number of planes moving from Sydney to Melbourne.

$$12 = 0.2 \times m$$
$$m = \frac{12}{0.2}$$
$$= 60$$

Step 3: Calculate the total number of planes that the airline has.

$$s + m = 25 + 60$$
$$= 85$$

Answer

E

27% of students chose option A, which was the correct value for s, however the question required further calculations to also find m. This question proved very difficult, as it required students to work backwards multiple steps from the information given.

Questions from multiple lessons

17. D

18. D

19. a. $\begin{bmatrix} 51.15 \\ 1423.97 \\ 126.68 \end{bmatrix}$

b. $\begin{bmatrix} 1.003 & 0 & 0 \\ 0 & 1.0035 & 0 \\ 0 & 0 & 0.9975 \end{bmatrix}$

Back to contents

8A Introduction to graphs and networks

Identifying properties of a graph

1. D

2. a. 5 b. 2 c. 2

3. a. 6 b. 3 c. 3 d. 5
   e. 6 f. No

4. a. 8 b. 2 c. 4
   d. There are direct roads between Croatia and three other countries.
   e. 6
   f. Yes, Kosovo and Macedonia.

Identifying and constructing graphs

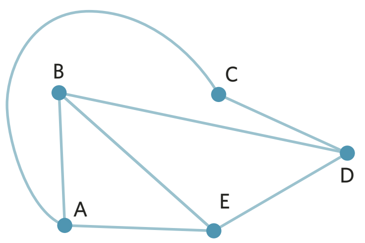
5. C

6. A

7. There are no edges and all vertices are isolated. Therefore, it is a degenerate graph.

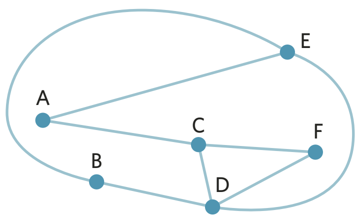
8. Planar graph
   Simple graph
   Complete graph
   Connected graph

9. a.



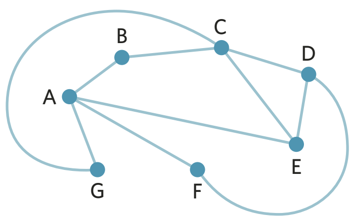
Answers may vary.

b.



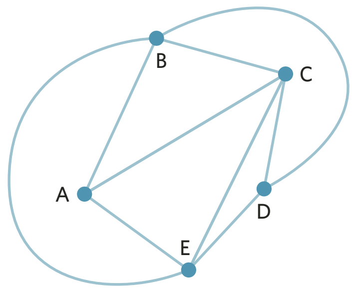
Answers may vary.

c.



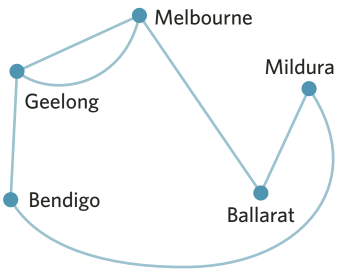
Answers may vary.

d.



Answers may vary.

10.



Answers may vary.

Using Euler's rule

11. C

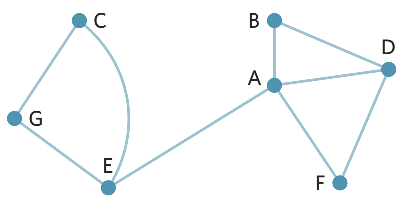
12. a. $v = 3$ b. $e = 8$ c. $f = 4$ d. $e = 11$
    e. $f = 5$ f. $v = 9$

13. 5 areas

14. a. $4 - 6 + 4 = 2$ b. $5 - 8 + 5 = 2$
    c. $5 - 8 + 5 = 2$ d. $7 - 10 + 5 = 2$

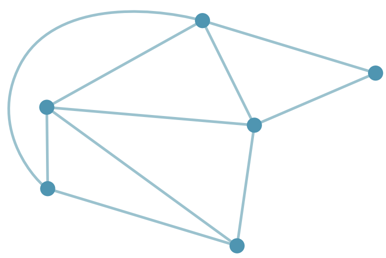
Joining it all together

15. a. A and E
    b.



Answers may vary.

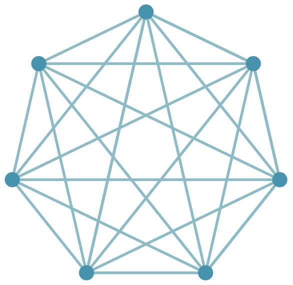
16. a. Planar graph
    Simple graph
    Connected graph
    b.



Answers may vary.

c. 6 d. 20
e. 2 vertices

17. a.



b. Simple graph
Complete graph
Connected graph

c. 6

Exam practice

18. a. Explanation

In the floor plan, there is a door between room A and room B. However, this edge is missing from the graph.

Answer

A and B

b. Explanation

There are 2 edges connected to vertex E.

Answer

2

19. Explanation

Dale and Cameron have edges connected to both Alex and Bo.

Answer

Dale and Cameron

> Many students gave one of the two names but not both.

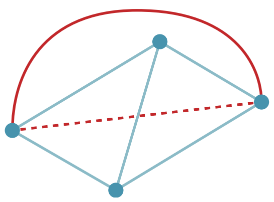
20. Explanation

Step 1: Identify any overlapping edges.

There are two overlapping edges in the middle of the graph.

Step 2: Remove an overlapping edge and redraw it.

One of the overlapping edges can be redrawn on the outside so it is not crossing any other edges.



Step 3: Count the number of faces.

When the graph is redrawn as planar, there are four faces. This also includes the space outside the graph.

Answer

C

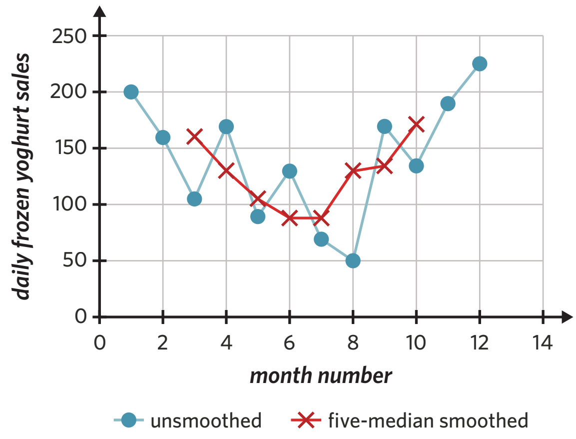
> 61% of students incorrectly selected D. These students did not redraw the graph in its planar form before counting the faces.

Questions from multiple lessons

21. D

22. A

23. a.



b. Mean of June and July: $\frac{139 + 81}{2} = 110$

Mean of July and August: $\frac{81 + 126}{2} = 103.5$

Smoothed value of July: $\frac{103.5 + 110}{2} = 106.75$

8B Graphs, networks and matrices

Constructing an adjacency matrix from a graph

1. C

2. a.

$$\begin{array}{cc} \begin{matrix} A & B & C & D \end{matrix} & \\ \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} & \begin{matrix} A \\ B \\ C \\ D \end{matrix} \end{array}$$

b.

$$\begin{array}{cc} \begin{matrix} A & B & C & D \end{matrix} & \\ \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} & \begin{matrix} A \\ B \\ C \\ D \end{matrix} \end{array}$$

3. a.

$$\begin{array}{cc} \begin{matrix} A & B & C & D & E \end{matrix} & \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix} & \begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix} \end{array}$$

b.

$$\begin{array}{cc} \begin{matrix} A & B & C & D & E \end{matrix} & \\ \begin{bmatrix} 0 & 2 & 0 & 0 & 0 \\ 2 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix} & \begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix} \end{array}$$

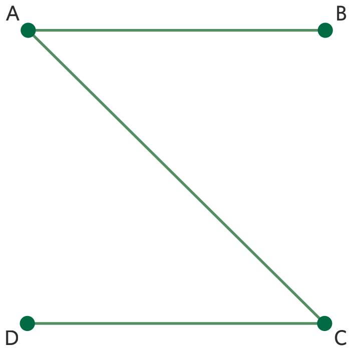
4.

$$\begin{array}{cc} \begin{matrix} A & B & C & D \end{matrix} & \\ \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} & \begin{matrix} A \\ B \\ C \\ D \end{matrix} \end{array}$$

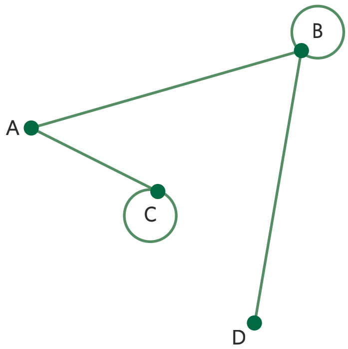
Back to contents

Constructing a graph from an adjacency matrix

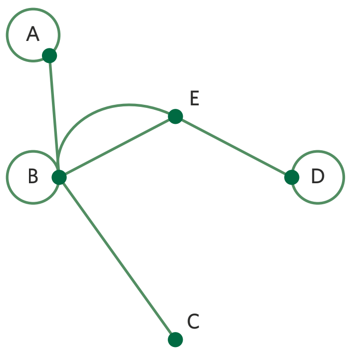
5. B

6. a. 

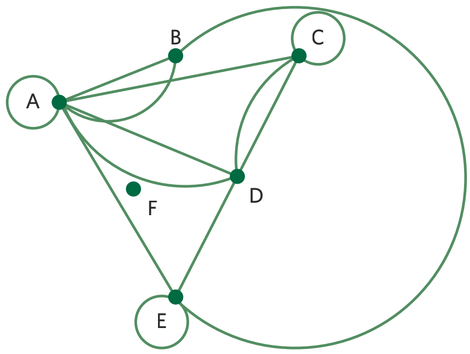
Answers may vary.

b. 

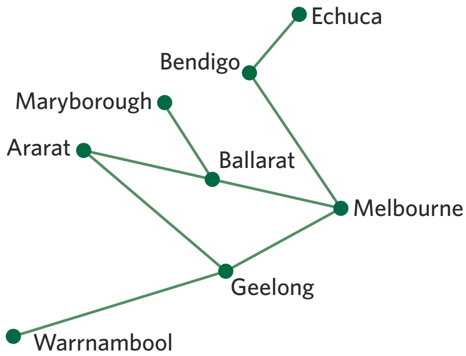
Answers may vary.

7. a. 

Answers may vary.

b. 

Answers may vary.

8. 

Answers may vary.

Representing directed graphs

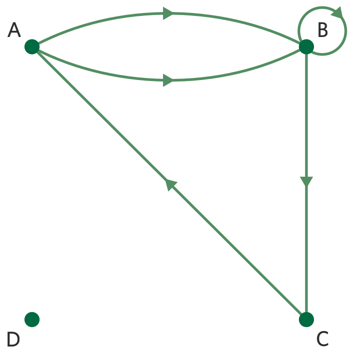
9. C

10. a.

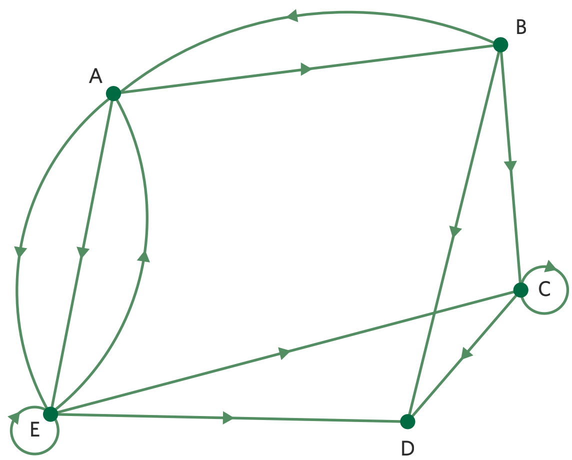
$$\begin{array}{c c} \text{to} & \\ \begin{array}{cccc} A & B & C & D \end{array} & \\ \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} & \begin{array}{c} A \\ B \\ C \\ D \end{array} \text{ from} \end{array}$$

b.

$$\begin{array}{c c} \text{to} & \\ \begin{array}{ccccc} A & B & C & D & E \end{array} & \\ \begin{bmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix} & \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} \text{ from} \end{array}$$

11. a. 

Answers may vary.

b. 

Answers may vary.

8B ANSWERS

Back to contents

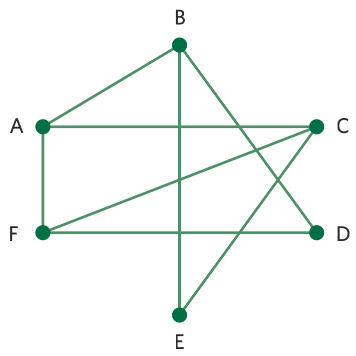
12.

to

$$\begin{array}{cc} \begin{matrix} A & B & C & D & E & F \end{matrix} & \\ \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} & \begin{matrix} A \\ B \\ C \\ D \\ E \\ F \end{matrix} \end{array} \text{ from}$$

Joining it all together

13. a.



Answers may vary.

b.

$$\begin{array}{cc} \begin{matrix} A & B & C & D & E & F \end{matrix} & \\ \begin{bmatrix} 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \end{bmatrix} & \begin{matrix} A \\ B \\ C \\ D \\ E \\ F \end{matrix} \end{array}$$

c.

to

$$\begin{array}{cc} \begin{matrix} A & B & C & D & E & F \end{matrix} & \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} & \begin{matrix} A \\ B \\ C \\ D \\ E \\ F \end{matrix} \end{array} \text{ from}$$

Exam practice

14. Explanation

Step 1: Determine the edges connecting to vertex J.

$m_{JJ} = 1$. There should be a loop at vertex J.

$m_{JK} = 3$. There should be three pathways between vertices J and K.

$m_{JL} = 0$. There should be no edges between vertices J and L.

$m_{JM} = 2$. There should be two edges between vertices J and M.

Step 2: Determine which network contains the information in the adjacency matrix.

Only options A, D and E have a loop at vertex J.

Of these, only option E has three pathways between J and K.

Answer

E

15. Explanation

Step 1: Determine the order of the matrix.

All adjacency matrices are square.

There are five vertices.

The order will be 5 × 5.

Step 2: Label the matrix.

Each vertex will have its own row and column.

$$\begin{array}{cc} \begin{matrix} P & Q & R & S & T \end{matrix} & \\ \begin{bmatrix} & & & & \\ \\ \\ \\ \\ \end{bmatrix} & \begin{matrix} P \\ Q \\ R \\ S \\ T \end{matrix} \end{array}$$

Step 3: Fill in row P of the matrix.

There is one connection from vertex P back to vertex P. Element m_{PP} is 1.

There are three connections between vertex P and vertex Q. Element m_{PQ} is 3.

There are no connections between vertex P and vertex R. Element m_{PR} is 0.

There are two connections between vertex P and vertex S. Element m_{PS} is 2.

There are two connections between vertex P and vertex T. Element m_{PT} is 2.

Step 4: Repeat for the remaining rows.

$$\begin{array}{cc} \begin{matrix} P & Q & R & S & T \end{matrix} & \\ \begin{bmatrix} 1 & 3 & 0 & 2 & 2 \\ 3 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 2 & 1 & 1 & 0 & 2 \\ 2 & 1 & 0 & 2 & 0 \end{bmatrix} & \begin{matrix} P \\ Q \\ R \\ S \\ T \end{matrix} \end{array}$$

Answer

A

31% of students incorrectly chose option B. There were three different ways of travelling directly between towns P and Q which is not represented by the matrix in option B. Another common mistake was not realising that since it was possible to return to town P without passing through another town, this loop needed to be shown as a 1 in the first row and first column of the matrix.

16. Explanation

Step 1: Determine which buildings vertex J is not connected to.

Vertex J does not have a loop.

Vertex J is not connected to vertex M.

Therefore row J in the adjacency matrix will have 2 zeros.

Step 2: Determine which buildings vertex K is not connected to.

Vertex K has a loop and is not connected to vertex N.

Therefore row K in the adjacency matrix will have 1 zero.

Step 3: Determine which buildings vertex L is not connected to.

Vertex L does not have a loop and is connected to all other vertices.

Therefore row L in the adjacency matrix will have 1 zero.

Back to contents

Step 4: Determine which buildings vertex M is not connected to.

Vertex M does not have a loop.

Vertex M is not connected to vertices J and N.

Therefore row M in the adjacency matrix will have 3 zeros.

Step 5: Determine which buildings vertex N is not connected to.

Vertex N does not have a loop.

Vertex N is not connected to vertices K and M.

Therefore row N in the adjacency matrix will have 3 zeros.

Step 6: Determine the total number of zeros in the adjacency matrix.

$2 + 1 + 1 + 3 + 3 = 10$ zeros

Answer

C

A zero in an adjacency matrix means there is no direct connection. If there is no loop, this also means there is no connection since the vertex does not connect to itself. Students who incorrectly chose D or E might have overlooked that there was a loop at building K.

Questions from multiple lessons

17. A

18. C

19. **a.** Louis and Natalia

b. **i.** N and O

ii. Ollie can travel from his house and return without passing any of the other houses.

8C Exploring and travelling problems

Identifying types of walks

1. B

2. **a.** Circuit **b.** Trail **c.** Path **d.** Walk

e. Trail **f.** Cycle

3. Cycle

Identifying Eulerian trails and circuits

4. B

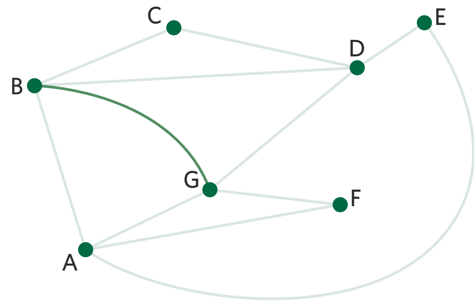
5. **a.** Eulerian circuit
A–B–C–E–D–C–A
Answers may vary.

b. Eulerian trail
A–C–E–F–B–A–B–D–E
Answers may vary.

c. Eulerian trail
A–B–F–C–A–D–E–G–F–E–C
Answers may vary.

d. Eulerian circuit
A–F–E–C–D–B–C–A–B–A
Answers may vary.

6. **a.**



b. A–B–C–D–E–A–F–G–D–B–G–A
Answers may vary.

7. **a.** Ramp M

b. K–N–M–H–G–I–K–J–I–H–J–L–M
Answers may vary.

c. Ramp K and M

Identifying Hamiltonian paths and cycles

8. C

9. **a.** Yes, A–B–D–C–F–E
Answers may vary.

b. No

c. Yes, F–H–G–E–C–B–D–A
Answers may vary.

10. **a.** Yes, C–E–D–B–A–C
Answers may vary.

b. No

c. No

11. **a.** Hamiltonian cycle

b. A–B–C–D–E
Answers may vary.

Joining it all together

12. **a.** III **b.** II and V

c. IV and V **d.** I

13. **a.** Trail **b.** Khloe **c.** Rob

d. Hamiltonian cycle
Kim–Kourtney–Khloe–Kendall–Rob–Kylie–Kim

Exam practice

14. Explanation

Step 1: Determine where the walk starts and ends.

The tour starts and finishes at the office.

Since it starts and ends at the same vertex, it is either a cycle or a circuit.

Step 2: Determine whether any edges or vertices are repeated.

The tour will visit each building only once. There are no repeated edges or vertices as a result.

Since there are no repeated edges and vertices, it is a cycle.

Step 3: Determine if it meets the definition of a Hamiltonian cycle.

A Hamiltonian cycle is a cycle that includes every vertex.

The school tour will visit each building.

Back to contents

Answer

Hamiltonian cycle

Some students incorrectly named the route as a circuit or path. The route was a cycle instead of a circuit as it did not have repeated vertices. While a path does not have repeated vertices, students needed to recognise that by starting and ending at the same vertex, it would be a cycle. It was also important to check, after determining that the route was a cycle, whether or not it was a Hamiltonian cycle.

15. Explanation

Step 1: Count the degrees of all of the vertices.

There are two vertices of degree 2.

There are three vertices of degree 3.

There is one vertex of degree 5.

Step 2: Identify what the degree of each vertex should be for an Eulerian circuit to be possible.

For an Eulerian circuit to be possible, all vertices must be of an even degree.

Step 3: Determine how many extra edges are needed.

There are four vertices of odd degree.

If an extra edge is added between two of these, the degree of each would increase by 1 so that both are of an even degree. Therefore, if two extra edges are added, each between each two vertices of an odd degree, an Eulerian circuit would be possible.

Answer

C

11% of students incorrectly chose option B. If only one extra edge was added, there would still be two vertices of an odd degree. This would make an Eulerian trail possible, but not an Eulerian circuit.

16. Explanation

Step 1: Determine the starting vertex.

The postal worker will start from vertex F.

Step 2: Follow the edges from the starting vertex to create a Hamiltonian path.

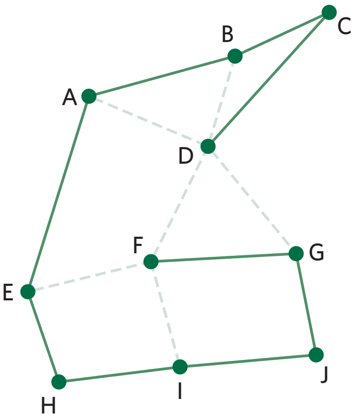
The postal worker can start by travelling to either vertex D, E, I or G.

She must then travel to every other vertex exactly once without repeating any edges.

She can end at any house.

Answer

F–G–J–I–H–E–A–C–D



Answers may vary.

Students needed to ensure that the route had 9 edges and included all other vertices. Some students' paths returned to F, creating a Hamiltonian cycle, which was incorrect.

17. Explanation

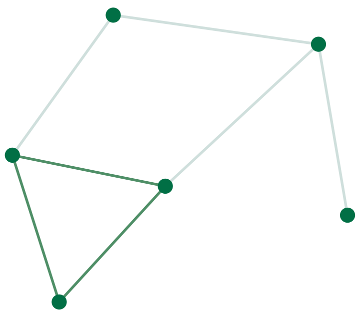
Determine whether each statement is true or false.

Statement 1:

This statement is true. A planar graph has no overlapping edges.

Statement 2:

This statement is true. A cycle must start and end at the same vertex and contain no repeated edges or vertices. A possible cycle is shown.



Statement 3:

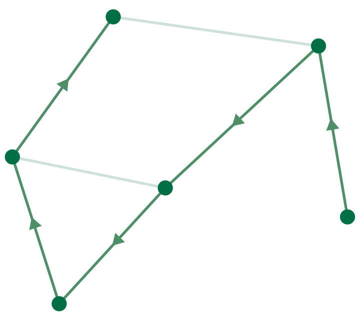
This statement is true. A bridge is an edge that when removed prevents every vertex from being reached.

Statement 4:

This statement is false. In order for an Eulerian trail to be possible, there must be exactly two vertices of an odd degree. This graph has four vertices of an odd degree.

Statement 5:

This statement is true. It is possible to reach every vertex without repeating any edges or vertices by starting and ending at a different vertex. A possible Hamiltonian path is shown.



Answer

D

Back to contents

Questions from multiple lessons

18. D

19. E

20. **a.** 4.56%

b. **i.** $2080

ii. $38 340.70

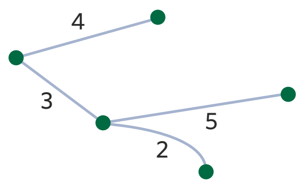
8D Minimum connector problems

Identifying and finding the weight of a spanning tree

1. C

2. **a.** 5 vertices, 4 edges

b.



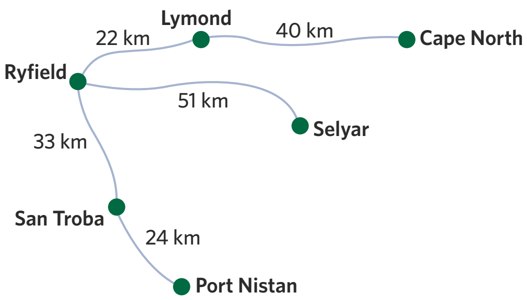
Note: There are multiple possible options.

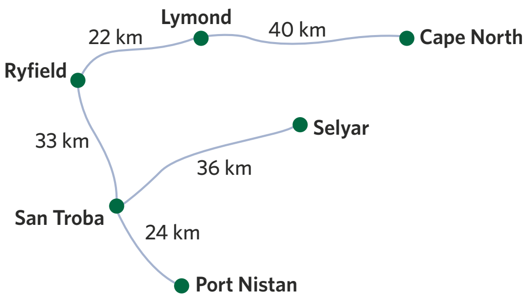
c. 14

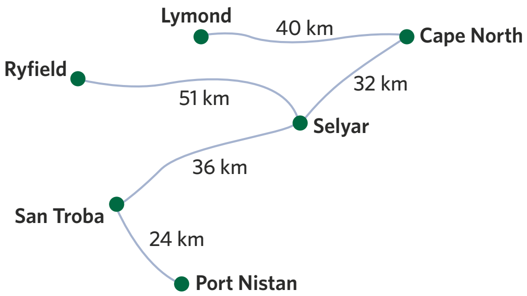
Note: There are multiple possible options.

3. 6

4.





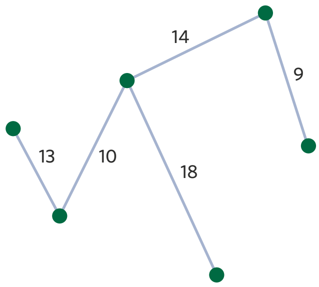


Note: There are multiple possible options.

Finding the minimum spanning tree

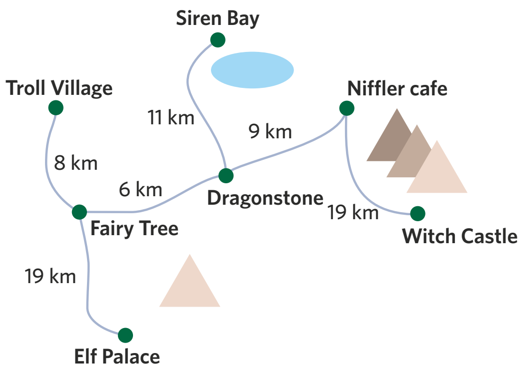
5. D

6.



7. **a.** 10 **b.** 110 **c.** 35

8. **a.**

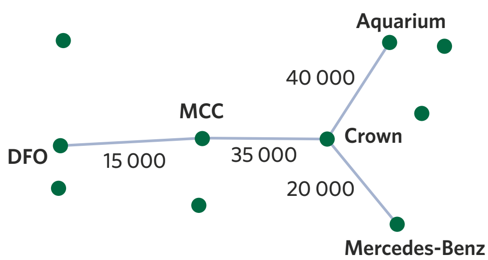


b. 72 km

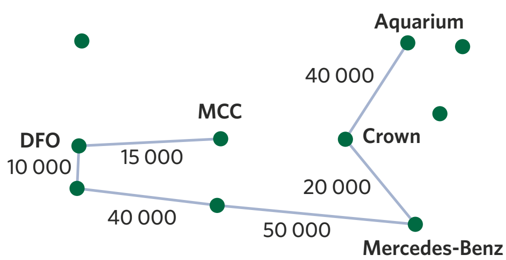
Joining it all together

9. **a.** 5 **b.** C

10. **a.**



b.



c. $180 000

Exam practice

11. **Explanation**

Determine the number of vertices and edges required.

A spanning tree connects all vertices. Therefore the spanning tree will use all 5 vertices.

$e = v - 1$

$= 5 - 1$

$= 4$

8D ANSWERS

Back to contents

Answer

B

12. Explanation

To solve this question, check whether each option is a spanning tree.

A: This is a spanning tree. There are no loops, duplicate edges or cycles and all vertices are connected. ✗

B: This is not a spanning tree. Although there are no loops, duplicate edges or cycles, the graph has an edge connecting vertex 3 and vertex 5 that was not in the original network. ✓

C: This is a spanning tree. There are no loops, duplicate edges or cycles and all vertices are connected. ✗

D: This is a spanning tree. There are no loops, duplicate edges or cycles and all vertices are connected. ✗

E: This is a spanning tree. There are no loops, duplicate edges or cycles and all vertices are connected. ✗

Answer

B

13. a. Explanation

The graph must connect all the rides. This is a spanning tree.

The shortest total length of cable will be used which means it will be the spanning tree with the lowest possible total weight.

Answer

Minimum spanning tree

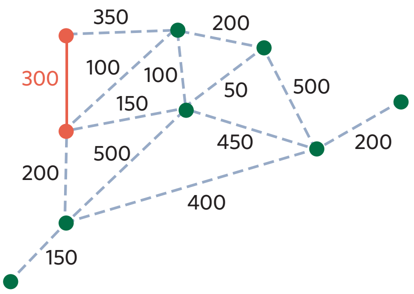
Some students only gave 'spanning tree'. Students were required to refer to the fact that the spanning tree would use the shortest total length of cable.

b. Explanation

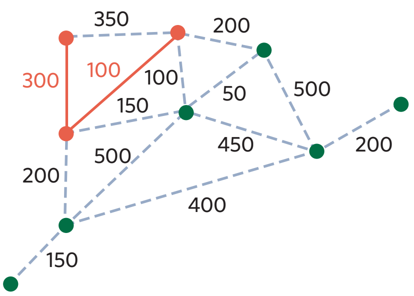
Step 1: Select any vertex.

For this example, start with the top-left vertex.

Step 2: Select the edge with the lowest weight connected to that vertex.

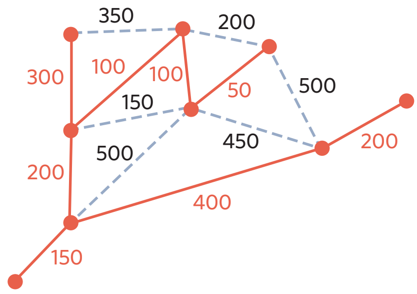


Step 3: Inspect all edges connected to either of the two connected vertices. Select the edge with the lowest weight that will connect a new vertex to the tree.



Step 4: Continue this process until all vertices are connected.

Answer



Questions from multiple lessons

14. B **15.** E

16. a. The *number of fish* in the pond increases by 0.85, on average, when the *pond size* increases by 1 m^2.

b. 4

c. 53.29%

8E Flow problems

Navigating directed graphs

1. B

2. a. Yes, B–C–D–E, 19 **b.** No

c. No **d.** Yes, B–C–H–A–D–G–E, 900

3. D

Calculating cut capacities

4. B

5. a. 15 **b.** 17 **c.** 28 **d.** 78

6. a. Lines A, C, and D **b.** Line A: 28
Line C: 42
Line D: 48

Determining the maximum flow

7. A

8. a. Line 4 **b.** Line 1 **c.** Line 2 **d.** Line 2

9. a. 11 **b.** 18

10. 19 sheep

Joining it all together

11. a. D **b.** Lines 4 and 7

c. 18 people

12. a. 19 people **b.** 26 people

c. 2 fewer people

13. a. 370 L **b.** 375 L

Back to contents

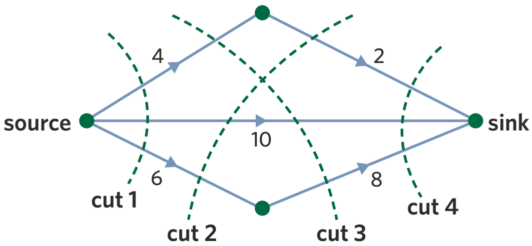
Exam practice

14. Explanation

Step 1: Identify all possible cuts through the network.

A cut must separate the source from the sink and completely stop the network flow.

There are four possible cuts for this network.



Step 2: Calculate the capacity of each cut.

Cut 1: $4 + 10 + 6 = 20$

Cut 2: $2 + 10 + 6 = 18$

Cut 3: $4 + 10 + 8 = 22$

Cut 4: $2 + 10 + 8 = 20$

Step 3: Identify the minimum cut.

The minimum cut is cut 2, with a capacity of 18 L/min.

The maximum flow is equal to the capacity of the minimum cut.

Answer

C

15. a. Explanation

The capacity of a cut is calculated by summing the capacities of all edges that flow from source to sink included in the cut.

All edges that the cut passes through flow from the source side of the cut to the sink side of the cut.

$20 + 12 + 20 = 52$

Answer

52 people

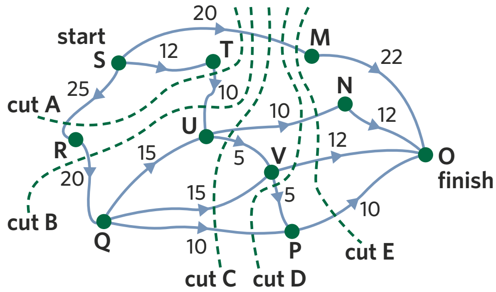
b. Explanation

Step 1: Simplify the problem by focussing on individual paths that must be cut.

The minimum cut will pass through the path S–M–O. The edge with the minimum capacity on this path is S–M, hence the minimum cut will pass through this edge.

The minimum cut will then either pass through the path S–T–U or U–N–O. The edges with the minimum capacity on these paths are T–U and U–N, hence the minimum cut will pass through one of these edges.

Step 2: Identify some feasible minimum cuts through the network given that the minimum cuts must pass through edge S–M and either edge T–U or U–N.



Cut A: $20 + 10 + 25 = 55$

Cut B: $20 + 10 + 20 = 50$

Cut C: $20 + 10 + 5 + 15 + 10 = 60$

Cut D: $20 + 10 + 12 + 5 + 10 = 57$

Cut E: $20 + 10 + 12 + 10 = 52$

Step 3: Identify the minimum cut.

The minimum cut has a capacity of 50 people per minute.

The maximum flow is equal to the capacity of the minimum cut.

Answer

50 people per minute

16. Explanation

Step 1: Calculate the capacity of each cut.

Cut A: $11 + x + 10 + 4 = 25 + x$

Cut B: $7 + 2 + x + 10 + 4 = 23 + x$

Cut C: $7 + 2 + x + 8 + 3 + 4 = 24 + x$

Cut D: $5 + 3 + 5 + 9 + 10 = 32$

Cut E: $5 + 3 + 9 + 10 = 27$

Step 2: Eliminate cuts A, C, and D.

Cuts A and C cannot ever be the minimum cut because cut B will always have a smaller capacity than them.

Cut D cannot be the minimum cut because cut E will always have a smaller capacity than it.

Step 3: Trial options B and E.

If $x = 2$:

Cut B: $23 + 2 = 25$

Cut E: 27

The maximum flow is given by the capacity of cut B.

If $x = 3$:

Cut B: $23 + 3 = 26$

Cut E: 27

The maximum flow is given by the capacity of cut B.

Answer

B

19% of students incorrectly chose option D. This may be because they interpreted the maximum flow to be equal to the cut with the greatest capacity rather than the smallest capacity.

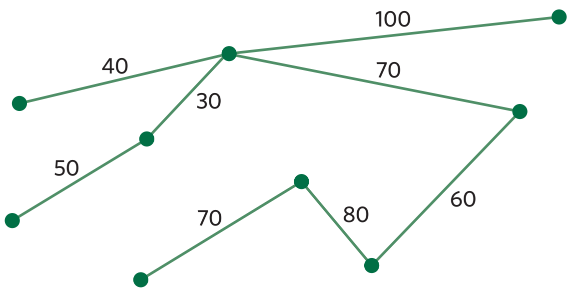
Back to contents

Questions from multiple lessons

17. E **18.** D

19. a. Minimal spanning tree

b.



8F Shortest path problems

Determining the shortest path by inspection

1. A

2. a. 17, A–B–C–E–D **b.** 20, A–G–F–C–D

c. 13, A–B–D **d.** 22, A–H–G–C–D

3. a. 1.65 km **b.** 1.15 km

Applying Dijkstra's algorithm

4. a. B **b.** 11

c. A–C–D–F

5. a. D **b.** 13

c. A–D–F–G

6. C

7. a. 27
A–G–C–F–E

b. 37
A–H–G–D–E

8. a. 16 minutes **b.** start–C–E–H–X

Joining it all together

9. a. B **b.** $135 **c.** D **d.** $140

Exam practice

10. Explanation

To solve this question, check whether each option is true or false.

A: This is false. A minimum cut is used to find maximum flow. ✗

B: This is false. Prim's algorithm is used to find the minimum spanning tree. ✗

C: This is true. Dijkstra's algorithm is used to find the shortest path between two vertices. ✓

D: This is false. A minimum spanning tree is used to find the minimum weight to connect all vertices. ✗

E: This is false. An Eulerian trail is a type of walk that includes every edge exactly once. This cannot be the shortest path in this network. ✗

Answer

C

11. Explanation

Step 1: Apply Dijkstra's algorithm to the weighted network.

N is the starting vertex and town S is the destination vertex.

	O	Q	P	R	S
N	$\boxed{15}$	30	30	X	X
O	$\boxed{15}$	$\boxed{30}$	30	90	120
Q	$\boxed{15}$	$\boxed{30}$	$\boxed{30}$	50	90
P	$\boxed{15}$	$\boxed{30}$	$\boxed{30}$	$\boxed{50}$	90
R	$\boxed{15}$	$\boxed{30}$	$\boxed{30}$	$\boxed{50}$	$\boxed{80}$

Step 2: Identify the route of the shortest path.

Start from the square in column S.

Draw a line up the column from this value until you reach the highest row with this same value. This is row R.

Draw a horizontal line to this vertex's column (R).

Repeat until you reach the row of the starting vertex.

The horizontal lines show the route of the shortest path.

	O	Q	P	R	S
N	$\boxed{15}$	30	30	X	X
O	$\boxed{15}$	$\boxed{30}$	30	90	120
Q	$\boxed{15}$	$\boxed{30}$	$\boxed{30}$	50	90
P	$\boxed{15}$	$\boxed{30}$	$\boxed{30}$	$\boxed{50}$	90
R	$\boxed{15}$	$\boxed{30}$	$\boxed{30}$	$\boxed{50}$	$\boxed{80}$

The shortest path is N–Q–R–S.

Answer

Quigley and Rosebush

Questions from multiple lessons

12. B **13.** B

14. a. $V_0 = 5000, \quad V_{n+1} = 1.03 \times V_n$

b. $V_n = 1.03^n \times 5000$ **c.** $5796.37

8G Matching problems

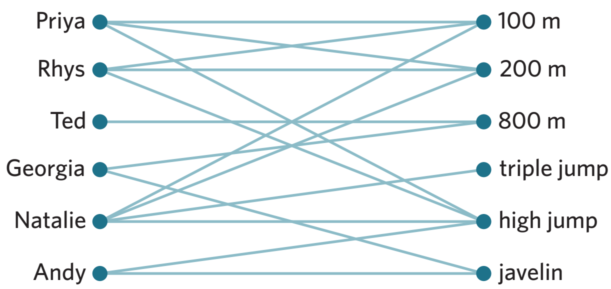
Representing matching problems using a bipartite graph

1. A

2. a. 3 students **b.** General Maths

c. 3 study groups **d.** English and General Maths

3.



Back to contents

Solving matching problems

4. D

5. a.

	research	writing	editing	presentation
Khan	0	10	5	5
Henry	10	20	0	30
Jamie	0	15	5	25
Mya	5	0	5	20

b. Only three lines are required to cover all of the zeros. As there are four allocations, allocation is not possible until a minimum of four lines are required to cover all of the zeros.

c. Whilst Jamie and Khan are equal fastest at research, Khan is also the fastest at presentation. To minimise time, Khan should complete the presentation section and let Jamie complete the research section.

d. Khan: presentation, Henry: editing, Jamie: research, Mya: writing

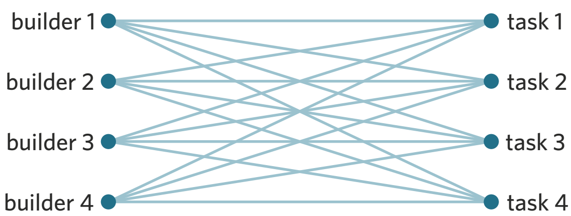
e. 205 minutes

6. Anna: jazz, Margot: hip hop, Tina: ballet, Svetlana: contemporary

7. Sophie: mathematical errors, Zephyr: clarity, Anna: other, Cassie: grammar

Joining it all together

8. a.



b.

	task A	task B	task C	task D
builder 1	20	0	40	10
builder 2	30	10	0	10
builder 3	40	30	40	0
builder 4	10	0	50	20

c. builder 1: task B, builder 2: task C, builder 3: task D, builder 4: task A

d. $350

9. 73 km

Exam practice

10. **Explanation**

Determine the people who can complete task 4.

Task 4 can only be completed by Mandy and Lee. While Mandy can only complete task 4, Lee can also complete task 1. Therefore, task 4 must be completed by Mandy.

Answer

B

11. **Explanation**

Step 1: Determine the highest averages for each batting position.

For batting position 1, Bo has the highest average.

For batting position 2, Bo and Cameron have the highest average.

For batting position 3, Alex has the highest average.

Step 2: Find the optimal allocation.

Batting position 1 should be allocated to Bo.

As Bo is allocated to batting position 1, batting position 2 should be allocated to Cameron.

Batting position 3 should be allocated to Alex.

Answer

Alex: 3, Bo: 1, Cameron: 2

> Some students attempted to use the Hungarian algorithm and solved the question as if it was a minimisation problem.

12. **Explanation**

Interpret the zero in terms of the tour Colin can plan.

In Colin's column there is only a zero in the tour 2 row. Therefore, Colin must plan tour 2 to minimise the total planning time.

Answer

Colin must plan tour 2.

> Some students made references to Colin planning tour 2 because he is the fastest. This is not correct, as Diane could plan tour 2 faster than Colin.

13. **Explanation**

Step 1: Calculate the completion time based on the allocation by the supervisor.

$4 + 3 + 3 + 2 = 12$

Step 2: Apply the Hungarian algorithm for when $k = 0$.

	task 1	task 2	task 3	task 4
Annie	4	0	3	0
Buddhi	0	0	1	1
Chuck	2	3	4	0
Dorothy	1	5	0	1

Note: The table will vary depending on where the lines that cover the zeros are drawn.

Step 3: Allocate the tasks based on the location of zeros.

Task 1 must be allocated to Buddhi.

Task 3 must be allocated to Dorothy.

Task 4 must be allocated to Chuck.

As Buddhi is allocated to task 1, task 2 must be allocated to Annie.

Step 4: Calculate the minimum completion time using the cost matrix.

$0 + 3 + 5 + 2 = 10$

Back to contents

Step 5: Determine the minimum value of k.

The completion time under the supervisor's allocation is 2 minutes longer than the minimum completion time when $k = 0$.

If $k = 2$, the minimum completion time would increase to 12 minutes, and the supervisors allocation would achieve the minimum completion time. This can be also found through the application of the Hungarian algorithm.

Note: When $k = 3$ or $k = 4$ the completion time is also 12 minutes. However, $k = 2$ is the smallest (least) value in which the supervisor's allocation is the minimum completion time.

Answer

C

58% of students incorrectly chose options D or E. These students likely found that these values of k lead to a minimum completion time that was equal to the supervisors allocation, however they failed to recognise that this was not the minimum value of k that did so.

Questions from multiple lessons

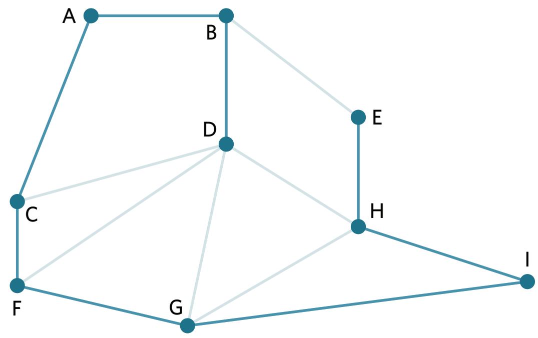
14. D

15. C

16. a. Vertex D

b. 1

c.



8H Activity networks and precedence tables

Interpreting precedence tables

1. B

2. A, B, C, D, E

Constructing activity networks

3. D

4. a.

activity	immediate predecessor(s)
A	–
B	A
C	A
D	A
E	B
F	D

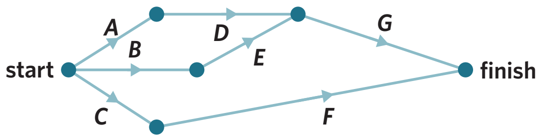
b.

activity	immediate predecessor(s)
A	–
B	–
C	A
D	A
E	B, C
F	D, E
G	D, E
H	G
I	F, H
J	G

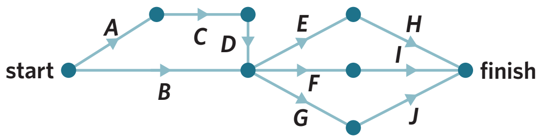
c.

activity	immediate predecessor(s)
A	–
B	A
C	A
D	B
E	B
F	C
G	E, F
H	E, F
I	C
J	D, G

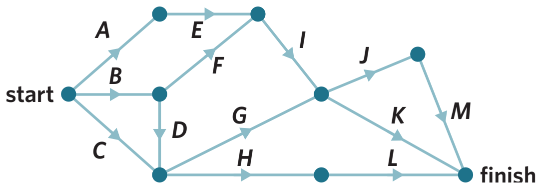
5. a.



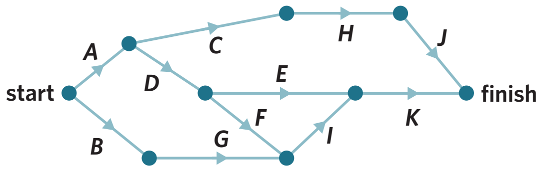
b.



c.



6.



Constructing activity networks with dummy activities

7. B

8. C

9. a. 2 **b.** 0 **c.** 2

10. a.

activity	immediate predecessor(s)
A	–
B	–
C	A
D	B
E	B
F	C, D
G	C, D
H	F
I	E, F, G

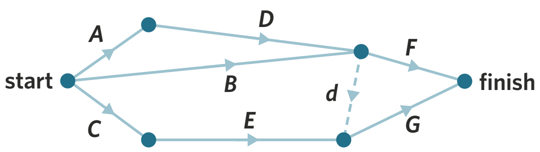
b.

activity	immediate predecessor(s)
A	–
B	–
C	–
D	B
E	A, D
F	B
G	C
H	I
I	F, G
J	B, E, H
K	I

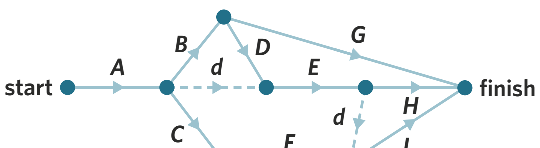
c.

activity	immediate predecessor(s)
A	–
B	–
C	–
D	B, C
E	A, D
F	A, D
G	B, C
H	C
I	E
J	F, G
K	F, G, I
L	J, H

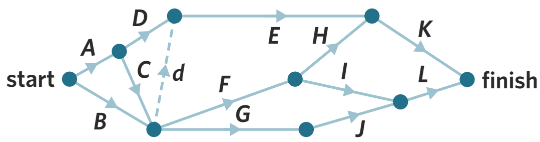
11. a.



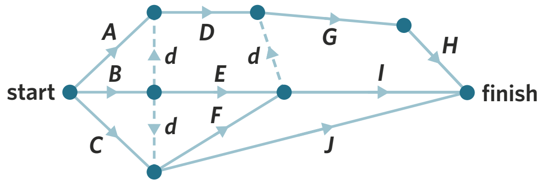
b.



c.



12.



Joining it all together

13. a.

activity	immediate predecessor(s)
A	–
B	A
C	A
D	A
E	B
F	D
G	C, E
H	C
I	F
J	G, H, I

b. Sew the pant legs and sew the jumpsuit sleeves.

c. D

Exam practice

14. Explanation

Find the activities that end at the start of activity I.

Activity D ends at the start of activity I.

A dummy activity ends at the start of activity I. It begins at the end of activity E.

Answer

D, E

A number of students incorrectly included the dummy as one of the immediate predecessors of activity I. These students failed to recognise that the dummy is not an activity, and is drawn to show that activity E is an immediate predecessor of activity I.

15. Explanation

Step 1: Find activities that share some, but not all, of the same immediate predecessors

Activity F has activity B as its immediate predecessor.

Activity C has activites A and B as its immediate predecessors.

Step 2: Find the activity that the dummy activity should start at.

A dummy activity is drawn from the end of the shared immediate predecessor.

The shared immediate predecessor is activity B.

Back to contents

Step 3: Find the activity that the dummy activity should end at.

A dummy activity is drawn to the start of the activity with an additional immediate predecessor.

The activity with an additional shared immediate predecessor is activity *C*.

Answer

This dummy activity could be drawn as a directed edge from the end of activity *B* to the start of activity *C*.

> Students need to be careful not to incorrectly assume that the dummy would be drawn to the start of activity *I* because it has the most immediate predecessors.

Questions from multiple lessons

16. D **17.** C

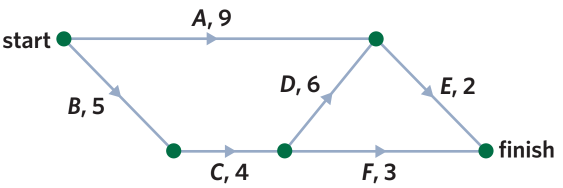
18. **a.** **i.** 8 L/s **ii.** 9 L/s **b.** 4 L/s

8I Critical path analysis

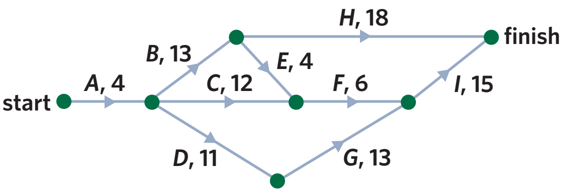
Interpreting weighted precedence tables

1. **a.** C

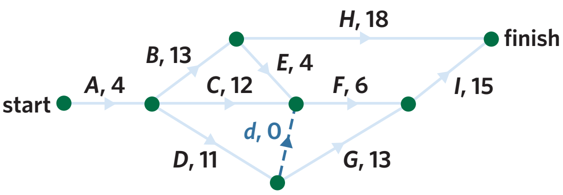
b.



2. **a.**



b.



Forward scanning to determine the EST and minimum completion time

3. **a.** B **b.** 7 **c.** 17

4. **a.** 14 minutes **b.** 30 minutes

Backward scanning to determine the LST

5. **a.** A **b.** 4

6. 11 minutes

Determining the critical path and float times

7. E

8. **a.** *A–D–H–I* **b.** 5

9. **a.** *B–C–D–H* **b.** No **c.** 8 hours

d. Yes **e.** No

Joining it all together

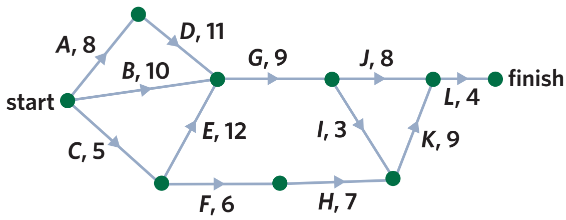
10. B

11. **a.** 37 minutes **b.** 0 minutes **c.** 6 minutes

d. *A–D–H–J* and *C–G–I–L*

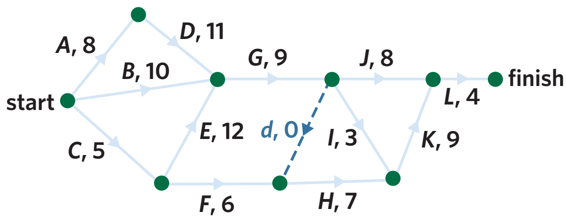
12. C

13. **a.**



b. 44 minutes **c.** *A–D–G–I–K–L*

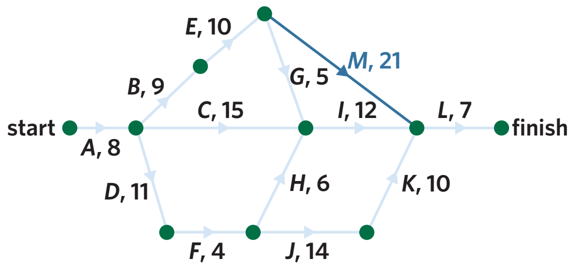
d.



e. 48 minutes **f.** *A–D–G–H–K–L*

14. **a.** *A–D–F–J–K–L* **b.** 54 days

c.



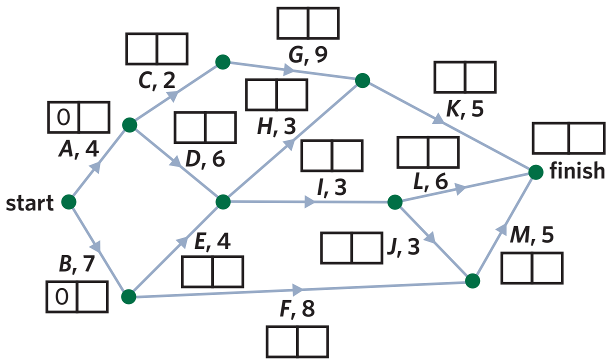
d. *A–B–E–M–L* **e.** 1 day

Exam practice

15. **Explanation**

Step 1: Draw a box, split into two cells, at each activity, as well as the finish vertex.

Fill in the EST as 0 in the left-hand cell of each of the starting activities (*A* and *B*).

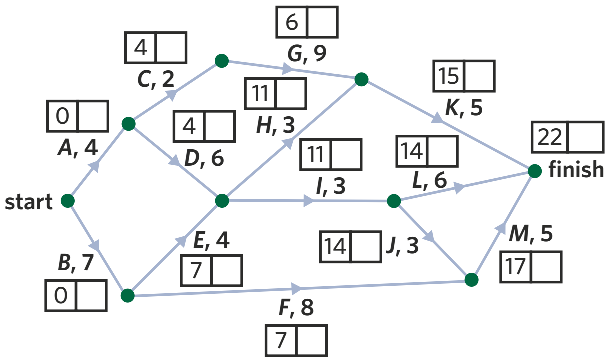


Back to contents

Step 2: Fill in the EST in the left-hand cell for all remaining activities and the finish vertex.

EST = EST of predecessor + duration of predecessor

If an activity has two or more immediate predecessors, its EST will be the largest value found when looking at the predecessors individually.



The minimum completion time is the EST of the finish vertex, 22 days.

Answer

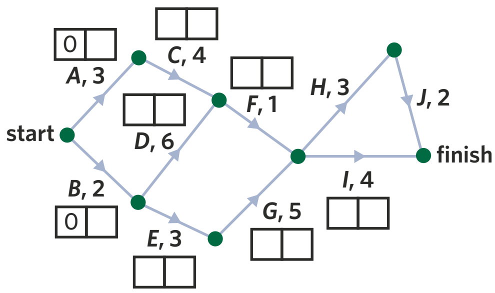
E

16% of students incorrectly chose option A. This was the minimum completion time calculated if, when an activity had two immediate predecessors, the EST was the smallest EST value found when looking at the predecessors individually.

16. a. **Explanation**

Step 1: Draw a box, split into two cells, at each activity that is on a possible path to activity *I*, including activity *I*.

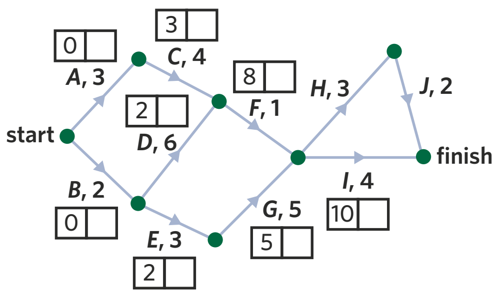
Fill in the EST as 0 in the left-hand cell of each of the starting activities (*A* and *B*).



Step 2: Fill in the EST in the left-hand cell for all remaining activities.

EST = EST of predecessor + duration of predecessor

If an activity has two or more immediate predecessors, its EST will be the largest value found when looking at the predecessors individually.

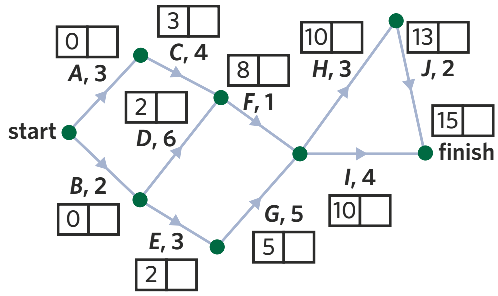


Answer

10 hours

b. **Explanation**

Step 1: Draw a box, split into two cells, at each remaining activity and the finish vertex and fill in the EST.

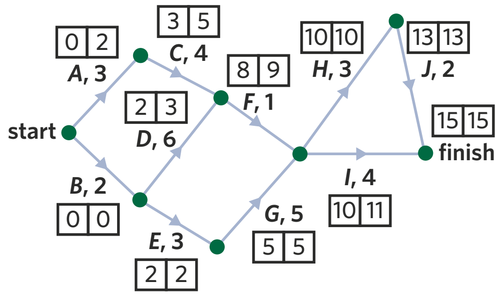


Step 2: Fill in the LST in the right-hand cell for all activities and the finish vertex.

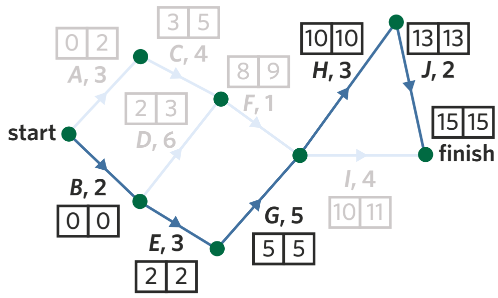
LST = LST of following activity − duration of activity

The LST for the finish vertex is the same as the EST.

If an activity has two or more activities following it, its LST will be the smallest value found when looking at each of the activities immediately following it.



Step 3: Highlight the path(s) that connects the start vertex to the finish vertex using only activities that have an EST and LST that are equal.



Answer

B–E–G–H–J

c. **Explanation**

Find the float times of non-critical activities.

float time = LST − EST

Activity *A*: 2 − 0 = 2

Activity *C*: 5 − 3 = 2

Activity *D*: 3 − 2 = 1

Activity *F*: 9 − 8 = 1

Answer

A and *C*

Back to contents

d. Explanation

Step 1: Find the starting point of the activity.

The activity has an EST of 5 hours. This means it will start at the same point as another activity that has an EST of 5 hours. There is only one current activity with an EST of 5 hours, activity *G*, and this starts at the end of activity *E*.

Step 2: Find the ending point of the activity.

The activity has a LST of 12 hours and a duration of 1 hour.

$$\begin{aligned} \textit{LST of following activity} &= \textit{LST of activity} + \textit{duration of activity} \\ &= 12 + 1 \\ &= 13 \end{aligned}$$

The only activity with a LST of 13 is activity *J*, so the activity ends at the start of activity *J*.

Answer

The extra activity could be represented on the network by a directed edge from the end of activity *E* to the start of activity *J*.

> Many students identified activity *E* as the starting point but did not identify *J* as the end point.

17. Explanation

Step 1: Identify possible paths that include activity *E*:

B–*E*–*J*

C–*G*–*F*–*E*–*J*

Step 2: Sum the durations of activities in these paths without activity *E*.

B–*E*–*J*: $5 + 7 = 12$

C–*G*–*F*–*E*–*J*: $4 + 3 + 1 + 7 = 15$

If the duration of activity *E* was any longer than 3 hours, the duration of the path *C*–*G*–*F*–*E*–*J* would be greater than 18 hours and the minimum completion time would increase.

Answer

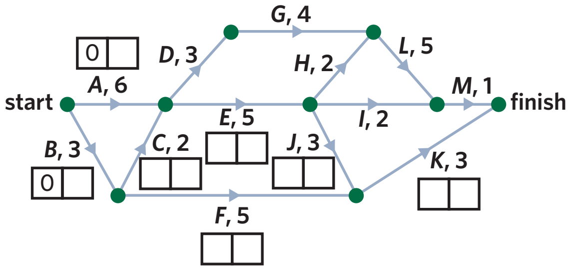
B

> 31% of students answered option E. This was the difference between the sum of known activity durations in path *B*–*E*–*J* and the minimum total completion time. Path *C*–*G*–*F*–*E*–*J* needs to be considered to arrive at the correct answer.

18. a. Explanation

Step 1: Draw a box, split into two cells, at each activity that is on a possible path to activity *K*, including activity *K*.

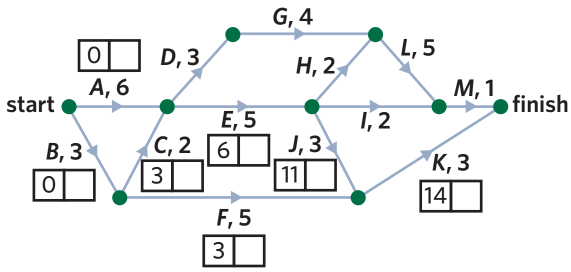
Fill in the EST as 0 in the left-hand cell of each of the starting activities (*A* and *B*).



Step 2: Fill in the EST in the left-hand cell for all remaining activities.

EST = EST of predecessor + duration of predecessor

If an activity has two or more immediate predecessors, its EST will be the largest value found when looking at the predecessors individually.

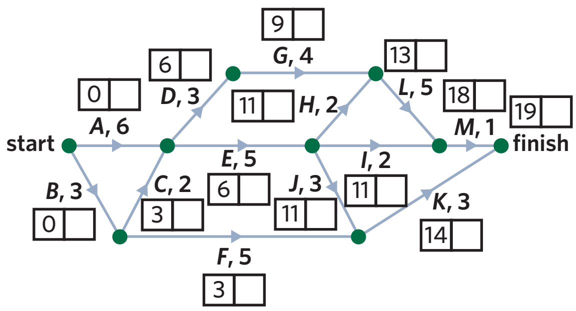


Answer

14 weeks

b. Explanation

Step 1: Draw a box, split into two cells, at each remaining activity, as well as the finish vertex. Fill in the EST of all activities as well as the finish vertex.

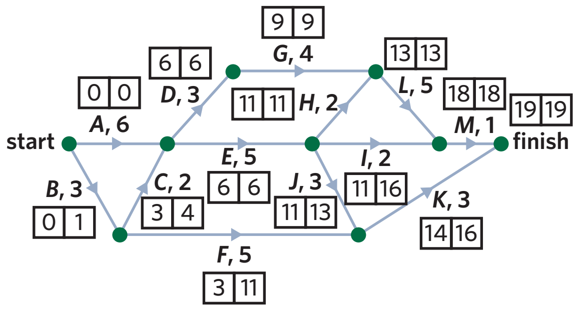


Step 2: Fill in the LST in the right-hand cell for all activities and the finish vertex.

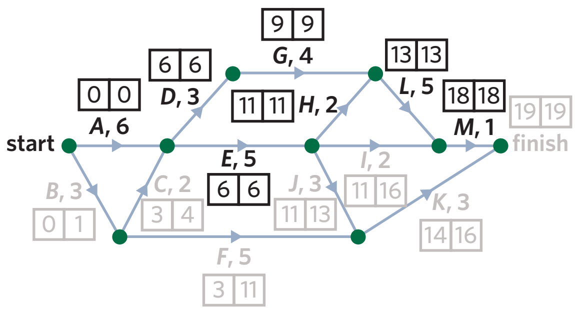
LST = LST of following activity − duration of activity

The LST for the finish vertex is the same as the EST.

If an activity has two or more activities following it, its LST will be the smallest value found when looking at each of the activities immediately following it.



Step 3: Highlight the activities that have an EST and LST that are equal.



Answer

7 activities

> Many students did not realise that there were two critical paths, leading to an answer of 5 activities.

Questions from multiple lessons

19. A **20.** A

21. **a.** *F* and *G* **b.** **i.** *C–G–J–N* **ii.** 4 days

8J Crashing

Crashing an activity network

1. C

2. **a.** D **b.** 18 days

c. *A–C–D–F–G* and *B–D–F–G*

3. **a.** *A–B–E–G* and *A–D–H*

b. No. Reducing the duration of activity *B* will not reduce the total duration of the critical path *A–D–H*.

c. 2 minutes **d.** *A–D–H*

Crashing an activity network with restrictions

4. D

5. **a.** *A–D–F–H* **b.** $1300

c. 23 days **d.** *C–E–G–H* and *A–D–F–H*

6. 41 emeralds

Joining it all together

7. **a.** *A*, *E* and *G* **b.** 15 weeks

c. $25 000

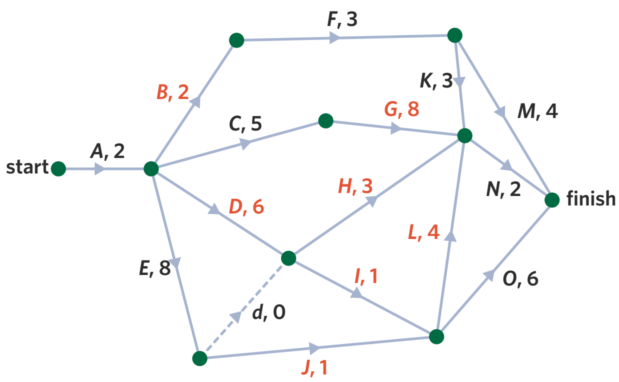
8. **a.** 28 hours **b.** *A–B–D–H–I–K*

c. 23 hours **d.** 120 galleons

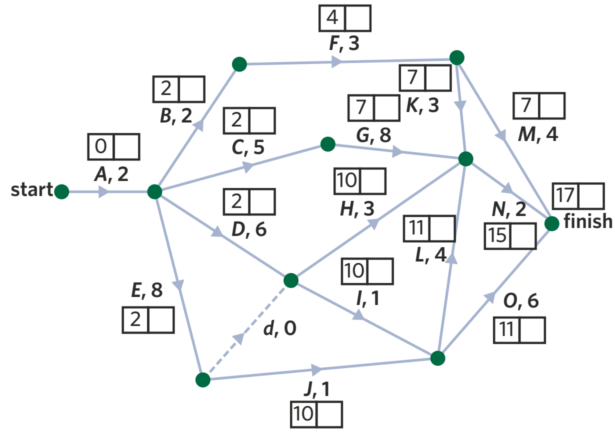
Exam practice

9. **a.** **Explanation**

Step 1: Reduce the duration of each of the activities by the maximum number of days.



Step 2: Complete forward scanning to find the new minimum completion time.



The minimum completion time is the EST of the finish vertex.

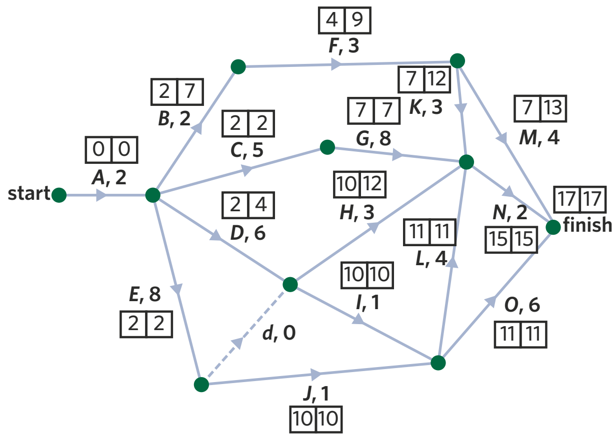
Answer

17 days

> Some students likely used incorrect reasoning to only reduce the duration of some of the activities. When finding the minimum completion time after crashing without considering cost, it is easiest to reduce the duration of all activities.

b. **Explanation**

Step 1: Complete backward scanning.



Activities *G*, *I*, *J* and *L* have a float time of 0.

Activities *B* and *D* have a float time greater than 0.

Step 2: Determine any redundant reduction in the duration of activities.

It is inefficient to pay for the reduction in the duration of activities when there will be float time.

The minimum completion time of the project would not change if activities *B* and *D* were each 1 day longer. Therefore, the durations of activities *B* and *D* do not need to be reduced.

Step 3: Identify the cost of reducing the minimum completion time by as much as possible.

The duration of activities *G*, *I*, *J* and *L* should each be reduced by 1 day.

Each activity costs $1000.

$4 \times 1000 = 4000$

Back to contents

Answer

$4000

Some students recognised the need to reduce the duration of the critical activities, but did not realise that activity *G* also had to be reduced as it was part of a new critical path.

10. Explanation

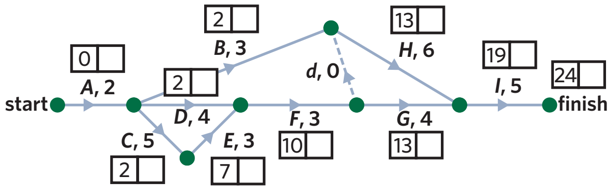
Step 1: Find the original minimum completion time of the project.

The critical path is *A–C–E–F–G–I*.

$2 + 5 + 3 + 6 + 4 + 5 = 25$

Step 2: Find the minimum completion time if activity *F* is completed before activity *H* is started.

Recall that the duration of activity *F* is now 3 weeks.

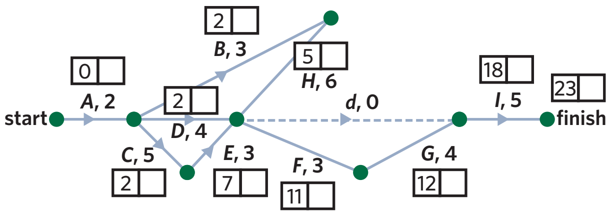


The minimum completion time has been reduced by 1 week.

Option A is correct and option B is incorrect.

Step 3: Find the minimum completion time if activity *H* is completed before activity *F* is started.

Recall that the duration of activity *F* is now 3 weeks.



The minimum completion time has been reduced by 2 weeks.

Options C, D and E are incorrect.

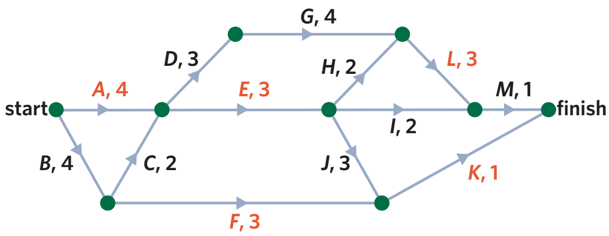
Answer

A

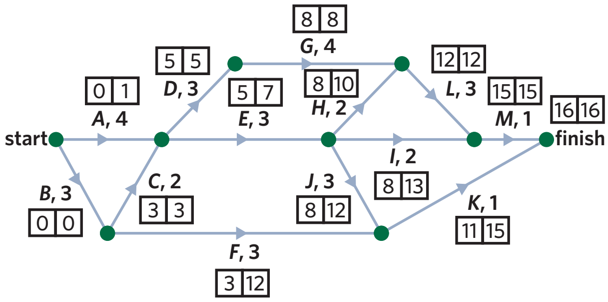
56% of students incorrectly chose options B or D. These students made the assumption that the reduction in the duration of activity *F* reduced the completion time of the project by the same amount.

11. Explanation

Step 1: Reduce the duration of each of the activities by the maximum number of weeks.



Step 2: Complete forward and backward scanning for the crashed activity network.



Activity *L* has a float time of 0.

Activity *A* has a float time of 1.

Activities *E*, *F* and *K* have a float time that is 2 or greater.

Step 3: Determine any redundant reduction in the duration of activities.

The minimum completion time of the project would not change if the duration of activity *F* was 2 weeks longer.

The minimum completion time of the project would not change if the duration of activities *E* and *K* were a total of 4 weeks longer.

The minimum completion time of the project would not change if the duration of activity *A* was 1 week longer.

Therefore, the durations of each of activity *E*, *F* and *K* do not need to be reduced at all, and the duration of activity *A* only needs to be reduced by 1 week.

Step 4: Calculate the cost of reducing the minimum completion time by as much as possible.

The duration of activity *L* should be reduced by 2 weeks and the duration of activity *A* should be reduced by 1 week.

$120\ 000 \times 2 + 140\ 000 = 380\ 000$

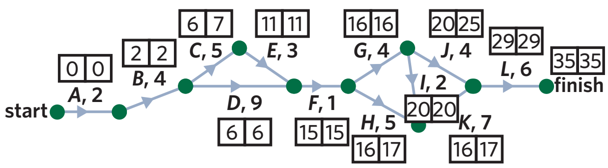
Answer

$380 000

Many students failed to recognise the activities that should be reduced.

12. Explanation

Step 1: Use forward and backward scanning to find the critical paths and float times.

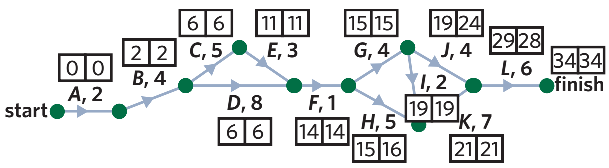


Step 2: Crash the project efficiently by 1 week.

Activities *D*, *G* and *K* are on the critical path.

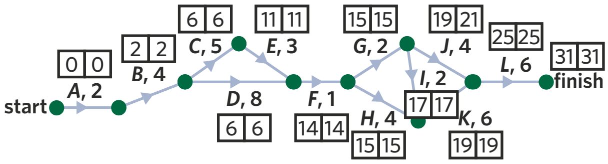
Activity *D* is the cheapest, so reduce its duration by 1 week.

Step 3: Complete forward and backward scanning to find the new critical paths and float times.



Back to contents

Step 4: Repeatedly crash the project efficiently by a week until the minimum completion time is 31 weeks.



Step 5: Fill in the table with the number of weeks each activity has been reduced by.

Answer

activity	C	D	G	H	K
reduction in completion time (0, 1 or 2 weeks)	0	1	2	1	1

Many students likely calculated the reduction in duration for each activity that led to the minimum cost for the maximum amount of reduction in project completion time.

Questions from multiple lessons

13. B

14. E

15. a. 97.5%

b. 320 schools

c. 214 students

Back to contents

GLOSSARY

A

activity network A directed network that represents a precedence table using edges as activities. p. 609

actual score An individual data value in a data set, often converted from or into a score that is standardised. p. 97

adjacency matrix A square, symmetrical matrix containing elements that represent the number of edges between pairs of vertices. p. 532

amortisation table A table that can be used to determine the reducing balance of a loan or investment. These tables can model a range of financial problems by showing step-by-step calculations for the resultant balances in a series of consecutive compounding periods. p. 354

amortising annuity A type of annuity where the investor receives a fixed sum of money every compounding period until the value of the investment reaches zero or the account is closed. p. 383

annuity An investment that involves a fixed sum of money paid to an investor at regular intervals, and typically earns compound interest. p. 383

annuity investment A type of annuity where the investor adds a fixed sum of money every compounding period in addition to the interest earned. This allows the investment to increase in value over time. p. 405

arithmetic A sequence or recurrence relation that increases or decreases by a constant amount. p. 298

B

back-to-back stem plot A stem plot that displays and compares the distribution of two categories. p. 119

bar chart A graphical display that is commonly used to display categorical data. The frequency or percentage frequency of each category is represented by columns of varied height. p. 10

bimodal When a data set has two distinct peaks. p. 64

binary matrix A matrix in which all elements are either 0 or 1. p. 460

bipartite graph A visual representation showing the connections between vertices that fall into two seperate groups. p. 596

boxplot A graphical representation of a five-number summary and any outliers. p. 51

bridge An edge that, if removed, would turn a connected graph into a disconnected graph. p. 519

C

categorical data Data that is organised into categories or groups. It is also referred to as qualitative data. p. 2

causation When a change in the explanatory variable definitively causes the observed change in the response variable. p. 148

centre The middle of a distribution, typically measured by the mean and median. p. 64

centring A process used with moving mean smoothing where there is an even number of data points. This process is used to align the smoothed data with the time series. p. 239

circuit A trail that starts and ends at the same vertex. No edges are repeated and vertices may be repeated. p. 548

coded time The simplified numerical representation of time. p. 222

coefficient of determination The degree to which the response variable can be predicted from the explanatory variable for a given linear relationship, usually expressed as a percentage. It is denoted as r^2. p. 179

coincidence Used in the context of data analysis, this relates to an association observed between two unrelated variables. p. 149

column A vertical list of numbers. p. 418

column matrix A matrix with one column and any number of rows. p. 419

column permutation A new matrix formed from the rearranged columns of another matrix. p. 460

common difference The amount being added or subtracted from one term to the next in a number pattern, denoted as d in a recurrence relation. p. 294

common ratio The amount that each term is being multiplied by from one term to the next in a number pattern, denoted as R in a recurrence relation. p. 294

common response When an observed association is caused by a third variable that produces the same response in the initial two variables. p. 148

communication matrix A square binary matrix where the 1s represent connections in a communication system. p. 469

complete graph A graph that directly connects every vertex to every other vertex. p. 523

compound interest Interest that is charged or earned at regular time intervals. It is usually calculated on the previous balance and changes exponentially. p. 339

confounding variables External variables that can also contribute to an observed association. p. 149

connected graph A graph in which every vertex is connected to each other, either directly or indirectly. p. 522

continuous data Numerical data can consist of any value. This type of data can contain decimals and fractions, is part of a continuous scale and would typically be collected by measuring. p. 3

Back to contents

crashing The process of reducing the minimum completion time of a project by reducing the duration of an activity on the critical path(s). p. 636

critical path A sequence of activities, each of which cannot be delayed without increasing the minimum completion time of the project. p. 627

culling The reduction of an animal population by slaughter. In a matrix recurrence relation scenario, this is represented by a subtraction to the population from one state to the next. p. 506

cut A line passing through a network that cuts the network in two, completely stopping any network flow from the source to the sink. p. 575

cut capacity The total capacity, or weighting, of all edges that a cut passes through. p. 575

cycle In the context of graphs and networks, a path that starts and ends at the same vertex. No edges or vertices are repeated. p. 548

cyclical variation Regular peaks and troughs in a time series plot. p. 226

D

data A set of values, words or responses, that is collected and ordered by variables. p. 2

degenerate graph A graph with no edges. All of the vertices are isolated. p. 523

degree The number of edges connected to a vertex. p. 519

depreciation The loss of value of an asset. p. 306

deseasonalising The process of removing the seasonal variation of a time series, to reveal any underlying trends. p. 270

determinant A number that can be calculated for a square matrix which determines whether the inverse of the matrix is defined. p. 451

diagonal matrix A square matrix in which all elements not in the leading diagonal are '0'. p. 419

Dijkstra's algorithm A method used to find the shortest path between two specified vertices in a weighted network. p. 586

directed graph (digraph) A network containing arrows on each edge that show directional information between vertices. p. 535

direction An indication of the relationship between two variables and how one responds to the other. The direction can be positive or negative. p. 134

discrete data Numerical data can only consist of a set of fixed values within a range. This type of data usually contains whole numbers and would typically be collected by counting. p. 3

dominance matrix A square binary matrix where the 1s represent one-step connections within the associated network. p. 472

dot plot A visual display that can be used for discrete numerical data, where each data point is represented by a dot above a single axis. The number of dots above the value on the axis represents its frequency. p. 20

dummy activity A directed edge that is drawn on an activity network to show additional predecessors when it is impossible to correctly represent a precedence table. p. 611

E

earliest start time (EST) The earliest time, relative to the beginning of the project, that an individual activity can start. p. 624

edge A line connecting one vertex to either another vertex or itself. p. 519

effective interest rate The adjusted or 'real' rate, respective to the number of compounding periods in the year. It can be used to compare the total amount of interest earned or charged on different investment or loan options with different compounding periods. p. 348

element An entry in a matrix. p. 418

equilibrium state matrix The state matrix (S_n) which has no difference compared to the matrix occurring after it (S_{n+1}). p. 496

Euler's rule A rule that shows the relationship between the number of vertices, edges and faces for all connected planar graphs. This is expressed as $v - e + f = 2$. p. 524

Eulerian circuit An Eulerian trail that starts and ends at the same vertex. p. 550

Eulerian trail A trail that includes every edge in a graph exactly once. p. 550

experimentation This involves observing or modifying changes in the explanatory variable to observe changes in the response variable, while all possible external variables remain constant. p. 149

explanatory variable The variable that is used to explain or predict the changes observed in the response variable. It is also known as the independent variable. p. 131

extrapolation Making predictions that are outside the range of the data set. These predictions have limited reliability. p. 170

F

face A separate area on a graph that is bordered by edges. The space outside the graph is counted as an additional face. p. 519

fence The boundary used to determine what is and what is not an outlier in a data set. p. 50

financial solver A program in the CAS calculator that can be used to solve complex financial problems. p. 356

five-number summary A summary that provides key information about a set of data and its distribution, including centre and spread. The summary includes the minimum, Q_1, median, Q_3, and maximum. p. 47

Back to contents

flat rate depreciation When the value of an asset decreases by a constant amount each period. p. 306

float time The maximum amount of time that an activity can be delayed without changing the completion time of the project. It can be calculated by finding the difference between the EST and LST. p. 627

flow The capacity of movement through a network, starting at one vertex (the source) and ending at another vertex (the sink). p. 573

form A determination of whether a relationship is linear or non-linear in a data analysis of two variables. p. 134

frequency table A table that tallies how often each value in a data set occurs. p. 9

FV (future value) The future value of the loan or investment, after n compounding periods. p. 356

G

geometric A sequence or recurrence relation that increases or decreases exponentially. p. 298

geometric growth/decay When each term in a sequence increases or decreases by a constant ratio. p. 294

graph A diagram that is used to show the connections between a group of common elements, such as objects, locations, people or activities. p. 518

grouped bar chart An extension of a bar chart that displays two categorical variables. The categories of one variable are displayed on one axis, while the categories of the other variable are represented with different bars for each category. p. 107

grouped frequency table A table that groups data into regular intervals, displaying the frequency and percentage frequency of each interval. p. 24

H

Hamiltonian cycle A Hamiltonian path that starts and ends at the same vertex. p. 552

Hamiltonian path A path that includes every vertex in a graph exactly once, without repeating edges. It does not need to include every edge. p. 552

histogram A visual representation of a grouped frequency distribution table. It can display either frequency or percentage frequency on its vertical axis. p. 25

Hungarian algorithm A method used to find the optimal allocation (minimum cost) of a cost matrix. p. 597

I

identity matrix A square matrix where all of the elements in the leading diagonal are one and the rest of the elements are zero. Multiplying a matrix by the identity matrix will result in the original matrix. It is denoted as I. p. 419

immediate predecessor An activity that must be completed directly before another activity can begin. p. 608

initial state matrix A column matrix that represents the initial, or starting, state of a system. It is commonly denoted as S_0, but can also be denoted as S_1. p. 481

interest The cost of borrowing money. For loans, it is the fee that banks and lenders charge the borrower. For investments, it is the money investors receive for their investment. p. 333

interest-only loan A compound interest loan where the borrower only repays the interest that is charged. As a result of this, the principal that must be paid back remains the same after each compounding period. p. 375

interpolation Making predictions that are within the range of the data set. p. 170

interquartile range (IQR) A measure of the spread of the middle 50% of a data set. p. 49

inverse matrix A matrix that has the property such that when it is pre-multiplied or post-multiplied with the original matrix, the result is the identity matrix, I. p. 452

irregular fluctuations Random variations in a time series plot that cannot be explained by trends, seasonality, cycles or structural change. p. 227

isolated vertex A vertex that is not connected by an edge to any other vertices on the graph. p. 520

isomorphic graphs Graphs that display the same information as each other but are drawn differently. p. 522

iteration Each term in a repeated process or recurrence relation, where u_n is the value of the term after n iterations. p. 298

L

latest start time (LST) The latest time, relative to the beginning of the project, that an individual activity can start without increasing the minimum completion time of the project. p. 626

leading diagonal The series of elements in a square matrix that goes from the top left element to the bottom right element. p. 419

least squares regression line The line which creates the minimum sum of the squared residuals. This is expressed in the form $y = a + bx$, where b represents the slope of the line, and a represents the y-intercept. p. 157

Leslie matrix A unique application of transition matrices, denoted as L, used to model the growth of a population and its age distribution over time. p. 486

linear growth/decay When each term in a sequence increases or decreases by a constant amount. p. 294

linearise To use a transformation to make data linear. p. 192

log x transformation A type of transformation that can be used to linearise non-linear data, where the logarithm of each x-value is taken. It 'compresses' larger x-values more than smaller x-values. p. 196

Back to contents

log y transformation A type of transformation that can be used to linearise non-linear data, where the logarithm of each y-value is taken. It 'compresses' larger y-values more than smaller y-values. p. 196

logarithmic scale A scale used to display numerical data that contains numbers of very different sizes. It can be used to reveal details about the underlying distribution of data sets with multiple orders of magnitude. p. 39

logarithms A function that returns a value representing the power to which a fixed number (the base) must be raised to produce a given number. p. 37

loop An edge that connects a vertex back to itself. p. 519

lower triangular matrix A matrix where all elements are '0' above the leading diagonal. p. 419

M

matrix A tool for displaying a collection of numerical values. p. 418

matrix power The resultant matrix when a matrix is raised to an index or power. p. 443

matrix product The resultant matrix when two or more matrices are multiplied. p. 439

maximum The largest value in a set of data. p. 47

maximum flow The minimum capacity through a network. p. 576

mean A measure of centre that averages out all values in a data set into even groupings. p. 80

median The middle value in an ordered set of data. p. 47

minimum The smallest value in a set of data. p. 47

minimum completion time The shortest time in which all activities within a project can be completed. p. 624

minimum cut The cut with the smallest cut capacity. p. 576

minimum spanning tree The spanning tree with the lowest total weight when the weights on all the edges are summed. p. 564

mode The most frequently occurring value in the data set. p. 13

moving mean smoothing Removing fluctuations in time series data by calculating the mean of each data point and the data points around it. p. 234

moving median smoothing Removing fluctuations in time series data by calculating the median of each data point and the data points around it. p. 254

N

negatively skewed A distribution that trails off in the negative direction. p. 63

nominal data Categorical data that cannot be sorted into a logical ordered list or hierarchy. p. 3

nominal interest rate The quoted interest rate for a loan or investment. Also known as the annual interest rate. p. 348

normal distribution A symmetrical (or approximately symmetrical) numerical data set that is centred around the mean, with a width determined by the standard deviation. p. 87

numerical data Data that can be counted or measured. It is also referred to as quantitative data. p. 2

O

observation Used in the context of data analysis, this relates to the collection of data to determine an association between two variables. p. 149

one-step communication links Direct communication links between two points. p. 469

one-step dominance Direct dominance of one point over another. p. 472

order Describes the dimensions of a matrix. It is expressed as *number of rows* × *number of columns*. p. 418

ordinal data Categorical data that can be sorted into a logical ordered list or hierarchy. p. 3

outliers Values that fall far outside the range of regular data within a set of numbers. p. 50

P

parallel boxplots A sequence of boxplots that display and compare the distribution of two or more data sets. p. 122

parallel dot plots A sequence of dot plots that display and compare the distribution of two or more data sets. p. 121

path A walk in which no edges or vertices are repeated. p. 548

Pearson's correlation coefficient A measure of the strength and direction of a linear relationship between two variables. It is also known as the correlation coefficient and is denoted as r. p. 145

percentage frequency The proportion of occurrences each value or category occurs in relation to the rest of the data, represented as a percentage. p. 9

percentage segmented bar chart A segmented bar chart that displays the percentage frequency of each category. p. 11

period The distance between two adjacent peaks. It may be regular or irregular. p. 226

permutation matrix A square binary matrix in which each row and column must contain the number 1 only once. p. 460

perpetuity A type of annuity where the payment received by the investor is equal to the interest earned. As a result of this, the investment value remains the same after each compounding period. p. 395

planar graph A graph that can be drawn with no overlapping edges. p. 522

PMT (payment amount) The payment amount made per compounding period. p. 356

population An entire group from which to collect data. p. 80

positively skewed A distribution that trails off in the positive direction. p. 63

post-multiplication The multiplication of one matrix after another matrix. p. 439

pre-multiplication The multiplication of one matrix before another matrix. p. 439

precedence table A table that displays the immediate predecessors of all activities within a project. p. 608

Prim's algorithm A method used to find the minimum spanning tree of a graph. p. 564

principal The initial value of an investment or loan. p. 306

PV (present value) The present value of the loan or investment. It is also known as the current value. p. 356

Q

quartiles The division of a distribution of data into quarters. p. 47

R

range A measure of the spread of the entire data set, found by calculating the difference between the maximum and minimum values. p. 49

reciprocal The result of taking a number and dividing 1 by that number. p. 198

recurrence relation A formula that links each term in a pattern-based sequence to the term after it. p. 294

reducing balance depreciation When the value of an asset decreases by a constant ratio, rather than by a constant amount, each time period. p. 316

reducing balance loan A compound interest loan with payments made at regular intervals to reduce the amount owing. p. 362

redundant communication links Communication links that start and finish at the same point, represented by non-zero elements in the leading diagonal. p. 470

reseasonalising The process of restoring the normal seasonal variation of previously deseasonalised data. p. 271

residual The vertical distance between the actual recorded data value, and the corresponding point on the least squares line. p. 157

residual plot A graph of the residual values plotted against the explanatory variable. p. 180

response variable The variable that is explained or predicted by changes in the explanatory variable. It is also known as the dependent variable. p. 131

restocking The increase of stock with a new supply. In a matrix recurrence relation scenario, this is represented by an addition to the population from one state to the next. p. 506

route A list of the vertices travelled through when moving from one vertex to another, showing the pathway travelled in a graph. p. 547

row A horizontal list of numbers. p. 418

row matrix A matrix with one row and any number of columns. p. 419

row permutation A new matrix formed from the rearranged rows of another matrix. p. 460

S

sample A subset of a population. p. 80

scalar multiplication Each element in a matrix is multiplied by a single number (the scalar). p. 431

scatterplot A display used to represent data relating to two numerical variables. Each point is determined by the value of one variable and the corresponding variable of the second variable, creating a coordinate. p. 132

scheduling In the context of graphs and networks, the allocation of an estimated completion time for each activity in a project. p. 622

seasonal index A measure of how a particular season compares to the average season. p. 268

seasonality Regular peaks and troughs in a time series plot that occur at the same time each period, within a calendar year. p. 226

segmented bar chart A variation of a bar chart, with each category combined into one column. p. 11

sequence A number pattern written as a list, in a logical order. p. 294

shortest path The shortest distance connecting two specified vertices in a weighted network. p. 585

simple graph A graph that does not contain any loops or duplicate edges. p. 522

simple interest Interest that is charged or earned at regular time intervals. It is usually calculated as a percentage of the principal amount and changes at a linear rate. p. 333

singular matrix A square matrix that does not have an inverse, where the determinant is equal to zero. p. 452

sink The vertex at which flow ends. p. 573

slope The average change in the response variable for every one-unit increase in the explanatory variable. It is also known as the gradient. p. 169

smoothing The process of removing fluctuations in time series data to reveal underlying trends. p. 234

source The vertex at which flow begins. p. 573

spanning tree A tree which connects all vertices in the original graph. p. 563

spread A measure that refers to how similar or varied a set of numerical data is. p. 49

square matrix A matrix with an equal number of rows and columns. p. 419

standard deviation A measure of spread that is based on the average difference between each data point and the mean of the data set. p. 81

standardised score A measure of the number of standard deviations between the mean and a data value. p. 97

Back to contents

state matrix A column matrix that is a snapshot of a system at a point in time. The state matrix at time period n is denoted as S_n. p. 481

steady state matrix See equilibrium state matrix. p. 496

stem plot A visual display of numerical data, where data points are split into two parts and rewritten such that the first digit(s) are grouped into one 'stem', and the remaining digits are listed in the 'leaf'. p. 21

strength An indication of how close the data points are to the general trend of the scatterplot when two variables are plotted against each other in data analysis. The strength can be weak, moderate or strong. p. 134

structural change When an established pattern of a time series plot is permanently and fundamentally altered due to an external factor. p. 227

summing matrix A row or column matrix that only has the number 1 for each element, used to calculate the sum of rows or columns in another matrix. p. 442

symmetric A distribution that is evenly distributed and does not trail off in either direction more than the other. p. 64

symmetric matrix A square matrix that is symmetric across the leading diagonal. p. 419

T

term In the context of number patterns, a term refers to any individual number within the list. p. 294

time series A subset of bivariate data where the explanatory variable is time. p. 222

trail A walk in which no edges are repeated, though vertices may be repeated. p. 548

transition diagram A visual representation of how a transition matrix functions. p. 504

transition matrix A square matrix that is used to represent the movement or changes of a system between states. It is denoted as T. p. 482

transpose The transpose of a matrix can be determined by swapping its rows and columns. p. 432

tree A type of graph that has no loops, duplicate edges or cycles. It uses the least number of edges to connect the vertices. p. 562

trend A general upwards (increasing) or downwards (decreasing) movement. p. 226

two-step communication links Indirect communication links between two points via a third point. p. 470

two-step dominance Indirect dominance of one point over another, via a third point. p. 472

two-way frequency table A table that can be used to compare different frequencies for two sets of categorical variables. p. 106

two-way percentage frequency table A table that can be used to compare different percentage frequencies for two sets of categorical variables. p. 106

U

unit cost depreciation When the value of an asset decreases by a constant amount for each unit of use. p. 306

upper triangular matrix A matrix where all elements are '0' below the leading diagonal. p. 419

V

vertex A point on a graph or network. p. 518

W

walk A route that passes through any number of vertices, in any order, starting and finishing at any vertex. p. 547

weighted graph A graph that has numeric values attached to each edge to provide further information about the connections between the vertices. Also known as a weighted network. p. 562

X

x-reciprocal transformation A type of transformation that can be used to linearise non-linear data, where 1 is divided by each x-value. It 'compresses' x-values that are greater than one, and 'stretches' x-values that are less than one. p. 198

x-squared transformation A type of transformation that can be used to linearise non-linear data, where each x-value is squared. It 'stretches' larger x-values more than smaller x-values. p. 193

Y

y-intercept The value of the response variable (y) when the explanatory variable (x) is equal to 0. p. 169

y-reciprocal transformation A type of transformation that can be used to linearise non-linear data, where 1 is divided by each y-value. It 'compresses' y-values that are greater than one, and 'stretches' y-values that are less than one. p. 198

y-squared transformation A type of transformation that can be used to linearise non-linear data, where each y-value is squared. It 'stretches' larger y-values more than smaller y-values. p. 193

Z

z-score See 'Standardised score'. p. 97

zero matrix A matrix in which every element is '0'. p. 419