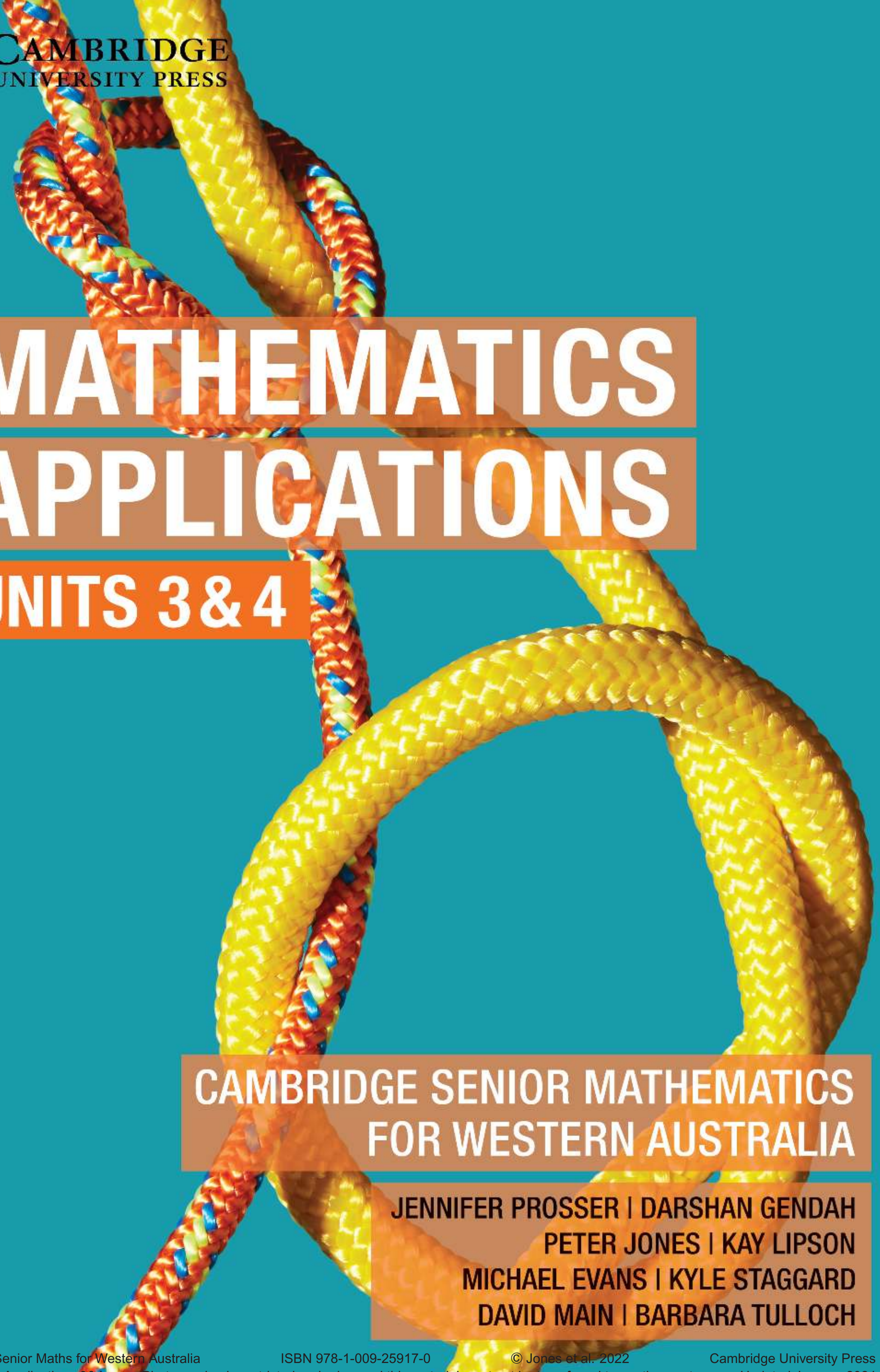
CAMBRIDGE
UNIVERSITY PRESS



MATHEMATICS APPLICATIONS

UNITS 3 & 4

CAMBRIDGE SENIOR MATHEMATICS
FOR WESTERN AUSTRALIA

JENNIFER PROSSER | DARSHAN GENDAH
PETER JONES | KAY LIPSON
MICHAEL EVANS | KYLE STAGGARD
DAVID MAIN | BARBARA TULLOCH

CAMBRIDGE
UNIVERSITY PRESS

Shaftesbury Road, Cambridge CB2 8EA, United Kingdom
One Liberty Plaza, 20th Floor, New York, NY 10006, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India
103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment, a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org

© Peter Jones, Kay Lipson, Michael Evans, Kyle Staggard, David Main, Barbara Tulloch 2022

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press & Assessment.

First published 2022
20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2

Cover design by Denise Lane (Sardine)
Text design by Jane Pitkethly
Typeset by diacriTech
Printed in China by C & C Offset Printing Co., Ltd.

A catalogue record for this book is available from the National Library of Australia at www.nla.gov.au

ISBN 978-1-009-25917-0 Paperback

Additional resources for this publication at www.cambridge.edu.au/GO

Reproduction and communication for educational purposes
The Australian *Copyright Act 1968* (the Act) allows a maximum of one chapter or 10% of the pages of this publication, whichever is the greater, to be reproduced and/or communicated by any educational institution for its educational purposes provided that the educational institution (or the body that administers it) has given a remuneration notice to Copyright Agency Limited (CAL) under the Act.

For details of the CAL licence for educational institutions contact:

Copyright Agency Limited
Level 12, 66 Goulburn Street
Sydney NSW 2000
Telephone: (02) 9394 7600
Facsimile: (02) 9394 7601
Email: memberservices@copyright.com.au

Reproduction and communication for other purposes
Except as permitted under the Act (for example a fair dealing for the purposes of study, research, criticism or review) no part of this publication may be reproduced, stored in a retrieval system, communicated or transmitted in any form or by any means without prior written permission. All inquiries should be made to the publisher at the address above.

Cambridge University Press & Assessment has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate. Information regarding prices, travel timetables and other factual information given in this work is correct at the time of first printing but Cambridge University Press & Assessment does not guarantee the accuracy of such information thereafter.

Cambridge University Press & Assessment acknowledges the Australian Aboriginal and Torres Strait Islander peoples of this nation. We acknowledge the traditional custodians of the lands on which our company is located and where we conduct our business. We pay our respects to ancestors and Elders, past and present. Cambridge University Press & Assessment is committed to honouring Australian Aboriginal and Torres Strait Islander peoples' unique cultural and spiritual relationships to the land, waters and seas and their rich contribution to society.

Contents

3 Regression: fitting lines to data 56

4 Number patterns and recursion 100

5 Applications of sequences and recurrence relations 145

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

9 Connector, assignment and flow problems 356

10 Project planning 417

Online appendices accessed through the Interactive Textbook or PDF Textbook

Introduction

Cambridge Mathematics Applications for Western Australia Units 3 & 4 is a new edition aligned specifically to the Western Australian Mathematics Applications Year 12 syllabus. Covering both Units 3 and 4 in one resource, it has been written with practical contexts and worded questions as its priority alongside ample practice offered through worked examples and exercises. The course is a continuation from the Western Australian Mathematics Applications Year 11 course (covered in the new Cambridge Units 1 & 2 textbook) and builds upon the knowledge and skills acquired within these units.

The textbook starts by introducing bivariate statistics, building upon the understanding of univariate data from the Year 11 Applications course, before leading to regression analysis and time series data. The study of number patterns and recursion are covered in terms of the general applications of sequences and specifically embedded within the finance chapters. Graphs and networks are presented with a focus on definitions and terminology, prior to adding context and making connections to flow problems and project planning.

Worked examples utilising CAS calculators are provided throughout, with screenshots and detailed user instructions for both ClassPad and TI-Nspire included for each CAS example. The CAS 'ribbons' used within each exercise direct students to questions which should be completed with the aid of technology and are intended to help prepare students for examinations and other assessments.

Compared to the previous Australian Curriculum edition, this WA edition has undergone a number of revisions. Careful adjustments to notation and language have been made throughout to match that used in the WA syllabus, the Year 12 WACE exam and in WA classrooms more generally. Sections on trivial sequences, isomorphic graphs, smoothing time series using moving medians, and interest-only loans have all been removed. Multiple-choice questions that were formerly located in the chapter reviews and revision chapters have also been removed for this edition.

The TI-Nspire calculator examples and instructions have been completed by Russell Brown and those for the Casio ClassPad have been completed by Maria Schaffner.

The integration of the features of the textbook and the new digital components of the package, powered by Cambridge HOTmaths, are illustrated on pages viii to xi.

About Cambridge HOTmaths

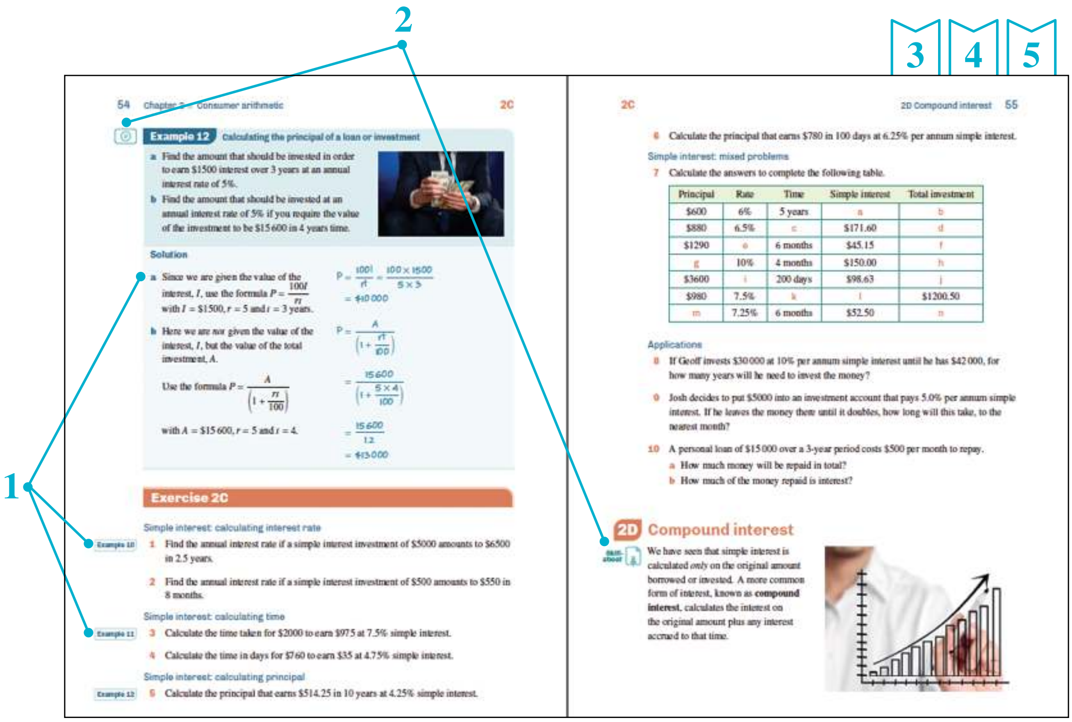
Cambridge HOTmaths is a comprehensive, award-winning mathematics learning system – an interactive online maths learning, teaching and assessment resource for students and teachers, for individuals or whole classes, for school and at home. Its digital engine or platform is used to host and power the Interactive Textbook and the Online Teaching Suite, and selected topics from HOTmaths' own Years 9 and 10 courses area are available for revision of prior knowledge. All this is included in the price of the textbook.

Overview

Overview of the print book

1. Graded step-by-step worked examples with precise explanations (and video versions) encourage independent learning, and are linked to exercise questions.
2. Additional linked resources in the Interactive Textbook are indicated by icons, such as skillsheets and video versions of examples.
3. Questions that suit the use of a CAS calculator to solve them are identified within exercises.
4. Chapter reviews contain a chapter summary and short-answer and extended-response questions.
5. The glossary includes page numbers of the main explanation of each term.

Numbers refer to descriptions above.



54 Chapter 2 – Consumer arithmetic 2C

Example 12 Calculating the principal of a loan or investment

a Find the amount that should be invested in order to earn $1500 interest over 3 years at an annual interest rate of 5%.

b Find the amount that should be invested at an annual interest rate of 5% if you require the value of the investment to be $15 600 in 4 years time.

Solution

a Since we are given the value of the interest, I, use the formula $P = \frac{100I}{rt}$ with $I = \$1500$, $r = 5$ and $t = 3$ years.

$$P = \frac{100I}{rt} = \frac{100 \times 1500}{5 \times 3} = \$10\,000$$

b Here we are *not* given the value of the interest, I, but the value of the total investment, A.

Use the formula $P = \frac{A}{\left(1 + \frac{rt}{100}\right)}$ with $A = \$15\,600$, $r = 5$ and $t = 4$.

$$P = \frac{A}{\left(1 + \frac{rt}{100}\right)} = \frac{15\,600}{\left(1 + \frac{5 \times 4}{100}\right)} = \frac{15\,600}{1.2} = \$13\,000$$

Exercise 2C

Simple interest: calculating interest rate

Example 10 1 Find the annual interest rate if a simple interest investment of $5000 amounts to $6500 in 2.5 years.

2 Find the annual interest rate if a simple interest investment of $500 amounts to $550 in 8 months.

Simple interest: calculating time

Example 11 3 Calculate the time taken for $2000 to earn $975 at 7.5% simple interest.

4 Calculate the time in days for $760 to earn $35 at 4.75% simple interest.

Simple interest: calculating principal

Example 12 5 Calculate the principal that earns $514.25 in 10 years at 4.25% simple interest.

2C 2D Compound interest 55

6 Calculate the principal that earns $780 in 100 days at 6.25% per annum simple interest.

Simple interest: mixed problems

7 Calculate the answers to complete the following table.

Principal	Rate	Time	Simple interest	Total investment
$600	6%	5 years	a	b
$880	6.5%	c	$171.60	d
$1290	e	6 months	$45.15	f
g	10%	4 months	$150.00	h
$3600	i	200 days	$98.63	j
$980	7.5%	k	l	$1200.50
m	7.25%	6 months	$52.50	n

Applications

8 If Geoff invests $30 000 at 10% per annum simple interest until he has $42 000, for how many years will he need to invest the money?

9 Josh decides to put $5000 into an investment account that pays 5.0% per annum simple interest. If he leaves the money there until it doubles, how long will this take, to the nearest month?

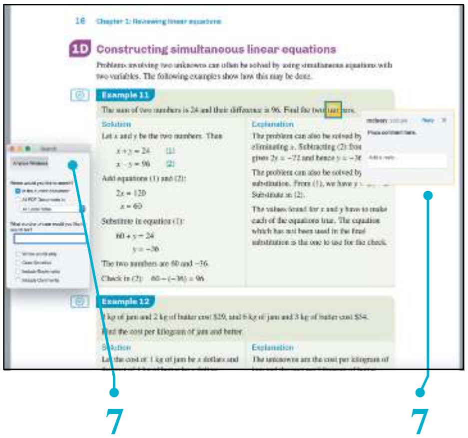
10 A personal loan of $15 000 over a 3-year period costs $500 per month to repay.
  a How much money will be repaid in total?
  b How much of the money repaid is interest?

2D Compound interest

Skill-sheet

We have seen that simple interest is calculated *only* on the original amount borrowed or invested. A more common form of interest, known as **compound interest**, calculates the interest on the original amount plus any interest accrued to that time.

Overview of the downloadable PDF textbook



16 Chapter 1: Reviewing linear equations

1D Constructing simultaneous linear equations

Problems involving two unknowns can often be solved by using simultaneous equations with two variables. The following examples show how this may be done.

Example 11

The sum of two numbers is 24 and their difference is 96. Find the two numbers.

Solution

Let x and y be the two numbers. Then

$x + y = 24$ (1)

$x - y = 96$ (2)

Add equations (1) and (2):

$2x = 120$

$x = 60$

Substitute in equation (1):

$60 + y = 24$

$y = -36$

The two numbers are 60 and −36.

Check in (2): $60 - (-36) = 96$

Explanation

The problem can also be solved by eliminating x. Subtracting (2) from (1) gives $2y = -72$ and hence $y = -36$.

The problem can also be solved by substitution. From (1), we have y … Substitute in (2).

The values found for x and y have to make each of the equations true. The equation which has not been used in the final substitution is the one to use for the check.

Example 12

3 kg of jam and 2 kg of butter cost $29, and 6 kg of jam and 3 kg of butter cost $54.

Find the cost per kilogram of jam and butter.

Solution

Let the cost of 1 kg of jam be x dollars and

Explanation

The unknowns are the cost per kilogram of

6 The convenience of a downloadable PDF textbook has been retained for times when users cannot go online.

7 PDF annotation and search features are enabled.

Overview of the Interactive Textbook

The **Interactive Textbook (ITB)** is an online HTML version of the print textbook powered by the HOTmaths platform, included with the print book or available as a separate purchase.

8 The material is formatted for on screen use with a convenient and easy-to-use navigation system and links to all resources.

9 **Workspaces** for all questions, which can be enabled or disabled by the teacher, allow students to enter working and answers online and to save them. Input is by typing, with the help of a symbol palette, handwriting and drawing on tablets, or by uploading images of writing or drawing done on paper.

10 **Self-assessment tools** enable students to check answers, mark their own work, and rate their confidence level in their work. This helps develop responsibility for learning and communicates progress and performance to the teacher. Student accounts can be linked to the learning management system used by the teacher in the Online Teaching Suite, so that teachers can review student self-assessment and provide feedback or adjust marks.

11 All worked examples have **video versions** to encourage independent learning.

12 **Worked solutions** are included and can be enabled or disabled in the student ITB accounts by the teacher.

13 An expanded and revised set of **Desmos interactives** and activities based on embedded graphics calculator and geometry tool windows demonstrate key concepts and enable students to visualise the mathematics.

14 The **Desmos graphics calculator, scientific calculator**, and **geometry tool** are also embedded for students to use for their own calculations and exploration.

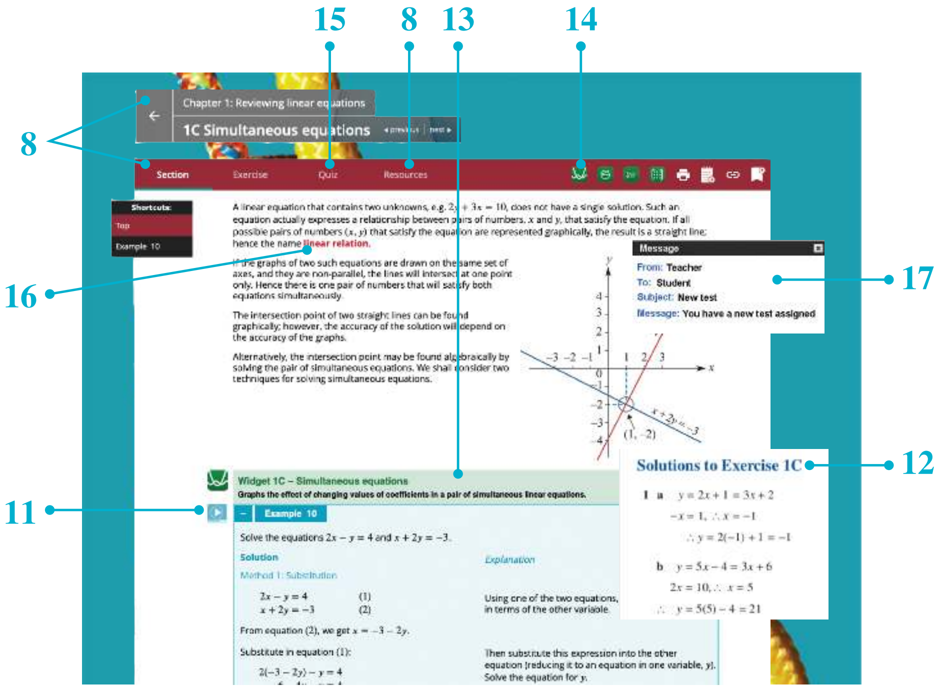
15 **Quick quizzes** containing automarked multiple-choice questions have been thoroughly expanded and revised, enabling students to check their understanding.

16 **Definitions** pop up for key terms in the text, and are also provided in a dictionary.

17 Messages from the teacher assign tasks and tests.

INTERACTIVE TEXTBOOK POWERED BY THE HOTmaths PLATFORM

A selection of features is shown. Numbers refer to the descriptions on pages xi–xii. HOTmaths platform features are updated regularly



A linear equation that contains two unknowns, e.g. $2y + 3x = 10$, does not have a single solution. Such an equation actually expresses a relationship between pairs of numbers, x and y, that satisfy the equation. If all possible pairs of numbers (x, y) that satisfy the equation are represented graphically, the result is a straight line; hence the name **linear relation**.

If the graphs of two such equations are drawn on the same set of axes, and they are non-parallel, the lines will intersect at one point only. Hence there is one pair of numbers that will satisfy both equations simultaneously.

The intersection point of two straight lines can be found graphically; however, the accuracy of the solution will depend on the accuracy of the graphs.

Alternatively, the intersection point may be found algebraically by solving the pair of simultaneous equations. We shall consider two techniques for solving simultaneous equations.

From: Teacher
To: Student
Subject: New test
Message: You have a new test assigned

Graphs the effect of changing values of coefficients in a pair of simultaneous linear equations.

Solve the equations $2x - y = 4$ and $x + 2y = -3$.

Solution

Method 1: Substitution

$2x - y = 4$ (1)
$x + 2y = -3$ (2)

From equation (2), we get $x = -3 - 2y$.

Substitute in equation (1):

$2(-3 - 2y) - y = 4$

Explanation

Using one of the two equations, in terms of the other variable.

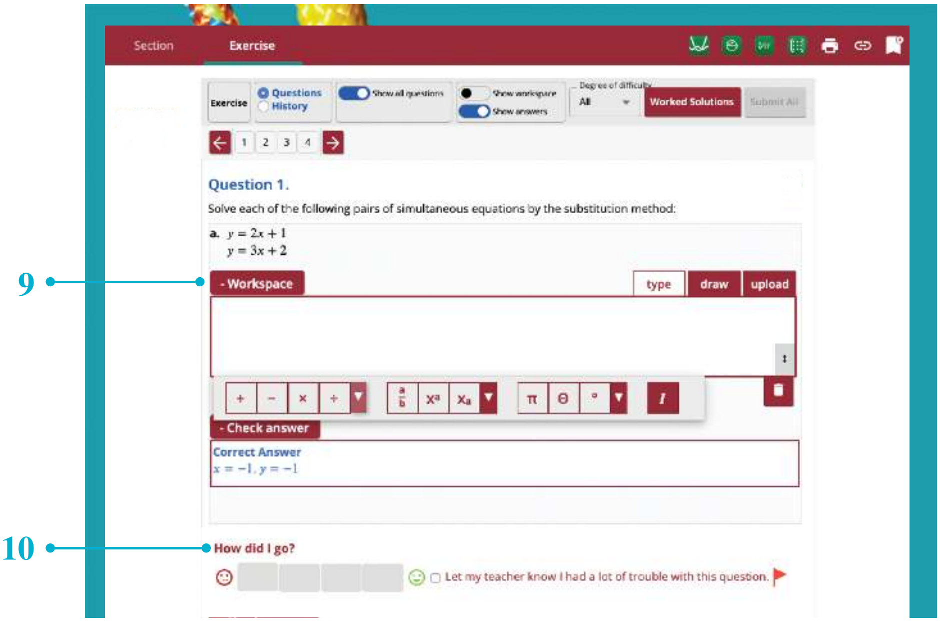
Then substitute this expression into the other equation (reducing it to an equation in one variable, y). Solve the equation for y.

Solutions to Exercise 1C

1 a $y = 2x + 1 = 3x + 2$
$-x = 1, \therefore x = -1$
$\therefore y = 2(-1) + 1 = -1$

b $y = 5x - 4 = 3x + 6$
$2x = 10, \therefore x = 5$
$\therefore y = 5(5) - 4 = 21$

WORKSPACES AND SELF-ASSESSMENT



Question 1.

Solve each of the following pairs of simultaneous equations by the substitution method:

a. $y = 2x + 1$
$y = 3x + 2$

Correct Answer
$x = -1, y = -1$

☐ Let my teacher know I had a lot of trouble with this question.

ISBN 978-1-009-25917-0

Overview of the Online Teaching Suite powered by the HOTmaths platform

The Online Teaching Suite is automatically enabled with a teacher account and is integrated with the teacher's copy of the Interactive Textbook. All the teacher resources are in one place for easy access. The features include:

18 The HOTmaths learning management system with class and student analytics and reports, and communication tools.

19 Teacher's view of a student's working and self-assessment which enables them to modify the student's self-assessed marks, and respond where students flag that they had difficulty.

20 A HOTmaths-style test generator.

21 A suite of chapter tests and assignments.

22 Editable curriculum grids and teaching programs.

23 A brand-new **Exam Generator**, allowing the creation of customised printable and online trial exams (see below for more).

More about the Exam Generator

The Online Teaching Suite includes a comprehensive bank of SCSA exam questions, augmented by exam-style questions written by experts, to allow teachers to create custom trial exams.

Custom exams can model end-of-year exams, or target specific topics or types of questions that students may be having difficulty with.

Features include:

- Filtering by question-type, topic and degree of difficulty
- Searchable by key words
- Answers provided to teachers
- Worked solutions for all questions
- SCSA marking scheme
- All custom exams can be printed and completed under exam-like conditions or used as revision.

ISBN 978-1-009-25917-0

Acknowledgements

The author and publisher wish to thank the following sources for permission to reproduce material:

Cover: © Getty Images / Richard Drury

Images: © Getty Images / MR.Cole_Photographer, Chapter 1 Opener / Yuichiro Chino, Chapter 2 Opener / diyun Zh, Chapter 3 Opener / Oxygen, Chapter 4 & 5 Opener / Abstract Aerial Art, Chapter 6 Opener / author unknown (Getty), Chapter 7 Opener / mfto, Chapter 8 Opener / Marco Bottigelli, Chapter 9 Opener / boonchai wedmakawandm Chapter 10 Opener / author unknown (Getty Images), Chapter 1 review; © Google Maps, Example 11, p.226; gl0ck, 6A (2).

Subtopics in chapter openers © School Curriculum and Standards Authority (2017) Mathematics Applications Year 12 Syllabus. The School Curriculum and Standards Authority does not endorse this publication or product.

Every effort has been made to trace and acknowledge copyright. The publisher apologises for any accidental infringement and welcomes information that would redress this situation.

1 Introduction to associations between two variables

In this chapter

Syllabus references

Topics: Identifying and describing associations between two categorical variables; Identifying and describing associations between two numerical variables; Fitting a linear model to numerical data

Subtopics: 3.1.1 – 3.1.5, 3.1.8

The statistical investigation process

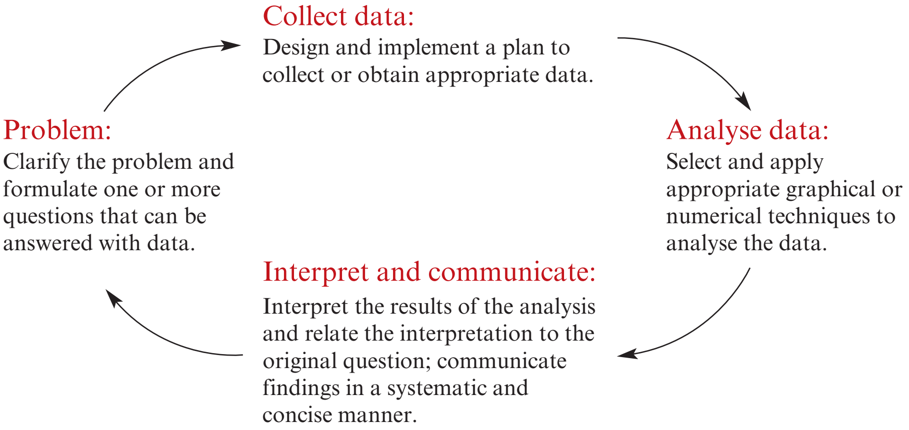
We have access to so much information about the modern world. Industry and governments rely upon an understanding of this information in order to plan for the future or to improve the work they do and how they do it. The environment in which we live benefits from analysing information about the effects of human habitation on other species, and the information gathered from medical research can have rewards in supporting our health.

Statisticians are specialist mathematicians who are skilled in the collection and analysis of this information, or **data**, about real-world situations. They work in particular ways to ensure that the information collected is valid and that the analysis of this data results in meaningful conclusions. This can be described as the **statistical investigation process**.

The statistical investigation process begins with the need to solve a real-world problem. Data related to this problem are collected, analysed using a variety of techniques and then interpreted. The interpretation of the data results in conclusions made about the original problem, which must then be communicated and explained to those involved with that problem.

The statistical investigation process is a cyclic one, which means that it might be necessary to repeat it over and over again before the problem is solved. In the course of solving one problem, a conclusion could be drawn that requires further investigation, which means more data is collected, analysed and interpreted.

The diagram below shows one way of representing the statistical investigation process.



Collect data:
Design and implement a plan to collect or obtain appropriate data.

Problem:
Clarify the problem and formulate one or more questions that can be answered with data.

Analyse data:
Select and apply appropriate graphical or numerical techniques to analyse the data.

Interpret and communicate:
Interpret the results of the analysis and relate the interpretation to the original question; communicate findings in a systematic and concise manner.

In earlier studies of mathematics, you will have learned some techniques for collecting and analysing data. In this course, you will be exploring questions about the relationships between **variables**. Examples of these types of questions are ‘Is a new treatment for a cold more effective than the old treatment?’ or ‘Do younger people spend more time using social media than older people?’ You will represent the data collected graphically and use specific techniques for the analysis of **bivariate data** to analyse and draw conclusions from this analysis. Most importantly, you will learn how to interpret the results of this analysis and how to report these findings to other people.

To begin this analysis, the role of the variables needs to be clearly identified.

1A Response and explanatory variables

When two variables are linked together in some way (associated) so that they vary together, the data generated is called bivariate data.

When we analyse bivariate data, we try to answer questions such as: ‘Is there a relationship between these two variables?’ and more specifically ‘Does knowing the value of one of the variables tell us anything about the value of the other variable?’

For example, let us take as our two variables the *mark* a student obtained on a test and the amount of *time* they spent studying for that test. It seems reasonable that the more time they spend studying, the better mark they will achieve. That is, the amount of *time* spent studying may help to explain the *mark* obtained. For this reason, in statistics we call this the **explanatory variable (EV)**. And, since the *mark* may go up or down in response to the amount of *time* spent studying, we call *mark* the **response variable (RV)**. In general, the value of the explanatory variable for a case is thought to partially explain the value of the response variable for that individual.

Response and explanatory variables

When investigating associations (relationships between two variables), the explanatory variable (EV) is the variable we expect to explain or predict the value of the response variable (RV).

Note: The explanatory variable is sometimes called the independent variable (IV) and the response variable the dependent variable (DV).

Identifying the response and explanatory variables

It is important to be able to identify the explanatory and response variables before starting to explore the relationship between two numerical variables. Consider the following examples.

Example 1 Identifying the response and explanatory variables

We wish to investigate the question ‘Do older people sleep less?’ The variables here are *age* and *time spent sleeping*. Which is the response variable (RV) and which is the explanatory variable (EV)?

Solution

When looking to see if the length of time people spent sleeping is explained by their age, *age* is the EV and *time spent sleeping* is the RV.

EV: age
RV: time spent sleeping

ISBN 978-1-009-25917-0

Example 2 Identifying the response and explanatory variables

We wish to investigate the relationship between kilojoule consumption and weight loss. The variables in the investigation are *kilojoule consumption* and *weight loss.* Which is the response variable (RV) and which is the explanatory variable (EV)?

Solution

Since we are looking to see if the weight loss can be explained by the amount people eat, *kilojoule consumption* is the EV and *weight loss* is the RV.

EV: kilojoule consumption
RV: weight loss

Example 3 Identifying the response and explanatory variables

We wish to investigate the question 'Can we predict people's height from their wrist circumference?' The variables in this investigation are *height* and *wrist circumference*. Which is the response variable (RV) and which is the explanatory variable (EV)?

Solution

Since we wish to predict height from wrist circumference, *wrist circumference* is the EV. *Height* is then the RV.

EV: wrist circumference
RV: height

It is important to note that, in Example 3, we could have asked the question the other way around, that is 'Can we predict people's wrist circumference from their height?' In that case *height* would be the EV and *wrist circumference* would be the RV. The way we ask our statistical question is an important factor when there is no obvious EV and RV.

Example 4 Identifying the response and explanatory variables

We wish to investigate the question, 'Does the time it takes a student to get to school depend on their mode of transport?' The variables here are *time* and *mode of transport*. Which is the response variable (RV) and which is the explanatory variable (EV)?

Solution

In asking the question in this way we are suggesting that a student's *mode of transport* might explain the differences we observe in the time it takes students to get to school.

EV: mode of transport
RV: time

ISBN 978-1-009-25917-0

Exercise 1A

Identifying the response and explanatory variables

Example 1 **1** Identify the explanatory variable and response variable in each of the following situations. The variable names are italicised.

- **a** We wish to investigate the relationship between the *age* of a certain type of tree and its *diameter*. We want to be able to predict the diameter of a tree from its age.
- **b** A study is to be made of the relationship between *weight loss* and the number of *weeks* a person is on a diet.
- **c** Data is collected to investigate the relationship between *age* of a second-hand textbook and its selling *price*.
- **d** The relationship between the number of *hours* a gas heating system is used and the *amount* of gas used is to be investigated.
- **e** A study is to be made of the relationship between the number of *runs* a cricketer scores and the number of *balls bowled* to them.
- **f** We wish to investigate whether a fish's *toxicity* can be predicted from its *colour*. We want to be able to predict *toxicity* from *colour*.
- **g** The relationship between *weight loss* and type of *diet* is to be investigated.
- **h** It is suggested that the *cost* of heating in a house depends on the type of *fuel* used.
- **i** The relationship between the *price* of a house and its *location* is to be investigated.

2 The following pairs of variables are related. Which is likely to be the explanatory variable? The variable names are italicised.

- **a** *exercise level* and *age*
- **b** *years of education* and *salary level*
- **c** *comfort level* and *temperature*
- **d** *time of year* and *incidence of hay fever*
- **e** *age group* and *musical taste*
- **f** *AFL team supported* and *state of residence*

ISBN 978-1-009-25917-0

1B Investigating associations between categorical variables

If two variables are related or linked in some way, we say they are **associated**. The statistical tool used to investigate associations between two **categorical variables** is a **two-way frequency table**.

Using a two-way frequency table to investigate an association

Analysing a situation where two categorical variables are associated

Table 2.1 was constructed from data collected to answer the question: ‘Is a person’s attitude to gun control associated with their sex?’

In this investigation, *attitude* to gun control is the response variable and *sex* is the explanatory variable.

Table 2.1

	Sex		
Attitude	*Male*	*Female*	*Total*
For	32	30	62
Against	26	12	38
Total	58	42	100

RV

EV

The number ‘100’ in the bottom corner of the table is the sum of the totals in the last column. It is also the sum of the totals in the bottom row. This is called the **grand sum** and can be a useful check that the table contains the correct numbers because these two sums should be the same.

The grand sum is the total number of people surveyed.

From the table we see that more males than females favoured gun control. However, this doesn’t tell us very much, because there were more males in the sample.

To solve this problem, we turn our table entries into percentages by calculating column percentages. See Table 2.2.

Table 2.2

	Sex (%)		
Attitude	*Male*	*Female*	*Total %*
For	55.2	71.4	62.0
Against	44.8	28.6	38.0
Total	100.0	100.0	100.0

The percentage of:

- males who are for gun control $= \dfrac{32}{58} \times 100\% = 55.2\%$
- females who are for gun control $= \dfrac{30}{42} \times 100\% = 71.4\%$
- males who are against gun control $= \dfrac{26}{58} \times 100\% = 44.8\%$
- females who are against gun control $= \dfrac{12}{42} \times 100\% \ = 28.6\%$

Observing and comparing the percentage values in a two-way frequency table can help decide if there is an association between the variables.

If there is *no association* between *attitude* to gun control and *sex*, approximately *equal percentages* of males and females would be ‘for’ gun control.

This is not the case, so we conclude that *attitude* to gun control and *sex* are *associated*.

To complete our analysis we might write a brief report of our finding.

Report

From Table 2.2 we see that a much higher percentage of females were for gun control than males, 71.4% to 55.2%, indicating that attitude to gun control is associated with sex.

Example 5 Identifying and describing associations between two categorical variables from a two-way frequency table

A survey was conducted with 100 people.

As part of this survey, people were asked whether or not they supported banning mobile phones in cinemas. The results are summarised in the table.

Ban mobile phones	*Sex*	
	Male	*Female*
Yes	87.9%	65.8%
No	12.1%	34.2%
Total	100.0%	100.0%

Is there is an association between support for banning mobile phones in cinemas and the sex of the respondent? Write a brief response quoting appropriate percentages.

Solution

A large difference in the percentages of males and females supporting the banning of mobile phones indicates an association.

Yes; the percentage of males in support of banning mobile phones in cinemas (87.9%) was much higher than for females (65.8%).

Note: Finding a single row in the two-way frequency distribution in which percentages are clearly different is sufficient to identify an association between the variables.

In the next example, there is *no* association between the variables.

Example 6 Identifying and describing associations between two categorical variables from a two-way frequency table (no association)

In the same survey people were asked whether or not they supported Sunday racing. The results are summarised in the table.

Is there an association between support for Sunday racing and the sex of the respondent? Write a brief response quoting appropriate percentages.

	Sex	
Sunday racing	*Male*	*Female*
For	55.6%	54.5%
Against	44.4%	45.5%
Total	100.0%	100.0%

Solution

The similar percentage of males and females supporting the banning of Sunday racing is consistent with there being no association.

No; the percentage of males (55.6%) supporting Sunday racing is similar to the percentage of females supporting Sunday racing (54.5%).

Note: As a rule of thumb, a difference of at least 5% would be required to classify a difference as significant.

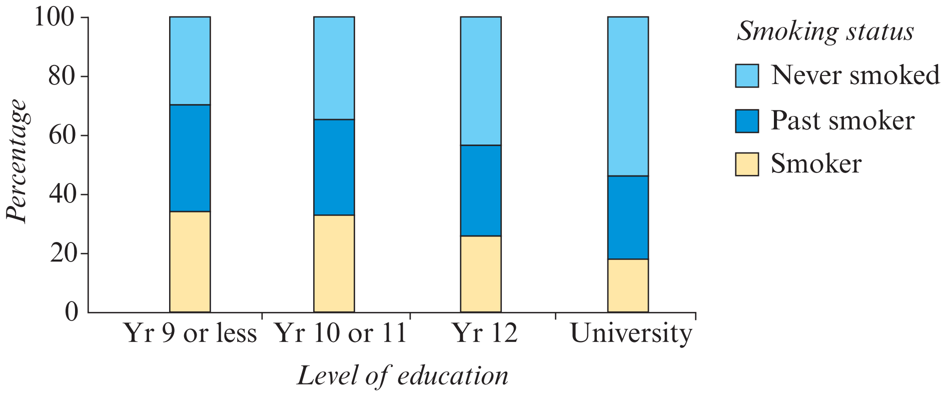
Analysing two categorical variables with multiple categories

Table 2.3 displays the smoking status of adults (smoker, past smoker, never smoked) by level of education (Year 9 or less, Year 10 or 11, Year 12, university).

Table 2.3

	Education level (%)			
Smoking status	*Year 9 or less*	*Year 10 or 11*	*Year 12*	*University*
Smoker	33.9	31.7	26.5	18.4
Past smoker	36.0	33.8	30.9	28.0
Never smoked	30.1	34.5	42.6	53.6
Total	100.0	100.0	100.0	100.0

The following **segmented bar chart** has been constructed from this table to help us with the analysis. Each column represents a column from the purple-shaded part of the table.



From the segmented bar chart, it is clear that as education level increases the percentage of smokers decreases, indicating there is an association between smoking and education level. We could report this finding as follows, using the table to determine exact percentages.

Report

From Table 2.3 we see that the percentage of smokers steadily decreases with education level, from 33.9% for Year 9 or below to 18.4% for university. This indicates that smoking is associated with level of education.

Note: A similar conclusion could be drawn by focusing attention on the top segment of each column, which shows that the percentage of non-smokers increases with education level.

Example 7 Identifying and describing associations between two categorical variables from a two-way frequency table

A survey was conducted with 1000 males under 50 years old. As part of this survey, they were asked to rate their interest in sport as 'high', 'medium', and 'low'. Their age group was also recorded as 'under 18','19–25', '26–35' and '36–50'. The results are displayed in the table.

	Age group (%)			
Interest in sport	*Under 18 years*	*19–25 years*	*26–35 years*	*36–50 years*
High	56.5	50.2	40.7	35.0
Medium	30.1	34.4	36.8	44.7
Low	13.4	15.4	22.5	20.3
Total	100.0	100.0	100.0	100.0

a Which is the explanatory variable, *interest in sport* or *age group*?

b Is there an association between *interest in sport* and *age group*? Write a brief response quoting appropriate percentages.

Solution

a In this situation, *age group* is the obvious explanatory variable.

Age group

b A significant difference in the percentages of the various age groups for any particular level of interest in sport indicates an association. We will choose those with a 'high level' of interest for analysis.

Yes; the percentage of males with a high level of interest in sport decreases with age group from 56.5% for the 'under 18 years' age group to 35.0% for the '36–50 years' age group.

Deciding whether to use row or column percentages

In most surveys or sources of data collection, the sum of the row and column totals do not equal 100. In these situations, we want to convert the raw data into percentages as we did in Table 2.2.

	Sex (%)		
Attitude	*Male*	*Female*	*Total*
For	55.2	71.4	62.0
Against	44.8	28.6	38.0
Total	100.0	100.0	100.0

ISBN 978-1-009-25917-0

Cambridge University Press

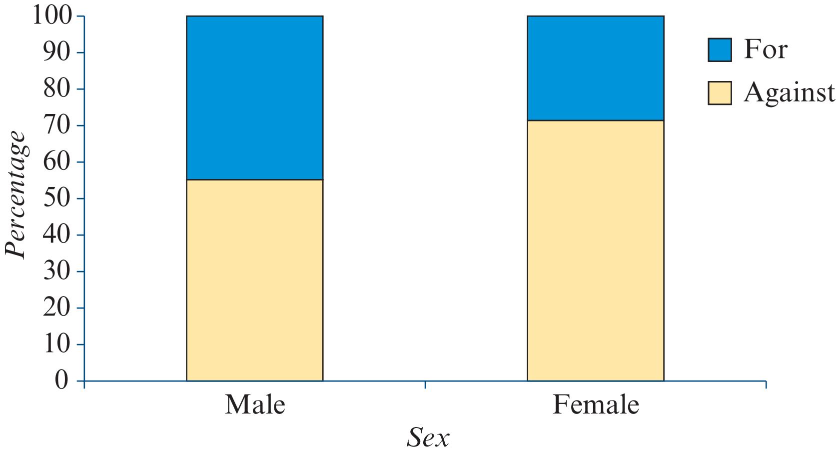
Updated January 2024

However, it is important to correctly decide whether to calculate row or column percentages.

We use the *explanatory variable* to make this decision:

- If the explanatory variable uses **rows** to show its different categories then use **row percentages**.
- If the explanatory variable uses **columns** to show its different categories then use **column percentages**.

For example, the original question for Table 2.2 was ‘Is a person’s attitude to gun control associated with their sex?” That is, given a person is male or female, what is their attitude to gun control? Therefore, the percentages need to be the distribution of attitude towards gun control based on the condition that the person is male or female. These are called **conditional percentages** and can be seen in the stacked column graph below.



Conditional percentages are calculated separately for each value of the explanatory variable. When we try to understand the association between two categorical variables, we compare the distributions of the response variable for values of the explanatory variable. In particular, we look at how the pattern of conditional percentages differs between the values of the explanatory variable.

Note: It is important to identify the explanatory variable because we always use the totals for the explanatory variable to calculate the percentages.

Example 8 Deciding whether to use row or column percentages

An online survey questioned people of different ages about their listening preferences whilst exercising. The ages of the participants were grouped by age as either ‘younger’ (aged 18–34 years), ‘middle-aged’ (35–54 years) and ‘older’ (55–65+ years). The results are shown in the table below.

Listening Preference	*Age*			*Total*
	Younger	*Middle-aged*	*Older*	
Pop	94	116	31	241
Hip-hop	35	8	1	44
Country	7	1	7	15
Total	136	125	39	300

a What is the explanatory variable and what is the response variable?

b Construct a percentage two-way table, using row or column percentages as appropriate. Round to the nearest whole per cent.

c Is there an association between listening preference during exercise and age group? Write a brief response quoting appropriate percentages.

Solution

a In considering this study we are asking if *age* might explain the differences in listening preference during exercise.

Explanatory variable: age group
Response variable: listening preference

b Since the explanatory variable is age group and these are the columns of the table, use column percentages to construct the percentage two-way table.

For example:

- Percentage of 'younger' people who listen to pop = $\frac{94}{136} \times 100 = 69\%$
- Percentage of 'younger' people who listen to hip-hop = $\frac{35}{136} \times 100 = 26\%$
- Percentage of 'younger' people who listen to country = $\frac{7}{136} \times 100 = 5\%$

Column percentage two-way table:

Listening Preference	Age (%)		
	Younger	Middle-aged	Older
Pop	69	93	79
Hip-hop	26	6	3
Country	5	1	18
Total	100	100	100

ISBN 978-1-009-25917-0

Cambridge University Press

Updated January 2024

c There is a significant difference between the percentages of the various age groups for each of the listening preferences. If there was no association then we would expect the percentage of each of the listening preferences to be the same in each of the age groups.

Yes, there is an association between age group and exercise, since 69% of younger people listen to pop, whilst this increases to 93% of middle-aged people.

Example 9 Deciding whether to use row or column percentages

A study is being conducted to determine if there is an association between the economic status of a passenger (crew, or first, second, or third class) on the Titanic and their survival status.

	Survival status		
	Yes	*No*	*Total*
Crew	212	673	885
First Class	203	122	325
Second Class	118	167	285
Third Class	178	528	706
Total	711	1490	2201

a What is the explanatory variable and what is the response variable?

b Explain whether row or column percentages would be used to help identify if there is an association between the variables.

c Construct an appropriately percentaged two-way table that can be used to help identify whether economic status can explain the survival status of the passenger. Round all percentages to the nearest whole number.

d Does the survival of passengers on the Titanic appear to be associated with economic status? Write a brief response quoting appropriate percentages.

Solution

a In considering this study we are asking if economic status might explain the survival status of the passenger

Explanatory variable: Economic status
Response variable: Survival status

b The use of row or column percentages is determined by the location of the explanatory variable in the two-way table

The explanatory variable (economic status) uses rows to show its different categories, so use row percentages.

c Since row percentages are being used, percentages must be calculated out of the total for each row.
For example:

- Percentage of crew who survived $= \frac{212}{885} \times 100 = 24\%$
- Percentage of crew who did not survive $= \frac{673}{885} \times 100 = 76\%$

	Survival status (%)		
	Yes	No	Total
Crew	24	76	100
First Class	62	38	100
Second Class	41	59	100
Third Class	25	75	100
Total	32	68	100

d There is a significant difference between the percentages for each economic status and the survival status.
If there was no association then we would expect the percentage of survival for each of the economic status' to be the same.

Yes there is an association.
There is a higher survival rate in first class (62%), then decreasing to second class (41%) and lowest survival rate in third class (25%).
Crew had the lowest survival rate (24%) compared to all three other economic status'.

Exercise 1B

Using two-way frequency tables to identify associations between two categorical variables

1 A survey was conducted with 242 university students. For this survey, data were collected on the students' *enrolment status* (full-time, part-time) and whether or not each *drinks alcohol* ('Yes' or 'No'). Their responses are summarised in the table opposite.

Drinks alcohol	*Enrolment status (%)*	
	Full-time	*Part-time*
Yes	80.5	81.8
No	19.5	18.2
Total	100.0	100.0

a Which variable is the explanatory variable?

b Is there an association between drinking alcohol and enrolment status? Write a brief report quoting appropriate percentages.

2 The table opposite was constructed from data collected to see if *handedness* (left, right) was associated with *sex* (male, female).

	Sex	
Handedness	*Male*	*Female*
Left	22	16
Right	222	147

a Which variable is the response variable?

b Convert the table to percentages by calculating column percentages.

c Is *handedness* associated with *sex*? Write a brief explanation using appropriate percentages.

Example 5 **3** A survey was conducted with 59 male and 51 female university students to determine whether, each day, they exercised, ‘regularly’, ‘sometimes’ or ‘rarely’. Their responses are summarised in the table.

	Sex	
Exercised	*Male*	*Female*
Rarely	17	20
Sometimes	31	28
Regularly	11	3
Total	59	51

- **a** Which is the explanatory variable?
- **b** What percentage of females exercised sometimes?
- **c** Construct a percentage two-way table, using row or column percentages as appropriate.
- **d** Is there an association between how regularly these students exercised and their sex? Write a brief response quoting appropriate percentages.

Example 7 **4** As part of the General Social Survey conducted in the US, respondents were asked to say whether they found life exciting, pretty routine or dull. Their marital status was also recorded as married, widowed, divorced, separated or never married. The results are organised into a table as shown.

	Marital status (%)				
Attitude to life	*Married*	*Widowed*	*Divorced*	*Separated*	*Never*
Exciting	47.6	33.8	46.7	45.9	52.3
Pretty routine	48.7	54.3	47.6	44.6	44.4
Dull	3.7	11.9	5.7	9.5	3.3
Total	100.0	100.0	100.0	100.0	100.0

- **a** What percentage of widowed people found life ‘dull’?
- **b** What percentage of people who were never married found life ‘exciting’?
- **c** What is the explanatory variable in this investigation?
- **d** Does the information you have been given support the contention that a person’s attitude to life is related to their marital status? Justify your argument by quoting appropriate percentages.

Constructing a two-way frequency table

5 A survey asked students about their school group (primary, secondary) and their engagement in competitive sport (play, do not play). The survey showed that:

- 125 primary school students play sport and 28 do not
- 36 secondary school students play sport and 11 do not

	Group		
Engagement	*Primary*	*Secondary*	*Total*
Play			
Do not play			
Total			

a Copy and complete the two-way frequency table shown above.

b How many students in total were surveyed?

c How many surveyed students in total play competitive sport?

d Construct an appropriate percentage two-way table given that *school group* is the explanatory variable.

e Which group of students appears to play more sport?

6 A survey was conducted with 350 people at an airport. Data were collected on the *sex* (male, female) and *attitude* to flying (enjoy, do not enjoy, have never flown) for each person in the sample.

Of the females in the sample: 52 enjoy flying, 30 do not enjoy flying and 4 have never flown.

Of the males in the sample: 164 enjoy flying, 86 do not enjoy flying and 14 have never flown.

a Explain why the variable *attitude* is the response variable for this survey.

b Construct a two-way frequency table with row and column sums.

c Re-draw the two-way table from part **b** showing row or column percentages as appropriate.

d What percentage of males in the sample do not enjoy flying?

e Explain why there does not seem to be an association between *sex* and *attitude* to flying.

1C Scatterplots and their construction

The first step in investigating an association between two **numerical variables** is to construct a visual display of the data, which we call a **scatterplot**.

Constructing a scatterplot manually

We will illustrate the process by constructing a scatterplot of the *mark* students obtained on an examination (the response variable) and the *time* they spent studying for the examination (the explanatory variable).

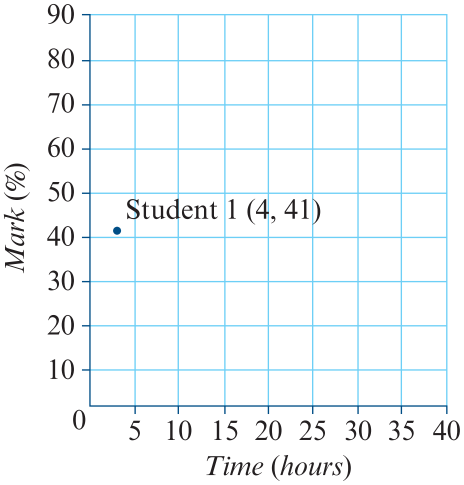
Student	1	2	3	4	5	6	7	8	9	10
Time (hours)	4	36	23	19	1	11	18	13	18	8
Mark (%)	41	87	67	62	23	52	61	43	65	52

In a scatterplot, each point represents a single case, in this instance a student.

When constructing a scatterplot, it is conventional to use the *vertical* or *y-axis* for the response variable (RV) and the *horizontal* or *x-axis* for the explanatory variable (EV).

- The horizontal or x-coordinate of the point represents the time spent studying.
- The vertical or y-coordinate of the point represents the mark obtained.

The scatterplot opposite shows the point for Student 1, who studied 4 hours for the examination and obtained a mark of 41.

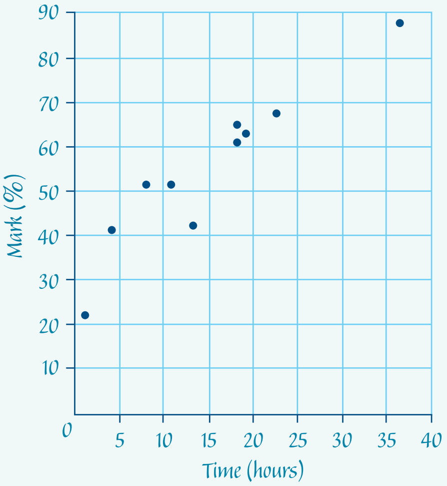


Example 10 Constructing a scatterplot manually

Use the data in the table above to construct a scatterplot manually.

Solution

To complete the scatterplot, the points for each remaining student are plotted as shown.



Example 11 Elements of a scatterplot

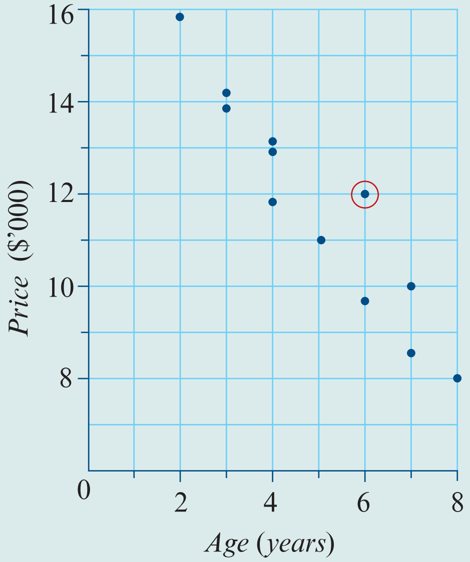
The scatterplot shown has been constructed from data collected to investigate the relationship between the *price* of a second-hand car and its *age*.

Use the scatterplot to answer the following questions.

a Which is the explanatory variable and which is the response variable?

b How many cars are in the data set?

c How old is the car circled? What is its price?



Solution

a The EV will be on the horizontal axis and the RV on the vertical axis.

EV: Age
RV: Price

b The number of cars is equal to the number of points on the scatterplot.

12 Cars

c The x-coordinate of the point will be the car's age and the y-coordinate its price.

The car is 6 years old, and its price is $12 000.

Using a graphics calculator to construct a scatterplot

You need to understand the principles of constructing a scatterplot by hand but a CAS calculator can also do the task.

How to construct a scatterplot using the TI-Nspire CAS

The data below shows the marks that 10 students obtained on an examination and the time they spent studying for the examination.

Time (hours)	4	36	23	19	1	11	18	13	18	8
Mark (%)	41	87	67	62	23	52	61	43	65	52

Use a calculator to construct a scatterplot. Use *time* as the explanatory variable.

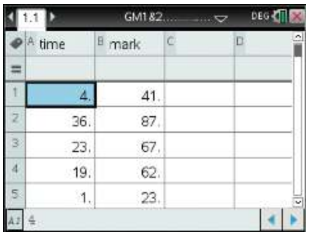
ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

Steps

1 Start a new document (ctrl + N) and select **Add Lists & Spreadsheet**.

Enter the data into lists named **time** and **mark**.



	A time	B mark	C	D
=				
1	4.	41.		
2	36.	87.		
3	23.	67.		
4	19.	62.		
5	1.	23.		

A1 4

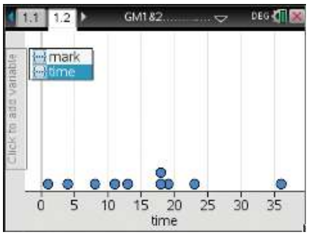
2 Statistical graphing is done through the **Data & Statistics** application.

Press ctrl + I and select **Add Data & Statistics** (or press on and arrow to and press enter).

Note: A random display of dots will appear – this is to indicate list data is available for plotting. It is not a statistical plot.

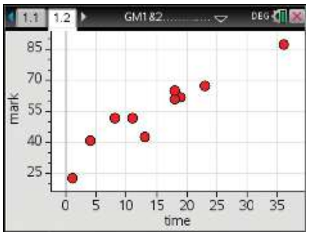
3 To construct a scatterplot.

a Press tab and select the variable **time** from the list. Press enter to paste the variable **time** to the x-axis.



b Press tab again and select the variable **mark** from the list. Press enter to paste the variable **mark** to the y-axis axis to generate the required scatterplot. The plot is automatically scaled.

Note: To change colour, move cursor over the plot and press ctrl + menu**>Color>Fill Color**.



How to construct a scatterplot using the ClassPad

The data below give the marks that 10 students obtained on an examination and the times they spent studying for the examination.

Time (hours)	4	36	23	19	1	11	18	13	18	8
Mark (%)	41	87	67	62	23	52	61	43	65	52

Use a calculator to construct a scatterplot. Use *time* as the explanatory variable.

ISBN 978-1-009-25917-0

Cambridge University Press

Updated January 2024

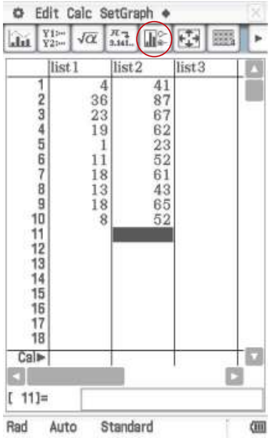
Steps

1 Open the **Statistics** application .

2 Enter the values into lists with **time** in list1 and **mark** in list2.

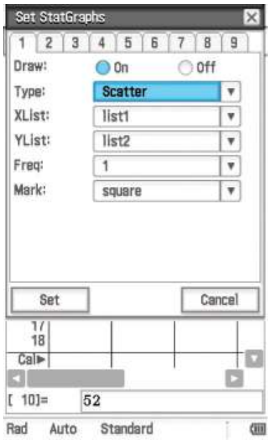
3 Tap to open the **Set StatGraphs** dialog box.

4 Complete the dialog box as shown and Tap SET.



	list1	list2	list3
1	4	41	
2	36	87	
3	23	67	
4	19	62	
5	1	23	
6	11	52	
7	18	61	
8	13	43	
9	18	65	
10	8	52	

[11]=



Draw: ● On ○ Off
Type: Scatter
XList: list1
YList: list2
Freq: 1
Mark: square

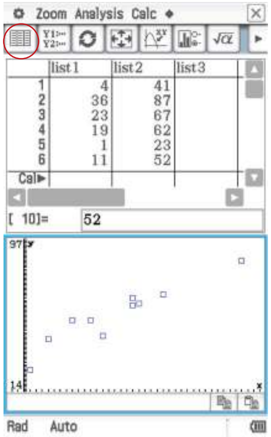
[10]= 52

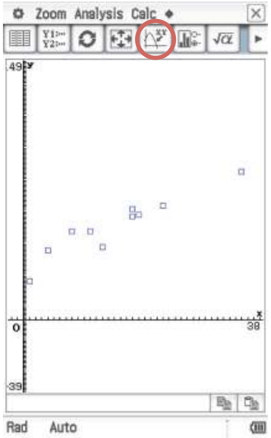
5 Tap to plot a scaled graph in the lower-half of the screen.

6 Tap to give a full-screen sized graph.Tap to return to a half-screen.

7 Tap to place a marker on the first data point ($x_c = 4$, $y_c = 41$).

8 Use the horizontal cursor arrow to move from point to point.





Exercise 1C

Constructing scatterplots

Note: Your teacher may tell you which questions to do manually and which questions to use a calculator.

Example 10

1 The table below shows the height and weight of eight people.

Height (cm)	190	183	176	178	185	165	185	163
Weight (kg)	77	73	70	65	65	65	74	54

Construct a scatterplot with the variable *height* as the explanatory variable and the variable *weight* as the response variable.

ISBN 978-1-009-25917-0

2 The table below shows the ages of 11 couples when they got married.

Age of wife	26	29	27	21	23	31	27	20	22	17	22
Age of husband	29	43	33	22	27	36	26	25	26	21	24

Construct a scatterplot with the variable *wife* (age of wife) as the explanatory variable and the variable *husband* (age of husband) as the response variable.

3 The table below shows the response time of 10 patients (in minutes) given a pain relief drug and the drug dosage (in milligrams).

Drug dosage	0.5	1.2	4.0	5.3	2.6	3.7	5.1	1.7	0.3	0.6
Response time	65	35	15	10	22	16	10	18	70	50

a Which variable is the explanatory variable?

b Construct an appropriate scatterplot.

4 The table below shows the number of people in a cinema at 5-minute intervals after the advertisements started.

Number in cinema	87	102	118	123	135	137
Time	0	5	10	15	20	25

a Which is the explanatory variable?

b Construct an appropriate scatterplot.

Using a CAS calculator to construct a scatterplot

5

Minimum temperature (x)	17.7	19.8	23.3	22.4	22.0	22.0
Maximum temperature (y)	29.4	34.0	34.5	35.0	36.9	36.4

The table above shows the maximum and minimum temperatures (in °C) during a hot week in Melbourne. Using a calculator, construct a scatterplot with minimum temperature as the explanatory variable. Name the variables *min temp* and *max temp*.

6

Balls faced	29	16	19	62	13	40	16	9	28	26	6
Runs scored	27	8	21	47	3	15	13	2	15	10	2

The table above shows the number of runs scored and the number of balls faced by batsmen in a one-day international cricket match. Use a calculator to construct an appropriate scatterplot. Remember to identify the explanatory variable.

Using scatterplots

7 For which of the following pairs of variables would it be appropriate to construct a scatterplot to investigate a possible association?

A Car *colour* (blue, green, black, . . .) and its *size* (small, medium, large)

B A food's *taste* (sweet, sour, bitter) and its *sugar content* (in grams)

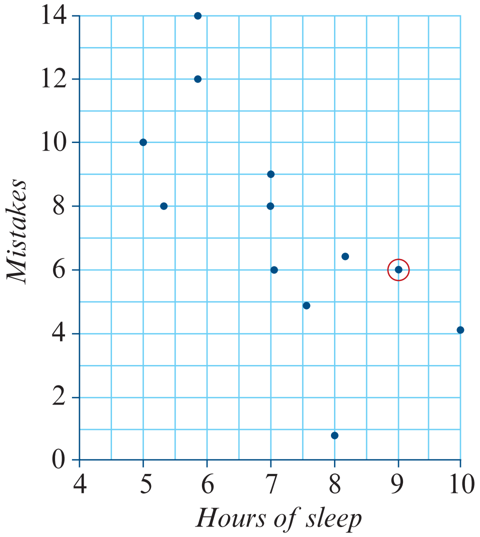
C The *weights* (in kg/cm) of 12 oranges and *lengths* (in centimetres) of nine bananas

D The *time* people spend exercising each day and their *resting pulse rate*

E The *arm spans* (in centimetres) and *gender* (male, female) of a group of students

Example 11

8 The scatterplot shown has been constructed from data collected to investigate the relationship between the amount of sleep a person has the night before a test and the number of mistakes they make on the test. Use the scatterplot to answer the following questions.



a Which is the explanatory variable and which is the response variable?

b How many people are in the data set?

c How much sleep had the individual circled had, and how many mistakes did they make?

Cambridge Senior Maths for Western Australia Mathematics Applications 3&4 ISBN 978-1-009-25917-0 Cambridge University Press
Updated January 2024

Key ideas and chapter summary

Assignment

Bivariate data	**Bivariate data** are data in which each observation involves recording information about two variables for the same person or thing. An example would be the height and weight of the children in a preschool.
Explanatory and response variables	The **explanatory variable** is used to explain or predict the value of the **response variable**.
Association	**Association** is a general term used to describe the relationship between variables.
Two-way frequency tables	**Two-way frequency tables** are used as the starting point for investigating the association between two **categorical variables**.
Grand sum	The **grand sum** is the sum of the row totals, or the column totals, in a two-way frequency table. The grand sum is also the number of data points contained in the frequency table
Identifying associations between two variables	*Associations* between *two categorical variables* are identified by *comparing* appropriate percentages in a two-way frequency table. *Associations* between *two numerical variables* are identified using a scatterplot.
Segmented bar charts	A **segmented bar chart** can be used to graphically display the information contained in a two-way frequency table. It is a useful tool for identifying relationships between two categorical variables.
Scatterplot	A two-dimensional data plot where each point represents the value of two related variables in a bivariate data set. In a scatterplot, the **response variable (RV)** is plotted on the *vertical* axis and the **explanatory variable (EV)** on the *horizontal* axis. A scatterplot is used to help identify and describe the relationship between two **numerical variables**.

Skills check

Having completed this chapter you should be able to:

- identify the explanatory variable and the response variable when investigating association between those variables
- construct two-way frequency tables to investigate association between two categorical variables

- construct two-way frequency tables with percentages to investigate association between two categorical variables
- analyse association between categorical variables using two-way frequency tables and segmented bar charts
- construct a scatterplot from a table of data manually and using a CAS calculator.

Short-answer questions

1 The table below was constructed from data to see if sex (male, female) was associated with playing sport.

	Sex	
Plays sport	*Male*	*Female*
Yes	68	79
No	34	
Total	102	175

a Identify the variables *plays sport* and *sex* as categorical or numerical variables.

b How many females do not play sport?

c Determine the percentage of males who do not play sport, to the nearest per cent.

d Is there any association between a person's sex and the likelihood they play sport? Write a brief response quoting appropriate percentages.

2 The data in the table below was collected from 150 males and 90 females about their attendance at football matches.

Attends football matches	*Sex (%)*	
	Male	*Female*
Regularly	42.8	39.6
Occasionally	35.7	53.5
Never	21.5	6.9
Total	100.0	100.0

a How many of these females said that they never attend football matches?

b How many people in total said that they attend football matches regularly?

c From the information given, is there an association between the variables *attends football matches* and *sex*?

ISBN 978-1-009-25917-0

Cambridge University Press

Updated January 2024

3 For which one of the following pairs of variables would it be appropriate to construct a scatterplot?

A *eye colour* (blue, green, brown, other) and *hair colour* (black, brown, blonde, other)
B *test score* and *sex* (male, female)
C *political party preference* (Labor, Liberal, Other) and *age* in years
D *age* in years and *blood pressure* in mmHg
E *height* in cm and *sex* (male, female)

Extended-response questions

1 One thousand drivers who had an accident during the past year were classified according to age and the number of accidents.

Number of accidents	*Age* < 30	*Age* ≥ 30
At most one accident	130	170
More than one accident	470	230
Total	600	400

a What are the variables shown in the table? Are they categorical or numerical?
b Determine the response and explanatory variables.
c How many drivers under the age of 30 had more than one accident?
d Convert the table values to percentages by calculating the column percentages. Round the percentages to one decimal place.
e Use these percentages to comment on the statement: ‘Of drivers who had an accident in the past year, younger drivers (age < 30) are more likely than older drivers (age ≥ 30) to have had more than one accident.’

ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

2 The data in the table below is based on a study of dolphin behaviour. In this study, the main activities of dolphins observed in the wild were classified as ‘travelling, ‘feeding’, and ‘socialising’.

	Time of observation				
Activity	*Morning*	*Noon*	*Afternoon*	*Evening*	*Total*
Travelling	6	6	14	13	39
Feeding	28	4	0	56	88
Socialising	38	5	9	10	62
Total	72	15	23	79	189

The time of the observation was recorded as ‘morning’, ‘noon’, ‘afternoon’ or ‘evening’.

a How many dolphins were observed feeding at noon?

b What is the dolphin activity most frequently observed in the morning?

c What is the explanatory variable for this study?

d What percentage of the dolphins observed in the evening were socialising? Round your answer to one decimal place.

e Construct a percentage two-way table, using row or column percentages as appropriate from the table on the previous page. Round your answers to one decimal place.

f Does there appear to be an association between dolphin activity and time of observation? Quote appropriate percentages to support your answer.

3 It was suggested that day and evening students differed in their satisfaction with a course in psychology. *Day* students and *evening* students were asked about their satisfaction levels and responded to a survey with *satisfied*, *neutral* or *dissatisfied*. Of the day students, 90 were satisfied, 18 were neutral and 12 were dissatisfied. Of the evening students, 22 were satisfied, 5 were neutral and 3 were dissatisfied.

a Construct a two-way frequency table with row and column sums.

b How many evening students were surveyed?

c Given that *type of student* is the explanatory variable. Construct a percentage two-way table using row or column percentages as appropriate.

d What percentage of day students are satisfied with their course?

e Does there appear to be a relationship between satisfaction with the course and the type of student in the sample?

f Comment on the statement: ‘There was greater satisfaction with the psychology course among day students, as 90 day students were satisfied with the course while only 22 evening students were satisfied.’

4

Temperature (°*C*)	0	10	50	75	100	150
Diameter (cm)	2.00	2.02	2.11	2.14	2.21	2.28

The table above shows the changing diameter of a metal ball as it is heated. Use a calculator to construct an appropriate scatterplot. Temperature is the explanatory variable.

5 The table below shows the number of seats and airspeed (in km/h) of eight aircraft.

Airspeed (*km/h*)	830	797	774	736	757	765	760	718
Number	405	296	288	258	240	193	188	148

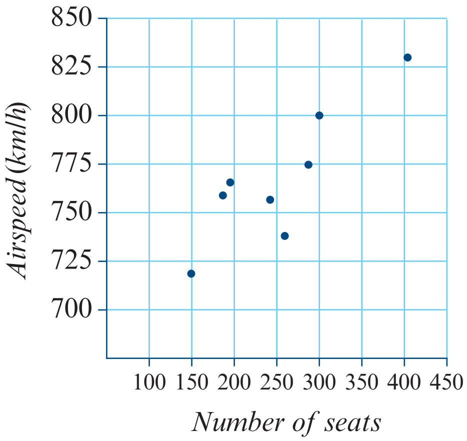
Construct a scatterplot with the variable *number* as the explanatory variable and the variable *airspeed* as the response variable.

6 The scatterplot opposite has been constructed to investigate the association between the airspeed (in km/h) of commercial aircraft and the number of passenger seats.
Use the scatterplot to answer the following questions.

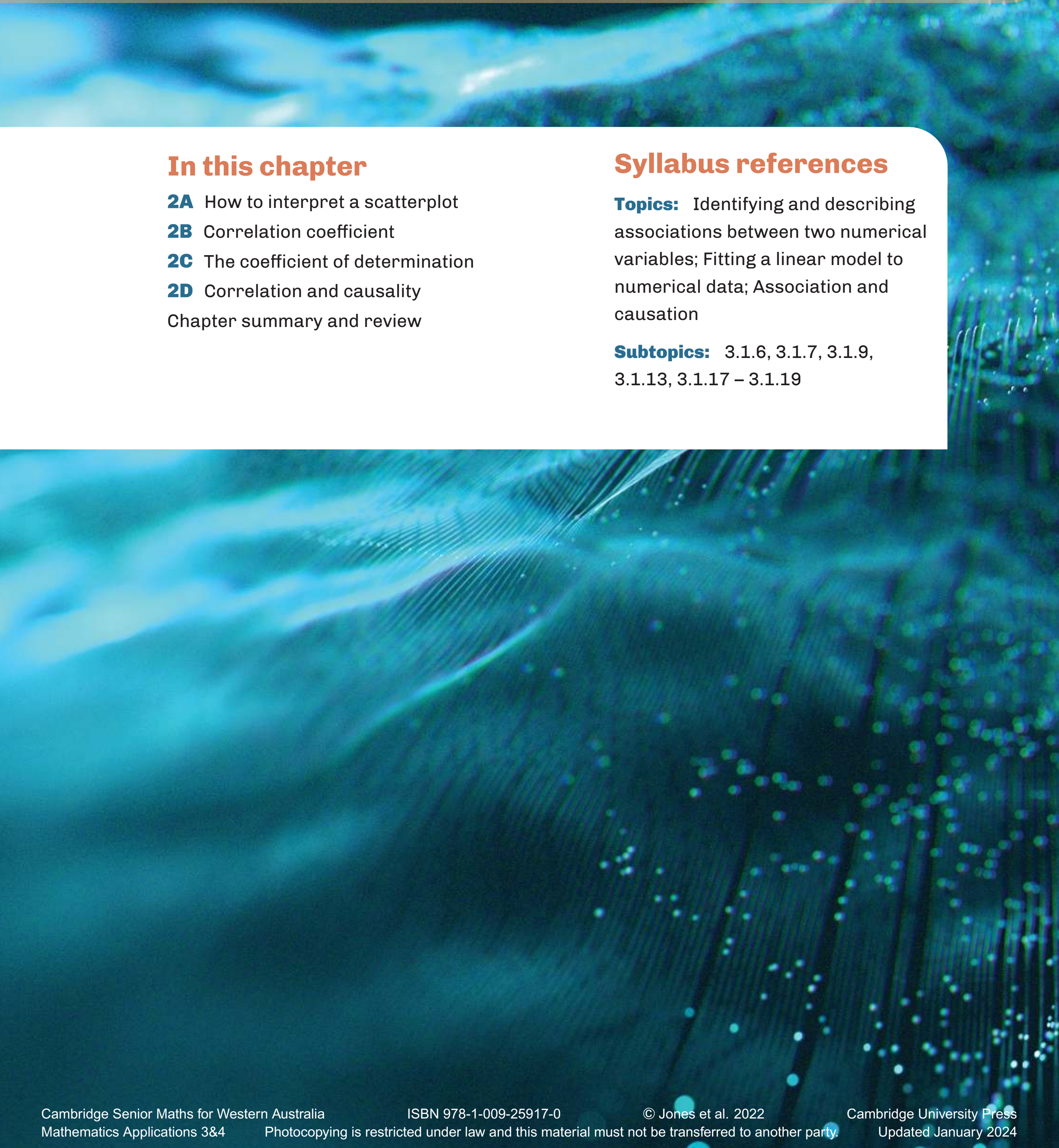
a Which is the explanatory variable?

b How many aircraft were investigated?

c What was the airspeed of the aircraft that could seat 300 passengers?



2 Interpreting associations between two variables



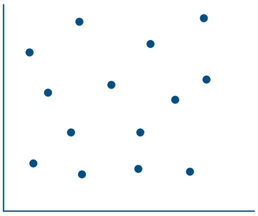
In this chapter

Syllabus references

Topics: Identifying and describing associations between two numerical variables; Fitting a linear model to numerical data; Association and causation

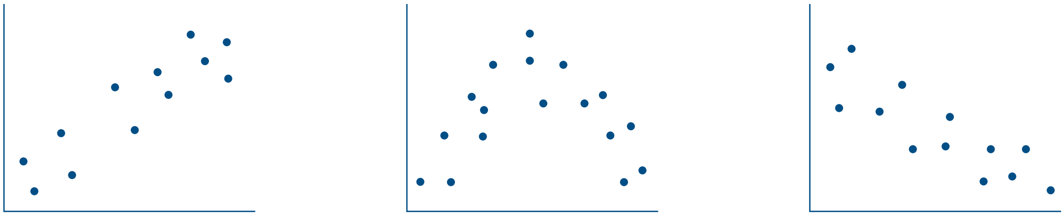
Subtopics: 3.1.6, 3.1.7, 3.1.9, 3.1.13, 3.1.17 – 3.1.19

2A How to interpret a scatterplot



What features do we look for in a scatterplot to help us identify and describe any associations present? First we look to see if there is a *clear pattern* in the scatterplot. In the example on the right, there is *no clear pattern* in the points. The points are *randomly scattered* across the plot, so we conclude that there is *no association*.

For the three examples below, there is a *clear* (but different) *pattern* in each set of points, so we conclude that there is an *association* in each case.



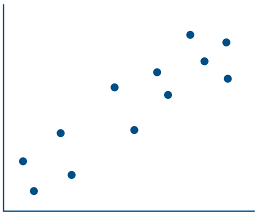
Having found a clear pattern, we need to be able to describe these associations clearly, as they are obviously quite different. There are four things we look for in the pattern of points:

- direction
- strength
- form
- outliers

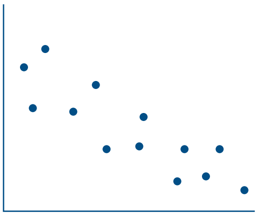
Direction of an association

We begin by looking at the overall pattern in the scatterplot.

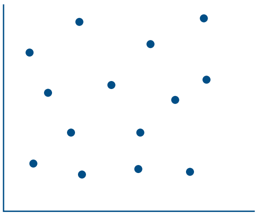
- If the points in the scatterplot tend to move up the plot as we go from left to right, then we say there is a *positive association* between the variables. That is, the values of the explanatory variable and the response variable tend to increase together.



- If the points in the scatterplot tend to move down the plot as we go from left to right, then we say there is a *negative association* between the variables. That is, as the values of the explanatory variable increase, the values of the response variable tend to decrease.



- If there is *no pattern* in the scatterplot; that is, the points just seem to randomly scatter across the plot, then we say there is *no association* between the variables.

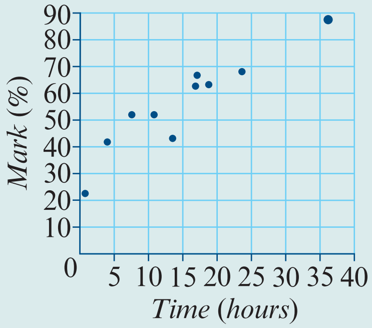


ISBN 978-1-009-25917-0

Example 1 Direction of an association

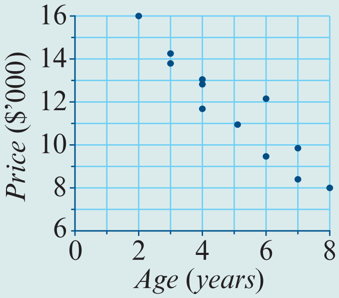
Determine whether each of the following scatterplots exhibits an association. Where there is an association, describe the direction of the association in the scatterplot and what it means in terms of the variables involved.

a



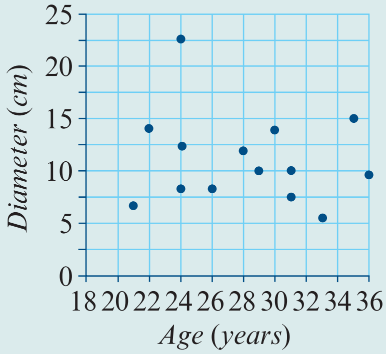
mark: mark on an exam
time: time spent studying

b



price: price of a car
age: age of car

c



diameter: calf diameter
age: age of person

Solution

a There is a *clear pattern* in the scatterplot. The points move *upwards* from left to right.

The direction of the association is positive. Students who spend more time studying for the exam tended to get higher marks.

b There is a *clear pattern* in the scatterplot. The points move *downwards* from left to right.

The direction of the association is negative. The price of a car tends to decrease with age.

c There is no clear pattern in the scatterplot.

There is no association between calf diameter and age.

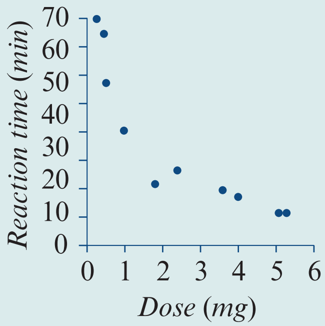
ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

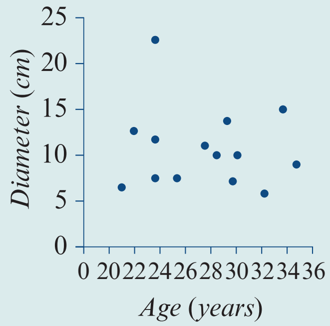
Example 2 Direction of an association

Determine whether each of the following scatterplots exhibits an association. Where there is an association, describe the direction of the association in terms of the variables in the scatterplot and what it means in terms of the variables involved.

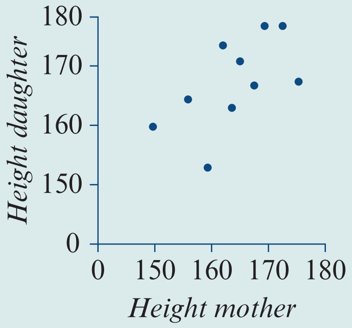
a



b



c



Solution

a There is a *clear pattern* in the scatterplot. The points in the scatterplot drift *downwards* from left to right.

The direction of the association is negative. Reaction times tend to decrease as the drug dose increases.

b There is no pattern in the scatterplot.

There is no association between diameter and age.

c There is a *clear pattern* in the scatterplot. The points in the scatterplot drift *upwards* from left to right.

The direction of the association is positive. Taller mothers have taller daughters.

Direction of an association

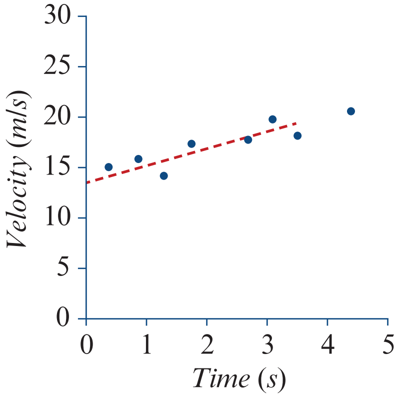
- Two variables have a *positive association* when the value of the response variable tends to increase as the value of the explanatory variable increases.
- Two variables have a *negative association* when the value of response variable tends to decrease as the value of the explanatory variable increases.
- Two variables have *no association* when there is no consistent change in the value of the response variable when the value of the explanatory variable increases.

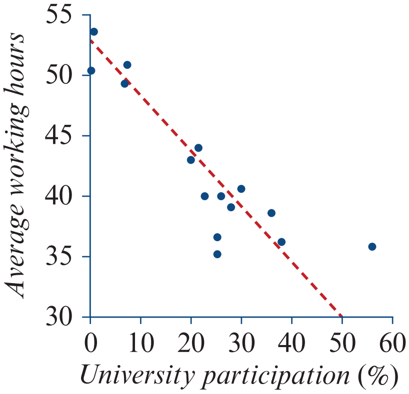
ISBN 978-1-009-25917-0

Form of an association

What we are looking for is whether the pattern in the points has a *linear form*. If the points in a scatterplot appear to be random fluctuations around a *straight line*, then we say that the scatterplot has a linear form. If the scatterplot has a *linear form*, then we say that the variables are *linearly associated*.

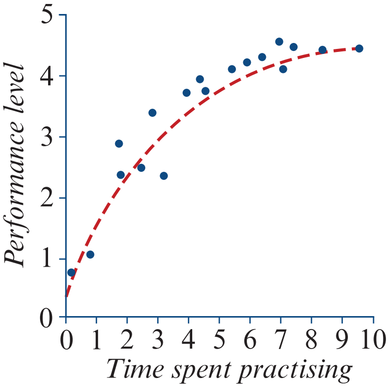
For example, both of the scatterplots below can be described as having a *linear form*; that is, the scatter in the points can be thought of as random fluctuations around a straight line. We can then say that the associations between the variables involved are linear. (The dotted lines have been added to the graphs to make it easier to see the linear form.)





By contrast, consider the scatterplot opposite, plotting performance level against time spent practising a task. There is an association between performance level and time spent practising, but it is clearly non-linear.

This scatterplot shows that while level of performance on a task increases with practice, there comes a time when the performance level will no longer improve substantially with extra practice.



While non-linear relationships exist (and we must always check for their presence by examining the scatterplot), many of the relationships we meet in practice are linear. We will restrict ourselves to the analysis of scatterplots with linear forms for now.

Form of an association

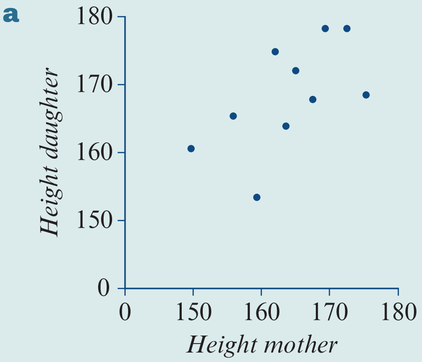
A scatterplot is said to have a *linear form* when the points tend to follow a straight line. A scatterplot is said to have a *non-linear form* when the points tend to follow a curved line.

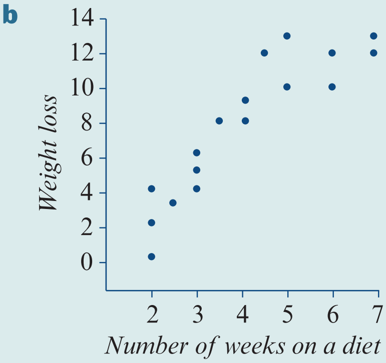
ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Example 3 Form of an association

Classify the *form* of the association in each scatterplot as linear or non-linear.





Solution

a There is a *clear pattern*.
The points in the scatterplot can be imagined to be scattered around a *straight line*.

The association is linear.

b There is a *clear pattern*.
The points in the scatterplot can be imagined to be scattered around a *curved line* rather than a straight line.

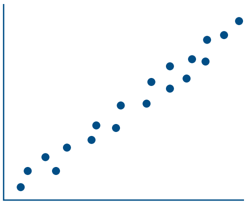
The association is non-linear.

Strength of an association

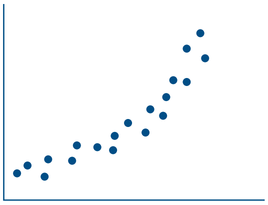
The **strength of an association** is a measure of how much scatter there is in the scatterplot.

Strong association

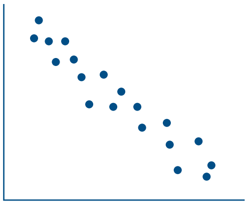
When there is a *strong association* between the variables, the points will tend to follow a single stream. A pattern is clearly seen. There is only a small amount of scatter in the plot.



Strong positive association



Strong positive association

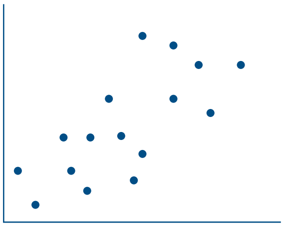


Strong negative association

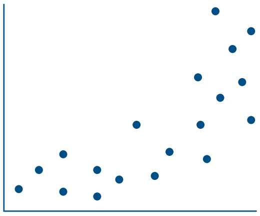
ISBN 978-1-009-25917-0

Moderate association

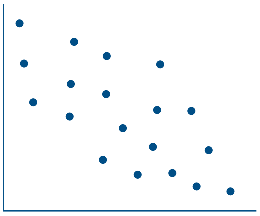
As the amount of scatter in the plot increases, the pattern becomes less clear. This indicates that the association is less strong. In the examples below, we might say that there is a *moderate association* between the variables.



Moderate positive association



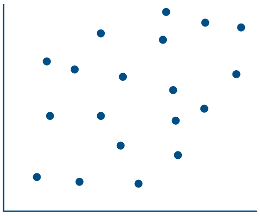
Moderate positive association



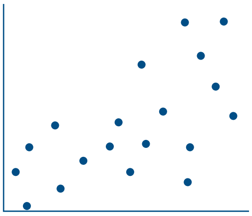
Moderate negative association

Weak association

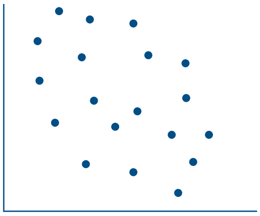
As the amount of scatter increases further, the pattern becomes even less clear. This indicates that any association between the variables is weak. The scatterplots below are examples of *weak association* between the variables.



Weak positive association



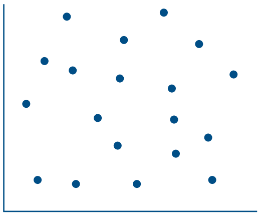
Weak positive association



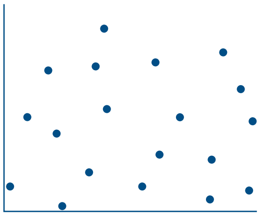
Weak negative association

No association

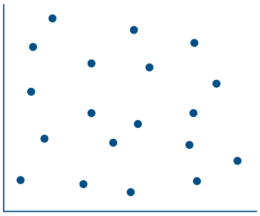
Finally, when all we have is scatter, as seen in the scatterplots below, no pattern can be seen. In this situation we say that there is *no association* between the variables.



No association



No association



No association

The scatterplots on the previous page and above should help you to get a feel for the strength of an association from the amount of scatter in a scatterplot.

At the moment, you only need to be able to estimate the strength of an association as strong, moderate, weak or none, by comparing it with the standard scatterplots shown on the previous pages. Later in this chapter, you will learn about a statistic, the **correlation coefficient**, which can be used to give a value to the strength of linear association.

Strength of an association

The strength of an association is a measure of how much scatter there is in the scatterplot.

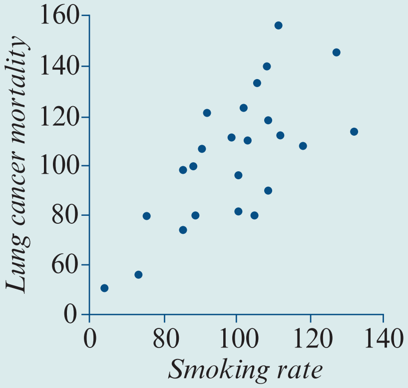
The strength can be described as:

- *strong* if a pattern can clearly be seen
- *moderate* if a pattern is evident, but not clearly seen
- *weak* if some evidence of a pattern is observed, but not obvious
- *no association* if there is no pattern observed at all.

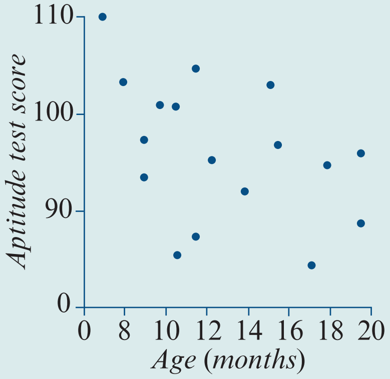
Example 4 Strength of an association

For each of the following scatterplots, describe the strength of any association between the variables.

a

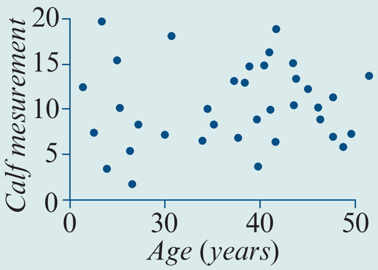


b



c

d



Solution

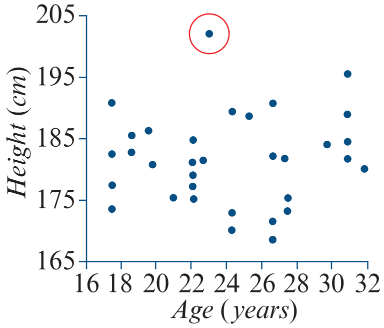
a	This graph has a clear linear form, but some of the points vary quite a bit from it.	This scatterplot shows moderate association between the variables.
b	This graph appears to have a general negative form, but most of the points vary quite a bit from it.	This scatterplot shows weak association between the variables.
c	This graph has a very clear pattern with all points consistent with the linear pattern.	This scatterplot shows strong association between the variables.
d	There does not seem to be any pattern in this graph.	There is no association evident in this scatterplot.

Outliers

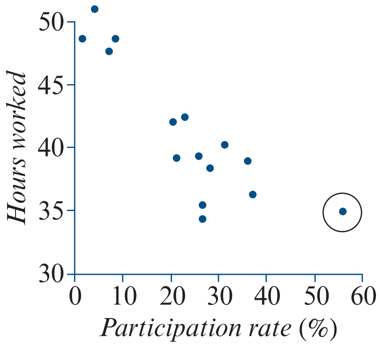
An **outlier** is a data point that seems to be standing out from the rest of the data. Outlier points generally do not fit any pattern that is evident as clearly as the other data points.

The scatterplot of height against age for a group of footballers (shown opposite) has one possible outlier. There is a footballer aged around 23 with height close to 205 cm. The point for this footballer is shown circled.

This point seems to sit apart from the rest of the data and can be considered as an outlier.



By contrast, there seems to be a point that stands out from the rest of the data in the scatterplot of hours worked against participation rate (shown opposite). This point represents hours worked of approximately 35 and a participation rate that is much higher than any other. This point is shown circled.



This point is not considered an outlier. It fits within the pattern that is evident and its location is consistent with the observation of negative association. Higher participation rates would be expected to have lower hours worked.

Outliers

An outlier is a data point that stands in some level of isolation from the rest of the data.

Outliers tend not to follow any pattern that is evident in the rest of the data.

ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

Exercise 2A

Identifying and describing associations

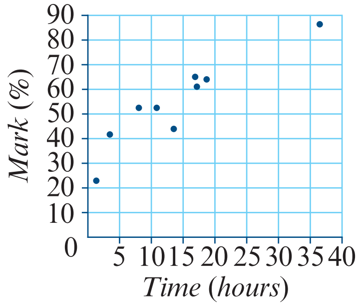
1 For each of the following pairs of variables, indicate whether you expect an association to exist and, if so, whether you would expect the association to be positive or negative.

- **a** *fitness level* and *amount* of daily exercise
- **b** *foot length* and *height*
- **c** *comfort level* and *temperature* above 30°C
- **d** *foot length* and *intelligence*
- **e** *time* taken to get to school and *distance* travelled
- **f** *number of pages* in a book and its *price*

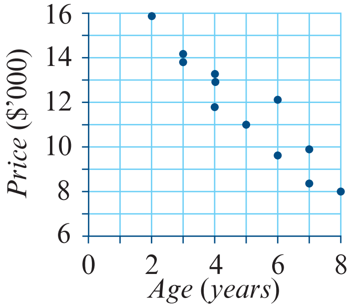
Example 2–4

2 The variables in each of the following scatterplots are associated. In each case classify the association according to its strength (strong/medium/weak), form (linear/non-linear) and direction (positive/negative). Also note the presence of outliers (if any).

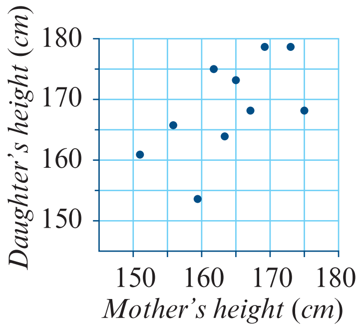
a



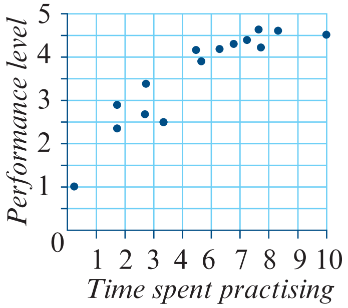
b



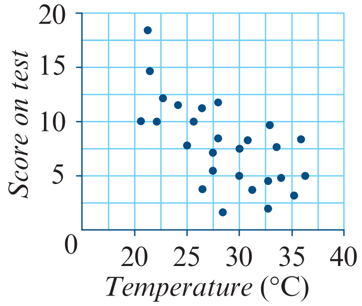
c



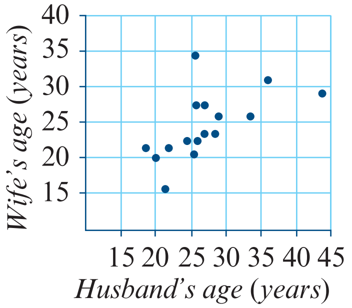
d



e



f



ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

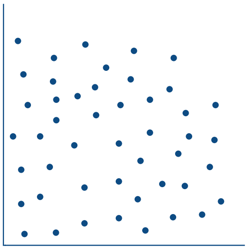
2B Correlation coefficient

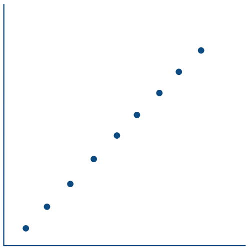
Strength of a linear relationship: the correlation coefficient

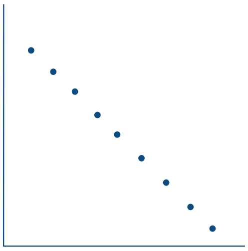
The strength of a linear association is an indication of how closely the points in the scatterplot fit a straight line. If the points in the scatterplot lie exactly on a straight line, we say that there is a perfect linear association. If there is no fit at all we say there is no association. In general, we have an imperfect fit, as seen in all of the scatterplots to date.

To measure the *strength of a linear relationship*, a statistician called Carl Pearson developed a *correlation coefficient*, r, which has the following properties.

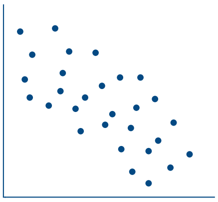
- If there is *no linear* association, $r = 0$.
- If there is a *perfect positive linear* association, $r = +1$.
- If there is a *perfect negative linear* association, $r = -1$.

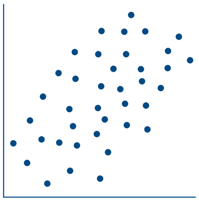

$r = 0$

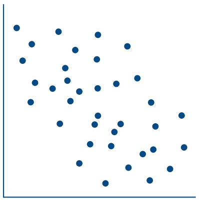

$r = +1$

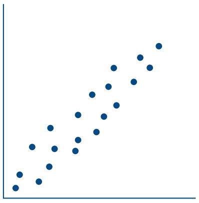

$r = -1$

In practice, the value of r will be somewhere between -1 and $+1$ and rarely exactly zero as shown in the selection of scatterplots below.


$r = -0.7$


$r = +0.5$


$r = -0.3$


$r = +0.9$

These scatterplots illustrate an important point – the stronger the association, the larger the magnitude of the correlation coefficient.

Calculating the value of the correlation coefficient

A CAS calculator can be used to calculate the value of the correlation coefficient for a given set of data.

How to calculate the correlation coefficient using the TI-Nspire CAS

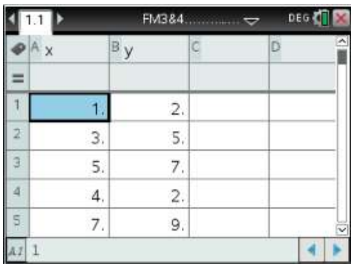
Determine the value of the correlation coefficient, r, for the given data. Give the answer correct to two decimal places.

x	1	3	5	4	7
y	2	5	7	2	9

Steps

1 Start a new document by pressing ctrl + N.

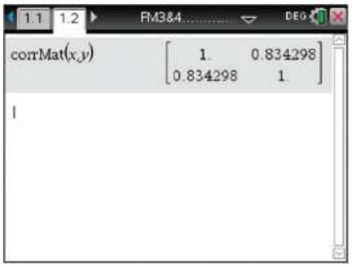
2 Select **Add Lists & Spreadsheet**.
Enter the data into lists named ***x*** and ***y***.



	A x	B y	C	D
=				
1	1.	2.		
2	3.	5.		
3	5.	7.		
4	4.	2.		
5	7.	9.		

A1 1

3 Press ctrl + I and select **Add Calculator**.
Using the **correlation matrix** command: type in **corrmat(*x*, *y*)** and press enter .



corrMat(x,y) $\begin{bmatrix} 1. & 0.834298 \\ 0.834298 & 1. \end{bmatrix}$

Alternatively:

a Press 📖 1 C to access the **Catalog**, scroll down to **corrMat(** and press enter.

b Complete the command by typing in ***x*, *y*** and press enter.
The value of the correlation coefficient is $r = 0.8342\ldots$ or 0.83 (2 d.p.)

ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

How to calculate the correlation coefficient using the ClassPad

The following data show the per capita income (in $'000) and the carbon dioxide emissions (in tonnes) of 11 countries.
Determine the value of the correlation coefficient correct to two decimal places.

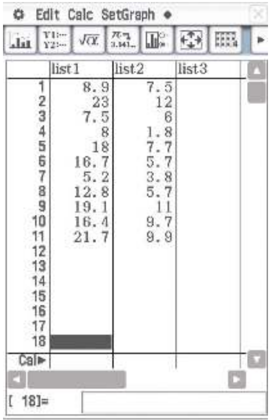
Income ($'000)	8.9	23.0	7.5	8.0	18.0	16.7	5.2	12.8	19.1	16.4	21.7
CO_2 (tonnes)	7.5	12.0	6.0	1.8	7.7	5.7	3.8	5.7	11.0	9.7	9.9

Steps

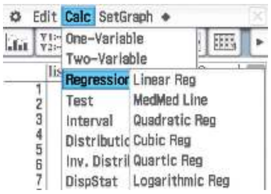
1 Open the **Statistics** application .

2 Enter the data into the columns:
- **Income** in list1
- **CO2** in list2.

3 Select **Calc>Regression>Linear Reg** from the menu bar.



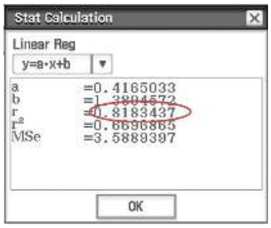
	list1	list2	list3
1	8.9	7.5	
2	23	12	
3	7.5	6	
4	8	1.8	
5	18	7.7	
6	16.7	5.7	
7	5.2	3.8	
8	12.8	5.7	
9	19.1	11	
10	16.4	9.7	
11	21.7	9.9	
12			
13			
14			
15			
16			
17			
18			



4 Press EXE.

5 Tap **OK** to confirm your selections.
The value of the correlation coefficient is
$r = 0.818\ldots$ or 0.82 (to 2 d.p.).

Linear Reg
XList: list1
YList: list2
Freq: 1
Copy Formula: Off
Copy Residual: Off



```
Linear Reg
y=a•x+b
a        =0.4165033
b        =1.3804672
r        =0.8183437
r²       =0.6696865
MSe      =3.5889397
```

Correlation coefficient

The correlation coefficient, r:

- measures the *strength* of *a linear relationship*, with larger values indicating stronger relationships
- has a value between –1 and +1
- is positive if the direction of the linear relationship is positive
- is negative if the direction of the linear relationship is negative.

Interpreting the correlation coefficient

The correlation coefficient, *r*, can be used to classify the strength of a linear association as follows:

strong positive linear association *r* between 0.75 and 0.99
moderate positive linear association *r* between 0.5 and 0.74
weak positive linear association *r* between 0.25 and 0.49
no linear association *r* between −0.24 and 0.24
weak negative linear association *r* between −0.25 and −0.49
moderate negative linear association *r* between −0.5 and −0.74
strong negative linear association *r* between −0.75 and −0.99

The correlation coefficient is not a perfect measure of the strength of an association. For certain sets of data, the correlation coefficient gives misleading indications of strength.

Warning!

If you use the value of the *correlation coefficient* as a measure of the strength of an association, you are implicitly assuming that:

1. the variables are *numerical*
2. the association is *linear*
3. there are *no outliers* in the data.

The *correlation coefficient* can give a *misleading* indication of strength if there are outliers present.

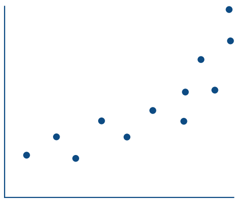
Skillsheet

Exercise 2B

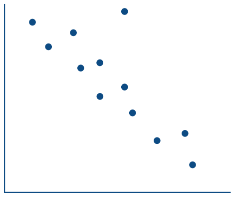
Basic ideas

1 What three assumptions do you make when using the value of the correlation coefficient as a measure of the strength of an association?

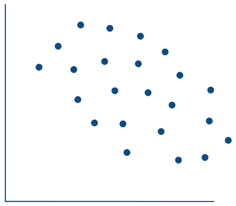
2 The scatterplots of three sets of related variables are shown.



Scatterplot A



Scatterplot B



Scatterplot C

a For each scatterplot, describe the association in terms of strength, direction, form and outliers (if any).

b For which of these scatterplots would it be inappropriate to use the correlation coefficient, *r*, to give a measure of the strength of the association between the variables? Give reasons.

ISBN 978-1-009-25917-0

3 Use the guidelines on page 40 to classify the strength of a linear relationship for which the correlation coefficient is calculated to be:

a $r = 0.20$ **b** $r = -0.30$ **c** $r = -0.85$ **d** $r = 0.33$

e $r = 0.95$ **f** $r = -0.74$ **g** $r = 0.65$ **h** $r = -0.24$

i $r = -0.48$ **j** $r = 0.29$ **k** $r = 1$ **l** $r = -1$

Calculating r using a CAS calculator

CAS

4 **a** The table below shows the maximum and minimum temperatures during a heat-wave. The *maximum* and *minimum* temperature each day are linearly associated. Use your calculator to show that $r = 0.818$, correct to three decimal places.

Day	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday
Maximum (°*C*)	29.4	34.0	34.5	35.0	36.9	36.4
Minimum (°*C*)	17.7	19.8	23.3	22.4	22.0	22.0

b This table shows the number of runs scored and balls faced by batsmen in a cricket match. *Runs scored* and *balls faced* are linearly associated. Use your calculator to show that $r = 0.8782$, correct to four decimal places.

Batsman	1	2	3	4	5	6	7	8	9	10	11
Runs scored	27	8	21	47	3	15	13	2	15	10	2
Balls faced	29	16	19	62	13	40	16	9	28	26	6

c This table shows the hours worked and university participation rate (%) in six countries. *Hours worked* and university *participation rate* are linearly associated. Use your calculator to show that $r = -0.6727$, correct to four decimal places.

Country	Australia	Britain	Canada	France	Sweden	US
Hours worked	35.0	43.0	38.2	39.8	35.6	34.8
Participation rate (%)	26	20	36	25	37	55

2C The coefficient of determination

If two variables are associated, it is possible to estimate the value of one variable from that of the other. For example, people's weights and heights are associated. Thus, given a person's height, we can roughly predict their weight. The degree to which we can make such predictions depends on the value of r. If there is a perfect linear association ($r = 1$) between two variables, we can make an exact prediction.

ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

For example, when you buy cheese by the gram there is an exact association between the weight of the cheese and the amount you pay ($r = 1$). At the other end of the scale, there is no association between an adult's height and their IQ ($r \approx 0$). So knowing an adult's height will not enable you to predict their IQ any better than guessing.

The coefficient of determination

The *degree* to which one variable can be predicted from another linearly related variable is given by a statistic called the **coefficient of determination**.

The coefficient of determination is *calculated* by squaring the correlation coefficient:

$$\text{coefficient of determination} = r^2$$

Calculating the coefficient of determination

Numerically, the coefficient of determination is r^2. Thus, if the correlation between weight and height is $r = 0.8$, then the

$$\text{coefficient of determination} = r^2 = 0.8^2 = 0.64 \quad \text{or} \quad 0.64 \times 100 = 64\%$$

Note: We have converted the coefficient of determination into a percentage (64%) as this is the most useful form when we come to interpreting the coefficient of determination.

Interpreting the coefficient of determination

We now know how to calculate the coefficient of determination, but what does it tell us?

Interpreting the coefficient of determination

The coefficient of determination (as a percentage) tells us the *variation in the response variable* that is *explained* by the *variation in the explanatory variable*.

When interpreting the coefficient of determination, a general model can be used:

- $r^2 \times 100\%$ of the variation in the *response variable* can be explained by the *explanatory variable*.

But what does this mean in practical terms?

Let's take the relationship between weight and height that we just considered. Here the coefficient of determination is 0.64 (or 64%).

The coefficient of determination tells us that 64% of the variation in people's weights is explained by the variation in their heights.

Note: The r^2 value does not indicate the direction (positive or negative) of a relationship, as it is always between 0 and 1. It only informs us about how the variation in the two variables can be explained.

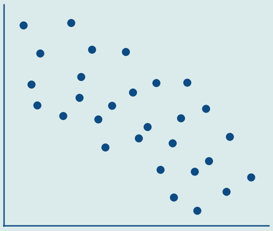
What do we mean by 'explained'?

If we take a group of people, their weights and heights will vary. One explanation is that taller people tend to be heavier and shorter people tend to be lighter. The coefficient of determination tells us that 64% of the variation in people's weights can be explained by the variation in their heights. The rest of the variation (36%) in their weights will be explained by other factors, such as sex, lifestyle, build.

Example 5 Calculating the correlation coefficient from the coefficient of determination

For the relationship described by this scatterplot, the coefficient of determination = 0.5210.

Determine the value of the correlation coefficient, r.



Solution

1. The coefficient of determination = r^2. Use the value of the coefficient of determination to set up an equation for r^2. Solve.

$r^2 = 0.5210$

$\therefore r = \pm\sqrt{0.5210} = \pm 0.7218$

2. There are two solutions, one positive and the other negative. Use the scatterplot to decide which applies.

Scatterplot indicates a negative association.

3. Write your answer.

$\therefore r = -0.7218$

Example 6 Calculating and interpreting the coefficient of determination

Carbon monoxide (CO) levels in the air and traffic volume are linearly related, with:

$$r = +0.985$$

Determine the value of the coefficient of determination, write it as a percentage and then interpret this value. In this relationship, *traffic volume* is the explanatory variable.

Solution

The coefficient of determination is:

$$r^2 = (0.985)^2 = 0.970\ldots \text{ hence } 0.970 \times 100 = 97.0\%$$

Therefore, 97% of the variation in carbon monoxide levels in the air can be explained by the variation in traffic volume.

ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

Clearly, traffic volume is a very good predictor of carbon monoxide levels in the air. Thus, knowing the traffic volume enables us to predict carbon monoxide levels with a high degree of accuracy. This contrasts with the next example, which concerns predicting mathematical ability from verbal ability.

Example 7 Calculating and interpreting the coefficient of determination

Scores on tests of verbal ability and mathematical ability are linearly related with:

$r = +0.275$

Determine the value of the coefficient of determination, write it as a percentage and then interpret this value. In this relationship, *verbal ability* is the explanatory variable.

Solution

The coefficient of determination is:

$r^2 = (0.275)^2 = 0.0756\ldots$ hence $0.076 \times 100 = 7.6\%$

Therefore, only 7.6% of the variation observed in scores on the mathematical ability test can be explained by the variation in scores obtained on the verbal ability test.
Clearly, scores on the verbal ability test are not good predictors of the scores on the mathematical ability test; 92.4% of the variation in mathematical ability is explained by other factors.

Exercise 2C

Calculating the coefficient of determination from r

1 For each of the following values of r, calculate the value of the coefficient of determination (correct to three decimal places) and convert to a percentage (correct to one decimal place).

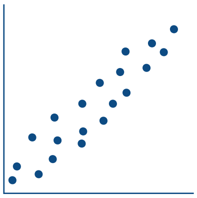
a $r = 0.675$ **b** $r = 0.345$ **c** $r = -0.567$ **d** $r = -0.673$ **e** $r = 0.124$

Calculating r from the coefficient of determination given a scatterplot

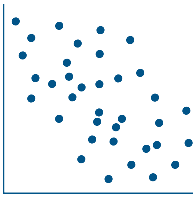
Note: The scatterplots have been included in Questions 2 to help you decide the sign of r.

Example 5

2 **a** For the relationship described by the scatterplot shown, the coefficient of determination, $r^2 = 0.8215$.
Determine the value of the correlation coefficient, r (correct to three decimal places).



b For the relationship described by the scatterplot shown, the coefficient of determination $r^2 = 0.1243$.
Determine the value of the correlation coefficient, r (correct to three decimal places).



ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

Calculating and interpreting the coefficient of determination

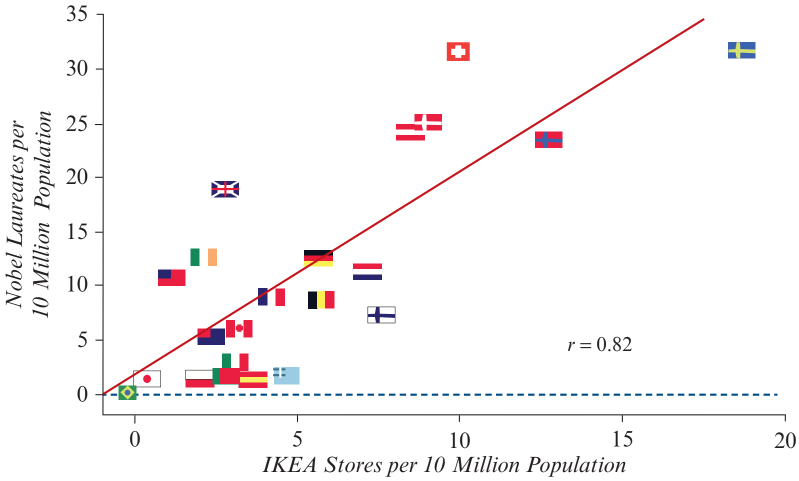
Example 6

3 For each of the following, determine the value of the coefficient of determination (correct to three decimal places), write it as a percentage (correct to one decimal place), and interpret.

- **a** Scores on hearing tests and age (EV) are linearly related, with $r_{\text{hearing, age}} = -0.611$.
- **b** Mortality rates and smoking rates (EV) are linearly related, with $r_{\text{mortality, smoking}} = 0.716$.
- **c** Life expectancy and birth rates (EV) are linearly related, with $r_{\text{life expectancy, birth rate}} = -0.807$.
- **d** Daily maximum (RV) and minimum temperatures are linearly related, with $r_{\text{max, min}} = 0.818$
- **e** Runs scored (EV) and balls faced by a batsman are linearly related, with $r_{\text{runs, balls}} = 0.8782$

2D Correlation and causality

Recently there has been interest in the strong association between the number of Nobel prizes a country has won and the number of IKEA stores in that country ($r = 0.82$). This strong association is evident in the scatterplot below. Here country flags are used to represent the data points.



Does this mean that one way to increase the number of Australian Nobel prize winners is to build more IKEA stores?

Almost certainly not, but this association highlights the problem of assuming that a strong *correlation* between two variables indicates the association between them is *causal*.

Correlation does not imply causality

A correlation tells you about the strength of the association between the variables, but no more. It tells you nothing about the source or cause of the association.

ISBN 978-1-009-25917-0

Cambridge University Press

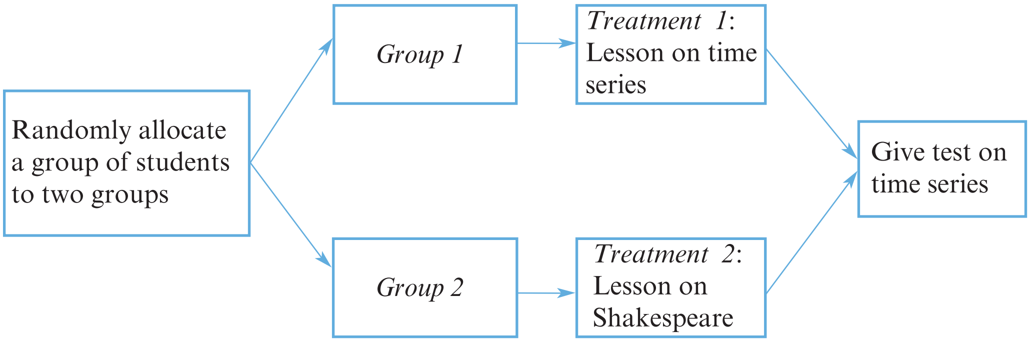
Updated January 2024

Video

To help you with this concept, you should watch the video ‘The Question of Causation’, which can be accessed through the link below. It is well worth 15 minutes of your time. http://cambridge.edu.au/redirect/?id=6103

Establishing causality

To establish **causality**, you need to conduct an *experiment*. In an experiment, the value of the *explanatory variable* is *deliberately manipulated*, while all other possible explanatory variables are kept constant or controlled. A simplified version of an experiment is displayed below.



In this experiment, a class of students is randomly allocated into two groups. Random allocation ensures that both groups are as similar as possible.

Next, group 1 is given a lesson on time series (treatment 1), while group 2 is given a lesson on Shakespeare (treatment 2). Both lessons are given under the same classroom conditions. When both groups are given a test on time series the next day, group 1 does better than group 2.

We then conclude that this was because the students in group 1 were given a lesson on time series.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Is this conclusion justified?

In this experiment, the students' test score is the response variable and the type of lesson they were given is the explanatory variable. We randomly allocated the students to each group while ensuring that all other possible explanatory variables were controlled by giving the lessons under the same classroom conditions. In these circumstances, the observed difference in the response variable (*test score*) can reasonably be attributed to the explanatory variable (*lesson type*).

Unfortunately, it is extremely difficult to conduct properly controlled experiments, particularly when the people involved are going about their everyday lives.

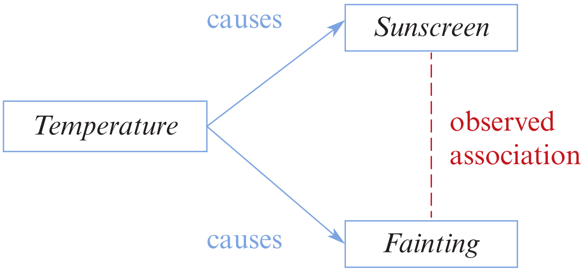
When data are collected through observation rather than experimentation, we must accept that strong association between two variables is insufficient evidence by itself to conclude that an observed change in the response variable has been caused by an observed change in the explanatory variable. It may be, but unless all of the relevant variables are under our control there will always be alternative non-causal explanations to which we can appeal. We will now consider the various ways this might occur.

Possible non-causal explanations for an association

Common response

Consider the following. There is a strong positive association between the number of people using sunscreen and the number of people fainting. Does this mean that applying sunscreen causes people to faint?

Almost certainly not. On hot and sunny days, more people *apply sunscreen* and more people *faint* due to heatexhaustion. The two variables are associated because they are both strongly associated with a common third variable, *temperature*. This phenomenon is called a *common response*. See the diagram below.

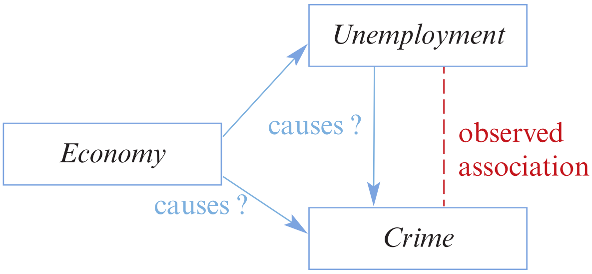


Unfortunately, being able to attribute an association to a single third variable is the exception rather than the rule. More often than not, the situation is more complex.

Confounding variables

Statistics show that *crime* rates and *unemployment* rates in a city are strongly correlated. Can you then conclude that a decrease in unemployment will lead to a decrease in crime rates?

It might, but other possible causal explanations could be found. For example, these data were collected during an economic downturn. Perhaps the state of the economy caused the problem. See the diagram below.



In this situation, we have at least two possible causal explanations for the observed association, but we have no way of disentangling their separate effects. When this happens, the effects of the two possible explanatory variables are said to be *confounded*, because we have no way of knowing which is the actual cause of the association.

Coincidence

It turns out that there is a strong correlation ($r = 0.99$) between the consumption of margarine and the divorce rate in the US state of Maine. Can we conclude that eating margarine causes people in Maine to divorce?

A better explanation is that this association is purely coincidental.

Occasionally, it is almost impossible to identify any feasible confounding variables to explain a particular association. In these cases we often conclude that the association is 'spurious' and it has just happened by chance. We call this *coincidence*.

Conclusion

However suggestive a strong association may be, this alone does not provide sufficient evidence for you to conclude that two variables are causally related. Unless the association is totally spurious and devoid of meaning, it will always be possible to find at least one variable 'lurking' in the background that could explain the association.

Association (correlation) and causation

By itself, an observed association between two variables is *never enough* to justify the conclusion that two variables are causally related, no matter how obvious the causal explanation may appear to be.

Exercise 2D

1 A study of primary school children aged 5 to 11 years finds a strong positive correlation between height and score on a test of mathematics ability. Does this mean that taller people are better at mathematics? What common cause might counter this conclusion?

2 There is a clear positive correlation between the number of churches in a town and the amount of alcohol consumed by its inhabitants. Does this mean that religion is encouraging people to drink? What common cause might counter this conclusion?

3 There is a strong positive correlation between the amount of ice-cream consumed and the number of drownings each day. Does this mean that eating ice-cream at the beach is dangerous? What common cause might explain this association?

4 The number of days a patient stays in hospital is positively correlated with the number of beds in the hospital. Can it be said that bigger hospitals encourage patients to stay longer than necessary just to keep their beds occupied? What common cause might counter this conclusion?

5 Suppose we found a high correlation between smoking rates and heart disease across a group of countries. Can we conclude that smoking causes heart disease? What confounding variable(s) could equally explain this correlation?

6 There is a strong correlation between cheese consumption and the number of people who died after becoming tangled in their bed sheets. What do you think is the most likely explanation for this correlation?

7 There is a strong positive correlation between the number of fire trucks attending a house fire and the amount of damage caused by the fire. Is the amount of damage in a house fire caused by the fire trucks? What common cause might explain this association?

Key ideas and chapter summary

Assignment

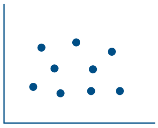
Describing associations in scatterplots

Associations are described in terms of:

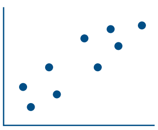
- **direction** (positive or negative)
- **form** (linear or non-linear)
- **strength** (strong, moderate, weak or none)
- existence of **outliers**.

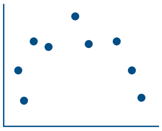
Identifying associations (relationships) between two numerical variables

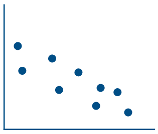
A random cluster of points (no clear pattern) indicates that there is *no association* between the variables.



A *clear pattern* in the scatterplot indicates that there is an *association between the variables*.







Outlier

An **outlier** is a data point that stands in some level of isolation from the rest of the data.

Correlation coefficient (*r*)

Correlation coefficient (r) is a statistic that measures the direction and strength of a linear relationship between a pair of variables.

Assumptions for the use of the correlation coefficient

If the value of the *correlation coefficient* is used as a measure of strength of an association, it is assumed that:

- the variables are *numerical*
- the association is *linear*
- there are *no outliers* in the data.

The coefficient of determination

Coefficient of determination $= r^2$

The coefficient of determination gives the percentage of variation in the response variable that can be explained by the variation in the explanatory variable.

Correlation and causation

A *correlation* between two variables does not automatically imply that the association is *causal*. Alternative *non-causal explanations* for the association include a *common response* to a common third variable, a *confounded* variable or simply *coincidence*.

ISBN 978-1-009-25917-0

Skills check

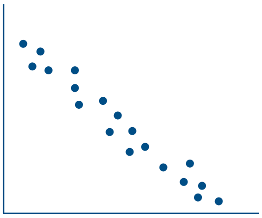
Having completed this chapter you should be able to:

- use a scatterplot to describe an association between two numerical variables in terms of:
  - direction (positive or negative association)
  - form (linear or non-linear)
  - strength (weak, moderate, strong)
  - possible outliers
- calculate and interpret the correlation coefficient, r
- know the three key assumptions made when using the correlation coefficient r as a measure of the strength of the linear association between two variables; that is:
  - the variables are numerical
  - the association is linear
  - there are no clear outliers
- calculate and interpret the coefficient of determination
- understand that finding an association between two variables does not automatically indicate a causal association
- identify situations where unjustified statements about causality could be (or have been) made and recognise possible non-causal explanations as examples of a *common response, confounding* or *coincidence*.

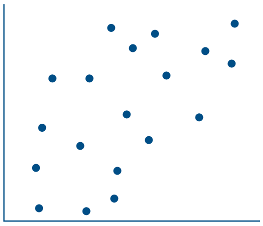
Short-answer questions

1 For each of the scatterplots below, describe the association between the variables in terms of strength, direction and form.

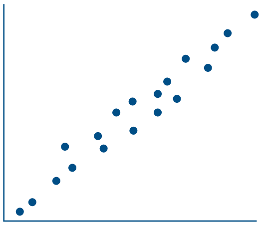
a



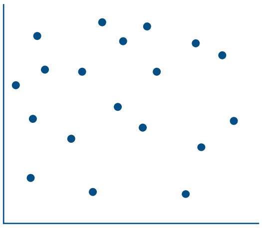
b



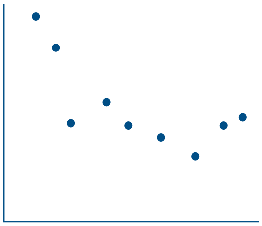
c



d



e

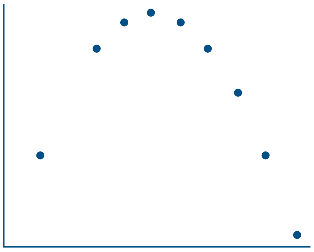


ISBN 978-1-009-25917-0

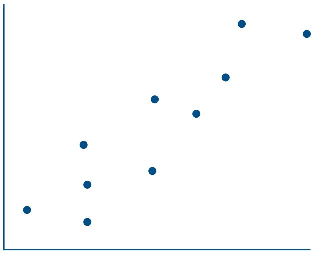
2 The association between birth weight and infant mortality rate is negative. Interpret the direction of the association in the context of the variables.

3 For each of the following scatterplots plots state whether it would make sense to calculate the correlation coefficient (r) to indicate the strength of the association between the variables?

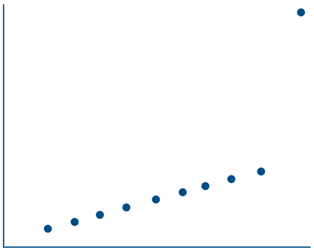
a



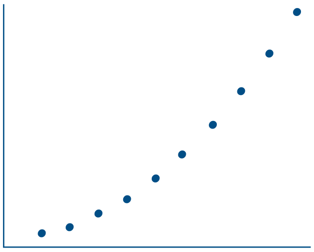
b



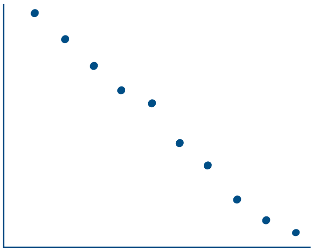
c



d



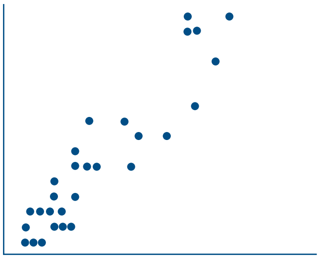
e



4 For the scatterplot shown, the value of the correlation coefficient, r, is closest to:

A 0.28 **B** 0.41 **C** 0.63

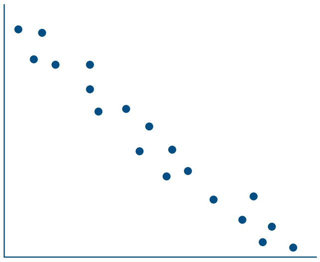
D 0.86 **E** 0.99



5 For the scatterplot shown, the value of the correlation coefficient, r, is closest to:

A −0.90 **B** −0.64 **C** −0.23

D 0.64 **E** 0.90



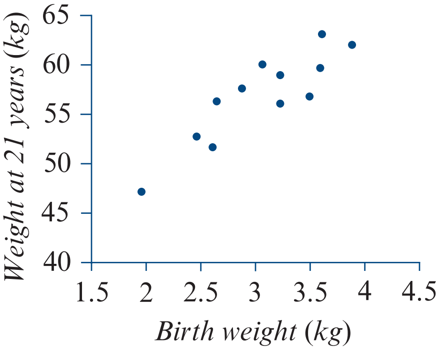
6 Describe the strength and direction of a linear relationship with a correlation coefficient of $r = -0.32$.

7 The weekly *income* and weekly *expenditure* on food for a group of 10 university students is given in the following table.

Income ($/week)	150	250	300	600	300	380	950	450	850	1000
Expenditure ($/week)	40	60	70	120	130	150	200	260	460	600

Determine the value of the correlation coefficient, r, for these data to one decimal place.

8 The scatterplot shows the weights at age 21 and at birth of 12 women. Describe the association between the variables in terms of strength, direction and form.



9 The variables *response time* to a drug and *drug dosage* are linearly associated, with $r = -0.9$.

a Describe the association in the context of the variables.

b Determine the coefficient of determination.

c Interpret the value of the coefficient of determination as a percentage, given that drug dosage is the explanatory variable.

10 The correlation coefficient between heart weight and body weight in a group of mice is $r = 0.765$.

a Using body weight as the explanatory variable, determine the coefficient of determination and interpret this value.

b Given that heart weight and body weight of mice are strongly correlated, interpret this association in the context of the variables.

11 A local council survey finds that there is a strong linear correlation ($r = 0.85$) between the amount of rubbish that is recycled per household and a person's annual salary.

a Describe the direction of the association and what is means in terms of the variables involved.

b The mayor of the council concludes that "to improve recycling in our local area, peoples salaries should be increased since higher salaries cause more household recycling." Comment on the appropriateness of this conclusion.

ISBN 978-1-009-25917-0

Cambridge University Press

Updated January 2024

12 There is a strong linear positive correlation ($r = 0.952$) between the marriage rate in Kentucky and the number of people who drown falling out of a fishing boat. From this information, a person might conclude that reducing the number of marriages in Kentucky will decrease the number of people who drown falling out of a fishing boat. Comment on the validity of this conclusion and give a possible non-causal explanation for this association.

Extended-response questions

CAS

1 The following table gives the *number* of times the ball was inside the team's 50 metre line in an AFL football game and the team's final *score* (in points) in that game.

Number	64	57	34	61	51	52	53	51	64	55	58	71
Score (points)	90	134	76	92	93	45	120	66	105	108	88	133

a Which variable is the response variable?

b Construct a scatterplot of *score* against *number*.

c Use the scatterplot to describe the association in terms of strength and direction.

2 The *distance* travelled to work and the *time* taken for ten company employees are given in the following table. *Distance* is the response variable. Determine the value of the correlation coefficient, r, for this set of data. Write your answer rounded to four decimal places.

Distance (km)	*Time (min)*
12	15
50	75
40	50
25	50
45	80
20	50
10	10
3	5
10	10
30	35

3 In a survey of nine problem gamblers, the respondents were asked the *amount* (in dollars) they had spent on gambling and the number of *hours* that they had spent gambling in the past week. This data collected is recorded in the table below.

Hours	10	11	12	15	20	21	25	35	40
Amount	500	530	300	750	1000	1200	2000	2300	5000

a The aim is to predict the amount of money spent on gambling from the time spent gambling. Which is the explanatory variable and which is the response variable?

b Construct a scatterplot of these data.

c Determine the value of the correlation coefficient, r, to three decimal places.

d Describe the association between the variables *amount* and *hours* in terms of strength, direction and form.

4 The following data was recorded through the National Health Survey:

Region	*Percentage with eye disease*	
	Male (%)	*Female* (%)
Australia	40.7	49.1
Other Oceania countries	46.1	66.2
United Kingdom	74.5	75.0
Other North-West Europe	71.2	71.5
Southern & Eastern Europe	71.6	74.6
North Africa & the Middle East	52.2	57.5
South-East Asia	47.7	54.8
All other countries	56.0	62.0

a Construct a scatterplot of these data, with percentage of males on the horizontal axis and percentage of females on the vertical axis.

b Determine the value of the correlation coefficient, r, to three decimal places.

c Describe the association between the male and female eye disease percentages for these countries in terms of strength, direction and form and outliers (if any).

d Determine the value of the coefficient of determination as a percentage and interpret.

5 The data below give the hourly pay *rates* (in dollars per hour) of 10 production-line workers along with their years of *experience* on initial appointment.

Rate ($/h)	15.90	15.70	16.10	16.00	16.79	16.45	17.00	17.65	18.10	18.75
Experience (years)	1.25	1.50	2.00	2.00	2.75	4.00	5.00	6.00	8.00	12.00

a Use a CAS calculator to construct a scatterplot of the data, with the variable *rate* plotted on the vertical axis and the variable *experience* on the horizontal axis. Why has the vertical axis been used for the variable *rate*?

b Comment on direction, outliers, form and strength of any association revealed.

c Determine the value of the coefficient of determination (r^2) and interpret.

d Determine the value of the correlation coefficient (r) correct to three decimal places.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

3

Regression: fitting lines to data

In this chapter

Syllabus references

Topic: Fitting a linear model to numerical data

Subtopics: 3.1.10 – 3.1.12, 3.1.14 – 3.1.16

3A Determining the equation of the least squares line

The process of fitting a straight line to bivariate data is known as **linear regression**.

The aim of linear regression is to model the association between two numerical variables by using a simple mathematical relation, the straight line. Knowing the equation of this line gives us a better understanding of the nature of the association. It also enables us to make predictions from one variable to another – for example, a young boy's adult height from his father's height.

The easiest way to fit a line to bivariate data is to construct a scatterplot and draw the line 'by eye'. We do this by placing a ruler on the scatterplot so that it seems to follow the general trend of the data. You can then use the ruler to draw a straight line. Unfortunately, unless the points are very tightly clustered around a straight line, the results you get by using this method will differ a lot from person to person.

The most common approach to fitting a straight line to data is to use the **least squares method**. This method assumes that the variables are linearly related, and works best when there are no clear outliers in the data.

Some terminology

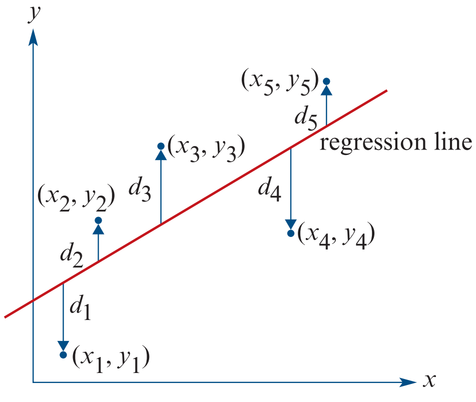
To explain the least squares method, we need to define several terms.

The scatterplot shows five data points, (x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4) and (x_5, y_5).

A regression line (not necessarily the least squares line) has also been drawn on the scatterplot.

The vertical distances d_1, d_2, d_3, d_4 and d_5 of each of the data points from the least squares line are also shown.

These vertical distances, d, are known as **residuals**.



The least squares line

The least squares line is the line where the sum of the squares of the residuals is as small as possible; that is, it minimises:

$$\text{the sum of the squares of the residuals} = (d_1)^2 + (d_2)^2 + (d_3)^2 + (d_4)^2 + (d_5)^2$$

Why do we minimise the sum of the *squares* of the residuals and not the sum of the residuals? This is because the sum of the residuals for the least squares line is always zero. The least squares line is like the mean. It balances out the data values on either side of itself. Some residuals are positive and some negative, and in the end they add to zero. Squaring the residuals solves this problem.

Residuals and the least squares line

The *residual* of a data point is its vertical distance from a least squares line.

The *least squares line* is the line that *minimises* the *sum of the squares* of the *residuals*.

The sum of the residuals for a least squares line is zero.

The *assumptions* for fitting a least squares line to data are the same as for using the correlation coefficient, r. These are that:

- the data is numerical
- the association is linear
- there are no clear outliers.

How do we determine the least squares line?

One method is 'trial-and-error'. We could draw a series of lines, each with a different **slope** and intercept. For each line, we could then work out the value of each of the residuals, square them, and calculate their sum. The least squares line would be the one that minimises these sums.

To see how this might work, you can simulate the process of fitting a least squares line using the interactive 'Regression line'.

The trial-and-error method does not guarantee that we get the exact solution. The exact solution can be found mathematically, using the techniques of calculus. However, the mathematics is beyond this course.

You will use your CAS calculator to do the computations.

From our study of linear equations in previous years we know that the equation of a straight line always takes the form of $y = mx + c$ where m represents the gradient and c is the **y-intercept**.

Equation of the least squares line

The equation of the least squares line has the form:

$$y = ax + b$$

where y is the response variable and x is the explanatory variable, and

- a is the value of slope (or gradient)
- b is the value of the y-intercept (or vertical intercept).

Typically, we write the values of a and b to three significant figures.

Note: If the explanatory and response variables are given specific algebraic variables within the question, then these must be used when writing the equation of the least squares line, instead of x and y.

Warning!

If you do not correctly decide which is the explanatory variable (the x-variable) and which is the response variable (the y-variable) before you start calculating the equation of the least squares line, you may get the wrong answer.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

How to determine and graph the equation of a least squares line using the TI-Nspire CAS

The following data give the height (in cm) and weight (in kg) of 11 people.

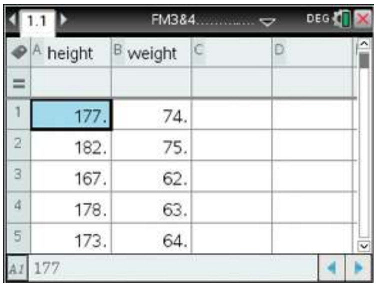
Height (x)	177	182	167	178	173	184	162	169	164	170	180
Weight (y)	74	75	62	63	64	74	57	55	56	68	72

Determine and graph the equation of the least squares line that will enable weight to be predicted from height. Write the intercept and slope correct to three significant figures.

Steps

1 Start a new document by pressing ctrl + N.

2 Select **Add Lists & Spreadsheet**. Enter the data into lists named *height* and *weight*, as shown.



	A height	B weight	C	D
=				
1	177.	74.		
2	182.	75.		
3	167.	62.		
4	178.	63.		
5	173.	64.		

A1 177

3 Identify the explanatory variable (EV) and the response variable (RV).

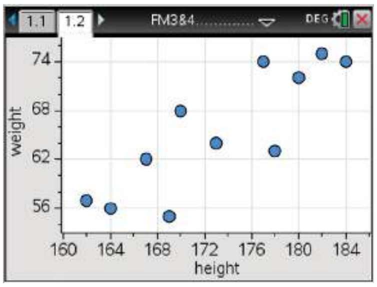
EV: *height*

RV: *weight*

Note: In saying that we want to predict *weight* from *height*, we are implying that *height* is the EV.

4 Press ctrl + I and select **Add Data & Statistics** and construct a scatterplot with *height* (EV) on the horizontal (or x-) axis and *weight* (RV) on the vertical (or y-) axis.

Press menu**>Settings** and click the **Diagnostics** box. Select **Default** to activate this feature for *all* future documents. This will show the coefficient of determination (r^2) whenever a regression is performed.



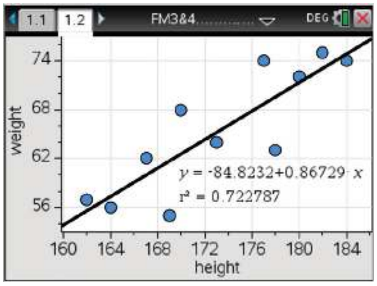
5 Press menu>**Analyze**>**Regression**>**Show Linear (a + bx)** to plot the least squares line on the scatterplot.

Note that, simultaneously, the equation of the least squares line is shown on the screen.

The equation of the least squares line is:

$$y = -84.8 + 0.867x$$

$$\text{or } weight = -84.8 + 0.867 \times height$$



$y = -84.8232+0.86729 \cdot x$

$r^2 = 0.722787$

The coefficient of determination is $r^2 = 0.723$, correct to three significant figures.

ISBN 978-1-009-25917-0

Cambridge University Press

Updated January 2024

How to determine and graph the equation of a least squares line using the ClassPad

The following data give the height (in cm) and weight (in kg) of 11 people.

Height (x)	177	182	167	178	173	184	162	169	164	170	180
Weight (y)	74	75	62	63	64	74	57	55	56	68	72

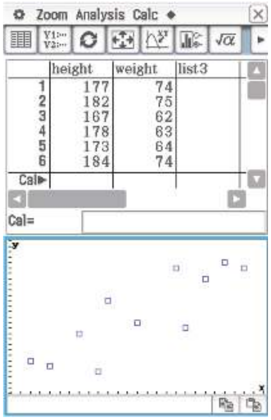
Determine and graph the equation of the least squares line that will enable weight to be predicted from height. Write the intercept and slope correct to three significant figures.

Steps

1 Open the **Statistics** application Statistics and enter the data into columns labelled **height** and **weight**.

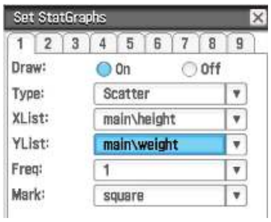
2 Tap to open the **Set StatGraphs** dialog box and complete as shown. Tap **Set** to confirm your selections.

3 Tap in the toolbar at the top of the screen to plot the scatterplot in the bottom half of the screen.



	height	weight	list3
1	177	74	
2	182	75	
3	167	62	
4	178	63	
5	173	64	
6	184	74	
Cal▸			

Cal=

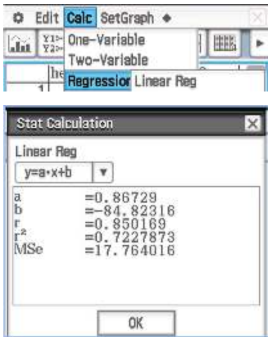

Set StatGraphs

Draw: ● On ○ Off
Type: Scatter
XList: main\height
YList: main\weight
Freq: 1
Mark: square

4 To calculate the equation of the least squares line:

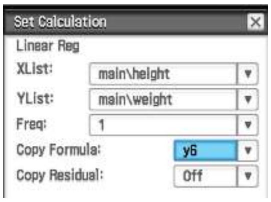
- Tap **Calc** from the menu bar.
- Tap **Regression** and select **Linear Reg**.
- Complete the **Set Calculations** dialog box as shown.
- Tap **OK** to confirm your selections in the **Set Calculations** dialog box. This also generates the results shown opposite.
- Tapping **OK** a second time automatically plots and displays the least squares line.

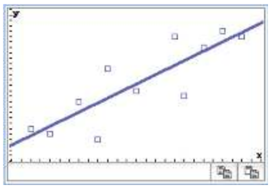
Note: **y6** as the formula destination is an arbitrary choice.



Stat Calculation
Linear Reg
y=a·x+b

```
a    =0.86729
b    =-84.82316
r    =0.850169
r²   =0.7227873
MSe  =17.764016
```


Set Calculation
Linear Reg
XList: main\height
YList: main\weight
Freq: 1
Copy Formula: y6
Copy Residual: Off



5 Use the values of the slope *a* and intercept *b* to write the equation of the least squares line in terms of the variables *weight* and *height*.

$y = 0.867x - 84.8$ (to three significant figures)

The coefficient of determination is $r^2 = 0.723$, correct to three significant places.

ISBN 978-1-009-25917-0

Cambridge University Press

Updated January 2024

Exercise 3A

If you feel that you need some help with significant figures see the video on the topic, accessed through the Interactive Textbook.

Some big ideas

1 The following questions relate to the formulas used to calculate the slope and intercept of the least squares line.

a A least squares line is calculated and the slope is found to be negative. What does this tell us about the sign of the correlation coefficient?

b The correlation coefficient is zero. What does this tell us about the slope of the least squares line?

Using a CAS calculator to determine the equation of a least squares line from raw data

CAS

2 The table shows the number of sit-ups and push-ups performed by six students.

Sit-ups (x)	52	15	22	42	34	37
Push-ups (y)	37	26	23	51	31	45

Let the number of *sit-ups* be the explanatory (x) variable. Use your calculator to show that the equation of the least squares line is:
$y = 0.566x + 16.5$ (correct to three significant figures)

3 The table shows average hours worked and university participation rates (%) in six countries.

Hours (h)	35.0	43.0	38.2	39.8	35.6	34.8
Rate (r)	26	20	36	25	37	55

Use your calculator to show that the equation of the least squares line that enables participation *rates* to be predicted from *hours* worked is:

$$r = -2.6\,h + 130$$ (correct to two significant figures)

4 The table shows the number of runs scored and balls faced by batsmen in a cricket match.

Balls faced (b)	27	8	21	47	3	15	13	2	15	10	2
Runs (r)	29	16	19	62	13	40	16	9	28	26	6

a Use your calculator to show that the equation of the least squares line enabling *runs* scored to be predicted from *balls faced* is:

$$y = 1.06x + 8.30$$

b Rewrite the equation in terms of the variables involved.

ISBN 978-1-009-25917-0

5 The table below shows the number of TVs and cars owned (per 1000 people) in six countries.

Number of TVs (t)	378	404	471	354	381	624
Number of cars (c)	417	286	435	370	357	550

We wish to predict the *number of TVs* from the *number of cars*.

a Which is the response variable?

b Show that, in terms of x and y, the equation of the least squares line is:

$$y = 0.930x + 61.2 \text{ (correct to three significant figures).}$$

c Rewrite the equation in terms of the variables involved.

3B Performing a regression analysis

Having learned how to calculate the equation of the least squares line, you are well on the way to learning how to perform a full regression analysis. In the process, you will need to use of many of the skills you have so far developed when working with scatterplots and correlation coefficients.

Elements of a regression analysis

A full regression analysis involves several processes, which include:

- constructing a *scatterplot* to investigate the nature of an association
- calculating the *correlation coefficient* to indicate the strength of the relationship
- determining the equation of the *least squares line*
- *interpreting the coefficients*, the slope (a) and the y-intercept (b), of the least squares line $y = ax + b$
- using the *coefficient of determination* to indicate the *predictive power* of the association
- using the *least squares line* to make *predictions*
- calculating residuals and using a *residual plot* to test the *assumption of linearity*
- writing a *report* on your findings.

An analysis using some data

We wish to investigate the nature of the association between the price of a second-hand car and its age. The ultimate aim is to find a mathematical model that will enable the price of a second-hand car to be predicted from its age.

To this end, the age (in years) and price (in dollars) of a selection of second-hand cars of the same brand and model have been collected and are recorded in a table (shown).

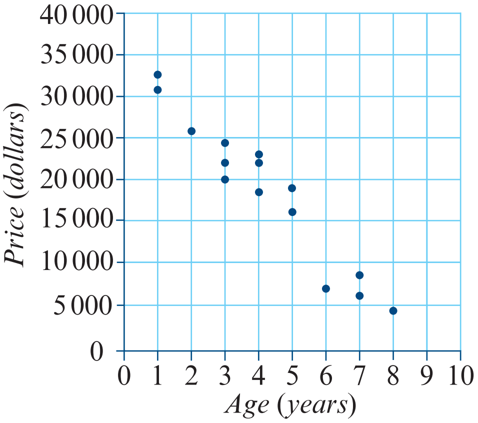
Age (years)	*Price (dollars)*	*Age (years)*	*Price (dollars)*
1	32 500	4	19 200
1	30 500	5	16 000
2	25 600	5	18 400
3	20 000	6	6 500
3	24 300	7	6 400
3	22 000	7	8 500
4	22 000	8	4 200
4	23 000		

Scatterplot and correlation coefficient

We start our investigation of the association between price and age by constructing a scatterplot and using it to describe the association in terms of strength, direction and form. In this analysis, *age* is the explanatory variable.

From the scatterplot, we see that there is a *strong negative linear association* between the price of the car and its age. There are no clear outliers. The correlation coefficient is $r = -0.964$.

We would communicate this information in a report as follows.



Report

There is a strong negative linear association between the price of these second-hand cars and their age ($r = -0.964$).

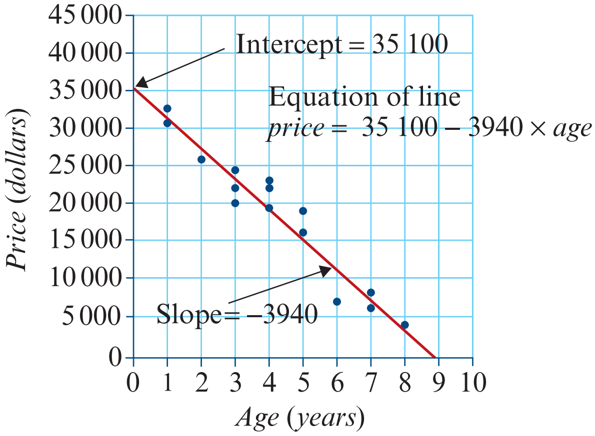
Fitting a least squares line to the data

Because the association is linear, it is reasonable to use a least squares line to model the association.

Using a calculator, the equation of the least squares line for this data is:

$$price = 35\,100 - 3940 \times age$$

This line has been plotted on the scatterplot as shown.



We now have a mathematical model to describe, *on average*,[1] how the price of this type of second-hand car changes with time.

Interpreting the slope and the intercept of a least squares line

The two key values in our mathematical model are the slope of the line (−3940) and the y-intercept (35 100), and a key step in performing a regression analysis is to know how to interpret these values in terms of the variables *price* and *age*.

Interpreting the slope and the intercept of a least squares line

For the least squares line $y = ax + b$

- the slope (a) estimates the average change (increase/decrease) in the *response variable* (y) for each one-unit increase in the *explanatory variable* (x)
- the y-intercept (b) estimates the average value of the *response variable* (y) when the *explanatory variable* (x) equals 0.

Note: The interpretation of the y-intercept in a data context can be problematic when $x = 0$ is not within the range of observed x-values.

---

[1] We say ‘on average’ because the line does not pass through every point. Rather, it is a line that is drawn through the average price of this type of car at each age.

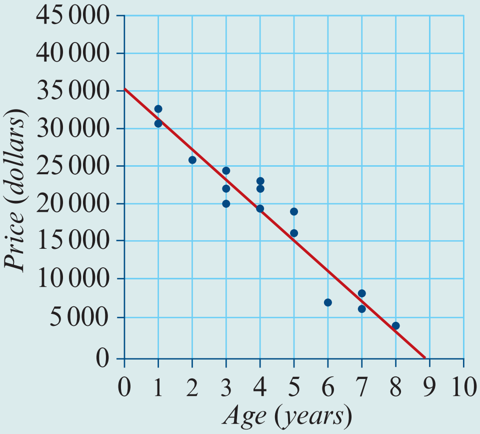
Example 1 Interpreting the slope and the intercept of a least squares line

The equation of a least squares line that enables the *price* of a second-hand car to be predicted from its *age* is:

$$price = 35\,100 - 3940 \times age$$

a Interpret the slope in terms of the variables *price* and *age*.

b Interpret the intercept in terms of the variables *price* and *age*.



Solution

a The *slope* predicts the average change (increase/decrease) in the *price* for each 1-year increase in the *age*. Because the slope is negative, it will be a decrease.

On average, the price of these cars decreases by $3940 each year.[2]

b The *intercept* predicts the value of the *price* of the car when *age* equals 0; that is, when the car is new.

On average, the price of these cars when new is $35 100.

[2] A general model for writing about the slope is that, on average, the Response variable increases/decreases by the value of the 'slope' for each unit increase in the Explanatory variable.

Using a least squares line to make predictions

The aim of an analysis of association is to model the relationship between two numerical variables by using the equation of a straight line. This equation can then be used to make predictions.

Example 2 Using a least squares line to make predictions

The equation of a least squares line that enables the *price* of a second-hand car to be predicted from its *age* is:

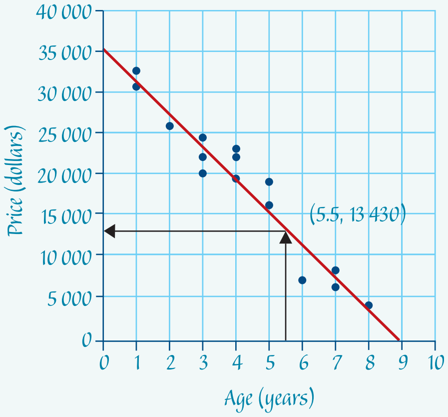
$$price = 35\,100 - 3940 \times age$$

Use this equation to predict the price of a car that is 5.5 years old.

Solution

There are two ways this can be done. One is to draw a vertical arrow at *age* = 5.5 up to the graph and then horizontally across to the *price* axis as shown, to get an answer of around $14 000. A more accurate answer is obtained by substituting *age* = 5.5 into the equation to obtain $13 430, as shown below.

$$\text{Price} = 35\,100 - 3940 \times 5.5$$
$$= \$13\,430$$



The data below shows the times that 10 students spent studying for an exam and the marks they subsequently obtained.

Time (t, hours)	4	36	23	19	1	11	18	13	18	8
Mark (m, %)	41	87	67	62	23	52	61	43	65	52

If we fitted a linear model to this data using the least squares method, we would have an equation close to:

$$m = 1.62t + 30.8$$

Using this equation, and rounding off to the nearest whole number, we would predict that a student who spent:

- 0 hours studying would obtain a mark of 31% ($m = 1.62 \times 0 + 30.8 = 31\%$)
- 8 hours studying would obtain a mark of 44% ($m = 1.62 \times 8 + 30.8 = 44\%$)
- 12 hours studying would obtain a mark of 50% ($m = 1.62 \times 12 + 30.8 = 50\%$)
- 30 hours studying would obtain a mark of 79% ($m = 1.62 \times 30 + 30.8 = 79\%$)
- 80 hours studying would obtain a mark of 160% ($m = 1.62 \times 80 + 30.8 = 160\%$)

This last result, 160%, points to one of the limitations of substituting into an equation of a least squares line without thinking carefully. Using this equation, we predict that a student who studies for 80 hours will obtain a mark of more than 100%: impossible. Something is wrong!

Interpolation and extrapolation

The problem is that we are using the least squares line equation to make predictions well outside the range of values used to calculate this equation. The maximum time any student spent studying for this exam was 36 hours; yet, we are using the equation we calculated to try to predict the exam mark for someone who studies for 80 hours. Without knowing that the model works equally well for someone who spends 80 hours studying. We are venturing into unknown territory and can have little faith in our predictions.

As a general rule, a least squares equation only applies to the range of data values used to determine the equation. Thus, we are reasonably safe using the line to make predictions that lie roughly within this data range, say from 1 to 36 hours. The process of making a prediction within the range of data used to derive the equation is called **interpolation** and we can have some faith in these predictions.

However, we must be extremely careful about how much faith we put into predictions made outside the data range. Making predictions outside the data range is called **extrapolation**.

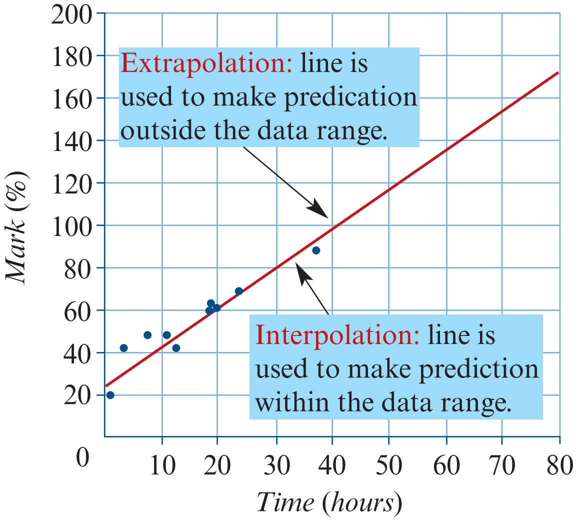
Predicting *within* the range of data is called interpolation.

In general, we can expect a reasonably reliable result when interpolating.

Predicting *outside* the range of data is called extrapolation.

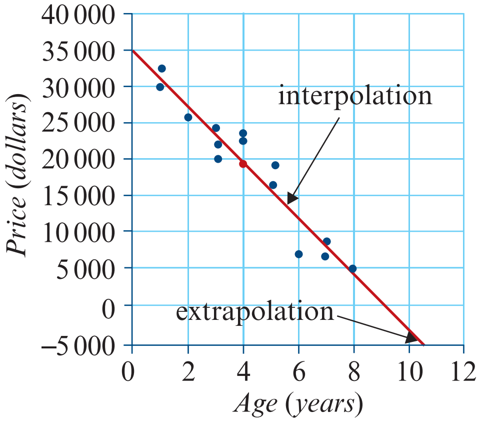
With extrapolation, we have no way of knowing whether our prediction is reliable or not.

For example, if we use the least squares line to predict the examination mark for 30 hours of studying time, we would be interpolating. However, if we use the least squares line to predict the examination mark for 50 hours of studying time, we would be extrapolating. Extrapolation is a less reliable process than interpolation because we are going beyond the original data.



ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024



For example, using the least squares line to predict the price of a car less than 5.5 years old would be an example of *interpolation*. This is because we are making a prediction *within* the data.

However, using the least squares line to predict the price of a 10-year-old car would be an example of *extrapolation*. This is because we are working *outside* the data.

The fact that the least squares line would predict a negative value for the car highlights the potential problems of making predictions beyond the data (extrapolating).

Example 3 Using a least squares line to make predictions

The equation relating the weights (w, in kg) and heights (h, in cm) of a group of students whose heights ranged from 163 cm to 190 cm is:

$$w = 0.6h - 40$$

Use this equation to predict the weight (in kg) of students with the following heights. Are you interpolating or extrapolating?

a 170 cm **b** 65 cm

Solution

a Substitute 170 into the equation and evaluate.	The weight of a person of height 170 cm is predicted to be: $w = 0.6 \times 170 - 40 = 62$ kg Interpolating: predicting within range of data.
b Substitute 65 into the equation and evaluate.	The weight of a person of height 65 cm is predicted to be: $w = 0.6 \times 65 - 40 = -1$ kg which is not possible. Extrapolating: predicting well outside the range of data.

The coefficient of determination

The coefficient of determination is a measure of the predictive power of a regression equation. While the association between the price of a second-hand car and its age does not explain all the variation in price, knowing the age of a car does give us some information about its likely price.

For a perfect relationship, the least squares line explains 100% of the variation in prices. In this case, with $r = -0.964$ we have the:

$$\text{coefficient of determination} = r^2 = 0.964^2 \approx 0.930 \text{ or } 93.0\%$$

Thus, we can conclude that:

> 93% of the variation in price of the second-hand cars can be explained by the variation in the ages of the cars.[3]

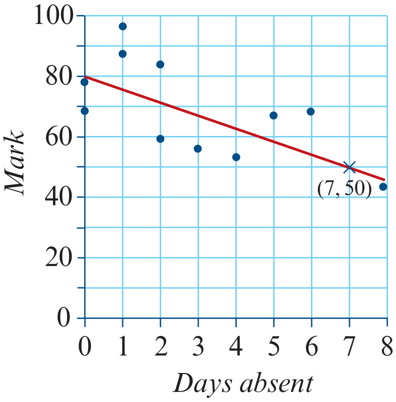
In this case, the regression equation has highly significant (worthwhile) predictive power. As a guide, any relationship with a coefficient of determination greater than 30% can be regarded as having significant predictive power.

Skill-sheet

Exercise 3B

Some basics

1 Use the line on the scatterplot opposite to determine the equation of the least squares line in terms of the variables, mark, *m*, and days absent, *d*.
Give the intercept and the slope correct to the nearest whole number.



Fitting a least squares line and calculating the correlation coefficient

CAS

2 The table below shows the weight (in kg) and blood glucose level (in mg/100 mL) of eight adults.

Weight (w, kg)	82.1	70.1	76.6	82.1	83.9	73.2	66.0	77.5
Glucose (g, mg/100 mL)	101	89	98	100	108	104	94	89

a Construct a scatterplot to display these data with *weight* as the Explanatory variable.

b Fit a least squares line to the scatterplot and determine its equation.

c Write the equation of the least squares line in terms of the variables *glucose* and *weight* with the coefficients given correct to three significant figures.

d Determine the correlation coefficient to three significant figures.

3 The table below shows the scores a group of nine students obtained on two class tests, Test *A* and Test *B*, as part of their school based assessment.

Test A	33	45	27	42	50	38	17	35	29
Test B	43	46	36	34	48	34	29	41	28

---

[3] A general model for writing about the coefficient of determination is: '$r^2 \times 100$ of the variation in the Response variable is explained by the variation in the Explanatory variable'.

a Construct a scatterplot to display these data with *Test A* as the Explanatory variable.

b Fit a least squares line to the scatterplot and determine its equation.

c Write the equation of the least squares line using the variables *A* and *B* to represent the two tests and with the coefficients given correct to three significant figures.

d Determine the correlation coefficient to three significant figures.

4 The table below shows the carbohydrate content (carbs) and the fat content (fat) in 100 g of nine breakfast cereals.

Carbs (c)	88.7	67.0	77.5	61.7	86.8	32.4	72.4	77.1	86.5
Fat (f)	0.3	1.3	2.8	7.6	1.2	5.7	9.4	10.0	0.7

a Construct a scatterplot to display these data with *carbs* as the Explanatory variable.

b Fit a least squares line to the scatterplot and determine its equation.

c Write the equation of the least squares line in terms of the variables *fat* and *carbs* with the coefficients given correct to three significant figures.

d Determine the correlation coefficient to three significant figures.

5 The table below shows the age and height of six young girls.

Age (months)	36	40	44	52	56	60
Height (cm)	84	87	90	92	94	96

a Construct a scatterplot to display these data with *age* as the Explanatory variable.

b Fit a least squares line to the scatterplot and determine its equation.

c Write the equation of the least squares line in terms of the variables *height* and *age* with the coefficients given correct to three significant figures.

d Determine the correlation coefficient to three significant figures.

6 The following table gives the shoe size and weight in kilograms of 10 adult males.

Shoe size	9.5	10.0	10.5	10.5	11	9.0	8.5	9.5	7.5	8
Weight (kg)	64	85	70	80	82	73	70	66	55	70

a Construct a scatterplot to display these data with *shoe size* as the Explanatory variable.

b Fit a least squares line to the scatterplot and determine its equation.

c Write the equation of the least squares line in terms of the variables *weight* and *shoe size* with the coefficients given correct to three significant figures.

d Determine the correlation coefficient to three significant figures.

ISBN 978-1-009-25917-0

Using a least squares line to make predictions: interpolation and extrapolation

7 Complete the following sentences.

Using a least squares line to make a prediction:

a within the range of data is called ☐.

b outside the range of data is called ☐.

Example 3 **8** For children between the ages of 36 and 60 months, the equation relating their height (h, in cm) to their age (a, in cm) is:

$$h = 0.4a + 72$$

Use this equation to predict the height (to the nearest cm) of a child with the following age. Are you interpolating or extrapolating?

a 40 months old **b** 55 months old **c** 70 months old

9 For shoe sizes between 6 and 12, the equation relating a person's *weight* (w, in kg) to shoe size (s) is:

$$w = 2.2s + 48.1$$

Use this equation to predict the weight (to the nearest kg) of a person with the following shoe size. Are you interpolating or extrapolating?

a 5 **b** 8 **c** 11

10 When preparing between 25 and 100 meals (m), a cafeteria's *cost* (c, in dollars) is given by the equation:

$$c = 5.8m + 175$$

Use this equation to predict the cost (to the nearest dollar) of preparing the following meals. Are you interpolating or extrapolating?

a no meals **b** 60 meals **c** 89 meals

11 For women of heights from 150 cm to 180 cm, the equation relating a *daughter's height* (y, in cm) to her *mother's height* (x, in cm) is:

$$y = 0.91x + 18.3$$

Use this equation to predict (to the nearest cm) the adult height of a woman whose mother is the following heights. Are you interpolating or extrapolating?

a 168 cm tall **b** 196 cm tall **c** 155 cm tall

Reading an equation of a least squares line and making predictions

12 The equation of a least squares line that enables hand span (y) to be predicted from height (x) is:

$$y = 0.33x + 2.9$$

Complete the following sentences, by filling in the boxes.

a The explanatory variable is ☐.

ISBN 978-1-009-25917-0

b The slope equals ☐ and the intercept equals ☐.

c A person is 160 cm tall. The least squares line predicts a hand span of ☐ cm.

13 For a 100 km trip, the equation of a least squares line that enables fuel consumption of a car, f (in litres) to be predicted from its weight, w (kg) is:

$$f = 0.01w - 0.1$$

Complete the following sentences.

a The response variable is ☐.

b The slope is ☐ and the intercept is ☐.

c A car weighs 980 kg. The least squares line predicts a fuel consumption of ☐ litres.

Interpreting an equation of a least squares line and its coefficient of determination

14 In an investigation of the relationship between the food energy content (y, in calories) and the fat content (x, in g) in a standard-sized packet of chips, the least squares line was found to be:

$$y = 14.7x + 27.8 \quad r^2 = 0.7569$$

Use this information to complete the following sentences.

a The slope is ☐ and the intercept is ☐.

b The least squares line equation predicts that the food energy content in a packet of chips increases by ☐ calories for each additional gram of fat it contains.

c $r =$ ☐

d ☐ % of the variation in food energy content of a packet of chips can be explained by the variation in their ☐.

e The fat content of a standard-sized packet of chips is 8 g. The least squares line equation predicts its food energy content to be ☐ calories.

15 In an investigation of the relationship between the success rate (s, %) of sinking a putt and the distance from the hole (d, in cm) of amateur golfers, the least squares line was found to be:

$$s = 98.5 - 0.278\,d \quad r^2 = 0.497$$

a Write the slope of this least squares line equation and interpret.

b Use the equation to predict the success rate when a golfer is 90 cm from the hole.

c At what distance (in metres) from the hole does the least squares line equation predict an amateur golfer to have a 0% success rate of sinking the putt?

d Calculate the value of r, correct to three decimal places.

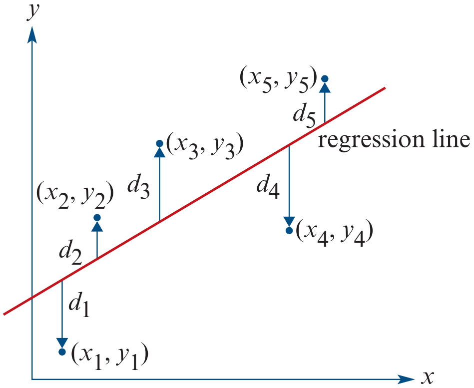
e Write the value of the coefficient of determination as a percentage and interpret.

3C Residuals

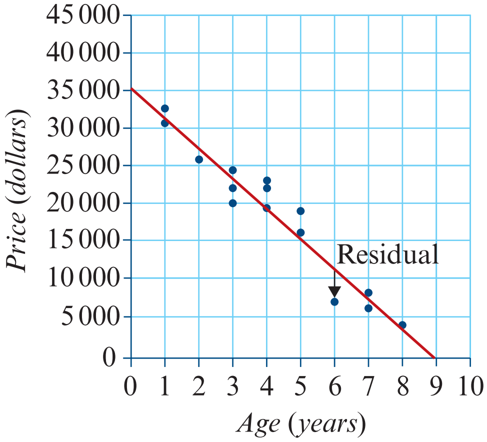
Calculating residuals

As we saw earlier (Section 3A), the residuals are the vertical distances between the individual data points and the least squares line.

We can estimate the value of a residual directly from the scatterplot.



For example, from the scatterplot, we can see that the *residual value* for the 6-year-old car is around –\$5000. That is, the actual price of the car is approximately \$5000 *less* than we would predict.



Residuals

residual value = actual data value − predicted value

Residuals can be positive, negative or zero.

- Data points *above* the least squares line have a *positive residual*.
- Data points *below* the least squares line have a *negative residual*.
- Data points *on* the line have a *zero residual*.

ISBN 978-1-009-25917-0

To determine the value of a residual more precisely, we need to do a calculation.

Calculating residuals

A residual can be calculated as:

$$residual = y - \hat{y}$$

where y denotes the actual data value and $\hat{y}$ denotes the predicted value.

The predicted value ($\hat{y}$) is obtained from substituting a given x value into the equation of least squares line.

$$\hat{y} = ax + b$$

Example 4 Calculating a residual

The actual price of the 6-year-old car is \$6500. Calculate the residual when its price (y) is predicted from its age (x) using the least squares line:
$y = -3940x + 35\,100$

Solution

1 Write the actual price.	Actual price: $y = \$6500$
2 Determine the predicted price using the regression equation: $\hat{y} = -3940x + 35\,100$	Predicted price $\hat{y} = -3940 \times 6 + 35\,100$ $= 11\,460$ $\therefore$ Predicted price $= \$11\,460$
3 Determine the residual. Residual = actual − predicted	Residual $= y - \hat{y}$ $= 6500 - 11\,460$ $= -\$4960$

The residual plot: testing the assumption of linearity

A key *assumption* made when calculating a least squares line is that the relationship between the variables is *linear*.

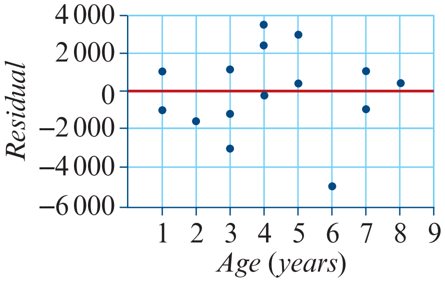
One way of testing this assumption is to plot the least squares line on the scatterplot and see how well a straight line fits the data. However, a better way is to use a residual plot, as this plot will show small departures from linearity.

Using a previous example of the association between the age and price of a selection of second-hand cars residual values can be calculated for each of the data points.

The least squares regression equation for these data is $y = -3940x + 35\,100$

Age (x, years)	1	1	2	3	3	3	4	4	4	5	5	6	7	7	8
Price (y, dollars)	32 500	30 500	25 600	20 000	24 300	22 000	22 000	23 000	19 200	16 000	18 400	6500	6400	8500	4200
Predicted value ($\hat{y}$)	31 160	31 160	27 220	23 280	23 280	23 280	19 340	19 340	19 340	15 400	15 400	11 460	7520	7520	3580
Residual $y - \hat{y}$	1340	−660	−1620	−3280	1020	−1280	2660	3660	−140	600	3000	−4960	−1120	980	620

A **residual plot** is a plot of the residual value for each data value against the independent variable (in this case, *age*). Because the mean of the residuals is always zero, the horizontal zero line (red) helps us to orient ourselves. This line corresponds to regression in the previous scatterplot.



From the residual plot, we see that there is *no clear pattern*[4] in the residuals. Essentially they are *randomly scattered* around the zero least squares line.

Thus, from this residual plot we can report as below.

Report

The lack of a clear pattern in the residual plot confirms the assumption of a linear association between the price of a second-hand car.

---

[4] From a visual inspection, it is difficult to say with certainty that a residual plot is random. It is easier to see when it is not random, as you will see in the next chapter. For present purposes, it is sufficient to say that a clear lack of pattern in a residual plot indicates randomness.

ISBN 978-1-009-25917-0

If the residual plot shows a clear pattern in the residuals then we say that a linear model is *not appropriate* for the data, as in the next example.

Example 5 Using the residual plot

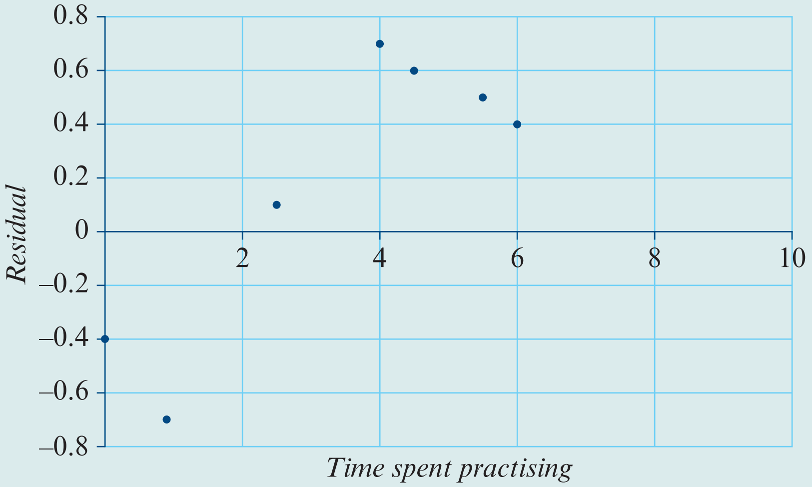
The table below shows data relating to time spent practising and performance level.

Time spent practising (x)	0.0	0.9	2.5	4.0	4.5	5.5	6.0	7.0	9.5
Performance level (y)	1.0	1.1	2.5	3.7	3.8	4.1	4.2	4.0	4.5
Predicted value ($\hat{y}$)	1.4	1.8	2.4	3	3.2	3.6	3.8	*A*	5.2
Residual $y - \hat{y}$	–0.4	–0.7	0.1	0.7	0.6	0.5	0.4	*B*	*C*

The least squares regression line for this data has been determined as $y = 0.4x + 1.4$

a Determine the values of *A*, *B* and *C* to complete the table above.

b Complete the residual plot below by adding the last two residual values.



c Use the residual plot to comment on the appropriateness of assuming that time spent practising and performance level are linearly associated.

Solution

a Predicted value A:
Calculated by substituting the corresponding *x* value into the equation of least squares regression line.

$$y = 0.4x + 1.4$$

Residual values B & C:
Residual calculated by actual value – predicted
Residual = $y - \hat{y}$

When $x = 7.0$:

$$\hat{y} = 0.4 \times 7.0 + 1.4 = 4.2$$

$\therefore$ Value of $A = 4.2$

Residual $= 4.0 - 4.2 = -0.2$

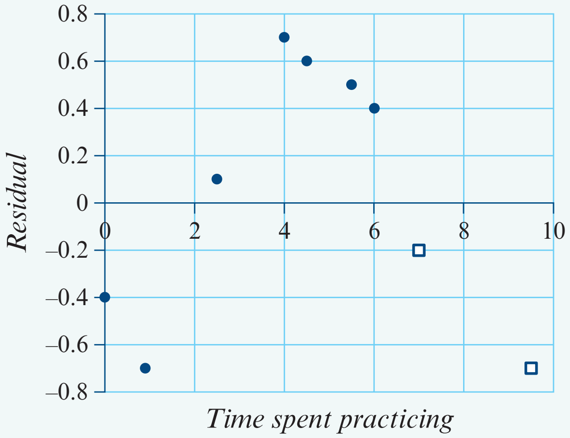
$\therefore$ Value of $B = -0.2$

When $x = 9.5$

Residual $= 4.5 - 5.2 = -0.7$

$\therefore$ Value of $C = -0.7$

b Plot residual points. Negative residuals are below the x-axis.



c A linear association is appropriate if the residual plot has no pattern (i.e. points randomly scattered).

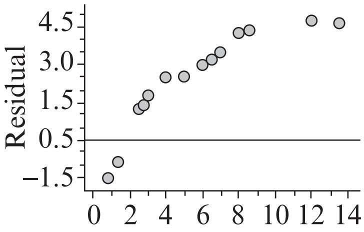
The residual plot shows a clear pattern in the residuals (approximately quadratic in nature). Therefore a linear model is not suitable for this data.

Exercise 3C

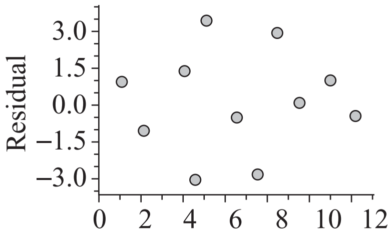
Interpreting residual plots

1 Each of the following residual plots has been constructed after a least squares line has been fitted to a scatterplot. Determine whether each residual plot suggests that the use of a linear model is appropriate for the data. Justify your answer.

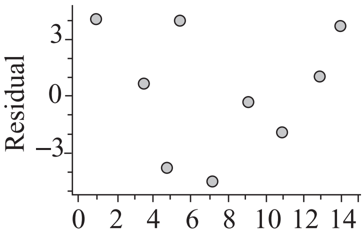
a



b



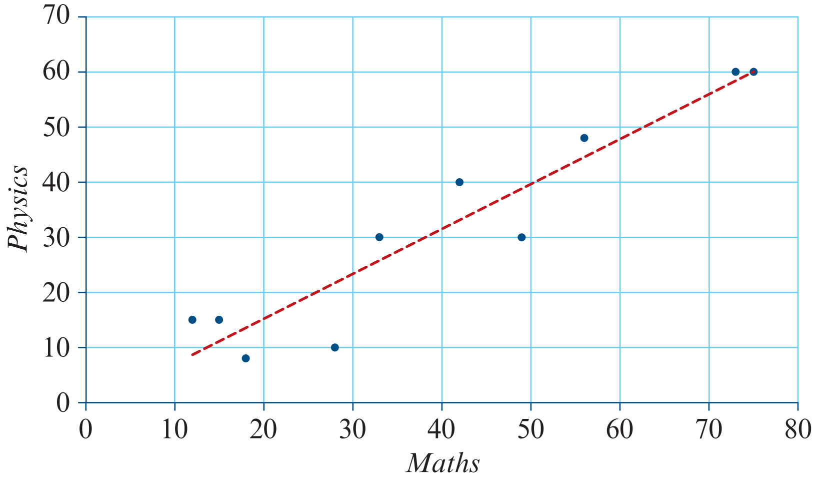
c



ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

2 The scatterplot below show the marks obtained by students in mathematics and physics.



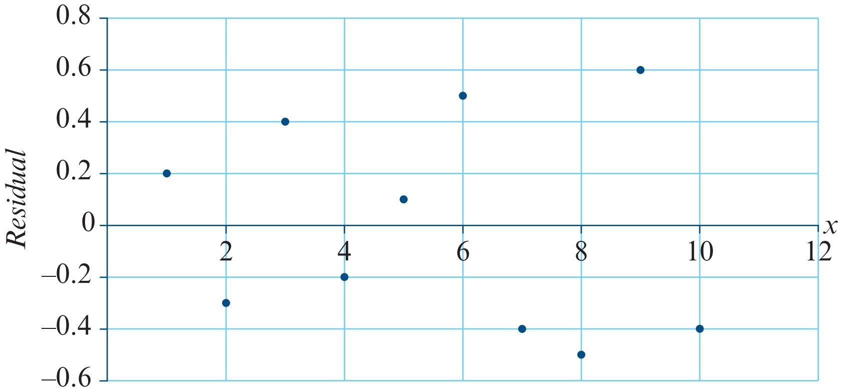
Determine whether the residual would be positive or negative for each of the following maths marks:

a 56

b 28

c 49

3 The graph below displays a residual plot that was constructed after completing a least squares regression on a set of bivariate numerical data.



Determine if the following statements are true or false for these data. Give a reason for your answer.

a When $x = 3$, the least squares regression equation underestimates the value of y.

b There appears to be a linear relationship between x and y.

4 A city council recorded data over an extended time. The data recorded the number of workmen used to lay brick paving for a public footpath (per kilometre), and the number of days required to complete the path.

Number of workmen (x)	2	4	5	6	8	11
Number of days (y)	18	12	10	8	7	5

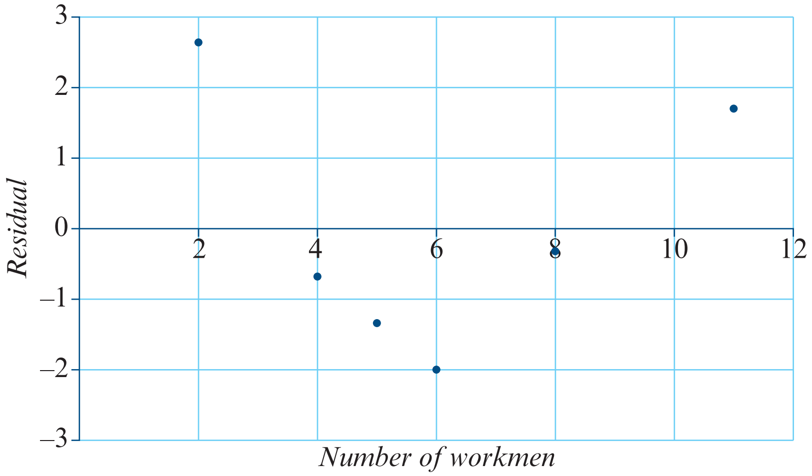
a Use your CAS calculator to determine the equation for the least squares line for these data, giving coefficients to two decimal places.

b **i** Use the least squares regression line to predict the number of days it will take 6 workmen to lay brick paving.

ii Determine the residual for 6 workmen.

iii Explain what this residual value tells you about how the predicted number of days to lay the brick paving compares to the actual value.

c By considering the residual plot below, comment on the appropriateness of a linear model for this data.



5 A student believed that the fuel consumption of a car is closely related to the mass of the vehicle. The following data shows a summary of the mass and fuel consumption of seven types of cars.

Mass (x)	1342	1651	880	1170	911	1152	1250
Fuel Consumption (y)	14.9	16.9	9.0	15.2	9.8	10.3	12.8
Predicted value ($\hat{y}$)	14.23	17.41	9.47	B	9.79	12.27	13.28
Residual ($y - \hat{y}$)	0.67	A	–0.47	C	0.01	–1.97	–0.48

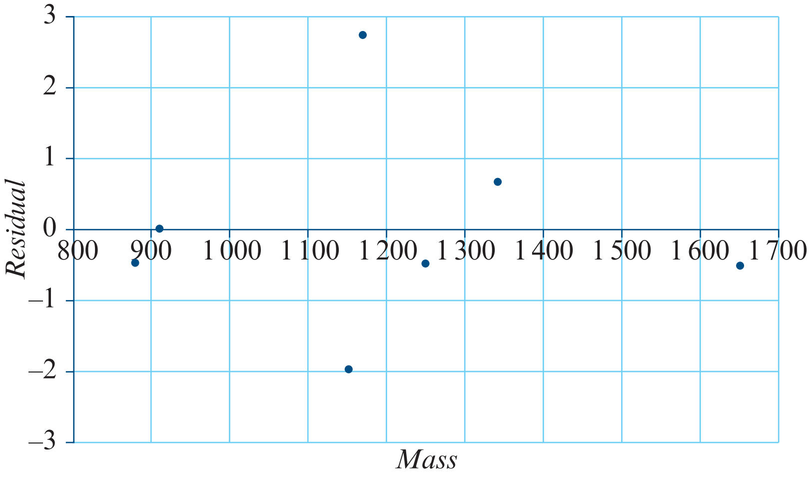
a Use your CAS calculator to determine the equation for the least squares line for these data, giving coefficients to three significant figures.

b Complete the table by determining the values of A, B and C.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

c A residual plot for this data is given below.

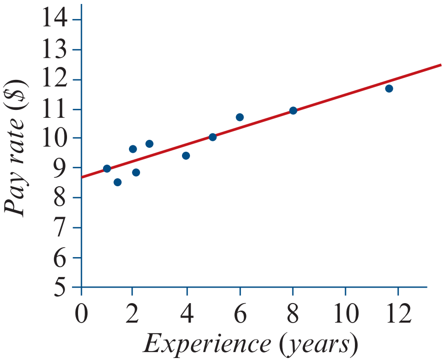


Use the residual plot to comment on the appropriateness of assuming that a car's mass and fuel consumption are linearly associated.

Conducting a regression analysis including the use of residual plots

6 The scatterplot opposite shows the pay rate (dollars per hour) paid by a company to workers with different years of work experience. Using a calculator, the equation of the least squares line is found to have the equation:

$$y = 8.56 + 0.289x \quad \text{with } r = 0.967$$



a Is it appropriate to fit a least squares line to the data? Why?

b Work out the coefficient of determination.

c What percentage of the variation in a person's pay rate can be explained by the variation in their work experience?

d Write the equation of the least squares line in terms of the variables *pay rate* and years of *experience*.

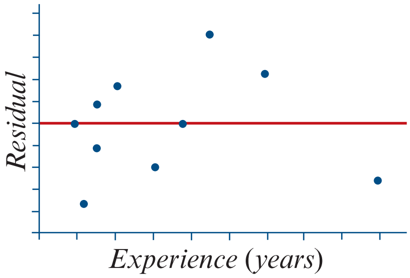
e Interpret the y-intercept in terms of the variables *pay rate* and years of *experience*. What does the y-intercept tell you?

f Interpret the slope in terms of the variables *pay rate* and years of *experience*. What does the slope of the least squares line tell you?

g Use the least squares line equation to:

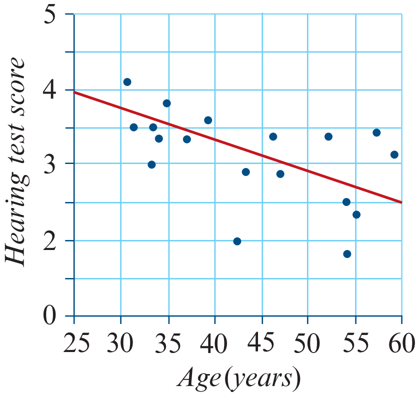
- **i** predict the hourly wage of a person with 8 years of experience
- **ii** determine the residual value if the person's actual hourly wage is \$11.20 per hour.

h The residual plot for this regression analysis is shown opposite. Does the residual plot support the initial assumption that the relationship between *pay rate* and years of *experience* is linear?
Explain your answer.



7 The scatterplot opposite shows scores on a hearing test against age. In analysing the data, a statistician produced the following statistics:

- coefficient of determination: $r^2 = 0.370$
- least squares line: $y = 4.9 - 0.043x$



a Determine the value of the correlation coefficient, r, for the data.

b Interpret the coefficient of determination in terms of the variables *hearing test score* and *age*.

c Write the equation of the least squares line in terms of the variables *hearing test score* and *age*.

d Write the slope and interpret.

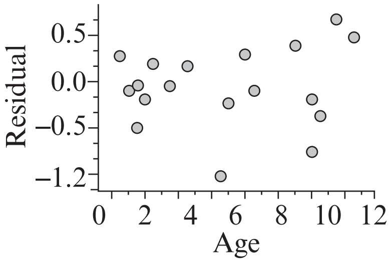
e Use the least squares line equation to:

- **i** predict the hearing test score of a person who is 20 years old
- **ii** determine the residual value if the person's actual hearing test score is 2.0.

f Use the graph to estimate the value of the residual for the person aged:

- **i** 35 years
- **ii** 55 years

g The residual plot for this regression analysis is shown opposite.
Does the residual plot support the initial assumption that the relationship between hearing test score and age is essentially linear?
Explain your answer.



ISBN 978-1-009-25917-0

3D Conducting a regression analysis using raw data

In a statistical investigation project you will need to be able to conduct a full regression analysis from raw data. This section is designed to help you with this task.

How to conduct a regression analysis using the TI-Nspire CAS

This analysis is concerned with investigating the association between life expectancy (in years) and birth rate (in births per 1000 people) in 10 countries.

Birth rate	30	38	38	43	34	42	31	32	26	34
Life expectancy (years)	66	54	43	42	49	45	64	61	61	66

Steps

1 Write the explanatory variable (EV) and response variable (RV). Use the variable names *birth* and *life*.

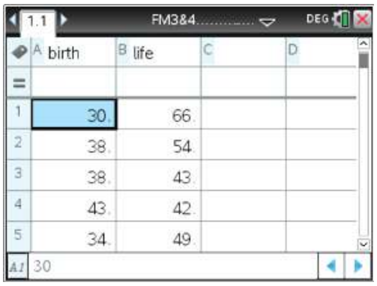
EV: birth

RV: life

2 Start a new document by pressing ctrl + N.

Select **Add Lists & Spreadsheet**.

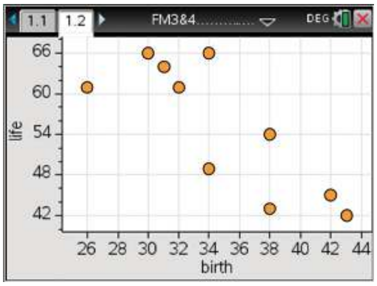
Enter the data into the lists named *birth* and *life*, as shown.



	A birth	B life	C	D
=				
1	30	66		
2	38	54		
3	38	43		
4	43	42		
5	34	49		

A1 30

3 Construct a scatterplot to investigate the nature of the relationship between life expectancy and birth rate.



4 Describe the association shown by the scatterplot. Mention direction, form, strength and outliers.

There is a strong, negative, linear relationship between life expectancy and birth rate. There are no obvious outliers.

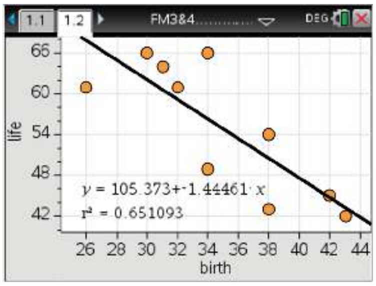
ISBN 978-1-009-25917-0

Cambridge University Press

Updated January 2024

5 Find and plot the equation of the least squares line and r^2 value.

Note: Check if **Diagnostics** is activated using `menu`>**Settings**.

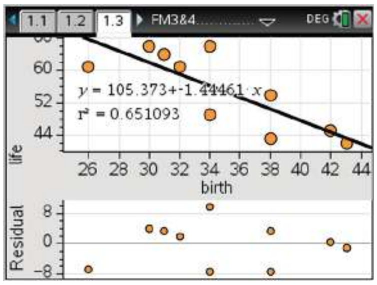


$y = 105.373+-1.44461 \cdot x$
$r^2 = 0.651093$

6 Generate a residual plot to test the linearity assumption.

Use `ctrl` + ◀ (or click on the page tab) to return to the scatterplot.

Press `menu`>**Analyze>Residuals>Show Residual Plot** to display the residual plot on the same screen.



$y = 105.373+-1.44461 \cdot x$
$r^2 = 0.651093$

7 Use the values of the intercept and slope to write the equation of the least squares line. Also write the values of r and the coefficient of determination.

Least squares line equation:

life = 105.4 – 1.445 × birth

Correlation coefficient: $r = -0.8069$

Coefficient of determination: $r^2 = 0.651$

How to conduct a regression analysis using the ClassPad

This analysis is concerned with investigating the association between life expectancy (in years) and birth rate (in births per 1000 people) in 10 countries.

Birth rate (per thousand)	30	38	38	43	34	42	31	32	26	34
Life expectancy (years)	66	54	43	42	49	45	64	61	61	66

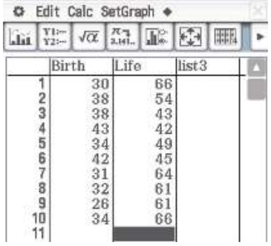
Steps

1 Write the explanatory variable (EV) and response variable (RV). Use the variable names *birth* and *life*.

EV: birth

RV: life

2 Enter the data into lists as shown.



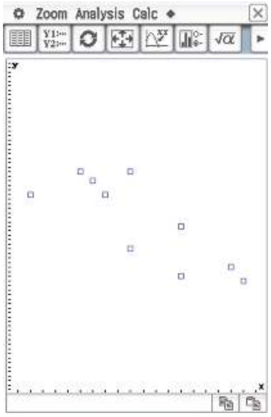
	Birth	Life	list3
1	30	66	
2	38	54	
3	38	43	
4	43	42	
5	34	49	
6	42	45	
7	31	64	
8	32	61	
9	26	61	
10	34	66	
11			

3 Construct a scatterplot to investigate the nature of the relationship between life expectancy and birth rate.

a Tap and complete the **Set Calculations** dialog box as shown.

b Tap to view the scatterplot.

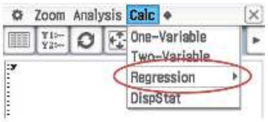
Draw: ● On ○ Off
Type: Scatter
XList: main\Birth
YList: main\Life
Freq: 1
Mark: square

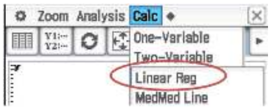


4 Describe the association shown by the scatterplot. Mention direction, form, strength and outliers.

There is a strong negative, linear association between life expectancy and birth rate. There are no obvious outliers.

5 Find the equation of the least squares line and generate all regression statistics, including residuals.



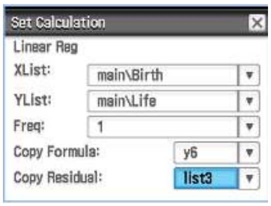


a Tap **Calc** in the toolbar.

Tap **Regression** and select **Linear Reg**.

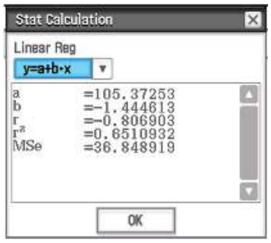
b Complete the **Set Calculations** dialog box as shown.

Note: **Copy Residual** copies the residuals to **list3**, where they can be used later to create a residual plot.



Linear Reg
XList: main\Birth
YList: main\Life
Freq: 1
Copy Formula: y6
Copy Residual: list3

c Tap **OK** in the Set **Calculation** box to generate the regression results.



Linear Reg
y=a+b·x

```
a      =105.37253
b      =-1.444613
r      =-0.806903
r²     =0.6510932
MSe    =36.848919
```

d Write the key results.

Least squares line equation:

life = 105.4 − 1.445 × birth

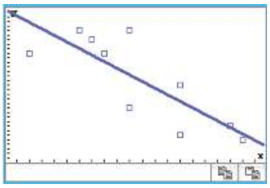
Correlation coefficient:

$r = -0.8069$

Coefficient of determination:

$r^2 = 0.651$

6 Tapping **OK** a second time automatically plots and displays the regression line on the scatterplot.



To obtain a full-screen plot, tap from the icon panel.

7 Generate a residual plot to test the linearity assumption.

Tap and complete the **Set Calculations** dialog box as shown.

Tap to view the residual plot.

Set StatGraphs

Draw: ● On ○ Off

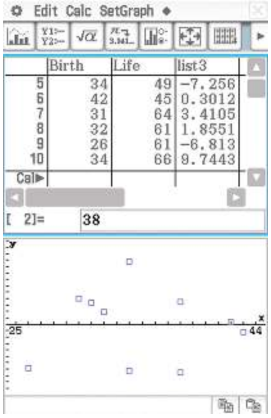
Type: Scatter

XList: main\Birth

YList: list3

Freq: 1

Mark: square



	Birth	Life	list3
5	34	49	−7.256
6	42	45	0.3012
7	31	64	3.4105
8	32	61	1.8551
9	26	61	−6.813
10	34	66	9.7443

[2]= 38

Inspect the plot and write your conclusion.

The random residual plot suggests linearity.

Note: When you performed a regression analysis earlier, the residuals were calculated automatically and stored in **list3**. The residual plot is a scatterplot with **list3** on the vertical axis and **birth** on the horizontal axis.

Exercise 3D

CAS

1 The table below shows the scores obtained by nine students on two tests. We want to be able to predict test B scores from test A scores.

Test A score (x)	18	15	9	12	11	19	11	14	16
Test B score (y)	15	17	11	10	13	17	11	15	19

Use your calculator to perform each of the following steps of a regression analysis.

a Construct a scatterplot.

b Determine the equation of the least squares line along with the values of r and r^2.

c Display the least squares line on the scatterplot.

d Obtain a residual plot.

2 The table below shows the number of careless errors made on a test by nine students. Also given are their test scores. We want to be able to predict test score from the number of careless errors made.

Test score	18	15	9	12	11	19	11	14	16
Careless errors	0	2	5	6	4	1	8	3	1

Use your calculator to perform each of the following steps of a regression analysis.

a Construct a scatterplot.

b Determine the equation of the least squares line along with the values of r and r^2. Write answers correct to three significant figures.

c Display the least squares line on the scatterplot.

d Obtain a residual plot.

3 How well can we predict an adult's weight from their birth weight? The weights of 12 adults were recorded, along with their birth weights. The results are shown.

Birth weight (kg)	1.9	2.4	2.6	2.7	2.9	3.2	3.4	3.4	3.6	3.7	3.8	4.1
Adult weight (kg)	47.6	53.1	52.2	56.2	57.6	59.9	55.3	58.5	56.7	59.9	63.5	61.2

a In this investigation, which would be the Response variable and which would be the Explanatory variable?

b Construct a scatterplot.

c Use the scatterplot to:

  i comment on the relationship between adult weight and birth weight in terms of direction, outliers, form and strength

  ii *estimate* the value of the correlation coefficient, r.

d Determine the equation of the least squares line, the coefficient of determination and the value of the correlation coefficient, r. Write answers correct to three significant figures.

e Interpret the coefficient of determination in terms of adult weight and birth weight.

f Interpret the slope in terms of adult weight and birth weight.

g Use the least squares line equation to predict the weight of an adult with a birth weight of:

i 3.0 kg **ii** 2.5 kg **iii** 3.9 kg

Give answers correct to one decimal place.

h It is generally considered that birth weight is a 'good' predictor of adult weight. Do you think the data support this contention? Explain.

i Construct a residual plot and use it to comment on the appropriateness of assuming that adult weight and birth weight are linearly associated.

3E Writing statistical reports

The final step in a regression analysis is to report your findings. The report below, on the second-hand car price/age data in section 3B is in a form that is suitable for inclusion in a statistical investigation project.

Report

From the scatterplot we see that there is a strong negative, linear association between the price of a second-hand car and its age, $r = -0.964$. There are no obvious outliers.

The equation of the least squares line is: price = 35 100 − 3940 × age.

The slope of the least squares line predicts that, on average, the price of these second-hand cars decreased by $3940 each year.

The intercept predicts that, on average, the price of these cars when new was $35 100.

The coefficient of determination indicates that 93% of the variation in the price of these second-hand cars is explained by the variation in their age.

The lack of a clear pattern in the residual plot confirms the assumption of a linear association between the price and the age of these second-hand cars.

A full regression analysis requires all of the skills and knowledge that you have developed so far. All of the individual concepts are put together to perform a complete analysis of the relationship between two numerical variables, leading to the creation of a final statistical report.

For a summary of the analysis, refer to 'Elements of a regression analysis' from the start of section 3B.

Example 6 Performing an analysis of association and writing a report

To test the effect of driving instruction on driving skill, 10 randomly selected learner drivers were given a score on a driving skills test. The number of hours of instruction for each learner was also recorded. The results are displayed in the table below.

Hours (h)	19	2	5	9	16	4	19	26	14	8
Score (s)	32	12	17	19	23	16	28	36	30	23

The driving instructors would like to predict the *score* of a learner from the number of *hours* of instruction that learner received.

a Perform a regression analysis of this data.

b Compile a statistical report.

ISBN 978-1-009-25917-0

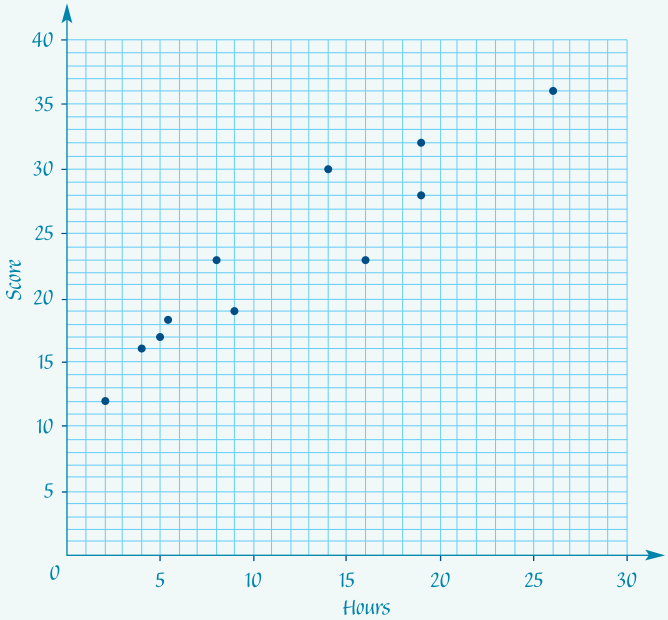
Cambridge University Press
Updated January 2024

Solution

a **1** Decide on the Explanatory variable and Response variable.

In this analysis, the EV is 'hours', so the RV is 'score'.

2 Construct a scatterplot and describe the nature of the association.



The association appears to be linear, strong and positive. There are no apparent outliers.

3 Determine the correlation coefficient and interpret.

$r = 0.9375$

The correlation coefficient confirms the observation of a strong and positive association.

4 Determine the least squares line and write in terms of the variables.

$s = 0.9297\,h + 12.2582$

5 Interpret the slope and intercept.

The slope is 0.9297. On average, the score increases by 0.9297 points for every 1 hour of instruction.

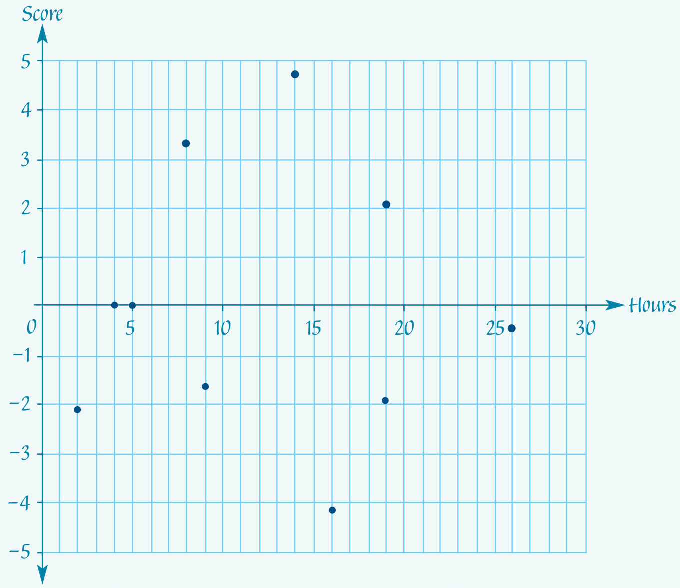
The intercept is 12.2582 and this suggests that, on average, the test score without any instruction at all is 12.2582

6 Determine the coefficient of determination and interpret.

$r^2 = 0.8789$

This indicates that, on average, 87.89% of the variation in score is explained by the variation in number of hours of instruction.

7 Construct a residual plot and interpret.



The residual plot shows no pattern and so the assumption of linearity is confirmed.

b Compile the statistical report.

From the scatterplot we see that there is a strong, positive and linear relationship between the test score and the number of hours of driving instruction. The correlation coefficient has value 0.8789 and there are no apparent outliers.

The equation of the least squares line is: $s = 0.9297\,h + 12.2582$

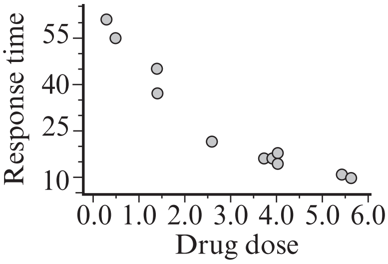
On average, the slope of this line predicts that the score increases by 0.9297 points for every hour increase in instruction time. The intercept predicts that a student who received no instruction would achieve a score of 12.2582 (or 12 rounded to the nearest whole number) on the test.

The coefficient of determination is 0.8789 and this predicts that, on average, 87.89% of the variation in test score is explained by the variation in hours of instruction.

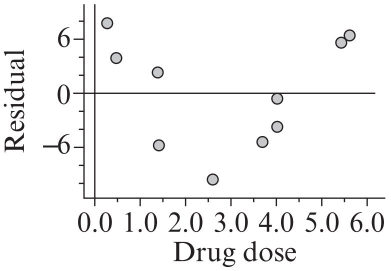
The residual plot shows no pattern and so the assumption of linearity is confirmed.

Exercise 3E

1 In a study of the effectiveness of a pain relief drug, the response time (in minutes) was measured for different drug doses (in mg). A least squares regression analysis was conducted to enable response time to be predicted from drug dose. The results of the analysis are displayed.



Regression equation: $y = ax + b$
$a = -9.306\,12$
$b = 55.8947$
$r^2 = 0.901\,028$
$r = -0.949\,225$



Use this information to complete the following report. Call the two variables *drug dose* and *response time*. In this analysis *drug dose* is the explanatory variable.

Report

From the scatterplot we see that there is a strong ☐ relationship between response time and ☐ : r = ☐. There are no obvious outliers.

The equation of the least squares line is:

response time = ☐ × drug dose + ☐

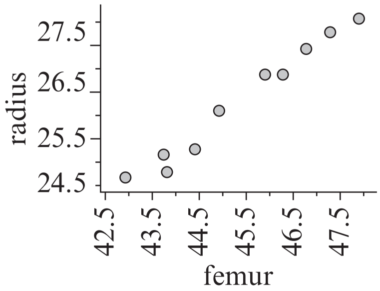
The slope of the least squares line predicts that, on average, response time increases/decreases by ☐ minutes for a 1-milligram increase in drug dose.

The y-intercept of the least squares line predicts that, on average, the response time when no drug is administered is ☐ minutes.

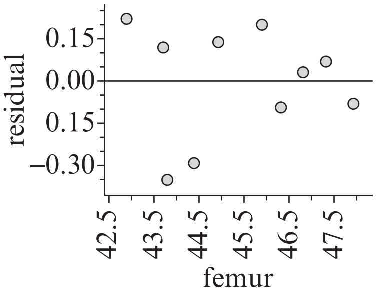
The coefficient of determination indicates that, on average, ☐ % of the variation in ☐ is explained by the variation in ☐.

The residual plot shows a ☐, calling into question the use of a linear equation to describe the relationship between response time and drug dose.

2 A regression analysis was conducted to investigate the nature of the relationship between femur (thigh bone) length and radius (the short thicker bone in the forearm) length in 18-year-old males. The bone lengths are measured in centimetres. The results of this analysis are reported below. In this investigation, femur length was treated as the independent variable.



Regression equation $y = ax + b$
$a = 0.739\,556$
$b = -7.249\,46$
$r^2 = 0.975\,291$
$r = 0.987\,568$



ISBN 978-1-009-25917-0

Use the format of the report given in the previous question to summarise findings of this investigation. Call the two variables *femur length* and *radius length*.

3 The table gives details of Australian Test Cricket Captains, based on all tests from 1930 up to and including the Test against Sri Lanka in July 2022.

Name	*Years*	*Played*	*Won*	*Name*	*Years*	*Played*	*Won*
W M Woodfull	1930–34	25	14	B N Jarman	1968	1	0
V Y Richardson	1935–36	5	4	I M Chappell	1970–75	30	15
D G Bradman	1936–48	24	15	G S Chappell	1975–83	48	21
W A Brown	1945–46	1	1	G N Yallop	1978–79	7	1
A L Hassett	1949–53	24	14	K J Hughes	1978–85	28	4
A R Morris	1951–55	2	0	A R Border	1984–94	93	32
I W Johnson	1954–57	17	7	M A Taylor	1994–1999	50	26
R R Lindwall	1956–57	1	0	S R Waugh	1999–2004	57	41
I D Craig	1957–58	5	3	R Ponting	2004–2010	77	48
R Benaud	1958–64	28	12	M Clarke	2010–2015	47	24
R N Harvey	1961	1	1	S Smith	2015–2018	35	19
R B Simpson	1963–78	39	12	T Paine	2018–2021	23	11
B C Booth	1965–66	2	0	P Cummins	2021–	8	5
W M Lawry	1967–71	25	9				

a Construct a scatterplot for matches won (*won*) against the matches played (*played*) by each captain. *Played* is the Explanatory variable. Estimate the value of correlation coefficient r.

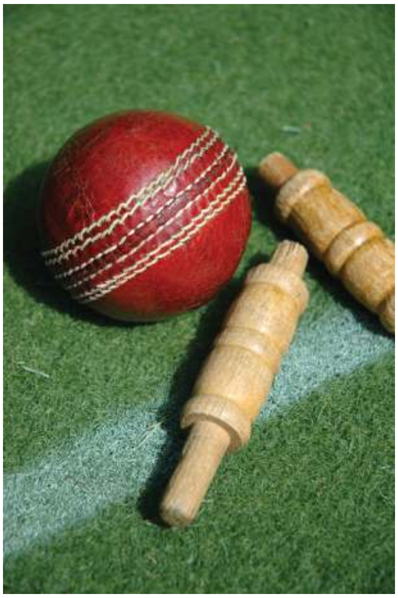
b Describe the association between matches played and matches won in terms of strength, direction, form and outliers if any.

c Fit a least squares line to the data and write its equation in terms of the variables *played* and *won*.

d Write the slope of the least squares line and interpret it in terms of the variables *played* and *won.*

e Determine the correlation coefficient, r, and compare to your earlier estimate.

f Write a statistical report on your analysis of the data.



ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Key ideas and chapter summary

Assign-ment

Linear regression	A straight line can be used to model a linear association between two numerical variables. The relationship can then be described by a rule of the form $y = ax + b$ In this equation: ■ y is the **response variable** ■ x is the **explanatory variable** ■ a is the **slope of the line** ■ b is the **y-intercept**
Least squares line	The least squares line is a linear regression line where the squares of the residual values of the data points is minimised.
Residuals	The vertical distance from a data point to the straight line is called a **residual**: residual value = data value − predicted value.
Least squares method	The **least squares method** is one way of finding the equation of a regression line. It minimises the sum of the squares of the residuals. It works best when there are no outliers. The equation of the least squares line is given by $y = ax + b$, where a represents the slope of the line and b the y-intercept.
Using the least squares line	The least squares line $y = ax + b$ enables the value of y to be determined for a given value of x. For example, the least squares line $cost = 0.06 \times number\ of\ pages + 1.20$ predicts that the cost of a 100-page book is: $cost = 0.06 \times 100 + 1.20 = \$7.20$
Key assumption of regression	Linear regression assumes that the underlying association between the variables is linear.
Interpolation and extrapolation	Predicting *within* the range of data is called **interpolation**. Predicting *outside* the range of data is called **extrapolation**.

Slope and intercept	The slope and intercept in the least squares model $y = ax + b$ are a and b respectively. The slope indicates the average increase/decrease in the **response variable** per unit increase in the **explanatory variable**, and the intercept predicts the **response variable** when the **explanatory variable** is zero. The **slope** of the least squares line above predicts that the cost of a textbook increases by 6 cents ($0.06) for each additional page. The **intercept** of the line predicts that a book with no pages costs $1.20 (this might be the cost of the cover).
Coefficient of determination	The **coefficient of determination** (r^2) gives a measure of the predictive power of a least squares line. For example, for the least squares line above, the coefficient of determination is 0.81. From this we conclude that 81% of the variation in the cost of a textbook can be explained by the variation in the number of pages.
Residual plots	**Residual plots** can be used to test the linearity assumption by plotting the residuals against the Explanatory variable. A residual plot that appears to be a random collection of points clustered around zero supports the linearity assumption. A residual plot that shows a clear pattern indicates that the association is not linear.
Statistical reports	A statistical report summarises a statistical analysis describing the association, or lack of it, between the variables and stating the value of the correlation coefficient and the equation of the least squares line. Its slope and intercept should be used to make predictions, if any, and the residual plot should be commented upon.

Skills check

Having completed this chapter you should be able to:

- for raw data, determine the equation of the least squares line using a CAS calculator
- interpret the slope and intercept of a least squares line
- interpret the coefficient of determination as part of a regression analysis
- use the least squares line for prediction
- evaluate the reliability of the predictions made using a least squares line
- calculate residuals
- construct a residual plot using a CAS calculator
- use a residual plot to determine the appropriateness of using the equation of the least squares line to model the association
- present the results of a regression analysis in report form.

ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

Short-answer questions

1 State the value of the slope and y-intercept for the least squares line $y = 0.52x - 1.2$

2 If the equation of a least squares line is $y = 8 - 9x$ and $r^2 = 0.25$, determine the value of the correlation coefficient, r.

3 For the least squares line $y = 8 - 9x$ predict the value of y when $x = 5$.

4 Write the equation for the least squares line in the form $y = ax + b$ for the data set shown.

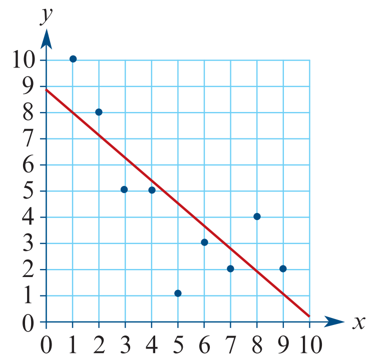
x	25	15	10	5
y	10	10	15	25

5 Write the equation for the least squares line for the data set shown.

y	30	25	15	10
x	40	20	30	10

6 Using a least squares line, the predicted value of a data value is 78.6. The residual value is –5.4. What is the actual data value?

7 The equation of the least squares line plotted on the scatterplot opposite is closest to:



A $y = 9 - 0.9x$

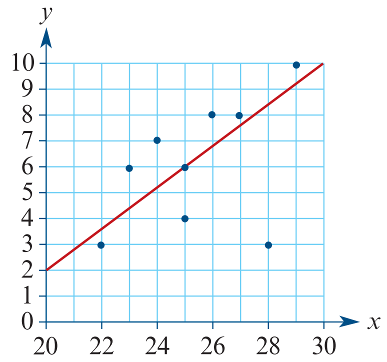
B $y = 0.9 - 9x$

C $y = 9 + 0.9x$

D $y = 0.9 + 9x$

E $y = 9 - x$

8 The equation of the least squares line plotted on the scatterplot opposite is closest to:



A $y = 0.8x - 14$

B $y = 0.3x + 2$

C $y = 1.25x - 23$

D $y = 0.8x + 2$

E $y = 5x + 2$

9 Weight (w, in kg) can be predicted from height (h, in cm) given the least squares line:

$$w = 0.95h - 96, \text{ with } r = 0.79$$

a What is the explanatory variable?

b Interpret the value of the slope of the least squares line in the context of the variables given.

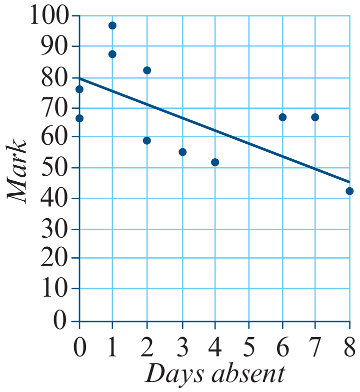
ISBN 978-1-009-25917-0

c Given the correlation coefficient is $r = 0.79$, determine the coefficient of determination and interpret.

d A person of height 179 cm weighs 82 kg. If the equation of the least squares line is used to predict their weight, calculate the residual to the nearest whole number.

10 The coefficient of determination for the data displayed in the scatterplot opposite is close to $r^2 = 0.5$.

State the value of the correlation coefficient, to one decimal place.



11 The weekly *income* and weekly *expenditure* on food for a group of 10 university students given in the following table

Income (n, \$/week)	150	250	300	600	300	380	950	450	850	1000
Expenditure (e, \$/week)	40	60	70	120	130	150	200	260	460	600

Use your CAS calculator to determine the least squares line that enables weekly expenditure on food to be predicted from weekly income.

12 The equation of a least squares line that enables weekly *amount* (y) spent on entertainment (x, in dollars) to be predicted from weekly *income* is given by:

$$y = 0.10x + 40$$

a Using this equation predict the amount spent an entertainment by an individual with a weekly income of \$600.

b On average, what is the weekly increase in spending on entertainment per extra dollar of weekly income?

Extended-response questions

1 The equation of the least squares line:

$$t = 7.3\,m - 22$$

can be used to predict the taste score (t) of a country town's drinking water from its *magnesium content* (m, in mg/litre).

a Which variable is the explanatory variable?

b Write and interpret the slope of the least squares line.

c Use the least squares line to predict the taste score of a country town's drinking water whose magnesium content is 16 milligrams/litre, correct to one decimal place.

ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

CAS

2 The *time* taken to complete a task and the number of *errors* on the task were recorded for a sample of 10 primary school children.

Time (s)	22.6	21.7	21.7	21.3	19.3	17.6	17.0	14.6	14.0	8.8
Errors	2	3	3	4	5	5	7	7	9	9

a Determine the equation of the least squares line that fits this data with *errors* as the response variable.

b Determine the value of the correlation coefficient to two decimal places.

3 In an investigation of the relationship between the hours of sunshine (s, per year) and days of rain (r, per year) for 25 cities, the least squares line was found to be:

$$s = 2850 - 6.88\,r, \text{ with } r^2 = 0.484$$

Use this information to complete the following sentences.

a In the equation of the least squares line, the explanatory variable is ☐.

b The slope is ☐ and the intercept is ☐.

c The equation of the least squares line predicts that a city that has 120 days of rain per year will have ☐ hours of sunshine per year.

d The slope of the least squares line predicts that the hours of sunshine per year will ☐ by ☐ hours for each additional day of rain.

e The correlation coefficient r = ☐, correct to three significant figures.

f ☐% of the variation in sunshine hours can be explained by the variation in ☐.

g One of the cities used to determine the equation of the least squares line had 142 days of rain and 1390 hours of sunshine.

  i The equation of the least squares line predicts that it has ☐ hours of sunshine.

  ii The residual value for this city is ☐ hours.

h Using a least squares line to make predictions within the range of data used to determine the equation of the least squares line is called ☐.

CAS

4 The cost of preparing meals in a school canteen is linearly related to the number of meals prepared. To help the caterers predict the costs, data were collected on the cost of preparing meals for different levels of demands. The data are shown below.

Number of meals (x)	30	35	40	45	50	55	60	65	70	75	80
Cost (y, dollars)	138	154	159	182	198	198	214	208	238	234	244

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

a Which is the response variable?

b Use your calculator to show that the equation of the least squares line that relates the cost of preparing meals to the number of meals produced is:

$$y = 2.10x + 81.5$$

c Use the equation to predict the cost of producing:

  i 48 meals. In making this prediction are you interpolating or extrapolating?

  ii 21 meals. In making this prediction are you interpolating or extrapolating?

d **i** Write and interpret the intercept of the least squares line.

  ii Write and interpret the slope of the least squares line.

e If $r = 0.978$, write the coefficient of determination and interpret.

5 The data below show the height (in cm) of a group of 10 children aged 2 to 11 years.

Height (cm)	86.5	95.5	103.0	109.8	116.4	122.4	128.2	133.8	139.6	145.0
Age (years)	2	3	4	5	6	7	8	9	10	11

The task is to determine the equation of a least squares line that can be used to predict height from age.

a In this analysis, which would be the Response variable and which would be the Explanatory variable?

b Use your calculator to confirm that the equation of the least squares line is: $height = 76.64 + 6.366 \times age$ and $r = 0.9973$.

c Use the least squares line to predict the height of a 1-year-old child. Give the answer correct to the nearest cm. In making this prediction are you extrapolating or interpolating?

d What is the slope of the least squares line and what does it tell you in terms of the variables involved?

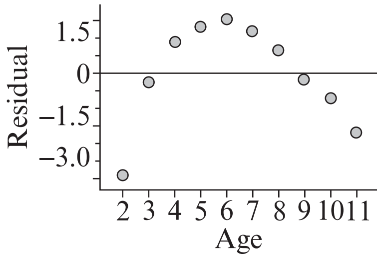
e Calculate the value of the coefficient of determination and interpret in terms of the relationship between age and height.

f Use the least squares equation to:

  i predict the height of the 10-year-old child in this sample

  ii determine the residual value for this child.

g Confirm that the residual plot for this analysis is shown below.



ISBN 978-1-009-25917-0

6 Write a statistical report on the analysis of the data in your answers to the previous questions. Explain why the residual plot suggests that a linear equation is not the most appropriate model for this relationship.

7 A marketing firm wanted to investigate the relationship between the number of times a song was played on the radio (*played*) and the number of downloads sold the following week (*weekly sales*).

CAS

The following data was collected for a random sample of ten songs.

Played	47	34	40	34	33	50	28	53	25	46
Weekly sales	3950	2500	3700	2800	2900	3750	2300	4400	2200	3400

a Which is the explanatory variable and which is the response variable?

b Construct a scatterplot of this data.

c Determine the value of the correlation coefficient, r, for this data.

d Describe the relationship between *weekly sales* and *played* in terms of direction, strength and form and outliers (if any).

e Determine the equation for the least squares line and write it down in terms of the variables *weekly sales* and *played*.

f Interpret the slope and intercept of the least squares line in the context of the problem.

g Use your equation to predict the number of downloads of a song when it was played on the radio 100 times in the previous week.

h In making this prediction, are you interpolating or extrapolating?

8 To test the effect of driving instruction on driving skill, 10 randomly selected learner drivers were given a *score* on a driving skills test. The number of *hours* of instruction for each learner was also recorded. The results are displayed in the table below.

Hours	19	2	5	9	16	4	19	26	14	8
Score	32	12	17	19	23	16	28	36	30	23

a Which is the explanatory variable and which is the response variable?

b Construct a scatterplot of these data.

c Determine the correlation coefficient, r, give answer correct to four decimal places.

d Describe the association between *score* and *hours* in terms of direction, strength and form and outliers (if any).

e Determine the equation for the least squares line and write it down in terms of the variables *score* and *hours*. Give coefficients correct to three significant figures.

f Interpret the slope and the intercept (if appropriate) of the regression line.

g Predict the score after 10 hours of instruction to the nearest point.

9 To investigate the association between marks on an assignment and the final examination mark, the following data was collected.

CAS

Assignment mark	80	77	71	78	65	80	68	64	50	66
Exam mark	83	83	79	75	68	84	71	69	66	58

a A scatterplot shows that there is a strong positive linear relationship between the *assignment mark* and the final *exam marks*. The correlation coefficient is $r = 0.76$. Given this information, a student wrote: 'Good final exam marks are the result of good assignment marks'. Comment on this statement.

b Determine the equation of the least square line and write it down in terms of the variables final *exam mark* and *assignment mark*. Write the coefficients correct to two significant figures.

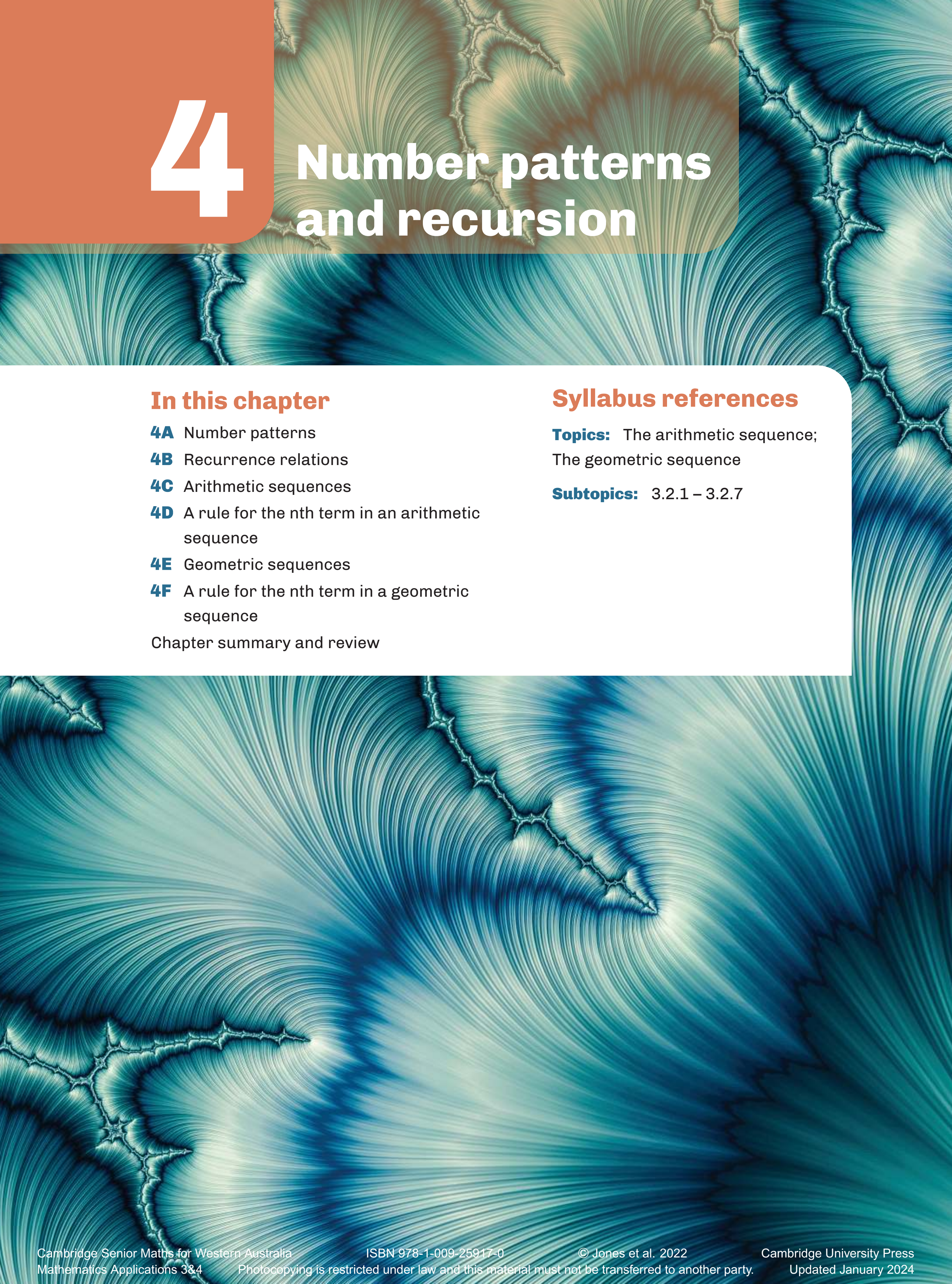
c Interpret the intercept and slope of the least squares line in terms of the variables in the study.

d Use your least squares line equation to predict the final exam mark for a student who scored 50 on the assignment. Give your answer correct to the nearest mark.

e How reliable is the prediction made in part **d**?

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024



4

Number patterns and recursion

In this chapter

Syllabus references

Topics: The arithmetic sequence; The geometric sequence

Subtopics: 3.2.1 – 3.2.7

4A Number patterns

A **sequence** is a list of numbers in a particular order. The numbers or items in a sequence are called the terms of the sequence.

We write the terms of a sequence as a list, separated by commas. If a sequence continues indefinitely, or if there are too many terms in the sequence to write them all, we use an *ellipsis*, '. . . ', at the end of a few terms of the sequence like this:

12, 22, 5, 6, 16, 43, . . .

Randomly generated sequences

Recording the numbers obtained while tossing a die would give a randomly generated sequence, such as:

3, 1, 2, 2, 6, 4, 3, . . .

Because there is no pattern in the sequence there is no way of predicting the next term.

Rule based sequences

Some sequences of numbers do display a pattern. For example, this sequence

1, 3, 5, 7, 9, . . .

has a definite pattern and so this sequence is said to be *rule-based*.

The sequence of numbers has a starting value. We add 2 to this number to generate the term 3. Then, add 2 again to generate the term 5, and so on.

The rule is 'add 2 to each term'.

$$\begin{tikzcd}1 \arrow[r, bend left, "+2"] & 3 \arrow[r, bend left, "+2"] & 5 \arrow[r, bend left, "+2"] & 7 \arrow[r, bend left, "+2"] & 9 & \dots\end{tikzcd}$$

In this chapter we will look at sequences that can be generated by a rule.

Example 1 Looking for a rule for a sequence of numbers

Look for a pattern or rule in each sequence and find the next number.

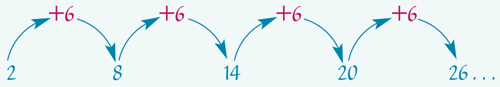
a 2, 8, 14, 20, . . .
b 5, 15, 45, 135, . . .
c 7, 4, 1, −2, . . .
d 1, 4, 9, 16, . . .
e 1, 1, 2, 3, 5, . . .

Solution

a 2, 8, 14, 20, . . .

1 Add 6 to make each new term.

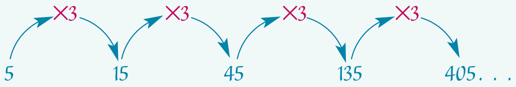
2 Add 6 to 20 to make the next term, 26.



b 5, 15, 45, 135, . . .

1 Multiply by 3 to make each new term.

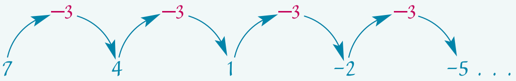
2 Multiply 135 by 3 to make the next term, 405.



c 7, 4, 1, −2, . . .

1 Subtract 3 each time to make the next term.

2 Subtract 3 from −2 to get the next term, −5.



d 1, 4, 9, 16, . . .

1 Each term is the square of its position in the sequence.

2 The fifth term is $5^2 = 25$.

1st,	2nd,	3rd,	4th,	5th,	
1^2,	2^2,	3^2,	4^2,	5^2,	
1,	4,	9,	16,	25,	. . .

e 1, 1, 2, 3, 5, . . .

1 Add the previous two terms to make each new term.

2 The next term is 3 + 5 = 8.

Note: This is called a **Fibonacci sequence**.

		$1 + 1$	$1 + 2$	$2 + 3$	$3 + 5$	
1,	1,	2,	3,	5,	8,	. . .

Example 2 Generating a sequence of terms

Write the first five terms of the sequence with a starting value of 5 and the rule ‘double the number and then subtract 3’.

Solution

1 Write the starting value.

5

2 Apply the rule (double, then subtract 3) to generate the next term.

$2 \times 5 - 3 = 7$

3 Calculate three more terms.

$2 \times 7 - 3 = 11$
$2 \times 11 - 3 = 19$
$2 \times 19 - 3 = 35$

4 Write your answer.

The sequence is 5, 7, 11, 19, 35, . . .

ISBN 978-1-009-25917-0

All of the calculations to generate sequences from a rule are repetitive. The same calculations are performed over and over again – this is called **recursion**.

Using a calculator to generate a sequence of numbers from a rule

A calculator can perform recursive calculations very easily, because it automatically stores the answer to the last calculation it performed, as well as the method of calculation.

Example 3 Generating a sequence of numbers with a calculator

Use a calculator to generate the first five terms of the sequence with a starting value of 5 and the rule 'double and then subtract 3'.

Solution

Steps

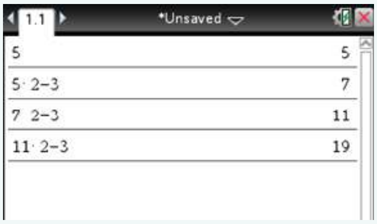
1 Start with a blank computation screen.

2 Type 5 and press enter or EXE.

3 Next type **× 2 –3** and press enter or EXE to generate the next term in the sequence, '7' is generated and displayed on the right-hand side of the screen. Note that the computation generating this value is shown of the left side of the screen as '5.2–3' on the TI-Nspire and 'ans × 2 –3' on the ClassPad (here 'ans' represents the answer to the previous calculation).

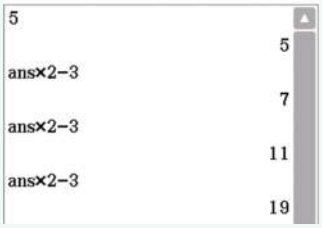
4 Pressing enter or EXE repeatedly applies the rule '× 2 –3' to the last calculated value, in the process generating successive terms of the sequence as shown.

TI-Nspire



```
5          5
5·2−3      7
7·2−3     11
11·2−3    19
```

ClassPad



```
5
                5
ans×2−3
                7
ans×2−3
               11
ans×2−3
               19
```

ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

Naming the terms in a sequence

The symbols T_1, T_2, T_3, are used as labels or names for the first, second and third terms in the sequence. In the labels T_1, T_2, T_3 the numbers 1, 2, 3 are called *subscripts*[1].

The subscripts tell us the position of each term in the sequence. So T_{10} is just a name for the tenth term in the sequence.

Example 4 Naming terms in a sequence

For the sequence: 2, 8, 14, 20, 26, 32, . . . state the values of:

a T_1 **b** T_4 **c** T_6

Solution

1 Write the name for each term under its value in the sequence.

2,	8,	14,	20,	26,	32,	. . .
T_1	T_2	T_3	T_4	T_5	T_6	

2 Read the value of each required term.

a $T_1 = 2$ **b** $T_4 = 20$ **c** $T_6 = 32$

Exercise 4A

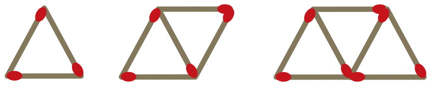
Looking for and applying patterns

Example 1

1 Find the next term in each sequence.

a 3, 7, 11, 15, . . .
b 10, 9, 8, 7, . . .
c 4, 8, 16, 32, . . .
d January, February, . . .
e 1, 2, 1, 2, . . .
f 48, 24, 12, 6, . . .
g 31, 24, 17, 10, . . .
h 1, 8, 27, 64, . . .
i a, c, e, g, . . .
j ♣, ♦, ♥, ♠, . . .
k 1, 3, 9, 27, . . .
l ⇒, ⇐, ⇒, ⇐, . . .
m 2, 6, 18, 54, . . .
n ↑, →, ↓, ←, . . .
o M, T, W, T, . . .
p O, T, T, F, . . .
q A, E, F, H, . . .

2 **a** Draw the next shape in this sequence of matchstick shapes.



Shape 1 Shape 2 Shape 3

b Write the sequence for the number of matchsticks used in each shape shown.

c Find the number of matchsticks needed to make shape 4 and shape 5.

d Give the first number and the rule for making each subsequent number in the sequence.

---

[1] Do not confuse the subscripts, which give the position of each term, with the powers we see in say, x^2, which means x multiplied by x. Also notice that T_3 is just the name for the term in the third position. It is not saying that the value of the term is 3. In the Example 3, $T_3 = 11$.

ISBN 978-1-009-25917-0

3 **a** Draw the next pattern in this sequence of dots.

b Complete the table.

Pattern number	1st	2nd	3rd	4th	5th
Number of dots	1	4			

c Give the first number of dots and the rule for making each subsequent number of dots in the sequence.

4 Describe how terms are generated in each number sequence and give the next two terms.

a 5, 8, 11, 14, . . .
b 19, 28, 37, 46, . . .
c 38, 34, 30, 26, . . .
d 66, 58, 50, 42, . . .
e 3, 6, 12, 24, . . .
f 4, 12, 36, 108, . . .
g 128, 64, 32, 16, . . .
h 3, –6, 12, –24, . . .
i 1, 2, 3, 5, . . .

Generating a sequence recursively

Example 2

5 Use the following starting values and rules to generate the first five terms of the following sequences recursively by hand.

a Starting value: 2
rule: add 6

b Starting value: 5
rule: subtract 3

c Starting value: 1
rule: multiply by 4

d Starting value: 10
rule: divide by 2

e Starting value: 6
rule: multiply by 2 add 2

f Starting value: 12
rule: multiply by 0.5 add 3

Example 3

6 Use the following starting values and rules to generate the first five terms of the following sequences recursively using a calculator.

a Starting value: 4
rule: add 2

b Starting value: 24
rule: subtract 4

c Starting value: 2
rule: multiply by 3

d Starting value: 50
rule: divide by 5

e Starting value: 5
rule: multiply by 2 add 3

f Starting value: 18
rule: multiply by 0.8 add 2

Naming terms in a sequence

Example 2

7 Find the required terms from the sequence: 6, 11, 16, 21, . . .

a T_1 **b** T_3 **c** T_2 **d** T_4 **e** T_5 **f** T_6

8 For each sequence state the value of the named terms:

i T_1 **ii** T_4 **iii** T_5

a 6, 10, 14, 18, . . .
b 2, 8, 32, 128, . . .
c 29, 22, 15, 8, . . .
d 96, 48, 24, 12, . . .

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

4B Recurrence relations

A **recurrence relation** is a mathematical rule that uses existing terms of a sequence to generate the terms that follow.

A *first-order recurrence relation* uses only one existing term of a sequence to generate the next. A second-order recurrence relation uses the previous two terms of a sequence to generate the next. Higher-order recurrence relations use more than the two previous terms of a sequence to generate the next.

Naming the terms in a recurrence relation

We have already seen how we refer to individual terms of a sequence using subscript notation. If T_6 is the sixth term in the sequence, then T_5 is the term before the sixth term and T_7 will be the term after the sixth term.

Part of the sequence that these terms belong to is shown below.

$$\ldots, T_5, T_6, T_7, \ldots$$

The subscript of T_5 (5) is one *less* than the subscript of T_6 (6).
The subscript of T_7 (7) is one *more* than the subscript of T_6 (6).

Example 5 Naming the terms in a recurrence relation

Write the symbol used to represent the term:

a after T_9

b before T_3

c three terms after T_4

d two terms before T_5

e the term before T_1

Solution

a The term after T_9 has a subscript that is one more than 9, or 10.

The term after T_9 is T_{10}.

b The term before T_3 has a subscript that is one less than 3, or 2.

The term before T_3 is T_2.

c The term that is three terms after T_4 has a subscript that is three more than 4, or 7.

The term three terms after T_4 is T_7.

d The term that is two terms before T_5 has a subscript that is two less than 5, or 3.

The two term before T_5 is T_3.

e The term that is before T_1 has a subscript that is one less than 1, or 0.

Note: You will learn about the significance of sequence terms that have subscript zero in the next chapter.

The term before T_1 is T_0.

ISBN 978-1-009-25917-0 © Jones et al. 2022 Cambridge University Press

Writing a recurrence relation

If we let T_n be any particular term in the sequence, we can call it the '*n*th term'.
The term before the *n*th term will be T_{n-1}.
The term after the *n*th term will be T_{n+1}.

The term before T_{n-1} will be T_{n-2}.
The term after T_{n+1} will be T_{n+2}.

A sequence can be written in general terms using these symbols like this:

$$\ldots, T_{n-3}, T_{n-2}, T_{n-1}, T_n, T_{n+1}, T_{n+2}, T_{n+3}, \ldots$$

Note: The value of n gives the position of the term in the sequence and has only integer values.

Consider the sequence with the starting term 5 and the rule 'multiply by three'.

$$5, 15, 45, \ldots$$

In symbols, we would write the starting term as T_1, the first term of the sequence.

$$T_1 = 5$$

The *n*th term of the sequence, T_n, is multiplied by three to give the term after it, T_{n+1}.
In symbols, this becomes:

$$T_{n+1} = 3 \times T_n \quad \text{or} \quad T_{n+1} = 3T_n$$

A recurrence relation consists of *both* the starting term and the rule, separated by a comma.

$$T_1 = 5, \quad T_{n+1} = 3T_n$$

Example 6 Generating a sequence from a recurrence relation

Write the first five terms of the sequence defined by the recurrence relation:

$$T_1 = 9, \quad T_{n+1} = T_n - 4$$

Solution

1 Write the starting value.

$T_1 = 9$

2 Use the rule to find the next term, T_2.

$$\begin{aligned} T_2 &= T_1 - 4 \\ &= 9 - 4 \\ &= 5 \end{aligned}$$

3 Use the rule to determine three more terms.

$$\begin{aligned} T_3 &= T_2 - 4 \\ &= 5 - 4 \\ &= 1 \end{aligned} \qquad \begin{aligned} T_4 &= T_3 - 4 \\ &= 1 - 4 \\ &= -3 \end{aligned} \qquad \begin{aligned} T_5 &= T_4 - 4 \\ &= -3 - 4 \\ &= -7 \end{aligned}$$

4 Write your answer.

The sequence is $9, 5, 1, -3, -7, \ldots$

Example 7 Writing a recurrence relation

Write a recurrence relation that generates the following sequences.

a 6, 13, 20, . . .

b 400, 100, 25, . . .

Solution

a 6, 13, 20, . . .

1 Write the first term.

$T_1 = 6$

2 Decide on the rule for the sequence. This sequence is generated by adding 7 to each term to generate the next.

The rule for this sequence is 'add 7'.

3 Write the rule in recurrence relation form.

The next term, T_{n+1}, is the current term, T_n, +7.

$T_{n+1} = T_n + 7$

4 Write the recurrence relation.

$T_1 = 6, \quad T_{n+1} = T_n + 7$

b 400, 100, 25, . . .

1 Write the first term.

$T_1 = 400$

2 Decide on the rule for the sequence. This sequence is generated by dividing each term by 4 to generate the next.

The rule for this sequence is 'divide by 4'.

3 Write the rule in recurrence relation form.

The next term, T_{n+1}, is the current term, T_n, ÷4.

$T_{n+1} = T_n \div 4$

4 Write the recurrence relation.

$T_1 = 400, \quad T_{n+1} = T_n \div 4$

Note: This recurrence relation can also be written as $T_1 = 400, T_{n+1} = \frac{1}{4}T_n$ since dividing by 4 is the same as multiplying by one quarter.

Recurrence relations using a calculator

You have seen how to generate the terms of a sequence using simple recursion on the main screen of your calculator. This technique can be very useful to quickly generate the first few terms of a sequence.

Example 8 Using a calculator to generate sequences from recurrence relations

A sequence is generated by the recurrence relation $T_1 = 300$, $T_{n+1} = 0.5T_n - 9$.

Use your calculator to generate this sequence and determine how many terms of the sequence are positive.

Solution

1 Start with a blank computation screen.

2 Type **300** and press enter (or EXE).

3 Next type × **0.5 –9** and press enter (or EXE) to generate the next term in the sequence, '141' is generated and displayed on the right-hand side of the screen.

4 Continue to press enter (or EXE) until the first negative terms appears.

5 Write your answer.

```
300                 300.
300 · 0.5 – 9       141.
141 · 0.5 – 9       61.5
61.5 · 0.5 – 9      21.75
21.75 · 0.5 – 9     1.875
1.875 · 0.5 – 9     –8.625
|
```

The first five terms of the sequence are positive.

The Casio ClassPad CAS calculator has a dedicated sequence function that is useful if many terms in a sequence need to be generated.

How to generate sequences from recurrence relations using the ClassPad

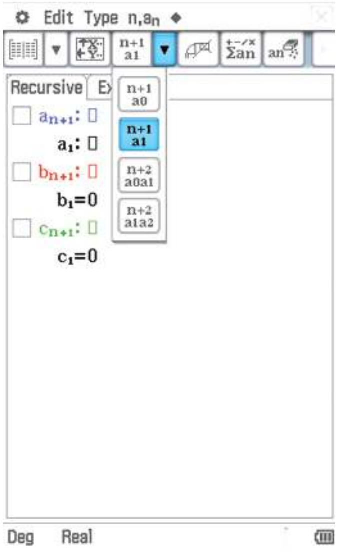
A sequence is generated by the recurrence relation $T_1 = 300$, $T_{n+1} = 0.5T_n - 9$.
Use your calculator to generate this sequence and determine how many terms of the sequence are positive.

Steps

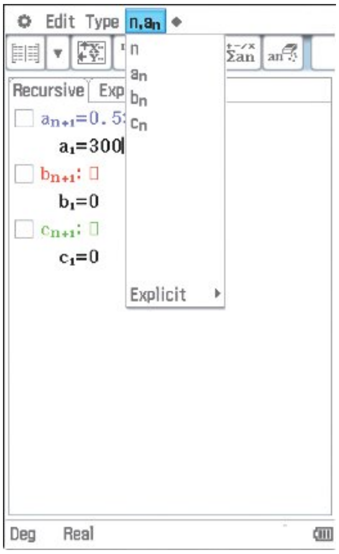
1 Tap menu and choose 'Sequence'.

2 Make sure the recursive tab is selected.
Select $\begin{smallmatrix}n+1\\a1\end{smallmatrix}$ from the menu as shown.

3 Enter the rule for the recurrence relation, substituting '*a*' for '*T*' and selecting a_n from the menu as shown.
Change the value of a_1 to 300.



a_{n+1}: ☐
a_1: ☐
b_{n+1}: ☐
$b_1=0$
c_{n+1}: ☐
$c_1=0$



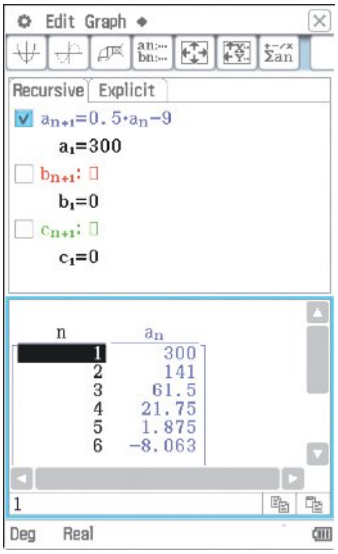
$a_{n+1}=0.5$
$a_1=300$
b_{n+1}: ☐
$b_1=0$
c_{n+1}: ☐
$c_1=0$

ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

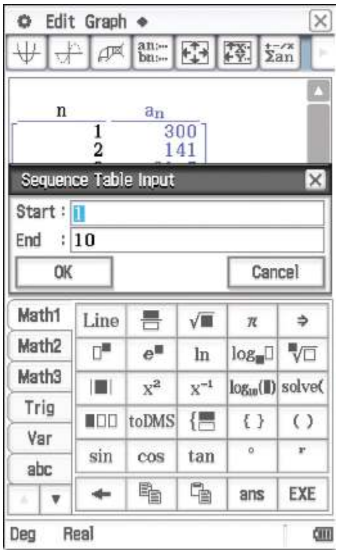
4 Select the recurrence relation by tapping in the box next to it to insert a ☑ and then tap ▦. The table at the bottom of the screen will now contain the term number, n, and the term values, a_n.

5 Tap ▦ to change the starting and ending term numbers.



☑ $a_{n+1}=0.5 \cdot a_n-9$
$a_1=300$
☐ b_{n+1}:
$b_1=0$
☐ c_{n+1}:
$c_1=0$

n	a_n
1	300
2	141
3	61.5
4	21.75
5	1.875
6	−8.063



Sequence Table Input
Start : 1
End : 10

Start = 1 and End = 10 will show the first 10 terms of the sequence in the table. Scroll the table to see all values.

6 Count the number of positive terms in the sequence and write your answer.

There are 5 positive terms in this sequence.

Exercise 4B

Naming the terms in a recurrence relation

Example 5

1 Write the symbol used to represent the term:

a before T_7
b after T_{12}
c two terms after T_9
d 5 terms after T_4

2 Some of the terms of a sequence are . . . , 7, 9, 11, 13, 15, 17, 19, . . .

a If $T_{n-1} = 9$, write the value of T_n
b If $T_{n-2} = 9$, write the value of T_n
c If $T_n = 15$, write the value of T_{n+2}
d If $T_n = 7$, write the value of T_{n+5}

Generating sequences using recurrence relations

Example 6

3 Without using your calculator, write the first five terms of the sequences generated by each of the recurrence relations below.

a $T_1 = 2, \quad T_{n+1} = T_n + 3$
b $T_1 = 50, \quad T_{n+1} = T_n - 5$
c $T_1 = 1, \quad T_{n+1} = 3T_n$
d $T_1 = 3, \quad T_{n+1} = -2T_n$
e $T_1 = 5, \quad T_{n+1} = 2T_n - 1$
f $T_1 = 2, \quad T_{n+1} = 2T_n + 3$
g $T_1 = -2, \quad T_{n+1} = 3T_n + 5$
h $T_1 = -10, \quad T_{n+1} = -3T_n + 5$

Generating sequences using recurrence relations and a calculator

Example 8

4 Using your calculator, write the first five terms of the sequence generated by each of the recurrence relations below.

CAS

a $T_1 = 12, \quad T_{n+1} = 6T_n - 15$

b $T_1 = 20, \quad T_{n+1} = 3T_n + 25$

c $T_1 = 2, \quad T_{n+1} = 4T_n + 3$

d $T_1 = 64, \quad T_{n+1} = 0.25T_n - 1$

e $T_1 = 48\,000, \quad T_{n+1} = T_n - 3000$

f $T_1 = 25\,000, \quad T_{n+1} = 0.9T_n - 550$

Exploring sequences with a calculator

5 How many terms of the sequence formed from the recurrence relation below are positive?

$$T_1 = 150, \qquad T_{n+1} = 0.6T_n - 5$$

6 How many terms of the sequence formed from the recurrence relation below are negative?

$$T_1 = -30, \qquad T_{n+1} = 0.8T_n + 2$$

4C Arithmetic sequences

As we have seen earlier, in some cases the rule that connects the values of the terms in a sequence is something like ‘each new term is made by adding or subtracting a fixed amount to or from the current term’.

Examples include:

Sequence	*Description*	*Recurrence relation*
A. 2, 7, 12, 17, 22, . . .	First term = 2 Rule: ‘add 5’	$T_1 = 2, \quad T_{n+1} = T_n + 5$
B. 200, 300, 400, 500, 600, . . .	First term = 200 Rule: ‘add 100’	$T_1 = 200, \quad T_{n+1} = T_n + 100$
C. 50, 45, 40, 35, 30, . . .	First term = 50 Rule: ‘subtract 5’	$T_1 = 50, \quad T_{n+1} = T_n - 5$
D. 100, 85, 70, 55, 40, . . .	First term = 100 Rule: ‘subtract 15’	$T_1 = 100, \quad T_{n+1} = T_n - 15$

Sequences that are generated by adding or subtracting a fixed amount to the previous term are called **arithmetic sequences**.

ISBN 978-1-009-25917-0

Common difference, *d*

The fixed amount we add or subtract to form an arithmetic sequence recursively is called the **common difference**. The symbol d is often used to represent the common difference.

If the sequence is *known* to be arithmetic, the common difference can be calculated by simply subtracting any pair of successive terms.

Common difference, *d*

In an *arithmetic sequence*, the fixed number added to (or subtracted from) each term to make the next term is called the *common difference, d,* where:

$$\begin{aligned} d &= \text{any term} - \text{its previous term} \\ &= T_2 - T_1 \\ &= T_3 - T_2 \\ &= T_4 - T_3 \end{aligned}$$

and so on.

For example, the common difference for the arithmetic sequence, 30, 25, 20, . . . is:

$$d = T_2 - T_1 = 25 - 30 = -5$$

Often it is not necessary to calculate the common difference in this formal way. It may be easy to see what amount has been repeatedly added (or subtracted) to make each new term.

Sequence	*Common difference calculations*	*Common difference*
A. 2, 7, 12, 17, 22, . . .	$7 - 2 = 5$ $\quad 17 - 12 = 5$ $12 - 7 = 5$ $\quad 22 - 17 = 5$	$d = 5$
B. 200, 300, 400, 500, 600, . . .	$300 - 200 = 100$ $\quad 500 - 400 = 100$ $400 - 300 = 100$ $\quad 600 - 500 = 100$	$d = 100$
C. 50, 45, 40, 35, 30, . . .	$45 - 50 = -5$ $\quad 35 - 40 = -5$ $40 - 45 = -5$ $\quad 30 - 35 = -5$	$d = -5$
D. 100, 85, 70, 55, 40, . . .	$85 - 100 = -15$ $\quad 55 - 70 = -15$ $70 - 85 = -15$ $\quad 40 - 55 = -15$	$d = -15$

In sequences A. and B., the common difference is positive. The terms of these sequences *increase* in size because the number is being added in the recurrence relation.

In sequences C. and D., the common difference is negative. The terms of these sequences *decrease* in size because the number is being subtracted in the recurrence relation.

Recurrence relations for arithmetic sequences

It is common to use the symbol a to represent the first term of an arithmetic sequence.

An arithmetic sequence is shown below.

$$10, 15, 20, \ldots$$

We know the arithmetic sequence information and we can write a recurrence relation that generates this sequence.

Arithmetic sequence information	*Recurrence relation*
first term: $a = 10$ common difference: $d = 5$	$T_1 = 10, \quad T_{n+1} = T_n + 5$

We can match a to 'T_1' and d to '+5' in the recurrence relation to determine the general recurrence relation for an arithmetic sequence.

In general, an arithmetic sequence will be generated by a recurrence relation of the form

$$T_1 = a, \quad T_{n+1} = T_n + d$$

where a is the first term of the sequence and d is the common difference.

Arithmetic sequences

A sequence is arithmetic if the terms increase or decrease by a fixed amount.

In any arithmetic sequence, the common difference, d, is the amount added or subtracted in the recurrence relation to calculate the next term.

$$d = \text{ any term } - \text{ previous term}$$

In general,

$$d = T_n - T_{n-1}$$

Arithmetic sequences are generated by a recurrence relation of the form

$$T_1 = a, \; T_{n+1} = T_n + d$$

where a is the first term and d is the common difference.

If $d > 0$, the terms of the sequence will increase.

If $d < 0$, the terms of the sequence will decrease.

Example 9 Finding the common difference in an arithmetic sequence

Find the common difference for the following arithmetic sequences and use it to find the 4th term in the sequence.

a 2, 5, 8, . . . **b** 25, 23, 21, . . .

Solution

1 Because we know the sequence is arithmetic, all we need to do is find the difference in value between terms 1 and 2.

2 To find the 4th term, add the common difference to the 3rd term.

a $d = T_2 - T_1 = 5 - 2 = 3$

$T_4 = T_3 + d = 8 + 3 = 11$

b $d = T_2 - T_1 = 23 - 25 = -2$

$T_4 = T_3 + d = 21 + (-2) = 19$

Identifying arithmetic sequences

If a sequence is arithmetic, the difference between successive terms will be constant. We can use this idea to see whether or not a sequence is arithmetic.

Example 10 Identifying an arithmetic sequence

Which of the following sequences is arithmetic?

a 21, 28, 35, 42, . . . **b** 2, 6, 18, 54, . . .

Solution

a 21, 28, 35, 42, . . .

1 Determine whether the difference between successive terms is constant.

Differences:

$28 - 21 = 7$

$35 - 28 = 7$

$42 - 35 = 7$

2 Write your conclusion.

As the differences between successive terms are constant, the sequence is arithmetic.

b 2, 6, 18, 54, . . .

1 Determine whether the difference between successive terms is constant.

Differences:

$6 - 2 = 4$

$18 - 6 = 12$

$54 - 18 = 36$

2 Write your conclusion.

As the differences between successive terms are not constant, the sequence is not arithmetic.

Example 11 Defining an arithmetic sequence with a recurrence relation

For the sequences below:

- **i** write the value of a and d
- **ii** write a recurrence relation that generates the terms of the sequence
- **iii** use the recurrence relation to find the value of the 5th term, T_5.

a $12, 20, 28, \ldots$

b $14, 11, 8, \ldots$

Solution

a $12, 20, 28, \ldots$

i **1** Write the first term.

$a = 12$

2 Check that the sequence is arithmetic by finding a common difference.

$20 - 12 = 8$

$28 - 20 = 8$

Both calculations have the same result.

$d = 8$

ii Write the recurrence relation.

$T_1 = a, \quad T_{n+1} = T_n + d$

$T_1 = 12, \quad T_{n+1} = T_n + 8$

iii Use recursion (by hand or calculator) to find T_5.

$T_3 = 28$

$T_4 = T_3 + 8 \quad T_5 = T_4 + 8$

$= 28 + 8 \quad = 36 + 8$

$= 36 \quad = 44$

b $14, 11, 8, \ldots$

i **1** Write the first term.

$a = 14$

2 Check that the sequence is arithmetic by finding a common difference.

$11 - 14 = -3$

$8 - 11 = -3$

Both calculations have the same result.

$d = -3$

ii Write the recurrence relation.

$T_1 = a, \quad T_{n+1} = T_n + d$

$T_1 = 14, \quad T_{n+1} = T_n - 3$

iii Use recursion (by hand or calculator) to find T_5.

$T_3 = 8$

$T_4 = T_3 - 3 \quad T_5 = T_4 - 3$

$= 8 - 3 \quad = 5 - 3$

$= 5 \quad = 2$

Graphs of arithmetic sequences

If we plot the values of the terms of an arithmetic sequence (T_n) against their number (n) or position in the sequence, we will find that the points lie on a straight line. We could anticipate this because, as we progress through the sequence, the value of successive terms increases by the same amount (the common difference, d).

The advantages of graphing a sequence are that the straight line required for an arithmetic sequence is immediately obvious and any exceptions would stand out very clearly. An upward slope indicates increasing terms and a downward slope indicates decreasing terms.

Note: When graphing the points of a sequence, do not connect the points as each term is a discrete value.

Example 12 Graphing the terms of an increasing arithmetic sequence ($d > 0$)

The sequence 4, 7, 10, . . . is arithmetic with common difference $d = 3$.

a Construct a table showing the term number (n) and its value (T_n) for the first four terms in the sequence.

b Use the table to plot the graph.

c Describe the graph.

Solution

a Show the term numbers and values of the first four terms in a table.

1 Write the term numbers in top row of the table.

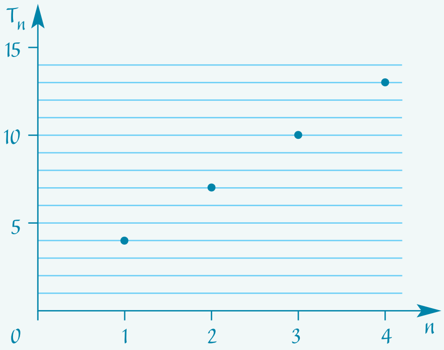
2 Write the values of the terms in the bottom row.

Term number, n	1	2	3	4
Term value, T_n	4	7	10	13

b Use the table to plot the graph.

1 Use the horizontal axis, n, the term numbers.
Use the vertical axis for the value of each term, T_n.

2 Plot each point from the table.



c Describe the graph.

1 Are the points along a straight line or a curve?

2 Is the line of the points rising (positive slope) or falling (negative slope)?

The points of an arithmetic sequence with $d = 3$ and $a = 4$ lie along a rising straight line.
The line has a positive slope.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Example 13 Graphing the terms of a decreasing arithmetic sequence ($d < 0$)

An arithmetic sequence is generated from the recurrence relation $T_1 = 9, \quad T_{n+1} = T_n - 2$

a Construct a table showing the term number (n) and its value (T_n) for the first four terms in the sequence.

b Use the table to plot the graph.

c Describe the graph.

Solution

a Show the term numbers and values of the first four terms in a table.

1 Write the term numbers in the top row of the table.

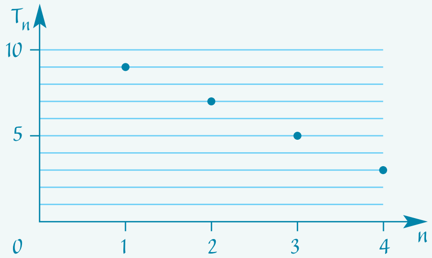
2 Write the values of the terms in the bottom row.

Term number, n	1	2	3	4
Term value, T_n	9	7	5	3

b Use the table to plot the graph.

1 Use the horizontal axis, n, for the term numbers.
Use the vertical axis for the value of each term, T_n.

2 Plot each point from the table.



c Describe the graph.

1 Are the points along a straight line or a curve?

The points of an arithmetic sequence with $d = -2$ and $a = 9$ lie along a falling straight line.

2 Is the line of the points rising (positive slope) or falling (negative slope)?

The line has a negative slope.

Graphs of arithmetic sequences

Graphs of arithmetic sequences are:

- points along a line with positive slope, when a constant amount is added ($d > 0$)
- points along a line with negative slope, when a constant amount is subtracted ($d < 0$).

A line with *positive* slope rises from left to right. A line with *negative* slope falls from left to right.

Exercise 4C

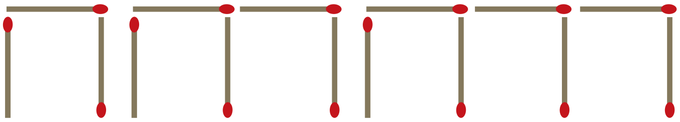
Basic ideas of arithmetic sequences

Example 10

1 Find out which of the sequences below is arithmetic. Give the common difference for each sequence that is arithmetic.

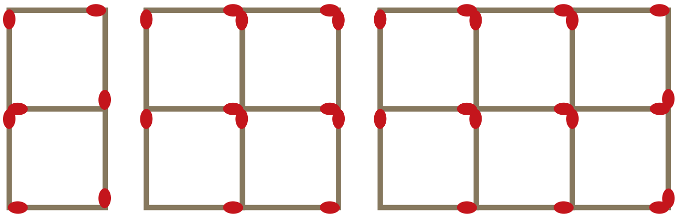
a 8, 11, 14, 17, . . .
b 7, 15, 22, 30, . . .
c 11, 7, 3, −1, . . .
d 12, 9, 6, 3, . . .
e 16, 8, 4, 2, . . .
f 1, 1, 1, 1, . . .

2 The numbers of matches used to construct the succession of shapes shown follow an arithmetic sequence.



a Write the sequence for the number of matches used in each structure.
b Find the common difference for this sequence.
c How many matches would be used in the next two structures?

3 Matchsticks have been used to make the shapes below.



a Give the sequence for the number of matchsticks used in each shape.
b Is the sequence for the number of matchsticks used arithmetic? If it is arithmetic, give the common difference.
c How many matchsticks would be used in each of the next two shapes?

Finding common differences and generating new terms

Example 9

4 For each of these arithmetic sequences, find the common difference and the 5th term.

a 5, 11, 17, 23, . . .
b 17, 13, 9, 5, . . .
c 11, 15, 19, 23, . . .
d 8, 4, 0, −4, . . .
e 35, 30, 25, 20, . . .
f 1.5, 2, 2.5, 3, . . .

5 Give the next two terms in each of these arithmetic sequences.

a 17, 23, 29, 35, . . .
b 14, 11, 8, 5, . . .
c 2, 1.5, 1.0, 0.5, . . .
d 27, 35, 43, 51, . . .
e 33, 21, 9, −3, . . .
f 0.8, 1.1, 1.4, 1.7, . . .

6 Find the missing terms in the following arithmetic sequences.

a $8, 13, 18, 23, \square, \square, \ldots$

b $14, 8, 2, -4, \square, \square, \ldots$

c $6, 15, \square, \square, 42, \ldots$

d $23, 18, \square, \square, 3, -2, \ldots$

e $3, \square, \square, 27, 35, 43, \ldots$

f $\square, \square, 29, 37, 45, 53, \ldots$

g $\square, \square, 7, -4, -15, -26, \ldots$

h $36, \square, 22, \square, 8, 1, \ldots$

i $15, \square, 31, \square, 47, \square, \ldots$

Defining an arithmetic sequence with a recurrence relation

7 Each of the following recurrence relations generates an arithmetic sequence. For each recurrence relation, write

i the first term **ii** the common difference **iii** T_4

for the sequence that it generates.

a $T_1 = 1, \quad T_{n+1} = T_n + 5$

b $T_1 = 45, \quad T_{n+1} = T_n - 2$

c $T_1 = 34, \quad T_{n+1} = T_n + 6$

d $T_1 = 0, \quad T_{n+1} = T_n + 3.5$

Example 11

8 Write a recurrence relation that generates the terms of the following arithmetic sequences.

a $3, 8, \ldots$

b $16, 9, \ldots$

c $1.6, 3.9, \ldots$

d $8.7, 5.6, \ldots$

e $293, 226, \ldots$

Graphing arithmetic sequences

Example 12

9 For the following arithmetic sequences:

a $3, 5, 7, \ldots$

b $11, 8, 5, \ldots$

i write the next term

ii show the term positions and values of the first four terms in a table

iii use the table to plot a graph

iv describe the graph

4D A rule for the *n*th term in an arithmetic sequence

Using recursion from a recurrence relation to find terms of an arithmetic sequence is convenient if the number of the term you want to find is small. It is easy to calculate T_6, for example, when you know the first term a and the common difference, d. You can determine this value with simple arithmetic.

But if the number of the term you want to find is large, say T_{150}, the calculations would be tedious and time consuming. Luckily, there is a rule that can be used to determine any term of an arithmetic sequence.

ISBN 978-1-009-25917-0

Identifying the pattern of successive terms of an arithmetic sequence

We know that an arithmetic sequence starts with a first term and then adds the common difference to each term to generate the next term. A pattern can be observed when we write these calculations out using the symbols a, d and n.

For an arithmetic sequence with first term a, common difference d and recurrence relation $T_l = a, \qquad T_{n+1} = T_n + d$:

$$\begin{aligned} T_1 &= a \\ T_2 &= T_1 + d = a + d \\ T_3 &= T_2 + d = (a + d) + d = a + 2d \\ T_4 &= T_3 + d = (a + 2d) + d = a + 3d \\ T_5 &= T_4 + d = (a + 3d) + d = a + 4d \end{aligned}$$

It can help identify the pattern in these terms to write the term number (n) and the term value (T_n) in a table.

n	T_n
1	a
2	$a + d$
3	$a + 2d$
4	$a + 3d$
5	$a + 4d$

Notice that the number of common difference values, d, that are added is one less than the term number n every time. For example,

$$T_4 = a + 3d$$

The term number is 4 and 3 times d is added. 3 is one less than 4.

This patterns means that the nth term of the sequence can be found by adding $(n - 1)$ times the common difference to the first term.

The *n*th term of an arithmetic sequence

The nth term, T_n, of an arithmetic sequence with

- first term a
- common difference d
- recurrence relation $T_1 = a, \qquad T_{n+1} = T_n + d$

can be found using the rule

$$T_n = a + (n - 1)d$$

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Example 14 Writing the rule for the *n*th term of an arithmetic sequence

Write the rule for the *n*th term of the arithmetic sequence: 15, 12, 9, 6, . . .

Solution

1 Write the first term.

The first term is 15, so $a = 15$.

2 Calculate the common difference.

Note: The question told us that the sequence is arithmetic, but you could still check this with other common difference calculations.

$d = 12 - 15 = -3$

Check: $d = 6 - 9 = -3$ $\quad d = 6 - 9 = -3$

3 Write the rule for the *n*th term of the sequence.

The nth term has rule: $T_n = a + (n - 1)d$

$T_n = 15 + (n - 1)(-3)$

Note: When substituting values into the *n*th term rule it is important to use brackets around the value for *d*, to indicate that this is being multiplied by $(n - 1)$.

Example 15 Finding the *n*th term of an arithmetic sequence

Find T_{30}, the 30th term in the arithmetic sequence: 21, 18, 15, 12, . . .

Strategy: To use $T_n = a + (n - 1)d$ we need to know *a*, *d* and *n*.

Solution

1 The first term is 21.

$a = 21$

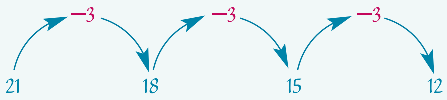
2 The term T_{30} is in the 30th position.

$n = 30$

3 Find the difference in the first two terms.

$$d = T_2 - T_1 = 18 - 21 = -3$$

4 Check that this difference generates the sequence.



5 Substitute the values of *a*, *d* and *n* into $T_n = a + (n - 1)d$.

$$T_n = a + (n - 1)d$$

$$T_{30} = 21 + (30 - 1)(-3)$$

6 Evaluate.

$$= 21 + 29(-3)$$

$$= -66$$

7 Write your answer.

The 30th term, T_{30}, is -66.

The rule for the *n*th term of an arithmetic sequence gives us a quick way of finding the required term in practical applications.

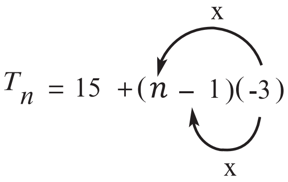
In this chapter we have found two general rules to help us find the *n*th term of a sequence. All general rules for an arithmetic sequence can be simplified into the form $T_n = a + bn$ by multiplying out the brackets and simplifying.

From Example 14, the rule was $T_n = 15 + (n - 1)(-3)$

This can be simplified by:

1 Multiplying the difference (d) into the bracket

2 Adding the constants $= 18 - 3n$



$$T_n = 15 + (n - 1)(-3)$$
$$= 15 - 3n + 3$$
$$= 18 - 3n$$

Therefore: $T_n = 18 - 3n$

This version of the nth term rule is useful for solving associated sequence problems including finding terms without a calculator.

Sketching a graph of an arithmetic sequence from a rule for the *n*th term using a calculator

Previously, graphs of sequences were drawn by plotting points from a table of values.

Graphs of arithmetic sequences can also be created using a CAS calculator.

How to draw a graph of an arithmetic sequence using the T1-Nspire CAS

An arithmetic sequence is generated from the recurrence relation $T_1 = 9,\ T_{n+1} = T_n - 2$.

a Write the rule for the nth term of the sequence.

b Use your calculator to draw a graph of the sequence terms.

Steps

a **1** Write the first term and the common difference for the sequence.

The first term is T_1, so $a = 9$.
The common difference is the amount added in the recurrence relation, so $d = -2$.

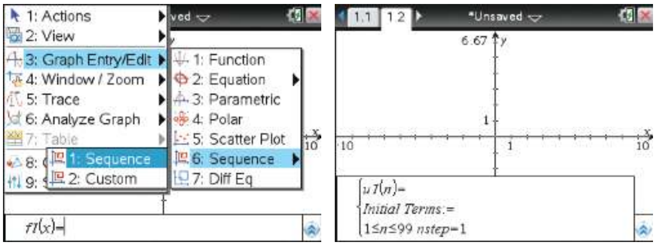
2 Write the rule for the nth term.

$T_n = a + (n - 1)\,d$

$T_n = 9 + (n - 1)(-2)$

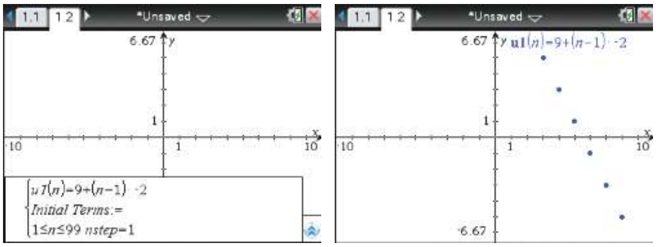
b **1** Start with a blank graph page.

2 Press MENU, then 3 : Graph Entry/Edit, then 6 : Sequence
Choose 1 : Sequence



$u1(n)=$
Initial Terms:=
$1\leq n\leq 99$ nstep=1

3 Use the keyboard to enter the rule for the *n*th term next to $u1\,(n)$ =.
Leave Initial Terms:= blank.
Leave $1 \leq n \leq 99$ $nstep = 1$

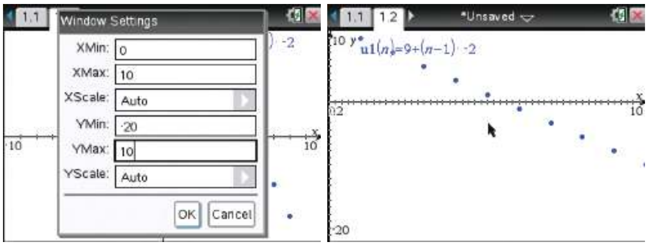


$u1(n)=9+(n-1)\cdot{-2}$
Initial Terms:=
$1\leq n\leq 99$ nstep=1

4 Press ENTER to see the graph.

Note: Make sure you use the (–) symbol for 'negative 2'

5 Change the window settings (Menu, 4: Window/Zoom, 1: Window Settings) to see a better view of the graph.



XMin: 0
XMax: 10
XScale: Auto
YMin: -20
YMax: 10
YScale: Auto

How to draw a graph of an arithmetic sequence using the ClassPad

An arithmetic sequence is generated from the recurrence relation $T_1 = 9, \quad T_{n+1} = T_n - 2$.

a Write the rule for the *n*th term of the sequence.

b Use your calculator to draw a graph of the sequence terms.

Steps

a **1** Write the first term and the common difference for the sequence.

The first term is T_1, so $a = 9$.
The common difference is the amount added in the recurrence relation so $d = -2$.

2 Write the rule for the *n*th term.

$T_n = a + (n-1)\,d$

$T_n = 9 + (n-1)\,(-2)$

ISBN 978-1-009-25917-0

Cambridge University Press

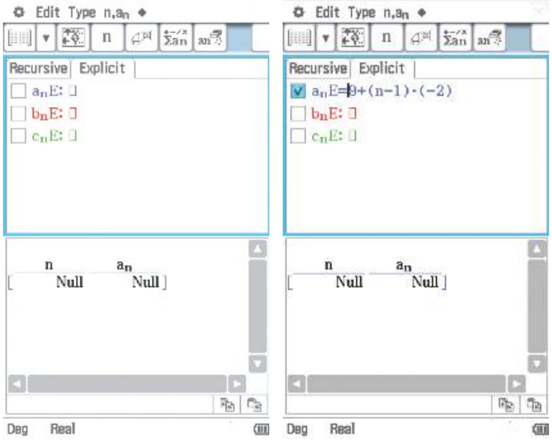
Updated January 2024

b **1** Tap menu and choose ‘Sequence’.

2 Make sure the explicit tab is selected.

Tap in the square after a_nE: and enter the rule for *n*th term of the sequence.

Use the dedicated ‘n’ tab at the top of the screen or the keyboard.



☐ a_nE:
☐ b_nE:
☐ c_nE:

n	a_n
Null	Null

☑ a_nE=9+(n−1)·(−2)
☐ b_nE:
☐ c_nE:

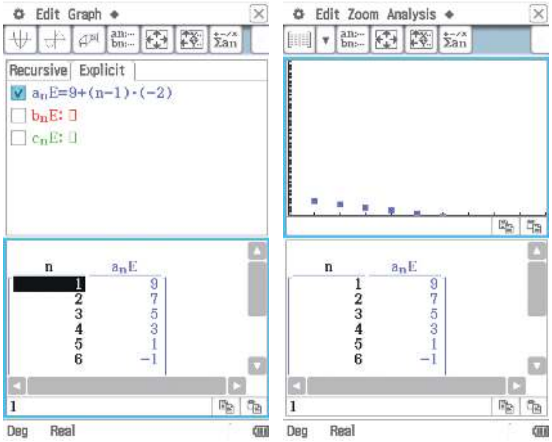
n	a_n
Null	Null

3 Select the recurrence relation by tapping in the box next to it to insert a ☑.

4 Tap ▦ to enter the values in the table.

The ▦ symbol will be replaced by some graph symbols.

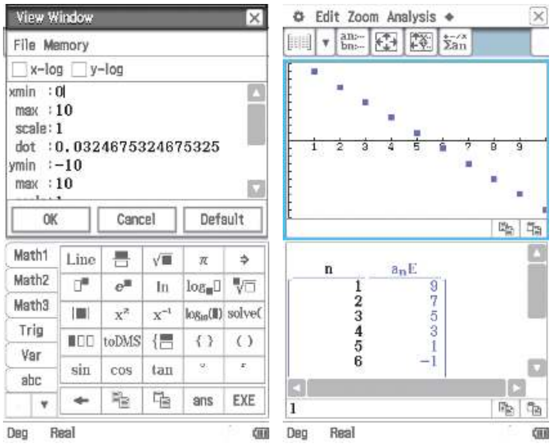
Tap ⊥ to plot a graph of the terms in the table.



☑ a_nE=9+(n−1)·(−2)
☐ b_nE:
☐ c_nE:

n	a_nE
1	9
2	7
3	5
4	3
5	1
6	−1

5 Tap ⊥ to change the view window settings to give a better view of the graph.



☐ x-log ☐ y-log
xmin : 0
max : 10
scale : 1
dot : 0.0324675324675325
ymin : −10
max : 10

n	a_nE
1	9
2	7
3	5
4	3
5	1
6	−1

Exercise 4D

Notation: working out *a* and *d*

1 Give the value of a and d in each of the following arithmetic sequences.

a $7, 11, 15, 19, \ldots$
b $8, 5, 2, -1, \ldots$
c $14, 23, 32, 41, \ldots$
d $62, 35, 8, -19, \ldots$
e $-9, -4, 1, 6, \ldots$
f $-13, -19, -25, -31, \ldots$

Writing the rule for the *n*th term of an arithmetic sequence

Example 14

2 Write the rule for the nth term of the following arithmetic sequences.

a $12, 21, 30, \ldots$
b $146, 196, 246, \ldots$
c $27, 22, 17, \ldots$
d $8, 3, -2, \ldots$

3 Write the rule for the nth term of the following recurrence relations that generate arithmetic sequences.

a $T_1 = 5, \quad T_{n+1} = T_n + 9$
b $T_1 = -8, \quad T_{n+1} = T_n + 3$
c $T_1 = 3, \quad T_{n+1} = T_n - 7$
d $T_1 = -1, \quad T_{n+1} = T_n - 6$

Finding the *n*th term of an arithmetic sequence

Example 15

4 Find the required term in each of these arithmetic sequences.

a $18, 21, 24, 27, \ldots$ Find T_{35}.
b $-14, -6, 2, 10, \ldots$ Find T_{41}.
c $27, 14, 1, -12, \ldots$ Find T_{37}.
d $16, 31, 46, 61, \ldots$ Find T_{29}.
e $-19, -23, -27, -31, \ldots$ Find T_{26}.
f $0.8, 1.5, 2.2, 2.9, \ldots$ Find T_{36}.
g $82, 68, 54, 40, \ldots$ Find T_{21}.
h $9.4, 8.8, 8.2, 7.6, \ldots$ Find T_{29}.

5 Find the 40th term in an arithmetic sequence that starts at 11 and has a common difference of 8.

6 The first term in an arithmetic sequence is 27 and the common difference is 19. Find the 100th term, T_{100}.

7 A sequence started at 100 and had 7 subtracted each time to make new terms. Find the 20th term, T_{20}.

ISBN 978-1-009-25917-0

Sketching a graph of an arithmetic sequence from the rule for the *n*th term.

CAS

8 Use your calculator to sketch a graph of the sequences that have the following rules for the *n*th term. Show 5 terms in your answer.

- **a** $T_n = 5 + (n - 1) \times 3$
- **b** $T_n = -4 + (n - 1) \times 5$
- **c** $T_n = 5 + (n - 1) \times -6$
- **d** $T_n = -18 + (n - 1) \times -2$

4E Geometric sequences

Common ratio, *r*

In a **geometric sequence**, each new term is made by multiplying the previous term by a fixed number called the **common ratio**, ***r***. This repeating or recurring process is another example of a sequence generated by *recursion*.

In the sequence:

$$\begin{tikzcd}
2, \arrow[r, bend left, "\times 3"] & 6, \arrow[r, bend left, "\times 3"] & 18, \arrow[r, bend left, "\times 3"] & 54, \arrow[r, bend left, "\times 3"] & \ldots \\
T_1 & T_2 & T_3 & T_4 &
\end{tikzcd}$$

each new term is made by multiplying the previous term by 3. The common ratio is 3.

In the sequence:

$$\begin{tikzcd}
20, \arrow[r, bend left, "\times \frac{1}{2}"] & 10, \arrow[r, bend left, "\times \frac{1}{2}"] & 5, \arrow[r, bend left, "\times \frac{1}{2}"] & 2.5, \arrow[r, bend left, "\times \frac{1}{2}"] & \ldots \\
T_1 & T_2 & T_3 & T_4 &
\end{tikzcd}$$

each new term is made by halving the previous term. In this sequence we are multiplying each term by $\times \frac{1}{2}$, which is equivalent to dividing by 2. The common ratio is $\frac{1}{2}$.

New terms in a geometric sequence $T_1, T_2, T_3, T_4, \ldots$ are made by multiplying the previous term by the common ratio, *r*.

$$\begin{tikzcd}
T_1 \arrow[r, bend left, "\times r"] & T_2 \arrow[r, bend left, "\times r"] & T_3 \arrow[r, bend left, "\times r"] & T_4 \arrow[r, bend left, "\times r"] & \ldots
\end{tikzcd}$$

So $T_1 \times r = T_2$ and $T_2 \times r = T_3$ and so on.

$$r = \frac{T_2}{T_1} \qquad r = \frac{T_3}{T_2}$$

ISBN 978-1-009-25917-0

Cambridge University Press

Updated January 2024

Common ratio, r

In a *geometric sequence*, the *common ratio, r,* is found by dividing the next term by the current term.

$$\text{Common ratio } r = \frac{\text{any term}}{\text{the previous term}} = \frac{T_2}{T_1} = \frac{T_3}{T_2} = \frac{T_4}{T_3} = \ldots$$

As preparation for our study of growth and decay we will be using $r > 0$.

Examples of geometric series

Examples of geometric sequences and their recurrence relations are shown in the table below.

Sequence	*Description*	*Recurrence relation*
A. $2, 6, 18, 54, 162, \ldots$	First term = 2 Rule: "multiply by 3"	$T_1 = 2, \quad T_{n+1} = 3T_n$
B. $1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \ldots$	First term = 1 Rule: "multiply by one half"	$T_1 = 1, \quad T_{n+1} = \frac{1}{2}T_n$

If a sequence is *known* to be geometric, the common ratio can be calculated by dividing any term by the term that is before it in the sequence. It doesn't matter which terms are chosen for the calculation, the common ratio will be the same for any pair of successive terms.

Sequence	*Common difference calculations*	*Common difference*
A $2, 6, 18, 54, 162, \ldots$	$6 \div 2 = 3 \quad 54 \div 18 = 3 \; 18 \div 6 = 3 \quad 162 \div 54 = 3$	$r = 3$
B $1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \ldots$	$\frac{1}{2} \div 1 = \frac{1}{2} \quad \frac{1}{8} \div \frac{1}{4} = \frac{1}{2} \; \frac{1}{4} \div \frac{1}{2} = \frac{1}{2} \quad \frac{1}{16} \div \frac{1}{8} = \frac{1}{2}$	$r = \frac{1}{2}$

In sequence **A**, the common ratio is greater than one. This causes the terms of this sequence to *increase* in size through the sequence.

In sequence **B**, the common ratio is between 0 and 1. This causes the terms of this sequence to *decrease* in size through the sequence, towards zero.

Note: We will only explore geometric sequences with a common ratio greater than zero, that is $r > 0$.

Recurrence relations for geometric sequences

It is common to use the symbol a to represent the first term of a geometric sequence, just like we did for an arithmetic sequence.

A geometric sequence is shown below.

$$10, 50, 250, \ldots$$

We know the geometric sequence information and we can write a recurrence relation that generates this sequence.

Arithmetic sequence information	*Recurrence relation*
first term: $a = 10$ common ratio: $r = 5$	$T_1 = 10, \quad T_{n+1} = 5T_n$

We can match a to 'T_1' and r to '5×' in the recurrence relation to determine the general recurrence relation for a geometric sequence.

Geometric sequences

A sequence is geometric if each term is multiplied by a fixed amount to generate the next.

In any geometric sequence, the common ratio, r, is the amount multiplied in the recurrence relation to calculate the next term.

$$r = \text{ any term } \div \text{ previous term}$$

In general,

$$r = T_n \div T_{n-1}$$

Geometric sequences are generated by a recurrence relation of the form

$$T_1 = a, T_{n+1} = rT_n$$

where a is the first term and r is the common ratio.

If $r > 1$, the terms of the sequence will increase.

If $0 < r < 1$, that is if r is between zero and 1, the terms of the sequence will decrease towards zero.

Example 16 Finding the common ratio, *r*, in a geometric sequence

Find the common ratio in each of the following geometric sequences.

a 3, 12, 48, 192, . . . **b** 16, 8, 4, 2, . . .

Solution

a 3, 12, 48, 192, . . .

1 The common ratio is equal to any term divided by its previous term.

Common ratio, $r = \frac{T_2}{T_1} = \frac{12}{3} = 4$

2 Check that multiplying by 4 makes each new term.

$$\begin{tikzcd}
3 \arrow[r, bend left, "\times 4"] & 12 \arrow[r, bend left, "\times 4"] & 48 \arrow[r, bend left, "\times 4"] & 192 \arrow[r, bend left, "\times 4"] & \ldots
\end{tikzcd}$$

3 Write your answer.

The common ratio is 4.

b 16, 8, 4, 2, . . .

1 Find the common ratio, *r*.

Common ratio, $r = \frac{T_2}{T_1} = \frac{8}{16} = \frac{1}{2}$

2 Check that multiplying by $\frac{1}{2}$ makes each new term.

$$\begin{tikzcd}
16 \arrow[r, bend left, "\times \frac{1}{2}"] & 8 \arrow[r, bend left, "\times \frac{1}{2}"] & 4 \arrow[r, bend left, "\times \frac{1}{2}"] & 2 \arrow[r, bend left, "\times \frac{1}{2}"] & \ldots
\end{tikzcd}$$

3 Write your answer.

The common ratio is $\frac{1}{2}$.

ISBN 978-1-009-25917-0

Identifying geometric sequences

To identify a sequence as geometric, it is necessary to find a common ratio between successive terms.

Example 17 Identifying a geometric sequence

Which of the following is a geometric sequence?

a 2, 10, 50, 250, . . .

b 3, 6, 18, 36, . . .

Solution

a 2, 10, 50, 250, . . .

1 Find the ratio (multiplier) between successive terms.

$$r = \frac{T_2}{T_1} \qquad r = \frac{T_3}{T_2} \qquad r = \frac{T_4}{T_3}$$

$$r = \frac{10}{2} \qquad r = \frac{50}{10} \qquad r = \frac{250}{50}$$

$$r = 5 \qquad r = 5 \qquad r = 5$$

2 Check that the ratios are the same.

The common ratio is 5.

3 Write your conclusion.

The sequence is geometric.

b 3, 6, 18, 36, . . .

1 Find the ratios of successive terms.

$$r = \frac{T_2}{T_1} \qquad r = \frac{T_3}{T_2} \qquad r = \frac{T_4}{T_3}$$

$$r = \frac{6}{3} \qquad r = \frac{18}{6} \qquad r = \frac{36}{18}$$

$$r = 2 \qquad r = 3 \qquad r = 2$$

2 Are the ratios the same?

The ratios are not the same.

3 Write your conclusion.

The sequence is not geometric.

Example 18 Defining a geometric sequence with a recurrence relation

For the sequences below:

- **i** write the value of a and r
- **ii** write a recurrence relation that generates the terms of the sequence
- **iii** use the recurrence relation to find the value of the 5th term, T_5.

a $7, 35, 175, \ldots$ **b** $3072, 768, 192, \ldots$

Solution

a $7, 35, 175, \ldots$

i **1** Write the first term.

$a = 7$

2 Check that the sequence is geometric by finding a common ratio.

$35 \div 7 = 5$

$175 \div 35 = 5$

Both calculations have the same result.

$r = 5$

ii Write the recurrence relation.

$T_1 = a, \quad T_{n+1} = r \times T_n$

$T_1 = 7, \quad T_{n+1} = 5 \times T_n$

iii Use recursion (by hand or calculator) to find T_5.

$T_3 = 175$

$$\begin{aligned} T_4 &= 5 \times T_3 \\ &= 5 \times 175 \\ &= 875 \end{aligned} \qquad \begin{aligned} T_5 &= 5 \times T_4 \\ &= 5 \times 875 \\ &= 4375 \end{aligned}$$

b $3072, 768, 192, \ldots$

i **1** Write the first term.

$a = 3072$

2 Check that the sequence is geometric by finding a common ratio.

$768 \div 3072 = 0.25$

$192 \div 768 = 0.25$

Both calculations have the same result.

$r = 0.25$ or $\frac{1}{4}$

ii Write the recurrence relation.

$T_1 = a, \quad T_{n+1} = r \times T_n$

$T_1 = 3072, \quad T_{n+1} = 0.25 \times T_n$

iii Use recursion (by hand or calculator) to find T_5.

$T_3 = 192$

$$\begin{aligned} T_4 &= 0.25 \times T_3 \\ &= 0.25 \times 192 \\ &= 48 \end{aligned} \qquad \begin{aligned} T_5 &= 0.25 \times T_4 \\ &= 0.25 \times 48 \\ &= 12 \end{aligned}$$

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Graphs of geometric sequences

In contrast with the straight-line graph of an arithmetic sequence, the values of a geometric sequence lie along a curve, of increasing values or decreasing values, depending on the value of the common ratio, *r*.

Example 19 Graphing an increasing geometric sequence ($r > 1$)

Consider the geometric sequence: 2, 6, 18, . . .

a Find the next term.

b Show the positions and values of the first four terms in a table.

c Use the table to plot the graph.

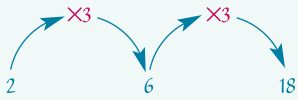
d Describe the graph.

Solution

a **1** Find the common ratio using $r = \frac{T_2}{T_1}$.

Common ratio, $r = \frac{T_2}{T_1} = \frac{6}{2} = 3$

2 Check that this ratio makes the given terms.



3 Multiply 18 by 3 to make the next term, 54.

4 Write your answer.

The next term is 54.

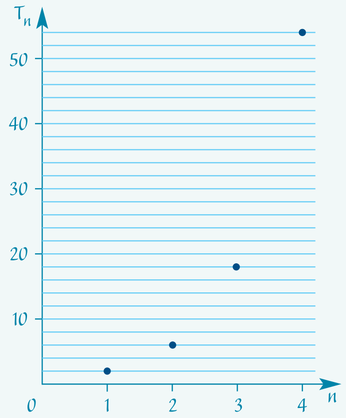
b **1** Number the positions along the top row of the table.

2 Write the terms in the bottom row.

Position, n	1	2	3	4
Term, T_n	2	6	18	54

c **1** Use the horizontal axis, *n*, for the position of each term.
Use the vertical axis, T_n, for the value of each term.

2 Plot each point from the table.



d Describe the pattern revealed by the graph.

The values lie along a curve and they are increasing.

Example 20 Graphing a decreasing geometric sequence ($0 < r < 1$)

Consider the geometric sequence: 32, 16, 8, . . .

a Find the next term.

b Show the positions and values of the first four terms in a table.

c Use the table to plot the graph.

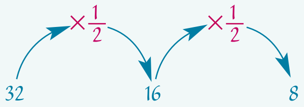
d Describe the graph.

Solution

a **1** Find the common ratio using $r = \frac{T_2}{T_1}$.

Common ratio, $r = \frac{T_2}{T_1} = \frac{16}{32} = \frac{1}{2}$

2 Check that this ratio makes the given terms.



3 Multiply 8 by $\times\frac{1}{2}$ to make the next term, 4.

4 Write your answer.

The next term is 4.

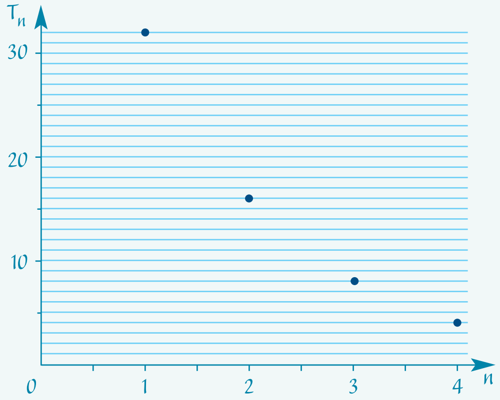
b **1** Number the positions along the top row of the table.

2 Write the terms in the bottom row.

Position, n	1	2	3	4
Term, T_n	32	16	8	4

c **1** Use the horizontal axis, n, for the position of each term.
Use the vertical axis, T_n, for the value of each term.

2 Plot each point from the table.



d Describe the pattern revealed by the graph.

The graph is a curve with values decreasing and approaching zero.

Graphs of geometric sequences (for r positive)

Graphs of geometric sequences for $r > 0$ are:

- *increasing* when r is greater than 1, $r > 1$
- *decreasing* towards zero when r is less than 1, $0 < r < 1$.

Exercise 4E

Identifying a geometric sequence

Example 17 **1** Find out which of the following sequences are geometric. Give the common ratio for each sequence that is geometric.

a 4, 8, 16, 32, . . . **b** 1, 3, 9, 27, . . . **c** 5, 10, 15, 20, . . .
d 5, 15, 45, 135, . . . **e** 24, 12, 6, 3, . . . **f** 3, 6, 12, 18, . . .
g 4, 8, 12, 16, . . . **h** 27, 9, 3, 1, . . . **i** 2, 4, 8, 16, . . .

Finding common ratios and generating new terms

Example 16 **2** Find the common ratio for each of the following geometric sequences.

a 3, 6, 12, 24, . . . **b** 64, 16, 4, 1, . . .
c 6, 30, 150, 750, . . . **d** 2, 8, 32, 128, . . .
e 32, 16, 8, 4, . . . **f** 2, 12, 72, 432, . . .
g 10, 100, 1000, 10 000, . . . **h** 3, 21, 147, 1029, . . .
i 7, 56, 448, 3584, . . .

3 Find the missing terms in each of these geometric sequences.

a 7, 14, 28, ☐, ☐, . . . **b** 3, 15, 75, ☐, ☐, . . . **c** 4, 12, ☐, ☐, 324, . . .
d ☐, ☐, 20, 40, 80, . . . **e** 2, ☐, 32, 128, ☐, . . . **f** 3, ☐, 27, ☐, 243, 729, . . .

Defining a geometric sequence with a recurrence relation

4 Each of the following recurrence relations generates a geometric sequence. For each recurrence relation, write

i the first term **ii** the common ratio **iii** T_4

for the sequence that it generates.

a $T_1 = 7, \quad T_{n+1} = 5 \times T_n$ **b** $T_1 = 3, \quad T_{n+1} = 6 \times T_n$
c $T_1 = 96, \quad T_{n+1} = \frac{1}{2} \times T_n$ **d** $T_1 = 160, \quad T_{n+1} = \frac{1}{4} \times T_n$

Example 18 **5** Write a recurrence relation that generates the terms of the following geometric sequences.

a 2, 14, . . . **b** 15, 3, . . . **c** 24, 3, . . . **d** 9, 27, . . .

Graphing geometric sequences

Example 19 **6** Consider each of the geometric sequences below.

i Find the next term. **ii** Show the first four terms in a table.
iii Use the table to plot a graph. **iv** Describe the graph.

a 3, 6, 12, . . . **b** 8, 4, 2, . . .

4F A rule for the *n*th term in a geometric sequence

Identifying the pattern using algebraic symbols

We know that a geometric sequence starts with a first term and then multiplies the common ratio to each term to generate the next term. A pattern can be observed when we write these calculations using the symbols a, r and n.

For a geometric sequence with first term a, common ratio r and recurrence relation $T_1 = a, \qquad T_{n+1} = r \times T_n$:

$$\begin{aligned} T_1 &= a \\ T_2 &= r \times T_1 = r \times a \\ T_3 &= r \times T_2 = r \times (r \times a) = r^2 \times a \\ T_4 &= r \times T_3 = r \times (r^2 \times a) = r^3 \times a \\ T_5 &= r \times T_4 = r \times (r^3 \times a) = r^4 \times a \end{aligned}$$

It can help identify the pattern in these terms to write the term number (n) and the term value (T_n) in a table.

n	T_n
1	a
2	$r \times a$
3	$r^2 \times a$
4	$r^3 \times a$
5	$r^4 \times a$

Notice that the power of the common ratio r is one less than the term number n every time. For example,

$$T_4 = r^3 \times a$$

The term number is 4 and the power of r is 3. 3 is one less than 4.

This patterns means that the nth term of the sequence can be found by multiplying the first term, a, by r to the power of $(n - 1)$.

The *n*th term of a geometric sequence

The nth term, T_n, of a geometric sequence with

- first term a
- common ratio r
- recurrence relation $T_1 = a, \qquad T_{n+1} = rT_n$

can be found using the rule

$$T_n = a \times r^{(n-1)}$$

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Example 21 Writing the rule for the *n*th term of a geometric sequence

Write the rule for the *n*th term of the arithmetic sequence: 2, 14, 98, 686, . . .

Solution

1 Write the first term.

The first term is 2, so $a = 2$.

2 Calculate the common ratio.

Note: The question told us that the sequence is geometric, but you could still check this with other common ratio calculations.

$r = 14 \div 2 = 7$

Check: $r = 98 \div 14$ $\quad r = 686 \div 98$

$= 7$ $\quad = 7$

3 Write the rule for the *n*th term of the sequence.

The nth term has rule: $T_n = a \times r^{(n-1)}$

$T_n = 2 \times 7^{(n-1)}$

Example 22 Finding the *n*th term of a geometric sequence

Find T_{15}, the 15th term in the geometric sequence: 3, 6, 12, 24, . . .

Strategy: To use $T_n = ar^{n-1}$, we need to know a, r and n.

Solution

1 The first term is 3.

$a = 3$

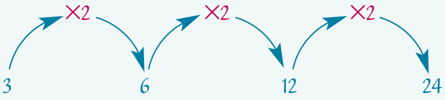
2 The term T_{15} is in the 15th position.

$n = 15$

3 Find the ratio of the first two terms.

$$r = \frac{T_2}{T_1} = \frac{6}{3} = 2$$

4 Check that this ratio generates the sequence.



5 Substitute the values of a, r and n into $T_n = ar^{n-1}$.

$T_n = ar^{n-1}$

6 Use a calculator to find T_{15}.

$$T_{15} = 3 \times 2^{15-1}$$
$$= 3 \times 2^{14}$$
$$= 49\,152$$

7 Write your answer.

The 15th term, T_{15}, is 49 152.

Sketching a graph of a geometric sequence from a rule for the *n*th term using a calculator

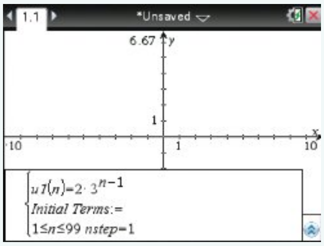
Exactly the same technique is used to create a graph of geometric sequences from a rule for the *n*th term as was used for arithmetic sequences. The only thing that changes is the format of the rule that is entered as $u1(n)$ on the TI-Nspire, or a_nE on the ClassPad.

Example 23 Using a calculator to sketch a graph of a geometric sequence from a rule for the *n*th term

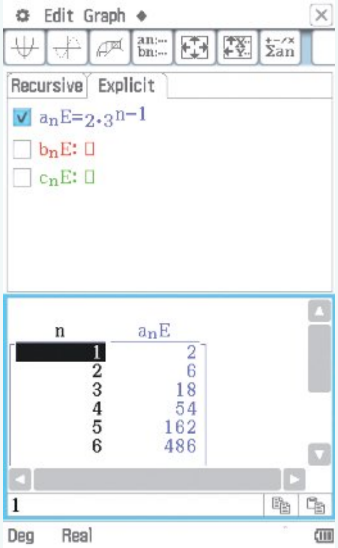
Use your calculator to sketch a graph of the sequence with *n*th term given by the rule $T_n = 2 \times 3^{(n-1)}$. Show 5 terms on your graph.

Solution

1 Enter the rule into the calculator.
Nspire: Menu,
3:Graph Entry/Edit,
6:Sequence,
1:Sequence
ClassPad Menu,
Sequence,
Explicit



$u1(n)=2\cdot 3^{n-1}$
Initial Terms:=
$1\le n\le 99$ nstep=1

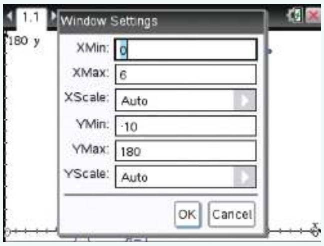


☑ $a_nE=2\cdot3^{n-1}$
☐ b_nE: □
☐ c_nE: □

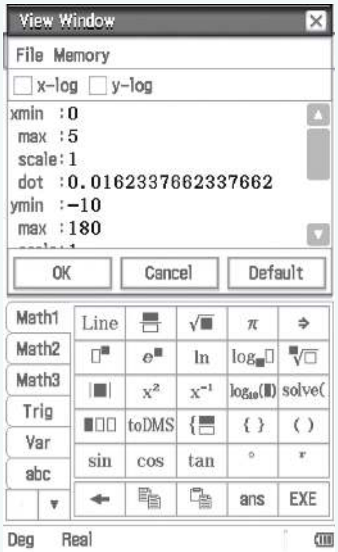
n	a_nE
1	2
2	6
3	18
4	54
5	162
6	486

1

2 Set the window to fit the graph.
This sequence increases because $r > 0$.
The first term ($n = 1$) is 2.
The fifth term ($n = 5$) is 162.
Choose window settings to fit.

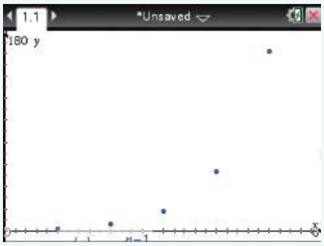


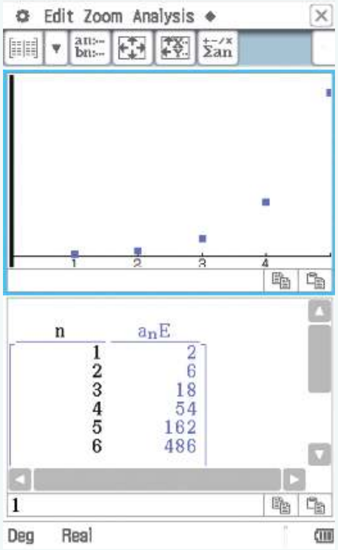
XMin: 0
XMax: 6
XScale: Auto
YMin: -10
YMax: 180
YScale: Auto



☐ x-log ☐ y-log
xmin :0
max :5
scale:1
dot :0.0162337662337662
ymin :-10
max :180

3 Sketch the graph.





n	anE
1	2
2	6
3	18
4	54
5	162
6	486

Exercise 4F

Notation: working out *a* and *r*

1 Give the value of the first term, a, and common ratio, r, in each of the following geometric sequences.

a 12, 24, 48, . . .
b 6, 18, 54, . . .
c 2, 8, 32, . . .
d 56, 28, 14, . . .
e 36, 12, 4, . . .
f 8, 56, 392, . . .
g 1, 10, 100, . . .
h 100, 10, 1, . . .
i 17, 221, 2873, . . .

2 State the values of the first term, a, and the common ratio, r, for the geometric sequences with the rules given.

a $T_n = 3 \times 4^{n-1}$
b $T_n = 5 \times 2^{n-1}$
c $T_n = 8 \times 7^{n-1}$
d $T_n = 200 \times 1.1^{n-1}$
e $T_n = 600 \times 0.9^{n-1}$
f $T_n = 3 \times 2^{n-1}$

Writing the rule for the *n*th term of a geometric sequence

Example 21

3 Give the rule for T_n for each of these geometric sequences.

a 9, 18, 36, 72, . . .
b 54, 18, 6, 2, . . .
c 4, 20, 100, 500, . . .
d 6, 42, 294, 2058, . . .
e 5, 20, 80, 320, . . .
f 8, 24, 72, 216, . . .

ISBN 978-1-009-25917-0

Finding the *n*th term of a geometric sequence

Example 22

4 Find the tenth term, T_{10}, in each of these geometric sequences.

a 4, 12, 36, . . .
b 3, 6, 12, . . .
c 2, 6, 18, . . .
d 1, 4, 16, . . .
e 10, 30, 90, . . .
f 512, 256, 128, . . .

5 Find the required term in each of these geometric sequences.

a 64, 32, 16, . . .	Find T_7.
b 9, 18, 36, . . .	Find T_8.
c 1, 2, 4, . . .	Find T_{20}.
d 1, 3, 9, . . .	Find T_{13}.
e 729, 243, 81, . . .	Find T_8.
f 100 000,10 000,1000, . . .	Find T_{10}.

6 Find the 10th term, T_{10}, in the geometric sequence that starts at 6 and has a common ratio of 2.

7 The first term in a geometric sequence is 200 and the common ratio is 1.1. Find the 30th term, T_{30}, correct to two decimal places.

8 A sequence starts at 1000 and each term is multiplied by 0.9 to make the next term. Find the 50th term, T_{50}, correct to two decimal places.

Using a recurrence relation to generate and graph the terms of a geometric sequence

CAS

9 **a** Generate and graph the first five terms of the geometric sequence defined by the recurrence relation: $T_1 = 1000, \quad T_{n+1} = 1.1T_n$.

b Calculate the value of the 13th term in the sequence correct to two decimal places.

10 **a** Generate and graph the first five terms of the geometric sequence defined by the recurrence relation: $T_1 = 256, \quad T_{n+1} = 0.5T_n$.

b Calculate the value of the 10th term in the sequence.

11 **a** Generate and graph the first five terms of the geometric sequence defined by the recurrence relation: $T_1 = 10\,000, \quad T_{n+1} = 1.25T_n$. Give values to the nearest whole number.

b Calculate the value of the 25th term in the sequence.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Key ideas and chapter summary

Assignment

Sequence	A **sequence** is a list of numbers in a particular order.
Term	An individual number in a sequence is called a **term** of that sequence. Terms are identified using notations with subscripts, where T_n represents the nth term of the sequence.
Recursion	**Recursion** is the repetitive application of the same rule, many times.
Recurrence relation	A **recurrence relation** gives the information needed to make each new term in a sequence using the previous term(s). A more general form is: $T_1 = a,$ $T_{n+1} = r \times T_n + d$ (starting value) (rule) Example: $T_{n+1} = 5T_n + 2, T_1 = 4$ tells us to multiply each term by 5 then add 2 to make each new term. Start at 4. 4, 22, 112, 562, . . . A *first-order* recurrence relation uses one term to generate the term that follows it. A *second-order* recurrence relation uses two consecutive terms of a sequence to generate the next term.
Arithmetic sequence and common difference	In an **arithmetic sequence**, each new term is made by adding a fixed number, called the common difference, d, to the previous term. Example: 3, 5, 7, 9, . . . is made by adding 2 to each term. The common difference, d, is found by taking any term and subtracting its previous term, e.g. $T_2 - T_1$. In our example above, $d = 5 - 3 = 2$.
Recurrence relation for an arithmetic sequence	A **recurrence relation for an arithmetic sequence** has the form $T_1 = a,$ $T_{n+1} = T_n + d$ where d = common difference and a = first term. In our example: $T_1 = 3,$ $T_{n+1} = T_n + 2$ Rule for finding T_n, the nth term in an arithmetic sequence: $T_n = a + (n - 1)d$ To find T_{10} in our example: put $n = 10, a = 3, d = 2$ $T_{10} = 3 + (10 - 1) \times 2$ $= 3 + (9) \times 2 = 21$
Rule for the *n*th term of an arithmetic sequence	The **rule for the *n*th term of an arithmetic sequence** has the form $T_n = a + (n - 1)d$ where a is the first term and d is the common difference.

ISBN 978-1-009-25917-0

Graph of an arithmetic sequence

The graph of an arithmetic sequence:

- values lie along a straight line
- increasing values when $d > 0$ (positive slope)
- decreasing values when $d < 0$ (negative slope)

Geometric sequence and common ratio

In a **geometric sequence**, each term is made by multiplying the previous term by a fixed number, called the common ratio, r.

Example: 5, 20, 80, 320, … is made by multiplying each term by 4.

The common ratio, r, is found by dividing any term by its previous term, e.g. $\frac{T_2}{T_1}$. In our example: $r = \frac{20}{5} = 4$

Recurrence relation for a geometric sequence

Recurrence relation for a geometric sequence:

$$T_1 = a, \quad T_{n+1} = r \times T_n$$

where r = common ratio and a = first term.

In our example: $T_1 = 5, \quad T_{n+1} = 4T_n$

Rule for finding T_n, the nth term, in a geometric sequence:

$$T_n = ar^{n-1}$$

where a = first term and r = common ratio.

To find T_7 in our example: put $n = 7$, $a = 5$, $r = 4$

$$\begin{aligned} T_7 &= 5 \times (4)^{7-1} \\ &= 5 \times (4)^6 = 20\,480 \end{aligned}$$

Rule for the *n*th term of a geometric sequence

The **rule for the *n*th term of a geometric sequence** has the form

$$T_n = a \times r^{(n-1)}$$

where a is the first term and r is the common ratio.

Graph of a geometric sequence

The graph of a geometric sequence:

- values increase when $r > 1$
- values decrease towards zero when $0 < r < 1$

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Skills check

Having completed the current chapter you should be able to:

- identify a pattern in a sequence of numbers
- use the naming convention for terms of a sequence
- decide whether a sequence is arithmetic, geometric or neither
- give the first term, a, and find the common difference, d, of an arithmetic sequence
- give the first term, a, and find the common ratio, r, of a geometric sequence
- find the nth term of an arithmetic or geometric sequence when given a few terms in the sequence
- generate a sequence using its recurrence relation
- generate the terms of a sequence using a calculator
- identify and sketch graphs of arithmetic and geometric sequences, by hand and with CAS.

Short-answer questions

1 Determine the 4th term in the sequence: 1, 4, 7, 10, 13, . . .

2 Which of the following is an arithmetic sequence?

a 2, 4, 8, . . .
b 2, 6, 18, . . .
c 2, 4, 6, . . .
d 2, 3, 5, . . .
e 2, 4, 7, . . .

3 Determine the value of the common difference, d, of the sequence: 27, 19, 11, 3, . . .

4 Determine the value of the 15th term of the sequence: 63, 56, 49, 42, . . .

5 Determine the value of the 7th term for the recurrence relation $T_1 = 5, T_{n+1} = T_n + 6$.

6 Find the sum of the first 5 terms of the sequence: 7, 11, 15, . . .

7 Determine the value of the common ratio, r, of the sequence: 17, 221, 2873, . . .

8 Which of the following sequences is geometric?

A 2, 6, 10, 14, . . .
B 2, 6, 12, 24, . . .
C 54, 18, 6, 2, . . .
D 54, 27, 9, 3, . . .
E 1, 3, 9, 18, . . .

9 Write the recurrence relation for the sequence: 3, 6, 12, 24, . . .

10 Determine the value of T_{20} for the sequence: 3, 6, 12, 24, . . .

11 Write the first 5 terms generated by the sequence with a recurrence relation $T_1 = 5$, $T_{n+1} = T_n - 3$.

12 A sequence of numbers is generated by the recurrence relation $T_1 = 2$, $T_{n+1} = 3T_n + 6$. Determine the value of T_4 in this sequence.

13 A sequence of numbers is generated by the recurrence relation $T_1 = 5$, $T_{n+1} = 3T_n - 2$. Determine the value of T_3 in this sequence.

14 A sequence is generated from the recurrence relation $T_1 = 40$, $T_{n+1} = T_n - 16$. Write the rule for the *nth* term of this sequence.

15 Find T_{20}, the 20th term in the sequence: 7, 11, 15, 19, . . .

16 Find T_{10}, the 10th term in the sequence: 3, 6, 12, . . .

17 Find T_5, the 5th term, of the sequence generated by the recurrence relation, $T_1 = 7$, $T_{n+1} = 2T_n$.

Extended-response questions

1 Match the descriptions given in **a** to **d** with the graphs of the sequences shown.

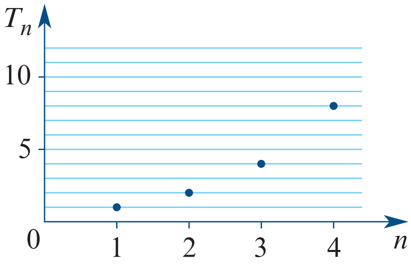
a Arithmetic with $d = 2$

b Arithmetic with $d = -2$

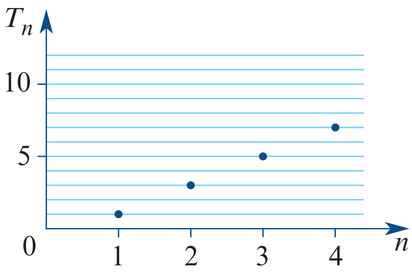
c Geometric with $r = 2$

d Geometric with $r = \frac{1}{2}$

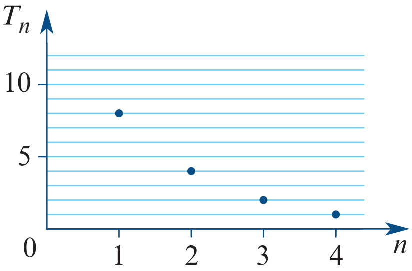
A



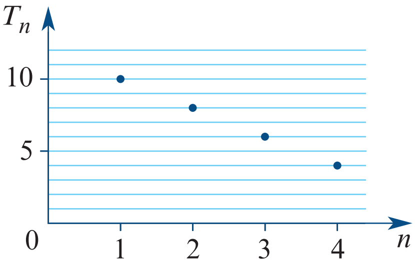
B



C



D



ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

2 A sequence is generated from the recurrence relation $T_1 = 800,\ T_{n+1} = 0.8T_n - 10$.

CAS

- **a** Use your calculator to generate the first five terms of the sequence.
- **b** Which term of the sequence will be the first negative term?

3 A sequence of numbers has starting value 30. Each term of the sequence is then generated using the rule 'subtract 3'. Let T_n be the value of the nth term.

- **a** Write a recurrence relation with $T_1 = 30$ that generates this sequence of numbers.
- **b** Sketch a graph of the value of the terms of the sequence against the value of n, the number of steps of the recurrence relation.
- **c** Which term of the sequence will be zero?

4 For each of the following arithmetic sequences, write:

- **i** the first term and common difference
- **ii** a recurrence relation that generates the terms of the sequence
- **iii** the rule for the nth term of the sequence
- **iv** the value of the 15th term

- **a** $2, 6, 10, \ldots$
- **b** $140, 131, 122, \ldots$
- **c** $1.5, 3.7, 5.9, \ldots$

5 For each of the following geometric sequences, write:

- **i** the first term and common ratio
- **ii** a recurrence relation that generates the terms of the sequence
- **iii** the rule for the nth term of the sequence
- **iv** the value of the 7th term

- **a** $3, 12, 48, \ldots$
- **b** $1400, 700, 350, \ldots$
- **c** $7200, 4320, 2592, \ldots$

5

Applications of sequences and recurrence relations

In this chapter

Syllabus references

Topics: The arithmetic sequence; The geometric sequence; Sequences generated by first-order linear recurrence relations

Subtopics: 3.2.4, 3.2.8 – 3.2.11

5A Applications of arithmetic sequences

Arithmetic sequences have applications to a variety of practical problems. Many everyday things can be **modelled** with arithmetic sequences, such as the cost of hiring a car, simple interest loans and investments, and the simple depreciation of assets.

General applications of arithmetic sequences

When arithmetic sequences are used to solve practical problems, the values of n, a and d have practical meanings connected with that problem.

For example, the cost of hiring a car is \$45 for the first day and \$35 for each extra day.
The cost for hiring the car for one day is \$45.
The cost for hiring the car for two days is \$45 + \$35 = \$80.
The cost for hiring the car for three days is an extra \$35, or \$115.
These costs form an arithmetic sequence: 45, 80, 115, . . .
The first term, a, is the cost for the first day, so $a = 45$.
The common difference, d, is the cost of every extra day, so $d = 35$.
The nth term represents the cost of hiring the car for n days.
The cost of hiring the car for n days can be found from the rule for the nth term of an arithmetic sequence.

Example 1 Application of an arithmetic sequence

The hire of a car costs \$180 for the first day and \$150 for each extra day.

a How much would it cost to hire the car for 7 days?

b Find a rule for the cost of hiring the car for n days.

Solution

Strategy: The cost of hiring the car for a given number of days follows an arithmetic sequence. The cost for the first day, \$180, is the first term. The common difference is \$150 because this is the amount added to find the cost for each extra day.

a **1** Identify values for a, n and d.

$a = 180, n = 7, d = 150$

2 Substitute the values for a, n and d into $T_n = a + (n - 1)d$.

$T_n = a + (n - 1)d$

$T_7 = 180 + (7 - 1)(150)$

3 Evaluate.

$= 180 + (6)(150)$

$= 1080$

4 Write your answer.

It would cost \$1080 to hire the car for 7 days.

b Substitute $a = 180$ and $d = 150$ into $T_n = a + (n - 1)d$.

$T_n = a + (n - 1)d$

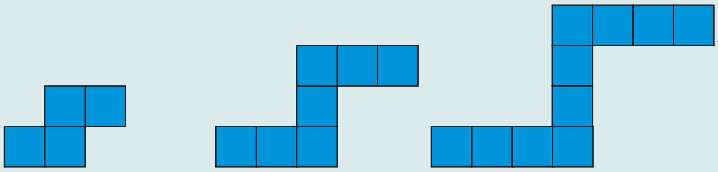
$T_n = 180 + (n - 1)(150)$

Note: This rule is saying that it costs \$180 for the first day and the additional $(n - 1)$ days each cost \$150.

ISBN 978-1-009-25917-0 Cambridge University Press Updated January 2024

Example 2 Application of arithmetic sequences using a recurrence relation and the rule for the *n*th term

The snake shapes below are made using blocks, each with a side length of 1 unit.



Snake 1 Snake 2 Snake 3

The perimeter of each snake shape can be found by counting the sides of the blocks around the outside of the shape. The perimeter of each new snake built is 6 units longer than the previous snake.

Let P_n be the perimeter of the snake shape n.

A recurrence relation that models this situation is:

$$P_1 = 10, \qquad P_{n+1} = P_n + 6$$

a **i** What does $P_1 = 10$ tell us?

ii What does the rule in the recurrence relation tell us?

b Use the recurrence relation to generate the perimeters of the first four snakes in this sequence and use these perimeters to construct a table showing the snake number (n) and its perimeter (P_n).

c Use the table to plot the snake perimeters against shape number and comment on the form of the graph.

d Use the rule for the nth term for this sequence to predict the perimeter of the 50th snake, in this sequence.

Solution

a **i** P_1 represents the perimeter of Snake 1.

i P_1 tells us that the perimeter of Snake 1 is ten units.

ii The rule is $P_{n+1} = P_n + 6$.

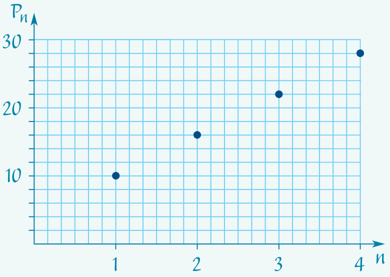
ii The rule tells us that the perimeter of each new snake in the sequence of snakes is 6 units longer than the perimeter of the previous snake.

b Starting with $P_1 = 10$, generate the next four perimeters using the recurrence rule.

$P_1 = 10$
$P_2 = P_1 + 6 = 10 + 6 = 16$
$P_3 = P_2 + 6 = 16 + 6 = 22$
$P_4 = P_3 + 6 = 22 + 6 = 28$

c Construct a table using the perimeters. Use the table to plot a graph with n on the horizontal axis and P_n on the vertical axis. Comment on the form of the graph.

Snake number	1	2	3	4
Perimeter	10	16	22	28



The plot is linear, indicating linear growth.

d Using the symbol P_n, rather than T_n, the rule for the nth term is $P_n = a + (n - 1)d$
Substitute $a = 10$, $n = 50$, $d = 6$ to obtain P_{50}.

$P_{50} = 10 + (50 - 1)6 = 304$
The perimeter of Snake 50 is 304 units.

Simple interest investments and loans

If you deposit money into a bank account, the bank is effectively borrowing money from you. The bank will pay you a fee for using your money and this fee is called **interest**. If a fixed amount of interest is paid into the account at regular time periods, it is called a **simple interest** investment.

If you borrow money from the bank and are charged a fixed amount of interest after regular time periods, it is called a *simple interest loan*.

Simple interest loans and investments have a starting value called the **principal**. The interest that is earned or charged is usually a percentage of this principal value and this is called the **interest rate**.

When we model simple interest loans and investments, it is usual to give the principal, or starting value, the symbol V_0. Note that the starting value has subscript zero for these models, not 1 as used previously. This represents the value of a loan or investment before any interest has been calculated.

In simple interest loans, it is usual to give the interest rate the symbol r. Take care not to confuse this with the common ratio used previously. The interest rate is a percentage per time period, usually one year or 'per annum' .

In a simple interest model, a principal of V_0, invested or borrowed at the rate of r% per time period, will have value V_n after n time periods.

Modelling simple interest loans and investments

V_0 = principal (amount borrowed or invested at simple interest rate r% per time period)

V_1 = value after 1 time period
V_2 = value after 2 time periods
V_n = value after n time periods

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Example 3 Determining the interest for a simple interest loan or investment

Cheryl invests $5000 with a bank and will be paid simple interest at the rate of 4.8% per annum.

a How much interest does Cheryl earn every year?

b How much will Cheryl have saved in the bank at the end of three years?

Solution

a **1** Write the starting value and the interest rate.

$V_0 = 5000$

$r = 4.8$

2 Use the interest rate and starting value to calculate the amount of interest.

$$\begin{aligned}\text{Interest} &= 4.8\% \text{ of } \$5000 \\ &= \frac{4.8}{100} \times \$5000 \\ &= \$240\end{aligned}$$

3 Write your answer.

Every year, $240 interest is earned.

b **1** Write the amount that is invested after each year.

Starting value $= V_0 = \$5000$

After 1 Year: $V_1 = \$5000 + \$240 = \$5240$

After 2 Year: $V_2 = \$5240 + \$240 = \$5480$

After 3 Year: $V_3 = \$5480 + \$240 = \$5720$

2 Write your answer.

After three years, Cheryl will have $5720 in the bank.

The calculations used to find the amount owed after three years are recursive calculations.

A fixed amount is added to the value of the loan or investment after each year and so simple interest loans and investments can be modelled using a recurrence relation similar to those that generate arithmetic sequences.

Recurrence relation model for simple interest investments and loans

Let V_n be the value of the loan or investment after n time periods.

Let r be the percentage interest rate, per time period.

The recurrence relation for the value of the loan or investment after n time periods is

$$V_0 = \text{principal}, \qquad V_{n+1} = V_n + D$$

where $D = \frac{r}{100} \times V_0$.

Example 4 Modelling a simple interest investment with a recurrence relation

Cheryl invests \$5000 in an investment account that pays 4.8% per annum simple interest.

Model this simple interest investment using a recurrence relation of the form:

$$V_0 = \text{the principal}, \qquad V_{n+1} = V_n + D \qquad \text{where} \qquad D = \frac{r}{100}V_0$$

Solution

1 Define the symbol V_n in this model.

V_n is the value of the investment after n years.

2 Write the value of V_0, the principal of the investment.

$V_0 = 5000$

3 Write the interest rate r and use it to determine the value of $D = \frac{r}{100}V_0$.

$r = 4.8$

$D = \frac{4.8}{100} \times 5000 = 240$

4 Use the values of V_0 and D to write the recurrence relation.

$V_0 = 5000, \qquad V_{n+1} = V_n + 240$

Once we have a recurrence relation, we can use it to determine things such as the value of the investment after a given number of years.

Example 5 Using a recurrence relation to analyse a simple interest investment

Cheryl's simple interest investment is modelled by

$$V_0 = 5000, \qquad V_{n+1} = V_n + 240$$

where V_n is the value of the investment after n years.

a Use the model to determine the value of Cheryl's investment after 3 years.

b When will Cheryl's investment first exceed \$6000, and what will its value be then?

Solution

a **i** Write the recurrence relation.

$V_0 = 5000, \qquad V_{n+1} = V_n + 240$

ii On a blank calculation screen, type **5000** and press enter (or EXE).

iii Type **+240** and press enter (or EXE) three times to obtain the value of Cheryl's investment after 3 years. Write your answer.

b **i** Continue pressing enter (or EXE) until the value of the investment first exceeds \$6000.

ii Write your answer.

```
5000              5000
5000 + 240        5240
5240 + 240        5480
5480 + 240        5720
5720 + 240        5960
5960 + 240        6200
```

a \$5720

b After 5 years; \$6200.

ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

Depreciation

Over time, the value of large items gradually decreases. A car bought new this year will not be worth the same amount of money in a few years' time. A new television bought for $2000 today is unlikely to be worth anywhere near this amount in 5 years.

Large equipment, machinery and other assets used in a business also lose value, or *depreciate*, over time. The depreciating value of equipment is often taken into account when calculating the actual costs of the business operations. It is important for businesses to be able to estimate the likely value of an asset after a certain amount of time. This is called the *future value* of the asset.

After a certain amount of time, or when the value of an item is depreciated to a certain amount, called its **scrap value**, the item will be sold or disposed of. At this point, the item has reached the end of its useful life and will be *written off*. This means the item is no longer an asset for the business.

Individuals often use the depreciation value of equipment to calculate tax refunds. The taxation law allows individuals to buy equipment, such as computers or other items, necessary for their work and then to claim tax refunds based on the depreciating value of those assets.

There are a number of techniques for estimating the future value of an asset. Two of them, **flat-rate depreciation** and **unit-cost depreciation**, can be modelled using a linear decay recurrence relation.

Flat-rate depreciation

Flat-rate depreciation is very similar to simple interest, but instead of adding a constant amount of interest, a constant amount is *subtracted* to decay the value of the asset after each time period. This constant amount is called the **depreciation** amount and, like simple interest, it is often given as a percentage of the initial purchase price of the asset.

Recurrence relation model for flat-rate depreciation

Let V_n be the value of the asset after n time periods.

Let r be the percentage depreciation rate, per time period.

The recurrence relation for the value of the asset after n time periods is

$$V_0 = \text{initial value of the asset}, \qquad V_{n+1} = V_n - D$$

where $D = \frac{r}{100} \times V_0$.

ISBN 978-1-009-25917-0

Example 6 Modelling a flat-rate depreciation with a recurrence relation

A new car was purchased for \$24 000 in 2014. The car depreciates by 20% of its purchase price each year. Model the depreciating value of this car using a recurrence relation of the form:

$$V_0 = \text{intial value}, \qquad V_{n+1} = V_n - D \qquad \text{where } D = \frac{r}{100}V_0$$

Solution

1 Define the symbol V_n in this model.	V_n is the value of the car after n years depreciation.
2 Write the value of V_0. Here, V_0 is the value of the car when new.	$V_0 = 24\,000$
3 Write the annual rate of depreciation, r, and use it to determine the value of $D = \frac{r}{100}V_0$.	$r = 20$ $D = \frac{20}{100} \times 24\,000 = 4800$
4 Use the values of V_0 and D to write the recurrence relation.	$V_0 = 24\,000, \quad V_{n+1} = V_n - 4800$

Once we have a recurrence relation, we can use it to determine things such as the value of an asset after a given number of years of flat-rate depreciation.

Example 7 Using a recurrence relation to analyse a flat-rate depreciation

The flat rate depreciation of a car is modelled by

$$V_0 = 24\,000, \qquad V_{n+1} = V_n - 4800$$

where V_n is the value of the car after n years.

a Use the model to determine the value of the car after 2 years.

b If the car was purchased in 2014, in what year will the car's value depreciate to zero?

Solution

a **i** Write the recurrence relation.

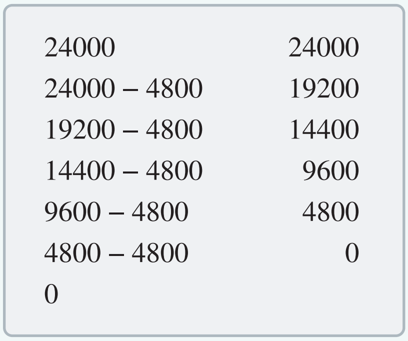
$V_0 = 24\,000, \quad V_{n+1} = V_n - 4800$

ii On a blank calculation screen, type **24 000** and press enter (or EXE).

iii Type **–4800** and press enter (or EXE) twice to obtain the value of the car after 2 years' depreciation. Write your answer.

b **i** Continue pressing enter (or EXE) until the car has no value.

ii Write your answer.



```
24000                24000
24000 – 4800         19200
19200 – 4800         14400
14400 – 4800          9600
9600 – 4800           4800
4800 – 4800              0
0
```

a \$14 400

b After 5 years depreciation

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Unit-cost depreciation

Some items lose value because of how often they are used, rather than because of their age. A photocopier that is 2 years old but has never been used could still be considered to be in 'brand new' condition and therefore worth the same, or close to, that it was 2 years ago. But if that photocopier was 2 years old and had printed many thousands of papers over those 2 years, it would be worth much less than its original value.

Cars can also depreciate according to their use rather than time. People often look at the number of kilometres a car has travelled before they consider buying it. An older car that has travelled few kilometres overall could be considered a better buy than a newer car that has travelled a large distance.

When the future value of an item is based upon use rather than age, we use a *unit-cost depreciation* method. Unit-cost depreciation can be modelled using a recurrence relation.

Recurrence relation model for unit-cost depreciation

Let V_n be the value of the asset after n units of use.

Let D be the cost per unit of use.

The recurrence relation for the value of the asset after n units of use is:

$$V_0 = \text{initial value of the asset}, \qquad V_{n+1} = V_n - D$$

Example 8 Modelling a unit-cost depreciation with a recurrence relation

A professional gardener purchased a lawn mower for \$270. The mower depreciates in value by \$3.50 each time it is used.

Model the depreciating value of this mower using a recurrence relation of the form:

$$V_0 = \text{intial value}, \quad V_{n+1} = V_n - D$$

where D = the depreciation in value per use

Solution

1 Define the symbol V_n in this model.	V_n is the value of the mower after being used to mow n lawns.
2 Write the value of V_0. Here, V_0 is the value of the mower when new.	$V_0 = 270$
3 Write the unit cost rate of depreciation, D.	$D = 3.50$
4 Write your answer.	$V_0 = 270$ $V_{n+1} = V_n - 3.50$

Once we have a recurrence relation, we can use it to determine things such as the value of an asset after a given number of years of unit-cost depreciation.

Example 9 Using a recurrence relation to analyse a unit-cost depreciation

The depreciated value of the lawn mower is modelled by

$$V_0 = 270, \qquad V_{n+1} = V_n - 3.50$$

where V_n is the value of the mower after being used to mow n lawns.

a Use the model to determine the value of the mower after it has been used three times.

b After how many uses would the value of the mower be less than $250?

Solution

a **i** Write the recurrence relation.

$V_0 = 270, \qquad V_{n+1} = V_n - 3.50$

ii On a blank calculation screen, type **270** and press `enter` (or `EXE`).

iii Type **–3.50** and press `enter` (or `EXE`) three times to obtain the value of the mower after three mows. Write your answer.

b **i** Continue pressing `enter` (or `EXE`) until the value of the lawn mower is first less than $250.

ii Write your answer.

```
270              270
270 – 3.5        266.5
266.5 – 3.5      263
263 – 3.5        259.5
259.5 – 3.5      256
256 – 3.5        252.5
252.5 – 3.5      249
```

a $259.50

b After six mows

Skill-sheet

Exercise 5A

Applications of arithmetic sequences

Example 1

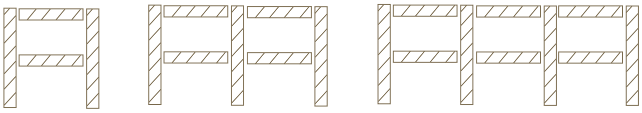
1 A canoe hire shop charges $15 for the first hour and $12 for each extra hour. How much would it cost to hire a canoe for 10 hours?

2 At the end of its first year after planting, a tree was 2.50 m high. It grew 0.75 m in each following year. How high was it 18 years after it was planted?

3 Bronwen swam her first race of 50 metres in 68.4 seconds. She hopes to reduce her time by 0.3 seconds each time she races. Give her times for the first four races if she succeeds.

4 Tristan had $250 on the first day of his holidays. If he spent $23 on each of the following days, how much did he have left after the 10th day of his holidays?

5 A single section of fencing is made from four logs. Two sections use seven logs. Examples of one, two and three-section fences are shown below.

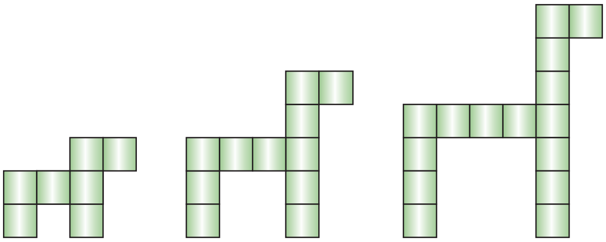


a How many logs are needed for a fence with three sections?

b How many logs would be needed to build a fence with 20 sections?

Applications of arithmetic sequence using a recurrence relation

Example 2 **6** The Llama shapes have been made using blocks.



Let b_n be the number of blocks used to make the nth Llama shape.
The number of blocks used to make each Llama shape is generated by the recurrence relation:

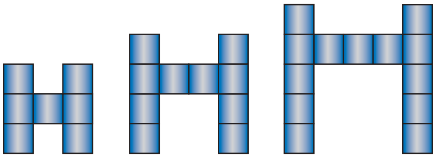
$$b_1 = 7, \qquad b_{n+1} = b_n + 4$$

a Count and record the number of blocks used to make the first, second and third Llama shapes.

b Use the recurrence relation for b_n to generate the first five terms of the sequence of perimeters for these shapes.

c Use a rule to calculate the number of blocks needed to make the Llama 10 shape.

7 The BBQ shapes have been made using blocks, each with a side length of 1 unit.



The perimeter of each BBQ shape can be found by counting the sides of the blocks around the outside of the shape. Let P_n be the perimeter of the nth BBQ shape.
The perimeters for each BBQ shape is generated by the recurrence relation:

$$P_1 = 16, \qquad P_{n+1} = P_n + 6$$

a Count and record the perimeters of the first, second and third BBQ shapes.

b Use the recurrence relation for P_n to generate the first four terms of this sequence of perimeters.

c Draw the fourth BBQ shape, find its perimeter and check if the recurrence relation correctly predicted the perimeter.

d Use the rule for the *n*th term of this sequence to predict the perimeter of the 10th BBQ shape.

Modelling a simple interest investment with recurrence relations

Example 4
Example 5

CAS

8 The following recurrence relation can be used to model a simple interest investment of \$2000 paying interest at the rate of 3.8% per annum.

$$V_0 = 2000, \qquad V_{n+1} = V_n + 76$$

In the recurrence relation, V_n is the value of the investment after n years.

a Use the recurrence relation to find the value of the investment after 1, 2 and 3 years.

b Use your calculator to determine how many years it takes for the value of the investment to be more than \$3000.

c Write a recurrence relation model if \$1500 was invested at the rate of 6.0% per annum.

9 The following recurrence relation can be used to model a simple interest loan of \$7000 charged interest at the rate of 7.4% per annum.

$$V_0 = 7000, \qquad V_{n+1} = V_n + 518$$

In the recurrence relation, V_n is the value of the loan after n years.

a Use the recurrence relation to find the value of the loan after 1, 2 and 3 years.

b Use your calculator to determine how many years it takes for the value of the loan to be more than \$10 000.

c Write a recurrence relation model if \$12 000 was borrowed at the rate of 8.2% per annum.

10 The following recurrence relation can be used to model a simple interest investment. In the recurrence relation, V_n is the value of the investment after n years.

$$V_0 = 15\,000, \qquad V_{n+1} = V_n + 525$$

a **i** What is the principal of this investment?
   ii How much interest is paid every year?
   iii What is the annual interest rate of this investment?

b Use your calculator to determine how many years it takes for the value of the investment to be more than double the principal.

Modelling flat-rate depreciation with recurrence relations

Example 6
Example 7

CAS

11 The following recurrence relation can be used to model the depreciation of a computer with purchase price \$2500 and annual depreciation of \$400.

$$V_0 = 2500, \qquad V_{n+1} = V_n - 400$$

In the recurrence relation, V_n is the value of the computer after n years.

a Use the recurrence relation to find the value of the computer after 1, 2 and 3 years.

b Use your calculator to determine how many years it takes for the value of the computer to be less than \$1000.

c Write a recurrence relation model if the computer was purchased for \$1800 and depreciated at \$350 per annum.

12 The following recurrence relation can be used to model the depreciation of a car purchased for \$23 000 and depreciated at 3.5% of its original value each year.

$$V_0 = 23\,000, \qquad V_{n+1} = V_n - 805$$

In the recurrence relation, V_n is the value of the car after n years.

a Use the recurrence relation to find the value of the car after 1, 2 and 3 years.

b Use your calculator to determine how many years it takes for the value of the car to be less than \$10 000.

c Write a recurrence relation model for a car purchased for \$37 000 and depreciated at \$700 per annum.

d Write a recurrence relation model for a car purchased for \$12 000 and depreciated at 4.5% of its purchase price per annum.

13 The following recurrence relation can be used to model the depreciation of a television. In the recurrence relation, V_n is the value of the television after n years.

$$V_0 = 1500, \qquad V_{n+1} = V_n - 102$$

a **i** What is the purchase price of this television?

ii What is the depreciation of the television each year?

iii What is the annual percentage depreciation of the television?

b Use your calculator to determine a selling price if the television is sold after 8 years.

Modelling unit-cost depreciation with recurrence relations

Example 8
Example 9

14 The following recurrence relation can be used to model the depreciation of a printer with purchase price \$450 and depreciated by 5 cents for every page printed.

$$V_0 = 450, \qquad V_{n+1} = V_n - 0.05$$

In the recurrence relation, V_n is the value of the printer after n pages printed.

a Use your calculator to find the value of the printer after 20 pages.

b Write a recurrence relation model if the printer was purchased for \$300 and depreciated at 8 cents per page printed.

ISBN 978-1-009-25917-0

15 The following recurrence relation can be used to model the depreciation of a delivery van with purchase price \$48 000 and depreciated by \$200 for every 1000 kilometres travelled.

$$V_0 = 48\,000, \qquad V_{n+1} = V_n - 200$$

In the recurrence relation, V_n is the value of the delivery van after n lots of 1000 kilometres travelled.

a Use the recurrence relation to find the value of the van after 1000, 2000 and 3000 kilometres.

b Use your calculator to determine the value of the van after 15 000 kilometres.

c Use your calculator to determine how many kilometres it takes for the value of the van to reach \$43 000.

CAS

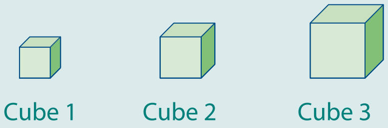
5B Applications of geometric sequences

Just like arithmetic sequences, geometric sequences can be applied to a variety of practical situations, such as population change and compound interest loans and investments.

Example 10 Application of geometric sequences using a recurrence relation and the rule for the *n*th term

The volume of cube 1 is 8 cm^3.

The volume of each successive cube is 1.5 times the volume of the previous cube.



Let c_n be the volume (in cm^3) of the nth cube in this sequence of cubes.

A recurrence relation that can be used to generate the volumes of this sequence of cubes is:

$$c_1 = 8\text{ cm}^3, \qquad c_{n+1} = 1.5c_n$$

a Use the recurrence relation to generate the volumes of the first four cubes in this sequence and use these volumes to construct a table showing the cube number (n) and its volume (c_n).

b Use the table to plot the volume of the cube against cube number and comment on the form of the graph.

c Use the rule for the nth term for this sequence to predict the volume of the 10th cube in this sequence.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Solution

a Starting with $c_1 = 8$, generate the next three volumes using the recurrence rule:

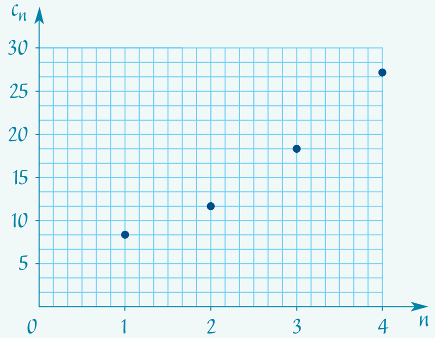
$$c_{n+1} = 1.5 \times c_n$$

Construct a table using these volumes.

$c_{n+1} = 1.5 \times c_n$
$c_1 = 8$
$c_2 = 1.5 \times c_1 = 1.5 \times 8 = 12$
$c_3 = 1.5 \times c_2 = 1.5 \times 12 = 18$
$c_4 = 1.5 \times c_3 = 1.5 \times 18 = 27$

Cube number	1	2	3	4
Volume	8	12	18	27

b Use the table to plot a graph with n on the horizontal axis and c_n on the vertical axis.



The plot is non-linear and curves upwards.

c Using the symbol c_n rather than T_n, the rule for the nth term is $c_n = ar^{n-1}$.
Substitute $a = 8, n = 10, r = 1.5$ to obtain c_{10}.

$c_{10} = 8 \times 1.5^{(10-1)} = 308\text{ cm}^3$ to the nearest cm^3.
The volume of cube 10 is 308 cm^3.

Percentage change

Let's investigate the sequence generated when each term is 10% less than the previous term. This sequence would arise if someone had \$100 and decided that on each new day they would spend 10% of whatever money they had left.

They start with \$100.	$T_1 = 100$
On the second day they spent 10% of \$100, which is \$10, leaving \$90.	$T_2 = 100 - 10\%$ of 100 $= 100 - 10$
The second term is \$90.	$= 90$
On the third day they spent 10% of \$90, which is \$9, leaving \$81.	$T_3 = 90 - 10\%$ of 90 $= 90 - 9$
The third term is \$81.	$= 81$

The sequence is 100, 90, 81, . . .

To find the common ratio, use $\frac{T_2}{T_1}$

$$\text{common ratio} = \frac{T_2}{T_1} = \frac{90}{100} = 0.90$$

We can make more terms in the sequence by multiplying by 0.90 to make each new term.

$$100 \xrightarrow{\times 0.90} 90 \xrightarrow{\times 0.90} 81 \xrightarrow{\times 0.90} 72.90 \xrightarrow{\times 0.90} 65.61\ldots$$

So on the fifth day the person had \$65.61 left.

Note: A 10% reduction means that there will only be 90% left. In other words, each new term will be 90% of the current term. So the common ratio $r = 90\% = 0.90$

In general, an r% reduction means that each new term will be $(100\% - r\%)$ or $1 - \frac{r}{100}$ of the current term.

Geometric sequences that involve percentage change have financial applications and also apply to situations of population change. We will use slightly different symbols to refer to the sequence elements. For general geometric sequences:

- The common ratio will be called R
- The percentage change will be called r.

It is important not to confuse the use of r, the percentage change, with the common ratio.

Percentage change

Let V_0 be the starting term of a sequence.
Let r be the percentage change from one term to the next.
Let R be the multiplying factor that generates the sequence (common ratio)

- For percentage *increase*: $R = 1 + \frac{r}{100}$
- For percentage *decrease*: $R = 1 - \frac{r}{100}$

Note: For general geometric sequences, R will take the place of the common ratio. It is important not to confuse the use of r, the percentage change, with the common ratio.

Example 11 Calculating the common ratio from percentage change

State, correct to two decimal places, the first four terms in each of these geometric sequences for the changes given.

a Starts at 200 and each new term is 4% *less* than the previous term.

b Starts at 500 and each new term is 12% *more* than the previous term.

Solution

a **1** The first term is 200.

$a = 200$

2 Use $R = 1 - \frac{r}{100}$ with $r = 4$ *or* as each new term is 4% *less* than the previous term, the new term is 96% of the previous term.
So $R = 96\% = 0.96$.

$R = 1 - \frac{4}{100} = 0.96$
or
$R = 96\% = 0.96$

3 Starting at 200, multiply by 0.96 to make each new term.

4 Write your answer correct to two decimal places.

The first four terms are:
200, 192, 184.32, 176.95.

b **1** The first term is 500.

$a = 500$

2 Use $R = 1 + \frac{r}{100}$ with $r = 12$ *or* as each new term is 12% *more* than the previous term, the new term is 112% of the previous term.
So $R = 112\% = 1.12$.

$R = 1 + \frac{12}{100} = 1.12$
or
$R = 112\% = 1.12$

3 Starting at 500, multiply by 1.12 to make each new term.

4 Write your answer correct to two decimal places.

The first four terms are:
500, 560, 627.2, 702.46.

Compound interest investments and loans

Most interest calculations are not as straightforward as simple interest. The more usual form of interest is **compound interest** where any interest that is earned after one time period is added to the principal and then contributes to the earning of interest in the next time period. The interest is said to *compound* after every time period.

This means that the value of the investment grows in ever increasing amounts, or grows geometrically, instead of by the same amount as in simple interest.

Consider an investment of $5000 that pays 8% interest per annum, compounding yearly. This means that the investment's value increases by 8% each year.

ISBN 978-1-009-25917-0

We can model the investment with a recurrence relation as follows:

Let V_n be the value of the investment after n years.

We can then write:

$$V_0 = \$5000$$

Each year, the investment will increase in value by 8%, so:

value next year = this year's value + interest earned

or symbolically:

$$V_{n+1} = V_n + 0.08 \times V_n$$

or more compactly:

$$V_{n+1} = 1.08V_n$$

We now have a recurrence rule that we can use to model and investigate the growth of investment over time.

Recurrence relation model for compound interest investments and loans

Let V_n be the value of the investment or loan after n compounding periods.

Let r be the percentage interest per compounding period.

The recurrence relation for the value of the investment or loan after n compounding periods is:

$$V_0 = \text{principal}, \qquad V_{n+1} = RV_n$$

where $R = 1 + \dfrac{r}{100}$

When using a recurrence relation to model compound interest investments and loans, it is important to take care with the rounding of any terms that are generated.

In general, rounding should only occur when the required term is calculated and written down as the answer to a question. Usually, this answer represents an amount of money so it is appropriate to round the value to two decimal places, or the nearest cent, as this is the smallest unit of currency that can actually be paid.

A rounded term should never be used to calculate further terms in the recurrence relation. Always use the decimal values stored in your calculator.

Example 12 Modelling a compound interest investment with a recurrence relation

The following recurrence relation can be used to model a compound interest investment of \$2000 paying interest at the rate of 7.5% per annum.

$$V_0 = 2000, \qquad V_{n+1} = 1.075 \times V_n$$

In the recurrence relation, V_n is the value of the investment after n years.

a Use the recurrence relation to find the value of the investment after:

 i 3 years **ii** 4 years

b Determine when the value of the investment will first exceed \$3000.

c Write the recurrence relation if \$1500 was invested at a compound interest rate of 6.0% per annum.

Solution

a i 1 Write the principal of the investment, V_0.

$V_0 = 2000$

2 Use the recurrence relation to calculate V_1, V_2 and V_3. Use your calculator if you wish.

$V_1 = 1.075 \times 2000 = 2150$

$V_2 = 1.075 \times 2150 = 2311.25$

$V_3 = 1.075 \times 2311.25 = 2484.593\,75$

3 V_3 represents the amount in the investment after 3 years. Round this answer to two decimal places.

After three years, the investment has value \$2484.59

ii 1 Use the recurrence relation to calculate V_4. Take care to use the unrounded value of V_3. This should be stored in your calculator.

$V_4 = 1.075 \times 2484.593\,75 = 2670.938\,281\,25$

2 V_4 represents the amount in the investment after 4 years. Round this answer to two decimal places.

After four years, the investment has value \$2670.94

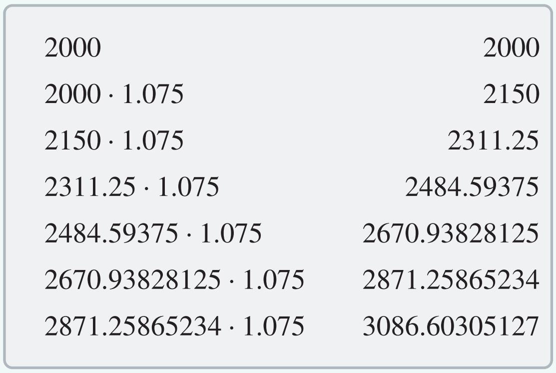
Note: If the rounded value of V_3 is used to calculate the value of V_4 the incorrect answer of \$2670.93 will be obtained.

b Steps

1 Type '2000' and press enter (or EXE).

2 Type **×1.075**.

3 Count how many times you press enter (or EXE) until the term value is greater than 3000.



```
2000                                 2000
2000 · 1.075                         2150
2150 · 1.075                      2311.25
2311.25 · 1.075                2484.59375
2484.59375 · 1.075          2670.93828125
2670.93828125 · 1.075       2871.25865234
2871.25865234 · 1.075       3086.60305127
```

4 Write your answer.

After 6 years, the investment will first exceed \$3000.

c **1** Identify the value of V_0 and r. — $V_0 = 1500$ and $r = 6$

2 Calculate the value or R. — The value of the investment will grow over time, so $R = 1 + \frac{r}{100} = 1 + \frac{6}{100}$

$R = 1.06$

3 Write your answer. — $V_0 = 1500, \quad V_{n+1} = 1.06 \times V_n$

Converting interest rates

Interest rates for one compounding time period can easily be converted to interest rates for other compounding time periods, by adjusting them with some simple arithmetic.

Converting interest rates

Assume that, in one year, there are:

- 365 days (ignore the possibility of leap years)
- 52 weeks (even though there are slightly more than this)
- 26 fortnights (even though there are slightly more than this)
- 12 months
- 4 quarters.

Convert an annual interest rate to another time period interest rate by dividing by these numbers.

Example 13 Compound interest with different compounding periods

Brian borrows \$5000 from a bank. He will pay interest at the rate of 4.5% per annum.

Let V_n be the value of the loan after n compounding periods.

Write a recurrence relation to model the value of Brian's loan if interest is compounded:

a yearly **b** quarterly **c** monthly

Solution

a **1** Define the variable V_n. The compounding period is *yearly*. — Let V_n be the value of Brian's loan after n years.

2 Determine the value of R. — The interest rate is 4.5% per annum.

$R = 1 + \frac{4.5}{100} = 1.045$

3 Write the recurrence relation. — $V_0 = 5000, \quad V_{n+1} = 1.045 \times V_n$

b	**1**	Define the variable V_n. The compounding period is *quarterly*.	Let V_n be the value of Brian's loan after n quarters.
	2	Determine the value of R.	The interest rate is 4.5% per annum. The quarterly rate is $\frac{4.5}{4} = 1.125$. $R = 1 + \frac{1.125}{100} = 1.011\,25$
	3	Write recurrence relation.	$V_0 = 5000, \; V_{n+1} = 1.011\,25 \times V_n$
c	**1**	Define the variable V_n. The compounding period is *monthly*.	Let V_n be the value of Brian's loan after n months.
	2	Determine the value of R.	The interest rate is 4.5% per annum. The monthly rate is $\frac{4.5}{12} = 0.375$. $R = 1 + \frac{0.375}{100} = 1.003\,75$
	3	Write the recurrence relation.	$V_0 = 5000, \qquad V_{n+1} = 1.003\,75 \times V_n$

Reducing-balance depreciation

Earlier in the chapter, we studied two different methods for depreciating the value of an asset, both of which were examples of linear decrease. **Reducing-balance depreciation** is another method of depreciation – one where the value of an asset decreases geometrically. Each year, the value will be reduced by a percentage, $r\%$, of the previous year's value. The calculations are very similar to compounding interest, but with a decrease in value, rather than an increase.

Recurrence relation model for reducing-balance depreciation

Let V_n be the value of the asset after n years.

Let r be the annual percentage depreciation.

The recurrence relation for the value of the asset after n years is:

$$V_0 = \text{initial value}, \qquad V_{n+1} = RV_n$$

where $R = 1 - \frac{r}{100}$

Example 14 Modelling a reducing-balance depreciation with a recurrence relation

The following recurrence relation can be used to model the value of office furniture with a purchase price of \$6900, depreciating at a reducing-balance rate of 7% per annum.

$$V_0 = 6900, \qquad V_{n+1} = 0.93 \times V_n$$

In the recurrence relation, V_n is the value of the office furniture after n years.

a Use the recurrence relation to find the value of the office furniture, correct to the nearest cent, after 1, 2 and 3 years.

b Determine when the value of the investment will first be less than \$5000.

c Write the recurrence relation if the furniture was initially valued at \$7500 and is depreciating at a reducing-balance rate of 8.4% per annum.

Solution

a **1** Write the purchase price of the furniture, V_0.

$V_0 = 6900$

2 Use the recurrence relation to calculate V_1, V_2 and V_3. Use your calculator if you wish.

$V_1 = 0.93 \times 6900 = 6417$

$V_2 = 0.93 \times 6417 = 5967.81$

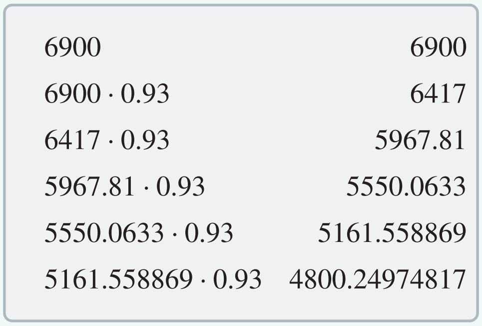
$V_3 = 0.93 \times 5967.81 = 5550.06$

b **Steps**

1 Type **6900** and press enter or EXE.

2 Type **× 0.93**.

3 Count how many times you press enter or EXE until the term value is less than 5000.



```
6900                                6900
6900 · 0.93                         6417
6417 · 0.93                      5967.81
5967.81 · 0.93                 5550.0633
5550.0633 · 0.93             5161.558869
5161.558869 · 0.93        4800.24974817
```

4 Write your answer.

The value of the furniture drops below \$5000 after 5 years.

c To write the recurrence relation:

1 Identify the value of V_0.

$V_0 = 7500$

2 Calculate the value of R.

The depreciation rate is 8.4% per annum.

$R = 1 - \frac{8.4}{100}$ or $R = 0.916$

3 Write your answer.

$V_0 = 7500, \qquad V_{n+1} = 0.916 \times V_n$

ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

Exercise 5B

Applications of geometric sequences

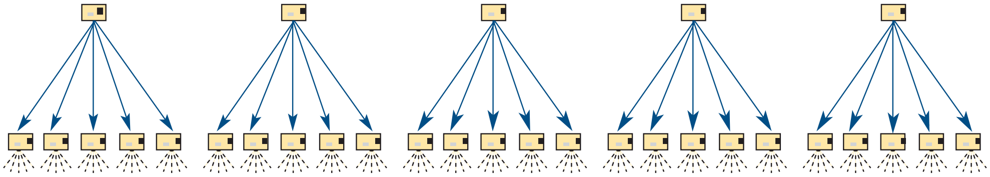
1 The zoom feature on a photocopy machine was set to 300%, making the new image 3 times larger than the original. A small photo, only 1 cm wide, of a person's face was photocopied. The new image was then photocopied to make a further enlargement, and the process was repeated.
Start with the original width and then list the widths of the next three images produced.

2 A gambler's strategy is to place a first bet of \$50 and to double the bet each time he loses. His first three bets are all losses.

a Write out the sequence of the bets he has made.

b What is the total of all his losses?

3 On the first day of the month, Dorian posted five letters to friends. Each friend was told to post 5 letters to their friends, with instructions to repeat the process.

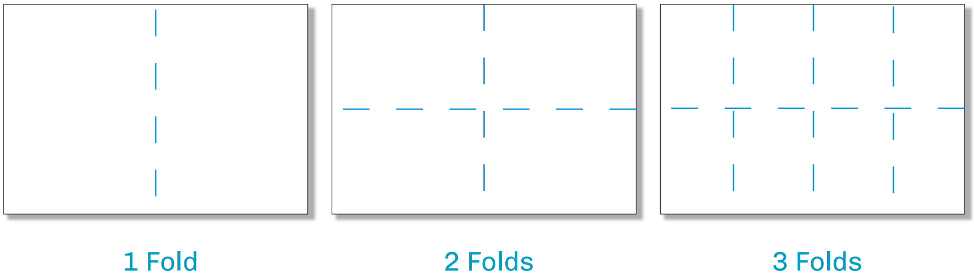


Assume that each letter posted is delivered the next day, and that each recipient posts their five letters immediately.

a Write the sequence for the number of letters posted on each of the first 4 days.

b What was the total number of letters posted over all of the 4 days?

4 A sheet of paper is in the shape of a rectangle. When the sheet is folded once and opened, 2 rectangles are formed either side of the crease. When a sheet is folded twice and opened, 4 rectangles are created, and so on.



Let f_n be the number of rectangles created by n folds.
The sequence for the number of rectangles created is generated by the recurrence relation:

$$f_1 = 2, \qquad f_{n+1} = 2f_n$$

ISBN 978-1-009-25917-0

a Use the recurrence relation for f_n to generate the first four terms of the sequence.

b Count and record the number of rectangles created by the first, second and third folds. Then fold a sheet of A4 paper four times to check that the recurrence relation correctly predicted the number of rectangles created.

c Use the rule for the *n*th term to calculate the number of rectangles after 10 folds.

d Using your calculator, generate the terms of the sequence to check your answer to **c**.

Percentage change and applications

Example 11

5 State, correct to two decimal places, the first four terms in each of these geometric sequences for the changes given.

a Starts at 100 and decreases by 5%

b Starts at 100 and increases by 20%

c Starts at 5000 and increases by 3%

d Starts at 7000 and decreases by 4%

6 Tom purchased his car for \$30 000 and expects it to depreciate in value by 15% each year. What will be the value of his car after 12 years?

7 Kathy invested \$60 000 so that it grows by 8% each year. What will her investment be worth after seven years?

Modelling a compound interest investment with recurrence relations

Example 12

CAS

8 An investment of \$6000 earns compounding interest at the rate of 4.2% per annum. A recurrence relation that can be used to model the value of the investment after *n* years is shown below.

$$V_0 = 6000, \qquad V_{n+1} = 1.042V_n$$

In the recurrence relation, V_n is the value of the investment after n years.

a Use the recurrence relation to find the value of the investment after 1, 2 and 3 years. Round your answers but use full decimals in calculations.

b Determine how many years it takes for the value of the investment to first exceed \$8000.

c Write a recurrence relation model for the value of an investment of \$5000 at a compounding interest rate of 6.8% per annum.

9 A loan of \$20 000 is charged compounding interest at the rate of 6.3% per annum. Round your answers but use full decimals in calculations. A recurrence relation that can be used to model the value of the loan after *n* years is shown below.

$$V_0 = 20\,000, \qquad V_{n+1} = 1.063V_n$$

In the recurrence relation, V_n is the value of the loan after n years.

a Use the recurrence relation to find the value of the loan after 1, 2 and 3 years.

b Determine how many years it takes for the value of the loan to first exceed \$30 000.

c Write a recurrence relation model for the value of a loan of \$18 000 at a compounding interest rate of 9.4% per annum.

Example 13 **10** Wayne invests \$7600 with a bank. He will be paid interest at the rate of 6% per annum, compounding monthly. Let V_n be the value of the investment after n months.

CAS

- **a** Write a recurrence relation to model Wayne's investment.
- **b** How much is Wayne's investment worth after 5 months?

11 Jessica borrows \$3500 from a bank. She will be charged compound interest at the rate of 8% per annum, compounding quarterly. Let V_n be the value of the loan after n quarters.

- **a** Write a recurrence relation to model the value of Jessica's loan from quarter to quarter.
- **b** If Jessica pays back everything she owes to the bank after 1 year, how much money will she need?

Modelling reducing-balance depreciation with recurrence relations

12 A motorcycle, purchased new for \$9800, will be depreciated using a reducing-balance depreciation method with an annual depreciation rate of 3.5%. Let V_n be the value of the motorcycle after n years.

- **a** Write a recurrence relation to model the value of the motorcycle from year to year.
- **b** Generate a sequence of numbers that represent the value of the motorcycle from year to year for 5 years in total. Write the values of the terms of the sequence correct to the nearest cent.
- **c** What is the value of the motorcycle after 5 years?
- **d** What is the depreciation of the motorcycle in the third year?

13 Office furniture was purchased new for \$18 000. It will be depreciated using a reducing-balance depreciation method with an annual depreciation rate of 4.5%. Let V_n be the value of the furniture after n years.

- **a** Write a recurrence relation to model the value of the furniture from year to year.
- **b** Generate a sequence of numbers that represent the value of the furniture from year to year for 5 years in total. Write the values of the terms of the sequence correct to the nearest cent.
- **c** What is the value of the furniture after 3 years?
- **d** What is the total depreciation of the furniture after 5 years?

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

5C First order linear relations

Growth and decay

If the successive terms of a sequence increase in size, the sequence shows a *growth* pattern.

If the successive terms of a sequence decrease in size, the sequence shows a *decay* pattern.

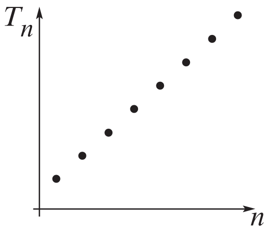
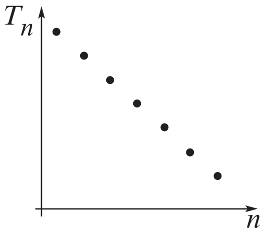
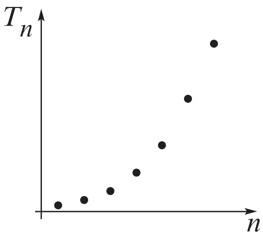
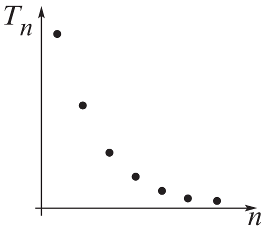
Arithmetic sequences grow or decay by a fixed amount called the common difference and they grow or decay *linearly*.

Geometric sequences grow or decay by a fixed ratio called the common ratio and they grow or decay *geometrically*.

The terms of an arithmetic sequence show **linear growth** if the common difference is greater than zero and show **linear decay** if the common difference is less than zero.

The terms of a geometric sequence show **geometric growth** if the common ratio is greater than one and show **geometric decay** if the common ratio is between zero and one.

Information about arithmetic and geometric sequences is summarised in the table below.

	Arithmetic sequence	*Geometric sequence*
	a = first term d = common difference	a = first term r = common ratio
Recurrence relation	$T_1 = a, \quad T_{n+1} = T_n + d$	$T_1 = a, \quad T_{n+1} = r \times T_n$
Rule for nth term	$T_n = a + (n - 1) \times d$	$T_n = a \times r^{(n-1)}$
Graphs of growth or decay	$d > 0$  Linear growth $d < 0$  Linear decay	$r > 1$  Geometric growth $0 < r < 1$  Geometric decay

Example 15 Describing the change in the terms of a sequence

Describe the nature (arithmetic, geometric) and direction (growth, decay) of each of the following sequences.

a 4, 8, 16, . . .

b $T_1 = 500, \quad T_{n+1} = 0.5 \times T_n$

c $T_n = 4 + (n - 1) \times 5$

Solution

a	**1**	Observe the change in the terms.	The terms of this sequence are increasing, so the change is one of growth.
	2	Decide on the nature of change.	difference: $d = 8 - 4 = 2$ $d = 16 - 8 = 8$ ratio: $r = 8 \div 4 = 2$ $r = 16 \div 8 = 2$ The difference is not common, but the ratio is common. The sequence is geometric.
	3	Write your answer.	This sequence shows geometric growth.
b	**1**	Identify the type of recurrence relation.	This recurrence relation involves multiplying a term by a number to generate the next term. The change is geometric.
	2	Identify the common ratio.	From the recurrence relation, the common ratio is $r = 0.5$
	3	Decide on the nature of the change.	Since r is between zero and one, the change is decay.
	4	Write your answer.	This sequence shows geometric decay.
c	**1**	Identify the type of sequence.	This is the general term of an arithmetic sequence.
	2	Identify the common difference.	From the rule for the nth term, the common difference is $d = 5$.
	3	Decide on the nature of the change.	Since d is greater than zero, the change is growth.
	4	Write your answer.	This sequence shows arithmetic growth.

First-order linear relations

So far, all of the practical applications of sequences have involved either arithmetic or geometric growth or decay only. There are many practical situations that involve aspects of both types of change – both arithmetic and geometric – and these are called **first-order linear recurrence relations**. These sequence are neither arithmetic nor geometric, but they do show general growth or decay.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

One example of a first-order linear relation is the change in the size of animal populations. Over time, animals reproduce and so the population grows in one sense, but at the same time, some animals will leave the population because they die or migrate to another area, meaning the population will also decay by a certain amount.

Example 16 Calculating an amount with combined geometric and arithmetic change

A trout farm currently has 10 000 trout in a pond.
In one month, breeding has caused the trout population to increase by 10%.
After one month, 3000 trout are caught and sold.
How many trout will be in the pond after one month?

Solution

1 The trout population increases by 10%. Calculate the multiplying ratio, R.	$R = 1 + \frac{10}{100} = 1.1$
2 Calculate the population after increase.	$\text{population} = 1.1 \times 10\,000$ $= 11\,000$
3 Subtract the number of trout caught and sold.	$\text{population} = 11\,000 - 3000$ $= 8000$
4 Write your answer.	After one month, there will be 8000 trout in the pond.

Modelling first-order linear recurrence relations

Practical situations, like the trout population change in the previous example, are often modelled with a *starting value*, or *current state*. The starting value of the trout population was 10 000.

The starting value can be represented by a capital letter, usually related to the quantity being modelled. A subscript of ‘0’ indicates a starting value.

The starting value of the trout population can be represented by P_0.

Every month, the population grows by 10%. This is geometric growth with $r = 1.1$, but in combined modelling, we will use a capital R. So, $R = 1.1$ will represent the geometric increase in the population.

Every month, the population decreases by 3000 trout. This is arithmetic decay with $d = -3000$, but in combined modelling, we will use a capital D. So, $D = -3000$ will represent the arithmetic decrease in the population.

A recurrence relation that models the population of trout in the pond can now be written:

$$P_0 = 10\,000, \quad P_{n+1} = 1.1 \times P_n - 3000$$

where P_n is the number of trout in the pond after n months.

Model for first-order linear recurrence relations

Let V_n = the value of a combined arithmetic and geometric sequence after n time periods.

Let V_0 = the starting value of the sequence.

Let r = the percentage change per time period.

Let D = the amount of arithmetic change.

The sequence can be generated by the recurrence relation

$$V_0 = \text{starting value}, \qquad V_{n+1} = R \times V_n + D$$

where $R = 1 + \dfrac{r}{100}$ for growth and $R = 1 - \dfrac{r}{100}$ for decay.

Example 17 Modelling a first-order linear recurrence relation

Each day, 1.4% of the water in a swimming pool is lost due to evaporation.

At the end of each day, 1000 litres of water are added to the pool.

There are currently 150 000 litres of water in the pool.

Let W_n be the amount of water in the pool after n days.

a Write a recurrence relation that models the amount of water in the pool.

b Use the recurrence relation to find the amount of water in the pool after 3 days. Write your answer to the nearest litre.

Solution

a **1** Calculate the value of R. Water is lost from the pool so there is geometric decay at the rate of 1.4% per day.

$$R = 1 - \frac{1.4}{100} = 0.986$$

2 Determine the value of D. Water is added to the pool, so D will be positive.

$$D = 1000$$

3 Determine the starting value, W_0. This is the current state, the amount of water in the pool at the start of the modelling process.

$$W_0 = 150\,000$$

4 Write your answer.

$$W_0 = 150\,000, \qquad W_{n+1} = 0.986 \times W_n + 1000$$

b **1** Use the recurrence relation to calculate the amount of water in the pool after 3 days. This will be W_3.

$$W_0 = 150\,000$$
$$W_1 = 0.986 \times 150\,000 + 1000 = 148\,900$$
$$W_2 = 0.986 \times 148\,900 + 1000 = 147\,815.4$$
$$W_3 = 0.986 \times 147\,815.4 + 1000 = 146\,745.9844$$

2 Write your answer.

After 3 days, there will be 146 746 litres of water in the pool.

ISBN 978-1-009-25917-0

Exercise 5C

Describing the change in the terms of a sequence

Example 15

1 Describe the nature (arithmetic, geometric) and direction (growth, decay) of each of the following sequences.

a 54, 44, 34, . . .
b 3, 12, 48, . . .
c 45, 15, 5, . . .
d 25, 34, 43, . . .

2 Describe the nature (arithmetic, geometric) and direction (growth, decay) of the sequences determined from the following rules for the *n*th term.

a $T_n = -4 + (n - 1) \times 8$
b $T_n = 2 + (n - 1) \times -5$
c $T_n = 0.8 \times 3^{(n-1)}$
d $T_n = 2 \times 0.6^{(n-1)}$

3 Describe the nature (arithmetic, geometric, combined) and direction (growth, decay) of the sequences determined from the following recurrence relations.

a $T_1 = 56, \quad T_{n+1} = 1.5 \times T_n$
b $T_1 = -34, \quad T_{n+1} = T_n + 7$
c $T_1 = 150, \quad T_{n+1} = 3 \times T_n - 5$
d $T_1 = 0.2, \quad T_{n+1} = 2.8 \times T_n$
e $T_1 = -16, \quad T_{n+1} = 0.5 \times T_n + 9$
f $T_1 = 27, \quad T_{n+1} = 0.5 \times T_n + 4$

Modelling first-order linear recurrence relations

4 The following recurrence relation can be used to model the number of native mice in a population after *n* years, M_n

$$M_0 = 8600, \qquad M_{n+1} = 1.15M_n - 750$$

This model takes into account percentage growth in the population due to breeding and linear decay due to mice leaving the population.

a How many mice were initially in the population?
b What is the annual percentage rate of increase in the mice population due to breeding?
c How many mice leave the population each year?
d Overall, is the population of mice growing or decaying?
e Use the recurrence relation to estimate, to the nearest whole number, the number of mice expected to be in the population after 5 years.

5 The following recurrence relation can be used to model the number of parrots in a population after n years, P_n

$$P_0 = 2600, \qquad P_{n+1} = 0.95P_n - 200$$

This model takes into account percentage decay in the population and linear decay due to parrots leaving the population.

- **a** How many parrots were initially in the population?
- **b** What is the annual percentage decay in the parrot population?
- **c** How many parrots leave the population each year?
- **d** Overall, is the population of parrots growing or decaying?
- **e** Use the recurrence relation to estimate, to the nearest whole number, the number of parrots expected to be in the population after 5 years.
- **f** After how many years is the population of parrots expected to reach zero?

6 Michelle currently owns a total of 3000 shares in a company. She would like to increase the number of shares she owns by 4% every month. Michelle will buy 40 shares each month as well. The following recurrence relation can be used to model the number of shares Michelle owns after n months, S_n

$$S_0 = A, \qquad S_{n+1} = B\,S_n + C$$

where A, B and C are constants.

- **a** Find the value of A, B and C.
- **b** Use the recurrence relation to find the number of shares Michelle will have in the company after 6 months. Write your answer to the nearest whole number of shares.
- **c** Use the recurrence relation to find how many months it will take Michelle to own more than 5000 shares.

Example 17

7 A real estate agent currently has a list of 260 houses for sale. On average, they sell 4.5% of the houses they have on the list every month. On average, they add 8 houses to the list every month.

Let H_n = the number of houses on the list after n months.

- **a** Write a recurrence relation that can be used to model the number of houses on the list each month.
- **b** Use the recurrence relation to determine:
  - **i** the number of houses on the list after 4 months, rounded to the nearest whole number
  - **ii** the number of months after which there are first less than 200 houses on the list.

8 The number of students in a school has been increasing by 3% of the previous year's number of students each year. Each year, an average of 20 students leave the school. There are currently 825 students in the school.

Let S_n = the number of students in the school after n years.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

a Write a recurrence relation that can be used to model the number of students in the school each year.

b Use the recurrence relation to determine the number of students in the school after 3 years, rounded to the nearest whole number.

c The school only has room for 900 students. After how many years will the school run out of room?

9 An internet service provider currently supplies internet services to 360 homes in a particular city suburb. Each month, an average of 12% of their customers cancel their service and an average of 3 new customers begin service.

Let C_n = the number of customers of the internet service provider after n months.

a Write a recurrence relation that can be used to model the number of customers of the internet service provider each month.

b Use the recurrence relation to determine the number of customers of the internet service provider after 4 months, rounded to the nearest whole number.

c After how many months will the internet service provider have less than 100 customers in the suburb?

5D Long-term steady state

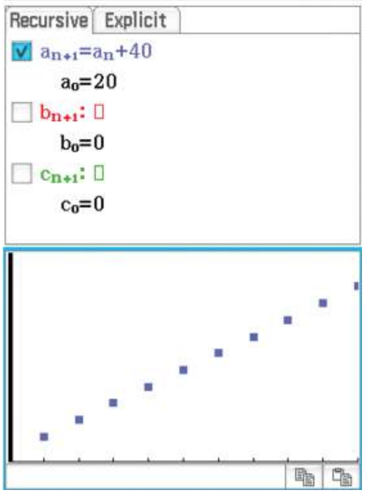
Long-term solutions for first-order linear relations

A knowledge of the long-term solution of first-order linear relations is important in some applications of arithmetic and geometric growth and decay. A first-order linear recurrence relation may give a long-term increasing, decreasing or steady-state solution.

Increasing long term solutions

Consider an oven which is initially at an ambient temperature of 20°C. The oven is turned on and increases in temperature by 40°C per minute. This scenario can be modelled by the recurrence relation $T_0 = 20,\ T_{n+1} = T_n + 40$, where T_n is the temperature of the oven n minutes after it has been turned on.

n	T_n
0	20
1	60
2	100
3	140
4	180
⋮	⋮



☑ $a_{n+1}=a_n+40$

$a_0=20$

☐ b_{n+1}: ▯

$b_0=0$

☐ c_{n+1}: ▯

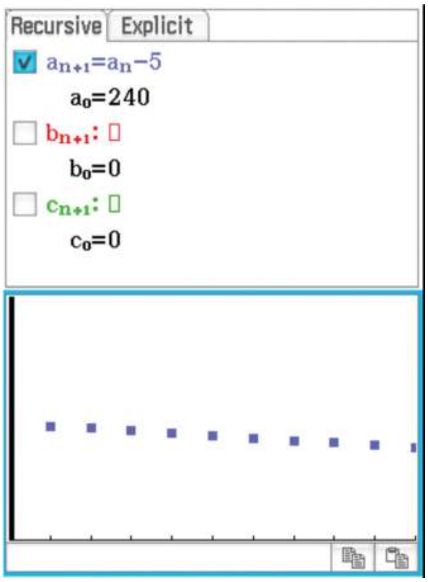
$c_0=0$

Graphing this sequence it can be seen that the temperature of the oven is increasing arithmetically, and will continue to increase indefinitely. This is an example of an **increasing long-term solution**.

Decreasing long term solutions

Conversely the oven is switched off at 240°C and its temperature drops 5°C per minute. This scenario can be modelled by the recurrence relation $T_0 = 240,\ T_{n+1} = T_n - 5$, where T_n is the temperature of the oven *n* minutes after it has been turned off.

n	T_n
0	240
1	235
2	230
3	225
4	220
⋮	⋮
47	5
48	0
49	−5



Graphing this new sequence it can be seen that the temperature of the oven is decreasing geometrically, and will continue to decrease indefinitely. This is an example of a **decreasing long-term solution**.

In this long-term sequence, the temperature of the oven will continue to decrease, eventually becoming negative. However, this is not possible in real life, as the temperature of the oven cannot decrease below ambient room temperature.

In this situation our model for temperature should be modified such that it reaches a constant value that does not change in the long run. This is known as a **long-term steady state.**

Long-term steady state for first-order linear relations

A recurrence relation that generates a first-order linear relation can sometimes generate terms of a sequence that are not identical, but as more and more terms are generated, the value of the terms tends to be very similar.

The recurrence relation $V_0 = 10,\ V_{n+1} = 0.5V_n + 2$ generates the following sequence:

$$10, 7, 5.5, \ldots$$

The terms of this sequence are decreasing in value, but the nature of this decrease isn't obvious until we generate more terms of the sequence. The table on the right shows some terms of the sequence and reveals that, as the number of the term increases, the value of the term gets closer to 4.

The reason why this happens can be understood by looking at the calculation for V_1, which occurs in two steps.

Geometric decay: $0.5 \times 10 = 5$

Arithmetic growth: $+ 2$

The component of geometric decay, 5, is larger than the arithmetic growth of 2 and so the terms of the sequence will gradually decrease in size.

n	V_n
0	10
1	7
2	5.5
3	4.75
4	4.375
5	4.1875
6	4.093 75
⋮	⋮
12	4.001 464 844
13	4.000 732 422
14	4.000 366 211

The constant steady state value in this example is 4. The sequence is said to *converge*, with a **long-term steady state** of 4.

The steady state value of the sequence terms, in the long run, is called the **limit** of the sequence.

The sequence above has rule $V_{n+1} = 0.5V_n + 2$, with $D = 2$ and $R = 0.5$, and so the long-term steady-state solution of the sequence will be

$$\text{limit} = \frac{2}{1 - 0.5} = \frac{2}{0.4} = 4$$

as observed.

Long-term steady state for a recurrence relation

Let V_n = the value of a combined arithmetic and geometric sequence after n periods.

Let V_0 = the starting value of the sequence.

Let R = the ratio for geometric change.

Let D = the amount of arithmetic change.

A recurrence relation of the form

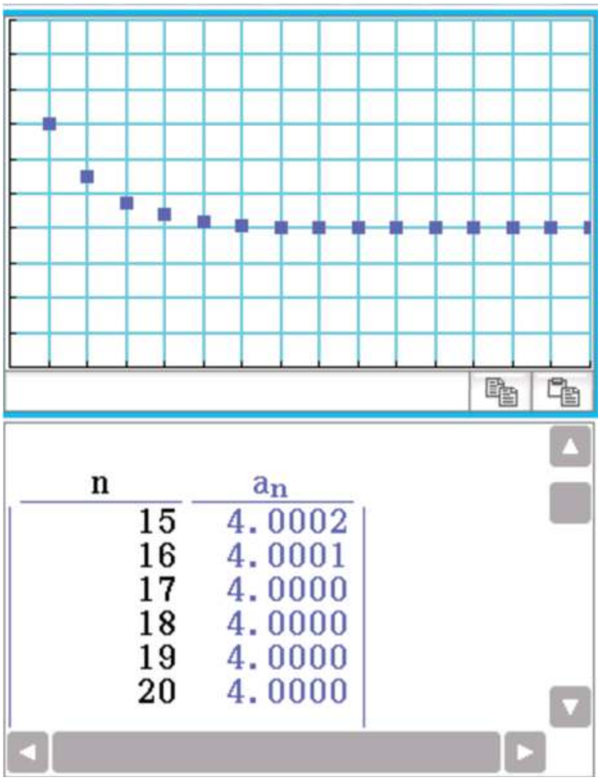
$$V_0 = \text{starting value}, \qquad V_{n+1} = R \times V_n + D$$

will generate a converging, steady state sequence if:

- $0 < R < 1$ (geometric decay)

For first-order linear relations where $0 < R < 1$, the long-term steady-state can be found by:

$$\frac{D}{1 - R}$$



n	a_n
15	4.0002
16	4.0001
17	4.0000
18	4.0000
19	4.0000
20	4.0000

Graphically, we can see that eventually the terms of the sequence form a plateau and will neither increase or decrease.

If the sequence generates a long-term steady-state then $V_n = V_{n+1} = V_{n+2}\ldots$

Using this understanding, we can also find the value of the long-term steady-state algebraically.

Sequence: $V_{n+1} = 0.5V_n + 2$

If there is a long-term steady-state, then $V_n = V_{n+1}$

$$V_n = 0.5V_n + 2$$

Let x represent V_n

$$x = 0.5x + 2$$

This is a linear equation that can be solved:

$$0.5x = 2$$

$$x = 2 \div 0.5$$

$$x = 4$$

Example 18 Determining the long-term steady state of a first-order linear relation

A trout farm currently has 8000 trout in a pond.

The population of trout increases by 12% each month.

Determine the number of trout that can be caught and sold if the number of trout in the pond is to remain at 8000.

Solution

1 Write the value of T_0 — $T_0 = 8000$

2 Determine value of R — $R = 1 + \frac{1.2}{100} = 1.12$

3 Write the recurrence relation for the sequence — Let D be the number of trout sold

$$T_0 = 8000$$

$$T_{n+1} = 1.12T_n - D$$

ISBN 978-1-009-25917-0

4 Set up equation for long-term steady state:

$T_n = T_{n+1} = 8000$

$$8000 = 1.12 \times 8000 - D$$
$$8000 = 8960 - D$$
$$D = 8960 - 8000$$
$$D = 960$$

5 Write your answer.

The trout farm can sell 960 trout each month to maintain a population of 8000.

Example 19 Determining the steady state of a recurrence relation

A gardener notices that 4% of her plants are being eaten by wild rabbits every week. She plants 6 new plants every week.

a How many plants can she expect to have in her garden, in the long run?

b How many new plants should she plant each week in order to maintain 100 plants in the garden, in the long run?

Solution

a **1** The number of plants added to the garden each week is the value of D.

$$D = 6$$

2 Calculate the value of R.

$$R = 1 - \frac{4}{100} = 0.96$$

3 Calculate the limit.

$$\text{limit} = \frac{6}{1 - 0.96} \text{ or limit} = \frac{6}{4} \times 100 = \frac{6}{0.04}$$
$$= 150$$

4 Write your answer.

The gardener can expect to have 150 plants in her garden, in the long run.

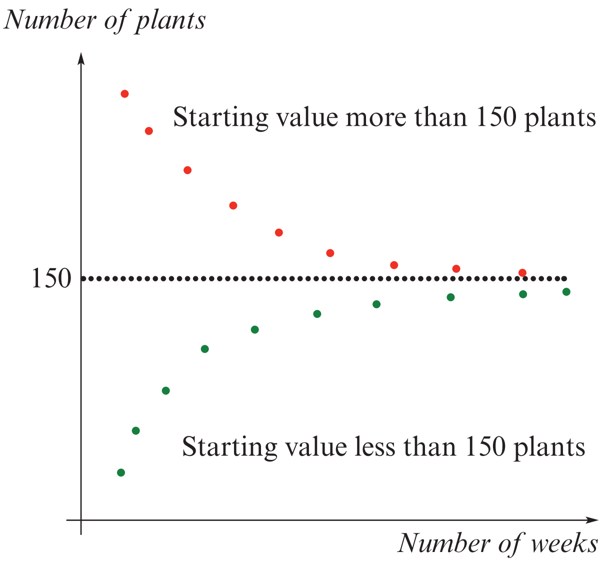
b **1** The limit is 100. Use the rule to determine D.

$$100 = \frac{D}{1 - 0.96}$$
$$D = (1 - 0.96) \times 100 = 4$$

2 Write your answer.

If the gardener plants 4 new plants each week, she will maintain 100 plants in the garden in the long run.

The long-term steady state for a sequence does not depend at all on the starting value of the sequence. The gardener above could have begun with more than the limit of the sequence, for example, 200 plants. In the long run, the number of plants in her garden would *decrease* to the limit of 150. The terms of the sequence will still converge, but from *above* rather than *below*.



The graph on the right shows the **convergence** of the number of plants in the garden for starting values that are more than and less than the limit.

Exercise 5D

Identifying steady state sequences

1 For each of the following recurrence relations, state whether the sequence will have a long-term increasing, decreasing or steady-state solution.

- **a** $V_0 = 5, \quad V_{n+1} = V_n + 0$
- **b** $V_0 = 6, \quad V_{n+1} = V_n + 2$
- **c** $V_0 = 12, \quad V_{n+1} = 0.6V_n + 2$
- **d** $V_0 = -5, \quad V_{n+1} = 0.95V_n + 2$
- **e** $V_0 = -10, \quad V_{n+1} = 0.5V_n$
- **f** $V_0 = 2.5, \quad V_{n+1} = 0.84V_n + 10$
- **g** $V_0 = 120, \quad V_{n+1} = 0.4V_n$

Determining limits of long-term steady state sequences

2 For each of the following recurrence relations, find the long-term steady-state solution.

- **a** $V_0 = 2, \quad V_{n+1} = 0.94V_n + 6$
- **b** $V_0 = 120, \quad V_{n+1} = 0.84V_n + 15$
- **c** $V_0 = -5, \quad V_{n+1} = 0.4V_n + 12$
- **d** $V_0 = 28, \quad V_{n+1} = 0.2V_n + 8$

ISBN 978-1-009-25917-0

Determining quantities for steady state sequences

3 Determine the value of D for which the recurrence relation

$$V_0 = 100, \qquad V_{n+1} = 0.8 \times V_n + D$$

will generate a long-term steady-state solution of 200.

4 Determine the value of R for which the recurrence relation

$$V_0 = 3, \qquad V_{n+1} = R \times V_n - 16$$

will generate a long-term steady-state solution of −40.

Applications of trivial and steady state sequences

Example 19

5 Dennis has a fish tank with a capacity of 250 litres. Every day, 2% of the water in the tank evaporates. Every day, Dennis adds P litres of water to the tank.
Let W_n be the amount of water in the tank after n days.

- **a** What is the value of the geometric decay factor R in this situation?
- **b** Write a recurrence relation that models the amount of water in the fish tank.
- **c** Calculate the value of P if:
  - **i** the amount of water in the tank after each day is a constant 250 L
  - **ii** the amount of water in the tank, in the long run, will be 200 L.

6 Brian has planted trees around his house. The trees are currently 100 cm tall. On average, the trees should grow by 80 cm each year.
To keep their growth under control, Brian will trim them by 20% of their height each year.
Let H_n be the height of a tree after n years.

- **a** Write a recurrence relation that models the height of a tree.
- **b** Explain why, in the long run, Brian can expect his trees to reach steady state height.
- **c** What height can Brian expect his trees to grow to, in the long run?
- **d** If Brian wanted to keep his trees at a height below 2.5 m, what percentage of the height should be trimmed each year?

Key ideas and chapter summary

Assignment

Modelling	**Modelling** is the use of a mathematical rule or formula to represent or model real-life situations. Recurrence relations can be used to model situations involving the *growth* (increase) or *decay* (decrease) in values of a quantity.
Interest	**Interest** is the fee that is added to a loan or the payment for investing money.
Simple interest	**Simple interest** is a fixed amount of interest that is paid at regular time intervals. Simple interest is an example of linear growth.
Principal	The **principal** is the initial amount that is invested or borrowed.
Balance	The **balance** is the value of a loan or investment at any time during the loan or investment period.
Recurrence relation model for simple interest	The **recurrence relation model for simple interest** has the form $V_0 = \text{principal}, \quad V_{n+1} = V_n + D$ where V_n is the value of the loan or investment after n time periods and D is the interest that is added after every time period. The value of the simple interest, D, can be calculated using $D = \frac{r}{100} \times V_0$ where r is the percentage interest rate.
Depreciation	**Depreciation** is the amount by which the value of an item decreases after a period of time.
Flat-rate depreciation	**Flat-rate depreciation** is a constant amount subtracted from the value of an item at regular time intervals. It is an example of linear decay.
Recurrence relation model for flat-rate depreciation	The **recurrence relation model for flat-rate depreciation** has the form $V_0 = \text{purchase price}, \quad V_{n+1} = V_n - D$ where V_n is the value of the item after n time periods and D is the depreciation after every time period. The value of the flat-rate depreciation, D, can be calculated using $D = \frac{r}{100} \times V_0$ where r is the percentage depreciation rate.
Unit-cost depreciation	**Unit-cost depreciation** is depreciation that is calculated based on units of use rather than time. It is an example of linear decay.

Recurrence relation model for unit-cost depreciation	The **recurrence relation model for unit-cost depreciation** has the form V_0 = purchase price, $V_{n+1} = V_n - D$ where V_n is the value of the item after n units of use and D is the depreciation per unit of use.
Percentage growth and decay	If a quantity grows by $r\%$ each year, then $R = 1 + \frac{r}{100}$. If a quantity decays by $r\%$ each year, then $R = 1 - \frac{r}{100}$.
Compound interest	When interest is added to a loan or investment and then contributes to earning more interest, the interest is said to compound. **Compound interest** is an example of geometric growth.
Compounding period	Interest rates are usually quoted as annual rates (per annum). Interest is sometimes calculated more regularly than once a year, for example, each quarter, month, fortnight, week or day. The time period for the calculation of interest is called the **compounding period**.
Recurrence relation model for compound interest investment and loans	The **recurrence relation model for compound interest** has the form V_0 = principal, $V_{n+1} = R \times V_n$ where V_n is the value of the loan or investment after n compounding time periods and $R = 1 + \frac{r}{100}$. Note: r is the percentage interest rate per compounding time period.
Reducing-balance depreciation	When the value of an item decreases as a percentage of its value after each time period, it is said to be depreciating using a reducing-balance method. It is an example of geometric decay.
Recurrence relation model for reducing-balance depreciation	The **recurrence relation model for reducing-balance depreciation** has the form V_0 = purchase price, $V_{n+1} = R \times V_n$ where V_n is the value of the item after n years and $R = 1 - \frac{r}{100}$. Note: r is the percentage depreciation rate per year.
Recurrence relation model for combined growth and decay	The **recurrence relation model for combined growth and decay** has the form V_0 = starting value, $V_{n+1} = R \times V_n + D$ where D is the amount of arithmetic change ($+D$ for growth, $-D$ for decay). The value of R can be determined from $R = 1 + \frac{r}{100}$ for geometric growth and $R = 1 - \frac{r}{100}$ for geometric decay.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

First-order linear recurrence relation	A first-order linear recurrence relation is defined by the rule: $T_1 = a, \quad T_{n+1} = bT_n + C, \quad \text{for } n \geq 1$ These are neither arithmetic nor geometric but show general growth or decay.
Long-term steady state	A sequence has a **long-term steady state** if the terms of the sequence either decrease or increase towards a value that remains almost constant. Long-term steady states only occur for sequences generated from a recurrence relation that involves geometric decay.
Long-term steady state for a recurrence relation	A recurrence relation that generates a sequence with a long-term steady state has the form $V_0 = \text{starting value}, \quad V_{n+1} = R \times V_n + D$ where $0 < R < 1$ (geometric decay) and $D > 0$ (arithmetic growth) The long-term steady state of a sequence can be determined using the rule: $\dfrac{D}{1 - R}$ The long-term steady state of a sequence is independent of the starting value.

Skills check

Having completed this chapter, you should be able to:

- apply arithmetic sequence theory to practical situations
- use a recurrence relation to model simple interest investments and loans
- use a recurrence relation to model flat-rate depreciation
- use a recurrence relation to model unit-cost depreciation
- apply geometric sequence theory to practical situations
- calculate percentage change and use this in recurrence relation models
- convert annual interest rates to rates for other compounding time periods
- use a recurrence relation to model compound interest investments and loans
- use a recurrence relation to model reducing-balance depreciation
- use recurrence relations to model situations involving combined arithmetic and geometric growth and decay
- identify the conditions for which a recurrence relation will generate a long-term steady state sequence

ISBN 978-1-009-25917-0

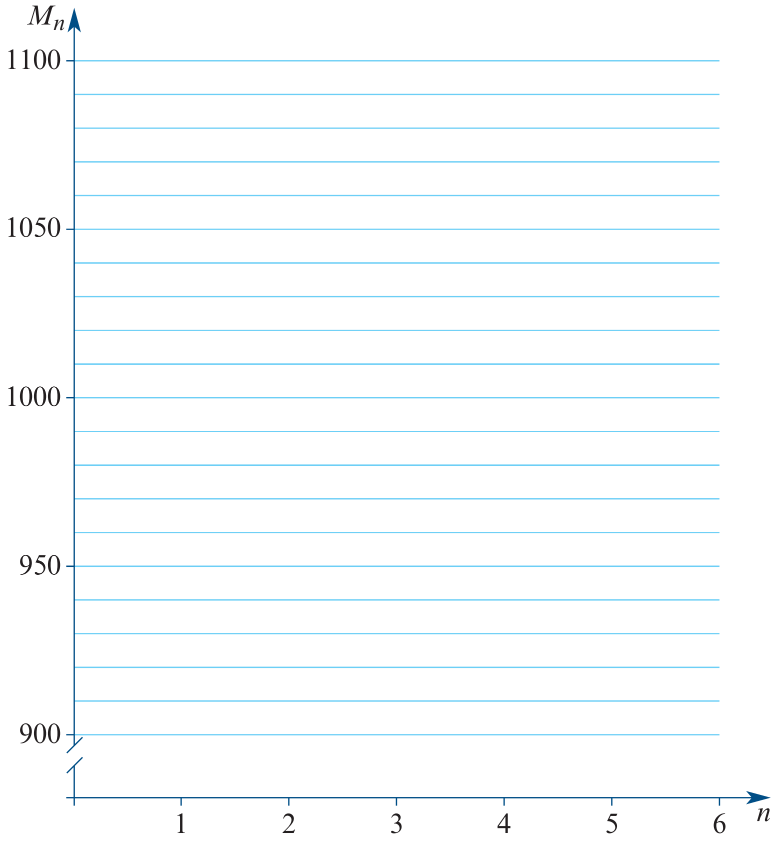
Short-answer questions

1 What common ratio should be used to generate a sequence where each term is 7% greater than the previous term?

2 Brian has two trees in his backyard. Every month, he will plant three more trees. Write a recurrence model for T_n, the number of trees in Brian's backyard after n months.

3 Complete the table and graph that shows the value of a simple interest investment of $1000, earning interest of $5 per month.

Month, n	0	1	2	3	4	5
M_n	1000					



4 Arthur invests $2000 with a bank. He will be paid simple interest at the rate of 5.1% per annum. If V_n is the value of Arthur's investment after n years, write a recurrence model for Arthur's investment.

5 The recurrence relation that generates a sequence of numbers representing the value of a car n years after it was purchased is $V_0 = 18\,000,\ V_{n+1} = V_n - 1098$. Given the car had a purchase price of $18 000, determine the type (flat-rate, unit cost) and value of the depreciation as a percentage.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

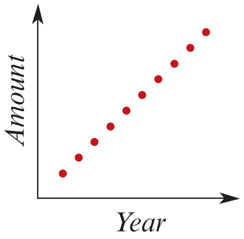
6 A computer is depreciating using a flat-rate depreciation method. It was purchased for \$2800 and depreciates at the rate of 8% per annum. Determine the amount of depreciation after 4 years.

7 A car is depreciating using a unit-cost depreciation method. It was purchased for \$18 990 and, after travelling a total of 20 000 kilometres, it has an estimated value of \$15 990. Calculate the amount of depreciation per kilometre.

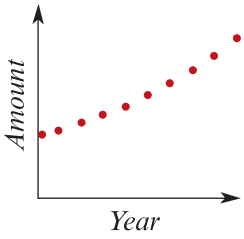
8 A population of penguins is decreasing by 8% every year. There are currently 2700 penguins in the population. Write a recurrence relation model for the number of penguins in the population after n years, P_n.

9 An investment of \$50 000 is made at a fixed rate of interest compounding annually over a number of years. Which graph best represents the value of the investment at the end of each year?

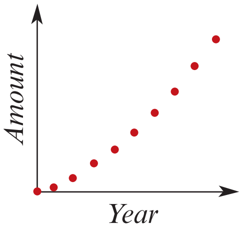
A



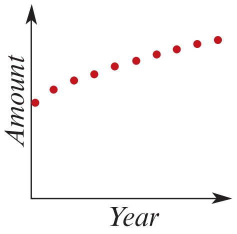
B



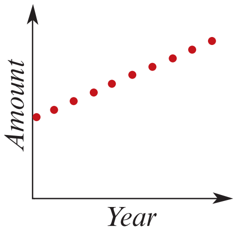
C



D



E



10 An item is depreciated using a reducing-balance depreciation method. The value of the item after n years, V_n, is modelled by the recurrence relation $V_0 = 4500$, $V_{n+1} = 0.86V_n$. Write an *nth* term rule for the value of the item after n years.

11 The interest rate on a compound interest loan is 12.6% per annum, compounding monthly. The value of the loan after n months, V_n, is modelled by the recurrence relation $V_0 = 400,\ V_{n+1} = R \times V_n$. Determine the value of R in this model.

CAS

12 Sandra invests \$6000 in an account that pays compounding interest at the rate of 4.57% per annum.

a Write a recurrence relation for the value of the investment in the account, V_n.

b Determine the number of years it takes the investment to exceed \$8000.

13 The wombats in a national park are breeding at a rate that increases their population by 8% each year. On average, 60 wombats die each year. There are currently 490 wombats in the national park. Write a recurrence relation that can be used to estimate the number of wombats, W_n, in the national park after n years.

14 Which of the following recurrence relations generate a sequence that has a long-term steady state?

A $T_o = 6, \quad T_{n+1} = T_n + 5$

B $T_o = 6, \quad T_{n+1} = T_n - 0.9$

C $T_o = 6, \quad T_{n+1} = 0.95T_n + 4$

D $T_o = 6, \quad T_{n+1} = 1.02T_n - 12$

E $T_o = 6, \quad T_{n+1} = 1.04T_n + 2$

15 Determine the long-term steady state for the sequence generated by the recurrence relation $T_0 = 120, \; T_{n+1} = 0.75T_n + 150$.

16 Murray is a sheep farmer with 800 sheep in his flock. Every year, Murray sells a percentage of his current flock of sheep. Every year, Murray keeps a fixed number of any lambs that are born into his flock.

a If Murray keeps 50 lambs per year and sells 10% of his flock every year, how many sheep can Murray expect to have in his flock in the long run?

b If Murray keeps 100 lambs per year, what percentage of sheep should he sell each year in order to maintain 800 sheep in his flock?

Extended-response questions

1 Microwave cooking instructions for heating muffins is to heat one muffin for 45 seconds and to allow another 30 seconds for each extra muffin. State the heating times for 1, 2, 3 and 4 muffins.

2 A basketball was dropped from a height of 48 metres. After each bounce, it reached only half of the previous height. Starting with the height of 48 metres, list the next three heights that it reached after each bounce.

3 The terms of a sequence start at 1000 and each new term is 8% less than the previous term. Find the 6th term, correct to two decimal places.

ISBN 978-1-009-25917-0

4 Jack borrows \$20 000 from a bank and is charged simple interest at the rate of 9.4% per annum. Let V_n be the value of the loan after n years.

CAS

a Write a recurrence relation model for the value of Jack's loan after n years.

b Use the model to find how much Jack will need to pay the bank after 5 years.

c Convert the annual interest rate to a weekly one, rounding your answer to two significant figures.

d Using the interest rate from part c, how many weeks does it take the value of Jack's loan to reach \$20 612?

5 Kelly bought her current car 5 years ago for \$22 500.
Let V_n be the value of Kelly's car after n years.

a If Kelly assumes a flat-rate depreciation of 12% per annum:

  i write a recurrence model for the value of Kelly's car after n years

  ii use the recurrence model to find the current value of Kelly's car.

b If Kelly assumes reducing value depreciation at 16% per annum:

  i write a recurrence model for the value of Kelly's car after n years

  ii use the recurrence model to find the current value of Kelly's car using reducing-balance depreciation.

c On the same axes, sketch a graph of the value of Kelly's car against the number of years for both flat-rate and reducing-balance depreciation.

6 A commercial cleaner bought a new vacuum cleaner for \$650. The value of the vacuum cleaner decreases by \$10 for every 50 offices that it cleans.

a By how much does the cleaning of one office depreciate the value of the vacuum cleaner?

b Write a recurrence model for the value of the vacuum cleaner after n offices have been cleaned.

c The cleaner has a contract to clean 10 offices, 5 nights a week for 40 weeks in a year. What is the value of the vacuum cleaner after 1 year?

CAS

7 Meghan has $5000 to invest.

a Company A offers her an account paying 6.3% per annum simple interest. How much will she have in this account at the end of 5 years?

b Company B offers her an account paying 6.1% per annum compound interest. How much will she have in this account at the end of 5 years?

c Find, correct to one decimal place, the simple interest rate that company A should offer if the two investments are to have equal values after 5 years.

8 The number of bacteria in a colony doubles each day. On the first day there were 300 bacteria.

a State the value of the first term, a, and the common ratio, r.

b How many bacteria will there be on the eighth day?

c On which day will the number of bacteria be 614 400?

9 Margaret started working for a company on an annual salary of $65 000 with a guaranteed increase of 9% each year.

a Give the value of the first term, a, and the common ratio, r, for the sequence of her salaries each year.

b Find Margaret's salary for the third year working with the company.

c In which year will her salary be $100 010.56?

10 A snail, starting from the bottom of a drainpipe, climbs 486 centimetres during the first day, 324 centimetres the next day and 216 centimetres the following day. Assuming that this pattern continues, answer the following.

a Find the common ratio for the sequence of distances travelled.

b How far will the snail travel on the fourth day?

c What will be the total distance climbed after 4 days?

11 A sum of $30 000 is borrowed at an interest rate of 7.2% per annum, compounding monthly.
Let V_n be the value of the loan after n months.

a Write a recurrence model for the value of this loan.

b Generate a sequence of numbers that represent the value of this loan each month over a period of 5 months. Round each value but use full decimals for calculations.

c What is the value of the loan after 1 year?

d If the loan is fully repaid after 18 months, how much money is needed?

12 Ilana uses her credit card to buy a dress costing $300, knowing that she will not be able to pay it off for some time. If she is charged interest on the amount for 6 months, how much will the dress finally cost her? (Assume that interest is charged at 18% per annum, compounded monthly, and that no payments are made during the 6-month period.)

CAS

13 A population of fish in a pond is growing at the rate of 6% every month. On average, each month there are 80 fish caught by people fishing in the pond. Initially, there were 840 fish in the pond.

Let F_n be the number of fish in the pond after n months.

a Write a recurrence model for the number of fish in the pond.

b Determine the number of fish in the pond after 5 months. Write your answer correct to the nearest fish.

c After how many months will there be no fish left in the pond?

14 Vikki is an artist who makes ceramic bowls. On average, she can produce 3 bowls every week in her studio. On average, she sells 5% of the bowls she has completed. Vikki currently has 20 bowls in the studio. Let B_n be the number of bowls Vikki has stored in her studio after n weeks.

a Write a recurrence relation that models the number of bowls in Vikki's studio.

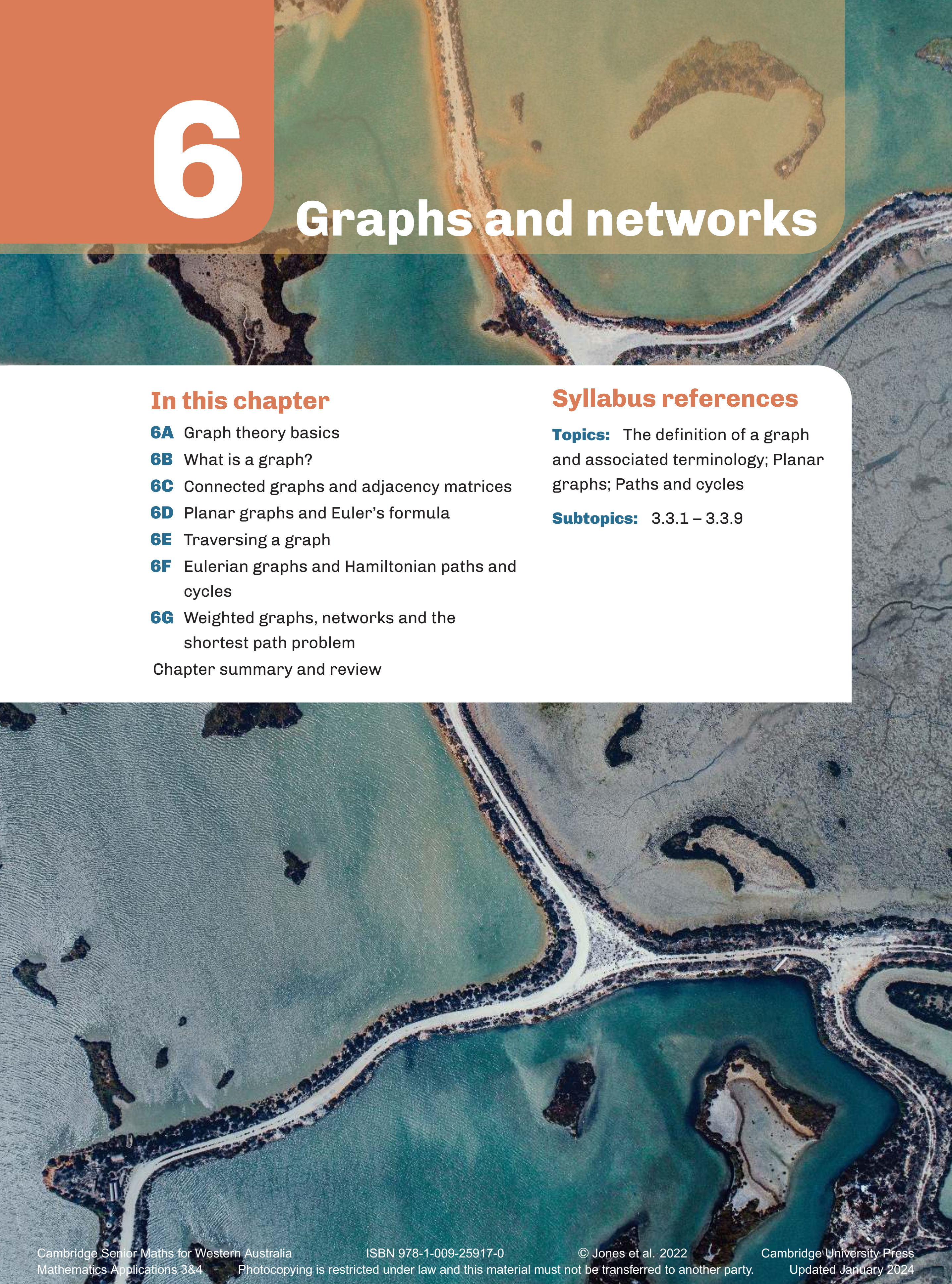
b In the long run, how many bowls can Vikki expect to be in her studio each week?

c How many bowls should Vikki produce per week in order to maintain 20 bowls in the studio, in the long run?

d Vikki has room to store no more than 30 bowls. If she produces 3 bowls every week, what percentage of the bowls she has in the studio should she sell to avoid running out of storage room?

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024



6

Graphs and networks

In this chapter

Syllabus references

Topics: The definition of a graph and associated terminology; Planar graphs; Paths and cycles

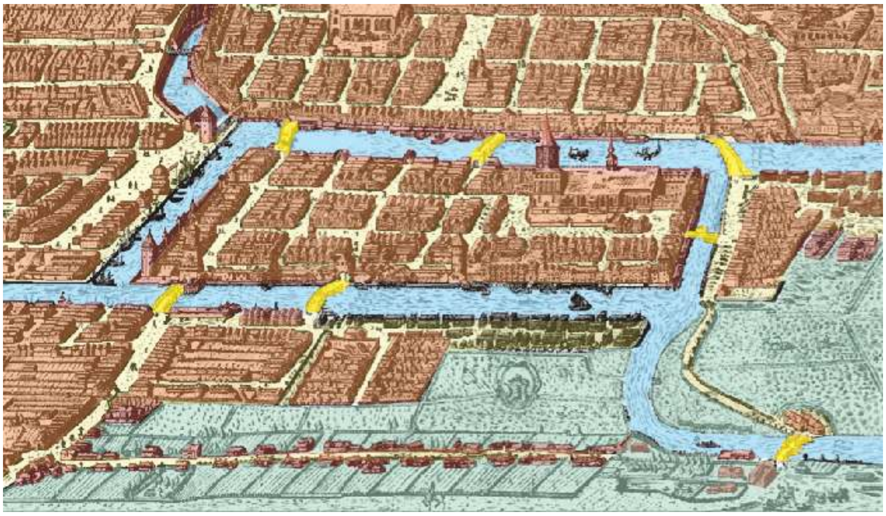
Subtopics: 3.3.1 – 3.3.9

6A Graph theory basics

The Königsberg bridge problem

The problem that began the scientific study of graphs and networks is known as the *Königsberg bridge problem*. The problem began as follows:

The centre of the old northern European city of Königsberg was located on an island in the middle of the Pregel River. The island was connected to the banks of the river and to another island by five bridges. Two other bridges connected the second island to the banks of the river, as shown below.

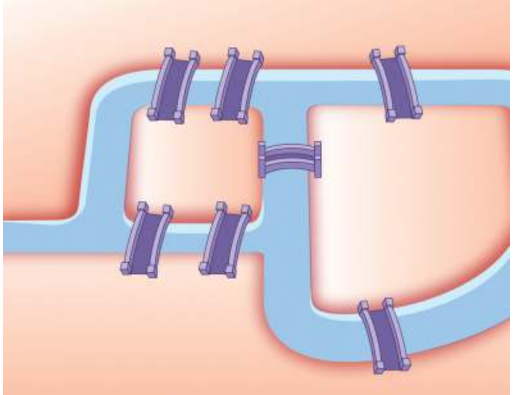


A view of Königsberg as it was in Euler's day.

Today Königsberg is Kaliningrad in Russia

ISBN 978-1-009-25917-0

The problem simplified

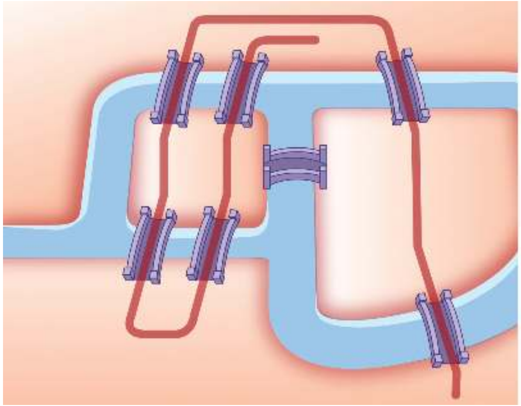


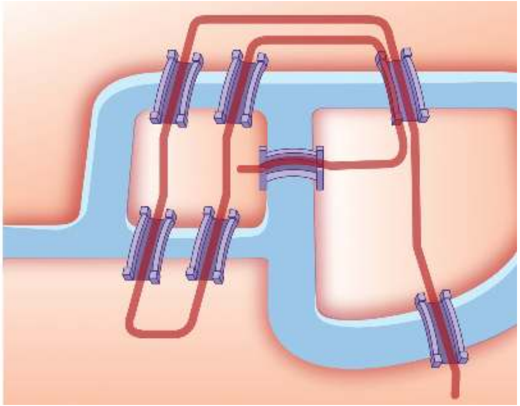
A simplified view of the situation is shown in the drawing opposite.

Can a continuous walk be planned so that all bridges are crossed only once?

Whenever someone tried to walk the route, they either ended up missing a bridge or crossing one of the bridges more than once.

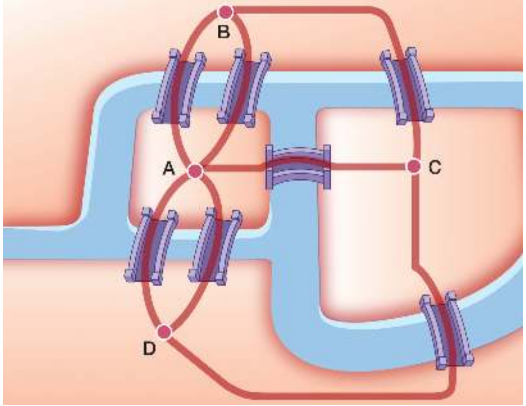
Two such failed attempts are marked on the diagrams that follow. See if you can trace out a walk on a copy of the diagram above that crosses every bridge, but only once.

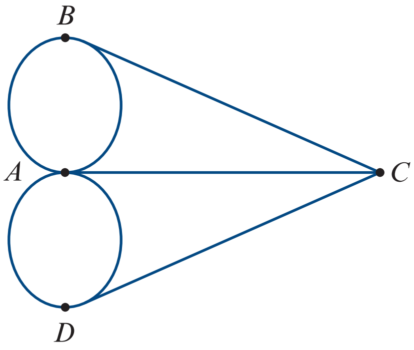




Enter the mathematician

The Königsberg bridge problem was well-known in 18th century Europe and attracted the attention of the Swiss mathematician Euler (pronounced ‘Oil-er’). He started analysing the problem by drawing a simplified diagram to represent the situation, as shown below. We now call this this type of simplified diagram a **graph**.

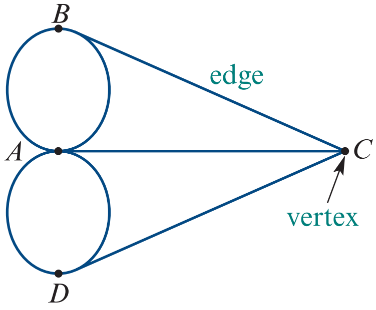




Euler’s diagram

Elements of a graph

A graph is made up of dots and lines joining these dots.



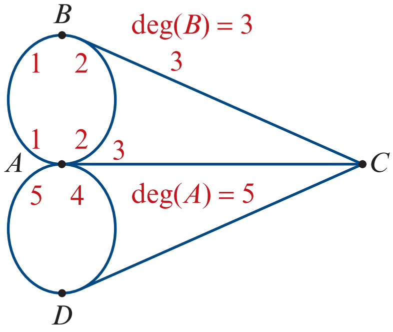
- The *dots* in a graph are called *vertices* (plural of **vertex**). These vertices represent the riverbanks and the islands.
- The *lines* in a graph are called **edges**. These edges represent the bridges.

Euler found that the solution of the Königsberg bridge problem was connected to the *degree* of each vertex.

Degree of a vertex

The **degree of a vertex** is the number of edges attached to the vertex.

For the Königsberg bridge graph:



- the degree of vertex A is 5 (because five edges are attached to vertex A): write this as $\deg(A) = 5$
- vertices B, C and D are all of degree 3 (three edges touch each of these vertices): write this as $\deg(B) = 3$, etc.

The degree of a vertex may be even or odd.

- The *degree* of a vertex will be *even* if there are an even number of edges (2, 4, 6, . . .) attached to the vertex.
- The *degree* of a vertex will be *odd* if there are an odd number of edges (1, 3, 5, . . .) attached to the vertex.

All four vertices in the Königsberg bridge graph have an odd degree.

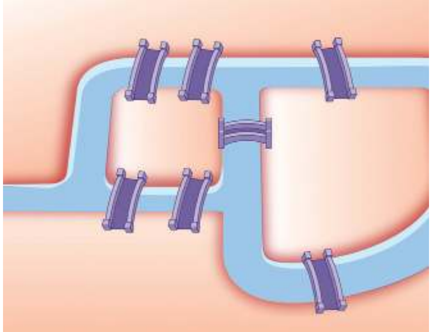
Euler's discovery

Euler was able to prove that a graph with *more than two odd* vertices cannot be traced or drawn without lifting the pencil or going over the same edge more than once. The problem was solved. The seven bridges of Königsberg could *not* be crossed in a single walk without either missing a bridge or crossing one bridge more than once.

With this analysis, a new area of mathematics was developed which has many practical applications in today's world. These include: analysing friendship networks, scheduling airline flights, designing electrical circuits, planning large-scale building projects, and many more. This relatively new area of mathematics is now called *graph theory*, some aspects of which we will explore in this chapter.

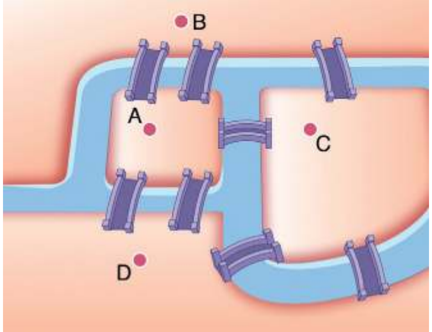
Exercise 6A

1 We now know that it is impossible to trace out a continuous walk that crosses each of the Königsberg bridges only once. If you don't believe this, try it for yourself on the diagram opposite.



In this exercise, you will investigate how things would change if the number of bridges is changed.

a The picture opposite shows a situation in which an eighth bridge has been added.



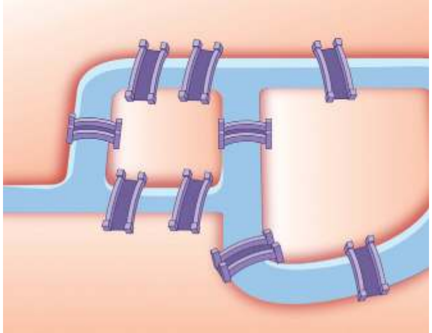
i With a pencil, or the tip of your finger, see whether you can trace out a continuous walk that crosses each of the bridges only once. Such a walk exists.

ii Construct a graph to represent this new situation with eight bridges. Labelled dots have been placed on the picture to help you draw your graph.

iii Your graph should have only two odd vertices. Check to see.

iv As you will learn later, when the graph has only two odd vertices, you can only complete the task if you start at the places represented by the odd vertices. You will then finish at the place represented by the other. Check to see.

b A ninth bridge has been added as shown opposite.



i With a pencil, or the tip of your finger, see whether you can trace out a continuous walk that crosses each of the bridges only once. Such a walk exists.

ii Construct a graph to represent this situation.

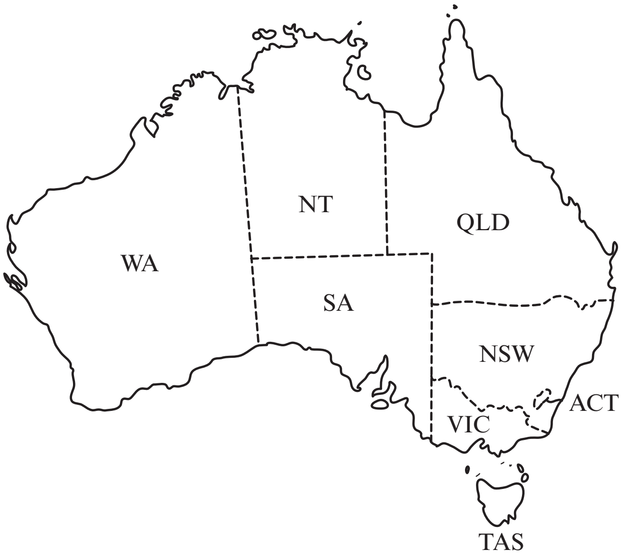
iii Your graph should *not* have any odd vertices; that is, they should all be even. Check to see.

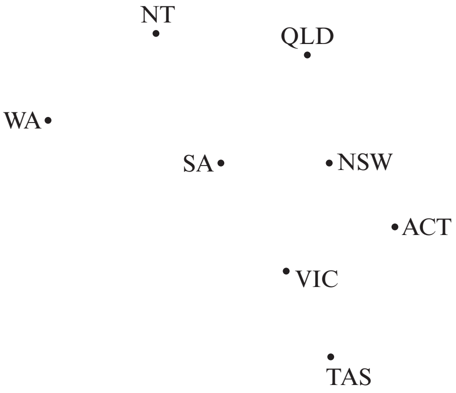
iv As you will learn later, when the graph has only even vertices, you can start your walk from any island or any river bank and still complete the task. Check to see.

ISBN 978-1-009-25917-0

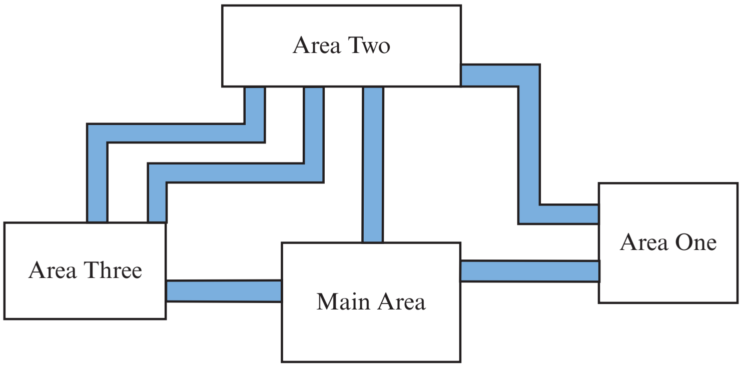
Cambridge University Press
Updated January 2024

2 The map below shows the states and territories of Australia. Draw a network that models which Australian states and territories share a common border. Use vertices to represent the states/territories and edges to represent common borders.





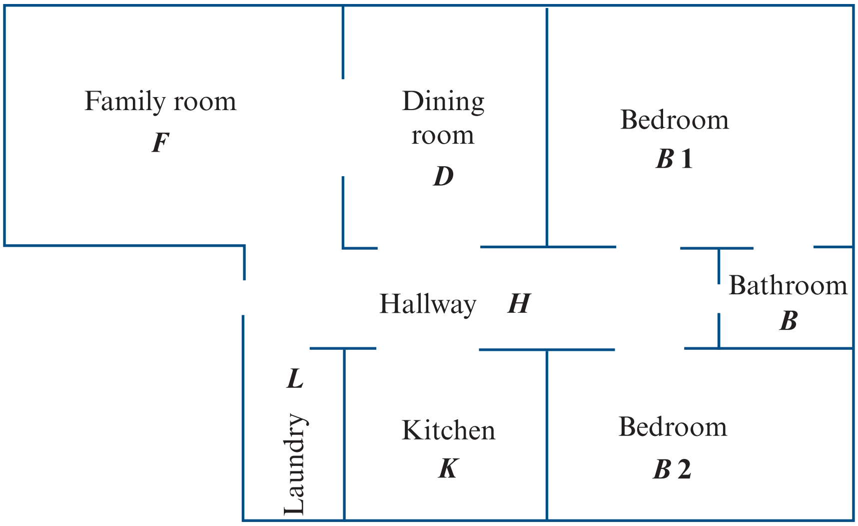
3 A video game has a map outlining the paths to travel along to visit different areas within the game world.



a Draw a network model showing the direct connections between the four areas, where each area is a vertex.

b How many ways are there for a player to travel from Area One to Area Three?

4 A plan of a house is shown below. Each room, including the hallway, is identified by a letter. Rooms are connected by doors.



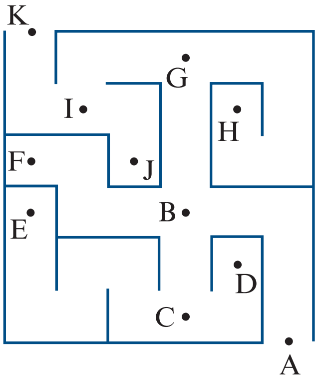
a Draw a network model showing the direct connections between rooms, where each room is a vertex.

b A real estate agent is conducting a walk-through of the house. Is it possible for the real estate agent to start in the hallway and walk through each room of the house only once? Justify your answer.

5 A park contains a giant hedge maze. The maze contains check points at each of its intersections and dead ends, denoted by the letters *A* – *K*.

a Draw a network model showing the layout of the maze.

b Determine the most efficient path through the maze and write the order of the checkpoints visited, given that the maze starts at *A* and ends at *K*.



6 A group of students are going on a school service trip. The students going on the trip are: Asher, Bree, Caitlin, Emily, Isobelle, Jorja, Meetal and Sarah. Prior to forming the group for the school trip Sarah was friends with everyone but Meetal. Meetal was friends with Bree, Asher, Caitlin and Emily. Jorja was friends with Sarah, Emily, Isobelle and Bree. Draw a network that models the pairs of friendships among the eight students prior to going on the school service trip.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

6B What is a graph?

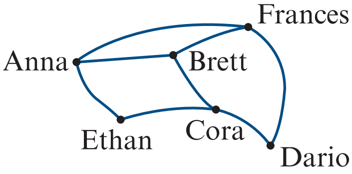
Representing connections with graphs

There are many situations in everyday life that involve connections between people or objects. Towns are connected by roads, computers are connected to the internet and people connect to each other through being friends on social media. A diagram that shows these connections is called a graph.

Edges and vertices

Six people – Anna, Brett, Cora, Dario, Ethan and Frances – have connections on a social media website. The graph shows these connections.

- Anna is a friend of Brett, Ethan and Frances.
- Brett is a friend of Anna, Cora and Frances.
- Cora is a friend of Brett, Dario and Ethan.
- Dario is a friend of Cora and Frances.
- Ethan is a friend of Anna and Cora.
- Frances is a friend of Anna, Brett and Dario.

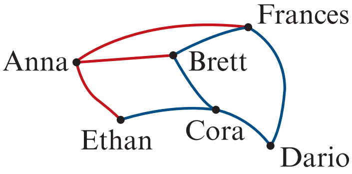


The graph shows each of the people as a dot called a vertex. The *vertices* (plural of vertex) are joined together by a line that indicates the social media friendship between the people. The lines that join the vertices in the graph are called edges.

Degree of a vertex

Anna has three friends. The vertex representing Anna has three edges attached to it. Each of these connects Anna to one of her friends.

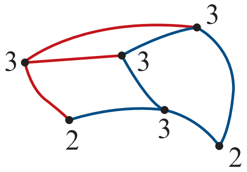
The number of times edges attach to a vertex is called the degree of that vertex.



The degree of the vertex representing Anna is odd, because there is an odd number of attachments (three) by the edges. The degree of the vertex representing Dario is even because there is an even number of attachments (two) by edges to it.

In symbolic form, we can let the letter A represent the vertex for Anna. The degree of this vertex can be written as deg(A). In this graph, deg(A) = 3.

The degrees of the vertices in this graph are:



ISBN 978-1-009-25917-0

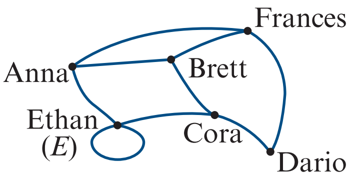
Cambridge University Press

Updated January 2024

Loops

Imagine that Ethan is able to add himself as a friend on the social media website.

The edge representing this connection would connect the vertex representing Ethan, E, back to itself. This type of edge is called a **loop**.

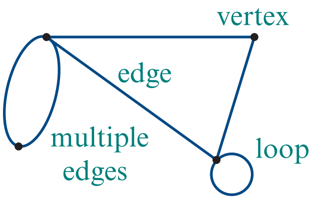


A loop is attached twice to a vertex and so it will count as two degrees. So $\deg(E) = 4$.

It is now time to introduce some definitions that we need to know in order to work with graphs. This chapter uses a specialised mathematical sense of the word 'graph' as opposed to its everyday use in statistics, science and other contexts.

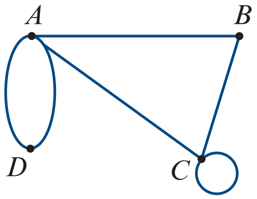
Graph elements: definitions

- A graph is a diagram that consists of a set of points called vertices that are joined by a set of lines called edges. Each edge joins two vertices.
- A loop is an edge in a graph that joins a vertex in a graph to itself.
- Two or more edges that connect the same vertices are called **multiple edges**.



- The degree of a vertex is the number of edges attached to the vertex. The degree of a vertex is denoted $\deg(V)$. For example, in the graph below, $\deg(A) = 4$, $\deg(B) = 2$, $\deg(C) = 4$ and $\deg(D) = 2$.

  Note: A loop contributes two degrees to a vertex because a loop is attached to its vertex at both ends.



- In any graph, the *sum of degrees* of the vertices is equal to *twice the number of edges.*

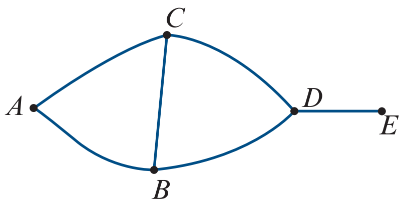
  Note: A *loop* contributes *one edge* to the graph.

Describing graphs

Graphs that represent connections between objects can take different forms and have different features. This means that there is a variety of ways to describe these graphs.

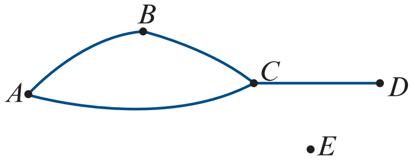
Simple graphs

Simple graphs do not have any loops. There are no duplicate or *multiple edges* either.



Disconnected graphs

A graph is connected if there is a path between each pair of vertices. A graph becomes **disconnected** if there is a vertex that is not connected to another vertex by an edge. That is, a graph is disconnected if it has at least one pair of vertices between which there is no path.

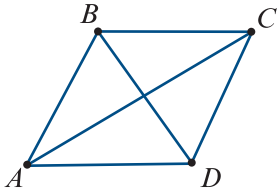


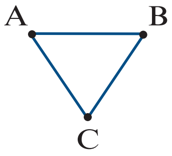
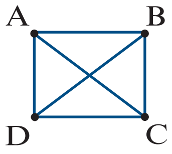
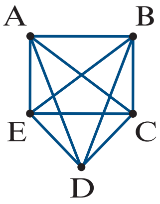
This graph is disconnected because vertex E is not connected to any other vertex by an edge.

Complete graphs

A **complete graph** is a simple graph in which every vertex is joined to every other vertex by an edge. A complete graph with n vertices is denoted by K_n.

The number of edges in a complete graph can be found using the rule $K_n = \frac{n(n-1)}{2}$.

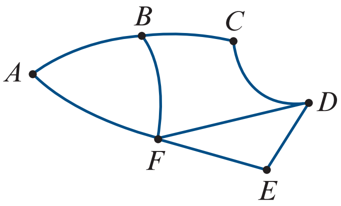


Complete graph	K_2	K_3	K_4	K_5
				
Vertices	2	3	4	5
Edges	1	3	6	10

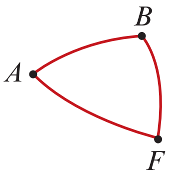
Note: In a complete graph each vertex has a degree that is one less than the total number of vertices.

Subgraphs

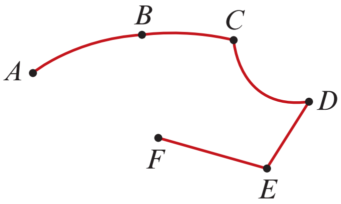
A **subgraph** is a part of a larger graph. All of the edges and vertices in the subgraph must exist in the original graph. If there are extra edges or vertices, the graph will not be a subgraph of the larger graph.



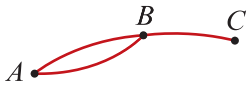
Graph 1



Graph 2



Graph 3



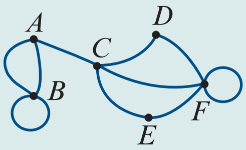
Graph 4

Graphs 2 and 3 above are subgraphs of graph 1. All of the vertices and edges in graphs 2 and 3 exist in graph 1.

Graph 4 above is not a subgraph of graph 1. There are two edges connecting vertex A to vertex B, but in graph 1 there is only one.

Example 1 Graph elements

A connected graph is shown on the right.



a What is the degree of vertex C?

b Which vertices have a loop?

c What is the degree of vertex F?

d Draw a subgraph of this graph that involves only vertices A, B and C.

Solution

a Count the number of times an edge connects to vertex C. There are four connections.

The degree of vertex C is 4.
deg(C) = 4

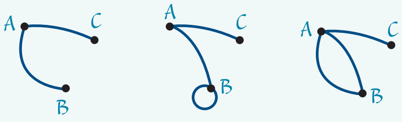
b A vertex has a loop if an edge connects it to itself.

Vertex B and vertex F have loops.

c Count the number of times an edge connects to vertex F. Remember that a loop counts as two degrees.

The degree of vertex F is 5.
deg(F) = 5

d There are a few possible answers for this question. Some are shown on the right.



Game of Sprouts

Rules for the game of Sprouts

Two or more points are drawn on a piece of paper. These are graph vertices. Players then take turns adding edges according to the following rules:

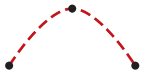
1. Each edge must join two vertices or itself.
2. Every time a new edge is drawn, a new vertex must be added on the edge.
3. Edges cannot cross nor pass through a vertex.
4. No vertex may have a degree greater than 3.
5. The last player able to add a new edge wins.

For more information see: http://cambridge.edu.au/redirect/?id=5922

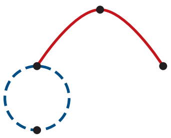
A sample game of Sprouts is played out below.

Sample game of Sprouts

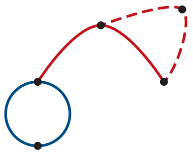
The starting vertices



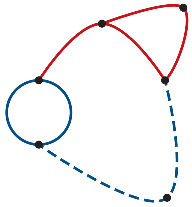
Step 1: Player 1



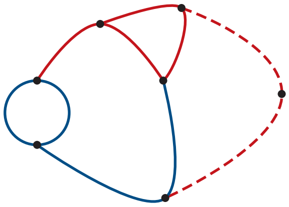
Step 2: Player 2



Step 3: Player 1



Step 4: Player 2



Step 5: Player 1 wins

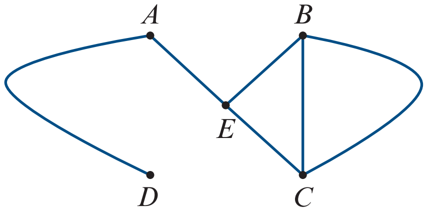
Player 1 wins because Player 2 cannot draw in a new edge without creating a vertex of degree greater than 3.

Exercise 6B

Graph elements and definitions

1 For each graph shown, complete the associated statements by filling in the boxes.

a



i The graph has ☐ vertices.

ii The graph has ☐ edges.

iii The graph has ☐ loops.

iv $\deg(A) =$ ☐

v $\deg(E) =$ ☐

vi The graph has ☐ odd vertices.

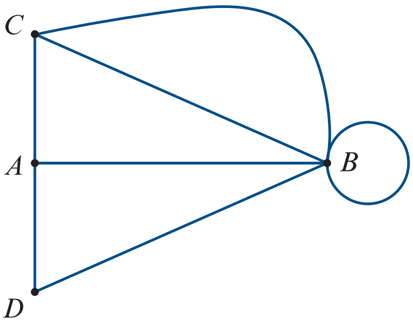
vii The graph has ☐ even vertices.

ISBN 978-1-009-25917-0

Cambridge University Press

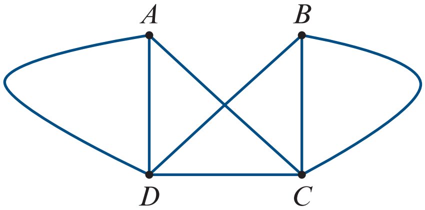
Updated January 2024

b



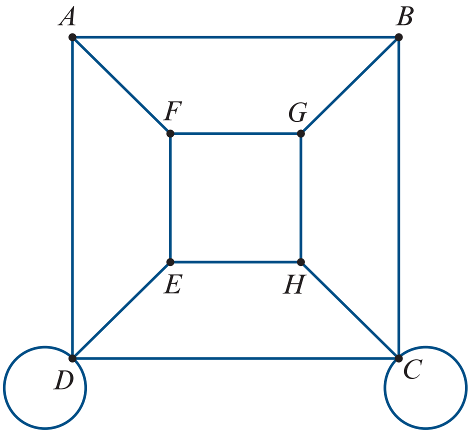
i The graph has ☐ vertices.
ii The graph has ☐ edges.
iii The graph has ☐ loops.
iv $\deg(B) =$ ☐
v $\deg(D) =$ ☐
vi The graph has ☐ odd vertices.
vii The graph has ☐ even vertices.

c



i The graph has ☐ vertices.
ii The graph has ☐ edges.
iii The graph has ☐ loops.
iv $\deg(A) =$ ☐
v $\deg(C) =$ ☐
vi The graph has ☐ odd vertices.
vii The graph has ☐ even vertices.

d



i The graph has ☐ vertices.
ii The graph has ☐ edges.
iii The graph has ☐ loops.
iv $\deg(C) =$ ☐
v $\deg(F) =$ ☐
vi The graph has ☐ odd vertices.
vii The graph has ☐ even vertices.

2 What is the sum of the degrees of the vertices of a graph with:

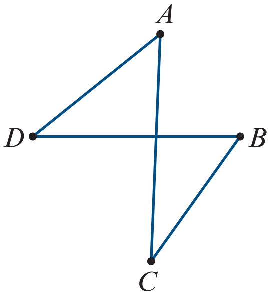
a five edges? **b** three edges? **c** one edge?

In each case draw an example of the graph and check your answer.

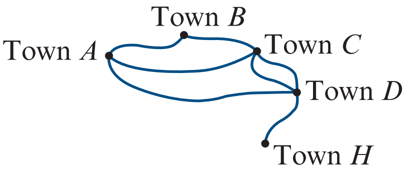
3 Why do you think that the sum of the vertex degrees of a graph will always equal twice the number of edges?

4 Consider the graph opposite.
A loop is added at vertex A:

a how will this change the degree of vertex A?
b how many edges are added to the graph?



5 This section of a road map can be considered as a graph, with towns as vertices and the roads connecting the towns as edges.



a Give the degree of:

i Town A **ii** Town B **iii** Town H

b What is the sum of the degrees of all the vertices of this graph?

c Draw a subgraph of this road map that contains only towns H, D and C.

6 Draw a graph that:

a has three vertices, two of which are odd

b has four vertices and five edges, one of which is a loop

c has six vertices, two of which are odd, and contains a subgraph that is a triangle

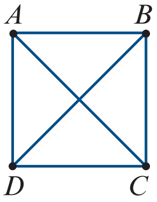
6C Connected graphs and adjacency matrices

Connected graphs and bridges

Connected graphs

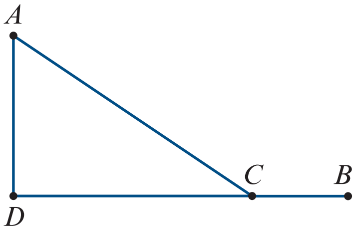
So far, all the graphs we have encountered have been *connected*. In a **connected graph**, every vertex is connected to every other vertex either directly or via another vertex. That is, every vertex in the graph can be reached from every other vertex in the graph.

The three graphs shown below are all connected.



Graph 1

Graph 2



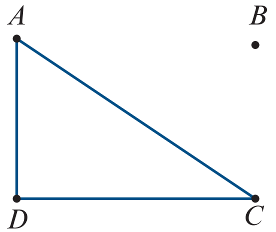
Graph 3

The graphs are connected because, starting at any vertex, say A, you can always find a path along the edges of the graph to take you to every other vertex. For example:

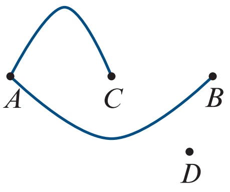
- In Graph 1 you can get directly from vertex A to vertex B by travelling along edge AB. A similar statement can be made about all the other vertices.

- In Graph 2 you can get from vertex A to vertex B indirectly via vertices C and D. All other vertices are accessible from vertex A in a similar manner.
- In Graph 3 you can get from vertex A to vertex B indirectly by travelling via vertex C, or via vertices D and C. All other vertices are directly accessible from vertex A.

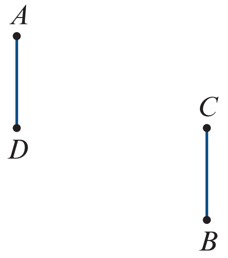
However, the three graphs below are *disconnected*, because there is not a path along the edges that connects vertex A (for example) to every other vertex in the graph.



Graph 1



Graph 2

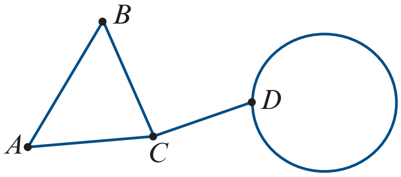


Graph 3

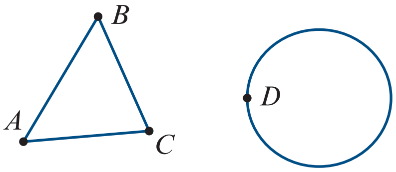
Bridges

Connected graphs have applications in a range of problems such as planning airline routes, communication systems and computer networks, where a single missing connection can lead to an inoperable system. Such critical connections are called bridges.

A **bridge** is an edge in a connected graph that, if removed, leaves the graph disconnected. In Graph 1 below, edge CD is a bridge because removing CD from the graph leaves it disconnected, see Graph 2.



Graph 1



Graph 2

Connected graphs and bridges

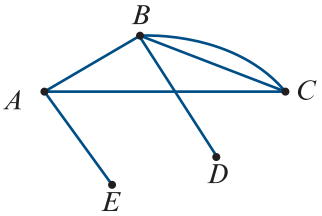
- A graph is *connected* if every vertex in the graph is accessible from every other vertex in the graph along a path formed by the edges of the graph.
- A *bridge* is a single edge in a connected graph that if removed leaves the graph disconnected. A graph can have more than one bridge.

Adjacency matrices

A **matrix** can be used to summarise the information in a graph. A matrix that records the number of connections between vertices of a graph is called an **adjacency matrix**.

ISBN 978-1-009-25917-0

A graph and the adjacency matrix for that graph are shown below.



$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 2 & 1 & 0 \\ 1 & 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

The adjacency matrix has:

- five rows and five columns, one for each vertex in the graph
- row and column labels that match the vertices in the graph, A, B, C, D, E
- a '0' in the intersection of row A and column D because no edge connects A to D
- a '0' in the intersection of row A and column A because there is no edge connecting A to itself; that is, there is no loop at vertex A
- a '1' in the intersection of row A column B because there is one edge connecting A to B
- a '2' in the intersection of row C and column B because two edges connect C to B.

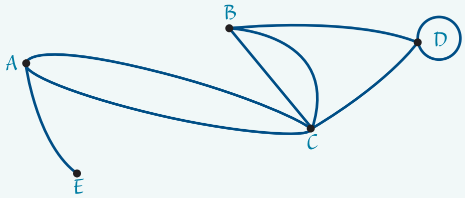
The number of edges between every other pair of vertices in the graph is recorded in the adjacency matrix in the same way.

Example 2 Drawing a graph from an adjacency matrix

Draw the graph that has adjacency matrix.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}\begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 0 & 0 & 2 & 0 & 1 \\ 0 & 0 & 2 & 1 & 0 \\ 2 & 2 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

Solution



1. Draw a dot for each vertex and label A to E.
2. There is a '2' in the intersection of row A and column C. This means there are two edges connecting vertex A and vertex C. Add these to the graph.
3. Note the '1' in the intersection of row D and column D. This shows that there is a loop at vertex D.
4. Look at every intersection of row and column and add edges to the graph, if they do not already exist.

Note: This graph is drawn as a planar graph (see page 211), but this is not strictly necessary unless required by the question.

Adjacency matrices

The adjacency matrix A of a graph is an $n \times n$ matrix in which, for example, the entry in row C and column F is the number of edges joining vertices C and F.

A loop is a single edge connecting a vertex to itself.

Loops are counted as one edge.

For all undirected graphs, the adjacency matrix will be symmetrical either side of the leading diagonal.

A directed graph (also called a **digraph**) is a network that has edges which only flow in one direction. That is, it might be possible to go from A to B but not B to A. In these types of graphs, the adjacency matrix will not be symmetrical.

An adjacency matrix can also be used to find the number ways to go between pairs of vertices within a given network. For example, the adjacency matrix, M, shows all direct connections of length 1 between each of the vertices (i.e. the number of way to go directly between two vertices). Squaring the adjacency matrix produces M^2 and gives the walks of length 2 between each of the vertices. That is, the number of ways to go between two vertices via another vertex. You may want to review how to square matrices using your CAS calculator to help with this process.

$$M = \begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array} \begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 2 & 1 & 0 \\ 1 & 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array} \rightarrow M^2 = \begin{bmatrix} 3 & 2 & 2 & 1 & 0 \\ 2 & 6 & 1 & 0 & 1 \\ 2 & 1 & 5 & 2 & 1 \\ 1 & 0 & 2 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{bmatrix}$$

In the above matrices it can be seen that:

- There are no ways to go directly from vertex A to vertex D
- There is 1 way to go from vertex A to vertex D via another vertex (i.e. 1 walk of length 3, A – B – D)
- There are 3 ways to go from vertex A back to vertex A via another vertex (i.e. 3 walks of length 2 between vertex A and vertex A)

This has many real-world applications including finding the number flights that go between two cities with one layover or determining the number of ways a someone may send a text message to another person via a friend.

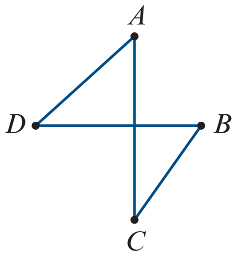
ISBN 978-1-009-25917-0

Exercise 6C

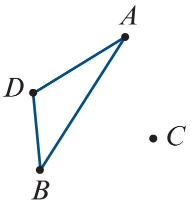
Connected graphs and bridges

1 Which of the following graphs are connected?

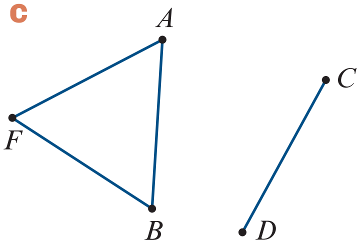
A



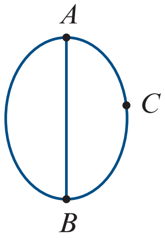
B



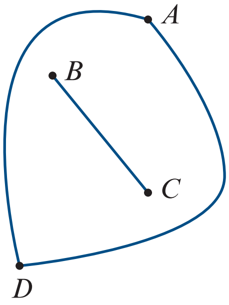
C



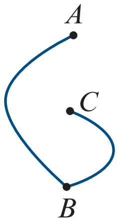
D



E



F



2 Draw a connected graph with:

- **a** three vertices and three edges
- **b** three vertices and five edges
- **c** four vertices and six edges
- **d** five vertices and five edges

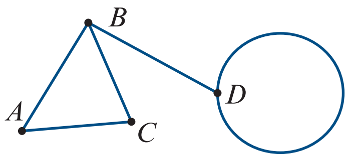
3 Draw a graph that is *not* connected with:

- **a** three vertices and two edges
- **b** four vertices and three edges
- **c** four vertices and four edges
- **d** five vertices and three edges

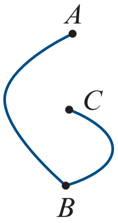
4 What is the smallest number of edges that can form a connected graph with four vertices?

5 Identify the bridge (or bridges) in the graphs below.

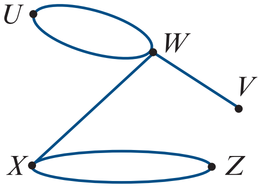
a



b



c



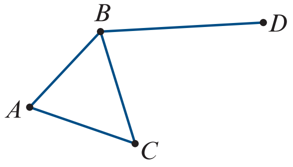
6 Draw a graph with four vertices in which every edge is a bridge.

ISBN 978-1-009-25917-0

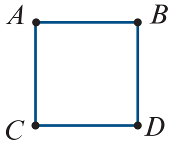
Writing adjacency matrices

7 For each of the following graphs, write down the adjacency matrix.

a

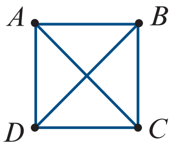


b

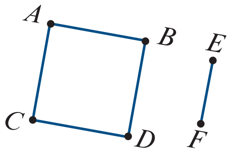


c

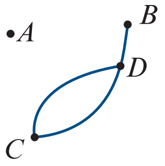
d



e



f



Drawing graphs from adjacency matrices

Example 2

8 Draw a graph from each of the following adjacency matrices.

a $\begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \end{array}$

b $\begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \end{array}$

c $\begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{bmatrix} 0 & 1 & 2 & 1 \\ 1 & 0 & 1 & 1 \\ 2 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix} \end{array}$

Miscellaneous

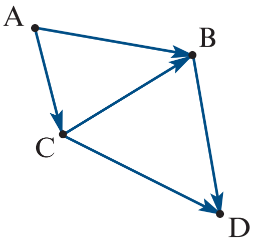
9 The adjacency matrix on the right has a row and column for vertex C that contains all zeroes. What does this tell you about vertex C?

$$\begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array}$$

10 Every vertex in a graph has one loop. What feature of the adjacency matrix would tell you this information?

11 A graph has five vertices: A, B, C, D and E. It has no duplicate edges and no loops. If this graph is complete, write down the adjacency matrix for the graph.

12 The network diagram to the right shows the bus routes that run between four cities, forming a network. The arrows indicate the direction the buses travel.



a Construct an adjacency matrix from the given digraph.

b How many ways can you travel from city A to city D?

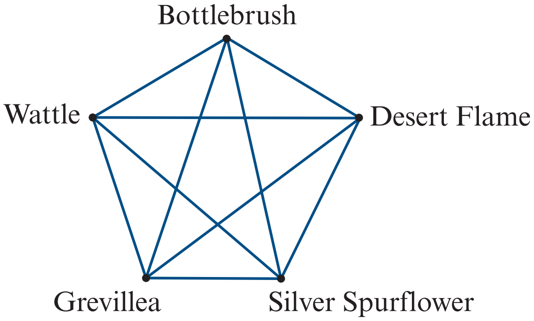
c What about these bus routes does not make sense?

13 A group of friends follow each other on a particular social media platform.
Wayne follows Oliver, Megan and Cedric
Oliver follows Megan and Cedric
Megan follows Cedric
Hannah follows Megan
Cedric does not follow anyone

a Draw a digraph to represent these interactions.

b Hannah wants to send a message to Cedric. If the platform only allows messages to be sent from those who are following another, how many ways can Hannah send her message to Cedric? State the people through whom the message will be sent.

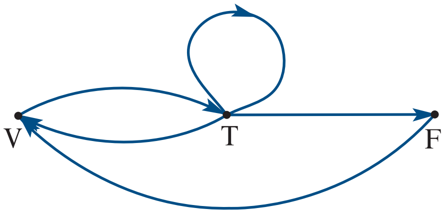
14 As a warm-up to their season, a cricket team plays a series of round-robin matches against other clubs in the league. Given that there are five teams in total, and each team only plays another once, then the games played can be represented by the network shown right.



a Construct an adjacency matrix showing all matches played in the round-robin.

b In one of the rounds the Silver Spurflower's are unable to play their match against Wattle whilst Bottlebrush forfeit to Grevillea. Instead, Wattle and Grevillea play against each other for that round. Adjust the adjacency matrix to show the changes in these matches.

15 A hiking track has various landmarks that are connected by marked paths. The landmarks are the Eagle View (V), Tunnel (T) and Falls (F). To ensure hiker's safety the paths are clearly marked and can only be accessed and walked in the direction given on the directed graph as shown on right.



a Construct an adjacency matrix from the given digraph.

b A hiker starts walking from the Eagle View. How many different routes can the hiker take, such that they end back where they started?

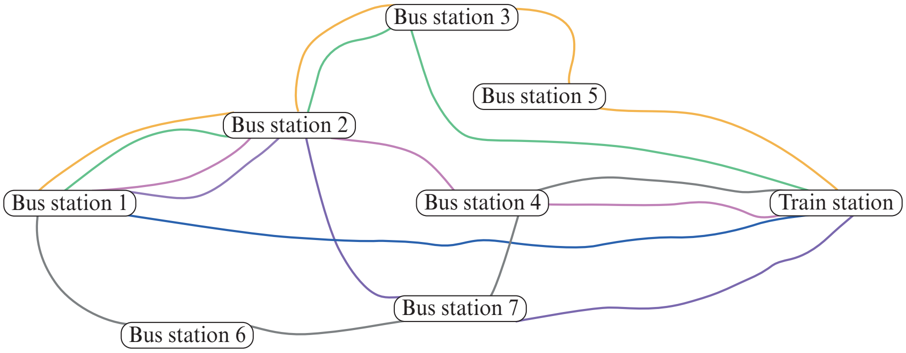
16 A university campus has a shuttle bus service to help students get around the campus. The shuttle bus operates between the following landmarks: Student Accommodation (A), Chancellery (C), Science (S), Maths Bus Stop (M), Design Bus Stop (D), Biomedical Centre (B).

The shuttle bus has 4 distinct routes as follows:

Route 1: A – S – B – D
Route 2: A – B – M – D
Route 3: A – M – D
Route 4: A – M – C – D

Construct one digraph that conveys all these routes. Note that the same edge may be used in different bus routes.

17 An efficient public transport system is needed in urban developments to alleviate traffic congestion. The Public Transport Authority is planning a new public transport system which connects seven bus stations to the train station within a residential area as per the mud map below.



a Construct an adjacency matrix from the given graph above to show the direct connections between the seven bus stations and the train station.

b How many routes go between Bus Station 2 and the Train Station; given,

  i that you do not want to change bus routes at each station?

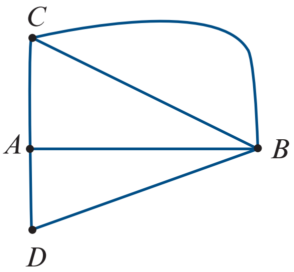
  ii that you can change bus routes at each station, but do not want to repeat bus stations?

c With use of your CAS calculator or otherwise, construct an adjacency matrix that shows the number of walks of length 2 between each of the stations.

6D Planar graphs and Euler's formula

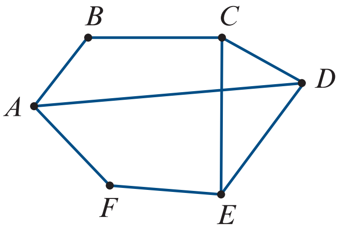
Planar graphs

A **planar graph** can be drawn on a plane (page surface) so that no edges intersect (cross), except at the vertices.

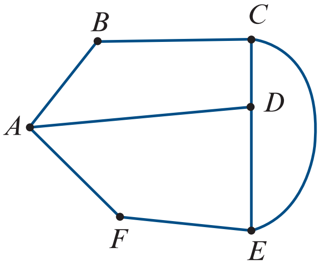


A planar graph: no intersecting edges

Some graphs do not initially appear to be planar; for example, Graph 1 shown below left. However, Graph 2 (below right) is equivalent to Graph 1. Graph 2 is clearly planar.



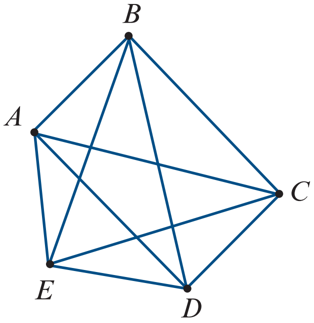
Graph 1: non-planar graph as drawn



Graph 2: planar form of Graph 1

Not all graphs are planar.

For example, the graph opposite cannot be redrawn in an equivalent planar form, no matter how hard you try.



Non- planar gra ph

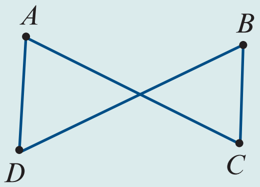
ISBN 978-1-009-25917-0

Cambridge University Press

Updated January 2024

Example 3 Redrawing a graph in planar form

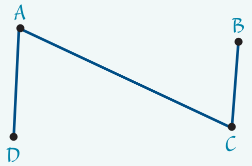
Redraw the graph shown opposite in a planar form.



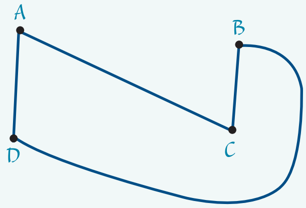
Solution (there are others)

1 Redraw the graph with edge *DB* removed.

Note: We have removed edge *DB* because it intersects edge *AC*.

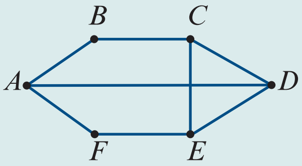


2 Replace edge *DB* as a curved line that avoids intersecting with the other three edges. The graph is now in an equivalent planar form: no edges intersect, except at vertices.



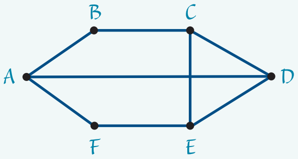
Example 4 Redrawing a graph in planar form

Show that this graph is planar by redrawing it so that no edges cross.

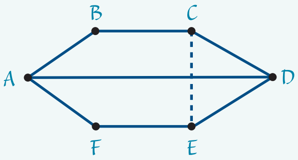


Solution

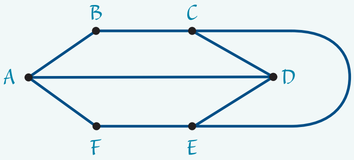
1 Choose one of the edges that crosses over another edge.



2 Remove it temporarily from the graph.



3 Redraw the edge between the same vertices but without crossing over another edge.



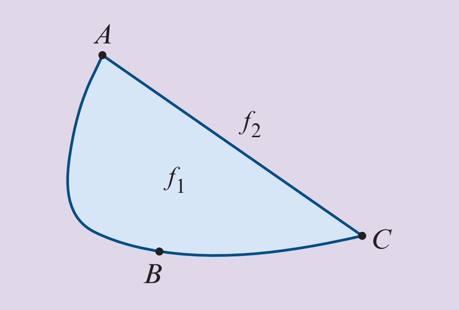
Faces of a graph

The graph opposite can be regarded as dividing the paper it is drawn on into two regions.

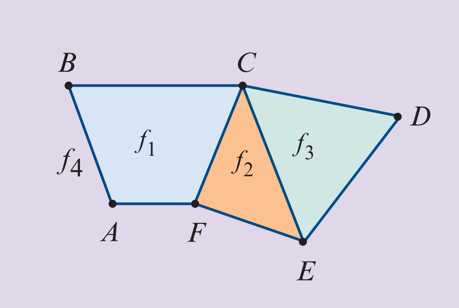
In the language of graphs, these regions are called the **faces** of the graph.

One face, f_1, is bounded by the graph.

The other face, f_2, is the region surrounding the graph. This ‘outside’ face is infinite.



The graph opposite divides the paper into four regions, so we say that it has four faces: f_1, f_2, f_3 and f_4. Here f_4 is an infinite face.



Euler’s formula

Euler discovered that, for connected planar graphs, there is a relationship between the *number of vertices*, v, the *number of edges*, e, and the *number of faces*, f. This relationship can be expressed in words as:

number of vertices − number of edges + number of faces = 2

or in symbols as:

$$v - e + f = 2$$

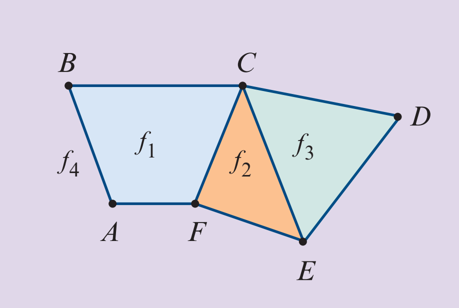
This is known as **Euler’s formula**.

For example, for the graph opposite:

$$v = 6, \ e = 8 \text{ and } f = 4.$$

So $\quad v - e + f = 6 - 8 + 4 = 2$

confirming Euler’s formula.



Euler’s formula

For a connected planar graph:

number of vertices − number of edges + number of faces = 2

or

$$v - e + f = 2$$

where v = number vertices, e = number of edges and f = number of faces.

ISBN 978-1-009-25917-0

Cambridge University Press

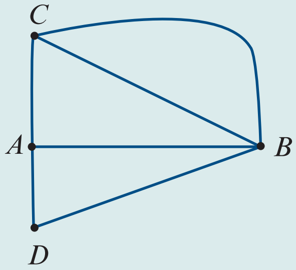
Updated January 2024

Example 5 Verifying Euler's formula

Consider the connected planar graph shown.

a Write the number of vertices, v, the number of edges, e, and the number of faces, f.

b Verify Euler's formula.



Solution

a **1** There are four vertices: A, B, C, D, so $v = 4$.

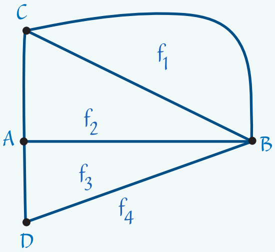
Number of vertices: $v = 4$

2 There are six edges: $AB, AC, AD, BC\ (\times 2)$ and BD, so $e = 6$.

Number of edges: $e = 6$

3 There are four faces, so $f = 4$.

Tip: Mark the faces on the diagram. Do not forget the infinite face f_4 that surrounds the graph.



Number of faces: $f = 4$

b **1** Write Euler's formula.

Euler's formula: $v - e + f = 2$

2 Substitute the values of v, e, and f. Evaluate.

$v - e + f = 4 - 6 + 4 = 2$

3 Write your conclusion.

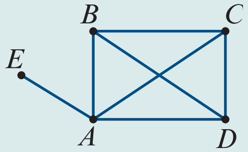
$\therefore$ Euler's formula is verified.

Example 6 Verifying Euler's formula

For the graph shown on the right:

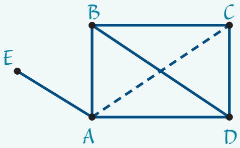
a redraw the graph into planar form

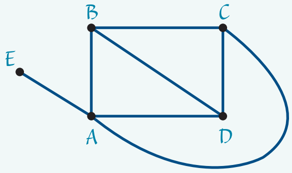
b verify Euler's formula for this graph



Solution

a Temporarily remove an edge that crosses another edge and redraw it so that it does not cross another edge.





ISBN 978-1-009-25917-0

b Count the number of vertices, edges and faces.

In the planar graph there are five vertices, seven edges and four faces.

$v - e + f = 5 - 7 + 4 = 2$

Euler's formula is verified.

Example 7 Using Euler's formula

A connected planar graph has six vertices and nine edges. How many faces does the graph have? Draw a connected planar graph with six vertices and nine edges.

Solution

a Write the known values.

$v = 6 \quad e = 9$

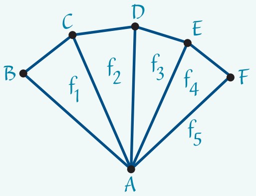
b Substitute into Euler's formula and solve for the unknown value.

$$v - e + f = 2$$
$$6 - 9 + f = 2$$
$$f - 3 = 2$$
$$f = 5$$

This graph has five faces, labelled f_1, f_2, f_3, f_4 and f_5.

c Sketch the graph.

Note: There are other possible graphs.



Example 8 Using Euler's formula to find the number of faces of a graph

A connected planar graph has four vertices and five edges. Find the number of faces.

Solution

1 Write the value of v and e.

$v = 4, e = 5$

2 Write Euler's formula.

Euler's formula:

$v - e + f = 2$

3 Substitute the values of v and e.

$$4 - 5 + f = 2$$
$$-1 + f = 2$$

4 Solve for f.

$$f = 3$$

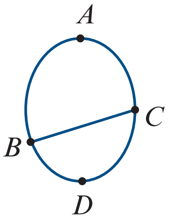
5 Write your answer.

The graph has three faces.

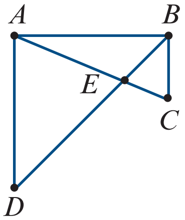
Exercise 6D

1 Which of the following graphs are drawn in planar form?

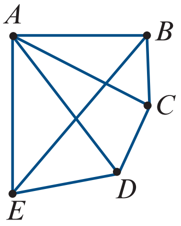
A



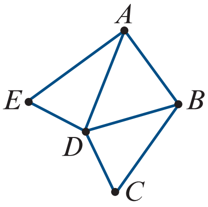
B



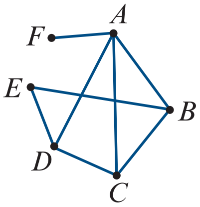
C



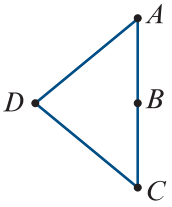
D



E



F

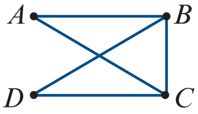


Redrawing planar graphs

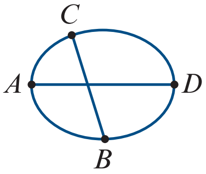
Example 3
Example 4

2 Redraw each graph in an equivalent planar form.

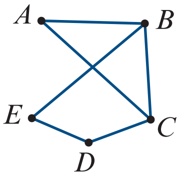
a



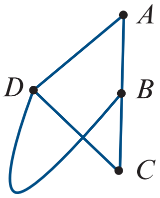
b



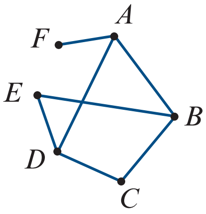
c



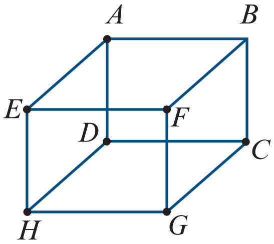
d



e



f



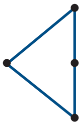
Verifying Euler's formula

Example 5

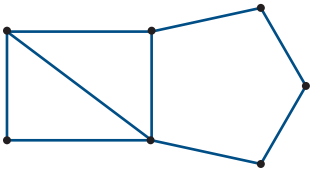
3 For each of the following graphs:

i state the values of v, e and f **ii** verify Euler's formula.

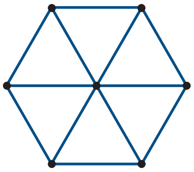
a



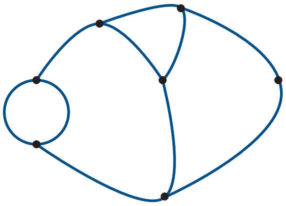
b



c



d



ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

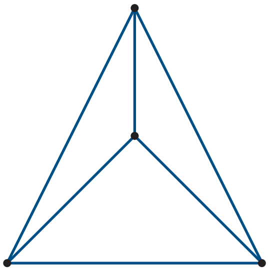
Using Euler's formula

Example 8

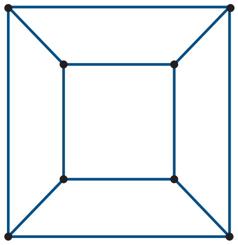
4 For a planar connected graph, find:

- **a** f given $v = 4$ and $e = 4$
- **b** v given $e = 3$ and $f = 2$
- **c** e given $v = 3$ and $f = 3$
- **d** v given $e = 6$ and $f = 4$
- **e** f given $v = 4$ and $e = 6$
- **f** f given $v = 6$ and $e = 11$
- **g** e given $v = 10$ and $f = 11$

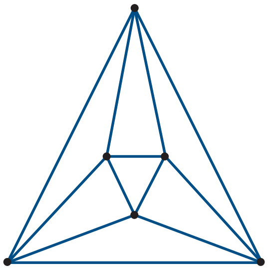
5 The five graphs shown below are known as the Platonic (after Plato) solids.



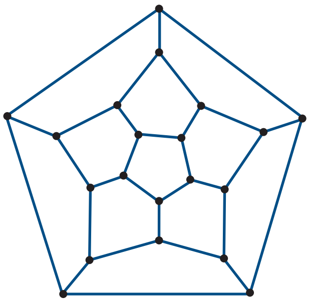
Graph 1 Tetrahedron



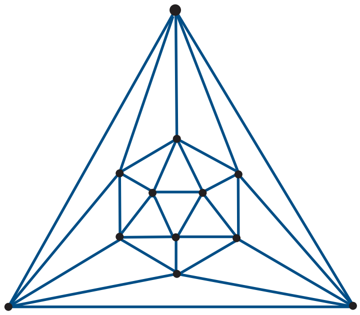
Graph 2 Cube



Graph 3 Octahedron



Graph 4 Dodecahedron



Graph 5 Icosahedron

For each graph, write the values of v, e and f and show that they satisfy Euler's formula.

6E Traversing a graph

Traversing a graph means starting at a particular vertex and then moving through the graph along the edges and passing through vertices to finish at another vertex.

Graphs can be used to model and analyse many practical problems. For example, a courier driver would like to know the shortest route to use for deliveries, and a tour guide would like to know the quickest route that allows tourists to see a number of sights without retracing their steps.

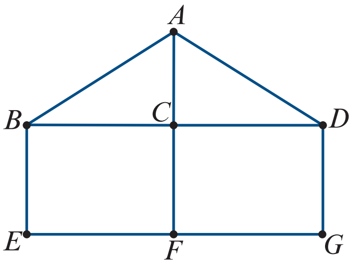
To solve these types of problems, you will need to learn the language we use to describe the different ways of navigating through a graph, from one vertex to another.

ISBN 978-1-009-25917-0

Walks, trails, paths, and cycles

The different ways of navigating through graphs, from one vertex to another, are described as *walks, trails, paths,* and *cycles*.

The graph opposite will be used to explain and define each of these terms.

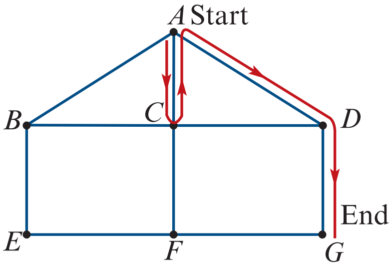


Walks

A **walk** is a sequence of edges, linking successive vertices in a graph.

A walk starts at one vertex and follows any route to finish at any other vertex.

The red line in the graph opposite traces out a walk. This walk can be written down by listing the vertices in the order they are visited: $A–C–A–D–G$.



- A walk that starts and finishes at different vertices is said to be an open walk.
- A walk that starts and finishes at the same vertex is said to be a closed walk.

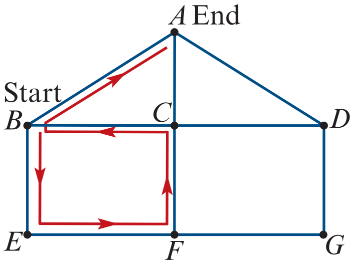
Trails

A **trail** is a walk with no repeated edges.

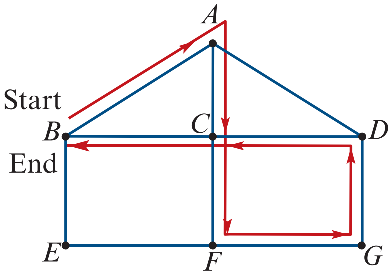
The red line in the graph below left traces out a trail. The trail in the bottom left can be written down by listing the vertices in the order they are visited: $B–E–F–C–B–A$.

Note: There are *no repeated edges* in this trail, but one vertex (B) is repeated.

- A trail that starts and ends at the same vertex is called a closed trail.



open trail



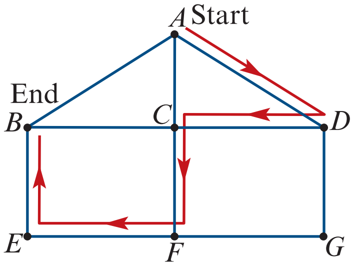
closed trail

Paths

A **path** is a walk with no repeated edges and no repeated vertices.

The red line in the graph opposite traces out a path. This path can be written down by listing the vertices in the order they are visited: $A–D–C–F–E–B$.

- A path that starts and finishes at different vertices is said to be open, while a path that starts and finishes at the same vertex is said to be closed.

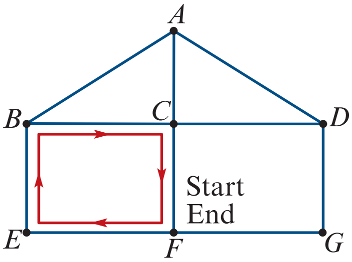


Cycles

A **cycle** is a path (no repeated edges, no repeated vertices) that starts and ends at the same vertex. The start and end vertex is an exception to repeated vertices. Cycles are also called *closed paths*.

The red line in the graph opposite traces out a cycle. This cycle can be written down by listing the vertices in the order they are visited: $F–E–B–C–F$.

Note: There are *no repeated edges* and *no repeated vertices* in this cycle, except for the start and end vertices.



Walk

A *walk* is a sequence of edges, linking successive vertices, that connects two different vertices in a graph.

Trail

A *trail* is a walk with no repeated edges.

Path

A *path* is a walk with no repeated edges or vertices, except when it is a closed path which starts and ends at the same vertex.

Note: Because a path has no repeated vertices, there can be no repeated edges.

Cycle

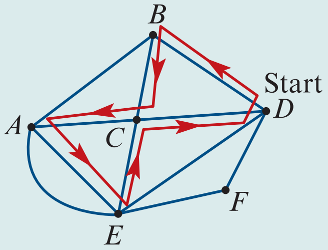
A *cycle* is a closed path which begins and ends at the same vertex and which has no repeated edges or vertices except the first.

Note: Because a cycle has no repeated vertices, there can be no repeated edges.

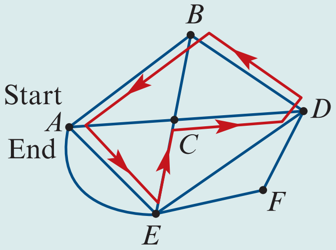
Example 9 **Identifying types of walks**

Identify the walk in each of the graphs below as a path, trail (open or closed), cycle or walk only.

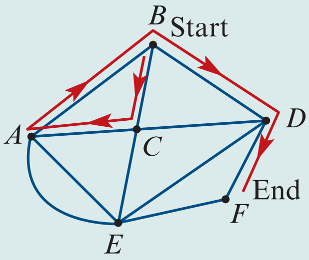
a



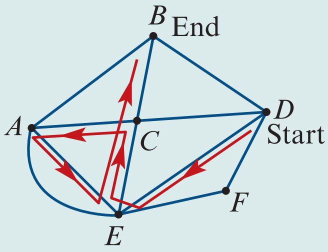
b



c



d



Solution

a This walk starts and ends at the same vertex so it is closed. The walk passes through vertex C twice without repeated edges, so it must be a closed trail.

b This walk starts and ends at the same vertex so it is closed. The walk has no repeated vertices or edges so it is a cycle.

c This walk starts at one vertex and ends at a different vertex, so it is open. There is one repeated vertex (B) and no repeated edges, so it is an open trail.

d This walk starts at one vertex and ends at a different vertex so it is open. There are repeated vertices (C and E) and repeated edges (the edge between C and E), so it must be an open walk.

Summary of the properties of walks, trails, circuits and cycles

A walk can be *any* route through a graph. Trails, paths, circuits and cycles are all types of walks that have particular properties. These properties are summarised in the table below.

Type of route	*Are repeated edges permitted?*	*Are repeated vertices permitted?*
walk (open)	yes	yes
trail (open)	no	yes
path (open)	no	no
closed walk	yes	yes
closed trail	no	yes
cycle (closed path)	no	no (except for the first and last)

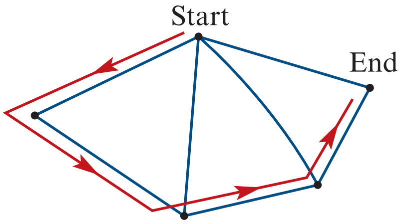
Exercise 6E

Identifying types of walks

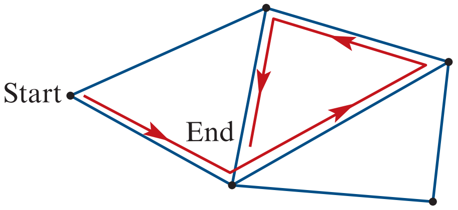
Example 9

1 Identify the walk in each of the graphs below as a path, trail (open or closed), cycle or walk only.

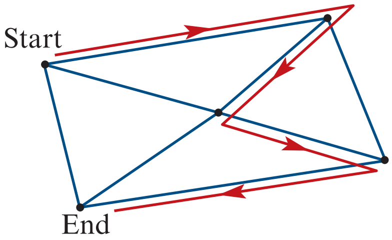
a



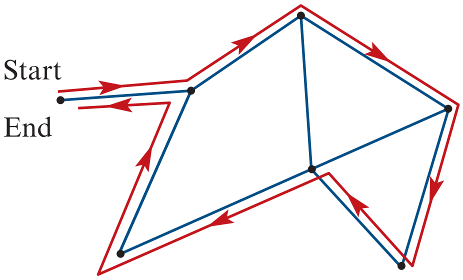
b



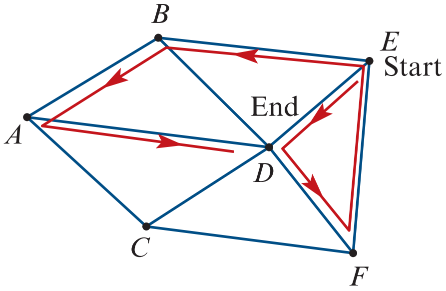
c



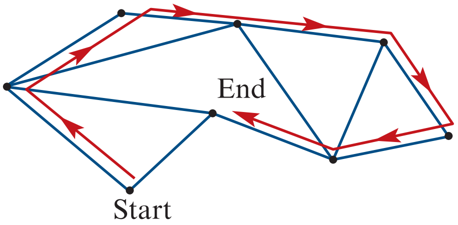
d



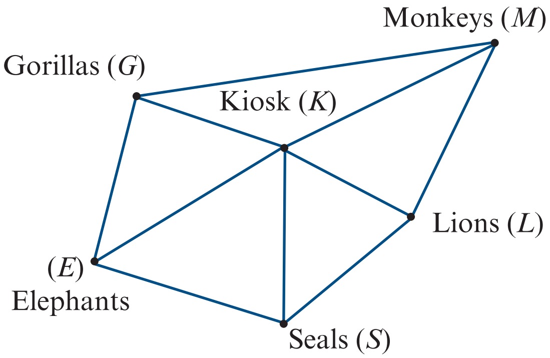
e



f



2 The graph opposite shows the pathways linking five animal enclosures in a zoo to each other and to the kiosk.



a Which of the following represents a trail in the graph?

- **i** *S-L-K-M-K*
- **ii** *G-K-L-S-E-K-M*
- **iii** *E-K-L-K*

b Which of the following represents a path in the graph?

i *K-E-G-M-L* **ii** *E-K-L-M* **iii** *K-S-E-K-G-M*

c Which of the following represents a closed trail in the graph?

i *K-E-G-M-K-L-K* **ii** *E-S-K-L-M-K-E* **iii** *K-S-E-K-G-K*

d Which of the following represents a cycle in the graph?

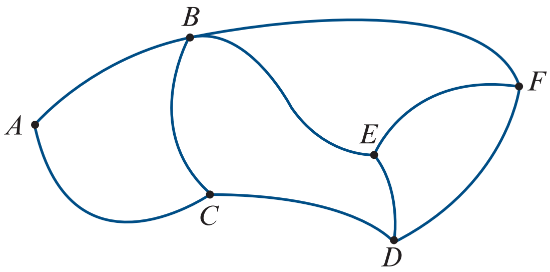
i *K-E-G-K* **ii** *G-K-M-L-K-G* **iii** *L-S-E-K-L*

ISBN 978-1-009-25917-0

Cambridge University Press

Updated January 2024

3 Using the graph opposite, identify the walks below as a path, trail (open or closed), cycle or walk only.



- **a** $A–B–E–B–F$
- **b** $B–C–D–E–B$
- **c** $C–D–E–F–B–A$
- **d** $A–B–E–F–B–E–D$
- **e** $E–F–D–C–B$
- **f** $C–B–E–F–D–E–B–C–A$

6F Eulerian graphs and Hamiltonian paths and cycles

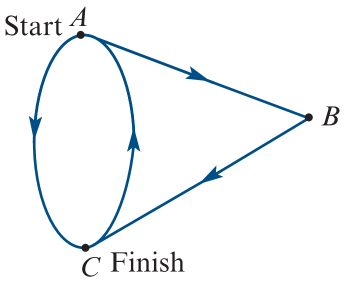
Traversable graphs

Many practical problems involve finding a trail in a graph that includes *every* edge. The Königsberg bridge problem is one such problem. Graphs that have this property are called traversable graphs.

A **traversable graph** has a trail that includes *every edge* in the graph.

The graph opposite is an example of a *traversable graph*. It has a trail that includes every edge.

The trail *A-B-C-A-C* is one such example.

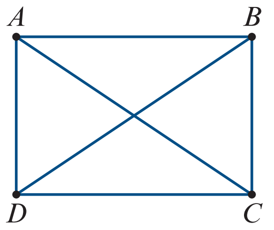


Graph 1: traversable

Not all graphs are traversable.

For example, it is impossible to find a trail in the graph opposite that includes every edge. Try it and see.

This graph is *not traversable*.



Graph 2: non-traversable

Identifying traversable graphs

One way of finding out whether a graph is traversable is to try a number of routes through the graph and see whether they work. When we do this, we are solving the problem by inspection. However, for all but the simplest graphs, this problem-solving method can become very tedious.

Fortunately, there is a more systematic problem-solving method that relates to the degree of the vertices and whether they are *odd* or *even*.

Rules for identifying traversable graphs

For a graph to be traversable, it must first be *connected*.

A connected graph is traversable if:

- *all* vertices are of *even* degree; or
- exactly *two* vertices are of *odd* degree and the rest are of even degree.

From this it follows that:

- a traversable graph will have either a trail (open or closed) or a cycle that involves the use of *every edge* in the graph
- if a graph has *more than two* vertices of odd degree, it is *not* traversable.

Using these rules, we can see why Graph 1 (previous page) is traversable: it has *exactly two odd* vertices (*A* and *C*) and the remaining vertex (*B*) is even. Likewise, Graph 2 is *not* traversable: it has four odd vertices (*A*, *B*, *C* and *D* are *all odd* vertices).

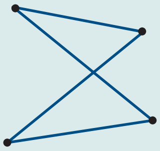
Example 10 Identifying traversable graphs

For each of the following graphs:

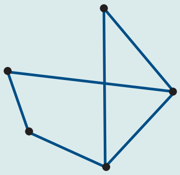
i determine whether the graph is traversable, and state why

ii if traversable, check by identifying an open or closed trail that traverses the graph.

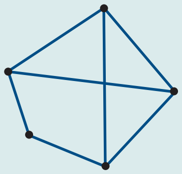
a



b

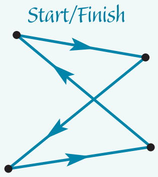


c

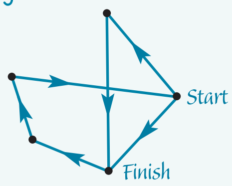


Solution

a Traversable: all even vertices. The graph has a cycle that involves every edge.



b Traversable: two odd vertices. The graph has a trail that involves every edge.



c Not traversable: more than two odd vertices

Eulerian and semi-Eulerian graphs

Because of the pioneering graph theory work completed by Euler, we describe traversable graphs as either Eulerian or semi-Eulerian.

Eulerian graphs

A graph is described as Eulerian if it has a closed trail that includes every edge once only.

To be Eulerian, a graph must:

- be connected
- have all vertices of an even degree.

An Eulerian graph can start at any vertex and will also finish at that vertex.

Semi-Eulerian graphs

A graph is described as **semi-Eulerian** if it has an open trail that includes every edge once only. The trail in an Eulerian graph is called an **Eulerian trail**.

To be semi-Eulerian, a graph must:

- be connected.
- have exactly two vertices of an odd degree.

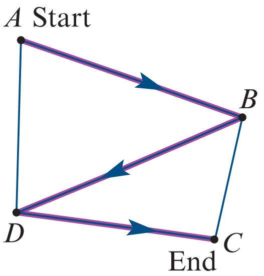
An Eulerian trail must start at one of the odd degree vertices and will finish at the other odd degree vertex.

Hint: Remember that *Eulerian* graphs are concerned with following *edges* and both begin with "*e*".

Hamiltonian paths and cycles

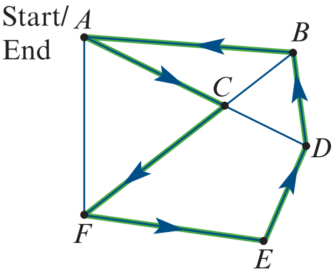
A **Hamiltonian path** involves all the vertices but not necessarily all the edges. For example, in the graph opposite, ***A***-***B***-***D***-***C*** is a Hamiltonian path. It starts at vertex *A* and ends at vertex *C*. (Follow the arrows.)

Note: A Hamiltonian path does not have to involve all edges.



A **Hamiltonian cycle** is a Hamiltonian path that starts and finishes at the same vertex. For example, in the second graph ***A***-***C***-***F***-***E***-***D***-***B***-***A*** is a Hamiltonian cycle. It starts and finishes at vertex *A*. (Follow the arrows.)

Note: A Hamiltonian cycle does not have to involve all edges.



ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

A Hamiltonian graph is a graph that contains a Hamiltonian cycle. A connected graph that contains an open Hamiltonian path, but not a Hamiltonian cycle, is said to be **semi-Hamiltonian**.

Unfortunately, unlike Eulerian trails and graphs, there are *no simple rules* for determining whether a network contains a Hamiltonian path or cycle. It is just a matter of ‘trial and error’.

Applications of Hamiltonian paths and cycles

Hamiltonian paths and cycles have many practical applications. In everyday life, a *Hamiltonian path* would apply to situations like the following:

- You plan a trip from Perth to Albany, with visits to Margaret River, Nannup, Bridgetown, Denmark and Mount Barker on the way, but do not want to visit any town more than once.

Hamiltonian cycles relate to situations like the following:

- A courier leaves her depot to make a succession of deliveries to a variety of locations before returning to her depot. She does not like to go past each location more than once.
- A tourist plans to visit all of the historic sites in a city without visiting each more than once.
- You are planning a trip from Adelaide to visit Melbourne, Canberra, Sydney, Broken Hill, Mildura and Renmark before returning to Adelaide. You don’t want to visit any town more than once.

In all these situations, there would be several suitable routes. However, other factors, such as time taken or distance travelled, may need to be taken into account in order to determine the best route. This is an issue addressed in the next section: weighted graphs and networks.

Hamiltonian paths and cycles

Hamiltonian paths

A Hamiltonian path visits every vertex of a graph.

Hamiltonian cycles

A Hamiltonian cycle is a Hamiltonian path (every vertex) that starts and ends at the same vertex.

Note: Inspection is the only way to identify Hamiltonian paths and cycles.

Remember: Eulerian trails and graphs do not repeat edges. Hamiltonian paths and cycles do not repeat vertices.

Hint: To remember the difference between Eulerian and Hamiltonian travels, remember that Eulerian refers to edges, and both start with ‘e’.

Example 11 Eulerian and Hamiltonian travel

A map showing the Victorian towns of St Andrews, Kinglake, Yarra Glen, Toolangi and Healesville is shown above.

a Draw a graph with a vertex representing each of these towns and edges representing the direct road connections between the towns.

b Explain why the graph is semi-Eulerian, but not Eulerian.

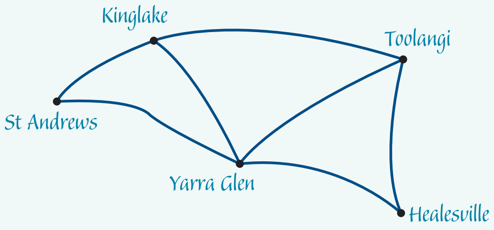
c Write an Eulerian trail that begins at Toolangi.

d Write a Hamiltonian cycle that begins at Healesville.

Solution

a A road connection exists between:

- St Andrews and Kinglake
- St Andrews and Yarra Glen
- Kinglake and Yarra Glen
- Kinglake and Toolangi
- Yarra Glen and Toolangi
- Yarra Glen and Healesville
- Healesville and Toolangi.



b The graph has two odd-degree vertices (Toolangi and Kinglake).

There are exactly two odd-degree vertices in this graph, so the graph is semi-Eulerian.

c There are a few different answers to this question. One of these is shown.

An Eulerian trail, starting at Toolangi is:
Toolangi Healesville–Yarra Glen–Toolangi–Kinglake–Yarra Glen–St Andrews–Kinglake

ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

d There are two different answers to this question. One of these is shown.

A Hamiltonian cycle that begins at Healesville is: Healesville–Yarra Glen–St Andrews–Kinglake–Toolangi–Healesville

Exercise 6F

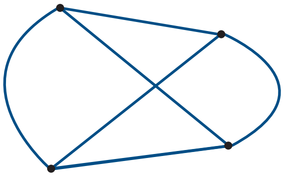
Identifying traversable graphs

Example 10

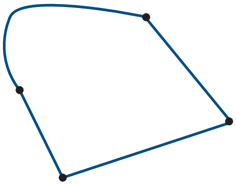
1 For each of the following graphs:

- **i** determine whether the graph is traversable, and state why
- **ii** if traversable, check by identifying a trail, path or cycle that involves every edge in the graph.

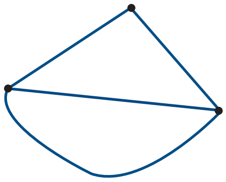
a



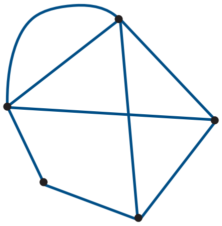
b



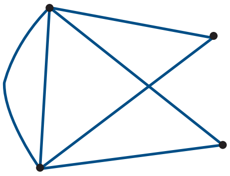
c



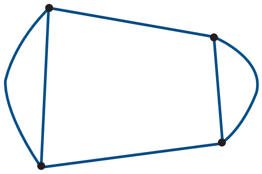
d



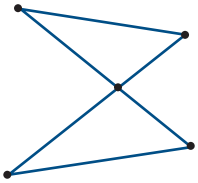
e



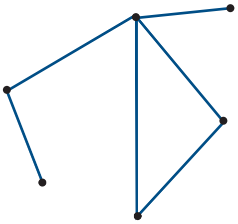
f



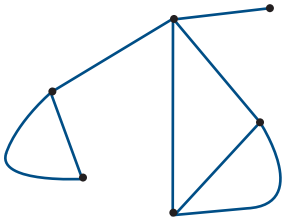
g



h



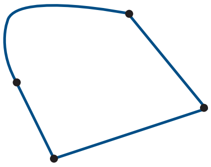
i



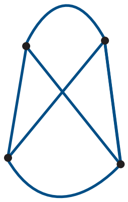
2 For each of the following graphs **a** to **i**, continued on next page:

- **i** determine whether the network is Eulerian, semi-Eulerian or neither, and state why
- **ii** if the graph is Eulerian or semi-Eulerian, show **one** example.

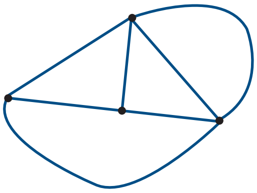
a



b

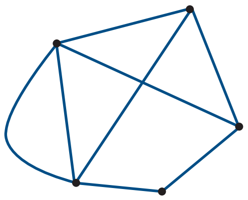


c

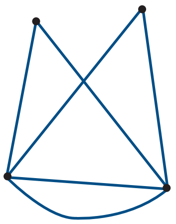


ISBN 978-1-009-25917-0

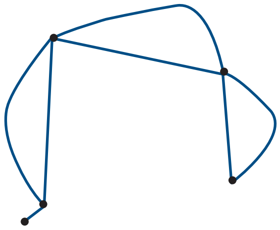
d



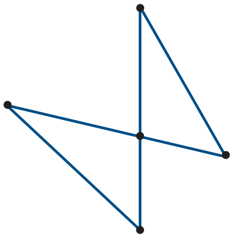
e



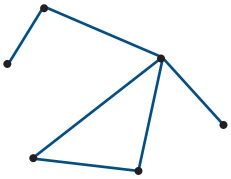
f



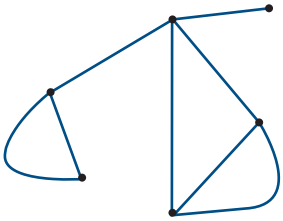
g



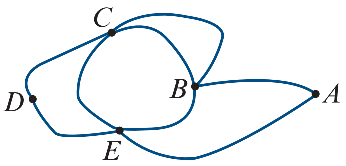
h



i



3 A road inspector lives in Town A and is required to inspect all roads connecting the neighbouring towns B, C, D and E. The network of roads is shown on the right.

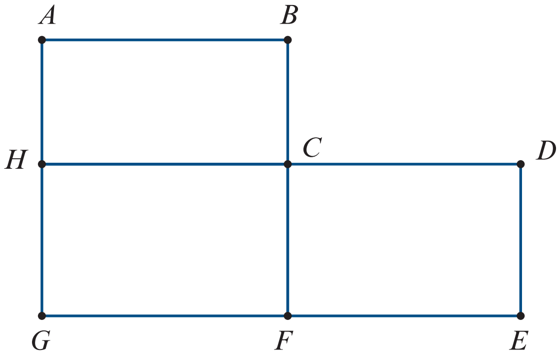


a Is it possible for the inspector to set out from Town A, carry out his inspection by travelling over every road linking the five towns only once, and return to Town A? Explain.

b Show one possible route he can follow.

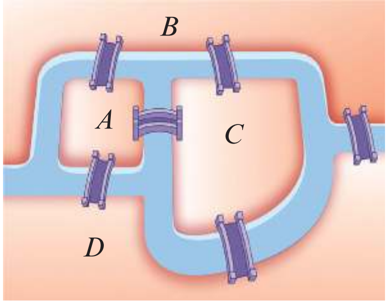
4 A postman has to deliver letters to the houses located on the network of streets shown on the right.

a Is it possible for the postman to start and finish his deliveries at the same point in the network without retracing his steps at some stage? If not, why not?



b It is possible for the postman to start and finish his deliveries at different points in the network without retracing his steps at some stage. Identify one such route.

5 Two islands are connected to the banks of a river by five bridges:



a Draw a graph to represent this situation. Label the vertices A, B, C and D to represent the river banks and the two islands. Use the edges of the graph to represent the bridges.

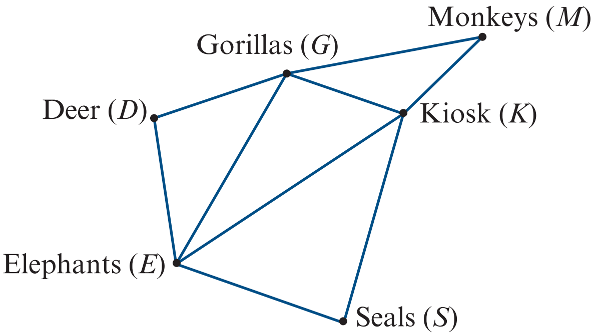
b It is not possible to plan a walking route that passes over each bridge once only. Why not?

c **i** Show where another bridge could be added to make such a walk possible.

ii Draw a graph to represent this situation.

iii Explain why it is now possible to find a walking route that passes over each bridge once only. Mark one such route on your graph.

6 The graph below models the pathways linking five animal enclosures in a zoo to the kiosk and to each other.



a Is it possible for the zoo's street sweeper to follow a route that enables its operator to start and finish at the kiosk without travelling down any one pathway more than once? If so, explain why.

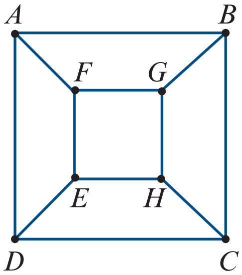
b If so, write down one such route.

Hamiltonian paths and cycles

7 List a Hamiltonian path for the network shown.

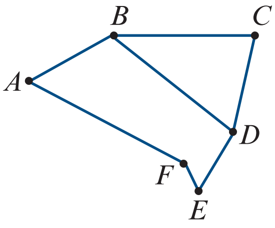
a Starting at A and finishing at D

b Starting at F and finishing at G

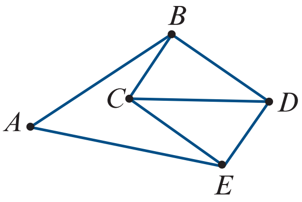


8 Identify a Hamiltonian cycle in each of the following graphs (if possible), starting at A each time.

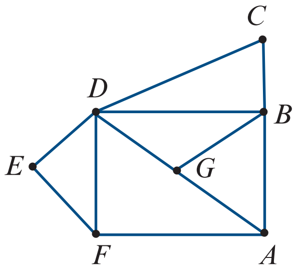
a



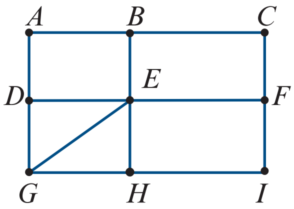
b



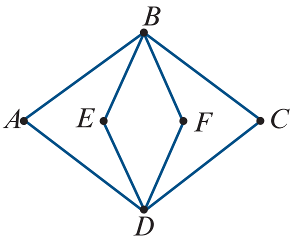
c



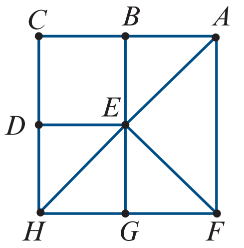
d



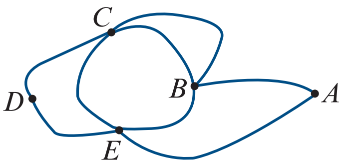
e



f



9 A tourist wants to visit a winery in each of five different towns Apsley (A), Berrigama (B), Cleverland (C), Donsley (D) and Everton (E) in a wine-growing region.



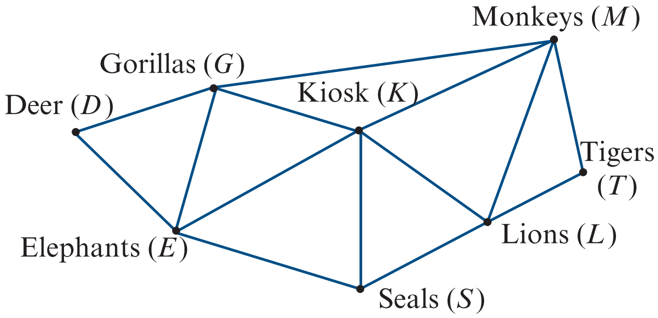
The network of roads connecting the towns is shown above. Can a tourist start a tour that visits each town only once by starting and finishing at the following towns? If so, identify one possible route in each case and give its mathematical name.

a Start at Cleverland and finish at Everton.

b Start at Cleverland and finish at Apsley.

c Start at Everton and finish at Everton.

10 The graph opposite models the pathways linking seven animal enclosures in a zoo to the kiosk and to each other.



a Is it possible for a visitor to the zoo to start their visit at the kiosk and see all of the animals without visiting any one animal enclosure more than once? If so, identify a possible route and give this route its mathematical name.

b Is it possible for a visitor to the zoo to start their visit at the deer enclosure and finish at the kiosk without visiting the kiosk or any enclosure more than once? If so, identify a possible route and give this route its mathematical name.

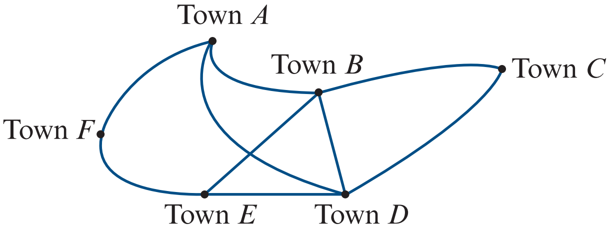
6G Weighted graphs, networks and the shortest path problem

Weighted graphs and networks

The graph opposite shows how six towns are connected by road.

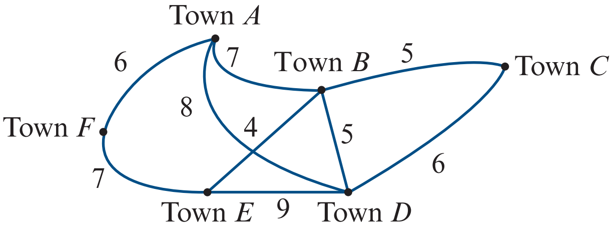
The towns are represented by the vertices of the graph.

The roads between towns are represented by the edges.



We can give more information about the situation we are representing with the graph by adding numbers to the edges.

When we do this, the resulting graph is called a **weighted graph**.



The weighted graph opposite shows the distances between the towns (in kilometres).

A **network** is a *weighted graph* that represents a real world situation. The weighted graph above could be called a network.

Weighted graphs and networks

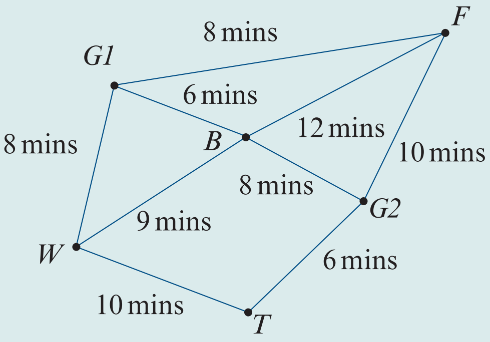
- A *weighted graph* is a graph where a number is associated with each edge. These numbers are called weights.
- A *network* is a weighted graph in which the weights are physical quantities, for example, distance, time, cost.

Example 12 Interpreting a network

The network opposite is used to model the tracks in a forest connecting a suspension bridge (*B*), a waterfall (*W*), a very old tree (*T*) and a fern gully (*F*).

Walkers can enter or leave the forest through either Gate 1 (*G1*) or Gate 2 (*G2*).

The numbers on the edges represent the times (in minutes) taken to walk directly between these places.



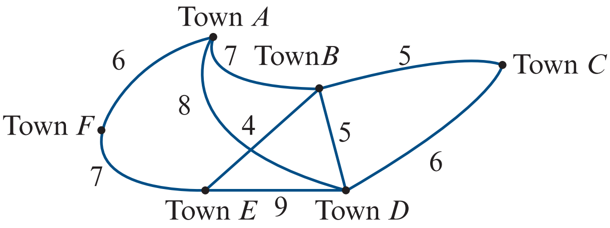
a How long does it take to walk from the bridge directly to the fern gully?

b How long does it take to walk from the old tree to the fern gully via the waterfall and the bridge?

Solution

a Identify the edge that directly links the bridge with the fern gully and read off the time.	The edge is B-F. The time taken is 12 minutes.
b Identify the path that links the old tree to the fern gully visiting the waterfall and the bridge on the way. Add up the times.	The path is T-W-B-F. The time taken is $10 + 9 + 12 = 31$ minutes.

The shortest path problem



Another question we might have when presented with a road network like the one shown is, 'What is the shortest distance between certain towns?'

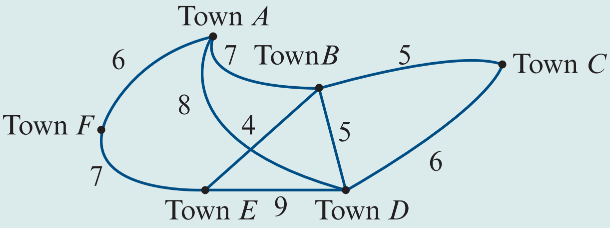
While this question is easily answered if all of the towns are directly connected by a road, for example, Town *A* and Town *B*, the answer is not so obvious if we have to travel through other towns to get there, for example, Town *F* and Town *C*.

The shortest path problem

Determining the shortest distance or time or the least cost to move around a network is called the **shortest path problem**.

Example 13 **Finding the shortest path by inspection**

Find the shortest route between Town *C* and Town *F* in the network shown right.



Solution

1 Identify all of the likely shortest routes between Town *C* and Town *F* and calculate their lengths.	C-D-E-F: The distance is $6 + 9 + 7 = 22$ km. C-B-E-F: The distance is $5 + 4 + 7 = 16$ km. C-B-A-F: The distance is $5 + 7 + 6 = 18$ km.

ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

Note: In theory, when using the 'by inspection' method to solve this problem we need to list all possible routes between Town *C* and Town *F* and determine their lengths. However, we can save time by eliminating any route that passes through any town more than once or any road more than once. We can also eliminate any route that 'takes the long way around' rather than using the direct route, for example, when travelling from Town *B* to Town *D* we can ignore the route that goes via Town *A* because it is longer.

2 Compare the different path lengths to identify the shortest path and write your answer.

The shortest path is C-B-E-F.

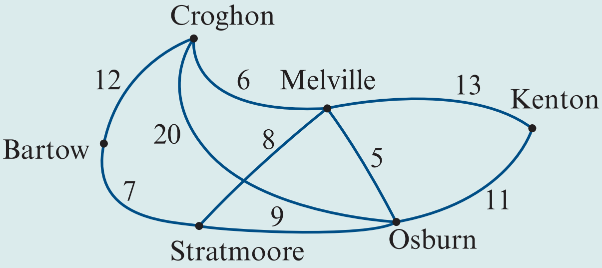
Note: You should compare the lengths of all likely paths because there can be more than one shortest path in a network.

When we know numerical information about the connections, travelling through a graph will have extra considerations. If the weights of a network represent time, we can choose a route that will allow us to travel in the shortest time. If the weights represent distance, we can determine a route that will allow us to travel the shortest distance.

These types of problems involve finding the shortest path from one vertex to another. In networks that have only a few vertices, it is often easy to find the shortest path between two vertices by inspection. All of the possible route options should be listed, but it is sometimes obvious that certain routes are going to be much longer than others.

Example 14 Finding the shortest path from one vertex to another

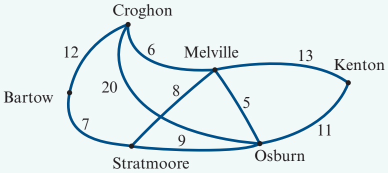
Find the shortest path from Bartow to Kenton in the network shown on the right.



Solution

1 List options for travelling from Bartow to Kenton.

B–S–M–O–K
B–S–M–K
B–S–O–K



Note: Routes via Croghon are likely to be much longer than routes via Stratmoore because of the larger distance from Bartow to Croghon (12 km).

2 Add the weights for each route.

B–S–M–O–K	7 + 8 + 5 + 11 = 31 km
B–S–M–K	7 + 8 + 13 = 28 km
B–S–O–K	7 + 9 + 11 = 27 km

3 Write your answer.

The shortest path from Bartow to Kenton is 27 km with route B–S–O–K.

ISBN 978-1-009-25917-0

Dijkstra's algorithm (optional)

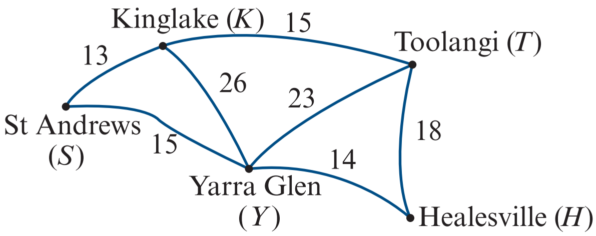
Finding the shortest path from one vertex of a graph to another is easy to determine if the graph is small and does not have too many vertices and edges. When there are many vertices and many edges, a systematic method, called an **algorithm**, can be used to find the shortest path.

Dutch computer scientist, Edsger Wybe Dijkstra (pronounced 'Dyke-stra') developed an algorithm for determining the shortest path through a graph. This algorithm, and others like it, have important applications to computerised routing and scheduling programs, such as GPS navigation devices.

Using Dijkstra's algorithm to find the shortest path between two vertices

Dijkstra's algorithm can be used to find the shortest path through a network, from a starting vertex to a destination vertex.

The weights in the graph opposite show the distances, in kilometres, by road between the towns. The algorithm will be used to find the shortest path between St Andrews (S) and Toolangi (T).



Step 1: Create a table

1. Write the starting vertex as the first row vertex.
2. Write the other vertices in the network as column vertices (order is unimportant).

	K	Y	T	H
S				

Note: St Andrews (S) is the starting vertex and is the first row vertex. All the other towns are column vertices.

Step 2: Complete the first row

Look at the graph to find the distance from the starting vertex (the row vertex) to the vertices that are directly connected to it (a column vertex).

1. Write the distance from the row vertex to a column vertex, directly underneath the column vertex as shown.
2. If a vertex is not directly connected to the starting vertex, mark a cross (×).

	K	Y	T	H
S	13	15	×	×

Note:

1. Kinglake (K) is 13 km from St Andrews (S).
2. Yarra Glen (Y) is 15 km from St Andrews (S).
3. Toolangi (T) and Healsville (H) are not directly connected to St Andrews (S) and have a cross (×).

ISBN 978-1-009-25917-0

3 Look for the smallest number in the first row and draw a box around it. If there are two or more the same, any one can be chosen.

4 The column vertex for this boxed number becomes the next row vertex.

	K	Y	T	H
S	$\boxed{13}$	15	×	×
K				

Note:

1 Put a box around 13 because it is the smallest number in the row.

2 The column vertex for this number is K so this becomes the next row vertex.

Step 3: Complete further rows

1 Copy all boxed numbers into the next row.

	K	Y	T	H
S	$\boxed{13}$	15	×	×
K	$\boxed{13}$			

Note: 13 has a box so it is copied into the next row.

2 For the remaining columns, add the boxed number for the row vertex to the distance from the row vertex to the column vertex.

- **a** If the value is *greater than* the value above it in the column, ignore the new value and copy the existing one.
- **b** If the value is *less than or equal to* the value above it in the column, write the new value.
- **c** If the row vertex is not directly connected to the column vertex, copy the existing value.

	K	Y	T	H
S	$\boxed{13}$	15	×	×
K	$\boxed{13}$	15	28	×

Note:

1 The box number for column K is 13. This must be added to the distance from K to every other vertex.

2 For Y: $\boxed{13} + 26 = 39$.
This is larger than the existing 15, so 15 is copied into the new row.

3 For T: $\boxed{13} + 15 = 28$.
This is now a possible connection so copy 28 into the new row.

4 For H: There is no direct connection between K and H, so it stays as a cross (×).

3 Look for the smallest *unboxed* number in the row and draw a box around it.

4 The column vertex for this new boxed number becomes the next row vertex.

	K	Y	T	H
S	$\boxed{13}$	15	×	×
K	$\boxed{13}$	$\boxed{15}$	28	×
Y				

Note:

1 Put a box around 15 because it is the smallest unboxed number in the row.

2 The 15 is in the column for Y, so Y is the next row vertex.

ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

5 Repeat step 3 until the destination vertex value has a box around it.

	K	Y	T	H
S	$\boxed{13}$	15	×	×
K	$\boxed{13}$	$\boxed{15}$	28	×
Y	$\boxed{13}$	$\boxed{15}$	$\boxed{28}$	29

Note:

1 The box number for column Y is 15. This must be added to the distance from Y to every other vertex.

2 For T: $\boxed{15} + 23 = 38$.
This is larger than the existing 28, so copy 28 into the new row.

3 For H: $\boxed{15} + 14 = 29$.
This is now a possible connection, so copy 29 into the new row.

4 The destination vertex (T) has a box around its value, so stop the algorithm.

Step 4: Backtrack to identify the shortest path and its length

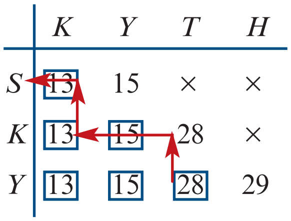
1 Start at the box value for the destination vertex. This is the length of the shortest path from the starting vertex to the destination vertex.

2 Draw a line up the column to the last number that is the same as the box number (it does not have to have a box around it).

3 Look at the row vertex for this number and draw a horizontal line to the column for this vertex.

4 Repeat until the starting vertex is reached.

5 The horizontal lines in the table indicate the shortest path.



	K	Y	T	H
S	$\boxed{13}$	15	×	×
K	$\boxed{13}$	$\boxed{15}$	28	×
Y	$\boxed{13}$	$\boxed{15}$	$\boxed{28}$	29

Note:

1 The shortest path length from St Andrews to Toolangi is 28 kilometres.

2 Start at the value 28 under the column vertex for T.

3 Draw a line up to the last value of 28 in the column. This occurs on row K.

4 Draw a horizontal line to column K.

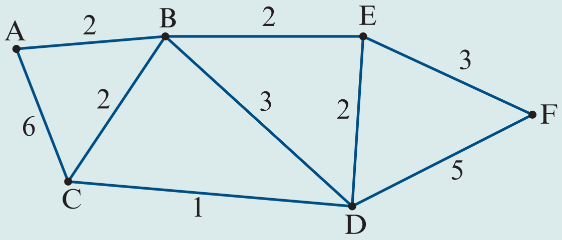
5 Draw a line up to the last value of 13 in the column. This occurs on row S.

6 There are horizontal lines between S and K and K and T so the shortest path is $S-K-T$.

ISBN 978-1-009-25917-0

Example 15 Finding the shortest path from one vertex to another using Dijkstra's algorithm

Find the shortest path from *A* to *F* in the weighted graph shown on the right.



Solution

Step 1

1 Create a table with the starting vertex (*A*) as the first row vertex.

2 Write the other vertices as column vertices.

	B	C	D	E	F
A					

Step 2

1 Write the distance from *A* to all the other vertices.

- *A*–*B* is distance 2.
- *A*–*C* is distance 6.

2 Put a cross (×) if the vertices are not directly connected.

- *A* is not directly connected to *D*, *E* nor *F* so insert a cross (×).

3 Put a box around the smallest number in the row.

- Put a box around 2, the smallest value in the row.

4 Look at the column vertex for this box number. This becomes the next row vertex.

- Vertex *B* is the next row vertex.

	B	C	D	E	F
A	$\boxed{2}$	6	×	×	×
B					

Step 3

1 Copy the boxed numbers into the next row.

2 For the remaining columns, add the box number for the row vertex to the distance from the row vertex to the column vertex.

- For column C: B–C is distance 2. New value is $\boxed{2} + 2 = 4$. This is less than 6, so write 4 in the row.
- For column D: B–D is distance 3. New value is $\boxed{2} + 3 = 5$. Write 5 in the row.
- For column E: B–E is distance 2. New value is $\boxed{2} + 2 = 4$. Write 4 in the row.
- For column F: B–F has no direct connection. Copy the cross (×).

	B	C	D	E	F
A	$\boxed{2}$	6	×	×	×
		2 + 2 = 4	2 + 3 = 5	2 + 2 = 4	
B	$\boxed{2}$	4	5	4	×

3 Put a box around the smallest unboxed number in the row.

4 Look at the column vertex for this box number. This becomes the next row vertex.

- Column C and column E both have the smallest number of 4. Either one can be boxed. Use E as the next row vertex.

	B	C	D	E	F
A	$\boxed{2}$	6	×	×	×
		2 + 2 = 4	2 + 3 = 5	2 + 2 = 4	
B	$\boxed{2}$	4	5	$\boxed{4}$	×
E					

5 Repeat until the destination vertex (F) value has a box around it.

	B	C	D	E	F
A	$\boxed{2}$	6	×	×	×
		2 + 2 = 4	2 + 3 = 5	2 + 2 = 4	
B	$\boxed{2}$	4	5	$\boxed{4}$	×
			4 + 2 = 6		4 + 3 = 7
E	$\boxed{2}$	$\boxed{4}$	5	$\boxed{4}$	7
			4 + 1 = 5		
C	$\boxed{2}$	$\boxed{4}$	$\boxed{5}$	$\boxed{4}$	7
					5 + 5 = 10
D	$\boxed{2}$	$\boxed{4}$	$\boxed{5}$	$\boxed{4}$	$\boxed{7}$

ISBN 978-1-009-25917-0

Step 4

1 The shortest distance from A to F is the boxed number in column F; that is, 7. Start backtracking at this number.

2 Draw a vertical line up the column to the last number; that is also 7. This occurs on row E.

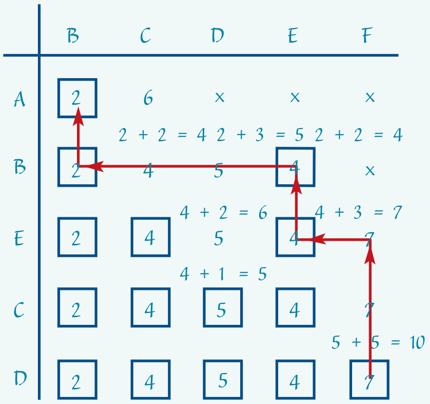
3 Draw a horizontal line across the table to column E. The value in this position is 4.

4 Draw a vertical line up the column to the last number that is also 4. This occurs on row B.

5 Draw a horizontal line across the table to column B. The value in this position is 2.

6 Draw a vertical line up the column to the last number that is also 2. This occurs on row A, the starting row.

7 Write your answer.



	B	C	D	E	F
A	$\boxed{2}$	6	x	x	x
		2 + 2 = 4	2 + 3 = 5	2 + 2 = 4	
B	$\boxed{2}$	4	5	$\boxed{4}$	x
			4 + 2 = 6		4 + 3 = 7
E	$\boxed{2}$	$\boxed{4}$	5	$\boxed{4}$	7
			4 + 1 = 5		
C	$\boxed{2}$	$\boxed{4}$	$\boxed{5}$	$\boxed{4}$	7
					5 + 5 = 10
D	$\boxed{2}$	$\boxed{4}$	$\boxed{5}$	$\boxed{4}$	$\boxed{7}$

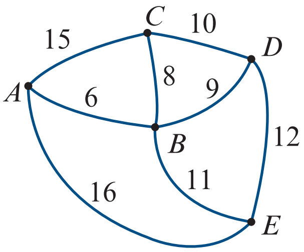
The shortest path from A to F is A–B–E–F.

Exercise 6G

Weighted graphs and networks

Example 12

1 The graph on the right shows towns A, B, C, D and E represented by vertices. The edges represent road connections between the towns. The weights on the edges are the average times, in minutes, it takes to travel along each road.



a Which two towns are 12 minutes apart by road?

b How long will it take to drive from C to D via B?

c A motorist intends to drive from D to E via B. How much time will they save if they travel directly from D to E?

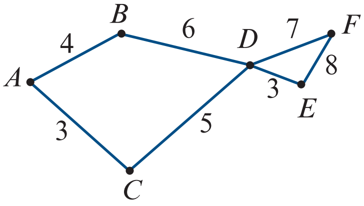
d Find the shortest time it would take to start at A, finish at E and visit every town exactly once.

ISBN 978-1-009-25917-0

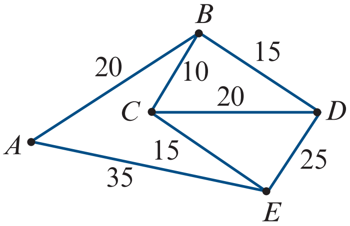
Finding the shortest path by inspection

Example 13
Example 14

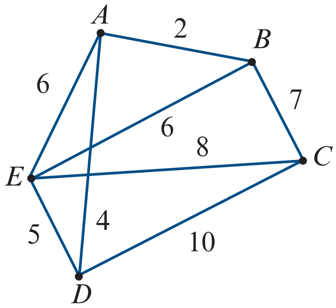
2 Find the shortest path from vertex A to vertex E in this network. The numbers represent time in hours.



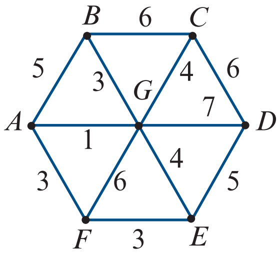
3 Find the shortest path from vertex A to vertex D in this network. The numbers represent lengths in metres.



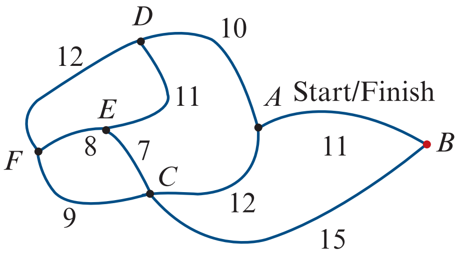
4 Find the shortest path from vertex B to vertex D in this network. The numbers represent cost in dollars.



5 Find the shortest path from vertex B to vertex F in this network. The numbers represent time in minutes.



6 The graph below shows a mountain bike rally course. Competitors must pass through each of the checkpoints, A, B, C, D, E and F. The average times (in minutes) taken to ride between the checkpoints are shown on the edges of the graph.



Competitors must start and finish at checkpoint A but can pass through the other checkpoints in any order they wish.
Which route would have the shortest average completion time?

ISBN 978-1-009-25917-0

Calculations within Dijkstra's algorithm (optional)

7 The table contains the first line first of a Dijkstra's algorithm solution to a shortest path problem.

	B	C	D	E	F
A	5	2	×	×	6

a Which vertex is the starting vertex for the problem?

b Which two vertices are not directly connected to the starting vertex?

c Which vertex will be the next vertex in the algorithm?

d What is the length of edge A–F?

8 The table on the right contains the first two lines of a Dijkstra's algorithm solution to a shortest path problem. Vertex R is a distance of 5 from Q, a distance of 6 from S, a distance of 4 from T and a distance of 3 from U. Complete the row for vertex R.

	Q	R	S	T	U
P	3	$\boxed{1}$	×	4	×
R		$\boxed{1}$			

9 The table on the right contains the first two lines of a Dijkstra's algorithm solution to a shortest path problem.

	N	P	W	V	U
M	×	5	7	$\boxed{5}$	×
V	8	5	6	$\boxed{5}$	×

a Vertex V was chosen as the second vertex in the algorithm. Which other vertex could have been chosen instead?

b Which vertex is definitely not directly connected to vertex V?

c What is the distance from vertex V to W?

10 A completed table of calculations for the shortest path through a network using Dijkstra's algorithm is shown on the right.

	B	C	D	E
A	$\boxed{8}$	14	×	×
B	$\boxed{8}$	$\boxed{10}$	11	×
C	$\boxed{8}$	$\boxed{10}$	$\boxed{11}$	16
D	$\boxed{8}$	$\boxed{10}$	$\boxed{11}$	$\boxed{16}$

a What is the length of the shortest path from A to C?

b What is the length of the shortest path from A to E?

c Write the shortest path taken from A to E.

11 A completed table of calculations for the shortest path through a network using Dijkstra's algorithm is shown on the right.

	B	C	D	E	F	G	H	I
A	2	$\boxed{1}$	×	5	×	×	×	×
C	$\boxed{2}$	$\boxed{1}$	×	5	4	×	×	×
B	$\boxed{2}$	$\boxed{1}$	6	5	$\boxed{4}$	×	×	×
F	$\boxed{2}$	$\boxed{1}$	6	$\boxed{5}$	$\boxed{4}$	×	8	×
E	$\boxed{2}$	$\boxed{1}$	$\boxed{6}$	$\boxed{5}$	$\boxed{4}$	7	7	×
D	$\boxed{2}$	$\boxed{1}$	$\boxed{6}$	$\boxed{5}$	$\boxed{4}$	$\boxed{7}$	7	×
G	$\boxed{2}$	$\boxed{1}$	$\boxed{6}$	$\boxed{5}$	$\boxed{4}$	$\boxed{7}$	$\boxed{7}$	8
H	$\boxed{2}$	$\boxed{1}$	$\boxed{6}$	$\boxed{5}$	$\boxed{4}$	$\boxed{7}$	$\boxed{7}$	$\boxed{8}$

a What is the length of the shortest path from A to G?

b What is the shortest path from A to G?

c What is the length of the shortest path from A to I?

d What is the shortest path from A to I?

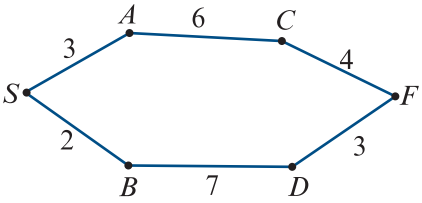
ISBN 978-1-009-25917-0

Finding the shortest path using Dijkstra's algorithm (optional)

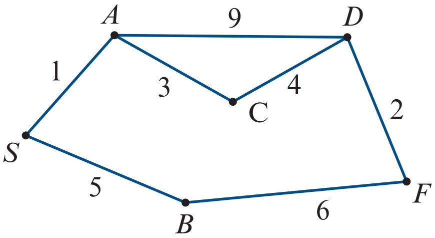
Example 15

12 Use Dijkstra's algorithm to determine the *shortest path from S to F in the following networks.* Write the length of the shortest path.

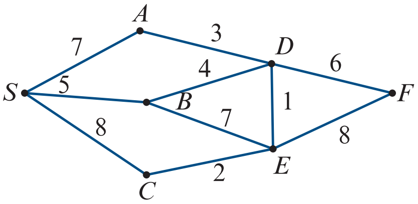
a



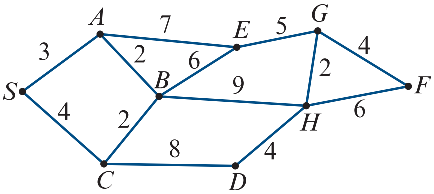
b



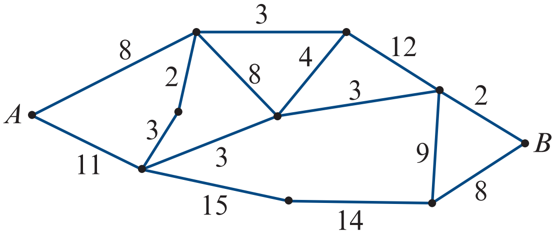
c



d



13 In the network below, the vertices represent small towns and the edges represent roads. The numbers on each edge indicate the distances (in kilometres) between towns. Determine the length of the shortest path between the towns labelled *A* and *B*.

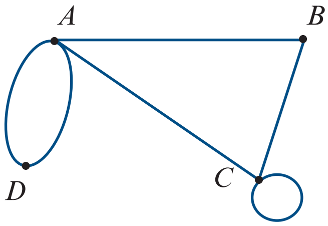
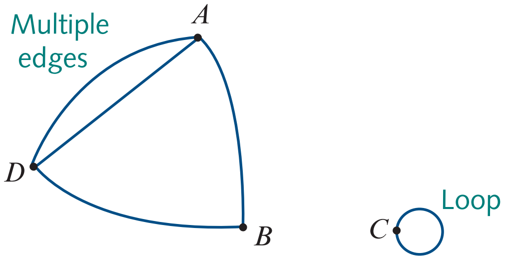


ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Key ideas and chapter summary

Assignment

Graph or network	A **graph** is a diagram that consists of a set of points called **vertices** and a set of lines called **edges**. Each edge joins two vertices. 
Vertices and edges	In the graph above, A, B, C, and D are **vertices** and the lines AB, AD, AC, and BC are **edges**.
Degree of a vertex	The **degree of vertex** A, written deg(A), is the *number of edges attached to the vertex*. The degree of a loop is two. For example, in the graph above: $\deg(B) = 2$ and $\deg(C) = 4$.
Loop	A **loop** is an edge that connects a vertex to itself. In the graph above, vertex C has a loop. The degree of a loop is two, so $\deg(C) = 4$.
Multiple edges and loops	 The graph above is said to have **multiple edges**, as there are two edges joining A and D. C has one edge, which links C to itself. This edge is a **loop**.
Simple graph	A **simple graph** is a graph that does not have loops and does not have multiple edges.
Disconnected graph	A **disconnected graph** has at least one pair of vertices between which there is no path.
Complete graph	A **complete graph** has every vertex connected to every other vertex by an edge.

ISBN 978-1-009-25917-0

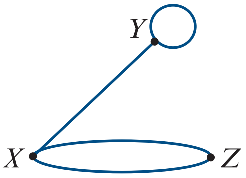
Cambridge University Press
Updated January 2024

Subgraph

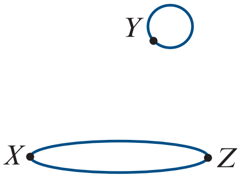
A **subgraph** is a graph that is part of a larger graph and has some of the same vertices and edges as that larger graph. A subgraph does not have any extra vertices or edges that do not appear in the larger graph.

Connected graph and bridges

A **graph is connected** if there is a path between each pair of vertices.
A **bridge** is a single edge in a connected graph that if removed leaves the graph disconnected. A graph can have more than one bridge.
Graph 1 is a connected graph.
Graph 2 is a disconnected graph.
Edge XY in Graph 1 is a bridge because removing it leaves Graph 1 disconnected.



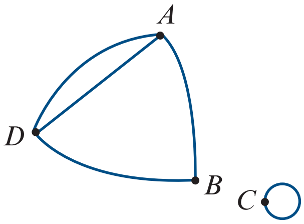
Graph 1



Graph 2

Adjacency matrix

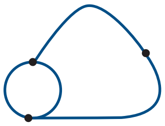
An **adjacency matrix** is a square matrix that uses a zero or an integer to record the number of edges connecting each pair of vertices in the graph. An example of a graph and its adjacency matrix:



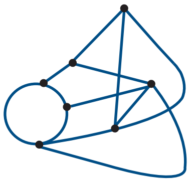
$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 2 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 2 & 1 & 0 & 0 \end{bmatrix} \end{array}$$

Planar graph

A graph that can be drawn in such a way that no two edges intersect, except at the vertices, is called a **planar graph**.



Planar graph



Non-planar graph

Face

A **face** in a graph is the area that is enclosed by edges. The faces are labelled f_1, f_2 and f_3 in the diagram following.
The area that surrounds a graph, f_3, is also considered to be a face.

Euler's formula

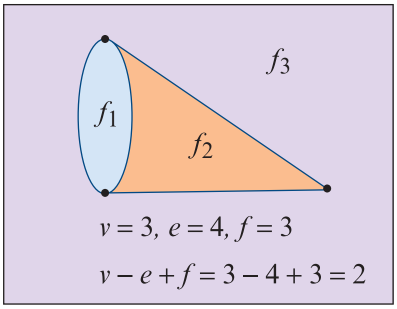
For any **connected planar** graph, **Euler's formula** states:

$v - e + f = 2$

v = the number of vertices

e = the number of edges

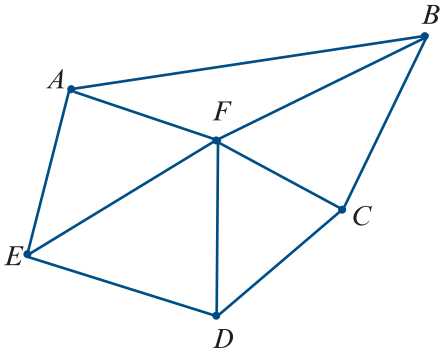
f = the number of faces



$v = 3,\ e = 4,\ f = 3$

$v - e + f = 3 - 4 + 3 = 2$

Walk, trail, path and cycle



A **walk** is a sequence of edges, linking successive vertices, in a graph. In the graph above *E-A-F-D-C-F-E-A* is a walk.

A **trail** is a walk with no repeated edges. *A-F-D-E-F-C* is a trail.

A **closed trail** is a walk that has no repeated edges that starts and ends at the same vertex. *A-F-D-E-F-B-A* is a closed trail.

A **path** is a walk with no repeated vertices. *F-A-B-C-D* is a path.

A **cycle** is a walk with no repeated vertices that starts and ends at the same vertex. *B-F-C-B* is a cycle.

Cycles are also called closed paths.

Traversable graph

A **traversable graph** has at least one trail that includes every edge.

Eulerian graph

An **Eulerian graph** has a closed trail that includes every edge just once. To be Eulerian, a graph must be connected and must have all vertices of an even degree.

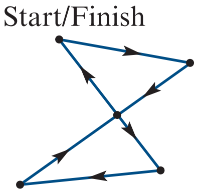
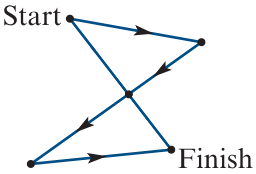
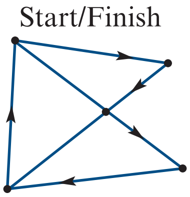
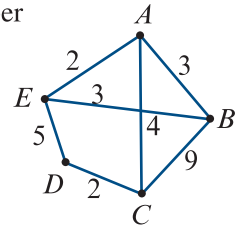
Semi-Eulerian graph

A **semi-Eulerian graph** has an open trail that includes every edge in the graph once only. To be semi-Eulerian, a graph must be connected and must have exactly two vertices of an odd degree.

ISBN 978-1-009-25917-0

Cambridge University Press

Updated January 2024

Condition for an Eulerian trail	To have an **Eulerian trail**, a graph must be connected and all vertices must be even. In the network shown, all vertices are even. It has an Eulerian trail. The trail starts and finishes at the same vertex.	
Hamiltonian path and cycle	A **Hamiltonian path** is a path through a graph that passes through every vertex exactly once, but does not necessarily start and finish at the same vertex.	
	A **Hamiltonian cycle** is a Hamiltonian path that starts and finishes at the same vertex. Inspection is the only way to identify Hamiltonian paths and cycles.	
Weighted graphs and networks	A **weighted graph** is one where a number is associated with each edge. These numbers are called weights. When the weights are physical quantities, for example, distance, time, cost, a weighted graph is often called a **network**.	
The shortest path problem	Determining the shortest distance or time or the least cost to move around a network is called the **shortest path problem**.	
Dijkstra's algorithm	**Dijkstra's algorithm** can be used to identify the shortest route between two vertices in a network.	

Skills check

Having completed this chapter you should be able to:

- recognise and determine the number of vertices, edges and faces of a graph
- recognise simple graphs, disconnected graphs, complete graphs, subgraphs, planar graphs, connected graphs and bridges.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

- construct and interpret an adjacency matrix
- construct a graph from an adjacency matrix
- use Euler's formula
- know and apply the definitions of walks, paths, trails (open or closed) and cycles
- identify Eulerian and semi-Eulerian graphs
- know and apply the definitions of Hamiltonian paths and cycles
- locate a Hamiltonian path or cycle in a graph
- identify weighted graphs
- solve shortest path problems by inspection
- solve shortest path problems using Dijkstra's algorithm.

Short-answer questions

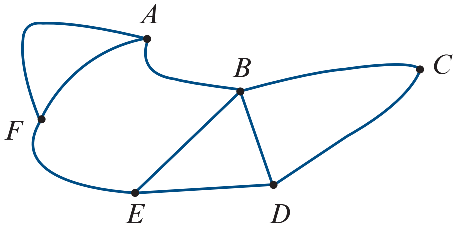
1 For the graph on the right, determine the:

a number of vertices

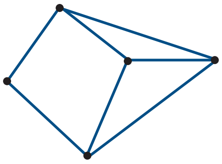
b number of edges

c degree of vertex *B*

d number of even vertices



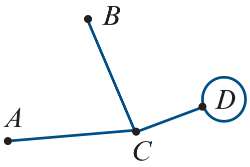
2 Determine the sum of the degree of the vertices of the graph on the right.



3 Draw the graph that matches the adjacency matrix:

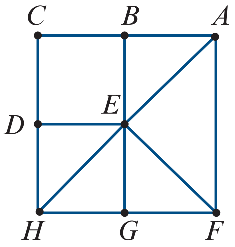
$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array}\begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \end{array}$$

4 Write the adjacency matrix that matches the graph shown.



ISBN 978-1-009-25917-0

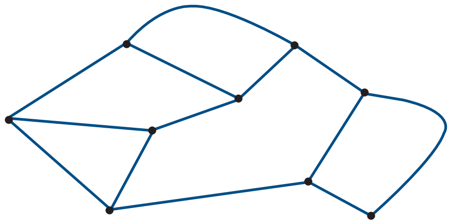
The graph below is to be used when answering Question 5



5 By using the key words walk, path, trail (open or closed) or cycle, describe what each sequence of vertices represents:

- a *C-B-E-A-E-G*
- b *D-E-H-G-E-A-B-C-D*
- c *C-B-E-A-F-E-G-H*
- d *D-E-A-F-G-H-D*

The graph below is to be used when answering Question 6



6 a What is the sum of the degrees of the vertices of the graph above?

b Write down the number of vertices, edges and faces for the graph above.

c Show, with use of Euler's formula, that the graph above is planar.

7 A connected graph with 10 vertices divides the plane into five faces. How many edges will this graph have?

8 For a connected graph to be semi-Eulerian, which of the following must be true?

- A All vertices must be odd.
- B All vertices must be even.
- C There must be exactly two odd vertices and the rest even.
- D There must be exactly two even vertices and the rest odd.
- E An odd vertex must be followed by an even vertex.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

9 For a connected graph to be an Eulerian graph, which of the following must be true?

A All vertices must be odd.

B All vertices must be even.

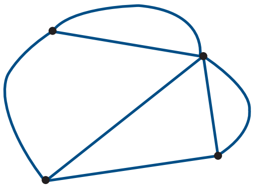
C There must be exactly two odd vertices and the rest even.

D There must be exactly two even vertices and the rest odd.

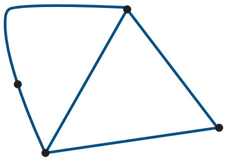
E An odd vertex must be followed by an even vertex.

The graphs below are to be used when answering Question 10

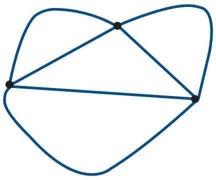
A



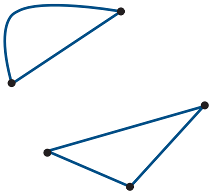
B



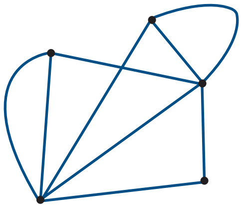
C



D



E

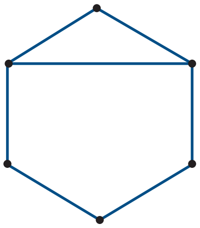


10 **a** Which one of the graphs above is semi-Eulerian?

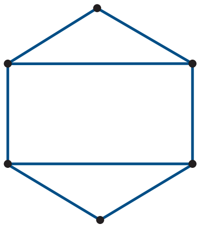
b Which one of the graphs above is Eulerian?

11 Which one of the following graphs is Eulerian?

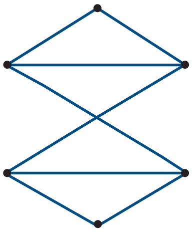
A



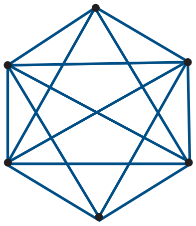
B



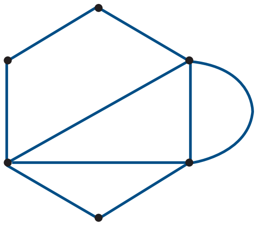
C



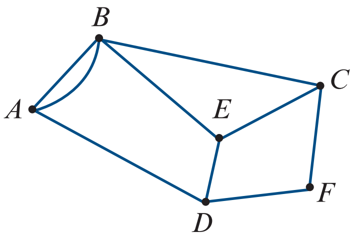
D



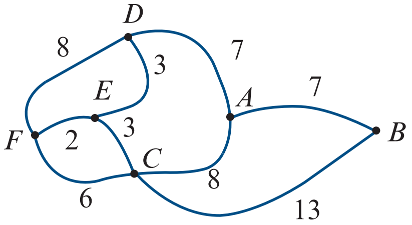
E



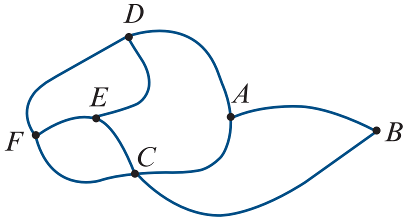
12 For the graph shown, where could an additional edge be added to the network so that the graph formed would be semi-Eulerian?



13 Determine the length of the shortest path from *F* to *B* in the network shown.

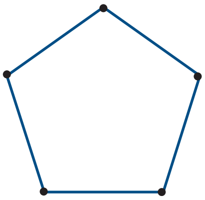


14 Write a Hamiltonian cycle for the graph shown, starting at vertex *F*.

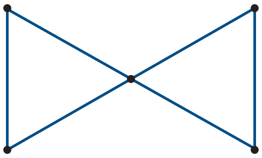


15 Of the following, which graph has both an Eulerian trail and Hamiltonian cycle?

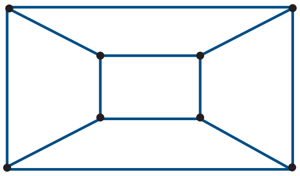
A



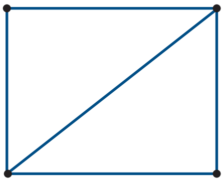
B



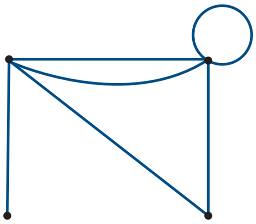
C



D



E



16 A park ranger wants to check all of the walking tracks in a national park, starting at and returning to the park office. She would like to do it without having to travel over the same walking track more than once. Give the mathematical name for the route she should follow.

17 A park ranger wants to check all of the campsites in a national park, starting at and returning to the park office. The campsites are all interconnected with walking tracks. She would like to check the campsites without having to visit each campsite more than once. Give the mathematical name for the route she should follow.

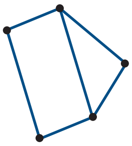
18 What is the minimum number of edges needed for a graph with seven vertices to be connected?

19 A connected planar graph with 15 vertices divides the plane into 12 regions. How many edges will be required to connect the vertices in this graph?

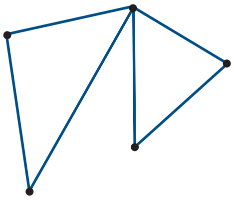
ISBN 978-1-009-25917-0

20 Which of the following graphs is *not* Eulerian?

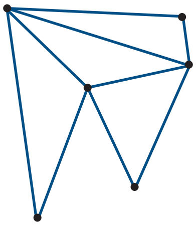
A



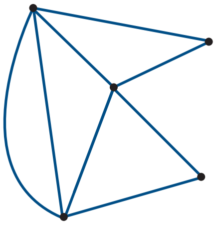
B



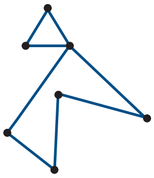
C



D



E



21 A connected planar graph divides the plane into a number of regions. If the graph has eight vertices and these are linked by 13 edges, how many regions will the graph have?

Extended-response questions

1 Draw a connected graph with:

a four vertices, four edges and two faces

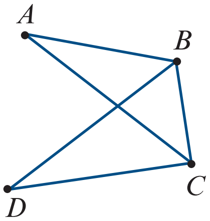
b four vertices, five edges and three faces

c five vertices, eight edges and five faces

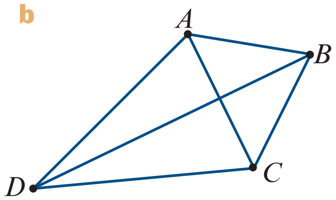
d four vertices, four edges, two faces and two bridges

2 Redraw each of the following graphs in a planar form.

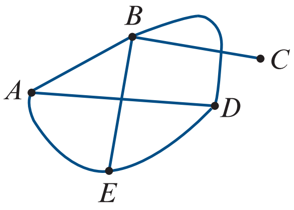
a



b



c

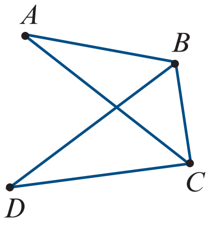


3 For the network shown, write:

a the degree of vertex C

b the numbers of odd and even vertices

c a route followed by an semi-Eulerian trail starting at vertex B

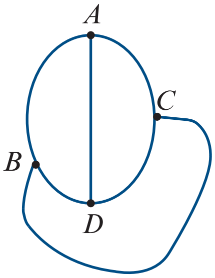


ISBN 978-1-009-25917-0

Cambridge University Press

Updated January 2024

4 Construct an adjacency matrix for the graph below.

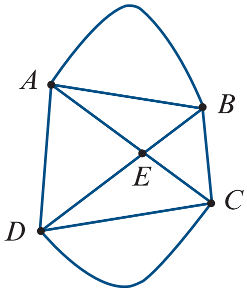


5 For the graph shown, write:

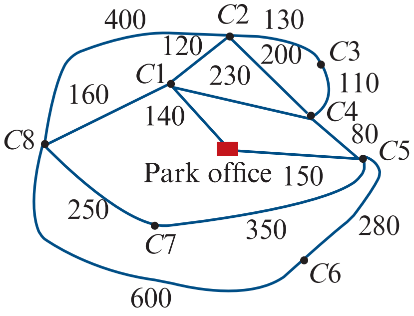
a the degree of vertex *C*

b the number of odd and even vertices

c a route followed by an Eulerian trail starting at vertex *A*



6 The diagram opposite shows the network of walking tracks in a small national park. These tracks connect the campsites to each other and to the park office. The lengths of the tracks (in metres) are also shown.



a The network of tracks is planar. Explain what this means.

b Verify Euler's formula for this network.

c A ranger at campsite *C8* plans to visit campsites *C1, C2, C3, C4* and *C5* on her way back to the park office. What is the shortest distance she will have to travel?

d How many even and how many odd vertices are there in the network?

e Each day, the ranger on duty has to inspect each of the tracks to make sure that they are all passable.

- **i** Is it possible for her to do this starting and finishing at the park office? Explain why.
- **ii** Identify one route that she could take.

f Following a track inspection after wet weather, the Head Ranger decides that it is necessary to put gravel on some walking tracks to make them weatherproof. What is the minimum length of track that will need to be gravelled to ensure that all campsites and the park office are accessible along a gravelled track?

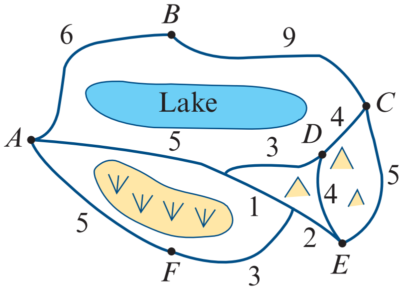
g A ranger wants to inspect each of the campsites but not pass through any campsite more than once on his inspection tour. He wants to start and finish his inspection tour at the park office.

ISBN 978-1-009-25917-0

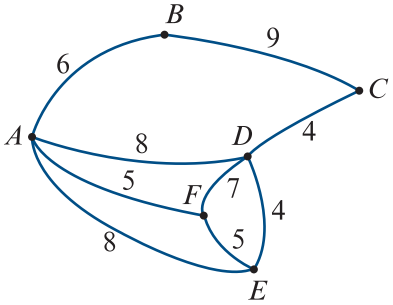
Cambridge University Press Updated January 2024

- **i** What is the technical name for the route he wants to take?
- **ii** With the present layout of tracks, he cannot inspect all the tracks without passing through at least one campsite twice. Suggest where an additional track could be added to solve this problem.
- **iii** With this new track, write down a route he could follow.

7 The map shows six campsites, A, B, C, D, E and F, which are joined by tracks. The numbers by the paths show lengths, in kilometres, of that section of track.



a **i** Complete the graph opposite, which shows the shortest direct distances between campsites. (The campsites are represented by vertices and tracks are represented by edges.)



ii A telephone cable is to be laid to enable phone calls to be made between campsites. For environmental reasons, it is necessary to lay the cable along as few of the existing tracks as possible. What is the minimum length of cable necessary to complete this task?

iii Fill in the missing entries for the adjacency matrix shown for the completed graph formed above.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \\ F \end{array}\begin{array}{c} \begin{array}{cccccc} A & B & C & D & E & F \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & – & – & – & – \\ 1 & 0 & – & – & – & – \\ 1 & 0 & – & – & – & – \end{bmatrix} \end{array}$$

b A walker follows the route $A–B–A–F–E–D–C–E–F–A$.

- **i** How far does this person walk?
- **ii** Why is the route *not* a Hamiltonian cycle?
- **iii** Write down a route that a walker could follow that is a Hamiltonian cycle.
- **iv** Find the distance walked in following this Hamiltonian cycle.

c It is impossible to start at A and return to A by going along each track exactly once. An extra track joining two campsites can be constructed so that this is possible. Which two campsites need to be joined by a track to make this possible?

ISBN 978-1-009-25917-0

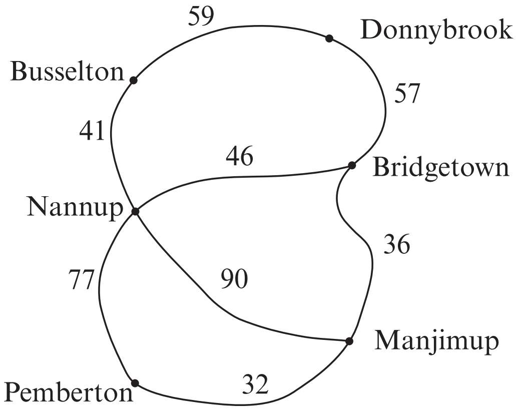
Cambridge University Press
Updated January 2024

8 The graph below shows part of the road network connecting towns in Western Australia.

The vertices represent the towns of Donnybrook, Bridgetown, Manjimup, Pemberton, Nannup and Busselton.

The edges represent a road connection between the towns.

The numbers on the edges show the distance, in kilometres, along the roads connecting the towns.



a Which two towns have the shortest distance between them?

b If a tourist drives along all of the roads shown in the graph only once, how far would they drive?

c Explain why it is not possible for the tourist to start at Busselton, drive along each of the roads in the graph only once and end back at Busselton.

d It is possible for the tourist to start in one town, travel along each of the roads only once and then end at a different town.

- **i** What is the name given to this type of walk?
- **ii** In what towns would the tourist start and end this walk?

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

7

Investigating and modelling time series

In this chapter

Syllabus references

Topics: Describing and interpreting patterns in time series data; Analysing time series data

Subtopics: 4.1.1 – 4.1.8

7A Time series data

Time series data are a special kind of bivariate data, where the explanatory variable is time. An example of time series data is the following table of Australian annual birth rates (average births per female) between 1931 and 2002.

Year	*Birth rate*	*Year*	*Birth rate*	*Year*	*Birth rate*	*Year*	*Birth rate*
1931	1.039	1949	1.382	1967	1.342	1985	0.920
1932	0.967	1950	1.415	1968	1.360	1986	0.894
1933	0.959	1951	1.409	1969	1.360	1987	0.883
1934	0.939	1952	1.468	1970	1.349	1988	0.877
1935	0.941	1953	1.477	1971	1.400	1989	0.882
1936	0.967	1954	1.497	1972	1.296	1990	0.908
1937	0.981	1955	1.532	1973	1.179	1991	0.887
1938	0.976	1956	1.546	1974	1.123	1992	0.906
1939	0.986	1957	1.598	1975	1.049	1993	0.893
1940	1.042	1958	1.603	1976	0.980	1994	0.884
1941	1.094	1959	1.614	1977	0.951	1995	0.875
1942	1.096	1960	1.613	1978	0.930	1996	0.861
1943	1.148	1961	1.668	1979	0.908	1997	0.855
1944	1.179	1962	1.609	1980	0.901	1998	0.848
1945	1.267	1963	1.572	1981	0.924	1999	0.846
1946	1.379	1964	1.480	1982	0.921	2000	0.844
1947	1.416	1965	1.400	1983	0.920	2001	0.833
1948	1.376	1966	1.355	1984	0.883	2002	0.848

This dataset is rather complex, and it is hard to see any patterns just by looking at the data.

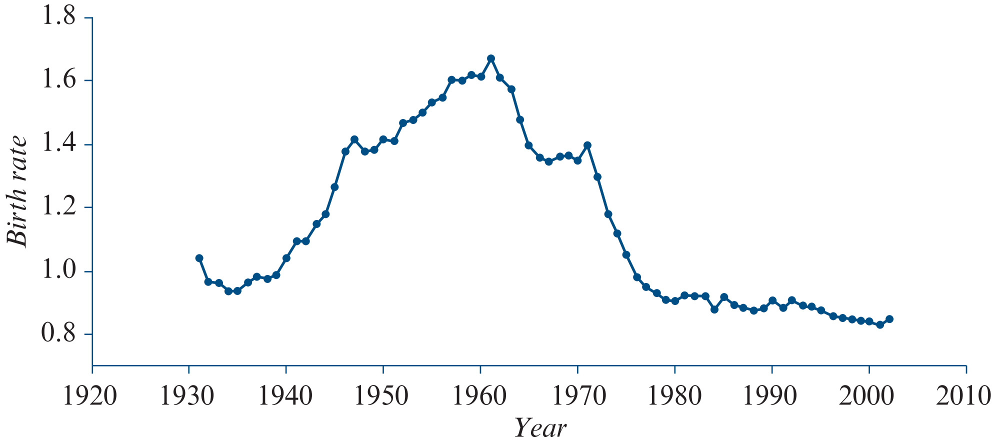
However, we can start to make sense of the data by displaying it graphically.

The graph we use for this purpose is called a **time series plot**.

A time series plot is a line graph with time plotted on the horizontal axis. The variable under investigation, the response variable, is plotted on the vertical axis.

The time series plot over page has been constructed from the birth rate data tabulated above.

In this time series plot, *birth rate* is the response variable.



Looking for patterns in time series plots

The features we look for in a time series are:

- trend
- cycles
- seasonality
- structural change
- possible outliers
- irregular (random) fluctuations.

One or all of these features can be found in a time series plot.

Trend

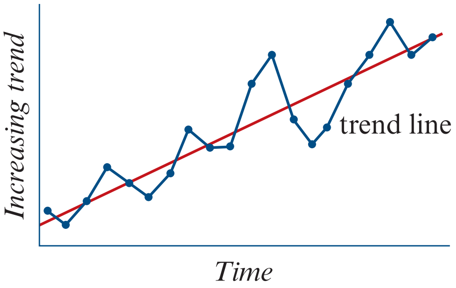
Examining a time series plot we can often see a general upward or downward movement over time. This indicates a long-term change over time that we call a trend.

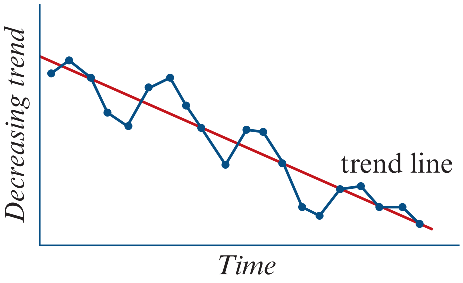
Trend

The tendency for values in a time series to generally increase or decrease over a significant period of time is called a **trend**.

One way of identifying trends on a time series graph is to draw a line that ignores the fluctuations, but which reflects the overall increasing or decreasing nature of the plot. These lines are called **trend lines**.

Trend lines have been drawn on the time series plots below to indicate an *increasing* trend (line slopes upwards) and a *decreasing* trend (line slopes downwards).

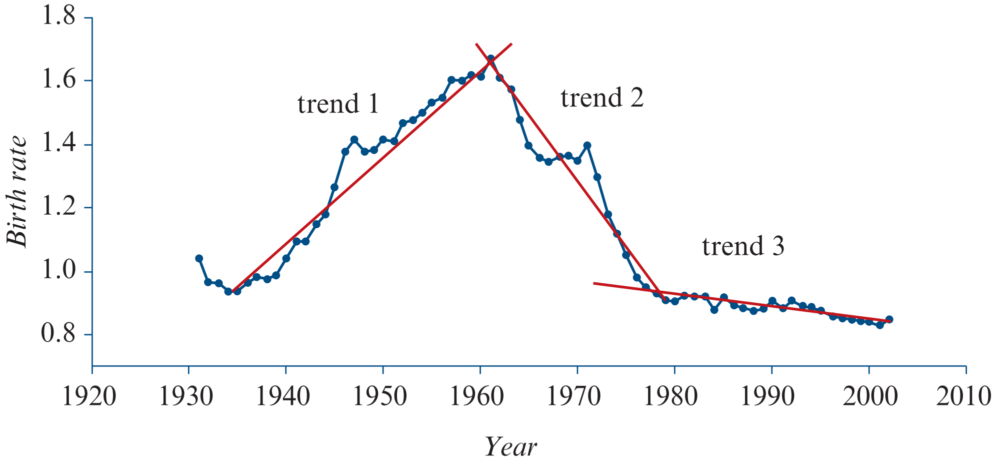




ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Sometimes, different trends are apparent in a time series for different time periods. For example, in the time series plot of the birth rate data, there are three distinct trends, which can be seen by drawing trend lines on the plot.



Each of these trends can be explained by changing socioeconomic circumstances.

Trend 1: Between 1940 and 1961 the birth rate in Australia grew quite dramatically. Those in the armed services came home from the Second World War, and the economy grew quickly. This rapid increase in the Australian birth rate during this period is known as the 'Baby Boom'.

Trend 2: From about 1962 until 1980 the birth rate declined very rapidly. Birth control methods became more effective, and women started to think more about careers. This period is sometimes referred to as the 'Baby Bust'.

Trend 3: During the 1980s, and up until the early 2000s, the birth rate continued to decline slowly for a complex range of social and economic reasons.

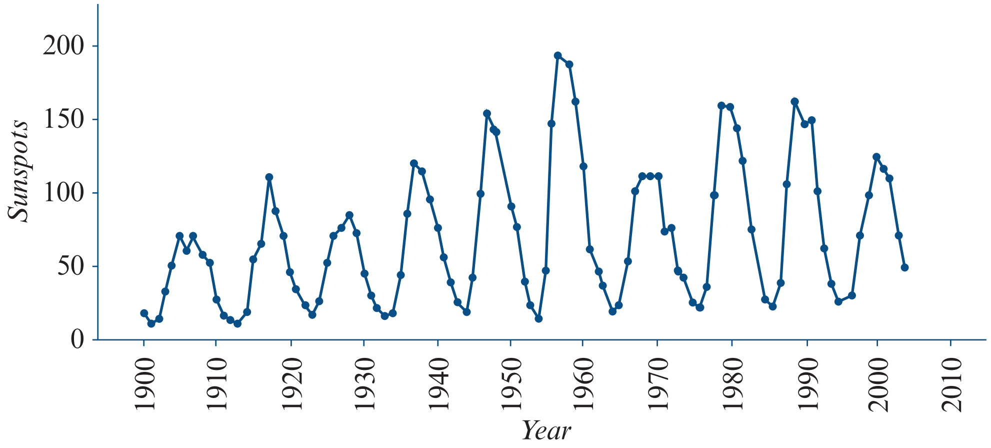
ISBN 978-1-009-25917-0

Cycles

Cycles

Cycles are periodic movements in a time series. The period is the time it takes for one complete up and down movement in the time series plot. In practice, this term is reserved for periods greater than 1 year.

Some cycles repeat regularly, and some do not. The following plot shows the sunspot[1] activity for the period 1900 to 2010. The period of this cycle is approximately 11 years.



Many business indicators, such as interest rates or unemployment figures, also vary in cycles, but their periods are usually less regular. Cycles with calendar-related periods of 1 year or less are of special interest and give rise to what is called ‘seasonality’ (or seasonal variation).

Seasonality

Seasonality

Seasonality is present when there is a periodic movement in a time series that has a calendar-related period, for example, a year, a month or a week.

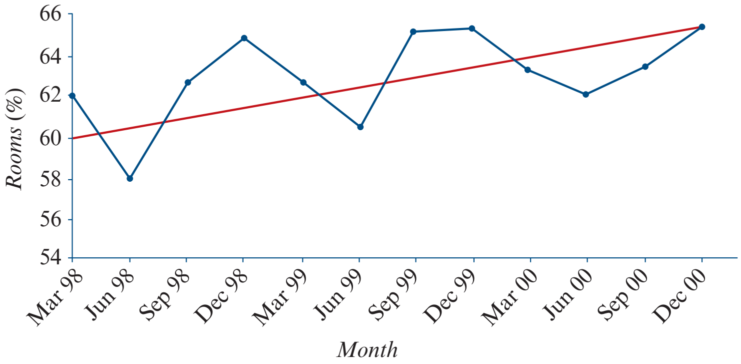
Seasonal variation is the regular rise and fall in the time series that recurs each year.

Seasonal movements tend to be more predictable than trends, and occur because of variations in the weather, such as ice-cream sales, or institutional factors, like the increase in the number of unemployed people at the end of the school year.

The next plot shows the total percentage of rooms occupied in hotels, motels and other accommodation in Australia by quarter, over the years 1998–2000.

---

[1] Sunspots are dark spots visible on the surface of the Sun that come and go over time.



This time series plot reveals both *seasonality* and *trend* in the demand for accommodation. The *regular peaks and troughs* in the plot that occur at the *same time each year* signal the presence of *seasonality*. In this case, the demand for accommodation is at its lowest in the June quarter and highest in the December quarter.

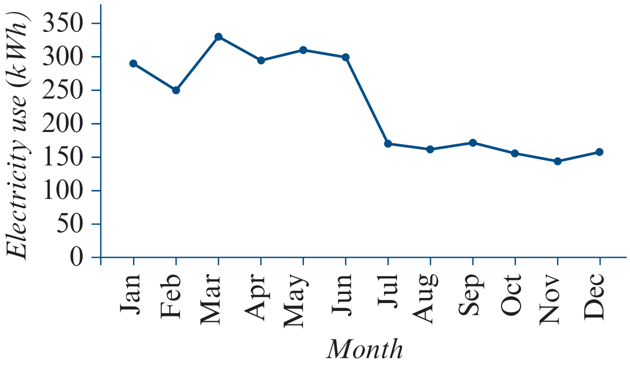
The *upward sloping trend line* signals the presence of a general increasing *trend*. This tells us that, even though demand for accommodation has fluctuated from month to month, demand for hotel and motel accommodation has increased over time.

Structural change

Structural change

Structural change is present when there is a sudden change in the established pattern of a time series plot.

The time series plot below shows the power bill for a rental house (in kWh) for the 12 months of a year. The plot reveals an abrupt change in power usage in June to July. During this period, monthly power use suddenly decreases from around 300 kWh per month from January to June to around 175 kWh for the rest of the year.



This is an example of structural change that can probably be explained by a change in tenants, from a family with two children to a person living alone.

ISBN 978-1-009-25917-0

Structural change is also displayed in the birth rate time series plot we saw earlier. This revealed three quite distinct trends during the period 1900–2010. These reflect significant external events (like a war) or changes in social and economic circumstances.

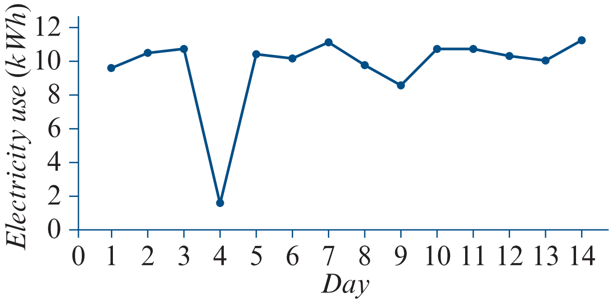
One consequence of structural change is that we can no longer use a single mathematical model to describe the key features of a time series plot.

Outliers

Outliers

Outliers are present when there are individual values that stand out from the general body of data.

The time series plot below shows the daily power bill for a house (in kWh) for a fortnight.



For this household, daily electricity use follows a regular pattern that, although fluctuating, averages about 10 kWh per day. In terms of daily power use, day 4 is a clear outlier, with less than 2 kWh of electricity used. A follow-up investigation found that, on this day, the house was without power for 18 hours due to a storm, so much less power was used than normal.

Irregular (random) fluctuations

Irregular (random) fluctuations

Irregular (random) fluctuations include all the variations in a time series that we cannot reasonably attribute to systematic changes like trend, cycles, seasonality and structural change or an outlier.

There can be many sources of irregular fluctuations, mostly unknown. A general characteristic of these fluctuations is that they are unpredictable.

One of the aims of time series analysis is to develop techniques to identify regular patterns in time series plots that are often obscured by irregular fluctuations. One of these techniques is smoothing, which you will meet in the next section.

Constructing time series plots

Most real-world time series data come in the form of large datasets that are best plotted with the aid of a spreadsheet or statistical package. The availability of the data in electronic form via the internet greatly helps this process. However, in this chapter, most of the time series datasets are relatively small and can be readily plotted using a CAS calculator.

How to construct a time series using the TI-Nspire CAS

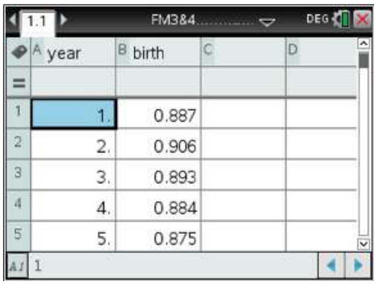
Construct a time series plot for the data presented below. The years have been recoded as 1, 2, . . . , 12, as is common practice.

2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
1	2	3	4	5	6	7	8	9	10	11	12
0.887	0.906	0.893	0.884	0.875	0.861	0.855	0.848	0.846	0.844	0.833	0.848

Steps

1 Start a new document by pressing ctrl + **N**.

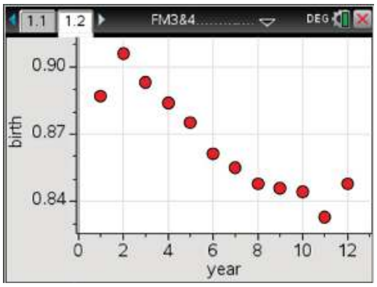
2 Select **Add Lists & Spreadsheet**. Enter the data into lists named *year* and *birth*.



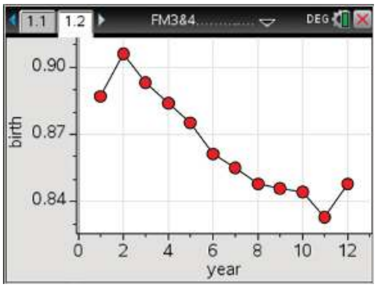
	A year	B birth	C	D
=				
1	1.	0.887		
2	2.	0.906		
3	3.	0.893		
4	4.	0.884		
5	5.	0.875		

A1 1

3 Press ctrl + **I** and select **Add Data & Statistics**. Construct a scatterplot of *birth* against *year*. Let *year* be the explanatory variable and *birth* the response variable.



4 To display as a connected time series plot, move the cursor to the main graph area and press ctrl + menu **>Connect Data Points**. Press enter.



ISBN 978-1-009-25917-0

Cambridge University Press

Updated January 2024

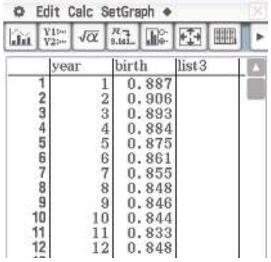
How to construct a time series using the ClassPad

Construct a time series plot for the data presented below. The years have been recoded as 1, 2, . . . , 12, as is common practice.

2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
1	2	3	4	5	6	7	8	9	10	11	12
0.887	0.906	0.893	0.884	0.875	0.861	0.855	0.848	0.846	0.844	0.833	0.848

Steps

1 Open the **Statistics** application and enter the data into the columns named *year* and *birth* as shown.



	year	birth	list3
1	1	0.887	
2	2	0.906	
3	3	0.893	
4	4	0.884	
5	5	0.875	
6	6	0.861	
7	7	0.855	
8	8	0.848	
9	9	0.846	
10	10	0.844	
11	11	0.833	
12	12	0.848	

2 Tap to open the **Set StatGraphs** dialog box and complete as follows.

- **Draw:** select **On**.
- **Type:** select **xyLine** (▼).
- **XList:** select **main/year** (▼).
- **YList:** select **main/birth** (▼).
- **Freq:** leave as **1**.
- **Mark:** leave as **square**.

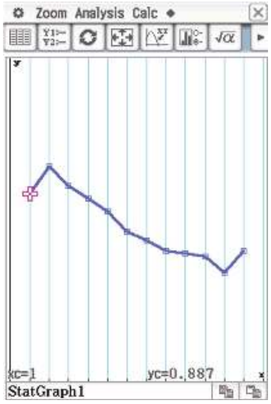
Tap **Set** to confirm your selections.

Draw: ● On ○ Off
Type: xyLine
XList: main\year
YList: main\birth
Freq: 1
Mark: square

3 Tap in the toolbar at the top of the screen to display the time series plot in the bottom half of the screen.

To obtain a full-screen display, tap Resize from the icon panel.

Tap from the toolbar, and use ▶ and ◀ to move from point to point to read values from the plot.



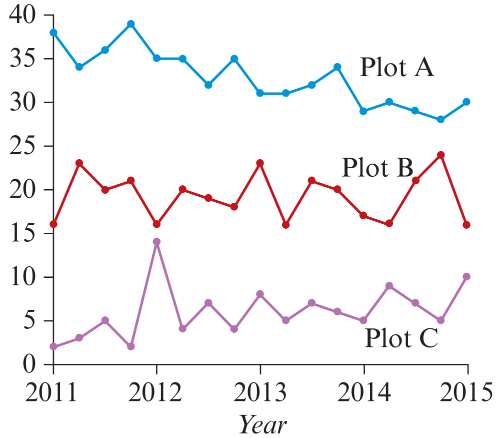
xc=1 yc=0.887

Exercise 7A

Identifying key features in a time series plot

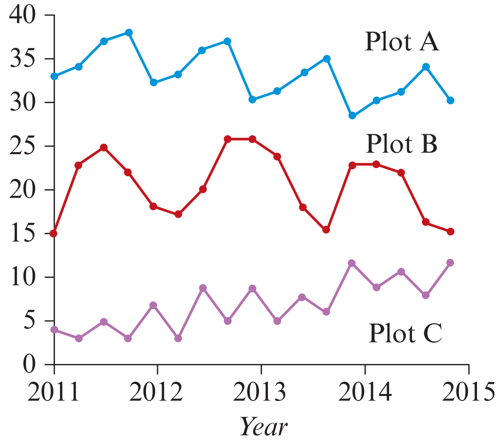
1 Complete the table below by indicating which of the listed features are present in each of the time series plots shown below.

	Plot		
Feature	A	B	C
Irregular fluctuations			
Increasing trend			
Decreasing trend			
Cycles			
Outlier			



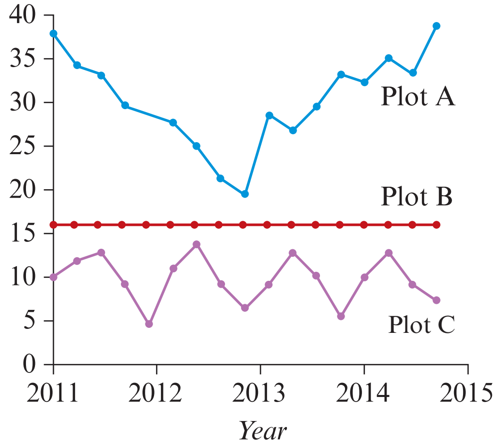
2 Complete the table below by indicating which of the listed features are present in each of the time series plots shown below.

	Plot		
Feature	A	B	C
Irregular fluctuations			
Increasing trend			
Decreasing trend			
Cycles			
Seasonality			



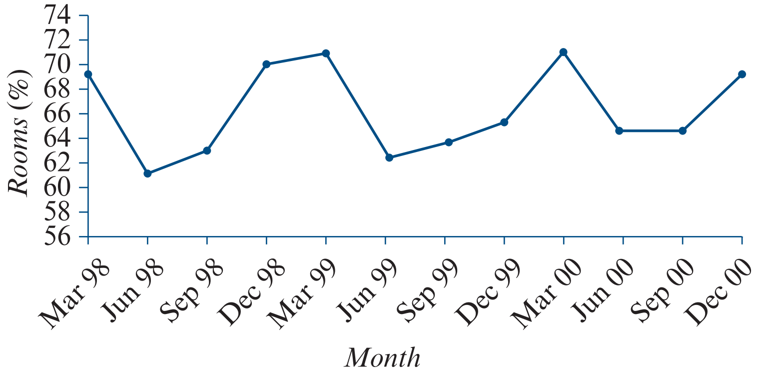
3 Complete the table below by indicating which of the listed features are present in each of the time series plots shown below.

	Plot		
Feature	A	B	C
Irregular fluctuations			
Structural change			
Increasing trend			
Decreasing trend			
Seasonality			



Describing time series plots

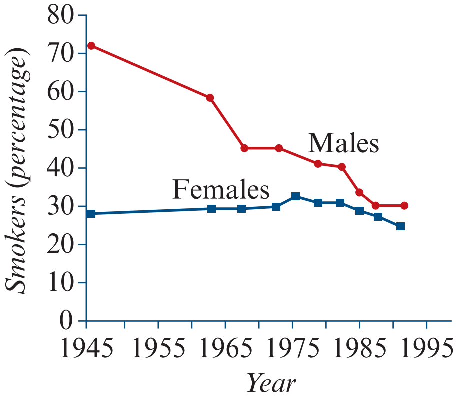
4 The time series plot for the hotel room occupancy rate (%) in Victoria over the period March 1998–December 2000 is shown below. Describe the features of the plot.



5 The time series plot opposite shows the smoking rates (%) of Australian males and females over the period 1945–92.

a Describe any trends in the time series plot.

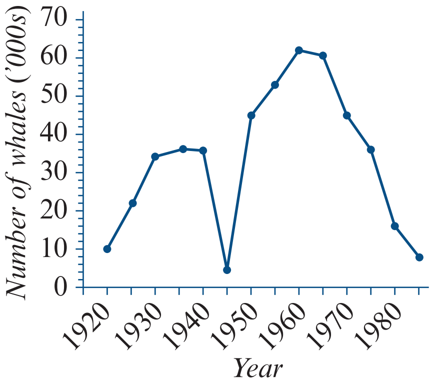
b Did the *difference* in smoking rates increase or decrease over the period 1945–92?



6 The time series plot opposite shows the number of whales caught during the period 1920–85. Describe the features of the plot.

Note: This time series exhibits structural change so cannot be described by a single trend. Here is some relevant historical information:

- The 1930s was the time of the Great Depression.
- 1939–45 was the period of the Second World War.
- 1960–85 was a time when countries began to accept that whales were endangered.



Using a CAS calculator to construct a time series plot

7 Use the data below to construct a time series plot of the Australian birth rate for 1960–70.

CAS

Year	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970
Rate	1.613	1.668	1.609	1.572	1.480	1.400	1.355	1.342	1.360	1.360	1.349

ISBN 978-1-009-25917-0

8 Use the data below to construct a time series plot of the population (in millions) in Australia over the period 1993–2003.

Year	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
Population	17.8	18.0	18.2	18.4	18.6	18.8	19.0	19.3	19.5	19.8	20.0

9 Use the data below to construct a time series plot for the number of teachers (in thousands) in Australia over the period 1991–2001.

Year	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
Teachers	213	217	218	218	221	223	228	231	239	244	250

10 The table below gives the number of male and female teachers (in thousands) in Australia over the years 1993–2001.

Year	1993	1994	1995	1996	1997	1998	1999	2000	2001
Males ('000s)	77.9	76.6	75.3	75.0	74.9	74.9	76.0	76.6	77.1
Females ('000s)	139.9	141.2	145.5	148.5	152.5	156.0	163.4	167.4	172.5

a Construct a time series plot showing both the male and female teachers on the same graph.

b Describe and comment on any trends you observe.

7B Smoothing a time series using moving averages

A time series plot can incorporate many of the sources of variation previously mentioned: trend, cycles, seasonality, structural change, outliers and irregular fluctuations. One effect of the irregular fluctuations and seasonality can be to obscure an underlying trend. The technique of **smoothing** can sometimes be used to overcome this problem.

Smoothing a time series plot using moving averages

This method of smoothing (**moving average smoothing**) involves replacing individual data points in the time series with their moving average (mean). The simplest method is to smooth over a small number of odd number points, for example, three or five.

The three-point moving average

To smooth by a three-point moving average (3-point MA), replace each data value with the mean of that value and the values of its two neighbours, one on each side. That is, if y_1, y_2 and y_3 are sequential data values, then:

$$\text{3-point MA } y_2 = \frac{y_1 + y_2 + y_3}{3}$$

The first and last values do not have values on each side, so leave them out.

ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

The five-point moving average

To smooth by a five-point moving average (5-point MA), replace each data value with the mean of that value and the values of its four neighbours, two each side. That is, if y_1, y_2, y_3, y_4, y_5 are sequential data values, then:

$$\text{5-point MA } y_3 = \frac{y_1 + y_2 + y_3 + y_4 + y_5}{5}$$

The first two and last two points do not have two values on each side, so leave them out.

If needed, these definitions can be readily extended for moving averages involving 7, 9, 11, . . . points. The larger the number of points we smooth over, the greater the smoothing effect.

Example 1 Three- and five-moving mean smoothing

The table below gives the temperature (°C) recorded at a weather station at 9 a.m. each day for a week.

Day	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Temperature	18.1	24.8	26.4	13.9	12.7	14.2	24.9

a Calculate the three-point moving average temperature for Tuesday.

b Calculate the five-point moving average temperature for Thursday.

Solution

a **1** Write the three temperatures centred on Tuesday.

18.1, 24.8, 26.4

2 Find their mean and write your answer.

$$\text{Mean} = \frac{(18.1 + 24.8 + 26.4)}{3} = 23.1$$

The three-point moving average temperature for Tuesday is 23.1°C.

b **1** Write the five temperatures centred on Thursday.

24.8, 26.4, 13.9, 12.7, 14.2

2 Find their mean and write your answer.

$$\text{Mean} = \frac{(24.8 + 26.4 + 13.9 + 12.7 + 14.2)}{5}$$
$$= 18.4$$

The five-point moving average temperature for Thursday is 18.4°C.

The next step is to extend these computations to smooth all terms in the time series.

Example 2 Three- and five-point moving averages of a time series

The following table gives the number of births per month over a calendar year in a country hospital. Use the three-point moving average (3-point MA) and the five-point moving average (5-point MA) methods, correct to one decimal place, to complete the table.

Month	*Number of births*	*3-point MA*	*5-point MA*
January	10	–	–
February	12		–
March	6		
April	5		
May	22		
June	18		
July	13		
August	7		
September	9		
October	10		
November	8		–
December	15	–	–

Solution

Complete the calculations as shown below.

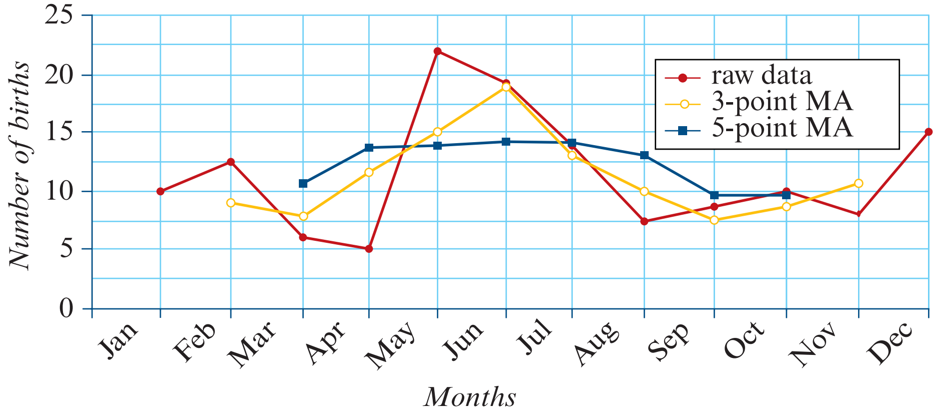
Month	*Number of births*	*3-point MA*	*5-point MA*
January	10		
February	12	$\frac{10+12+6}{3}=9.3$	
March	6	$\frac{12+6+5}{3}=7.7$	$\frac{10+12+6+5+22}{5}=11.0$
April	5	$\frac{6+5+22}{3}=11.0$	$\frac{12+6+5+22+18}{5}=12.6$
May	22	$\frac{5+22+18}{3}=15.0$	$\frac{6+5+22+18+13}{5}=12.8$
June	18	$\frac{22+18+13}{3}=17.7$	$\frac{5+22+18+13+7}{5}=13.0$
July	13	$\frac{18+13+7}{3}=12.7$	$\frac{22+18+13+7+9}{5}=13.8$
August	7	$\frac{13+7+9}{3}=9.7$	$\frac{18+13+7+9+10}{5}=11.4$
September	9	$\frac{7+9+10}{3}=8.7$	$\frac{13+7+9+10+8}{5}=9.4$
October	10	$\frac{9+10+8}{3}=9.0$	$\frac{7+9+10+8+15}{5}=9.8$
November	8	$\frac{10+8+15}{3}=11.0$	
December	15		

ISBN 978-1-009-25917-0

Cambridge University Press

Updated January 2024

The result of this smoothing can be seen in the plot below, which shows the raw data, the data smoothed with a three-point moving average and the data smoothed with a five-point moving average.



Note: In the process of smoothing, *data points are lost* at the beginning and end of the time series.

Two observations can be made from this plot:

1. smoothing using a five-point moving average is more effective in reducing the irregular fluctuations than using a three-point moving average.
2. the five-point moving average plot shows that there is no clear trend although the raw data suggest that there might be an increasing trend.

There are many ways of smoothing a time series. Moving averages of group size other than three and five are common and often very useful.

However, if we smooth over an even number of points, we run into a problem. The centre of the set of points is not at a time point belonging to the original series. Usually, we solve this problem by using a process called **centring**. Centring involves taking a two-point moving average of the already smoothed values so that they line up with the original time values. It is a two-step process.

Two-point centred moving average

We will illustrate the process by finding the two-point moving average, centred on Tuesday, for the daily temperature data opposite.

Day	*Temperature*
Monday	18.1
Tuesday	24.8
Wednesday	26.4

It is straightforward to calculate two point moving averages for this data by calculating the mean for Monday and Tuesday, followed by the mean for Tuesday and Wednesday. However, as we can see in the diagram over page, these means do not align with a particular day, but lie between days. We solve this problem by finding the average of these two means. This gives a smoothed value that is now centred on Tuesday.

We call this process smoothing using a two-point centred moving average (2-point CMA).

Day	*Temperature*	*2-point MA*	*2-point CMA*
Monday	18.1		
		$\frac{(18.1 + 24.8)}{2} = 21.45$	
Tuesday	24.8		$\frac{(21.45 + 25.6)}{2} = 23.525$
		$\frac{(24.8 + 26.4)}{2} = 25.60$	
Wednesday	26.4		

In practice, we do not have to draw such a diagram to perform these calculations. The purpose of doing so is to show how the centring process works. In practice, calculating two-point centred moving average is a much briefer and routine process as we illustrate in the following example.

Rather than using a two-step process for smoothing time series data involving centring the moving average, there is a short-cut for finding the centred moving average in one step.

This can be seen using the table below.

Day	*Temperature*	*2-point CMA*
Monday	18.1	
Tuesday	24.8	$\frac{(0.5 \times 18.1 + 24.8 + 0.5 \times 26.4)}{2} = 23.525$
Wednesday	26.4	

The idea behind this approach to centring a moving average when there is an even number of points is that we can pull the moving average (without centring) forward by half a point. So, rather than the moving average of two points occurring between the first and second point, the centred moving average can be centred on the second point using three points.

The two-point centred moving average

To smooth by a two-point centred moving average (2-point CMA) for three sequential data points y_1, y_2 and y_3, then:

$$\text{2-point CMA } y_2 = \frac{\frac{1}{2}y_1 + y_2 + \frac{1}{2}y_3}{2}$$

By using half of the first and last data point, this counts as only one data point, so effectively we have only used 2.

Example 3 Two-point centred moving averages

The temperatures (°C) recorded at a weather station at 9 a.m. each day for a week are displayed in the table.

Day	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Temperature	18.1	24.8	26.4	13.9	12.7	14.2	24.9

Calculate the two-point centred moving average temperature for Tuesday.

Solution

1 For two-point centred moving averages write the **three** data values centred on Tuesday (highlighted in red).

18.1 24.8 26.4

2 Calculate the mean of these three values by using half of the first value, the centre value and half of the last value, then dividing by 2.

$$\text{2-point CMA} = \frac{(0.5 \times 18.1 + 24.8 + 0.5 \times 26.4)}{2}$$

$$= 23.525$$

3 Write your answer.

The two-point centred moving average temperature on Tuesday is 23.5°C (to 1 d.p.)

The process for smoothing using a four-point centred moving average (4-point CMA) is the same as smoothing using a two-point centred moving average, except that the moving average is centred on the third point using five points.

The four-point centred moving average

To smooth by a four-point centred moving average (4-point CMA) for five sequential data points y_1, y_2, y_3, y_4, y_5, then:

$$\text{4-point CMA } y_3 = \frac{\frac{1}{2}y_1 + y_2 + y_3 + y_4 + \frac{1}{2}y_5}{4}$$

Example 4 Four-point centred moving averages

The table below gives the temperature (°C) recorded at a weather station at 9 a.m. each day for a week.

Day	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Temperature	18.1	24.8	26.4	13.9	12.7	14.2	24.9

Calculate the four-point centred moving average temperature for Thursday.

Solution

1 For four-point centred moving averages write the **five** data values centred on Thursday (highlighted in red).

24.8 26.4 13.9 12.7 14.2

2 Calculate the mean of these five values by using half of the first value, the centre three values and half of the last value, then dividing by 4.

$$\text{4-point CMA} = \frac{(0.5 \times 24.8 + 26.4 + 13.9 + 12.7 + 0.5 \times 14.2)}{4}$$

$$= 18.125$$

3 Write your answer.

The four-point centred moving average temperature on Thursday is 18.1°C (to 1 d.p.)

This process can be expanded for any even-point centred moving average. That is, to calculate an even-point centred moving average for a particular time period (n) we must use ($n + 1$) data points. We will also continue to divide by an even number when calculating the average.

The next step is to extend these computations to smooth all terms in the time series.

Skill-sheet

Exercise 7B

Basic skills

Example 1

1 The table below gives the temperature (°C) recorded at a weather station at 3.00 p.m. each day for a week.

Day	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Temperature (°C)	28.9	33.5	21.6	18.1	16.2	17.9	26.4

Calculate, to one decimal place, the:

a Three-point moving average for Wednesday

b Five-point moving average for Friday

c Seven-point moving average for Thursday

d Two-point centred moving average for Tuesday

e Four-point centred moving average for Friday

Calculating the smoothed value of individual data points

2

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

For the time series data in the table above, find:

a the three-point moving average y-value for $t = 4$

b the three-point moving average y-value for $t = 6$

c the three-point moving average y-value for $t = 2$

d the five-point moving average y-value for $t = 3$

e the five-point moving average y-value for $t = 7$

f the five-point moving average y-value for $t = 4$

g the two-point centred moving average y-value centred at $t = 3$

h the two-point centred moving average y-value centred at $t = 8$

i the four-point centred moving average y-value centred at $t = 3$

j the four-point centred moving average y-value centred at $t = 6$

Note: Copies of the tables in Questions 3 to 7 can be accessed via the skillsheet icon in the Interactive Textbook.

Smoothing a table of values

3 Complete the following table.

t	1	2	3	4	5	6	7	8	9
y	10	12	8	4	12	8	10	18	2
3-point MA of y	–								–
5-point MA of y	–	–						–	–

Smoothing and plotting a time series (three- and five-mean smoothing)

Example 2 **4** The maximum temperature of a city over a period of 10 days is given below.

CAS

Day	1	2	3	4	5	6	7	8	9	10
Temperature (°C)	24	27	28	40	22	23	22	21	25	26
3-point MA										
5-point MA										

a Use a CAS calculator to construct a time series plot of the temperature data.

b Use the three-point and five-point moving averages to smooth the data and complete the table.

c Use a CAS calculator to plot the smoothed temperature data, and compare and comment on the plots. This is best done if all plots are on the same graph.

5 The value of the Australian dollar in US dollars (exchange rate) over 10 days is given below.

Day	1	2	3	4	5	6	7	8	9	10
Exchange rate	0.743	0.754	0.737	0.751	0.724	0.724	0.712	0.735	0.716	0.711
3-point MA										
5-point MA										

ISBN 978-1-009-25917-0

a Construct a time series plot of the data. Label and scale the axes.

b Use the three-point and five-point moving averages to smooth the data and complete the table.

c Use a CAS calculator to plot the smoothed exchange rate data, and compare and comment on the plots. This is best done if all three plots are on the same graph.

Smoothing a time series (two- and four-mean smoothing)

Example 3 **6** Construct a table with four columns: 'Month', 'Number of births', 'Two-point moving average' and 'Two-point centred moving average' using the following data.

Month	Jan	Feb	Mar	Apr	May	Jun	July	Aug	Sep	Oct	Nov	Dec
Number of births	10	12	6	5	22	18	13	7	9	10	8	15

Example 4 **7** Construct a table with four columns: 'Month', 'Internet usage', 'Four-point moving average' and 'Four-point centred moving average' using the following data.

Month	Apr	May	Jun	July	Aug	Sep	Oct	Nov	Dec
Internet usage	21	40	52	42	58	79	81	54	50

7C Seasonal indices

When the data is seasonal, it is often necessary to **deseasonalise** it before further analysis. Firstly we will look at how they are used and why they are helpful, before we go on to calculate seasonal indices.

The concept of a seasonal index

Consider the (hypothetical) monthly seasonal indices for unemployment given in the table.

Jan	*Feb*	*Mar*	*Apr*	*May*	*Jun*	*Jul*	*Aug*	*Sept*	*Oct*	*Nov*	*Dec*	*Total*
1.1	1.2	1.1	1.0	0.95	0.95	0.9	0.9	0.85	0.85	1.1	1.1	12.0

Key fact 1

Seasonal indices are calculated so that their *average* is *1* (or 100%). This means that the *sum* of the seasonal indices equals the *number of seasons*.

Thus, if the seasons are months, the seasonal indices add to 12. If the seasons are quarters, then the seasonal indices would add to 4, and so on.

Key fact 2

Seasonal indices tell us how a particular season (generally a day, month or quarter) compares to the *average season*.

ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

For example:

- Seasonal index for unemployment for the month of February is 1.2 or 120%.
  This tells us that February unemployment figures tend to be 20% *higher* than the monthly average. Remember, the average seasonal index is 1 or 100%.
- Seasonal index for August is 0.90 or 90%.
  This tells us that the August unemployment figures tend to be only 90% of the monthly average. Alternatively, August unemployment figures are 10% *lower* than the monthly average.

We can use seasonal indices to remove the seasonal component (deseasonalise) from a time series, or to put it back in (**reseasonalise**).

Using seasonal indices to deseasonalise or reseasonalise a time series

To calculate deseasonalised figures, each entry is divided by its seasonal index as follows.

Deseasonalising data

Time series data are deseasonalised using the relationship:

$$\text{deseasonalised figure} = \frac{\text{actual figure}}{\text{seasonal index}}$$

Example 5 Using a seasonal index to deseasonalise data

The seasonal index (SI) for cold drink sales for summer is SI = 1.33.

Last summer a beach kiosk's actual cold drink sales totalled \$15 653.

What were the *deseasonalised* sales?

Solution

Use the rule

$$\text{deseasonalised sales} = \frac{\text{actual sales}}{\text{seasonal index}}$$

with actual sales = \$15 653
and SI = 1.33.

$$\text{Deseasonalised sales} = \frac{15\,653}{1.33} = 11\,769.17$$

The deseasonalised sales for summer were \$11 769.17.

The rule for determining deseasonalised data values can also be used to reseasonalise data – that is, convert a deseasonalised value into an actual data value.

Reseasonalising data

Time series data are reseasonalised using the rule:

actual figure = deseasonalised figure × seasonal index

ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

Example 6 Using a seasonal index to reseasonalise data

The seasonal index for cold drink sales for spring is SI = 0.85.

Last spring a beach kiosk's deseasonalised cold drink sales totalled \$10 870.

What were the *actual* sales?

Solution

Use the rule

actual sales = deseasonalised sales × seasonal index

with deseasonalised sales = \$10 870 and SI = 0.85.

$$\text{Actual sales} = 10\,870 \times 0.85$$
$$= 9239.50$$

The actual sales for spring were \$9239.50.

Example 7 Deseasonalising a time series

The quarterly sales figures for Mikki's shop over a 3-year period are given below.

Year	*Summer*	*Autumn*	*Winter*	*Spring*
1	920	1085	1241	446
2	1035	1180	1356	541
3	1299	1324	1450	659

Use the seasonal indices shown to deseasonalise these sales figures. Write answers correct to the nearest whole number.

Summer	*Autumn*	*Winter*	*Spring*
1.03	1.15	1.30	0.52

Solution

1 To deseasonalise each sales figure in the table, divide by the appropriate seasonal index.
For example, for summer, divide the figures in the 'Summer' column by 1.03.
Round results to the nearest whole number.

$$\frac{920}{1.03} = 893$$
$$\frac{1035}{1.03} = 1005$$
$$\frac{1299}{1.03} = 1261$$

2 Repeat for the other seasons.

Deseasonalised sales figures

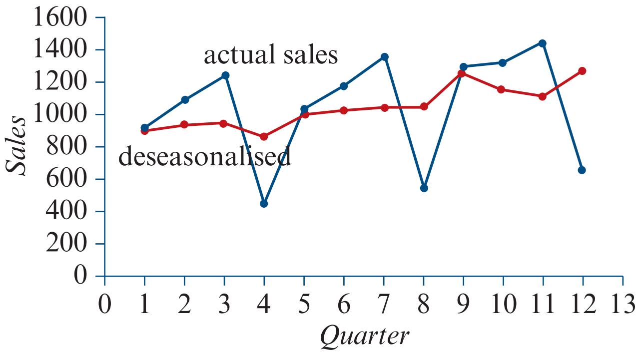
Year	Summer	Autumn	Winter	Spring
1	893	943	955	858
2	1005	1026	1043	1040
3	1261	1151	1115	1267

Comparing a plot of the raw data with the deseasonalised data

The plot below shows the time series deseasonalised sales.

Two things to be noticed are that deseasonalising has:

- removed the seasonality from the time series plot
- revealed a clear underlying trend in the data.



It is common to deseasonalise time series data before you fit a trend line.

Calculating seasonal indices by using the average percentage method

To complete this section, you will now learn to calculate a seasonal index. We will start by using only 1 year's data to illustrate the basic ideas and then move onto a more realistic example where several years' data are involved.

Example 8 Calculating seasonal indices (1 year's data)

Mikki runs a shop and she wishes to determine quarterly seasonal indices based on last year's sales, which are shown in the table opposite.

Summer	*Autumn*	*Winter*	*Spring*
920	1085	1241	446

Solution

1 The seasons are quarters. Write the formula in terms of quarters.

$$\text{Seasonal index} = \frac{\text{value for season}}{\text{seasonal average}}$$

2 Find the quarterly average for the year.

$$\text{Quarterly average} = \frac{920 + 1085 + 1241 + 446}{4}$$
$$= 923$$

3 Work out the seasonal index (SI) for each time period.

$$SI_{Summer} = \frac{920}{923} = 0.997$$
$$SI_{Autumn} = \frac{1085}{923} = 1.176$$
$$SI_{Winter} = \frac{1241}{923} = 1.345$$
$$SI_{Spring} = \frac{446}{923} = 0.483$$

4 Check that the seasonal indices sum to 4 (the number of seasons). The slight difference is due to rounding error.

Check: $0.997 + 1.176 + 1.345 + 0.483 = 4.001$

5 Write out your answers as a table of the seasonal indices.

Seasonal indices

Summer	Autumn	Winter	Spring
0.997	1.176	1.345	0.483

The next example illustrates how seasonal indices are calculated with 3 years' data. While the process looks more complicated, we just repeat what we did in Example 8 three times and average the results for each year at the end.

Example 9 Calculating seasonal indices (several years' data)

Suppose that Mikki has 3 years of data, as shown. Use the data to calculate seasonal indices, correct to two decimal places.

Year	*Summer*	*Autumn*	*Winter*	*Spring*
1	920	1085	1241	446
2	1035	1180	1356	541
3	1299	1324	1450	659

Solution

The strategy is as follows:

- Calculate the seasonal indices for years 1, 2 and 3 separately, as Example 8 (as we already have the seasonal indices for year 1 from Example 8, we will save ourselves some time by simply quoting the result).
- Average the three sets of seasonal indices to obtain a single set of seasonal indices.

1 Write the result for year 1.

Year 1 seasonal indices:

Summer	Autumn	Winter	Spring
0.997	1.176	1.345	0.483

2 Now calculate the seasonal indices for year 2.

a The seasons are quarters. Write the formula in terms of quarters.

$$\text{Seasonal index} = \frac{\text{value for quarter}}{\text{quarterly average}}$$

b Find the quarterly average for the year.

$$\text{Quarterly average} = \frac{1035 + 1180 + 1356 + 541}{4}$$
$$= 1028$$

c Work out the seasonal index (SI) for each time period.

$$SI_{Summer} = \frac{1035}{1028} = 1.007$$
$$SI_{Autumn} = \frac{1180}{1028} = 1.148$$
$$SI_{Winter} = \frac{1356}{1028} = 1.319$$
$$SI_{Spring} = \frac{541}{1028} = 0.526$$

d Check that the seasonal indices sum to 4.

Check: $1.007 + 1.148 + 1.319 + 0.526 = 4.000$

e Write your answers as a table of the seasonal indices.

Year 2 seasonal indices:

Summer	Autumn	Winter	Spring
1.007	1.148	1.319	0.526

3 Now calculate the seasonal indices for year 3.

a Find the quarterly average for the year.

$$\text{Quarterly average} = \frac{1299 + 1324 + 1450 + 659}{4}$$
$$= 1183$$

b Work out the seasonal index (SI) for each time period.

$$SI_{Summer} = \frac{1299}{1183} = 1.098$$
$$SI_{Autumn} = \frac{1324}{1183} = 1.119$$
$$SI_{Winter} = \frac{1450}{1183} = 1.226$$
$$SI_{Spring} = \frac{659}{1183} = 0.557$$

c Check that the seasonal indices sum to 4.

Check: $1.098 + 1.119 + 1.226 + 0.557 = 4.000$

d Write your answers as a table of the seasonal indices.

Year 3 seasonal indices:

Summer	Autumn	Winter	Spring
1.098	1.119	1.226	0.557

ISBN 978-1-009-25917-0

Cambridge University Press

Updated January 2024

4 Find the 3-year averaged seasonal indices by averaging the seasonal indices for each season.

Final seasonal indices:

$$S_{\text{Summer}} = \frac{0.997 + 1.007 + 1.098}{3} = 1.03$$

$$S_{\text{Autumn}} = \frac{1.176 + 1.148 + 1.119}{3} = 1.15$$

$$S_{\text{Winter}} = \frac{1.345 + 1.319 + 1.226}{3} = 1.30$$

$$S_{\text{Spring}} = \frac{0.483 + 0.526 + 0.557}{3} = 0.52$$

5 Check that the seasonal indices sum to 4.

Check: $1.03 + 1.15 + 1.30 + 0.52 = 4.00$

6 Write your answers as a table of the seasonal indices.

Summer	Autumn	Winter	Spring
1.03	1.15	1.30	0.52

Interpreting the seasonal indices

Having calculated these seasonal indices, what do they tell us in the above situation?

The seasonal index of:

- 1.03 for summer tells us that summer sales are typically 3% above average
- 1.15 for autumn tells us that autumn sales are typically 15% above average
- 1.30 for winter tells us that winter sales are typically 30% above average
- 0.52 for spring tells us that spring sales are typically 48% below average.

Correcting for seasonality

Also, using the rule

$$\text{deseasonalised figure} = \frac{\text{actual figure}}{\text{seasonal index}}$$

we can work out how much we need to increase or decrease the actual sales figures to correct for seasonality.

For example, we see that for winter:

$$\begin{aligned}\text{deseasonalised figure} &= \frac{\text{actual figure}}{1.30} \\ &= 0.769\ldots \times \text{actual figure} \approx 77\% \text{ of the actual figures}\end{aligned}$$

Thus, to correct the seasonality in winter, we need to decrease the actual sales by about 23%.

Similarly we can show that, to correct for seasonality in spring ($\text{SI}_{\text{spring}} = 0.52$), we need to increase the actual spring sales figure by around 92% $\left(\frac{1}{0.52} \approx 1.92\right)$.

ISBN 978-1-009-25917-0

Skill-sheet

Exercise 7C

Basic skills and interpretation

1 The table below shows the monthly sales figures (in $'000s) and seasonal indices (for January to November) for a product produced by the U-beaut company.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Seasonal index	1.2	1.3	1.1	1.0	1.0	0.9	0.8	0.7	0.9	1.0	1.1	
Sales ($'000s)	9.6	10.5	8.6		7.1	6.0	5.4		6.4	7.2	8.3	7.4

a Calculate the seasonal index for December, to one decimal place.

b Calculate the deseasonalised sales (in $'000s) for March, to one decimal place.

c Calculate the deseasonalised sales (in $'000s) for June, to one decimal place.

d The deseasonalised sales (in $'000s) for August are 5.6. Determine the actual sales for August.

e The deseasonalised sales (in $'000s) for April are 6.9. Determine the actual sales for April.

f By what percentage do the sales figures for February tend to exceed the average monthly sales?

g By what percentage do the sales figures for September tend to be less than the average monthly sales?

h The seasonal index for January is 1.2. To correct the actual monthly sales figure for seasonality we need to:

- **A** decrease the actual sales figures by around 20%
- **B** increase the actual sales figures by around 20%
- **C** decrease the actual sales figures by around 17%
- **D** increase the actual sales figures by around 17%
- **E** increase the actual sales figures by around 80%

2 The table below shows the quarterly newspaper sales (in $'000s) of a corner store. Also shown are the seasonal indices for newspaper sales for the first, second and third quarters.

	Quarter 1	*Quarter 2*	*Quarter 3*	*Quarter 4*
Sales		1060	1868	1642
Seasonal index	0.8	0.7	1.3	

a Calculate the seasonal index for quarter 4.

b Calculate the deseasonlised sales (in $'000s) for quarter 2, to the nearest whole number.

ISBN 978-1-009-25917-0

c Calculate the deseasonalised sales (in $'000s) for quarter 3, to the nearest whole number.

d The deseasonalised sales (in $'000s) for quarter 1 are 1256. Determine the actual sales for quarter 1, to the nearest whole number.

Deseasonalising a time series

3 The following table shows the number of students enrolled in a three-month computer systems training course along with some seasonal indices that have been calculated from the previous year's enrolment figures. Complete the table by calculating the seasonal index for spring and the deseasonalised student numbers for each course.

	Summer	*Autumn*	*Winter*	*Spring*
Number of students	56	125	126	96
Deseasonalised numbers				
Seasonal index	0.5	1.0	1.3	

4 The number of waiters employed by a restaurant chain in each quarter of one year, along with some seasonal indices that have been calculated from the previous year's data, are given in the following table.

	Quarter 1	*Quarter 2*	*Quarter 3*	*Quarter 4*
Number of waiters	198	145	86	168
Seasonal index	1.30		0.58	1.10

a What is the seasonal index for the second quarter?

b The seasonal index for quarter 1 is 1.30. Explain what this means in terms of the average quarterly number of waiters.

c Deseasonalise the data.

Calculating seasonal indices

Example 8

5 The table below records quarterly sales (in $'000s) for a shop.

Quarter 1	*Quarter 2*	*Quarter 3*	*Quarter 4*
60	56	75	78

Use the data to determine the seasonal indices for the four quarters. Give your results correct to two decimal places. Check that your seasonal indices add to 4.

6 The quarterly sales figures for Jen & Berry's Ice Cream over a four-year period are shown in the table below.

Year	*Summer*	*Autumn*	*Winter*	*Spring*
1	890	834	750	876
2	918	862	764	904
3	1030	848	778	883
4	988	960	869	946

Use the data to calculate the seasonal indices, correct to two decimal places.

7 The daily price of unleaded fuel (in cents per litre) for a petrol station in Yokine for three weeks in June is given in the table below.

Week	*Day*	*t*	*ULP Price*	*Weekly mean*	*Percentage of Weekly mean*	*Deseasonalised (d)*
1	Sunday, 6/6	1	127.7	*A*	93.5	135.9
	Monday, 7/6	2	126.7		92.8	136.2
	Tuesday, 8/6	3	126.7		92.8	137.0
	Wednesday, 9/6	4	149.7		109.6	136.8
	Thursday, 10/6	5	149.7		109.6	137.1
	Friday, 11/6	6	139.7		*B*	*D*
	Saturday, 12/6	7	135.7		99.4	137.4
2	Sunday, 13/6	8	129.7	138.3	93.8	138.0
	Monday, 14/6	9	128.7		93.1	138.4
	Tuesday, 15/6	10	126.7		91.6	*E*
	Wednesday, 16/6	11	149.7		108.3	136.8
	Thursday, 17/6	12	149.7		108.3	137.1
	Friday, 18/6	13	147.7		106.8	143.2
	Saturday, 19/6	14	135.7		98.1	137.4
3	Sunday, 20/6	15	131.7	139.3	*C*	140.2
	Monday, 21/6	16	129.7		93.1	139.5
	Tuesday, 22/6	17	129.7		93.1	140.2
	Wednesday, 23/6	18	153.7		110.4	140.5
	Thursday, 24/6	19	152.7		109.6	139.9
	Friday, 25/6	20	139.7		100.3	135.4
	Saturday, 26/6	21	137.7		98.9	139.4

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

a Determine the missing values A, B and C in the table.

b Complete the table below calculating the seasonal indices for Tuesday and Wednesday.

Day	Sun	Mon	Tue	Wed	Thu	Fri	Sat
Seasonal index	94.0	93.0			109.2	103.1	98.8

c Using the appropriate seasonal index, determine the missing values D and E to complete the table.

7D Fitting a trend line and forecasting

Fitting a trend line

If there appears to be a linear trend, we can use the least squares method to fit a line to the data to model the trend.

Example 10 Fitting a least squares trend line to a time series plot (no seasonality)

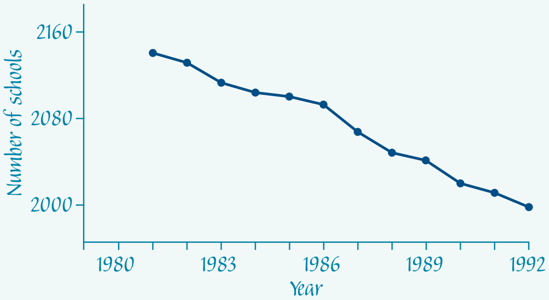
Fit a trend line to the data in the following table, which shows the number of government schools in Victoria over the period 1981–92, and interpret the slope.

Year (t)	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
Number (n)	2149	2140	2124	2118	2118	2114	2091	2064	2059	2038	2029	2013

Solution

1 Construct a time series plot of the data to confirm that the trend is linear.

Note: For convenience we let 1981 = 1, 1982 = 2 and so on when entering the data into a calculator.



2 Fit a least squares trend line to the data with *year* as the EV. Write its equation.

$n = -12.5t + 2169$

3 Write the slope and interpret.

Slope = −12.5
Over the period 1981–92 the number of schools in Victoria decreased at an average rate of 12.5 schools per year.

Forecasting

Using a trend line fitted to a time series plot to make predictions about future values is known as **trend line forecasting**.

Example 11 Forecasting

Example 10 shows a decreasing trend for the number of schools in Victoria over the period 1981–92. If that trend had continued, how many government schools do we predict would have been in Victoria in 2015? Give your answer correct to the nearest whole number.

Solution

Substitute the appropriate value for *year* in the equation determined using a least squares regression. Since 1981 was designated as year ‘1’, then 2015 is year ‘35’.

When $t = 35$;

$$\begin{aligned} n &= -12.5t + 2169 \\ &= -12.5 \times 35 + 2169 \\ &= 1731.5 \\ &\approx 1732 \text{ Schools} \end{aligned}$$

Note: As with any prediction involving extrapolation, the results obtained when predicting well beyond the range of the data should be treated with caution.

Forecasting taking seasonality into account

When time series data is seasonal, it is usual to deseasonalise the data before fitting the trend line.

Example 12 Fitting a least squares trend line to a time series with seasonality

The *deseasonalised* quarterly sales data from Mikki’s shop are shown below.

Season	Summer	Autumn	Winter	Spring	Summer	Autumn
Quarter (n)	1	2	3	4	5	6
Sales (s)	893	943	955	858	1005	1026

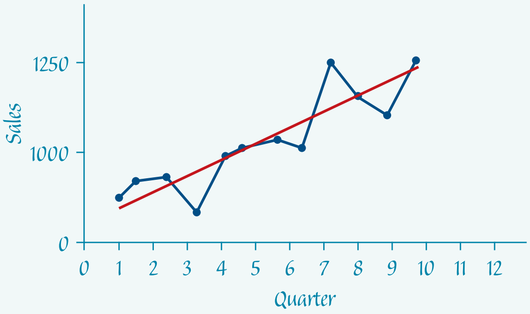
Season	Winter	Spring	Summer	Autumn	Winter	Spring
Quarter (n)	7	8	9	10	11	12
Sales (s)	1043	1040	1261	1151	1115	1267

Fit a trend line and interpret the slope.

Solution

1 Plot the time series.

2 Using the calculator (with *Quarter* as the explanatory and *Sales* as the response variable), find the equation of the least squares line. Plot it on the time series.



ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

3 Write the equation of the least squares trend line.	$S = 32.1n + 838.0$
4 Interpret the slope in terms of the variables involved.	Over the 3-year period, sales at Mikki's shop increased at an average rate of 32.1 sales per quarter.

Making predictions with deseasonalised data

When using deseasonalised data to fit a trend line, you must remember that the result of any prediction is a deseasonalised value. To be meaningful, this result must then be **reseasonalised** by multiplying by the appropriate seasonal index.

Example 13 Forecasting (seasonality)

What sales do we predict for Mikki's shop in the winter of year 4? Use Example 12 and 9 for the calculation. (Because many items have to be ordered well in advance, retailers often need to make such decisions.)

Solution

1 Substitute the appropriate value for the time period in the equation for the trend line. Since summer year 1 was designated as quarter '1', then winter year 4 is quarter '15'.	When $n = 15$; $\hat{S} = 32.1n + 838.0$ $= 32.1 \times 15 + 838.0$ $= 1319.5$ Deseasonalised sales predicted for winter of year 4 = 1319.5
2 The value just calculated is the deseasonalised sales figure for the quarter in question. To obtain the *actual* predicted sales figure we need to reseasonalise this predicted value. To do this, we multiply this value by the seasonal index for winter, which is 1.30 from Example 9.	Reseasonalise = prediction × seasonal index $= 1319.5 \times 1.30$ $= 1715.35$ Seasonalised sales prediction for winter of year 4 ≈ 1715

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Exercise 7D

Fitting a least squares trend line to a time series plot (no seasonality)

CAS

1 The data below shows the number of students enrolled (in thousands) at university in Australia for the period 1992–2001.

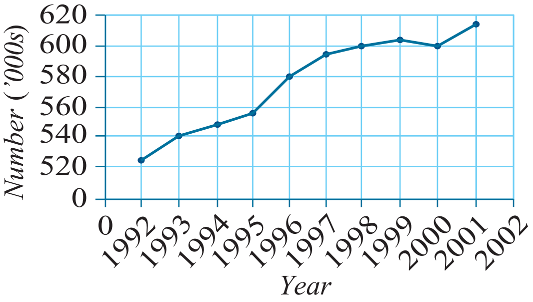
Year	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
Number	525	539	545	556	581	596	600	603	600	614

The time series plot of the data is shown below.

a Comment on the plot.

b Fit a least squares trend line to the data, using 1992 as year 1, and interpret the slope.

c Use this equation to predict the number of students expected to enroll at university in Australia in 2020.



Example 10
Example 11

2 The table below shows the percentage of total retail sales that were made in department stores over an 11-year period.

Year	1	2	3	4	5	6	7	8	9	10	11
Sales (%)	12.3	12.0	11.7	11.5	11.0	10.5	10.6	10.7	10.4	10.0	9.4

a Construct a time series plot.

b Comment on the time series plot in terms of trend.

c Fit a trend line to the time series plot, find its equation and interpret the slope.

d Draw the trend line on your time series plot.

e Use the trend line to forecast the percentage of retails sales which will be made by department stores in year 15.

3 The average ages of mothers having their first child in Australia over the years 1989–2002 are shown below.

Year	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Age	27.3	27.6	27.8	28.0	28.3	28.5	28.6	28.8	29.0	29.1	29.3	29.5	29.8	30.1

a Fit a least squares trend line to the data, using 1989 as year 1, and interpret the slope.

b Use this trend relationship to forecast the average ages of mothers having their first child in Australia in 2018 (year 30). Explain why this prediction is not likely to be reliable.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Fitting a least squares line to a time series with seasonality

4 **a** The table below shows the *deseasonalised* quarterly washing-machine sales of a company over 3 years. Use least squares regression to fit a trend line to the data.

	Year 1				*Year 2*				*Year 3*			
Quarter number	1	2	3	4	5	6	7	8	9	10	11	12
Deseasonalised	53	51	54	55	64	64	61	63	67	69	68	66

b Use this trend equation for washing-machine sales, with the seasonal indices below, to forecast the sales of washing machines in the fourth quarter of year 4.

Quarter	1	2	3	4
Seasonal index	0.90	0.81	1.11	1.18

5 The sale of boogie boards for a surf shop over a 2-year period is given in the table.

Year	*Quarter 1*	*Quarter 2*	*Quarter 3*	*Quarter 4*
1	138	60	73	230
2	283	115	163	417

The quarterly seasonal indices are given below.

Seasonal index	1.13	0.47	0.62	1.77

a Use the seasonal indices to calculate the deseasonalised sales figures for this period to the nearest whole number.

b Plot the actual sales figures and the deseasonalised sales figures for this period and comment on the plot.

c Fit a trend line to the deseasonalised sales data. Write the slope and intercept correct to three significant figures.

d Use the relationship calculated in **c**, together with the seasonal indices, to forecast the sales for the first quarter of year 4. Remember to reseasonalise your prediction.

6 The table shows the daily price of unleaded fuel (in cents per litre) for a petrol station.

	Sun	*Mon*	*Tue*	*Wed*	*Thu*	*Fri*	*Sat*
Week 1	137.7	135.7	133.7	175.7	165.7	149.7	141.7
Week 2	139.7	137.7	135.7	175.7	163.7	151.7	139.7
Week 3	135.7	135.7	135.7	175.7	163.7	155.7	143.7

The daily seasonal indices are given below.

	Sun	*Mon*	*Tue*	*Wed*	*Thu*	*Fri*	*Sat*
Seasonal index (%)	92.4	91.5	90.6	117.9	110.3	102.2	95.1

a Use the seasonal indices to calculate the deseasonalised fuel prices for this period to one decimal place.

b Determine the equation for the least squares regression line for the deseasonalised fuel prices. Let t represent the time, where Sunday Week 1 is $t = 1$. Write the slope and intercept correct to three significant figures.

c Predict the price of unleaded fuel, to one decimal place, for Monday of Week 4.

7 A university records its enrolment numbers each quarter over a four-year period.

Year	*Season*	*Quarter (t)*	*Enrolments (n)*	*Seasonal mean*	*Seasonal indices*	*Deseasonalised (d)*
2014	Summer	1	1201	1015.5	1.18	1018
	Autumn	2	1053		1.04	1012
	Winter	3	830		0.82	1012
	Spring	4	979		0.96	1020
2015	Summer	5	1221	1038.25	1.18	1035
	Autumn	6	1062		1.02	1041
	Winter	7	853		0.82	1040
	Spring	8	1017		0.98	1038
2016	Summer	9	1227	1051.75	1.17	1049
	Autumn	10	1084		1.03	1052
	Winter	11	876		0.83	1055
	Spring	12	1020		0.97	1052
2017	Summer	13	1238	1073.0	1.15	1077
	Autumn	14	1104		1.03	1072
	Winter	15	907		0.85	1067
	Spring	16	1043		0.97	1075

a Calculate the season indices for each quarter.

	Summer	Autumn	Winter	Spring
Seasonal index				

b Determine the equation for the least squares regression line for the deseasonalised enrolment figures. Write the slope and intercept correct to two decimal places.

c Predict the number of university enrolments for Winter 2018.

ISBN 978-1-009-25917-0

Key ideas and chapter summary

Assignment

Time series data	**Time series data** are a collection of data values along with the times (in order) at which they were recorded.
Time series plot	A **time series plot** is a line graph where the values of the response variable are plotted in time order.
Features to look for in a time series plot	■ Trend ■ Seasonality ■ Possible outliers ■ Cycles ■ Structural change ■ Irregular (random) fluctuations
Trend	**Trend** is present when there is a long-term upward or downward movement in a time series.
Cycles	**Cycles** are present when there is a periodic movement in a time series. The period is the time it takes for one complete up and down movement in the time series plot. This term is generally reserved for periodic movements with a period greater than one year.
Seasonality	**Seasonality** is present when there is a periodic movement in a time series that has a calendar related period – for example, a year, a month, a week.
Structural change	**Structural change** is present when there is a sudden change in the established pattern of a time series plot.
Outliers	**Outliers** are present when there are individual values that stand out from the general body of data.
Irregular (random) fluctuations	**Irregular (random) fluctuations** are always present in any real-world time series plot. They include all of the variations in a time series that we cannot reasonably attribute to systematic changes like trend, cycles, seasonality, structural change or the presence of outliers.
Smoothing	**Smoothing** is a technique used to eliminate some of the irregular fluctuations in a time series plot so that features such as trend are more easily seen.
Moving averages	In a time series, a **moving average** is a method used to smooth the data, whereby each original data value is replaced by the mean of itself and a number of data values on either side. When smoothing over an even number of data points, centring is required to ensure the moving average is centred on the chosen point of time.

Seasonal indices	**Seasonal indices** are used to quantify the seasonal variation in a time series.
Deseasonalise	The process of accounting for the effects of seasonality in a time series is called **deseasonalisation**.
Reseasonalise	The process of a converting seasonal data back into its original form is called **reseasonalisation**.
Trend line forecasting	**Trend line forecasting** uses the equation of a trend line to make predictions about the future.

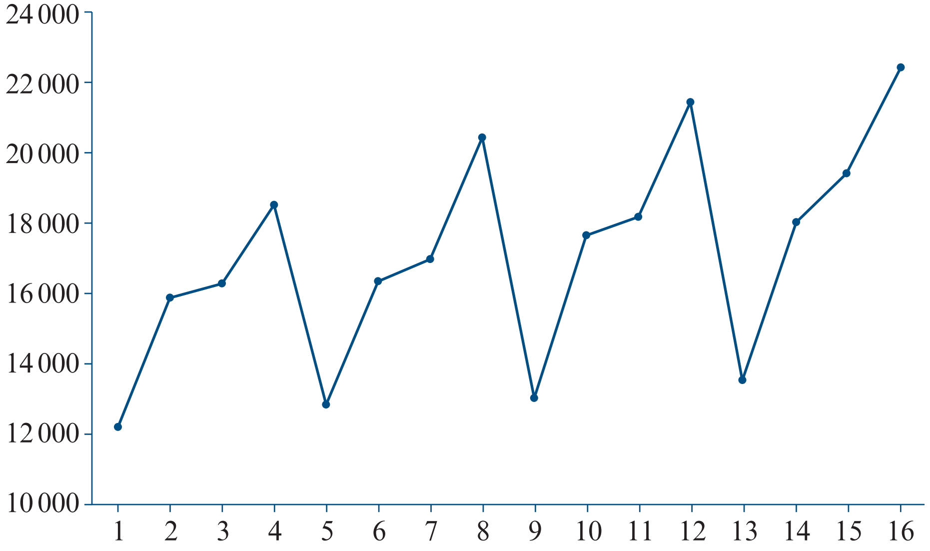
Skills check

Having completed this chapter you should be able to:

- recognise time series data
- construct a times series plot
- identify the presence of trend, cycles, seasonality, structural change and irregular (random) fluctuations in a time series plot
- smooth a time series using moving averges to help identify any overall trend
- calculate and interpret seasonal indices
- calculate and interpret a trend line for linear trends
- use a trend line to make forecasts.

Short-answer questions

1 Describe the pattern in the time series graph below in terms of trend and seasonality.



Use the following table to answer Question 2

Time period	1	2	3	4	5	6
Data value	2.3	3.4	4.4	2.7	5.1	3.7

2 For the time series data in the table above, calculate, to one decimal place, the:

a Three-point moving average for time period 2

b Five-point moving average for time period 3

c Two-point centred moving average for time period 5

d Four-point centred moving average for time period 4

3 The seasonal indices for the number of customers at a restaurant are as follows.

Jan	*Feb*	*Mar*	*Apr*	*May*	*Jun*	*Jul*	*Aug*	*Sep*	*Oct*	*Nov*	*Dec*
1.0	p	1.1	0.9	1.0	1.0	1.2	1.1	1.1	1.1	1.0	0.7

Determine the value of p.

4 The seasonal indices for the number of wetsuits sold as a surf shop are given in the table.

Quarter	Summer	Autumn	Winter	Spring
Seasonal index	1.8	0.4	0.3	1.5

a The number of wetsuits sold one summer is 432. Determine the deseasonalised number of wetsuits sold that summer.

b The deseasonalised number of wetsuits sold one winter was 380. Determine the actual number of wetsuits sold that winter.

c Describe what the value of the seasonal index for Spring indicates about the number of wetsuits sold in Spring compared to the yearly average.

5 The number of visitors to an information centre each quarter was recorded for one year. The results are tabulated below.

Quarter	Summer	Autumn	Winter	Spring
Visitors	1048	677	593	998

Using this data, determine the seasonal index for Autumn, to two decimal places.

6 A trend line is fitted to a time series plot displaying the average age at marriage of males (a, in years) for the period 1995–2002.

The equation of this line is: $a = 0.236t + 27.1$

Here, t is the year where $t = 1$ is 1995, $t = 2$ is 1996, and so on.

a Using this trend line, predict the average age of marriage of males in 2004.

b Interpret the value of the slope in context.

ISBN 978-1-009-25917-0

7 Suppose that the seasonal indices for the wholesale price of petrol are:

Day	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Index	1.2	1.0	0.9	0.8	0.7	1.2	1.2

The daily deseasonalised prices for a petrol outlet for a week (in cents/litre) are given in the following table.

Day	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Price	88.3	85.4	86.7	88.5	90.1	91.7	94.6

a Determine the equation of the least squares line that could enable us to predict the deseasonalised price.

b Determine the seven-point moving average price of petrol (in cents/litre) for Wednesday this week, to one decimal place.

c The deseasonalised (in cents/litre) price on Thursday was 90.1. Calculate the actual price on Thursday.

Extended-response questions

CAS

1 The number of workers in a city building who were smokers, as a percentage of the total number of workers in the building, was recorded every 5 years from 1985 to 2010. The results are shown in the table below.

Year	1985	1990	1995	2000	2005	2010
Smokers (%)	35.4	31.8	24.1	19.7	18.2	17.3

a Construct a time series plot of this data.

b What general trend exists in this data?

The table below shows the variable *year* rescaled and with a new variable name, *time*
The year 1985 = 0, 1990 = 5 and so on.

Year	1985	1990	1995	2000	2005	2010
Time (t)	0	5	10	15	20	25
Smokers (s, %)	35.4	31.8	24.1	19.7	18.2	17.3

c Write the equation of the least squares line that can be used to predict the percentage of workers in the building who were *smokers* from *time*. The explanatory variable is *time*.

d Use the least squares line calculated in **c** above to predict the percentage of workers in the building who were smokers in 2012.

e In 2007, 17.9% of the workers in the building were smokers. Calculate the residual in the prediction determined from the least squares line calculated in **c** above.

ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

2 The table below shows the average interest rate for the period 1987–1997. Also shown are the three-point moving average interest rates but with one missing.

Year	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
Rate (%)	15.50	13.50	17.00	16.50	13.00	10.50	9.50	8.75	10.50	8.75	7.55
3-point MA(%)		15.33	15.67	15.50	13.33		9.58	9.58	9.33	8.93	

a Complete the table by calculating the three-point moving average interest rate for 1992.

b Construct a time series plot for the average interest rate during the period 1987–1997.

c Plot the smoothed interest rate data on the graph and comment on any trend.

3 A real estate agent recorded the number of property sales in the table below.

Year	*Quarter*	*t*	*Sales*	*Yearly mean*	*Seasonal indices*	*Deseasonalised*
2011	Jan – Mar	1	225	*A*	*B*	185.52
	Apr – Jun	2	195		1.06	188.16
	Jul – Sep	3	181		0.98	181.93
	Oct – Dec	4	136		0.74	179.90
2012	Jan – Mar	5	215	176	1.22	177.27
	Apr – Jun	6	*C*		1.03	175.62
	Jul – Sep	7	177		1.01	177.91
	Oct – Dec	8	130		0.74	171.97
2013	Jan – Mar	9	205	167	1.23	169.03
	Apr – Jun	10	170		1.02	*D*
	Jul – Sep	11	164		0.98	164.84
	Oct – Dec	12	129		0.77	170.65
2014	Jan – Mar	13	186	157.5	1.18	153.36
	Apr – Jun	14	163		1.03	157.29
	Jul – Sep	15	159		1.01	*E*
	Oct – Dec	16	122		0.77	161.39

CAS

a Determine the missing values *A*, *B* and *C* in the table above.

b Complete the table below calculating the seasonal indices for the first quarter (January – March) and the last quarter (October – December)

Quarter	Jan – Mar	Apr – Jun	Jul – Sep	Oct – Dec
Seasonal index		1.04	0.99	

ISBN 978-1-009-25917-0

c Using the appropriate seasonal index, determine the missing values D and E to complete the table.

d Fit a least squares trend line to the deseasonalised value and write the equation for the trend line.

e **i** Use the trend line and seasonal indices to predict the sales for the first quarter of 2016.

ii Comment on the validity of this prediction.

4 The table below shows the seasonal indices for the number of customers at a restaurant.

Day	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Seasonal index	0.6344	0.6614	0.8332	1.0357	1.2825		1.0698

a The seasonal index for Saturday is missing from the table. Determine the value of the seasonal index for Saturday.

b The actual number of customers at the restaurant on Tuesday is 552. Calculate the deseasonalised number of customers at the restaurant on Tuesday.

A trend line can be used to forecast the deseasonalised number of customers at the restaurant. The equation for the trend line is given by:

$$\textit{number of customers} = 0.81 \times t + 831.23$$

where t is the day number and $t = 1$ is the Monday of Week 1.

c Calculate the actual number of customers at the restaurant on Friday of Week 6.

8

Loans, investments and annuities

In this chapter

Syllabus references

Topics: Compound interest loans and investments; Reducing balance loans (compound interest loans with periodic repayments); Annuities and perpetuities (compound interest investments with periodic payments made from investment)

Subtopics: 4.2.1 – 4.2.7

ISBN 978-1-009-25917-0

8A Nominal and effective interest rates

Nominal interest rate

Compound interest rates are usually quoted as annual rates, or interest rate per annum. This rate is called the **nominal interest rate** for the investment or loan. Sometimes an annual rate might be quoted, but the interest can be calculated and paid according to a different time period, such as monthly. The time period after which compound interest is calculated and paid is called the **compounding period**.

The terms of a compound interest loan or investment are usually quoted as a nominal interest rate followed by a compounding period. The interest rate for the compounding period is easily calculated using simple arithmetic.

It must be assumed that there are:

- 12 equal months in every year (even though some months have different numbers of days)
- 4 quarters in every year (a quarter is equal to 3 months)
- 26 fortnights in a year (even though there are slightly more than this)
- 52 weeks in a year (even though there are slightly more than this)
- 365 days in a year (ignore the existence of leap years).

A nominal interest rate is converted to a compounding interest rate by *dividing* by these numbers.

Example 1 Converting nominal interest rates to compounding interest rates

An investment account will pay interest at the rate of 3.6% per annum. Convert this interest rate to:

a a monthly rate

b a fortnightly rate

c a quarterly rate

Solution

Annual interest rates are converted to rates for other compounding periods by dividing by the relevant number.

a Divide by 12. $\text{Monthly interest rate} = \dfrac{3.6\%}{12} = 0.3\%$

b Divide by 26. $\text{Fortnightly interest rate} = \dfrac{3.6\%}{26} = 0.138\% \text{ to 3 d.p.}$

c Divide by 4. $\text{Quarterly interest rate} = \dfrac{3.6\%}{4} = 0.9\%$

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Effective interest rates

As a general principle with compound interest, the more frequently interest is calculated and added to your investment or loan (the compounding period), the more rapidly the value of your investment or loan increases.

This is illustrated in the next table, which compares the value of a $5000 investment paying a nominal interest rate of 4.8% per annum with the value of the investment if the interest is calculated on a quarterly or monthly basis, rather than just yearly.

Principal of investment: $5000

Nominal annual interest rate: 4.8%

	Value of investment for interest earned at the rate of:		
Month	*4.8% per annum*	*1.2% per quarter*	*0.4% per month*
0	5000.00	5000.00	5000.00
1			5020.00
2			5040.08
3		5060.00	5060.24
4			5080.48
5			5100.80
6		5120.72	5121.21
7			5141.69
8			5162.26
9		5182.17	5182.91
10			5203.64
11			5224.45
12	5240.00	5244.35	5245.35
Total interest earned*	240.00	244.35	245.35
Effective annual interest rate	4.80%	4.89%	4.91%

*Note that the total interest earned is the value of the investment at the end of the year less the principal.

As you can see from the table, the more frequently interest is calculated and added, the greater the value of the investment at the end of the year. For this investment, the amounts ranged from $240.00 when interest is calculated and added annually to $245.35 when interest is calculated and added monthly.

What this means is that an investment where the interest is calculated and added monthly has a higher effective annual interest rate than an investment where interest is calculated and added yearly. The effective interest rate is determined as follows.

The interest earned can be expressed as a percentage of the original investment. This value, called the **effective interest rate**, can be used to compare the investment performance.

$$\text{Compounding } \textit{yearly}\text{: effective rate} = \frac{240}{5000} \times 100\% = 4.80\% \text{ per annum}$$

$$\text{Compounding } \textit{quarterly}\text{: effective rate} = \frac{244.35}{5000} \times 100\% = 4.89\% \text{ per annum}$$

$$\text{Compounding } \textit{monthly}\text{: effective rate} = \frac{245.35}{5000} \times 100\% = 4.91\% \text{ per annum}$$

The effective rate for an investment at 4.8% per annum, compounding quarterly, is 4.89%. This means that, after 1 year, the interest earned will be 4.89% of the amount invested. After 1 year, the interest earned by the monthly compounding investment will be 4.91% of the amount invested. The investment that compounds monthly earns more interest in a year.

Instead of writing a table for a whole year in order to calculate the effective interest rates for different loans or investments, we can use the following rule.

Effective interest rate

The effective interest rate of a loan or investment is the interest earned after one year expressed as a percentage of the amount borrowed or invested.

Let:

- r be the nominal interest rate per annum
- $r_{\text{effective}}$ be the effective annual interest rate
- n be the number of times the interest compounds each year.

The effective annual interest rate is given by: $r_{\text{effective}} = \left(\left(1 + \dfrac{r}{100 \times n}\right)^n - 1\right) \times 100\%$

Example 2 Comparing loans and investments with effective interest rates

Brooke would like to borrow \$20 000. She is deciding between two loan options:

- option A: 5.95% per annum compounding weekly
- option B: 6% per annum compounding quarterly.

a Calculate the effective interest rate for each investment.

b Which investment option is the best and why?

Solution

a **1** Decide on the values of r and n for each option.

Option A

$r = 5.95$

There are 52 weeks in a year so $n = 52$.

Option B

$r = 6$

There are 4 quarters in a year so $n = 4$.

2 Apply the effective interest rate rule.

Option A

$$r_{\text{effective}} = \left(\left(1 + \frac{5.95}{100 \times 52}\right)^{52} - 1\right) \times 100\%$$

$= 6.13\%$

Option B

$$r_{\text{effective}} = \left(\left(1 + \frac{6}{100 \times 4}\right)^{4} - 1\right) \times 100\%$$

$= 6.14\%$

b Compare the effective interest rates.

Brooke is borrowing money, so the best option is the one with the lowest effective interest rate. She will pay less interest with option A.

The amount of interest earned over a particular time period depends on the number of compounds within that time period. In a short time period, the number of compounds has little effect on the total interest paid or earned; however, over long time periods the number of compounds can have a significant effect on the total interest earned or paid.

Remember

For any compound interest loan or investment, increasing the number of compounds per year will increase the total interest earned or paid.

Example 3 Calculating effective interest rates using a CAS calculator

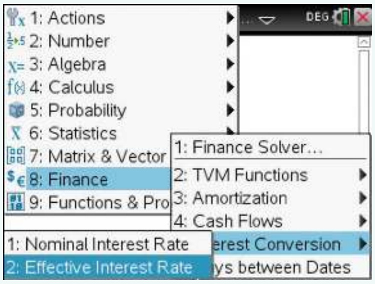
Marissa has $10 000 to invest. She chooses an account that will earn compounding interest at the rate of 4.5% per annum, compounding monthly.

Use a CAS calculator to find the effective rate for this investment, correct to three decimal places.

Solution for TI-Nspire

Steps

1 Press menu and then select
8: Finance ▶
5: Interest Conversion
2: Effective interest rate
to paste in the **eff(…)** command.
The parameters of this function are eff(nominal rate, number of compounds per year).



2 Enter the nominal rate (4.5) and compounds per year (12) into the function, separated by a comma. Press enter to get the effective interest rate.

```
eff(4.5,12)                4.59398
```

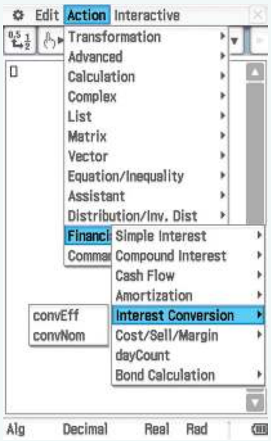
Solution for ClassPad

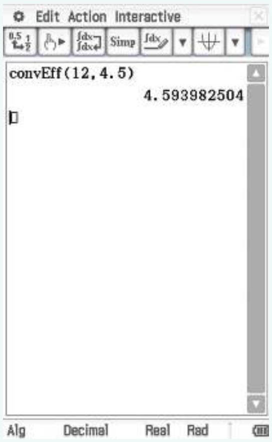
Steps

1 Select **Action**, **Financial**, **Interest Conversion**, **ConvEff** to paste in the **convEff(...)** command.
The parameters of this function are convEff(number of compounds per year, nominal rate).

2 Enter the nominal rate (4.5) and compounds per year (12) into the function, separated by a comma.
Press **EXE** to get the effective interest rate.

3 Write your answer.





```
convEff(12,4.5)
                4.593982504
```

The effective interest rate for this investment is 4.594%.

Exercise 8A

Interest rate conversions and effective interest rates

Example 1

1 Convert each of the annual interest rates below to an interest rate for the given time period. Write your answers, correct to three significant figures.

- **a** 4.95% per annum to monthly
- **b** 8.3% per annum to quarterly
- **c** 6.2% per annum to fortnightly
- **d** 7.4% per annum to weekly
- **e** 12.7% per annum to daily

2 Convert each of the interest rates below to an annual interest rate.

- **a** 0.54% monthly
- **b** 1.45% quarterly
- **c** 0.57% fortnightly
- **d** 0.19% weekly
- **e** 0.022% daily

Example 3

3 Use your calculator to determine the effective annual interest rate, correct to two decimal places, for the following nominal rates and compounding periods.

- **a** 6.2% per annum, compounding monthly
- **b** 8.4% per annum, compounding daily
- **c** 4.8% per annum, compounding weekly
- **d** 12.5% per annum, compounding quarterly
- **e** 7.5% per annum, compounding every 6 months

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Comparing loans and investments with effective interest rates

4 Brenda invests $15 000 in an account earning nominal compound interest of 4.60% per annum, compounding quarterly.

- **a** Explain why Brenda would be better off with *more frequent* compounds per year.
- **b** Calculate the effective interest rate for the current investment with quarterly compounds, correct to two decimal places.
- **c** Calculate the effective rate for this investment with monthly compounds, correct to two decimal places.
- **d** Explain how these effective rates support your answer to part **a**.

5 Stella borrows $25 000 from a bank and pays nominal compound interest of 7.94% per annum, compounding fortnightly.

- **a** Explain why Stella would be better off with *less frequent* compounds per year.
- **b** Calculate the effective rate for the current loan with fortnightly compounds, correct to two decimal places.
- **c** Calculate the effective rate for this loan with monthly compounds, correct to two decimal places.
- **d** Explain how these effective rates support your answer to part **a**.

Example 2

6 Luke is considering a loan of $35 000. His bank has two compound interest rate options:

A: 8.3% per annum, compounding monthly
B: 7.8% per annum, compounding weekly.

- **a** Calculate the effective interest rate for each of the loan options.
- **b** Calculate the amount of interest Luke would pay in the first year for each of the loan options.
- **c** Which loan should Luke choose and why?

7 Sharon is considering investing $140 000. Her bank has two compound interest investment options:

A: 5.3% per annum, compounding monthly
B: 5.5% per annum, compounding quarterly.

- **a** Calculate the effective interest rate for each of the investment options.
- **b** Calculate the amount of interest Sharon would earn in the first year for each of the investment options.
- **c** Which investment option should Sharon choose and why?

ISBN 978-1-009-25917-0

8B Compound interest loans and investments using a finance solver

The recurrence relations that have been used to model reducing-balance loans are very convenient and easy to use, particularly if you need to analyse a reducing-balance loan or investment over a small number of compounding periods.

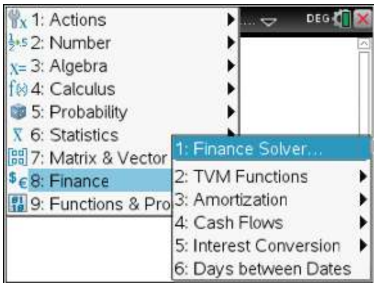
CAS calculators have a **finance solver** feature that makes analysing reducing-balance loans with a large number of repayments much easier than by hand.

How to use the finance solver on the TI-Nspire CAS

Steps

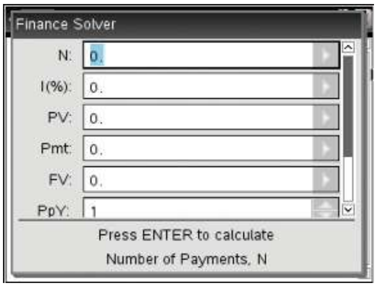
1 Start a new document by pressing ctrl + **N** (or on>**New Document**. You may be prompted to save your current document. (See Appendix A).

2 Select **Add Calculator**.
Press menu>**Finance**> **Finance Solver**.



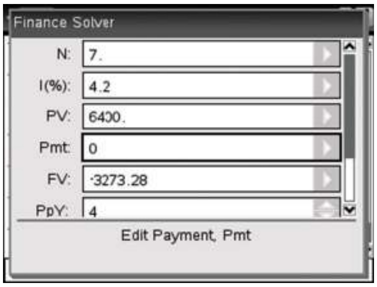
3 To use finance solver you need to know the meaning of each of its symbols. These are as follows:

- **N** is the total number of repayments.
- **I(%)** is the annual interest rate.
- **PV** is the present value of the loan.
- **Pmt** is the amount paid at each repayment. Leave this as zero for now.
- **FV** is the future value of the loan.
- **PpY** is the number of repayments per year.
- **CpY** is the number of times the interest is compounded per year. (It is almost always the same as **PpY**.)
- **PmtAt** is used to indicate whether the interest is compounded at the end or at the beginning of the time period. Leave this set at **END**.


Finance Solver
N: 0.
I(%): 0.
PV: 0.
Pmt: 0.
FV: 0.
PpY: 1
Press ENTER to calculate
Number of Payments, N

Note: Use tab or to ▼ move down boxes. Press ▲ to move up. For **PpY** and **CpY** press tab to move down to the next entry box.

4 When using finance solver to solve loan and investment problems, there will be one unknown quantity. To find its value, move the cursor to its entry field and press enter to solve.
In the example shown pressing enter will solve for **Pmt**.


Finance Solver
N: 7.
I(%): 4.2
PV: 6400.
Pmt: 0
FV: -3273.28
PpY: 4
Edit Payment, Pmt

ISBN 978-1-009-25917-0

Cambridge University Press

Updated January 2024

How to use the finance solver on the ClassPad

Steps

1 Tap **Financial** from the main menu screen.

2 Select the compound interest solver by tapping on **Compound Interest** from the solver screen.

3 To use finance solver you need to know the meaning of each of its symbols. These are as follows:

- **N** is the total number of repayments.
- **I%** is the annual interest rate.
- **PV** is the present value of the loan or investment.
- **PMT** is the amount paid at each payment. Leave this as zero for now.
- **FV** is the future value of the loan or investment.
- **P/Y** is the number of repayments per year.
- **C/Y** is the number of times interest is compounded per year. (It is almost always the same as **P/Y**.)

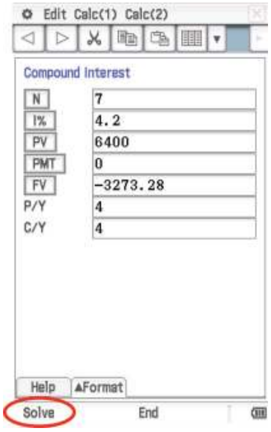
Compound Interest

N
I%
PV
PMT
FV
P/Y 1
C/Y 1

4 Tap **Format** and confirm that the setting for ‘Odd Period’ is set to ‘off’ and ‘Payment Date’ is set to ‘End of period’.

5 When using finance solver to solve loan problems, there will be one unknown quantity. To find its value, tap its entry field and tap **Solve**.

In the example shown tapping **Solve** will solve for **PMT**.



Compound Interest

N 7
I% 4.2
PV 6400
PMT 0
FV −3273.28
P/Y 4
C/Y 4

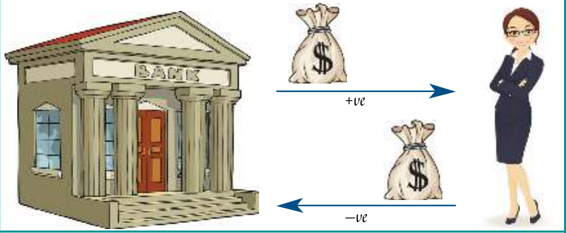
Sign convention for a finance solver

A finance solver is a powerful computation tool. However, you have to be very careful in the way you enter information because it needs to know which way the money is flowing. It does this by following a *sign convention*.

In general terms:

- if you receive money, or someone owes you money, we treat this as a positive (+ve)
- if you pay out money, or you owe someone money, we treat this as a negative (−ve).

The sign convention for a reducing-balance loan is summarised below. You will learn how to apply it in the worked examples that follow.

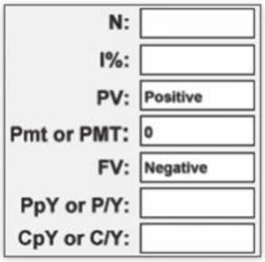
Reducing-balance loan	
Rules: ■ the bank gives you money: positive (+ve) ■ you give bank money: negative (−ve).	
PV: positive	Bank gives (lends) you money.
PMT: negative	You repay the loan by making regular repayments.
FV: negative, zero or positive	After the repayment is made, you still owe the bank money (FV negative), the loan is fully paid out (FV is zero), or you have overpaid your loan and the bank needs to repay you some money (FV positive).

Using a finance solver to analyse compound interest loans

In a compound interest loan, money is given to you by the bank. You must pay this money back, along with the interest that is charged.

Using the sign convention for compound interest loans:

- the principal or present value, PV, is the money borrowed, so it will be positive
- the future value of the loan, FV, is the amount you must pay back to the bank, so it will be negative
- there are no payments, so Pmt or PMT will be zero.



N:
I%:
PV: Positive
Pmt or PMT: 0
FV: Negative
PpY or P/Y:
CpY or C/Y:

The finance solver screen shown has a summary of the sign convention for compound interest loans.

Example 4 Using a finance solver to analyse a compound interest loan

Nelson borrows $12 000 to pay for a holiday. He will repay the loan, and the interest he has been charged, after 6 months.

Calculate the amount he will owe the bank if he is charged compounding interest at the rate of:

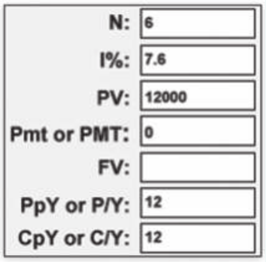
a 7.6% per annum, compounding monthly

b 4.6% per annum, compounding fortnightly.

Solution

a **1** Enter the following values into a finance solver:

- Nelson pays the loan back after 6 months, so N = 6.
- Interest rate is 7.6%, so I% = 7.6
- The amount borrowed is $12 000, so PV = 12 000.
- Pmt or PMT = 0
- The amount owed is FV which we need to find. Leave this blank.
- There are 12 compounds in a year so PpY or P/Y and CpY or C/Y are both 12.



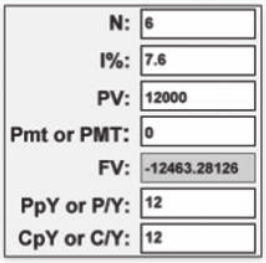
N: 6
I%: 7.6
PV: 12000
Pmt or PMT: 0
FV:
PpY or P/Y: 12
CpY or C/Y: 12

2 Solve for FV.

TI-Nspire: Move the cursor to the FV field and press `enter`.

ClassPad: Tap in the FV field and then tap **Solve**.

The solved value for FV is shown shaded.



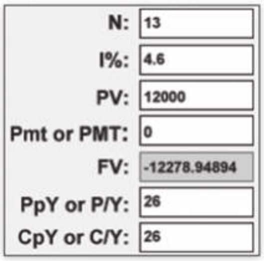
N: 6
I%: 7.6
PV: 12000
Pmt or PMT: 0
FV: -12463.28126
PpY or P/Y: 12
CpY or C/Y: 12

3 Write your answer, rounding to the nearest cent.

Nelson will owe the bank $12 463.28

b **1** Enter the following values into a finance solver and solve for FV:

- 6 months is half a year, or 13 fortnights so N = 13.
- Interest rate is 4.6%, so I% = 4.6
- The amount borrowed is $12 000, so PV = 12 000.
- Pmt or PMT = 0
- The amount owed is FV which we need to find. Solve for this value (shown shaded).
- There are 26 compounds in a year so PpY or P/Y and CpY or C/Y are both 26.



N: 13
I%: 4.6
PV: 12000
Pmt or PMT: 0
FV: -12278.94894
PpY or P/Y: 26
CpY or C/Y: 26

2 Write your answer, rounding to the nearest cent.

Nelson will owe the bank $12 278.95

Using a financial solver to analyse compound interest investments

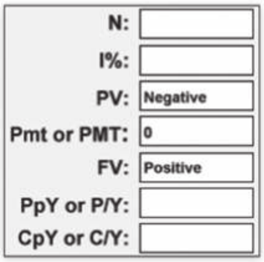
Compound interest investments can be analysed using a finance solver as well, but the sign convention means the principal value will be negative, because this is given to the bank. The future value will be positive because this is money that the bank will need to give back to you.

In a compound interest investment, money is given by you to the bank. The bank must pay you this money back, along with the interest that is earned.

Using the sign convention for compound interest investments:

- the principal or present value, PV, is the money invested, so it will be negative
- the future value of the loan, FV, is the amount the bank must pay back to you, so it will be positive
- there are no payments, so Pmt or PMT will be zero.

The finance solver screen shown has a summary of the sign convention for compound interest investments.



N:
I%:
PV: Negative
Pmt or PMT: 0
FV: Positive
PpY or P/Y:
CpY or C/Y:

Example 5 Using a finance solver to analyse a compound interest investment

Tito invested \$35 000 with a bank and will be paid compound interest at the rate of 5.2% per annum, compounding monthly.

Calculate the amount he will receive after 4 years.

Solution

1 Enter the following values into a finance solver:

- Tito invests for 4 years, or 48 months, so N = 48.
  Note: You can enter this as 4 × 12 and let the calculator work out how many months this is.
- Interest rate is 5.2%, so I% = 5.2
- The amount invested is \$35 000, so PV = −35 000.
- Pmt or PMT = 0
- The amount received at the end of the investment is FV which we need to find.
  Solve for this value (shown shaded).
- There are 12 compounds in a year so PpY or P/Y and CpY or C/Y are both 12.

N: 4x12
I%: 5.2
PV: -35000
Pmt or PMT: 0
FV: 43073.10087
PpY or P/Y: 12
CpY or C/Y: 12

2 Write your answer, rounding to the nearest cent.

Tito will receive \$43 073.10 after 4 years.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Solving financial problems

A finance solver can solve for any one of the seven different quantities that are entered, provided the other six are known. You just need to move the cursor to the field for the unknown quantity and press `enter` (Nspire), or tap in the field for the unknown quantity and then tap **Solve** (ClassPad).

Example 6 Solving financial problems

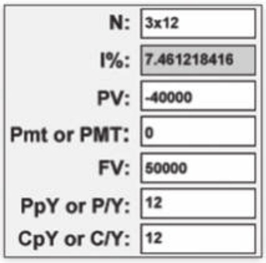
Antonella has \$40 000 to invest. She would like to have at least \$50 000 in the investment account after three years.

a What is the minimum annual interest rate, compounding monthly, that Antonella will need for her investment? Write your answer correct to one decimal place.

b If Antonella has an account paying compound interest at the rate of 7.8% per annum, compounding monthly, how many months will it take for the investment to grow to \$60 000? Round your answer up to the nearest month.

Solution

a **1** Enter the following values into a finance solver:



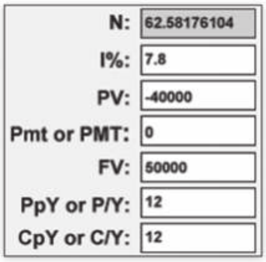
N: 3x12
I%: 7.461218416
PV: -40000
Pmt or PMT: 0
FV: 50000
PpY or P/Y: 12
CpY or C/Y: 12

2
- Antonella needs the investment to be \$50 000 in value after three years, or 36 months. N = 36.
  Note: You can enter this as 3 × 12 and let the calculator work out how many months this is.
- Interest rate is unknown. Solve for this quantity (shown shaded).
- The amount invested is \$40 000, so PV = −40 000.
- Pmt or PMT = 0
- Antonella needs a future value of \$50 000 at least. FV = 50 000
- There are 12 compounds in a year so PpY or P/Y and CpY or C/Y are both 12.

3 Write your answer, rounding to one decimal place.

Antonella will need an interest rate of at least 7.5% per annum.

b **1** Enter the following values into a finance solver:



N: 62.58176104
I%: 7.8
PV: -40000
Pmt or PMT: 0
FV: 50000
PpY or P/Y: 12
CpY or C/Y: 12

- Number of months is unknown. Solve for this quantity (shown shaded).
- Interest rate 7.8% per annum so I% = 7.8
- The amount invested is \$40 000, so PV = −40 000.

- Pmt or PMT = 0
- Antonella needs a future value of $60 000 so FV = 60 000
- There are 12 compounds in a year so PpY or P/Y and CpY or C/Y are both 12.

2 Write your answer, rounding up to the nearest month.

It will take 63 months for Antonella's investment to grow to $60 000.

Skill-sheet

Exercise 8B

Note: The answers to questions in this exercise have been verified on TI-Nspire and Casio Classpad CAS calculators. Other finance solvers may use rounding differently and so produce slightly different answers.

Using a finance solver

CAS

Example 4

1 Use a finance solver to determine the value of the following compound interest loans (PV is positive, FV is negative) after 2 years. Round your answer to the nearest cent.

a $20 000 borrowed at 3.5% per annum, compounding annually

b $8000 borrowed at 8.9% per annum, compounding quarterly

c $5800 borrowed at 5.7% per annum, compounding monthly

d $26 000 borrowed at 6.4% per annum, compounding fortnightly

e $4500 borrowed at 4.1% per annum, compounding weekly

2 Use a finance solver to determine the value of the following compound interest investments (PV is negative, FV is positive) after 5 years. Round your answer to the nearest cent.

a $25 000 invested at 5.6% per annum, compounding annually

b $10 000 invested at 4.3% per annum, compounding quarterly

c $8500 invested at 3.6% per annum, compounding monthly

d $2000 invested at 2.8% per annum, compounding fortnightly

e $8700 invested at 4.9% per annum, compounding weekly

Solving financial problems

Example 6

3 How many months will an investment of $25 000 take to double in value if it is invested at an interest rate of 5.9% per annum, compounding monthly? Round your answer up to the nearest month.

ISBN 978-1-009-25917-0

4 A loan with principal \$6000 and interest rate 5.8% per annum, compounding weekly, is fully repaid when its value is \$8500. After how many weeks will this occur? Round your answer to the nearest week.

5 An investment of \$4000 has a value of \$4500 after 2 years. At what interest rate, compounding monthly, has this money been invested? Round your answer to one decimal place.

6 Michael borrowed \$10 000 and after 2 years, he paid back this money and the interest charged. If Michael paid back \$10 961.72, what is the interest rate, compounding monthly, for his loan? Round your answer to one decimal place.

7 After 3 years, Cynthia would like to have \$50 000 saved. If she can invest money at an interest rate of 3.6% per annum, compounding fortnightly, how much money should she invest to reach her savings goal? Round your answer to the nearest cent.

8C Modelling compound interest loans with periodic repayments

Reducing-balance loans

When money is borrowed from a bank, it is very unusual for the borrower to wait until the term of the loan is complete before paying all of the money owed, including the interest, back to the bank. Generally loans are repaid by making regular repayments to reduce the amount owed, the **balance** of the loan, to zero over time. The amount originally borrowed is called the principal of the loan.

This kind of loan is called a **reducing-balance loan**. A reducing-balance loan is effectively a compound interest loan, with regular repayments. Personal loans and mortgages (home loans) are examples of reducing-balance loans.

Recurrence relation model for a reducing-balance loan

Let V_n be the *balance* of the loan after n repayments have been made.

Let r be the *interest rate* per compounding period.

Let D be the *repayment* made.

A recurrence relation that can be used to model a reducing balance loan is

$$V_0 = \text{the principal}, \qquad V_{n+1} = RV_n - D$$

where $R = 1 + \frac{r}{100}$

Example 7 Modelling a reducing-balance loan with a recurrence relation

Alyssa borrows \$1000 at an interest rate of 15% per annum, compounding monthly.

She will repay the loan by making four monthly repayments of \$257.85.

Construct a recurrence relation to model this loan, in the form

$$V_0 = \text{the principal}, \qquad V_{n+1} = RV_n - D$$

where V_n is the balance of the loan after n repayments.

Solution

1 Define the symbol V_n.

Let V_n be the balance of the loan after n repayments.

2 Write the value of V_0. Here V_0 is the principal of the loan, the balance of the loan before any repayments have been made.

$V_0 = 1000$

3 Because interest compounds monthly, calculate the monthly interest rate.

$r = \frac{15\%}{12} = 1.25\%$

4 Determine the value of R. Either recognise that a 1.25% increase gives an R of 1.0125, or use the formula $R = 1 + r/100 = 1 + 1.25/100 = 1.0125$

$R = 1.0125$

5 Write the value of D. In this context, D is the amount repaid each month.

$D = 257.85$

6 Use the values of V_0, R and D to write the recurrence relation.

$V_0 = 1000$

$V_{n+1} = 1.0125V_n - 257.85$

Once we have a recurrence relation, we can use it to determine things such as the balance of a loan after a given number of repayments.

Example 8 Using a recurrence relation to analyse a reducing-balance loan

Alyssa's loan can be modelled by the recurrence relation:

$$V_0 = 1000, \qquad V_{n+1} = 1.0125V_n - 257.85$$

a Use your calculator to determine recursively the balance of the loan after Alyssa has made each of the four repayments.

b What is the balance of the loan (the amount she still owes) after she has made two repayments? Give your answer to the nearest cent.

c Is the loan fully paid out after four repayments have been made? If not, how much will the last repayment have to be to ensure that the loan is fully repaid after four payments?

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

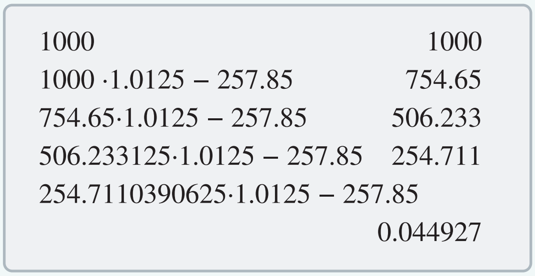
Solution

a **i** Write the recurrence relation.

$V_0 = 1000, V_{n+1} = 1.0125V_n - 257.85$

ii Type '1000' and press 'enter' or 'EXE'.

iii Type '× 1.0125-257.85' and press 'enter' (or EXE) 4 times to obtain the screen opposite.



```
1000                                  1000
1000 ·1.0125 – 257.85               754.65
754.65·1.0125 – 257.85              506.233
506.233125·1.0125 – 257.85          254.711
254.7110390625·1.0125 – 257.85
                                   0.044927
```

b The balance of the loan after two repayments corresponds to the third line on the screen.

Round the answer to the nearest cent.

$506.23 (to the nearest cent)

c No. Alyssa still owes 4 cents. Add this amount to the regular monthly repayment to find the value of the last repayment that ensures that the loan is fully repaid after 4 repayments.

Last payment = 257.85 + 0.04

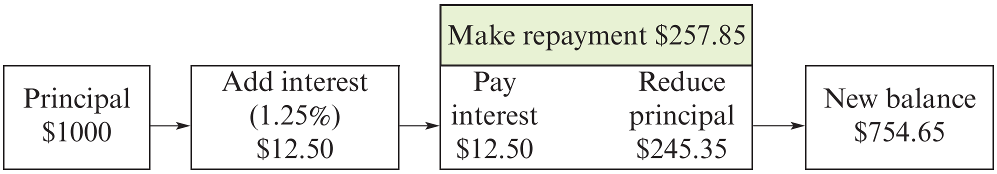
= $257.89

Note: In Example 7, the balance of Alyssa's loan is reduced to around 4 cents. The loan is almost fully repaid after four repayments, but not exactly. To repay the loan in exactly four repayments, Alyssa would have to repay $258.861 023 . . . each month. In practice this amount must be rounded to the nearest cent, so it is not possible to repay the loan exactly by making four equal repayments.

Amortisation tables

Loans that are repaid by making regular payments until the balance of the loan is zero are called **amortising** loans. In an amortising loan, part of each payment goes towards paying the interest owed on the unpaid balance of the loan with the remainder used to reduce the principal of the loan (the amount borrowed).

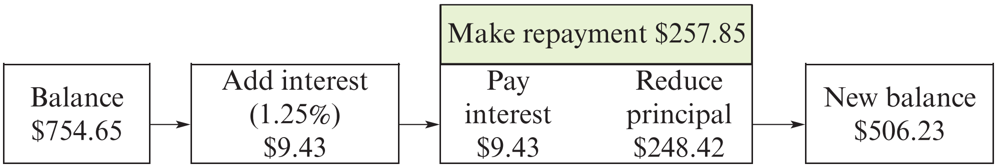
For example, consider Alyssa's loan from Example 7. Interest on the $1000 loan was charged at the rate of 1.25% per month and the loan was to be repaid with four monthly payments of $257.85. The first step in the amortisation process is shown here.



At the end of the first month, when the first repayment is due:

- 1 month's interest is charged on the full $1000; that is, 1.25% of $1000 = $12.50
- the interest owed is then deducted from the repayment, $257.85 – $12.50 = $245.35
- the remaining $245.35 is then used to reduce the balance of the loan to give a new balance of $754.65 ($1000 – $245.35).

The process is then repeated when the next repayment is made and continued until four repayments have been made.



The results of this analysis can then be summarised in an **amortisation table**. The amortisation table shows the repayment number, the repayment made, interest owing and the balance of the loan both before and after each repayment. The amortisation table for Alyssa's loan is shown below.

Month, n	*Balance of loan at start of month n ($)*	*Interest owing ($)*	*Repayment ($)*	*Balance owing at end of month ($)*
1	1000	12.50	257.85	754.65
2	754.65	9.43	257.85	506.23
3	506.23	6.33	257.85	254.71
4	254.71	3.18	257.85	0.04
5	0.04	0.00	0.04	0.00
	Total	31.44		

The value of the final repayment will always be less than the regular payment amount. Interest must be added to this value before the loan can be fully paid off. In this case, because the amount owing after repayment 4 is only 4 cents, the interest calculated on this amount is negligible, and so does not affect the final repayment amount.

Month, n	*Balance of loan at start of month n ($)*	*Interest owing ($)*	*Repayment ($)*	*Balance owing at end of month ($)*
1	1000	12.50	257.85	754.65
2	754.65	9.43	257.85	506.23
3	506.23	6.33	257.85	254.71
4	254.71	3.18	257.89*	0.00
	Total	31.44		

*The last repayment can be increased by 4 cents so that the balance of the loan is reduced to exactly $0.00 after 4 repayments had been made. Therefore, at month 4 the final repayment will be $257.89.

The amortisation table above can be used to summarise the key properties of a reducing balance loan.

ISBN 978-1-009-25917-0

Cambridge University Press

Properties of a reducing-balance loan

At each step of the loan:

1 interest owing = interest rate per payment period × balance at start of month

For example, before repayment 1 is made, interest owing

= 1.25% of 1000 = $12.50

2 balance owing before repayment = balance at start of month + interest owing

For example, before repayment 1 is made, principal owing

= 1000 + 12.50 = $1012.50

3 balance owing at end of month = principal owing before repayment – repayment

For example, when repayment 1 is made, balance owing at end of month

= 1012.50 – 257.85 = $754.65

4 total cost of repaying the loan = the sum of the repayments

For this loan, the total cost of repaying the loan

= 3 × 257.85 + 257.89 = $1031.44

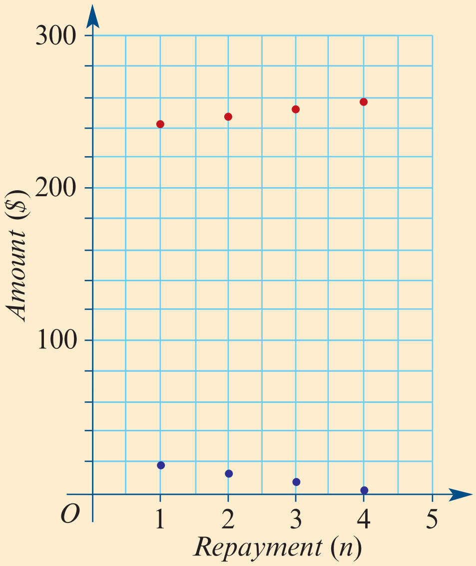
5 total interest paid = total cost of repaying the loan – principal

For this loan, the total interest paid

= 1031.44–1000 = $31.44

6 The amount of interest paid from each repayment decreases with repayment number while the amount of principal repaid increases.

For this loan, the graph on the right shows how the amount of interest paid each repayment (blue dots) decreases with repayment number, while the amount of principal paid increases (red dots).



Example 9 Reading and interpreting an amortisation table

A business borrows $10 000 at a rate of 8% per annum. The loan is to be repaid by making four quarterly payments of $2626.20. The amortisation table for this loan is shown over page.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Repayment number, n	*Balance at start of quarter*	*Interest owing*	*Repayment*	*Balance carried forward*
1	10 000.00	*A*	2626.20	7573.80
2	*B*	151.48	2626.20	5099.08
3	5099.08	101.98	2626.20	*C*
4		51.50	2626.20	
5		0.00	*D*	
	Total			

a Determine, to the nearest cent:

- **i** the quarterly interest rate
- **ii** the interest paid when repayment 1 is made using the quarterly interest rate, *A*
- **iii** the balance owing before repayment 2, *B*
- **iv** the balance of the loan after repayment 3, *C*
- **v** the value of the final repayment, *D*
- **vi** the total cost of repaying the loan

b Verify that the total amount of interest paid = total cost of the loan – principal

Solution

a i Quarterly rate = annual rate ÷ 4

Quarterly interest rate = 8% ÷ 4 = 2%

ii Interest paid = quarterly interest rate × balance at start of quarter

Interest paid = 2% × 10 000 = \$200

A = \$200

iii Balancing owing before repayment = balance carried forward after previous repayment

Balancing carried forward = \$7573.80

B = \$7573.80

iv Balance of loan after repayment = balance at start of quarter + interest owing − repayment

Balancing owing

= 5099.08 + 101.98 − 2626.20

= \$2574.86

C = \$2574.86

v Final repayment = balance at start of final quarter × quarterly interest rate

Balance owing from previous quarter

= 2574.86 + 51.50 − 2626.20 = \$0.16

D = \$0.16

vi Total cost of repaying the loan = sum of repayments made

Total cost = 4 × 2626.20 + 0.16

= \$10 504.96

b Use the rule: Total interest paid = total cost of the loan − principal

Total interest = 10 504.96 − 10 000

= \$504.96

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Exercise 8C

Reading and interpreting an amortisation table

Example 9

1 A student borrows $2000 at an interest rate of 12% per annum to pay off a debt. The loan is to be repaid by making six monthly repayments of $345.
The amortisation table for this loan is shown below.

Month, n	*Balance of loan at start of month n ($)*	*Interest owing ($)*	*Repayment ($)*	*Balance owing at end of month ($)*
1	2000.00	20.00	345.00	1675.00
2	1675.00	*A*	345.00	1346.75
3	1346.75	13.47	345.00	1015.22
4	1015.22	10.15	345.00	680.37
5	680.37	6.80	345.00	*B*
6		3.42	345.00	*C*

Determine, to the nearest cent:

a the monthly interest rate

b the interest paid when repayment 2 is made using the monthly interest rate, *A*

c the balance owing on the loan before repayment 4

d the balance of the loan after repayment 5 has been made, *B*

e the value of the final repayment, *C*

f the total cost of repaying the loan

g the total amount of interest paid

2 The amortisation table for a loan is shown below. The loan is to be repaid over 2 years by making eight quarterly repayments.

Repayment number, n	*Balance of loan at start of quarter n ($)*	*Interest owing ($)*	*Repayment ($)*	*Balance carried forward ($)*
1	4000.00	100.00	557.85	3542.15
2	3542.15	88.55	557.85	3072.85
3	3072.85	76.82	557.85	2591.83
4	2591.83	*A*	557.85	2098.77
5	2098.77	52.47	557.85	1593.39
6	1593.39	39.83	557.85	*B*
7		*C*	557.85	544.41
8	544.41	13.61	557.85	0.17

ISBN 978-1-009-25917-0

a Reading directly from the table:

  i what is the principal of this loan?

  ii what is the quarterly payment?

  iii how much interest was paid for repayment 1?

  iv how much of the principal is still owing after repayment 2?

  v what is the balance of the loan after repayment 3 is made?

  vi what must the final repayment be for the loan to be fully repaid after 8 quarters?

b What is:

  i the quarterly interest rate?

  ii the annual interest rate?

c Determine the values of A, B and C.

Using a recurrence relation to analyse a reducing-balance loan

Example 8 **3** A reducing-balance loan can be modelled by the recurrence relation

$$V_0 = 2500, \qquad V_{n+1} = 1.08V_n - 626.00$$

where V_n is the balance of the loan after n repayments have been made.
The loan is to be fully repaid after five repayments.

CAS

a Use your calculator to determine recursively the balance of the loan after three repayments have been made. Give your answer to the nearest cent.

b Is the loan fully paid out after five repayments of $626 have been made? If not, what is the value of the final repayment?

4 A reducing-balance loan can be modelled by the recurrence relation:

$$V_0 = 5000, \qquad V_{n+1} = 1.01V_n - 862.70$$

a Use your calculator to determine recursively the balance of the loan after two repayments have been made. Give your answer to the nearest cent.

b How many repayments will it take for the loan to be fully repaid and what is the value of the final repayment?

Modelling reducing-balance loans with recurrence relations

Example 7 **5** $2000 is borrowed at an interest rate of 6% per annum compounding monthly.
The loan will be repaid by making monthly repayments of $339.
Let V_n be the balance of the loan after n repayments have been made.

a Model this loan using a recurrence relation of the form:

$$V_0 = \text{the principal}, \qquad V_{n+1} = RV_n - D$$

b Use the recurrence relation to determine the balance of the loan after 4 months. Give your answer to the nearest cent.

c How long will it take to fully repay the loan and what is the value of the final repayment?

ISBN 978-1-009-25917-0

6 \$10 000 is borrowed at an interest rate of 12% per annum compounding quarterly. The loan will be repaid by making quarterly repayments of \$2690.27.
Let B_n be the balance of the loan after n repayments have been made.

- **a** Model this loan using a recurrence relation of the form:

$$B_0 = \text{the principal}, \qquad B_{n+1} = RB_n - D$$

- **b** Use the recurrence relation to determine the balance of the loan after two repayments have been made.
- **c** To the nearest cent, does a quarterly payment of \$2690.27 ensure that the loan is fully repaid after four repayments?

7 Let V_n be the balance of a loan after n repayments have been made. Write a recurrence relation model for the balance of a loan of:

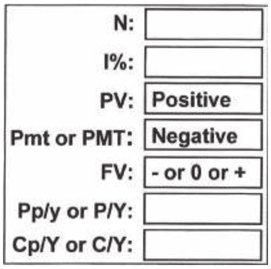
- **a** \$3500 borrowed at 4.8% per annum, compounding monthly, with repayments of \$280 per month
- **b** \$150 000 borrowed at 3.64% per annum, compounding fortnightly, with repayments of \$650 per fortnight.

8D Using a finance solver to analyse reducing-balance loans

One of the most common types of reducing-balance loans are home loans. A large amount of money is borrowed to buy a property and then repaid with many repayments over a number of years. A typical home loan would involve fortnightly or monthly repayments for a period of 25 to 30 years.

Reducing-balance loans involve a payment that is made after each compounding time period. Using a finance solver to analyse reducing-balance loans means we can now enter values in the Pmt or PMT field. This is the regular repayment amount.

Using the sign convention for a finance solver:



N:
I%:
PV: Positive
Pmt or PMT: Negative
FV: - or 0 or +
Pp/y or P/Y:
Cp/Y or C/Y:

- the principal or present value, PV, is the money given to you by the bank, so it will be positive
- repayments are paid to the bank, so Pmt or PMT will be negative
- the future value of the loan, FV, can be negative, zero or positive:
  - negative indicates an amount that is still owing
  - zero indicates that the loan has been fully repaid
  - positive indicates that more than required has been paid and the bank must refund this amount.

The finance solver screen shown has a summary of the sign convention for reducing-balance loans.

Example 10 Determining the balance of a reducing-balance loan after a given number of payments

Andrew borrows \$20 000 at an interest rate of 7.25% per annum, compounding monthly. This loan will be repaid with repayments of \$481.25 each month.

a How much, correct to the nearest cent, does Andrew owe after 3 years?

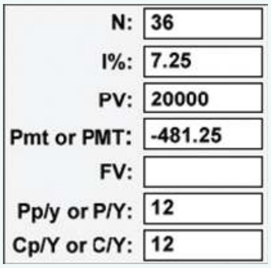
b How long will it take for Andrew to fully repay the loan?

c What is the final repayment amount Andrew must make to fully repay the loan?

Solution

a **1** Open the finance solver on your calculator and enter the information shown.

- **N**: 36 (number of monthly repayments in 3 years)
- **I%**: 7.25 (annual interest rate)
- **PV**: 20 000 (positive to indicate that this is money received by Andrew from the lender)
- **Pmt** or **PMT**: −481.25 (negative to indicate that this is money that Andrew is giving back to the lender)
- **Pp/Y**: 12 (monthly repayments)
- **Cp/Y**: 12 (interest compounds monthly)



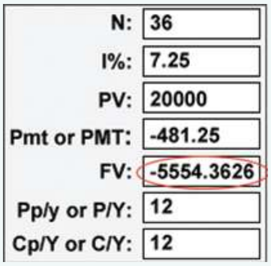
N: 36
I%: 7.25
PV: 20000
Pmt or PMT: -481.25
FV:
Pp/y or P/Y: 12
Cp/Y or C/Y: 12

2 Solve for the unknown future value (FV). On the:

- *TI-Nspire:* Move the cursor to the **FV** entry box and press to enter solve.
- *ClassPad:* Tap on the **FV** entry box and tap **Solve**.

  The amount −5554.3626 . . . now appears in the **FV** entry box.

Note: A negative FV indicates that Andrew will still owe the lender money after the repayment has been made.



N: 36
I%: 7.25
PV: 20000
Pmt or PMT: -481.25
FV: -5554.3626
Pp/y or P/Y: 12
Cp/Y or C/Y: 12

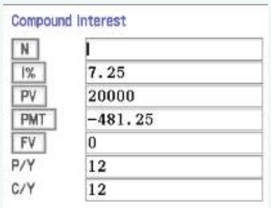
3 Write your answer, correct to the nearest cent.

Andrew owes \$5554.36.

b **1** Find the value of N when the future value (FV) is zero (0).

Enter the information below.

- **I%:** 7.25 (annual interest rate)
- **PV:** 20 000
- **Pmt or PMT:** -481.25
- **FV:** 0 (no money owing on the loan)
- **P/Y:** 12
- **C/Y:** 12



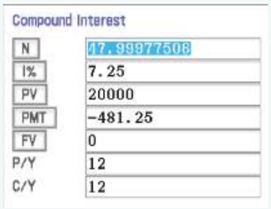
Compound Interest

N	
I%	7.25
PV	20000
PMT	−481.25
FV	0
P/Y	12
C/Y	12

2 *ClassPad:* Tap on the **N** entry box and tap **Solve**. The value 47.9997… now appears in the **N** entry box.

Round the value of N up to the next whole number.

Note: The value for N indicates the number of repayments required for the loan to be fully repaid.



Compound Interest

N	47.99977506
I%	7.25
PV	20000
PMT	−481.25
FV	0
P/Y	12
C/Y	12

Convert answer into months/years as appropriate.

Andrew requires 48 repayments to fully pay off the loan.

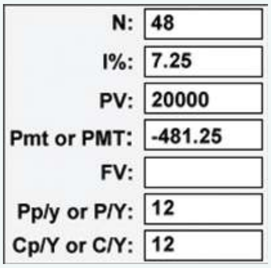
$48 \div 12 = 4$

3 Write your answer

Andrew requires 48 repayments (4 years) to fully repay the loan.

c **1** Find the amount Andrew owes after 48 repayments. Enter the information below, as shown opposite.

- **N**: 48 (number of monthly repayments in 4 years)
- **I%**: 7.25 (annual interest rate)
- **PV**: 20 000
- **Pmt** or **PMT** (the repayment amount is negative): −481.25
- **Pp/Y**: 12 (monthly repayments)
- **Cp/Y**: 12 (interest compounds monthly)



N:	48
I%:	7.25
PV:	20000
Pmt or PMT:	-481.25
FV:	
Pp/y or P/Y:	12
Cp/Y or C/Y:	12

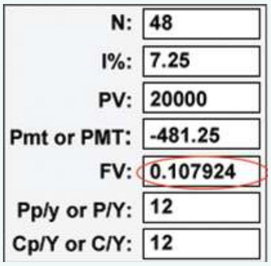
2 Solve for the unknown future value (FV). On the:

- *TI-Nspire:* Move the cursor to the **FV** entry box and press enter to solve.
- *ClassPad:* Tap on the **FV** entry box and tap **Solve**.

  The amount 0.1079 . . . (11 cents) now appears in the **FV** entry box.

The FV is positive (+11 cents). This means that the bank owes Andrew 11 cents. Andrew has repaid the bank slightly more than required to repay the loan exactly. In this case the regular repayment was rounded up to the nearest cent, meaning that each monthly repayment was slightly more than required to repay the loan with exactly 48 repayments.

To compensate, Andrew's final repayment will be reduced by 11 cents.



N: 48
I%: 7.25
PV: 20000
Pmt or PMT: -481.25
FV: 0.107924
Pp/y or P/Y: 12
Cp/Y or C/Y: 12

Final repayment
$= \$481.25 - \$0.11$
$= \$481.14$

3 Write your answer.

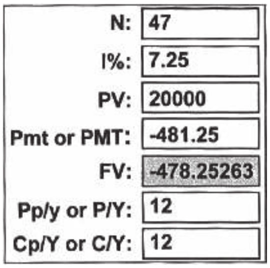
Andrew's final repayment will be $481.14.

The example above demonstrated that Andrew overpaid his loan by making 48 repayments of $481.25 and so the bank had to refund him 11 cents.

Another way to calculate this is to determine how much Andrew owes after 47 repayments, that is the balance of the loan with just one repayment left to make.

The finance solver screen on the right shows this calculation.

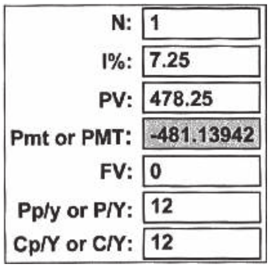
The $478.25 that Andrew owes (FV) will be repaid, with interest, after one further repayment.



N: 47
I%: 7.25
PV: 20000
Pmt or PMT: -481.25
FV: -478.25263
Pp/y or P/Y: 12
Cp/Y or C/Y: 12

We can calculate the final repayment that is required by using this $478.25 as the balance of the loan (PV) and calculating the repayment required to fully repay the loan (FV = 0) after one repayment.

The finance solver screen on the right shows this calculation and verifies the final repayment amount of $481.14 as calculated in Example 10 above.



N: 1
I%: 7.25
PV: 478.25
Pmt or PMT: -481.13942
FV: 0
Pp/y or P/Y: 12
Cp/Y or C/Y: 12

Example 11 Determining the repayment amount, total cost and total amount of interest paid

Sipho borrows $10 000 to be repaid with monthly repayments over a period of 5 years. Interest is charged at the rate of 8% per annum, compounding monthly.

Find:

a the monthly repayment amount, correct to the nearest cent

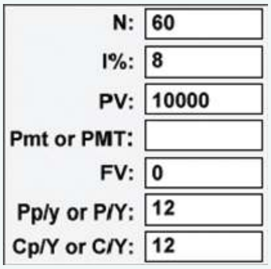
b the total cost of repaying the loan to the nearest dollar

c the total amount of interest paid to the nearest dollar.

Solution

a **1** Open the finance solver on your calculator and enter the information shown.

- **N**: 60 (number of monthly repayments in 5 years)
- **I%**: 8 (annual interest rate)
- **PV**: 10 000
- **FV**: 0 (the balance will be zero when the loan is repaid)
- **Pp/Y**: 12 (monthly repayments)
- **Cp/Y**: 12 (interest compounds monthly)



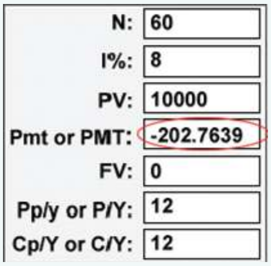
N: 60
I%: 8
PV: 10000
Pmt or PMT:
FV: 0
Pp/y or P/Y: 12
Cp/Y or C/Y: 12

2 Solve for the unknown future value (Pmt or PMT). On the:

- *TI-Nspire:* Move the cursor to the **Pmt** entry box and press enter to solve.
- *ClassPad:* Tap on the **PMT** entry box and tap **Solve**.

The amount −202.7639 . . . now appears in the **Pmt** or **PMT** entry box.

Note: The sign of the repayment is negative to indicate that this is money Sipho is giving back to lender.



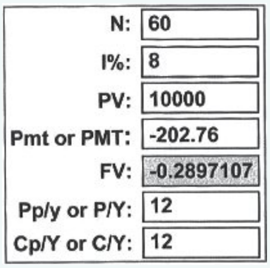
N: 60
I%: 8
PV: 10000
Pmt or PMT: -202.7639
FV: 0
Pp/y or P/Y: 12
Cp/Y or C/Y: 12

3 Write your answer.

Sipho repays $202.76 every month.

ISBN 978-1-009-25917-0

b **1** Use the rounded repayment amount to calculate the balance of the loan after 60 repayments.
A negative future value of −0.2897 . . . indicates that Sipho still owes \$0.29.
This adjustment amount must be added to the final repayment in order to fully repay the loan.



N: 60
I%: 8
PV: 10000
Pmt or PMT: -202.76
FV: -0.2897107
Pp/y or P/Y: 12
Cp/Y or C/Y: 12

2 Total cost of repaying the loan = number of repayments × rounded repayment amount + adjustment amount

Total cost = 60 × \$202.76 + \$0.29
= \$12 165.89

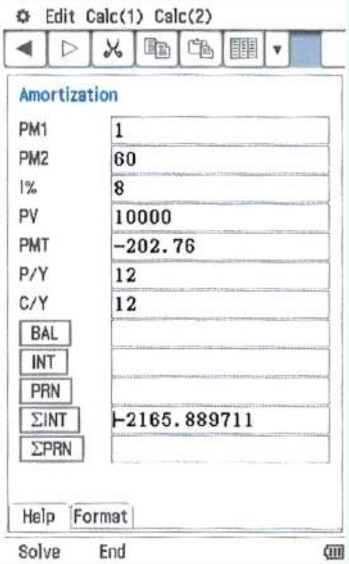
3 Round to the nearest dollar.

Total cost = \$12 166

c Total interest paid = total cost of repaying the loan − the principal

Interest paid = total cost − principal
= 12 166 − 10 000
= \$2166 (to the nearest dollar)

P/Y 12
C/Y 12



PM1 1
PM2 60
I% 8
PV 10000
PMT −202.76
P/Y 12
C/Y 12
ΣINT −2165.889711

Note: Classpad users can calculate the total interest paid for a loan using the finance solver.
From the compound interest screen, tap Calc(1) and then tap Amortization.
PM1 is the first repayment number (1) and PM2 is the last repayment (60).
Solve for Σ INT to calculate the total interest paid.

Example 12 Reducing-balance loans with changing conditions

An amount of \$150 000 is borrowed for 25 years at an interest rate of 6.8% per annum, compounding monthly. Round all answers to the nearest cent.

a What are the monthly repayments for this loan?

b Use the rounded monthly repayment to determine the amount still owing at the end of 3 years.

After 3 years, the interest rate rises to 7.2% per annum.

c Use the rounded answer from **b** to determine the new monthly repayments that will see the loan paid in 25 years.

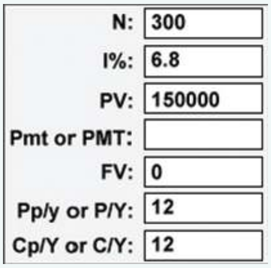
d What is the value of the final repayment on this loan?

Solution

a **1** Open the finance solver on your calculator and enter the information shown.

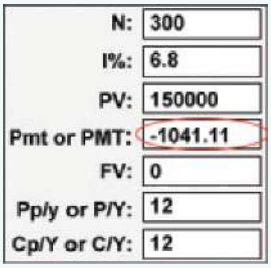
- **N**: 300 (number of monthly repayments in 25 years)
- **I%**: 6.8 (annual interest rate)
- **PV**: 150 000
- **FV**: 0 (the balance will be zero when the loan is repaid)
- **Pp/Y**: 12 (monthly repayments)
- **Cp/Y**: 12 (interest compounds monthly)

Note: You can enter **N** as 25 × 12 (25 years of monthly repayments). The finance solver will calculate this as 300 for you.



N: 300
I%: 6.8
PV: 150000
Pmt or PMT:
FV: 0
Pp/y or P/Y: 12
Cp/Y or C/Y: 12

2 Solve for **Pmt** or **PMT**.

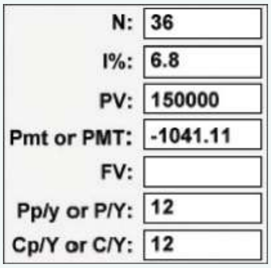


N: 300
I%: 6.8
PV: 150000
Pmt or PMT: -1041.11
FV: 0
Pp/y or P/Y: 12
Cp/Y or C/Y: 12

3 Write your answer.

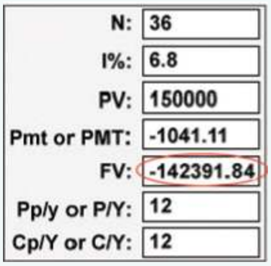
The repayments on this loan are \$1041.11 each month.

b **1** Change **N** to 3 × 12 or 36 (3 years of monthly repayments).



N: 36
I%: 6.8
PV: 150000
Pmt or PMT: -1041.11
FV:
Pp/y or P/Y: 12
Cp/Y or C/Y: 12

2 Clear **FV** and solve.



N: 36
I%: 6.8
PV: 150000
Pmt or PMT: -1041.11
FV: -142391.84
Pp/y or P/Y: 12
Cp/Y or C/Y: 12

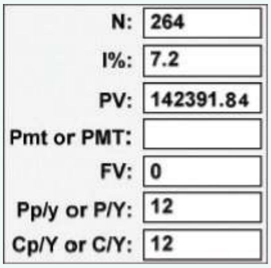
3 Write your answer.

After 3 years, there is $142 391.84 still owing on the loan.

c If the loan is still to be repaid in 25 years, there are still 22 years left.

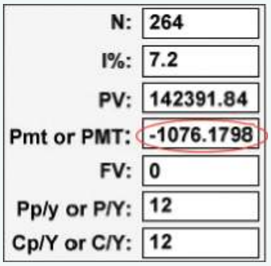
1 Change:

- **N** to 22 × 12 or 264 repayments
- **I(%)** to 7.2 (the new interest rate)
- **PV** to 142 391.84 (the balance after 3 years)
- **FV** to 0 (to pay out the loan).



N: 264
I%: 7.2
PV: 142391.84
Pmt or PMT:
FV: 0
Pp/y or P/Y: 12
Cp/Y or C/Y: 12

2 Clear **Pmt** or **PMT** and solve.

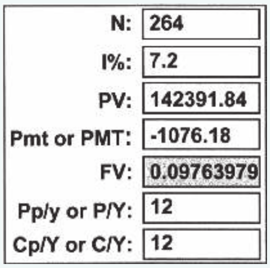


N: 264
I%: 7.2
PV: 142391.84
Pmt or PMT: -1076.1798
FV: 0
Pp/y or P/Y: 12
Cp/Y or C/Y: 12

3 Write your answer.

The new repayments will be $1076.18 each month.

d **1** Use the rounded repayment amount to calculate the balance of the loan after 264 repayments. A positive future value of 0.0976... indicates that the repayments have resulted in too much being paid and a refund of $0.10 must be made (rounded to nearest cent). This adjustment amount must be subtracted from the final repayment in order to fully repay the loan.


N: 264
I%: 7.2
PV: 142391.84
Pmt or PMT: -1076.18
FV: 0.09763979
Pp/y or P/Y: 12
Cp/Y or C/Y: 12

2 Final repayment = usual repayment – adjustment

Final repayment
= $1076.18 – $0.10
= $1076.08

Exercise 8D

Note: The answers to questions in this exercise have been verified on TI-Nspire and Casio ClassPad CAS calculators. Other finance solvers may use rounding differently and so produce slightly different answers.

Determining the balance of a reducing-balance loan using a finance solver

CAS

1 Use a finance solver to find the balance, correct to the nearest cent, of each of the following reducing-balance loans after the given number of compounding periods.

	Principal	*Annual interest rate*	*Compounding*	*Repayment per period*	*Balance after ...*
a	$8000	4.5%	Monthly	$350	6 months
b	$25 000	7.8%	Monthly	$1200	1 year
c	$240 000	8.3%	Quarterly	$7900	5 years
d	$75 000	6.9%	Quarterly	$4800	2 years
e	$50 000	4.6%	Weekly	$350	1 year

Determining the repayment amount of a reducing-balance loan using a finance solver

2 Use a finance solver to find the repayment amount, correct to the nearest cent, for each of the following reducing-balance loans.

	Principal	*Annual interest rate*	*Compounding*	*Time to fully repay the loan*
a	$17 000	6.8%	Monthly	30 months
b	$9500	4.2%	Monthly	2 years
c	$2800	9.6%	Quarterly	6 quarters
d	$140 000	8.6%	Quarterly	15 years
e	$250 000	5.2%	Fortnightly	25 years

ISBN 978-1-009-25917-0

Determining the number of repayments and the value of the final repayment of a reducing-balance loan using a finance solver

3 Use a finance solver to find:

- **i** the number of monthly repayments required to repay the loan (correct to the nearest repayment)
- **ii** the value of the final repayment (correct to the nearest cent)

for each of the following reducing-balance loans.

	Principal	*Annual interest rate*	*Monthly repayment*
a	\$15 000	6.2%	790.00
b	\$8500	3.8%	135.00
c	\$2900	7.5%	70.00
d	\$135 000	4.2%	1400.00
e	\$260 000	3.9%	1550.00

Determining the repayment amount, total cost and total interest paid using a finance solver

Example 11

4 A loan of \$90 000 is to be repaid over a period of 30 years. Interest is charged at the rate of 11% per annum, compounding monthly.
Find:

- **a** the monthly repayment, correct to the nearest cent
- **b** the value of the final repayment
- **c** the total cost of paying off the loan, correct to the nearest cent
- **d** the total amount of interest paid.

5 A building society offers \$240 000 home loans at an interest rate of 10.25%, compounding monthly.

- **a** If payments are \$2200 per month, calculate the amount still owing on the loan after 12 years. Write your answer, correct to the nearest cent.
- **b** If the loan is to be fully repaid after 12 years, calculate:
  - **i** the monthly repayment, correct to the nearest cent
  - **ii** the value of the final repayment
  - **iii** the total amount repaid, correct to the nearest cent
  - **iv** the total amount of interest paid, correct to the nearest cent.

6 Dan arranges to make repayments of \$450 per month to repay a loan of \$20 000, with interest being charged at 9.5% per annum, compounded monthly.
Find:

- **a** the number of monthly repayments required to pay out the loan (to the nearest month)
- **b** the amount of interest charged.

ISBN 978-1-009-25917-0

CAS

Reducing-balance loans with changing conditions

Example 12

7 An amount of $35 000 is borrowed for 20 years at 10.5% per annum, compounded monthly. Write all answers in this question correct to the nearest cent and use rounded repayment and balance values in subsequent calculations.

- **a** What are the repayments for the loan?
- **b** How much interest is paid on the loan over the 20-year period?
- **c** How much is still owing after 4 years?

After 4 years, the interest rate rises to 13.75% per annum.

- **d** What are the new repayments that will see the amount repaid in 20 years?

8 A couple negotiates a 25-year mortgage of $150 000 at a fixed rate of 7.5% per annum compounded monthly for the first 7 years, then at the market rate for the remainder of the loan. They agree to monthly repayments of $1100 for the first 7 years. Calculate:

- **a** the amount still owing after the first 7 years
- **b** the new monthly repayments required to pay off the loan if after 7 years the market rate has risen to 8.5% per annum.

9 A couple puts a $50 000 down-payment on a new home and arranges to pay off the rest in monthly payments of $1384 for 30 years at a monthly compounded interest rate of 8.5% per annum. Write all answers to this question correct to the nearest cent and use rounded repayment and balance values in subsequent calculations.

- **a** What was the selling price of the house?
- **b** How much interest will they pay over the term of the loan?
- **c** How much do they owe after 6 years?

After 6 years the interest rates increase by 0.9%. The couple must now extend the period of their loan in order to pay it back in full.

- **d** How much will they still owe after the original 30-year period?
- **e** Calculate the new monthly repayment amount required if the couple still wishes to pay off the loan in 30 years.

8E Annuities

An **annuity** is a special type of investment that earns compound interest, but also allows for regular payments to be withdrawn from the investment. It is common for people who have retired from work to invest some or all of their superannuation funds into an annuity (sometimes called ‘purchasing an annuity’) and withdraw a regular payment (sometimes called a ‘pension’).

The calculations used to model the values of reducing-balance loans and annuities are identical. The only difference is that the value of the loan represents how much is owed at any time, whereas the value of the annuity represents how much money that is left in the investment.

Analysing annuities with recurrence relations

Annuities can be modelled using recurrence relations in the same way as reducing-balance loans, with the value representing how much is left in the annuity rather than the amount that is still owing on a loan.

Modelling an annuity

Let r be the interest *rate* per compounding period.

Let D be the payment received.

A recurrence relation that can be used to model the value of an annuity after n payments, V_n, is

$$V_0 = \text{principal}, \qquad V_{n+1} = RV_0 - D$$

where $R = 1 + \dfrac{r}{100}$

Example 13 Using a recurrence relation to model an annuity

Reza plans to travel overseas for 6 months. He invests \$12 000 in annuity that earns interest at the rate of 6% per annum, providing him with a monthly income of \$2035 per month for 6 months.

Model this annuity using a recurrence relation of the form

$$V_0 = \text{the principal}, \qquad V_{n+1} = RV_n - D$$

where V_n is the value of the annuity after n payments have been received.

Solution

1 Define the symbol V_n.

Let V_n be the value of the annuity after n payments have been received.

2 Write the value of V_0. Here V_0 is the principal of the annuity, the original amount of money invested.

$V_0 = 12\,000$

3 Because interest compounds monthly, calculate the monthly interest rate.

$r = \dfrac{6\%}{12} = 0.5\% \text{ or } 0.005$

4 Determine the value of R. Either recognise that a 0.5% increase gives an R of 1.005 or use the formula $R = 1 + r/100 = R = 1 + 0.5/100 = 1.005$.

$R = 1.005$

5 Write the value of D. In this context, D is the payment received each month.

$D = 2035$

6 Use the values of V_0, R and D to write the recurrence relation.

$V_0 = 12\,000$

$V_{n+1} = 1.005V_n - 2035$

Once we have a recurrence relation, we can use it to determine things such as the value of the annuity after a given number of payments have been received.

Example 14 Using a recurrence relation to analyse an annuity

Reza's annuity can be modelled by the recurrence relation

$$V_0 = 12\,000, \qquad V_{n+1} = 1.005V_n - 2035$$

where V_n is the value of the annuity after n payments have been received.

a Use your calculator to determine recursively the value of the annuity after Reza has received three payments from the annuity.

b Is the annuity fully paid out after six monthly payments have been made? If not, how much will the last payment have to be to ensure that the annuity terminates after 6 months?

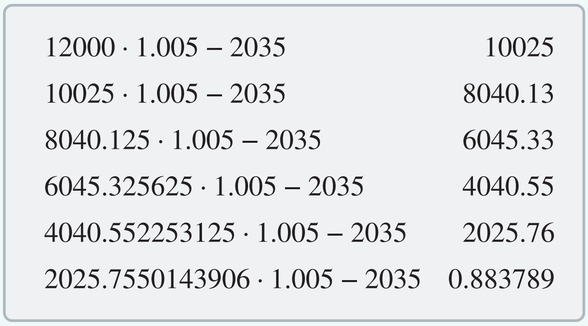
Solution

a **1** Write the recurrence relation.

$V_0 = 12\,000, \qquad V_{n+1} = 1.005V_n - 2035$

2 Type **12 000** and press enter or EXE.

3 Type **× 1.005-2035** and press enter or EXE six times to obtain the screen opposite.



```
12000 · 1.005 – 2035                          10025
10025 · 1.005 – 2035                        8040.13
8040.125 · 1.005 – 2035                     6045.33
6045.325625 · 1.005 – 2035                  4040.55
4040.552253125 · 1.005 – 2035               2025.76
2025.7550143906 · 1.005 – 2035             0.883789
```

4 Read the value from the screen. Round the answer to the nearest cent.

\$6045.33 (to the nearest cent)

b No. There will still be 0.88 cents left in the annuity. To terminate the annuity, the 88 cents is added to Reza's final payment.

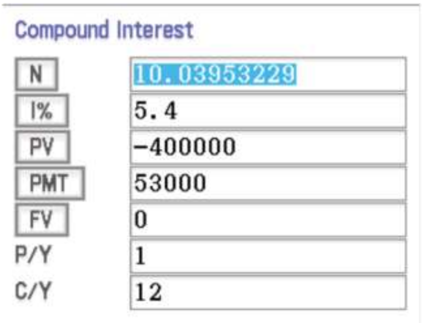
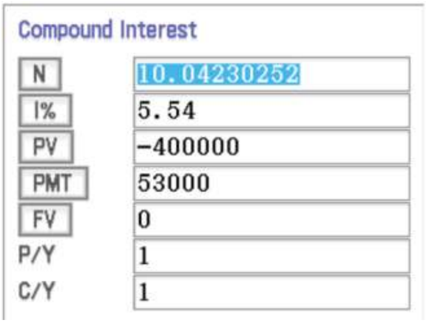
Last payment $= 2035 + 0.88$

$= \$2035.88$

When using recurrence relations to model an annuity it is crucial that both the interest rate compounds at the *same time intervals* as the payments. Most commonly, interest may be compounding annually with withdrawals made once a year. However, it may be possible for withdrawals to be made annually, while the investment accrues interest monthly. Effective interest rates may be used to align the compounding period with the withdrawals.

For example, an annuity paying an annual sum of \$53 000 is set up with an initial sum of \$400 000 and interest rate of 5.4% per annum compounding monthly. In this situation the interest is being calculated monthly, however withdrawals are only being made once per year.

To use a recursive relation to model this scenario the nominal interest rate needs to be converted to its **effective interest rate**.

Nominal interest rate: 5.4% p.a. compounding monthly	→	**Effective interest rate:** 5.54% p.a. annually
 Compound Interest N 10.03953229 I% 5.4 PV −400000 PMT 53000 FV 0 P/Y 1 C/Y 12		 Compound Interest N 10.04230252 I% 5.54 PV −400000 PMT 53000 FV 0 P/Y 1 C/Y 1 **Recursive rule:** $R = 1 + \frac{5.54}{100} = 1.0554$ $V_{n+1} = 1.0554V_n - 53\,000$ $V_0 = 400\,000$

The effective interest rate can be calculated using the formula:

$$i_{effective} = \left(1 + \frac{i}{100 \times n}\right)^n - 1$$

where i is the nominal annual interest rate and n is the number of compounding periods per annum.

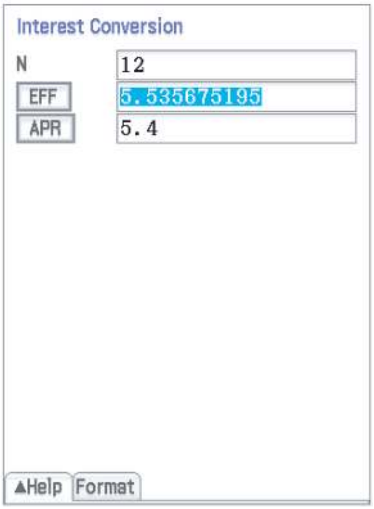
For example, if the annual interest rate for the annuity is quoted as 5.4% compounding monthly, then the effective annual interest rate charged is:

$$\begin{aligned} i_{effective} &= \left(1 + \frac{5.4}{100 \times 12}\right)^{12} - 1 \\ &= 0.05536 \end{aligned}$$

∴ Effective interest rate is 5.54% p.a.

The Financial Solver on ClassPad calculators also can convert nominal interest rates into their effective interest rates using the **Interest Conversion** option.

- **N** is the number of times interest is compounded per year
- **EFF** is the effective interest rate as a percent
- **APR** is the nominal interest rate as a percent



N 12
EFF 5.535675195
APR 5.4

Amortisation tables for annuities

An amortisation table can be used to summarise the key properties of an annuity.

An amortisation table for an annuity shows the payment number, the payment received, the interest earned and the balance of the annuity after each payment has been received. The amortisation table for Reza's annuity is shown below.

Payment number, n	*Balance at start of month, n*	*Interest earned*	*Payment*	*Balance of annuity carried forward*
1	12 000.00	60	2035.00	10 025.00
2	10 025.00	50.13	2035.00	8040.13
3	8040.13	40.20	2035.00	6045.33
4	6045.33	30.23	2035.00	4040.55
5	4040.55	20.20	2035.00	2025.76
6	2025.76	10.13	2035.00	0.89
7	0.89	0.00	0.89	0.00

Note: After the sixth payment there is 89 cents remaining in the annuity. If the annuity was to last for six payments, then the last payment could be increased by 89 cents so the annuity is reduced to exactly $0.00 after six payments have been made.

The amortisation table above can be used to summarise the key properties of an annuity.

Properties of an annuity

At each step of the loan:

1 interest earned = interest rate per compounding period × balance

For example, before payment 1 is received, interest earned

$= 0.5\% \times 12\,000 = \$60$

2 balance of annuity before payment = balance at start of month + interest earned

For example, before payment 1 is received, balance of annuity before payment

$= 12\,000 + 60 = \$12\,060$

3 balance of annuity carried forward = balance of annuity before payment − payment

For example, balance of annuity carried forward

$= 12\,060 - 2035 = 10\,025$

4 total return from the annuity = the sum of the payments received

For this annuity, the return from the annuity

$= 6 \times 2035 + 0.89 = \$12\,210.89$

Note: This amount can also be obtained by summing the payment column

5 total interest earned = total payments received − principal

For this annuity, the total interest earned

$= 12\,210.89 - 12\,000 = \$210.89$

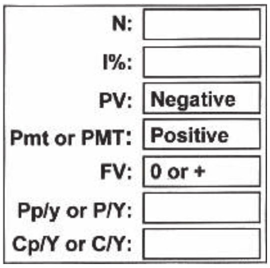
Note: This amount can also be obtained by summing the interest column

Using a finance solver to analyse annuities

Annuities usually last for many years, providing monthly payments and so it can be more convenient to analyse annuities using a finance solver.

Using the sign convention for a finance solver:

- the principal or present value, PV, is the money given by you to the bank, so it will be negative
- payments are given to you, so Pmt or PMT will be positive
- the future value of the annuity, FV, can be positive or zero:
  - zero indicates that all of the funds in the annuity have been used
  - positive indicates that some funds still exist in the annuity and further payments can be made.



N:
I%:
PV: Negative
Pmt or PMT: Positive
FV: 0 or +
Pp/y or P/Y:
Cp/Y or C/Y:

The finance solver screen shown has a summary of the sign convention for annuities.

Example 15 Solving annuity problems with a finance solver

Joe invests $200 000 into an annuity, paying 5% compound interest per annum, compounding monthly.

a If he wishes to be paid monthly payments for 10 years, how much will he receive each month?

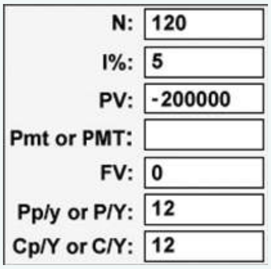
b If he receives a regular monthly payment of $3000, long will the annuity last? Give your answer correct to the nearest month.

c What interest rate, correct to one decimal place, would allow Joe to withdraw $2500 each month for 10 years?

Solution

a **1** Open the finance solver on your calculator and enter the information shown.

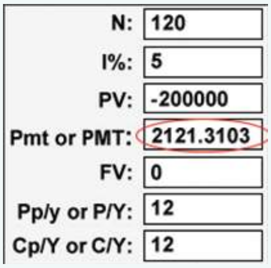
- **N**: 120 (10 years)
- **I%**: 5 (annual interest rate)
- **PV**: −200 000
- **FV**: 0 (the annuity will be exhausted after 10 years)
- **Pp/Y**: 12 (monthly payments)
- **Cp/Y**: 12 (interest compounds monthly)



N: 120
I%: 5
PV: -200000
Pmt or PMT:
FV: 0
Pp/y or P/Y: 12
Cp/Y or C/Y: 12

2 Solve for **Pmt** or **PMT**.

Note: The sign of Pmt or PMT is positive, because it is money received.



N: 120
I%: 5
PV: -200000
Pmt or PMT: 2121.3103
FV: 0
Pp/y or P/Y: 12
Cp/Y or C/Y: 12

3 Write your answer.

Joe will receive $2121.31 each month from the annuity.

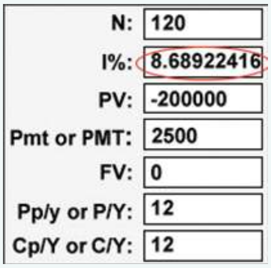
b **1** Change the payment **Pmt** or **PMT** to 3000 and solve for **N**.

N: 78.2639745
I%: 5
PV: -200000
Pmt or PMT: 3000
FV: 0
Pp/y or P/Y: 12
Cp/Y or C/Y: 12

2 Write your answer, round to the nearest month.

Joe's annuity will last 78 months.

c **1** Change the **Pmt** or **PMT** value to 2500 and the number of withdrawals, **N** to 120 (10 years). Solve for I%.



N: 120
I%: 8.68922416
PV: -200000
Pmt or PMT: 2500
FV: 0
Pp/y or P/Y: 12
Cp/Y or C/Y: 12

2 Write your answer.

Joe would require an interest rate of 8.7% per annum to make monthly withdrawals of $2500 for 10 years.

Exercise 8E

Using a recurrence relation to analyse an annuity

Example 14

CAS

1 An annuity can be modelled by the recurrence relation

$$V_0 = 5000, \qquad V_{n+1} = 1.01V_n - 1030$$

where V_n is the balance of the annuity after n monthly payments have been made. The annuity is to be fully repaid after five payments.

- **a** Use your calculator to determine recursively the balance of the annuity after three payments have been made. Give your answer to the nearest cent.
- **b** Is the annuity fully paid out after five payments of $1030 have been received? If not, how much will the last payment have to be to ensure that the balance of the annuity is zero after 5 payments?

2 An annuity can be modelled by the recurrence relation

$$V_0 = 6000, \qquad V_{n+1} = 1.005V_n - 1518$$

where V_n is the balance of the annuity after n payments have been made. The annuity is to be fully repaid after four payments.

- **a** Use your calculator to determine recursively the balance of the annuity after two payments have been made. Give your answer to the nearest cent.
- **b** Is the annuity fully paid out after four payments of $1518 have been received? If not, how much will the last payment have to be to ensure that the balance of the annuity is zero after four payments?

Using a recurrence relation to model an annuity

Example 13

3 Helen invests \$40 000 in an annuity paying interest at the rate of 6% per annum, compounding quarterly. She receives a payment of \$10 380 each quarter for 1 year. Let V_n be the balance of the loan after n payments have been received.

a Model this loan using a recurrence relation of the form:

$$V_0 = \text{the principal,} \qquad V_{n+1} = RV - D$$

b Use the recurrence relation to determine the balance of the annuity after 6 months.

CAS

Reading and interpreting an amortisation table

4 A university student is given a living allowance of \$6000 for her first year of study. She invests the money in an annuity paying an interest rate of 3% per annum, compounding monthly. From this annuity, she receives a monthly payment of \$508.

Part of the amortisation table for this annuity is shown below.

Payment number, n	*Balance of annuity, n*	*Interest earned*	*Payment*	*Balance of annuity carried forward*
1	6000.00	15.00	508.00	5507.00
2	5507.00	13.77	508.00	5012.77
3	5012.77	12.53	508.00	4517.30
4	4517.30	11.29	508.00	4020.59
5	4020.59	10.05	508.00	3522.64
6	3522.64	A	508.00	B
⋮	⋮	⋮	⋮	⋮
12	508.70	1.27	508.00	C

Determine, to the nearest cent:

a the monthly interest rate

b the interest earned before payment 1 is received

c the value of the annuity before payment 3 is received

d the balance of the annuity after payment 5 has been received

e the values of A, B and C

f the value of the last payment if the balance of the annuity is to be zero after the 12th payment is received

g the total return from the annuity

h the total amount of interest earned.

Solving annuity problems with a financial solver

Example 15

5 Leigh invests \$64 000 in an annuity, with interest of 6.25% per annum, compounding monthly. If he receives payments of \$1275 per month, how long will this annuity last? Write your answer, correct to the nearest month.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

6 Raj invests \$85 500 in an annuity, with interest of 7.25% per annum, compounding quarterly. If Raj receives a regular quarterly payment of \$5000, how long will the annuity last?
Write your answer, correct to the nearest quarter.

7 Stephanie invests \$40 000 in an annuity, with interest paid at 7.5% per annum, compounded monthly. If she wishes to receive a monthly payment for 10 years, how much will she receive each month, correct to the nearest cent?

8 Helen has \$80 000 to invest. She chooses an annuity that pays interest at the rate of 6.4% per annum, compounding monthly. Helen expects her investment to be fully exhausted after 15 years.

a Find the monthly withdrawal that Helen can make, correct to the nearest cent.

b Find the amount Helen has left in the investment after 2 years, correct to the nearest cent.

After 2 years, the interest rate of Helen's investment was reduced to 6.2% per annum, compounding monthly.

c If Helen continues to withdraw the same amount from the investment each month, how many more months, correct to the nearest month, will her investment last?

d If Helen would like her investment to still be exhausted after 15 years in total, what is the new monthly withdrawal, correct to the nearest cent, that she can make?

9 Sameep has \$150 000 to invest. He deposits the money into a savings account earning 6.2% per annum, compounding monthly, for 10 years. He makes no withdrawals during that time.

a How much money, correct to the nearest cent, is in this account after 10 years?

After 10 years Sameep invests the total amount in his savings account in an annuity investment. He will require this investment to provide monthly withdrawals of \$2800.

b What annual interest rate, correct to two decimal places, is required if Sameep's investment is to be exhausted after 12 additional years?

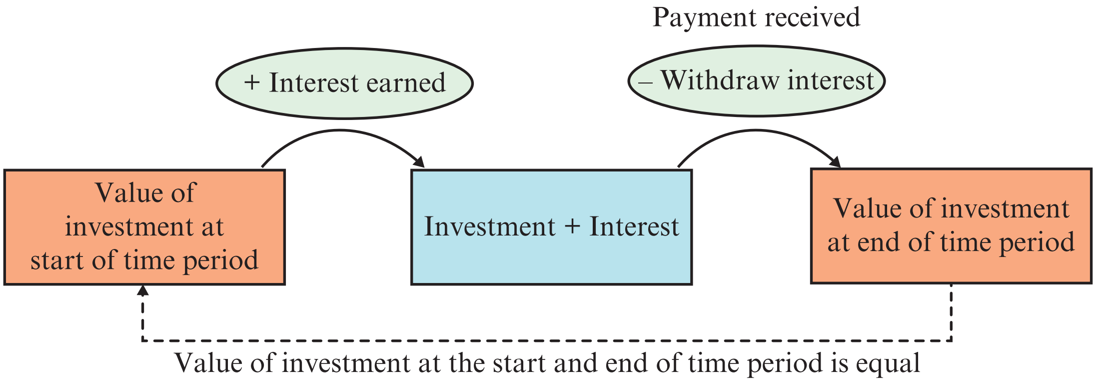
8F Perpetuities

An annuity will earn interest after every compounding period. If the payment received after each compounding period is smaller than this interest, the annuity will continue to grow. If the payment received after each compounding period is larger than this interest, the annuity will decay until nothing is left.

If the payment received is exactly the same as the interest earned in one compounding period, the annuity will maintain its value indefinitely. This type of annuity is called a **perpetuity**. Payments of the same amount as the interest earned can be made *in perpetuity*, or forever.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024



A perpetuity is an example of the trivial sequence for an annuity recurrence relation. The payment received is the same as the interest earned after each compounding period.

Perpetuities

Let V_0 be the initial value of the annuity.
Let r be the interest *rate* per compounding period.
Let D be the regular payment received.
For perpetuities:

payment received = interest earned

$$\text{so } D = \frac{r}{100} \times V_0$$

Example 16 Calculating the withdrawal from a perpetuity

Elizabeth invests her superannuation payout of \$500 000 into a perpetuity that will provide a monthly income without using any of the initial investment.

If the interest rate for the perpetuity is 6% per annum, what monthly payment will Elizabeth receive?

Solution

1 Write the values of V_0 and r.

Elizabeth invests \$500 000 so $V_0 = 500\,000$.
The annual interest rate is 6% so $r = \frac{6\%}{12} = 0.5\%$.

2 Find the monthly interest earned.

$$D = \frac{r}{100} \times V_0$$
$$D = \frac{0.5}{100} \times 500\,000$$
$$D = 2500$$

3 Write your answer, rounding as required.

Elizabeth will receive \$2500 every month from her investment.

Example 17 Calculating the investment required to establish a perpetuity

How much money will need to be invested in a perpetuity account, earning interest of 4.2% per annum compounding monthly, if $200 will be withdrawn every month? Write your answer to the nearest dollar.

Solution

1 Write the values of r and D.

The annual interest rate is 4.2%,

so $r = \frac{4.2\%}{12} = 0.35\%$.

The amount withdrawn each month is $200 so $D = 200$.

2 Use the rule $D = \frac{r}{100} \times V_0$ to write an expression that can be solved for V_0.

$$200 = 0.0035 \times V_0$$

$$V_0 = \frac{200}{0.0035}$$

$$= \$57\,142.857$$

3 Write your answer.

$57 143 will need to be invested to establish the perpetuity investment.

Problems involving perpetuities can be solved using a financial calculator.

Example 18 Calculating the interest rate of a perpetuity

A university mathematics faculty has $30 000 to invest. It intends to award an annual mathematics prize of $1500 with the interest earned from investing this money in a perpetuity.

What is the minimum interest rate that will allow this prize to be awarded indefinitely?

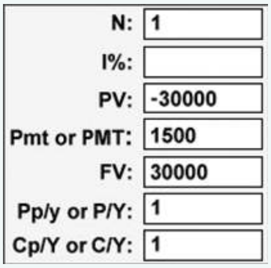
Solution

Using a financial solver

We will consider just one compounding period because all compounding periods will be identical.

1 Open the finance solver on your calculator and enter the information below.

- **N**: 1 (one payment)
- **PV**: –30 000
- **Pmt or PMT**: 1500 (prize is $1500 each year)
- **FV**: 30 000 (the balance will be the same after each payment)
- **Pp/Y**: 1 (yearly payment)
- **Cp/Y**: 1 (interest compounds yearly)



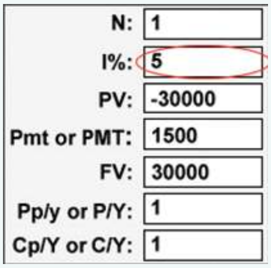
N: 1
I%:
PV: -30000
Pmt or PMT: 1500
FV: 30000
Pp/y or P/Y: 1
Cp/Y or C/Y: 1

ISBN 978-1-009-25917-0

2 Solve for the unknown interest rate (**I%**). On the:

- *TI-Nspire:* Move the cursor to the **I%** entry box and press `enter` to solve.
- *ClassPad:* Tap on the **I%** entry box and tap **Solve**.

  The amount **5** now appears in the **I%** entry box.



N: 1
I%: 5
PV: -30000
Pmt or PMT: 1500
FV: 30000
Pp/y or P/Y: 1
Cp/Y or C/Y: 1

3 Write your answer, rounding as required.

The minimum annual interest rate to award this prize indefinitely is 5%.

Alternative method

The prize is equal to the $r\%$ of the amount invested.

1 Set up an equation with r.

Prize = $r\%$ of investment

$$1500 = r\% \times 30\,000$$

$$1500 = \frac{r}{100} \times 30\,000$$

2 Solve the equation for r.

$$1500 = r \times 300$$

$$r = \frac{1500}{300} = 5$$

3 Write your answer.

The minimum annual interest rate to award this prize indefinitely is 5%.

Exercise 8F

Calculating the principal and interest rate of a perpetuity

Example 17
Example 18

CAS

1 Geoff wishes to set up a fund so that every year \$2500 is donated to the RSPCA.

a If the interest on his initial investment averages 2.5% per annum, compounding annually, how much should he invest?

b If Geoff has \$80 000 to invest, what is the minimum interest rate he requires to provide for the donation in perpetuity?

2 Barbara wishes to start a scholarship that will reward the top mathematics student each year with a \$500 prize.

a If the interest on the initial investment averages 2.7% per annum, compounding annually, how much should be invested? Answer correct to the nearest dollar.

b Barbara has \$12 000 to invest in a perpetuity to provide this prize. What is the minimum interest rate, correct to two decimal places, that she requires in order to pay the prize in perpetuity?

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

3 Cathy wishes to maintain an ongoing donation of \$5500 per year to the Collingwood Football Club. If the interest on the initial investment averages 2.75% per annum, compounding annually, how much should she invest?

Calculating the payment from a perpetuity

Example 16

4 Craig wins \$1 000 000 in a lottery and decides to place it in a perpetuity that pays 5.75% per annum interest, compounding monthly. What monthly payment, correct to the nearest cent, does he receive?

5 Suzie invests her inheritance of \$642 000 in a perpetuity that pays 6.1% per annum, compounding quarterly.

- **a** What quarterly payment does she receive?
- **b** After five quarterly payments, how much money remains invested in the perpetuity?
- **c** After 10 quarterly payments, how much money remains invested in the perpetuity?

8G Compound interest investments with regular additions to the principal (annuity investments)

Suppose you have some money invested in an account that pays compound interest. To increase the rate at which your investment grows, you decide to add to your investment by making additional payments on a regular basis. This type of investment is sometimes called an annuity invesment.

Analysing annuity investments with recurrence relations

Just like reducing-balance loans and annuities, annuity investments can be modelled using recurrence relations.

Modelling compound interest investments with regular additions to the principal

Let r be the *interest rate* per compounding period.
Let D be the *payment* made.
A recurrence relation that can be used to model the value of a compound interest investment, V_n, after n additional payments have been made is

$$V_0 = \text{the principal}, \qquad V_{n+1} = RV_n + D \quad \text{where} \quad R = 1 + \frac{r}{100}$$

ISBN 978-1-009-25917-0

Example 19 Modelling compound interest investments with additions to the principal

Nor plans to travel overseas when she finishes Year 12. She has already saved \$1200 and thinks that she can save an additional \$50 each month that she plans to add to her savings account. The account pays interest at a rate of 3% per annum, compounding monthly.

Model this investment using a recurrence relation of the form

$$V_0 = \text{the principal}, \qquad V_{n+1} = RV_n + D$$

where V_n is the value of the investment after n payments (additions to the principal) have been made.

Solution

1 Define the variable V_n.

Let V_n be the value of the investment after n payments.

2 Write the value of V_0. Here, V_0 is the principal of the investment, the original amount of money invested.

$V_0 = 1200$

3 Because interest compounds monthly, calculate the monthly interest rate.

$r = \frac{3\%}{12} = 0.25\%$

4 Determine the value of R. Either recognise that a 0.25% increase gives an R of 1.0025, or use the formula $R = 1 + r/100$, so $R = 1 + 0.25/100 = 1.0025$.

$R = 1.0025$

5 Write the value of D. In this context, D is the amount added each month.

$D = 50$

6 Use the values of V_0, R and D to write the recurrence relation.

$V_0 = 50$
$V_{n+1} = 1.0025V_n + 50$

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Once we have a recurrence relation, we can use it to determine things such as the value of the investment after a given number of additional payments have been made.

Example 20 Using a recurrence relation to analyse compound interest investments with additions to the principal

Nor's investment can be modelled by the recurrence relation

$$V_0 = 1200, \qquad V_{n+1} = 1.0025V_n + 50$$

where V_n is the value of the investment after n payments have been received.

a Use your calculator to determine recursively the value of the investment after Nor has made three additional payments to her investment.

b What will be the value of her investment after 1 year?

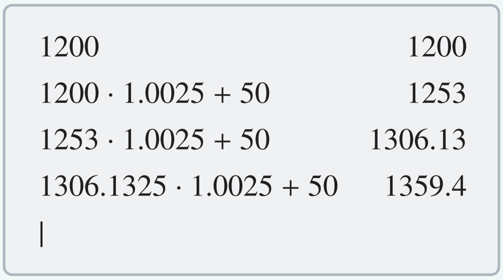
Solution

a **1** Write the recurrence relation.

$V_0 = 1200, \qquad V_{n+1} = 1.0025V_n + 50$

2 Type 1200 and press `enter` (or `EXE`).

3 Type × 1.0025 + 50 and press `enter` (or `EXE`) three more times to obtain the screen opposite.

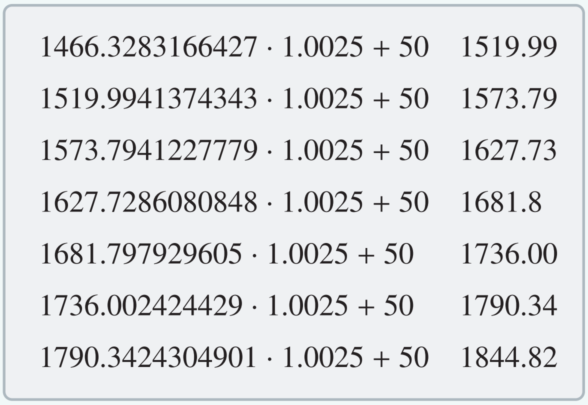


```
1200                          1200
1200 · 1.0025 + 50            1253
1253 · 1.0025 + 50         1306.13
1306.1325 · 1.0025 + 50     1359.4
|
```

4 Write your answer.

After three additional payments, the value of Nor's investment is $1359.40.

b **1** Continue pressing `enter` or `EXE` nine more times to obtain the screen opposite. The value of the investment after 12 additional payments can now be read from the screen.



```
1466.3283166427 · 1.0025 + 50    1519.99
1519.9941374343 · 1.0025 + 50    1573.79
1573.7941227779 · 1.0025 + 50    1627.73
1627.7286080848 · 1.0025 + 50    1681.8
1681.797929605 · 1.0025 + 50     1736.00
1736.002424429 · 1.0025 + 50     1790.34
1790.3424304901 · 1.0025 + 50    1844.82
```

2 Write your answer.

After 1 year, the value of Nor's investment is $1844.82.

ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

Amortisation tables for compound interest investments with additions to the principal

An amortisation table for an annuity shows the payment number, the payment made, the interest earned and the balance of the investment after each payment has been received. The amortisation table for Nor's investment annuity follows.

Payment number, n	*Balance of annuity, n*	*Interest earned*	*Payment*	*Balance of investment carried forward*
1	1200.00	3.00	50.00	1253.00
2	1253.00	3.13	50.00	1306.13
3	1306.13	3.27	50.00	1359.40
4	1359.40	3.40	50.00	1412.80
5	1412.80	3.53	50.00	1466.33
6	1466.33	3.67	50.00	1519.99
7	1519.99	3.80	50.00	1573.79
8	1573.79	3.93	50.00	1627.73
9	1627.73	4.07	50.00	1681.80
10	1681.80	4.20	50.00	1736.00
11	1736.00	4.34	50.00	1790.34
12	1790.34	4.48	50.00	1844.82

The amortisation table above can be used to summarise the key properties of a compound interest investment with addition payments.

Properties of an investment

At each step of the investment:

1 interest earned = interest rate per compounding period × previous balance

For example, before payment 1 is received, interest earned

$= 0.25\% \times 1200 = \$3.00$

2 balance of annuity before payment = balance at start of period + interest earned

For example, before payment 1 is received, balance of annuity before payment

$= 1200 + 3 = \$1203.00$

3 balance of investment carried forward = balance of annuity before payment + payment

For example, balance of investment carried forward

$= 1203 + 50 = \$1253.00$

4 total interest earned = balance of investment carried forward – (principal + total payments made)

After 12 months, the total interest earned

$= 1844.82 - (1200 + 12 \times 50) = \$44.82$

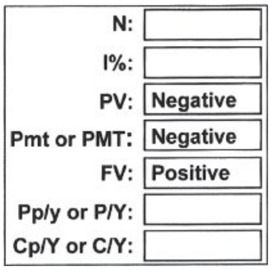
Note: This amount can also be obtained by summing the interest column

Using a financial solver to analyse a compound interest investment with regular additions to the principal

Just like reducing-balance loans and annuities, compound interest investments with regular additions often last for many compounding periods. It is once again very convenient to use a finance solver to analyse them.

Using the sign convention for a finance solver:

- the principal or present value, PV, is the money given by you to the bank, so it will be negative
- payments are given by you to the bank, so Pmt or PMT will be negative
- the future value of the annuity, FV, must be positive. The value will grow until you decide to stop making additions. The value is money that the bank will give back to you.



N:
I%:
PV: Negative
Pmt or PMT: Negative
FV: Positive
Pp/y or P/Y:
Cp/Y or C/Y:

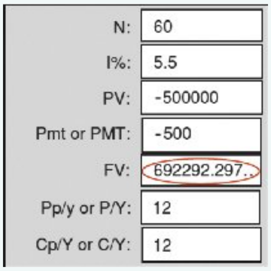
The finance solver screen shown has a summary of the sign convention for annuity investments.

Example 21 Determining the value of an investment with regular additions made to the principal using a financial solver

Lars invests $500 000 at 5.5% per annum, compounding monthly. He makes a regular deposit of $500 per month into the account. What is the value of his investment after 5 years?

Solution

1 Open the finance solver on your calculator and enter the information below, as shown opposite.
  - **N**: 60 (5 years)
  - **I%**: 5.5
  - **PV**: –500 000
  - **PMT**: –500
  - **FV**: to be determined
  - **Pp/Y**: 12 payments per year
  - **Cp/Y**: 12 compounding periods per year



N: 60
I%: 5.5
PV: -500000
Pmt or PMT: -500
FV: 692292.297...
Pp/y or P/Y: 12
Cp/Y or C/Y: 12

2 Solve for FV and write your answer correct to the nearest cent.

After 5 years, Lars' investment will be worth $692 292.30.

Exercise 8G

Using a recurrence relation to model and analyse an investment with additions to the principal

Example 20

CAS

1 A compound interest loan with regular additions to the principal can be modelled by the recurrence relation

$$V_0 = 2000, \qquad V_{n+1} = 1.08V_n + 1000$$

where V_n is the value of the investment after n yearly payments (additions to the principal) have been made.

a Use your calculator to determine recursively the balance of the investment after 5 years. Give your answer to the nearest cent.

b What is the principal of this investment?

c How much is added to the principal each year?

d What is the annual interest rate?

2 A compound interest loan with regular additions to the principal can be modelled by the recurrence relation

$$V_0 = 20\,000, \qquad V_{n+1} = 1.025V_n + 2000$$

where V_n is the value of the investment after n quarterly payments (additions to the principal) have been made.

a Use your calculator to determine recursively the balance of the investment after three quarterly payments have been made. Give your answer to the nearest cent.

b What is the principal of this investment?

c How much is added to the principal each quarter?

d What is the quarterly interest rate? What is the annual interest rate?

Example 19

3 Sarah invests \$1500 at 9% per annum, compounding monthly. She plans to add an additional \$40 to her investment each month.

Let V_n be the value of the investment after Sarah has made n monthly payments to her investment.

Write a recurrence relation to model Sarah's investment.

Reading and interpreting an amortisation table

4 The amortisation table below charts the growth of a compound interest investment with regular additions made to the principal each month.

Payment number, n	*Balance of annuity, n*	*Interest earned*	*Payment*	*Balance of annuity carried forward*
1	5000.00	50.00	100.00	5150.00
2	5150.00	51.50	100.00	5301.50
3	A	B	100.00	C
4		54.55	100.00	5609.06
5	5609.06	56.09	100.00	5765.15
6	5765.15	57.65	100.00	5922.80
7	5922.80	59.23	100.00	6082.03
8	6082.03	60.82	100.00	6242.85

a Write:
  - **i** the original amount invested
  - **ii** the amount added to the principal each month
  - **iii** the amount of interest earned in month 1
  - **iv** the monthly interest rate
  - **v** the annual interest rate
  - **vi** the value of the investment before payment 2
  - **vii** value of the investment after 8 months.

b Determine, to the nearest cent:
  - **i** the values of A, B and C
  - **ii** the total interest earned by the investment after 8 months.

Solving problems involving compound interest loans with additions to the principal using a financial solver

Example 21

5 Lee invests \$12 000 at 5.7% per annum, compounding monthly. He makes a regular deposit of \$250 per month into the account. What is the value of his investment after 10 months?

CAS

6 A sporting club invests \$10 000 at 5.2% per annum, compounding monthly. The club plans to add \$200 to the loan each month. How long will it take for the investment to be worth at least \$12 000? Give your answer as a whole number of months.

7 Bree has $25 000 in an account that pays interest at a rate of 6.15% per annum, compounding monthly.

a If she makes monthly deposits of $120 to the account, how much will she have in the account at the end of 5 years?

b If she makes monthly withdrawals of $120 from the account, how much will she have in the account at the end of 5 years?

8 Jarrod saves $500 per month in an account that pays interest at a rate of 6% per annum, compounding monthly.

a If he makes monthly deposits of $500 to the account, how much will he have in the account at the end of 10 years?

b Suppose that, after 10 years of making deposits, Jarrod starts withdrawing $500 each month from the account. How much will he have in the account at the end of another 10 years?

Key ideas and chapter summary

Nominal interest rate	A **nominal interest rate** is an annual interest rate for a loan or investment.
Effective interest rate	The **effective interest rate** is the interest earned or charged by an investment or loan, written as a percentage of the original amount invested or borrowed. Effective interest rates allow loans or investments with different compounding periods to be compared. Effective interest rates can be calculated using the rule $r_{\text{effective}} = \left(\left(1 + \frac{r}{100 \times n}\right)^n - 1\right) \times 100\%$ where r is the nominal interest rate and n is the number of compounding periods in 1 year.
Reducing-balance loan	A **reducing-balance loan** is a loan that attracts compound interest but is reduced in value by making regular payments. Each payment partly pays the interest that has been added and partly reduces the value of the loan.
Recurrence relation model for a reducing-balance loan	The **recurrence relation model for a reducing-balance loan** has the form $V_0 = principal, \quad V_{n+1} = R\,V_n - D$ where V_n is the balance of the loan after n repayments, r is the interest rate per compounding period, D is the repayment made and $R = 1 + \frac{r}{100}$.
Amortisation	An amortising loan is one that is paid back with periodic payments. An amortising investment is one that is exhausted by regular withdrawals. **Amortisation** of reducing-balance loans tracks the distribution of each periodic payment, in terms of the interest paid and the reduction in the value of the loan. Amortisation of an annuity tracks the source of each withdrawal, in terms of the interest earned and the reduction in the value of the investment.
Amortisation table	An **amortisation table** shows the amortisation (payment) of all or part of a reducing-balance loan or annuity. It has columns for the payment number, the payment amount, the interest paid or earned and the balance after the payment has been made.

Finance Solver	**Finance Solver** is a function on a CAS calculator that performs financial calculations. It can be used to determine any of the principal, interest rate, periodic payment, future value or number of payments given all of the other values.
Sign convention	When using a **finance solver**, it is important to observe the following sign convention for the values that are used. Values will be: ■ positive if they represent money that you receive or that someone owes you ■ negative if they represent money that you pay out or that you owe someone.
Annuity	An **annuity** is an investment that earns compound interest and from which regular payments are made.
Perpetuity	A **perpetuity** is an annuity where the regular payments or withdrawals are the same as the interest earned. The value of a perpetuity remains constant.
Annuity investment	An **annuity investment** is a compound interest investment to which additional payments are added after every compounding period.
Recurrence relation model for an annuity investment	The **recurrence relation model for an annuity investment** has the form $V_0 = principal, \qquad V_{n+1} = R\,V_n + D$ where V_n is the balance of the investment after n additional payments, r is the interest rate per compounding period, D is the additional payment made and $R = 1 + \frac{r}{100}$.

Skills check

Having completed this chapter, you should be able to:

- calculate effective interest rates using a rule and a CAS calculator
- compare loans and investments using effective interest rates
- analyse compound interest loans and investments using a finance solver
- analyse reducing-balance loans using recurrence relations and a finance solver
- analyse annuities using recurrence relations and a finance solver
- analyse perpetuities using recurrence relations and a finance solver
- analyse annuity investments using recurrence relations and a finance solver.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Short-answer questions

1 V_n is the value of an investment after n years. The value of this investment from *month to month* is modelled by the recurrence relation $V_0 = 25\,000$, $V_{n+1} = 1.007V_n - 400$. Determine the annual interest rate for this investment.

2 An investment of \$18 000, earning compound interest at the rate of 6.8% per annum, compounding yearly, and with regular additions of \$2500 every year can be modelled with a recurrence relation. If V_n is the value of the investment after n years, write a recurrence relation for this investment.

3 A loan of \$28 000 is charged interest at the rate of 6.4% per annum, compounding monthly. It is repaid with regular monthly repayments of \$1200.

CAS

a Correct to the nearest cent, determine the value of the loan after 5 months.

b Calculate the final repayment on the loan, correct to the nearest cent.

4 A scholarship will be set up to provide an annual prize of \$400 to the best mathematics student in a school. The scholarship is paid for by investing an amount of money into a perpetuity, paying interest of 3.4% per annum, compounding annually. Determine the amount that needs to be invested to provide this scholarship.

5 A loan of \$6000, plus interest, is to be repaid in full in 12 quarterly repayments. Interest at 10% per annum is calculated on the remaining balance each quarter. Calculate the amount of the repayment required to fully pay off the loan.

6 Monthly withdrawals of \$220 are made from an account that has an opening balance of \$35 300, invested at 7% per annum, compounding monthly. Calculate the balance of the account after 1 year, to the nearest dollar.

7 Paula borrows \$12 000 from a bank, to be repaid over 5 years. Interest of 12% per annum is charged monthly on the amount of money owed. If Paula makes monthly repayments, determine the amount she owes at the end of the second year.

8 The amortisation table for a reducing-balance loan is as follows:

Repayment number, n	*Balance of loan, n (\$)*	*Interest (\$)*	*Repayment (\$)*	*Balance owing at end (\$)*
1	40 000.00	160.00	400.00	39 760.00
2	39 760.00	159.04	400.00	39 519.04
3	39 519.04	158.08	400.00	39 277.12
4	39 277.12	157.11	400.00	39 034.22
5	39 034.22	156.14	400.00	38 790.36
6	38 790.36	155.16	400.00	38 545.52

ISBN 978-1-009-25917-0

Cambridge University Press

Updated January 2024

Determine:

a the principal of this loan

b the periodic repayment amount on this loan

c the amount owing on the loan prior to the fifth repayment

d as a percentage of the repayment amount, the interest paid from the sixth repayment

Extended-response questions

CAS

1 Barry is considering borrowing \$250 000 to buy a house. A home loan at his bank will charge interest at the rate of 4.8% per annum, compounding monthly. Barry can afford repayments of \$1800 every month.

Let V_n be the value of Barry's loan after n months.

a Write a recurrence model for the value of Barry's loan.

b After 12 months, how much would Barry owe on this loan?

c How many months would it take Barry to reduce the value of his loan below \$200 000?

2 Samantha inherited \$150 000 from her aunt. She decides to invest this money into an account paying 6.25% per annum interest, compounding monthly.

a If Samantha deposited her money into a perpetuity, what monthly payment would she receive?

b If Samantha deposited her money into an annuity and withdrew \$1000 per month, how much would she have in the account after 1 year?

c If Samantha deposited her money into an annuity and withdrew \$2000 per month, how long would it take for the value of her investment to drop below \$100 000?

d If Samantha deposited her money into an annuity and withdrew \$4000 per month:

  i how long would her investment last?

  ii including interest, what would be the value of her last withdrawal?

3 A loan of \$10 000 is to be repaid over 5 years. Interest is charged at the rate of 11% per annum, compounding quarterly.

Find:

a the quarterly repayment, correct to the nearest cent

b the total cost of paying off the loan, to the nearest dollar

c the total amount of interest paid.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

4 Interest on a reducing balance loan of $65 000 is compounded quarterly at an interest rate of 12.75% per annum.
Calculate the quarterly repayment if:

a the amount still owing after 10 years is $25 000

b the amount still owing after 20 years is $25 000

c the loan is fully repaid after 10 years

d the loan is fully repaid after 20 years.

5 The Andersons were offered a $24 800 loan to pay the total cost of a new car. Their loan is to be repaid in equal monthly repayments of $750, except for the last month when less than this will be required to fully pay out the loan. They will pay 10.8% interest per annum, calculated monthly on the reducing balance.

a Calculate the least number of months needed to repay this loan plus interest.

b Calculate, to the nearest cent, the amount of the final repayment.

c When the Andersons took out their loan they had the choice of making monthly payments of $750 or quarterly repayments of $2250. They chose to make monthly payments of $750. In either case they would have to pay 10.8% interest per annum calculated monthly on the reducing balance. In terms of the total amount of money they would have to pay to repay the loan, did they make the correct decision? Explain your answer without making any further calculations.

6 When a family bought their home they borrowed $100 000 at 9.6% per annum, compounded quarterly. The loan was to be repaid over 25 years in equal quarterly repayments.

a How much of the first quarterly repayment went towards paying off the principal?

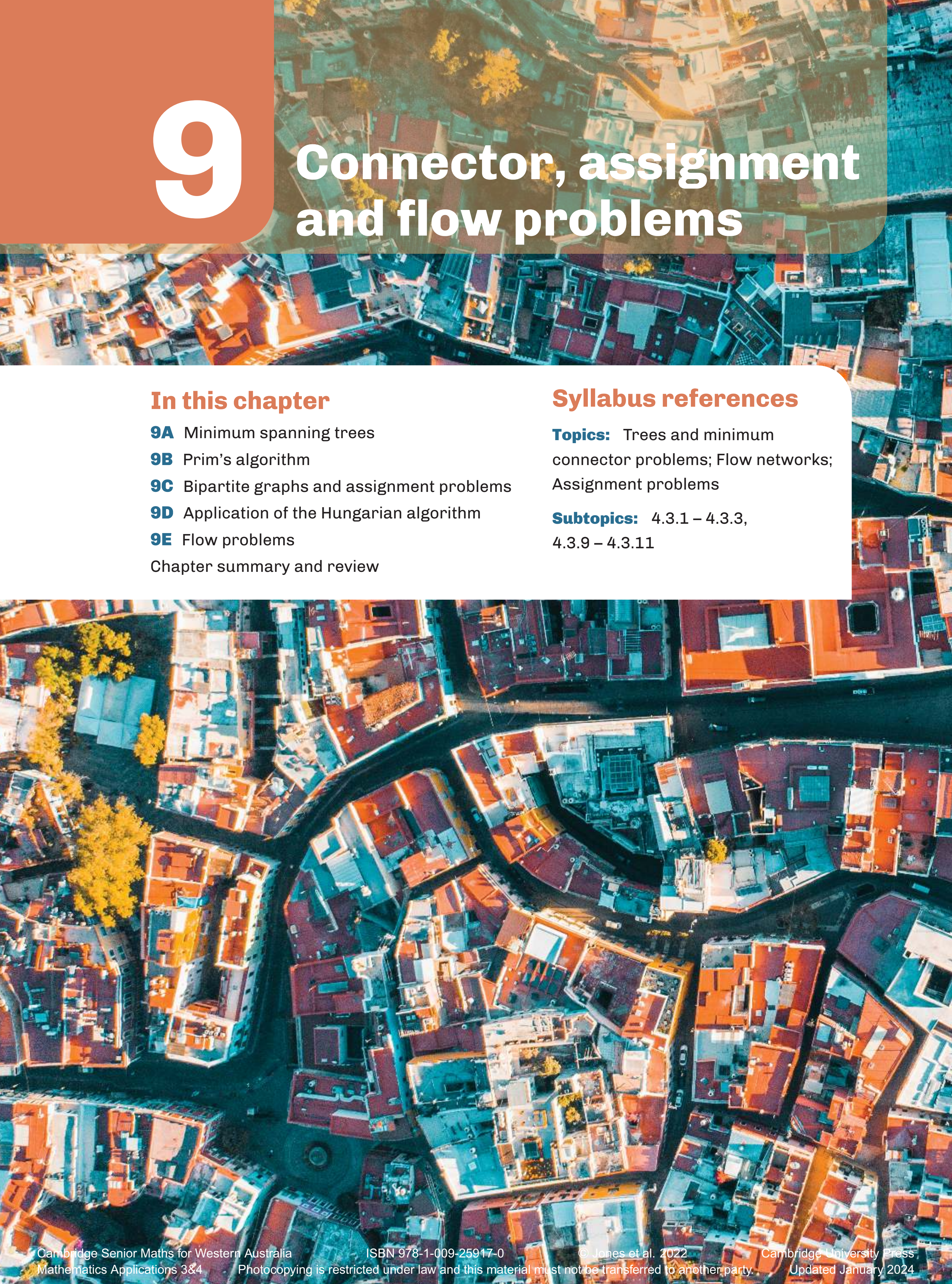
b The family inherit some money and decide to terminate the loan after 10 years and pay what is owing in a lump sum. How much will this lump sum be?

7 Helene has won $750 000 in a lottery. She decides to place the money in an investment account that pays 4.5% per annum interest, compounding monthly.

a How much will Helene have in the investment account after 10 years?

b After the 10 years are up, Helene decides to use her money to invest in an annuity, which pays 3.5% per annum, compounding monthly. If Helen requires $6000 per month for her living expenses, how long will the annuity last?

c Helene's accountant suggests that rather than purchase an annuity she places the money in a perpetuity so that she will be able to leave some money to her grandchildren. If the perpetuity pays 3.5% per annum compounding monthly, how much is the monthly payment that Helene will receive?



9

Connector, assignment and flow problems

In this chapter

Syllabus references

Topics: Trees and minimum connector problems; Flow networks; Assignment problems

Subtopics: 4.3.1 – 4.3.3, 4.3.9 – 4.3.11

Cambridge Senior Maths for Western Australia
Mathematics Applications 3&4
ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

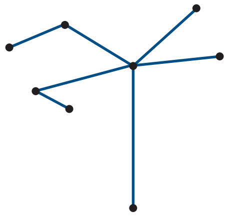
9A Minimum spanning trees

In the previous applications of networks, the weights on the edges of the graph were used to determine a minimum pathway through the graph. In other applications, it is more important to minimise the number and weights of the edges in order to keep all vertices connected to the graph. For example, a number of towns might need to be connected to a water supply. The cost of connecting the towns can be minimised by connecting each town into a network or water pipes only once, rather than connecting each town to every other town.

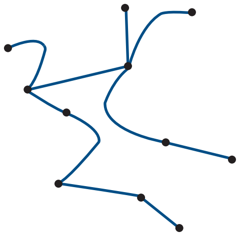
Problems of this type are called connector problems. In order to solve connector problems, you need to learn the language of networks that have as few edges as possible.

Trees

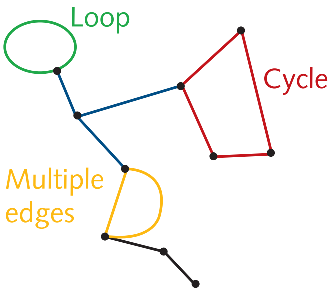
A **tree** is a connected graph that contains no cycles, multiple edges or loops. A tree may be part of a larger graph. For example, Graphs 1 and 2 below are examples of trees. Graph 3 is *not* a tree.



Graph 1: a tree



Graph 2: a tree



Graph 3: not a tree

Graphs 1 and 2 are trees: they are connected and have no cycles, multiple edges or loops.

Graph 3 is *not* a tree: it has several cycles (loops and multiple edges count as cycles).

For trees, there is a relationship between the number of vertices and the number of edges.

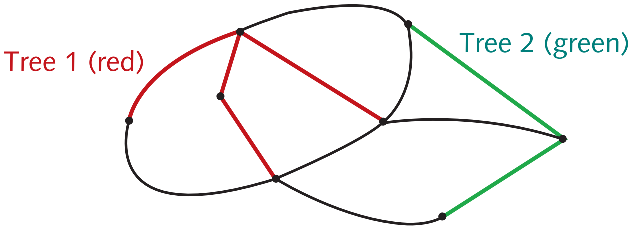
- Graph 1, a tree, has 8 vertices and 7 edges.
- Graph 2, a tree, has 11 vertices and 10 edges.

An inspection of other trees would show that, in general, the number of edges in a tree is one less than the number of vertices.

Rule connecting the number of edges of a graph to the number of vertices

A *tree* with n vertices has $n - 1$ edges.

Another important fact about trees is that every connected graph contains one or more trees. For example, the connected graph below contains multiple trees. Two of these are shown superimposed on the graph in colour.



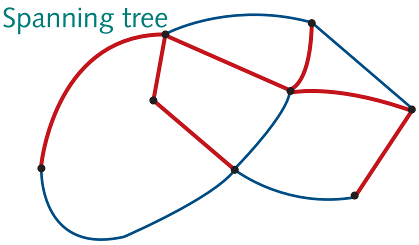
Spanning trees

Spanning tree

A **spanning tree** is a tree that connects *all* of the vertices in a connected graph.

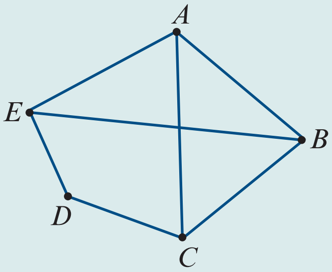
An example of a spanning tree is shown on the graph opposite. There are several other possibilities.

Note that the spanning tree, like the graph, has 8 vertices. From this it can be concluded that the spanning tree will have 8 − 1 = 7 edges, as we can observe.



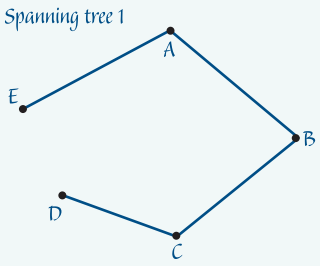
Example 1 Finding a spanning tree in a network

Find two spanning trees for the graph shown opposite.



Solution

1 The graph has five vertices and seven edges.
A spanning tree will have five (n) vertices and four ($n - 1$) edges.



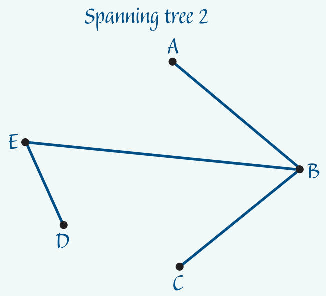
2 To form a spanning tree, remove any *three* edges, provided that:

- all the vertices remain connected
- there are no multiple edges or loops.

Spanning tree 1 is formed by removing edges *EB, ED* and *CA.*

Spanning tree 2 is formed by removing edges *EA, AC* and *CD.*

Note: Several other possibilities exist.



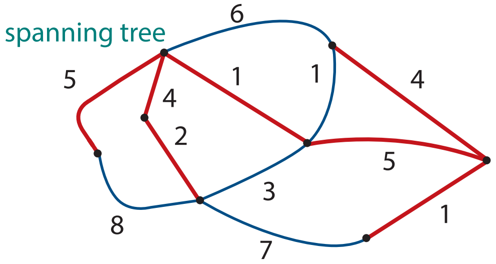
Minimum spanning trees

For weighted graphs or networks, it is possible to determine the 'length' of each spanning tree by adding up the weights of the edges in the tree.

For the spanning tree shown below:

Length = 5 + 4 + 2 + 1 + 5 + 4 + 1

= 22 units



Minimum spanning tree

A **minimum spanning tree** is the spanning tree of *minimum* length (may be minimum distance, minimum time, minimum cost, etc.). There may be more than one minimum spanning tree in a weighted graph.

Minimum spanning trees have many real world applications such as planning the layout of a computer network or a water supply system for a new housing estate. In these situations, we might want to minimise the amount of cable or water pipe needed for the job. Alternatively, we might want to minimise the time needed to complete the job or its cost.

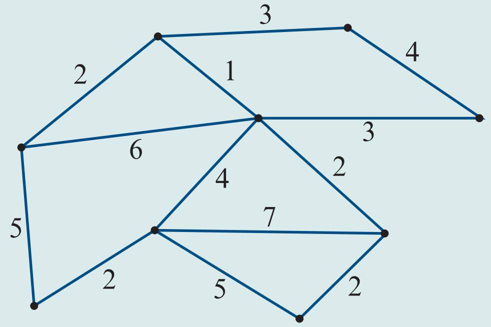
ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

Example 2 Finding the length of a spanning tree

a Draw one spanning tree for the graph shown.

b Calculate the length of the spanning tree.



Solution

a **1** Count the number of vertices and edges in the graph.

There are 9 vertices and 13 edges.

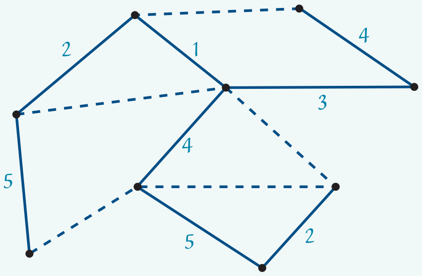
2 Calculate the number of edges in the spanning tree.

The spanning tree will have 8 edges.

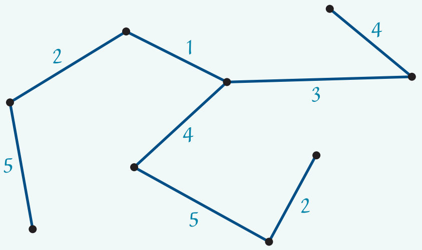
3 Calculate how many edges must be removed.

Remove 13 − 8 = 5 edges.

4 Choose edges to remove.



b Add the weights of the remaining edges.



$$\text{Length} = 5 + 2 + 1 + 4 + 5 + 2 + 3 + 4$$
$$= 26 \text{ units}$$

Prim's algorithm

Prim's algorithm for finding a minimum spanning tree

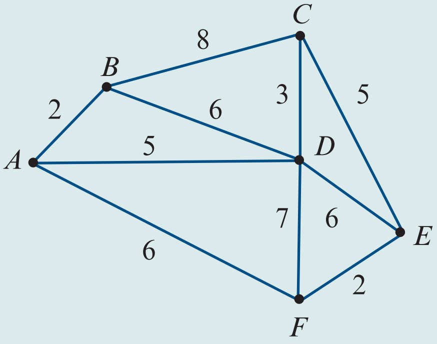
Prim's algorithm is a set of rules to determine a minimum spanning tree for a graph.

- Choose a starting vertex (any will do). Inspect the edges starting from this vertex and choose the one with the lowest weight. (If two edges have the same weight, the choice can be arbitrary.) You now have two vertices and one edge.
- Next, inspect the edges starting from the vertices. Choose the edge with the lowest weight, ignoring any that would create a cycle. (If two edges have the same weight, the choice can be arbitrary.) You now have three vertices and two edges.
- Repeat the process until all the vertices are connected, and then stop. The result is a minimum spanning tree.

Prim's algorithm can also be used to find the minimum spanning tree when the weights of the graph are given in a table or distance matrix. The steps for how to do this are explored in section 9B.

Example 3 Applying Prim's algorithm

Apply Prim's algorithm to obtain a minimum spanning tree for the network shown, and calculate its length.

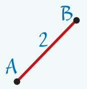


Solution

1 Start at A:

AB is the lowest-weighted edge (2).

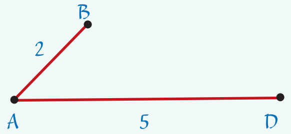
Draw it in.



2 From A or B:

AD is the lowest-weighted edge (5).

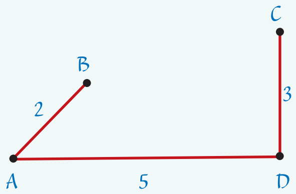
Draw it in.



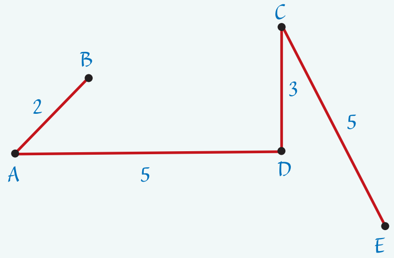
3 From A, B or D:

DC is the lowest-weighted edge (3).

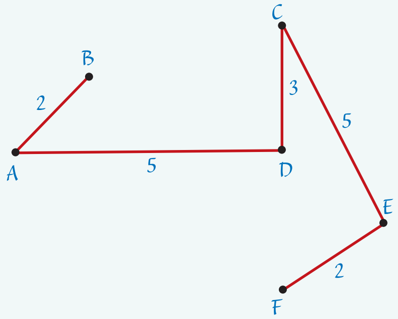
Draw it in.



4 From A, B, C or D:
CE is the lowest-weighted edge (5).
Draw it in.



5 From A, B, C, D or E:
EF is the lowest-weighted edge (2).
Draw it in.
All vertices have now been joined.
The minimum spanning tree is determined.



6 Find the length of the minimum spanning tree by adding the weights of the edges.

Minimum spanning tree
Length = 2 + 5 + 3 + 5 + 2 = 17 units

Connector problems

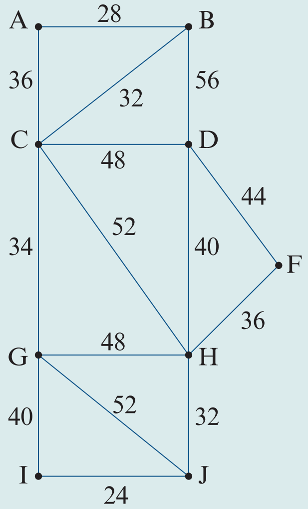
Minimum spanning trees represent the least weight required to keep all of the vertices connected in the graph. If the edges of a graph represent the cost of connecting towns to a gas pipeline, then the total weight of the minimum spanning tree would represent the minimum cost of connecting the towns to the gas. This is an example of a *connector problem*, where the cost of keeping towns or other objects connected is important to make as low as possible.

Example 4 Solving a connector problem

The map on the right shows the distances, in kilometres, between electric vehicle charging stations in Western Australia.

a Indicate on the diagram the minimum spanning tree for this network.

b Find the total length of the minimum spanning tree.



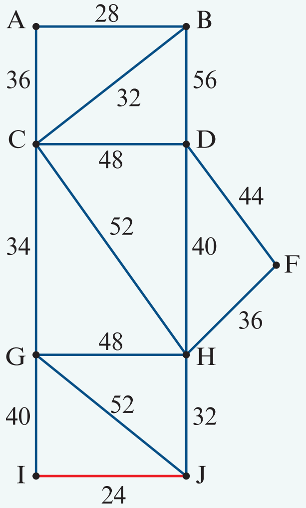
c The connections from CG and CH are removed to make way for a new charging station at vertex E. The new charging station will connect to existing stations C, D, G and H with the distances given in the table below.

	C	D	G	H
E	26	30	20	28

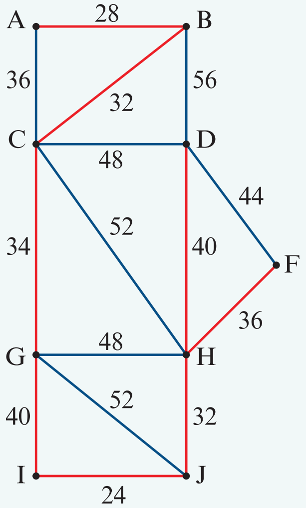
Determine how the addition of vertex E effects the overall length of cable needed to ensure all charging stations are connected to the network, but not necessarily by a direct connection.

Solution

a **1** To determine the minimum spanning tree for the network, find the edge with the lowest weighting and mark this on the diagram.



2 Follow Prim's algorithm to find the minimum spanning tree.



b Add the weights of the minimum spanning tree.

$28 + 32 + 34 + 40 + 24 + 32$
$+ 40 + 36 = 266$

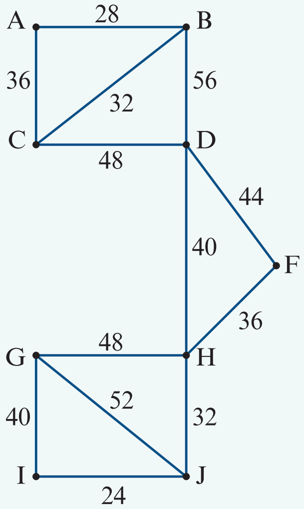
Write your answer with correct units.

The length of the minimum spanning tree is 266 km.

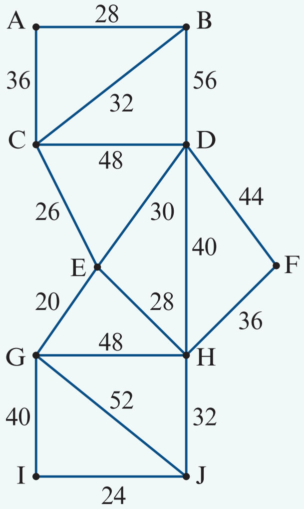
ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

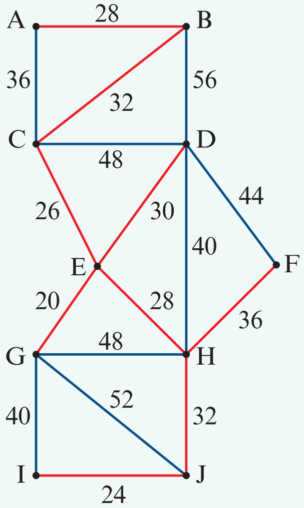
c **1** Remove edges CG and CH.



2 Re-draw the network with the addition of vertex E and its edges.



3 Use Prim's algorithm to re-draw the minimum spanning tree, with the inclusion of vertex E.



4 Describe the change in length of the minimum spanning tree in terms of an increase or decrease and by how much.	New length of minimum spanning tree $= 28 + 32 + 30 + 26 + 20 + 28 + 36 + 32 + 24$ $= 256$ km The addition of vertex E increases the minimum spanning tree by 10 km.

Exercise 9A

Trees and spanning trees

1 **a** How many edges are there in a tree with 15 vertices?

b How many vertices are there in a tree with 5 edges?

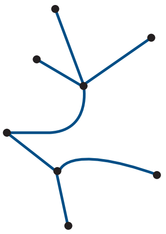
c Draw two different trees with four vertices.

d Draw three different trees with five vertices.

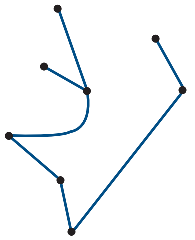
2 A connected graph has eight vertices and ten edges. Its spanning tree has ☐ vertices and ☐ edges.

3 Which of the following graphs are trees?

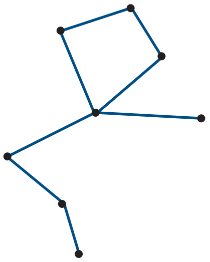
A



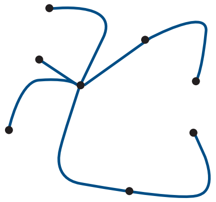
B



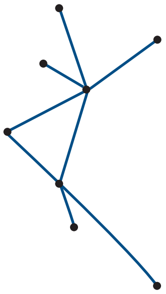
C



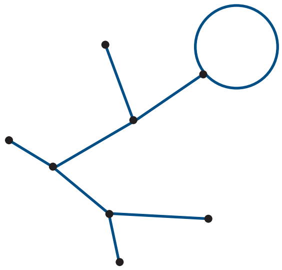
D



E



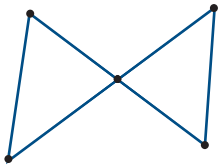
F



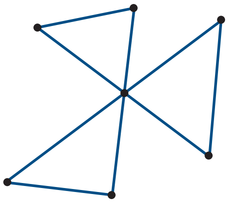
Finding a spanning tree in a network

Example 1 **4** For each of the following graphs, draw three different spanning trees.

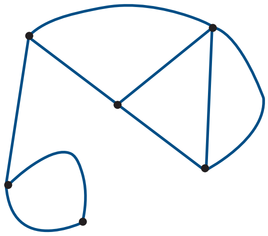
a



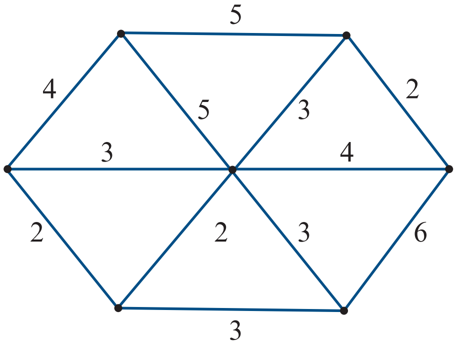
b



c



Example 2 **5** A network is shown on the right.



a How many edges must be removed in order to leave a spanning tree?

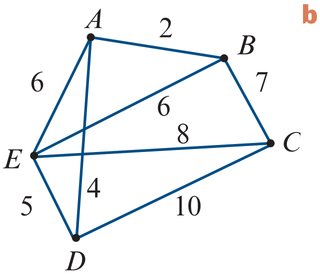
b Remove some edges to form three different trees.

c For each tree in part **b**, find the total weight.

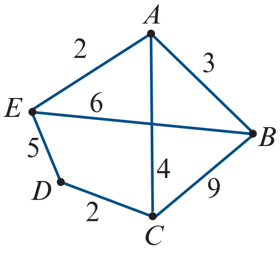
Applying Prim's algorithm

Example 3 **6** For each of the following connected graphs, use Prim's algorithm to determine the minimum spanning tree and its length.

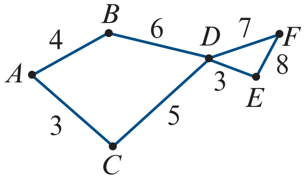
a



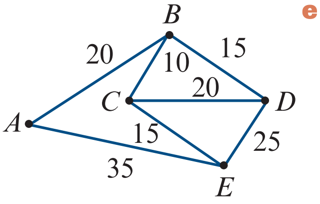
b



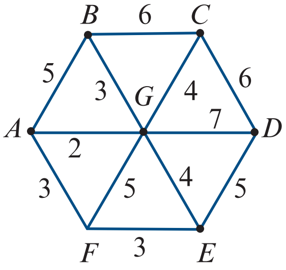
c



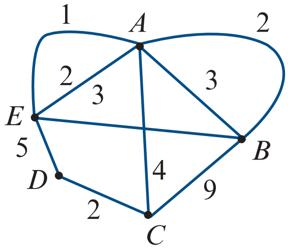
d



e



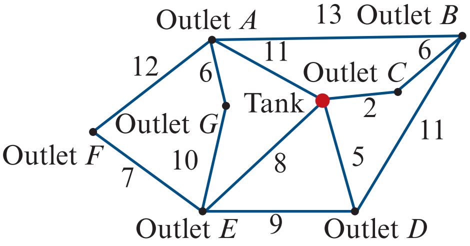
f



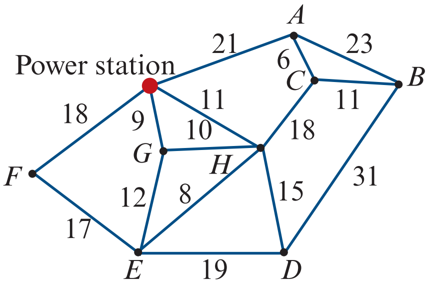
Solving connector problems

Example 4

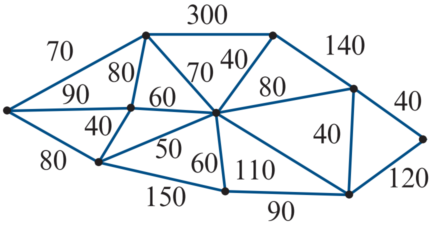
7 Water is to be piped from a water tank to seven outlets on a property. The distances (in metres) of the outlets from the tank and from each other are shown in the network opposite. Starting at the tank, determine the minimum length of pipe needed.



8 Power is to be connected by cable from a power station to eight substations (*A* to *H*). The distances (in kilometres) of the substations from the power station and from each other are shown in the network opposite. Determine the minimum length of cable needed.



9 In the network opposite, the vertices represent water tanks on a large property and the edges represent pipes used to move water between these tanks. The numbers on each edge indicate the lengths of pipes (in m) connecting different tanks.

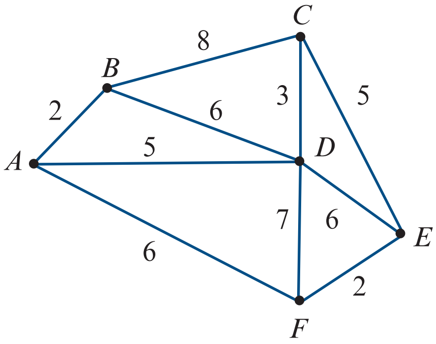


Determine the shortest length of pipe needed to connect all water storages.

9B Prim's algorithm

In the previous exercise, Prim's algorithm was used to find the minimum spanning tree for a weighted graph. Any connected weighted graph can also be represented in a table or distance matrix format, where the rows and columns show the weights between the different vertices.

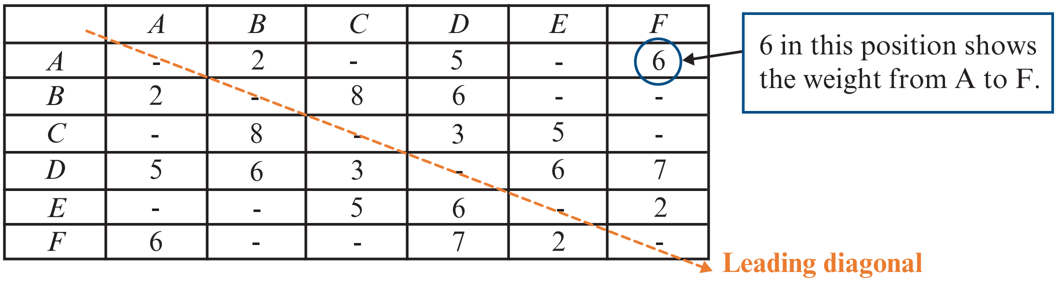
The network on the right was used to find the minimum spanning tree in Example 3.



Rather than drawing a graph, the same information can also be shown in the distance matrix at right.

	A	*B*	*C*	*D*	*E*	*F*
A	-	2	-	5	-	6
B	2	-	8	6	-	-
C	-	8	-	3	5	-
D	5	6	3	-	6	7
E	-	-	5	6	-	2
F	6	-	-	7	2	-

A '-' in the matrix indicates that there is no weight between two vertices. That is, the two vertices are not directly connected.



	A	*B*	*C*	*D*	*E*	*F*
A	-	2	-	5	-	6
B	2	-	8	6	-	-
C	-	8	-	3	5	-
D	5	6	3	-	6	7
E	-	-	5	6	-	2
F	6	-	-	7	2	-

6 in this position shows the weight from A to F.

From your understanding of matrices in Applications Units 1 & 2, this table is symmetrical either side of the leading diagonal. This indicates that the graph is non-directed. There are also no values in the leading diagonal, meaning there are no loops at any of the vertices.

If a distance matrix can be used to show the same information as displayed in a graph, then it follows that you can use the matrix to find the minimum spanning tree.

ISBN 978-1-009-25917-0

Cambridge University Press

Prim's algorithm for finding a minimum spanning tree

Prim's algorithm is a set of rules to determine a minimum spanning tree for a graph when presented as a distance matrix.

1. Choose any vertex to start. (Typically, it is easiest to start with the vertex in the first row.)
2. Cross out the row corresponding to this vertex from the table.
3. Label the column corresponding to this vertex ①.
4. Scan down the highlighted column ① and circle the lowest available value.
5. Cross out the rest of the values in the row of this lowest value and label the column corresponding to this vertex ②.
6. Scan down the values in columns ① and ② and circle the lowest available value.
7. Cross out the rest of the row, label the column and carry on in this manner until all columns are labelled.

The length of the minimum spanning tree is the sum of the numbers that have been circled. To draw the minimum spanning tree, follow the labels above each column to connect the vertices in order.

Example 5 Using Prim's algorithm to find the minimum spanning tree from an adjacency matrix

The distance, in units, between six vertices are given by the following matrix.

a Using Prim's algorithm and showing your working at each stage, find a minimum spanning tree for these six towns.

b State the length of the minimum spanning tree.

	A	*B*	*C*	*D*	*E*	*F*
A	-	2	-	5	-	6
B	2	-	8	6	-	-
C	-	8	-	3	5	-
D	5	6	3	-	6	7
E	-	-	5	6	-	2
F	6	-	-	7	2	-

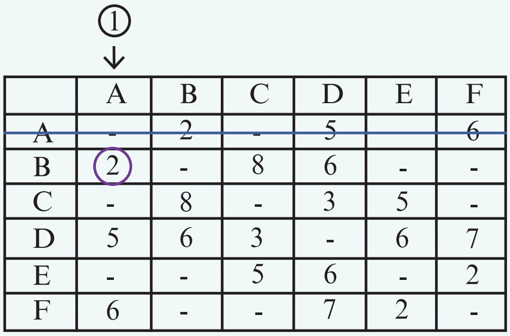
Solution

a **1** Choose a starting vertex and cross out all elements in that vertex's row and arrow its column.

	① ↓					
	A	B	C	D	E	F
~~A~~	~~-~~	~~2~~	~~-~~	~~5~~		~~6~~
B	2	-	8	6	-	-
C	-	8	-	3	5	-
D	5	6	3	-	6	7
E	-	-	5	6	-	2
F	6	-	-	7	2	-

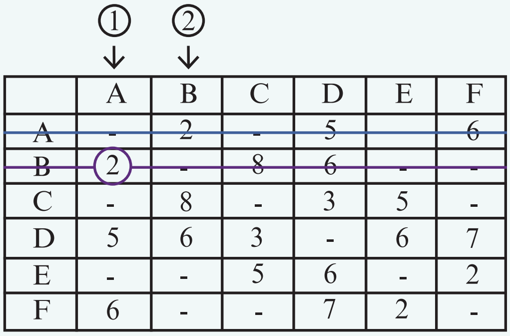
ISBN 978-1-009-25917-0

2 Neglecting all crossed out elements, scan all arrowed columns for the lowest available element and circle that element.



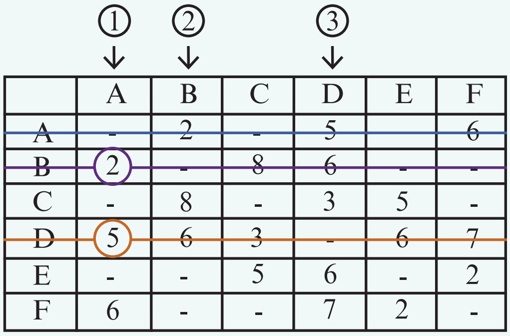
	A	B	C	D	E	F
~~A~~	~~-~~	~~2~~	~~-~~	~~5~~		~~6~~
B	2	-	8	6	-	-
C	-	8	-	3	5	-
D	5	6	3	-	6	7
E	-	-	5	6	-	2
F	6	-	-	7	2	-

3 Cross out the circled element's row and arrow its column.

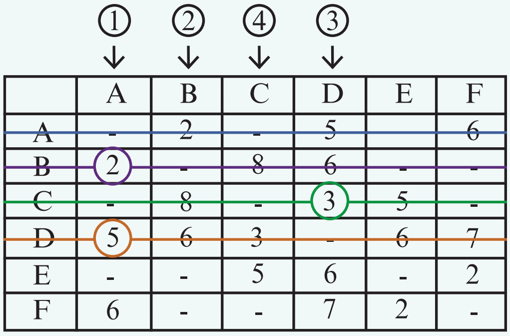


	A	B	C	D	E	F
~~A~~	~~-~~	~~2~~	~~-~~	~~5~~		~~6~~
~~B~~	~~2~~	~~-~~	~~8~~	~~6~~	~~-~~	~~-~~
C	-	8	-	3	5	-
D	5	6	3	-	6	7
E	-	-	5	6	-	2
F	6	-	-	7	2	-

4 Repeat steps 2 and 3 until all rows crossed out.



	A	B	C	D	E	F
~~A~~	~~-~~	~~2~~	~~-~~	~~5~~		~~6~~
~~B~~	~~2~~	~~-~~	~~8~~	~~6~~	~~-~~	~~-~~
C	-	8	-	3	5	-
~~D~~	~~5~~	~~6~~	~~3~~	~~-~~	~~6~~	~~7~~
E	-	-	5	6	-	2
F	6	-	-	7	2	-



	A	B	C	D	E	F
~~A~~	~~-~~	~~2~~	~~-~~	~~5~~		~~6~~
~~B~~	~~2~~	~~-~~	~~8~~	~~6~~	~~-~~	~~-~~
~~C~~	~~-~~	~~8~~	~~-~~	~~3~~	~~5~~	~~-~~
~~D~~	~~5~~	~~6~~	~~3~~	~~-~~	~~6~~	~~7~~
E	-	-	5	6	-	2
F	6	-	-	7	2	-

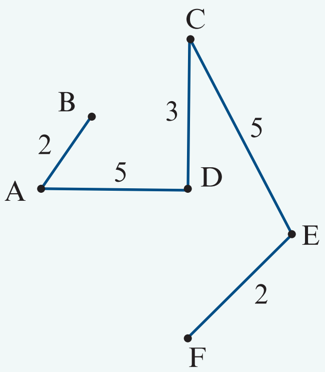
ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

	① ↓	② ↓	④ ↓	③ ↓	⑤ ↓	
	A	B	C	D	E	F
~~A~~	~~-~~	~~2~~	~~-~~	~~5~~		~~6~~
~~B~~	②	~~-~~	~~8~~	~~6~~	~~-~~	~~-~~
~~C~~	~~-~~	~~8~~	~~-~~	③	~~5~~	~~-~~
~~D~~	⑤	~~6~~	~~3~~	~~-~~	~~6~~	~~7~~
~~E~~	~~-~~	~~-~~	⑤	~~6~~	~~-~~	~~2~~
F	6	-	-	7	2	-

	① ↓	② ↓	④ ↓	③ ↓	⑤ ↓	⑥ ↓
	A	B	C	D	E	F
~~A~~	~~-~~	~~2~~	~~-~~	~~5~~		~~6~~
~~B~~	②	~~-~~	~~8~~	~~6~~	~~-~~	~~-~~
~~C~~	~~-~~	~~8~~	~~-~~	③	~~5~~	~~-~~
~~D~~	⑤	~~6~~	~~3~~	~~-~~	~~6~~	~~7~~
~~E~~	~~-~~	~~-~~	⑤	~~6~~	~~-~~	~~2~~
~~F~~	~~6~~		~~-~~	~~7~~	②	~~-~~

5 The spanning tree is formed by the circled vertices.



b The length of the minimum spanning tree is the sum of the circled elements.

Length = 2 + 5 + 3 + 5 + 2 = 17 units

Example 6 Applications of Prim's algorithm

A school wants to upgrade its internet cabling between some of its buildings. The matrix below shows the distances, in metres, between the seven buildings that the school wants to connect using the minimum cable length.

	S	*A*	*C*	*T*	*G*	*L*	*Q*
S	-	51	-	45	-	-	-
A	51	-	-	82	50	63	60
C	-	-	-	60	-	89	50
T	45	82	60	-	-	-	30
G	-	50	-	-	-	40	-
L	-	63	89	-	40	-	47
Q	-	60	50	30	-	47	-

a Apply Prim's algorithm to determine the path that connects all the school buildings while minimising the total distance and draw the minimum spanning tree.

b If expected cost for installation is $12 per metre, calculate the total cost for the school to connect all the buildings.

Solution

a Follow steps for using Prim's algorithm until all rows crossed out.

Check that there is one less element circled than number of vertices.

① ↓

	S	A	C	T	G	L	Q
~~S~~	~~-~~	~~51~~	~~-~~	~~45~~	~~-~~	~~-~~	~~-~~
A	51	-	-	82	50	63	60
C	-	-	-	60	-	89	50
T	45	82	60	-		-	30
G	-	50	-	-	-	40	-
L	-	63	89	-	40	-	47
Q	-	60	50	30	-	47	-

① ↓ ② ↓

	S	A	C	T	G	L	Q
~~S~~	~~-~~	~~51~~	~~-~~	~~45~~	~~-~~	~~-~~	~~-~~
A	51	-	-	82	50	63	60
C	-	-	-	60	-	89	50
~~T~~	~~45~~	~~82~~	~~60~~	~~-~~		~~-~~	~~30~~
G	-	50	-	-	-	40	-
L	-	63	89	-	40	-	47
Q	-	60	50	30	-	47	-

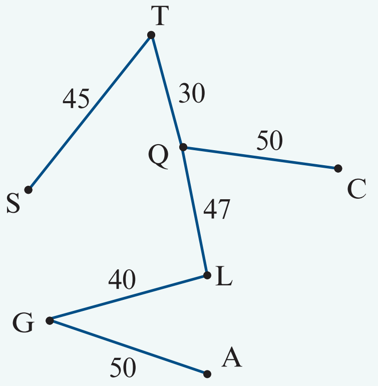
	① ↓			② ↓			③ ↓
	S	A	C	T	G	L	Q
~~S~~	~~-~~	~~51~~	~~-~~	~~45~~	~~-~~	~~-~~	~~-~~
A	51	-	-	82	50	63	60
C	-	-	-	60	-	89	50
~~T~~	~~45~~	~~82~~	~~60~~	~~-~~		~~-~~	~~30~~
G	-	50	-	-	-	40	-
L	-	63	89	-	40	-	47
~~Q~~	~~-~~	~~60~~	~~50~~	~~30~~	~~-~~	~~47~~	~~-~~

	① ↓			② ↓		④ ↓	③ ↓
	S	A	C	T	G	L	Q
~~S~~	~~-~~	~~51~~	~~-~~	~~45~~	~~-~~	~~-~~	~~-~~
A	51	-	-	82	50	63	60
C	-	-	-	60	-	89	50
~~T~~	~~45~~	~~82~~	~~60~~	~~-~~		~~-~~	~~30~~
G	-	50	-	-	-	40	-
~~L~~	~~-~~	~~63~~	~~89~~	~~-~~	~~40~~	~~-~~	~~47~~
~~Q~~	~~-~~	~~60~~	~~50~~	~~30~~	~~-~~	~~47~~	~~-~~

	① ↓			② ↓	⑤ ↓	④ ↓	③ ↓
	S	A	C	T	G	L	Q
~~S~~	~~-~~	~~51~~	~~-~~	~~45~~	~~-~~	~~-~~	~~-~~
A	51	-	-	82	50	63	60
C	-	-	-	60	-	89	50
~~T~~	~~45~~	~~82~~	~~60~~	~~-~~		~~-~~	~~30~~
~~G~~	~~-~~	~~50~~	~~-~~	~~-~~	~~-~~	~~40~~	~~-~~
~~L~~	~~-~~	~~63~~	~~89~~	~~-~~	~~40~~	~~-~~	~~47~~
~~Q~~	~~-~~	~~60~~	~~50~~	~~30~~	~~-~~	~~47~~	~~-~~

	① ↓	⑥ ↓	⑦ ↓	② ↓	⑤ ↓	④ ↓	③ ↓
	S	A	C	T	G	L	Q
~~S~~	~~-~~	~~51~~	~~-~~	~~45~~	~~-~~	~~-~~	~~-~~
~~A~~	~~51~~	~~-~~	~~-~~	~~82~~	㊿	~~63~~	~~60~~
~~C~~	~~-~~	~~-~~	~~-~~	~~60~~	~~-~~	~~89~~	㊿
~~T~~	㊺	~~82~~	~~60~~	~~-~~		~~-~~	~~30~~
~~G~~	~~-~~	~~50~~	~~-~~	~~-~~	~~-~~	㊵	~~-~~
~~L~~	~~-~~	~~63~~	~~89~~	~~-~~	~~40~~	~~-~~	㊼
~~Q~~	~~-~~	~~60~~	~~50~~	㉚	~~-~~	~~47~~	~~-~~

Draw the minimum spanning tree. There are many ways to draw the tree. Check that the correct vertices are connected.



b Calculate length of minimum spanning tree.

Total cost = length of tree × cost per metre

Length of minimum spanning tree

$= 45 + 30 + 50 + 47 + 40 + 50 = 262$ m

Cost $= 262 \times \$12 = \$3144$

Exercise 9B

Basic use of Prim's algorithm

1 Use Prim's algorithm to determine the minimum spanning tree and its length for each of the distance matrices below.

a

	A	*B*	*C*	*D*
A	-	11	10	9
B	11	-	9	6
C	10	9	-	5
D	9	6	5	-

b

	R	*S*	*T*	*V*	*W*
R	-	2	3	7	6
S	2	-	8	1	10
T	3	8	-	4	9
V	7	1	4	-	5
W	6	10	9	5	-

ISBN 978-1-009-25917-0

c

	A	*B*	*C*	*D*	*E*
A	-	2	10	5	7
B	2	-	13	8	10
C	10	13	-	12	14
D	5	8	12	-	15
E	7	10	14	15	-

d

	E	*F*	*G*	*H*	*J*
E	-	19	-	20	-
F	19	-	13	12	-
G	-	13	-	15	14
H	20	12	15	-	12
J	-	-	14	12	-

e

	K	*L*	*M*	*N*	*P*	*Q*
K	-	12	43	68	73	75
L	12	-	33	53	61	62
M	43	33	-	36	26	31
N	68	53	36	-	27	19
P	73	61	26	27	-	6
Q	75	62	31	19	6	-

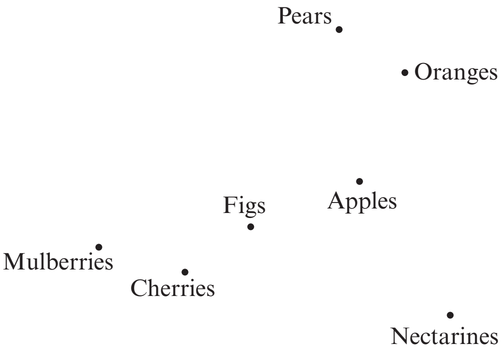
Applications of Prim's algorithm

2 A farmer is looking to upgrade their irrigation system for their orchard. The distances, in metres, between their different crops are shown in the matrix below.

	Mulberries	Cherries	Figs	Apples	Oranges	Pears	Nectarines
Mulberries	-	15	29	47	60	51	63
Cherries	15	-	14	33	47	40	48
Figs	29	14	-	19	32	30	35
Apples	47	33	19	-	16	18	22
Oranges	60	47	32	16	-	13	29
Pears	51	40	30	18	13	-	38
Nectarines	63	48	35	22	29	38	-

a Using Prim's algorithm and showing your working at each stage, find the minimum length of irrigation needed to ensure all these crops are connected to the system.

b Draw the minimum spanning tree on the layout of the orchard at the right.



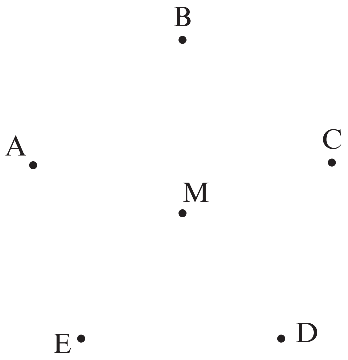
c If irrigation costs $9.50 per metre, calculate the total cost to upgrade the system.

3 An electrician has the task of completing a circuit which connects the meter (M) and the lights (A, B, C, D and E) of a house. Not all of these points are able to be connected directly – the length, in metres, of the possible direct connections are given in the table below.

	M	*A*	*B*	*C*	*D*	*E*
M	-	6	7.5	10.5	-	3
A	6	-	5	12	4.5	6
B	7.5	5	-	9	6	4.5
C	10.5	12	9	-	7	7.5
D	-	4.5	6	7	-	4
E	3	6	4.5	7.5	4	-

a What is the minimum length of electrical cable needed for the circuit?

b Draw the tree which will complete the circuit using the least amount of wiring.



4 A zoo needs to connect each of their enclosures with walking paths.

	Entrance	Savannah	Nocturnal House	Rainforest	Bushwalk	Main Lawn	Wetlands
Entrance	-	10	-	-	-	6.5	5
Savannah	10	-	5	5	-	6	-
Nocturnal House	-	5	-	9.5	-	-	-
Rainforest	-	5	9.5	-	4.5	2.5	-
Bushwalk	-	-	-	4.5	-	-	3
Main Lawn	6.5	6	-	2.5	4.5	-	4
Wetlands	5	-	-	-	3	4	-

The distances between the enclosures are measured in centimetres, based on the map, with a scale of 1:2500, given to the patrons upon arrival. Not all the enclosures have direct walking paths between them.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

a Using the fact that there are seven enclosures, determine the minimum number of paths needed to ensure that all enclosures are connected by a path.

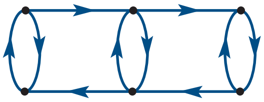
b Determine the length, in metres, of the minimum spanning tree for the zoo such that patrons can access all enclosures by a walking path.

c Due to the amount of foot traffic, the zoo administration has decided that the Entrance and Main Lawn needs to have a walking path between them. Given that this path must be included, how does this affect the minimum spanning tree previously found in part **b**?

9C Bipartite graphs and assignment problems

Directed graphs

In a previous chapter, graphs were used to represent connections between people, places or objects. The vertices of a graph represented objects, such as towns, and edges represented the connections between them, such as roads. Weighted graphs included extra numerical information about the connections, such as distance, time or cost. When a graph has this numerical information we call it a network.



A **directed graph**, or **digraph**, records directional information on networks using arrows on the edges. The network on the right shows roads around a city. The vertices are the intersections of the roads and the edges are the possible road connections between the intersections. The arrows show that some of the roads only allow traffic in one direction, while others allow traffic in both directions. The edges of a directed graph are sometimes called **arcs**.

Adjacency matrices for directed graphs

An adjacency matrix can be drawn to show the possible directed connections between vertices in a graph. Unlike the adjacency matrices that you have drawn previously, adjacency matrices for directed graphs are not necessarily symmetric.

The matrix labels for the *rows* of the matrix are the *origin* vertices. The *column* labels in the matrix are the *destination* vertices. The direction of travel is from a row vertex to a column vertex.

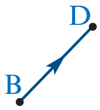
$$\begin{array}{rc} & \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{array}{r} A \\ B \\ \text{origin } C \\ D \\ E \end{array} & \begin{bmatrix} - & - & - & - & - \\ - & - & - & 1 & - \\ - & - & - & - & - \\ - & 0 & - & - & - \\ - & - & - & - & - \end{bmatrix} \end{array}$$

ISBN 978-1-009-25917-0

In the graph for this adjacency matrix:

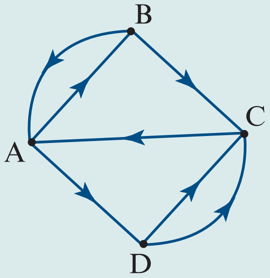
- there is one arc from vertex *B* to vertex *D*, shown by a ‘1’ in row *B*, column *D*
- there is no arc from vertex *D* back to vertex *B*, shown by a ‘0’ in row *D*, column *B*.

This means the arc between vertex B and vertex D has a directional arrow from B to D only.



Example 7 Writing an adjacency matrix for a directed graph

Write the adjacency matrix for the directed graph shown on the right.



Solution

1 Set up a matrix that has the vertices listed as row and column headings. There are 4 vertices, so the matrix will have order 4 × 4.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \left[\begin{array}{cccc} \; & \; & \; & \; \\ \; & \; & \; & \; \\ \; & \; & \; & \; \\ \; & \; & \; & \; \end{array}\right] \end{array}$$

2 Count the number of arcs that allow travel from each vertex to each other vertex. Take careful note of the direction of the arc.

There is:

- one arc from *A* to *B*
- one arc from *A* to *D*
- one arc from *B* to *A*
- one arc from *B* to *C*
- one arc from *C* to *A*
- two arcs from *D* to *C*.

Write these numbers in the matrix.

$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \left[\begin{array}{cccc} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \end{array}\right] \end{array}$$

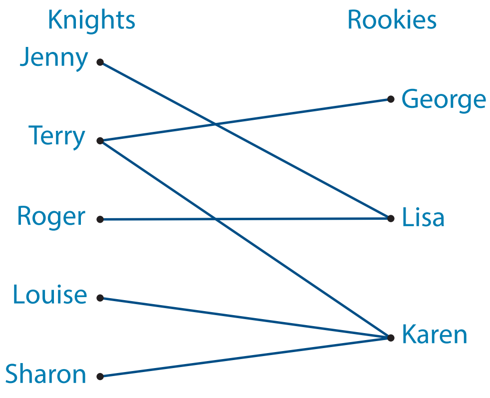
ISBN 978-1-009-25917-0

Cambridge University Press

Updated January 2024

Bipartite graphs

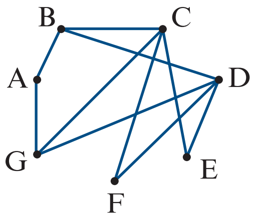
In some situations, the vertices of a graph belong in two separate sets. Consider the members of two different chess teams. The Knights team has five players and the Rookies team has three players. The members of the teams are represented by a vertex, arranged vertically underneath the team name as shown in the diagram opposite. The edges of the diagram connect team members that have played a chess match.



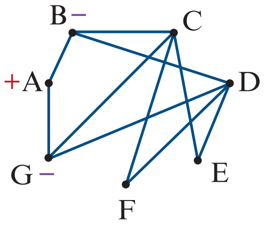
This type of graph is called a **bipartite graph**. Each edge in a bipartite graph joins one vertex from one group to a vertex in the other group.

Identifying bipartite graphs

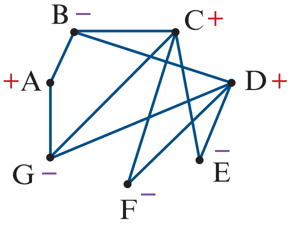
Sometimes bipartite graphs may be drawn such that the vertices are not arranged into two distinct groups. This can make it difficult to discern if the graph is bipartite or not.



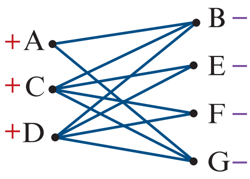
Since the primary feature of a bipartite graph is that each edge of the graph joins a vertex in the first group to a vertex in the second group, alternate colouring or labelling can be used to see if the vertices are split into two distinct groups. For example, if we allocated vertex A as belonging to group '+' then the two vertices B and G must be in the opposite group (–) to A, since A is joined to both B and G.



This process of assigning adjacent vertices with alternating labels continues until all vertices in the graph are assigned a label. During this process, if two adjacent vertices – vertices that are joined by an edge – have the same label, then the graph is *not bipartite*.



The graph can therefore be re-drawn as:



Example 8 Constructing and analysing a bipartite graph

Angie, Bev, Charlie, Dorian and Evelyn are presenters on a TV travel show. Each presenter will be assigned to film a story about one country that they have visited before.

- Angie has visited Senegal.
- Bev has visited Vietnam and Peru.
- Charlie has visited Ghana and Peru.
- Dorian has visited Senegal and Myanmar.
- Evelyn has visited Vietnam and Myanmar.

Construct a bipartite graph of the information above and use it to decide on the assignment of each presenter to a country.

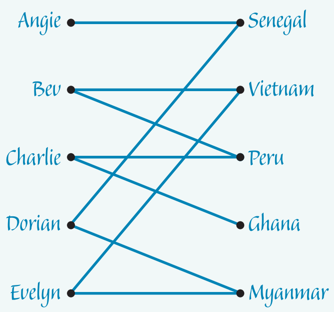
Solution

1 The two groups of item are: Presenters and Countries. Draw a vertex for each presenter in one column and each country in another.

2 Angie has visited Senegal so could be sent there to film her story. Join the vertices for Angie and Senegal with an edge.

3 Bev has visited Vietnam and Peru so join the vertex for Bev to the vertices for Vietnam and Peru.

4 Join the other presenter vertices to country vertices in a similar way.



5 Definite assignments are shown in red, impossible assignments are shown with dotted lines in the diagram.

a Angie is only connected to Senegal and so must visit this country. If Angie visits Senegal, Dorian cannot.

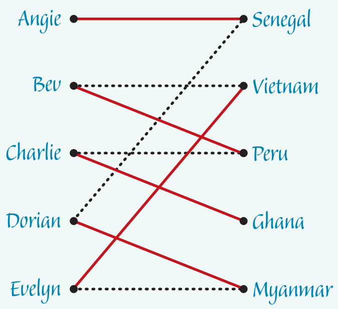
b If Dorian cannot visit Senegal, he must visit Myanmar.

c If Dorian visits Myanmar, then Evelyn cannot.

d If Evelyn cannot visit Myanmar, she must visit Vietnam.

e If Evelyn must visit Vietnam, Bev cannot and so she must visit Peru.

f If Bev must visit Peru, then Charlie cannot. Charlie must visit Ghana.



6 Write the assignments.

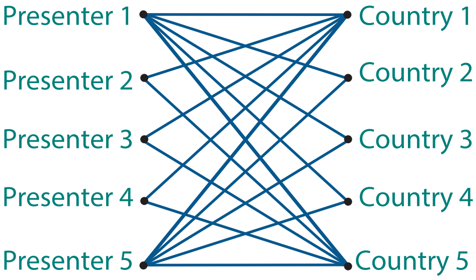
Angie will visit Senegal.
Bev will visit Peru.
Charlie will visit Ghana.
Dorian will visit Myanmar.
Evelyn will visit Vietnam.

ISBN 978-1-009-25917-0

Complete bipartite graphs

The **allocation** of a presenter to a country in the example above would be greatly simplified if every presenter could visit all of the countries. There would be many different allocations of presenter to country and if all of the possible allocations were represented in a bipartite graph, then that graph would be a complete bipartite graph because all vertices for the presenters would be connected to each vertex for the countries.

Such a graph is very complex.



Rather than just assign the presenters randomly, the producers could use information about the presenters preferences, or perhaps the number of times they have been to each of the countries to make the assignments with priority. This information would be weights on the edges of an already very complex diagram.

Rather than write the weights on a bipartite graph, we can write them in a table instead.

Cost matrix

The table of weights for a bipartite graph is called a **cost matrix**. Even though it is called a cost matrix, the 'cost' does not have to be in terms of money only. It could be the time it takes people to complete a task, or the distance that people need to travel.

As an example, a factory might need to assign each of four employees to one of four machines.

Employee	*A*	*B*	*C*	*D*
Wendy	30	40	50	60
Xenefon	70	30	40	70
Yolanda	60	50	60	30
Zelda	20	80	50	70

The cost matrix on the right shows the time each employee takes to complete the task on each machine. The cost matrix can be used to determine the best assignment of employee to machine so that the overall time taken to complete the tasks is minimised.

The **Hungarian algorithm** is used to do this.

ISBN 978-1-009-25917-0

The Hungarian algorithm

Performing the Hungarian algorithm

Since we are using a matrix to perform the Hungarian algorithm, we start by removing the row and column labels from the table and writing the costs as elements in a matrix (with square brackets).

Employee	*A*	*B*	*C*	*D*
Wendy	30	40	50	60
Xenefon	70	30	40	70
Yolanda	60	50	60	30
Zelda	20	80	50	70

$$\rightarrow \begin{bmatrix} 30 & 40 & 50 & 60 \\ 70 & 30 & 40 & 70 \\ 60 & 50 & 60 & 30 \\ 20 & 80 & 50 & 70 \end{bmatrix}$$

Step 1: Subtract the lowest value in each row, from every value in that row.

- 30 has been subtracted from every value in the first row
- 30 has been subtracted from every value in the second row
- 30 has been subtracted from every value in the third row
- 20 has been subtracted from every value in the fourth row

$$\begin{bmatrix} 0 & 10 & 20 & 30 \\ 40 & 0 & 10 & 40 \\ 30 & 20 & 30 & 0 \\ 0 & 60 & 30 & 50 \end{bmatrix}$$

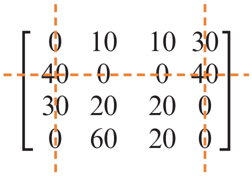
Step 2: If a column does not contain a zero, subtract the lowest value in the column from every value in that column.

- The third column does not have a zero.
- 10 has been subtracted from every value in the third column.

$$\begin{bmatrix} 0 & 10 & 10 & 30 \\ 40 & 0 & 0 & 40 \\ 30 & 20 & 20 & 0 \\ 0 & 60 & 20 & 50 \end{bmatrix}$$

Step 3: If the minimum number of lines required to cover all the zeroes in the table is equal to the number of allocations to be made, go to step 5. Otherwise, continue to step 4a.

- The zeroes can be covered with three lines. This is less than the number of allocations to be made (4).
- Continue to step 4.



$$\begin{bmatrix} 0 & 10 & 10 & 30 \\ 40 & 0 & 0 & 40 \\ 30 & 20 & 20 & 0 \\ 0 & 60 & 20 & 0 \end{bmatrix}$$

ISBN 978-1-009-25917-0

Cambridge University Press

Updated January 2024

Step 4a: Add the smallest uncovered value to any value that is covered by two lines. Subtract the smallest uncovered value from all the uncovered values.

- The smallest uncovered element is 10.
- 10 has been *added* to 40 (row 2, column 1) and 40 (row 2, column 4) because these values are covered by two lines.
- 10 has been *subtracted* from all the uncovered values.

$$\begin{bmatrix} 0 & 0 & 0 & 30 \\ 50 & 0 & 0 & 50 \\ 30 & 10 & 10 & 0 \\ 0 & 50 & 10 & 50 \end{bmatrix}$$

Step 4b: Repeat from step 3.

- The zeroes can be covered with a minimum of four lines. This is the same as the number of allocations to make.
- Continue to step 5.

$$\begin{bmatrix} 0 & 0 & 0 & 30 \\ 50 & 0 & 0 & 50 \\ 30 & 10 & 10 & 0 \\ 0 & 50 & 10 & 50 \end{bmatrix}$$

Step 5: Choose a set of zeroes such that each row and column only has one zero selected.

- There is one zero in the third row, so we must highlight this zero.
- There is one zero in the fourth row, so we must highlight this zero.
- There are two zeroes in the first column, but since we have already highlighted the zero in the fourth row, we cannot also highlight the zero in the first row. We can therefore highlight either zero in the second or third column in the first row.
- Based on the zero highlighted in the first row, we are left with one remaining zero in the second row.

$$\begin{matrix} 0 & \boxed{0} & 0 & 30 \\ 50 & 0 & \boxed{0} & 50 \\ 30 & 10 & 10 & \boxed{0} \\ \boxed{0} & 50 & 10 & 50 \end{matrix}$$

or

$$\begin{matrix} 0 & 0 & \boxed{0} & 30 \\ 50 & \boxed{0} & 0 & 50 \\ 30 & 10 & 10 & \boxed{0} \\ \boxed{0} & 50 & 10 & 50 \end{matrix}$$

Step 6: Make the allocation and calculate minimum cost.

- The zeroes in the final matrix correspond to the ideal allocation in the original matrix.
- Match the highlighted zeroes to the entries in the original table.
- Read off the ideal allocation and calculate the cost.

Employee	*A*	*B*	*C*	*D*
Wendy	30	$\boxed{40}$	50	60
Xenefon	70	30	$\boxed{40}$	70
Yolanda	60	50	60	$\boxed{30}$
Zelda	$\boxed{20}$	80	50	70

or

Employee	*A*	*B*	*C*	*D*
Wendy	30	40	$\boxed{50}$	60
Xenefon	70	$\boxed{30}$	40	70
Yolanda	60	50	60	$\boxed{30}$
Zelda	$\boxed{20}$	80	50	70

Allocation:

- Zelda must operate machine A (20 minutes).
- Yolanda must operate machine D (30 minutes).
- Wendy can operate either machine B (40 minutes) or C (50 minutes).
- Xenefon can operate either machine C (40 minutes) or B (30 minutes).

Minimum time taken to finish the work = 20 + 30 + 40 + 40 = 130 minutes,
or 20 + 30 + 50 + 30 = 130 minutes.

Note: Because Wendy and Xenefon can operate either B or C, there are two possible allocations. Both allocations will have the same minimum cost.

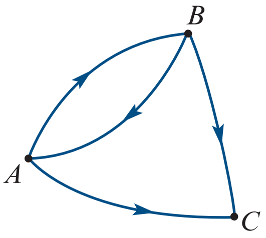
Exercise 9C

Adjacency matrices for directed graphs

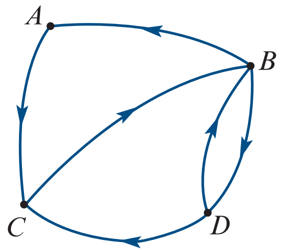
Example 7

1 Write an adjacency matrix for the following directed graphs.

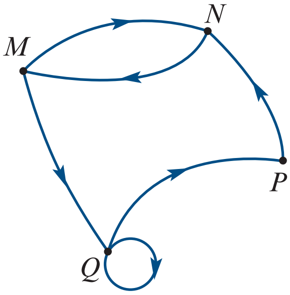
a



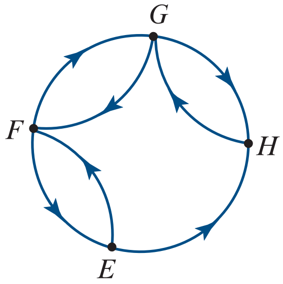
b



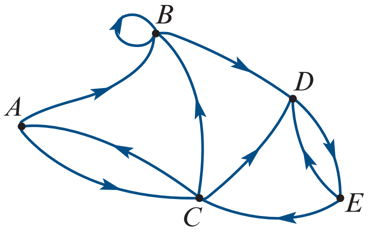
c



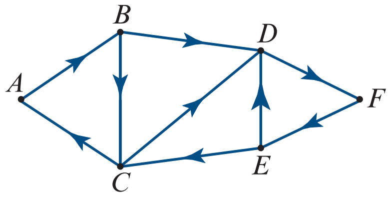
d



e



f



2 Draw a directed graph for the following adjacency matrices.

a

$$\begin{array}{c|ccc} & A & B & C \\ A & 0 & 1 & 1 \\ B & 0 & 1 & 0 \\ C & 1 & 1 & 1 \end{array}$$

b

$$\begin{array}{c|ccc} & S & T & U \\ S & 1 & 0 & 1 \\ T & 0 & 0 & 2 \\ U & 0 & 1 & 0 \end{array}$$

c

$$\begin{array}{c|cccc} & P & Q & R & S \\ P & 0 & 1 & 1 & 1 \\ Q & 1 & 0 & 2 & 0 \\ R & 1 & 0 & 0 & 1 \\ S & 0 & 0 & 1 & 0 \end{array}$$

d

$$\begin{array}{c|cccc} & P & Q & R & S \\ P & 1 & 1 & 0 & 1 \\ Q & 0 & 0 & 1 & 1 \\ R & 1 & 1 & 0 & 1 \\ S & 1 & 0 & 0 & 0 \end{array}$$

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Construction and analysing bipartite graphs

Example 8

3 Gloria, Minh, Carlos and Trevor are buying ice-cream. They have a choice of five flavours: chocolate, vanilla, peppermint, butterscotch and strawberry. Gloria likes vanilla and butterscotch, but not the others. Minh only likes strawberry. Carlos likes chocolate, peppermint and butterscotch. Trevor likes all flavours.

a Explain why a bipartite graph can be used to display this information.

b Draw an appropriate bipartite graph to show the people and the flavours of ice cream they like.

c What is the degree of the vertex representing Trevor?

The ice-cream shop has no butterscotch ice-cream available.

Gloria, Minh, Carlos and Trevor can have only one ice-cream each and must all have a different flavour.

d Who must have the vanilla ice-cream?

e If Carlos chooses peppermint, write down which ice-cream flavour the other people must have.

4 Joni, Ian, Dylan and Joshua are teachers in a school. The school has a maths class, an English class, a geography class and a science class that need a teacher. Each teacher can be allocated one class only.

Joni can teach English or geography.

Ian can teach maths or science.

Dylan can teach English or geography.

Joshua can teach geography or science.

a Draw a bipartite graph to show the teachers and the subject that they can teach.

b Explain why Joshua must take the science class.

c Write two different allocations of teachers to subjects.

5 The table on the right shows five people in the rows and five sports in the columns. A tick (✓) in a table cell indicates that the person in that row can coach the sport in that column. A cross (×) indicates that they cannot coach that sport.

	Hockey	*Cricket*	*Soccer*	*Rugby*	*Squash*
Rob	✓	×	✓	×	×
Janet	×	✓	✓	✓	✓
Tara	×	✓	×	✓	×
Diana	✓	×	×	✓	×
Jason	×	×	×	✓	×

Each sport in the table must be coached by only one of the people in the table.

a Draw a bipartite graph to represent the information contained in the table above.

b Explain why Diana must coach hockey.

c Write the allocation of people to sports.

Performing the Hungarian algorithm

6 **a** A cost matrix is shown. Find the allocation(s) by the Hungarian algorithm that will give the minimum cost.

	A	*B*	*C*	*D*
W	110	95	140	80
X	105	82	145	80
Y	125	78	140	75
Z	115	90	135	85

b Find the minimum cost for the given cost matrix and give a possible allocation.

	A	*B*	*C*	*D*
W	2	4	3	5
X	3	5	3	4
Y	2	3	4	2
Z	2	4	2	3

7 A school is to enter four students in four track events: 100 m, 400 m, 800 m and 1500 m. The four students' times (in seconds) are given in the table. The rules permit each student to enter only one event. The aim is to obtain the minimum total time.

Student	*100 m*	*400 m*	*800 m*	*1500 m*
Dimitri	11	62	144	379
John	13	60	146	359
Carol	12	61	149	369
Elizabeth	13	63	142	349

Use the Hungarian algorithm to select the 'best' student for each event.

8 Three volunteer workers, Joe, Meg and Ali, are available to help with three jobs. The time (in minutes) in which each worker is able to complete each task is given in the table opposite. Which allocation of workers to jobs will enable the jobs to be completed in the minimum time?

	Job		
Student	*A*	*B*	*C*
Joe	20	20	36
Meg	16	20	44
Ali	26	26	44

9 A company has four machine operators and four different machines that they can operate. The table shows the hourly cost in dollars of running each machine for each operator. How should the machinists be allocated to the machines to minimise the hourly cost from each of the machines with the staff available?

	Machine			
Operator	*W*	*X*	*Y*	*Z*
A	38	35	26	54
B	32	29	32	26
C	44	26	23	35
D	20	26	32	29

ISBN 978-1-009-25917-0

10 A football association is scheduling football games to be played by three teams (the Champs, the Stars and the Wests) on a public holiday. On this day, one team must play at their home ground, one will play away and one will play at a neutral ground.

The costs (in $’000s) for each team to play at each of the grounds are given in the table below.

Determine a schedule that will minimise the total cost of playing the three games and determine this cost.

Note: There are two different ways of scheduling the games to achieve the same minimum cost. Identify both of these.

Team	*Home*	*Away*	*Neutral*
Champs	10	9	8
Stars	7	4	5
Wests	8	7	6

11 A roadside vehicle assistance organisation has four service vehicles located in four different places. The table shows the distance (in kilometres) of each of these service vehicles from four motorists in need of roadside assistance.

	Motorist			
Service vehicle	*Jess*	*Mark*	*Raj*	*Karla*
A	18	15	15	16
B	7	17	11	13
C	25	19	18	21
D	9	22	19	23

Determine a service vehicle assignment that will ensure that the total distance travelled by the service vehicles is minimised. Determine this distance.

9D Applications of the Hungarian algorithm

In the previous exercise the Hungarian algorithm was used to solve allocation problems involving a square matrix. For example, assigning 4 employees to one of 4 machines. In this situation there was no employee left without a task. However, in some problems there may not be enough machines to assign to each employee, or too fewer employees for each machine.

When this occurs, our matrix will not be a square and so a **dummy row or column** must be added to convert the cost matrix into a square matrix, before proceeding with the Hungarian algorithm. A dummy row or column has all elements of 0.

Using the Hungarian algorithm for non-square arrays

As an example, a factory might need to assign four employees to one of three machines.

Employee	*D*	*E*	*F*
Wendy	30	50	60
Xenefon	70	40	70
Yolanda	60	60	30
Zelda	20	50	70

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Step 0: Add a dummy column where all the elements are 0 to make the matrix square.

- Matrix must be 4 × 4 since there are four employees to allocate.
- Continue to Step 1 of *Performing the Hungarian algorithm*

$$\begin{bmatrix} 30 & 50 & 60 & 0 \\ 70 & 40 & 70 & 0 \\ 60 & 60 & 30 & 0 \\ 20 & 50 & 70 & 0 \end{bmatrix}$$

Step 1: Subtract the lowest value in each row, from every value in that row.

- There is a zero in each row, so there is no need to do any row subtraction.

$$\begin{bmatrix} 30 & 50 & 60 & 0 \\ 70 & 40 & 70 & 0 \\ 60 & 60 & 30 & 0 \\ 20 & 50 & 70 & 0 \end{bmatrix}$$

Step 2: If a column does not contain a zero, subtract the lowest value in that column from every value in that column.

- Subtract 20 from the first column
- Subtract 40 from the second column
- Subtract 30 from the third column
- No column subtraction needed for the fourth column (all values are zero).

$$\begin{bmatrix} 10 & 10 & 30 & 0 \\ 50 & 0 & 40 & 0 \\ 40 & 20 & 0 & 0 \\ 0 & 10 & 40 & 0 \end{bmatrix}$$

Step 3: If the minimum number of lines required to cover all the zeroes in the table is equal to the number of allocations to be made, go to step 5. Otherwise, continue to step 4a.

- Zeroes can be covered with four lines. This is equal to the number of allocations to be made.
- Continue to step 5.

$$\begin{bmatrix} 10 & 10 & 30 & 0 \\ 50 & 0 & 40 & 0 \\ 40 & 20 & 0 & 0 \\ 0 & 10 & 40 & 0 \end{bmatrix}$$

Step 5: Choose a set of zeroes such that each row and column only has one zero selected.

- There is one zero in the first column.
- There is one zero in the second column.
- There is one zero in the third column.
- The final zero must be in the first row of the fourth column, so that each row and column only has one zero highlighted.

$$\begin{bmatrix} 10 & 10 & 30 & \boxed{0} \\ 50 & \boxed{0} & 40 & 0 \\ 40 & 20 & \boxed{0} & 0 \\ \boxed{0} & 10 & 40 & 0 \end{bmatrix}$$

Step 6: Make the allocation and calculate minimum cost.

- Match the highlighted zeroes to the entries in the original table and read off the ideal allocation.

Allocation:

- Zelda operates Machine D.
- Xenefon operates Machine E.
- Yolanda operates Machine F.
- Wendy does not have an assignment because the fourth column was a "dummy column" with no value.

Employee	*D*	*E*	*F*
Wendy	30	50	60
Xenefon	70	$\boxed{40}$	70
Yolanda	60	60	$\boxed{30}$
Zelda	$\boxed{20}$	50	70

Minimum time to finish work
= 20 + 40 + 30 = 90 minutes.

Using the Hungarian algorithm to maximise allocations

The Hungarian algorithm has been used to minimise the cost, however, it can also be used to determine a suitable allocation that will **maximise** the overall outcome. This might be useful when allocating workers to tasks, where companies may want to maximise productivity or profit.

To solve a maximisation problem using the Hungarian algorithm, *subtract all values in the matrix from the largest overall number in the matrix*. Then apply the standard Hungarian algorithm as outlined in 9C.

Example 9 Using the Hungarian algorithm to maximise allocations

A corporate office wants to assign three managers to three different store locations. The sales of the previous month for each of the managers at each of the locations are shown in the table.

	Store 1	*Store 2*	*Store 3*
Manager A	12	17	11
Manager B	11	20	7
Manager C	8	16	5

Given that each manager can only be assigned to one of the stores, determine the best allocation to maximise sales.

Solution

1 Subtract all values in the matrix from the largest overall number.
The largest value is 20, so each value is subtracted from 20.

$$\begin{bmatrix} 12 & 17 & 11 \\ 11 & 20 & 7 \\ 8 & 16 & 5 \end{bmatrix}$$

$$\begin{bmatrix} 8 & 3 & 9 \\ 9 & 0 & 13 \\ 12 & 4 & 15 \end{bmatrix}$$

ISBN 978-1-009-25917-0

2 Complete row subtraction
Row 1 – subtract 3
Row 2 – subtract 0
Row 3 – subtract 4

$$\begin{bmatrix} 5 & 0 & 6 \\ 9 & 0 & 13 \\ 8 & 0 & 11 \end{bmatrix}$$

3 Complete column subtraction
Column 1 – subtract 5
Column 2 – subtract 0
Column 3 – subtract 6

$$\begin{bmatrix} 0 & 0 & 0 \\ 4 & 0 & 7 \\ 3 & 0 & 5 \end{bmatrix}$$

4 **a** Cover zeroes with lines. If minimum number of lines not equal to number of allocations, then go to next step. Otherwise allocate.

$$\begin{bmatrix} 0 & 0 & 0 \\ 4 & 0 & 7 \\ 3 & 0 & 5 \end{bmatrix}$$

b Subtract the lowest uncovered value from all uncovered value and add to any values covered by 2 lines.

$$\begin{bmatrix} 0 & 3 & 0 \\ 1 & 0 & 4 \\ 0 & 0 & 2 \end{bmatrix}$$

5 Repeat step 4a.

$$\begin{bmatrix} 0 & 3 & 0 \\ 1 & 0 & 4 \\ 0 & 0 & 2 \end{bmatrix}$$

6 Allocate based on zeroes

$$\begin{bmatrix} 0 & 3 & \boxed{0} \\ 1 & \boxed{0} & 4 \\ \boxed{0} & 0 & 2 \end{bmatrix}$$

7 Write allocations

Store 1 – Manager C
Store 2 – Manager B
Store 3 – Manager A

Exercise 9D

Non-square arrays

1 A local council has three tasks to be performed by four subcontractors: A, B, C and D. The estimates of time, in hours, each subcontractor would take to perform each task is given in the table. Use the Hungarian algorithm to determine how each of the tasks (one to each subcontractor) is to be allocated to minimise the total number of hours.

		I	II	III
Subcontractors	A	12	29	18
	B	16	30	9
	C	38	23	18
	D	21	33	23

ISBN 978-1-009-25917-0

2 The coach of a school triathlon team needs to assign athletes to each leg of the race. The four best athletes and their average times (in minutes) they have achieved in each of the legs are shown below.

	Jasmine	Kim	Lauren	Melissa
Swim	26	19	23	28
Cycle	57	70	54	62
Run	26	23	25	36

a How should the athletes be assigned to make the fastest triathlon team?

b Not taking into consideration any transition time during the race, state the total average time it should take the team to complete the triathlon.

3 A courier company needs to send drivers to deliver four parcels to various locations. The delivery times for each of the possible drivers (in minutes) is given below.

	Parcel A	Parcel B	Parcel C	Parcel D
Driver 1	22	14	21	51
Driver 2	19	12	21	28
Driver 3	61	70	50	68
Driver 4	19	22	11	43
Driver 5	60	30	14	86

Assign each parcel to a courier driver to minimise the amount of time it would take to deliver the four parcels.

Maximisation problems

4 A company has three sales representatives who are to be assigned to three sales territories. The monthly sales estimates for each sale representative for the different territories are as follows.

	Territories		
	1	*2*	*3*
Adams	800	1100	1200
Brown	500	1600	1300
Cooper	500	1000	2300

Assign the sales representatives to the territories to maximise the monthly sales.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

5 A company makes four products. There are four operators who can make these products. The number of products that can be made per hour varies from operator to operator.

	Fred	Gigi	Hazel	Ian
Product 1	25	20	21	24
Product 2	30	25	20	37
Product 3	28	25	19	26
Product 4	37	25	25	34

a Find the optimal allocation such that the company can produce the maximum number of products per hour.

b If the operators work for 8 and a half hours a day and allow 30 minutes for lunch, calculate the maximum number of products that can be made by each operator per day.

6 A company has rented a new space and is attempting to determine where various departments should be located within the shop. The shop manager is considering five departments that might occupy the five locations. After a careful study of the layout of the store, the manager has made estimates of the expected annual profit for each department in each location. These estimates, in thousands of dollars, are shown below.

	Ground floor	Floor 1	Floor 2	Floor 3	Floor 4
Clothing	39	65	69	66	57
Toys	64	84	24	92	22
Homewares	49	50	61	31	45
Electrical	48	45	55	23	50
Beauty	59	34	30	34	18

Determine where each department should be located within the shop to maximise annual profit and state the value of this profit.

9E Flow problems

Connector problems are concerned with the length, time or cost of connecting vertices with arcs in a directed graph.

Flow problems are concerned with the transfer of material, such as water through pipes or cars along road. They are concerned with maximising the amount of water, cars or other objects that can pass through the network during a certain time period.

ISBN 978-1-009-25917-0

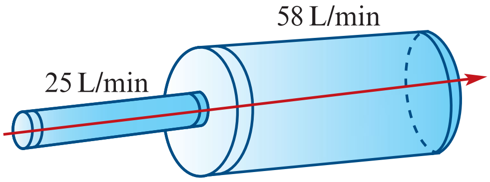
Understanding maximum flow

One of the applications of directed graphs to real-life situations is flow problems. Flow problems involve the transfer or **flow** of material from one point, called the **source**, to another point called the **sink**. Examples of this include water flowing through pipes, or traffic flowing along roads.

source → flow through network → sink

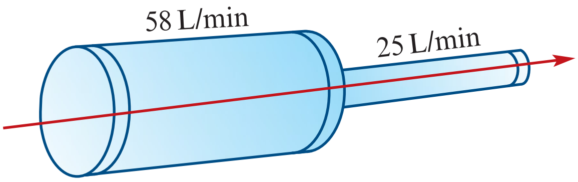
Water flows through pipes in only one direction. In a digraph representing water flow, the vertices are the origin and destination of the water and the arcs represent the pipes connecting them. The weights on the arcs would be the rate at which water can pass through the pipe at any given time. The weights of flow problem directed graphs are called **capacities**.

The diagram on the right shows two pipes that are joined together, connecting the source of water to the sink. There is a small pipe with capacity 25 litres per minute joined to a large pipe with capacity 58 litres per minute.



Even though the large pipe has a capacity greater than 25 litres per minute, the small pipe will only allow 25 litres of water through each minute. The flow through the large pipe will never be more than 25 litres per minute. The large pipe will experience flow below its capacity.

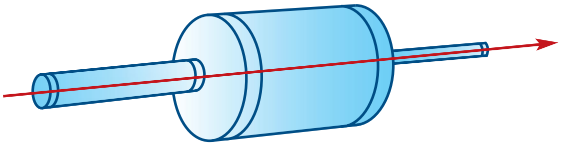
If we reverse the connection and direct water through the large capacity pipe into the smaller capacity pipe, there will be a 'bottleneck' of flow at the junction.



The large capacity pipe is delivering 58 litres of water every minute to the small pipe, but the small pipe will only allow 25 litres per minute to pass.

In both of these situations, the flow through the entire pipe system (both pipes from source to sink) is restricted to a maximum of 25 litres per minute. This is the capacity of the smallest pipe in the connection.

If we connect more pipes together, one after the other, we can calculate the overall capacity or **maximum flow** of the pipe system by looking for the *smallest capacity pipe* in that system.

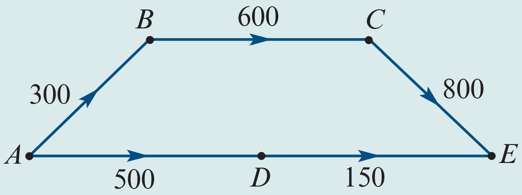


Maximum flow

If pipes of different capacities are connected one after the other, the *maximum flow* through the pipes is equal to the *minimum* capacity of the individual pipes.

Example 10 Calculating the maximum flow

In the digraph shown on the right, the vertices A, B, C, D and E represent towns. The arcs of the graph represent roads and the weights of those arcs are the maximum number of cars that can travel on the road each hour. The roads allow only one-way travel.



a Find the maximum traffic flow from A to E through town C.

b Find the maximum traffic flow from A to E overall.

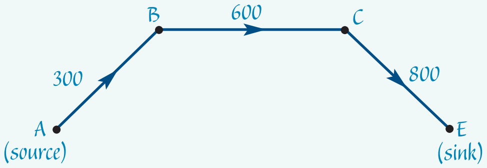
c A new road is being built to allow traffic from town D to town C. This road can carry 500 cars per hour.

  i Add this road to the digraph.

  ii Find the maximum traffic flow from A to E overall after this road is built.

Solution

a Look at the subgraph that includes town C. The smallest capacity of the individual roads is 300 cars per hour. This will be the maximum flow through town C.



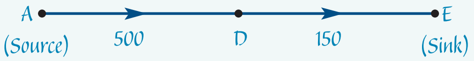
The maximum flow from A to E through town C is equal to the smallest capacity road along that route. The maximum flow is 300 cars per hour.

b Look at the two subgraphs from A to E.

The maximum flow through C will be 300 cars per hour.

The maximum flow through D will be 150 cars per hour (minimum capacity).

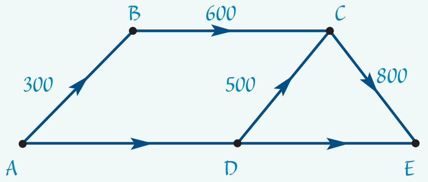
Add the maximum flow through C to the maximum flow through D.



The maximum flow from A to E overall is:

300 + 150 = 450 cars per hour

c **i** Add the arcs to the diagram.



ii Determine the maximum flow from A to E.	The maximum flow through A–B–C–E is 300. But C–E has capacity 800. If another 500 cars per hour come through D–C, they will be able to travel from C–D. The new maximum flow is now 800 cars per hour.

It is important that when we want to determine the maximum flow through a network that we show clear systematic working. We need to work through the network, considering all possible flow paths, and adding the total capacity for each path from source to sink. Using the **Exhaustion of Paths algorithm** we are able to identify the maximum flow for the graph as well as any edges that are left with surplus.

Exhaustion of Paths algorithm

Step 1: Identify the source and sink of the network

Step 2: Choose one possible path from source to sink and identify the edge with the smallest capacity.

- The edge with the smallest capacity = capacity of the chosen path
- Record the capacity and the order of the vertices on this path
- Subtract this value from each number on that path. The values remaining will be the residual capacity for each edge along that path. Update the diagram to show the remaining capacity.

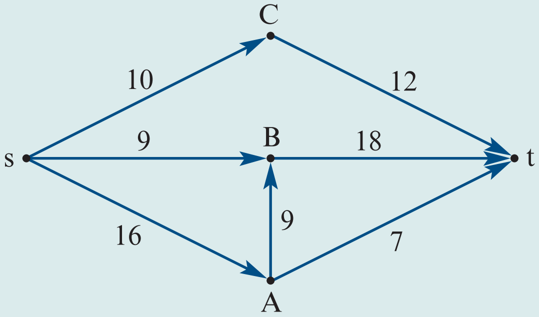
Step 3: Continue to work systematically (from top to bottom) through the flow network finding different paths with a non-zero capacity. Repeat step 2 until all possible paths have been exhausted.

Step 4: Add the recorded capacity of each path. This is the maximum flow through the network.

Note: Recording the capacity of these paths in a different order may find different capacities. The primary intent is to try to find the maximum capacity of flow through the network and this algorithm is one way to do this. The overall capacity of a network can be found by using the understanding that *maximum flow* = *minimum cut*.

Example 11 Using exhaustion of paths to find maximum flow

Use systematic working to find the maximum flow through the network on the right from source (*s*) to sink (*t*).

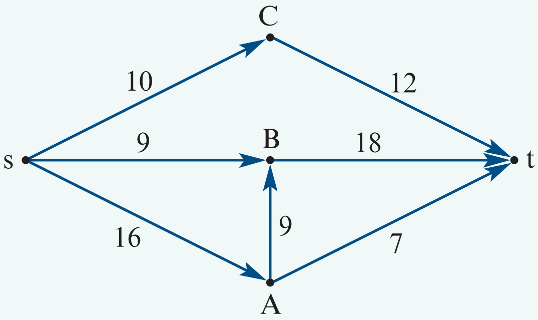


Solution

1 Choose a path with no zero edges:

$s - C - t$

The edge with the smallest capacity is $s - C = 10$

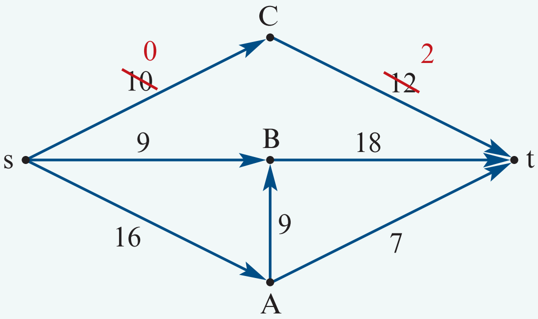


2 Subtract 10 from each of the edges on this path

$s - C = 10 - 10 = 0$

$C - t = 12 - 10 = 2$

3 Update the diagram to show the remaining capacity.



4 Choose a different path with no zero edges:

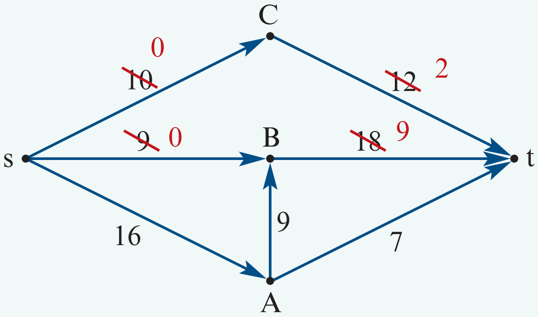
$s - B - t$

The edge with the smallest capacity is $s - B = 9$

$s - B = 9 - 9 = 0$

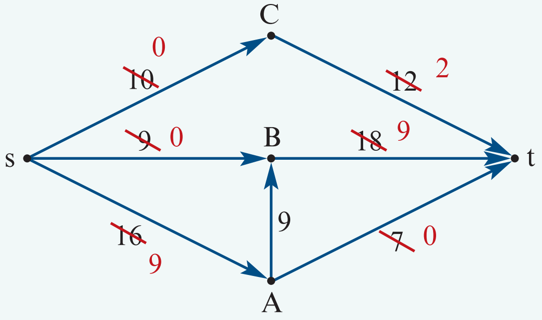
$B - t = 18 - 9 = 9$

Update the diagram to show the remaining capacity.

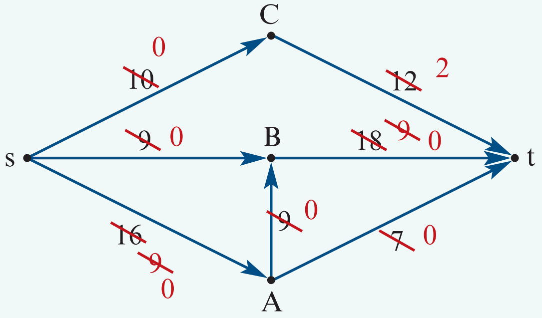


5 Continue to choose different paths, record the capacity of each and update the diagram.

$s - A - t = 7$



$s - A - B - t = 9$



6 There are no more paths that can be chosen. The maximum flow is obtained by adding up each flow that was pushed through the network

$s - C - t = 10$

$s - B - t = 9$

$s - A - t = 7$

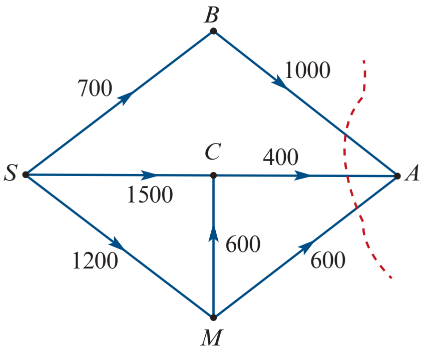
$s - A - B - t = 9$

Maximum flow $= 10 + 9 + 7 + 9 = 35$

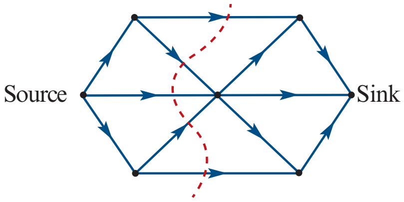
Cuts

It is difficult to determine the maximum flow by inspection for directed networks that involve many vertices and arcs. We can simplify the search for maximum flow by searching for **cuts** within the digraph.

A cut divides the network into two parts, completely separating the source from the sink. It is helpful to think of cuts as imaginary breaks within the network that completely block the flow through that network. For the network of water pipes shown in this diagram, the dotted line is a cut. This cut completely blocks the flow of water from the source (S) to the sink (A).

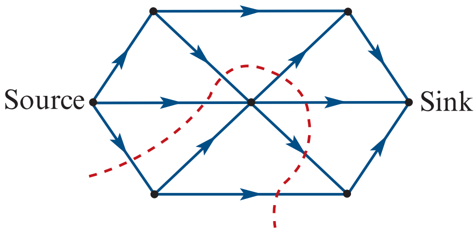


The dotted line on the graph on the right is a cut because it separates the source and the sink completely. No material can flow from the source to the sink.



ISBN 978-1-009-25917-0

The dotted line on the graph on the right is *not* a cut because material can still flow from the source to the sink. Not all of the pathways from source to sink have been blocked by the cut.

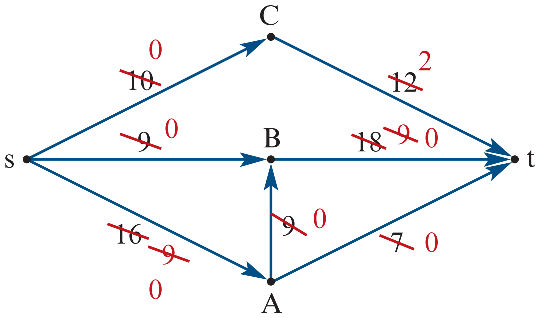


The capacity of a cut is important to help determine the maximum flow through any digraph. Look for the smallest, or minimum, **cut capacity** that exists in the graph. This will be the same as the maximum flow that is possible through that graph. This is known as the *maximum-flow minimum-cut theorem*.

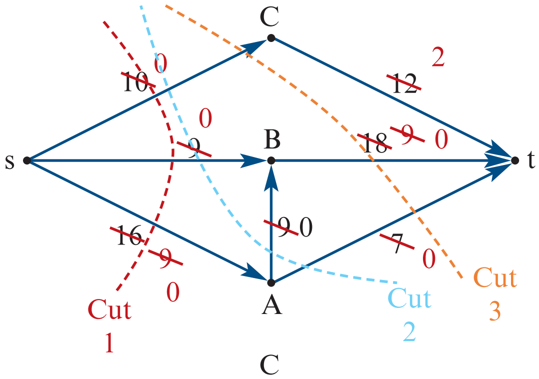
Maximum flow = Minimum cut

Previously the exhaustion of paths algorithm was used to find the maximum flow of a network. Once this algorithm has been followed and written on the network, if a cut is drawn through edges that only have a residual capacity of zero, then the capacity of these edges will be equal to the maximum flow.

From Example 11, our final network diagram looked like this:



Minimum cuts can be identified from this diagram. The minimum cut will only pass through edges of zero. There are multiple minimum cuts in this network.

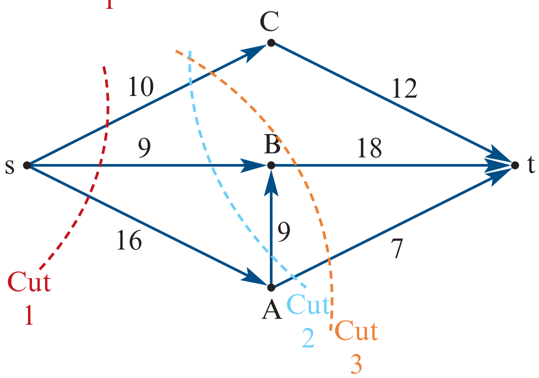


Value of the cuts:

Cut 1 = 10 + 9 + 16 = 35

Cut 2 = 10 + 9 + 9 + 7 = 35

Cut 3 = 10 + 18 + 7 = 35



ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

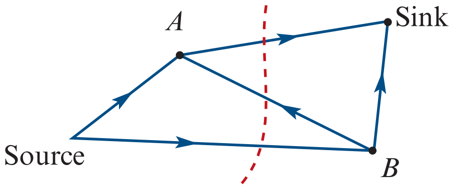
Cut, cut capacity and minimum cut capacity

A *cut* is an imaginary line across a directed graph that completely separates the *source* (start of the flow) from the *sink* (destination of the flow).

The *cut capacity* is the sum of the capacities of the arcs that are cut. Only arcs that flow from the source side of the cut to the sink side of the cut are included in a cut capacity calculation.

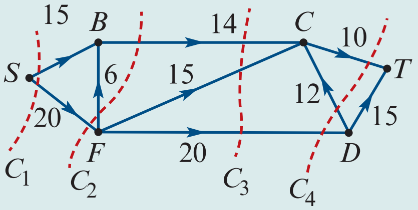
The **minimum cut** *capacity* possible for a graph equals the *maximum flow* through the graph.

In the simple network on the right, the cut passes through three edges. The edge B to A is not counted in the capacity of the cut because the flow for that arc is from the sink side to the source side of the cut.



Example 12 Calculating cut capacity

Calculate the capacity of the four cuts shown in the network on the right. The source is vertex S and the sink is vertex T.



Solution

All arcs in C_1 are counted.

The capacity of $C_1 = 15 + 20 = 35$

Note that the arc from F to B is not counted in C_2.

The capacity of $C_2 = 14 + 20 = 34$

All arcs in C_3 are counted.

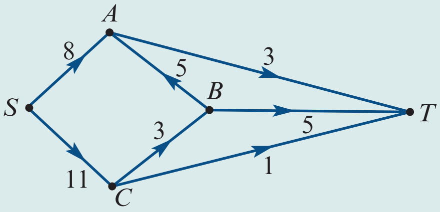
The capacity of $C_3 = 14 + 15 + 20 = 49$

Note that the arc from D to C is not counted in C_4.

The capacity of $C_4 = 20 + 10 = 30$

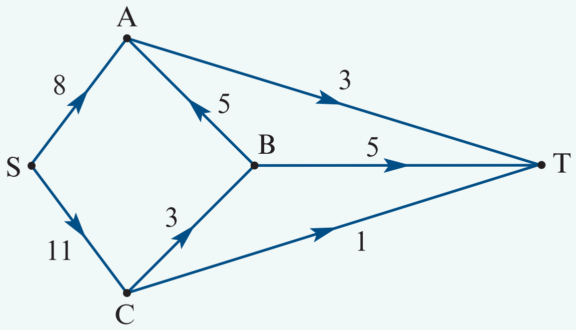
Example 13 Calculating minimum-cut maximum-flow

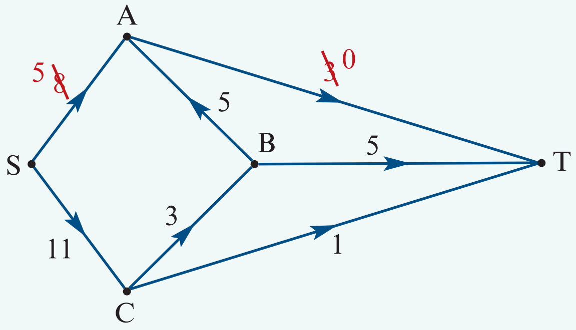
Determine the maximum flow from S to T for the digraph shown on the right.



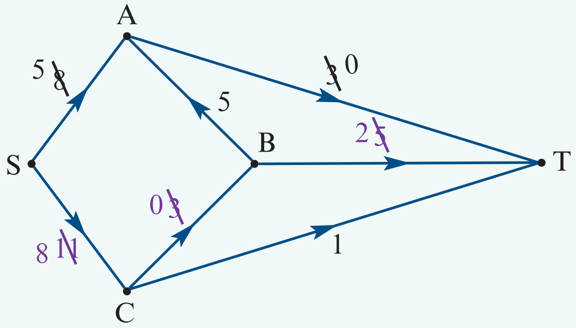
Solution

1 Follow exhaustion of paths algorithm to determine possible maximum flow for network.

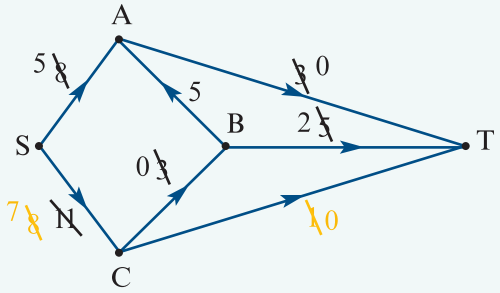




SAT = 3



SCBT = 3



SCT = 1

ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

2 Add up each flow that was pushed through the network.

$SAT = 3$

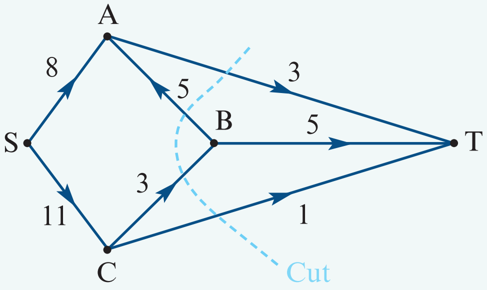
$SCBT = 3$

$SCT = 1$

Maximum flow $= 3 + 3 + 1 = 7$

3 Check maximum flow by finding minimum cut.
Identify the minimum cut from the diagram.
The minimum cut will only pass through edges of zero.
Add the capacity of each of the edges on the minimum cut.

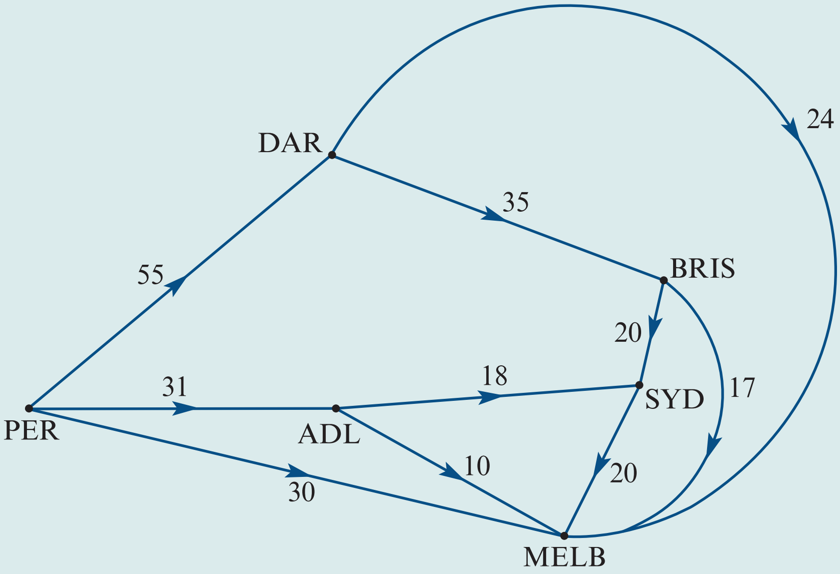
Note: The edge $B - A$ is not counted as the cut has a counter-flow that goes from sink to source (rather than source to sink).



Value of cut $= 3 + 3 + 1 = 7$

Example 14 Calculating maximum-flow

The network below shows the flow of luggage on flights between Australia's domestic airports, in thousands of bags per day.



a Determine the maximum number of bags that can be handled on all flights from Perth to Melbourne.

b Additional direct flights are added from Perth to Brisbane which will handle 16 000 bags per day in total. Determine the overall impact these flights will have on the maximum flow of luggage from Perth to Melbourne.

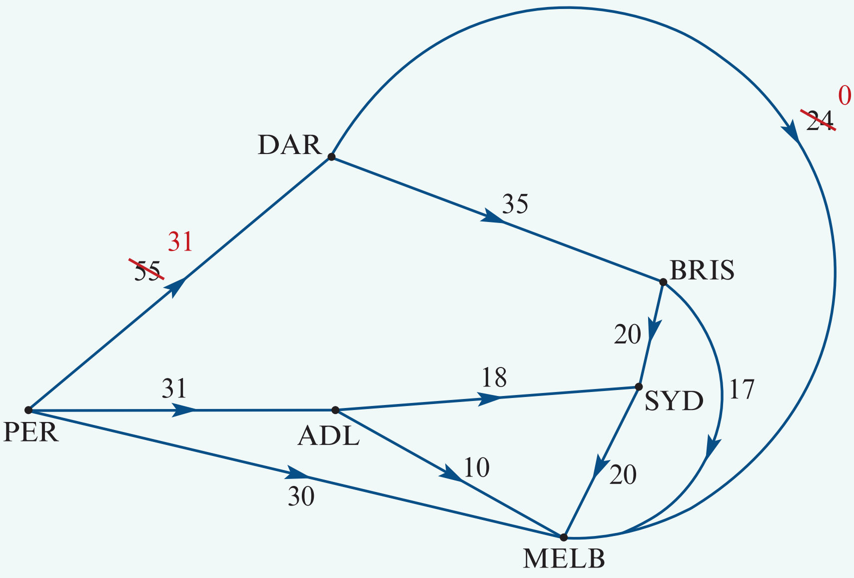
ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

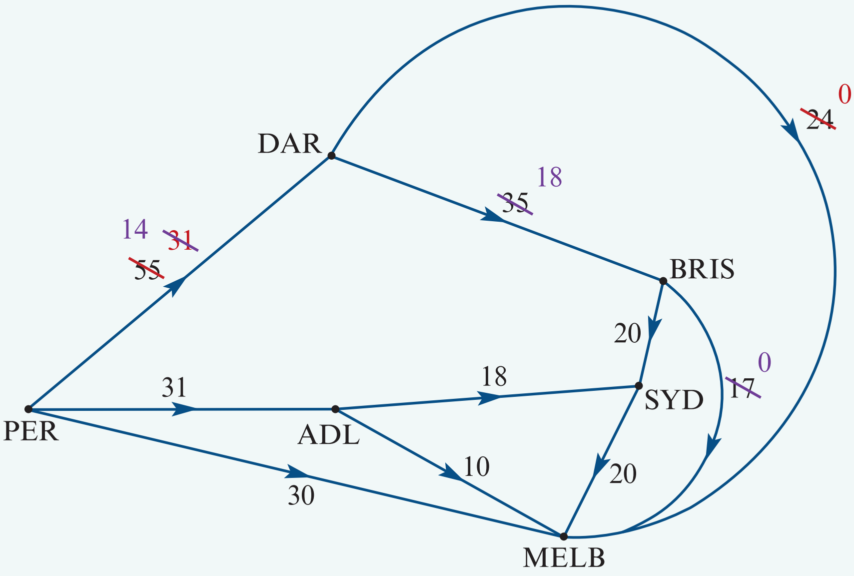
c During Summer more passengers, and their bags, need to be flown from either Adelaide or Sydney to Melbourne for the Australian Open. Determine which capital city should increase its baggage handling capabilities to have the greatest impact on the overall flow of luggage. Describe how this would affect the flow of luggage on other flights if this airport operated to its new capacity.

Solution

a Follow exhaustion of paths algorithm to determine maximum flow for network.

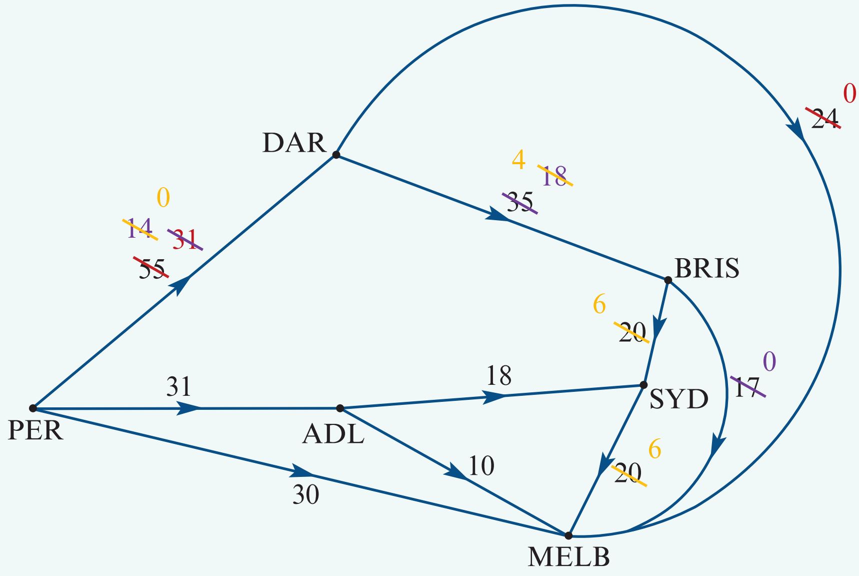


PER – DAR – MELB = 24

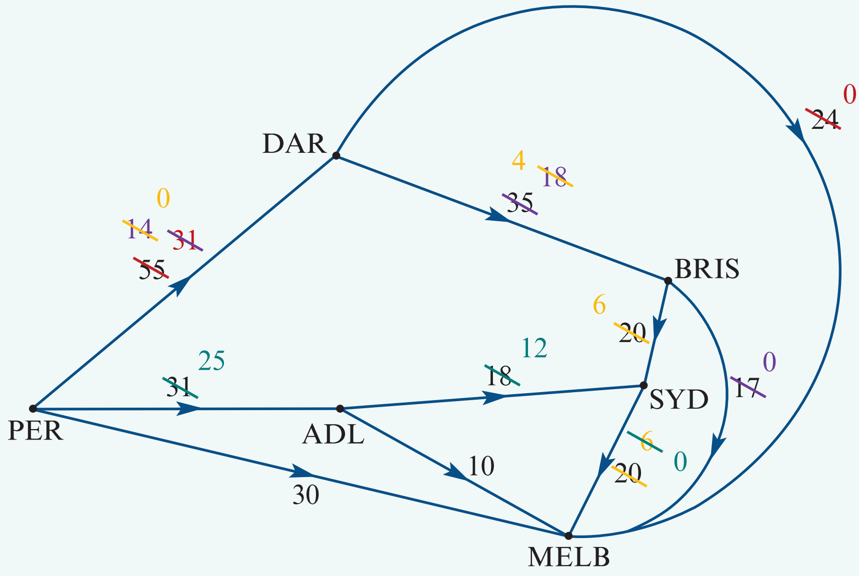


PER – DAR – BRIS – MELB = 17

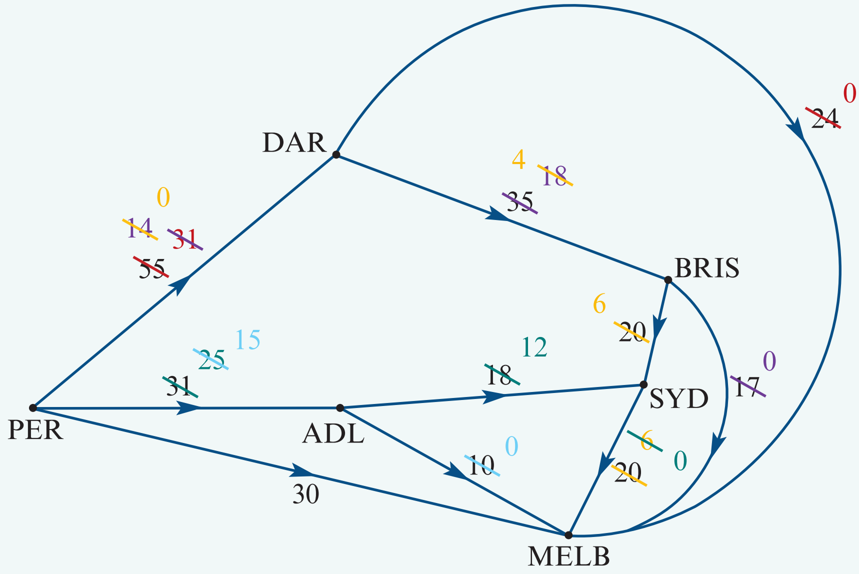
ISBN 978-1-009-25917-0



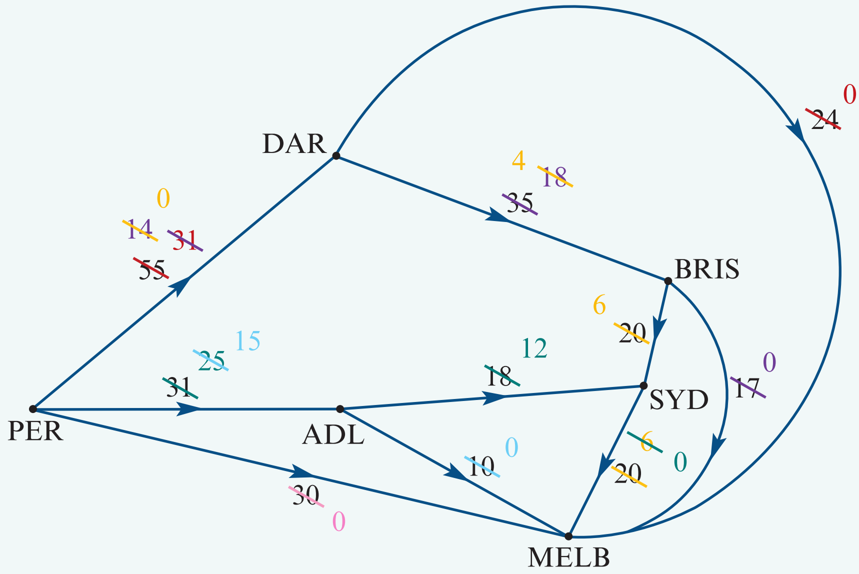
PER – DAR – BRIS – SYD – MELB = 14



PER – ADL – SYD – MELB = 6



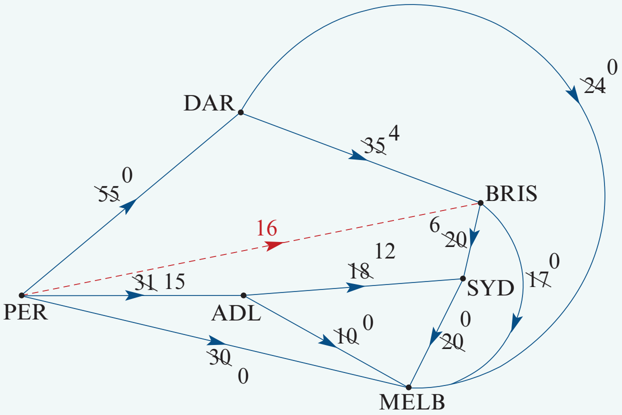
PER – ADL – MELB = 10



PER – MELB = 30

Add each flow pushed through the network.	24 + 17 + 14 + 6 + 10 + 30 = 101
Write answer in context of question.	Maximum flow is 101 000 bags per day.

b Draw additional flight and capacity of edge on network.

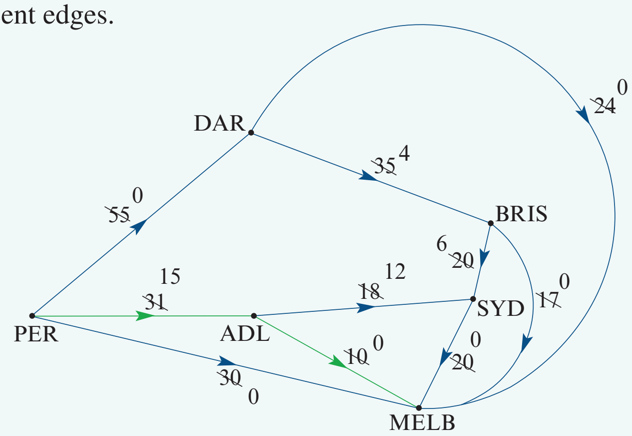


Consider flow out of Brisbane following non-zero edges (edges with excess and not operating at maximum capacity)

Consider path PER – BRIS – SYD – MELB:
Path SYD – MELB at capacity (has 0 excess) so no more luggage can be pushed through the network, despite BRIS – SYD having an excess of 6 (thousand bags per day) which could be used up.

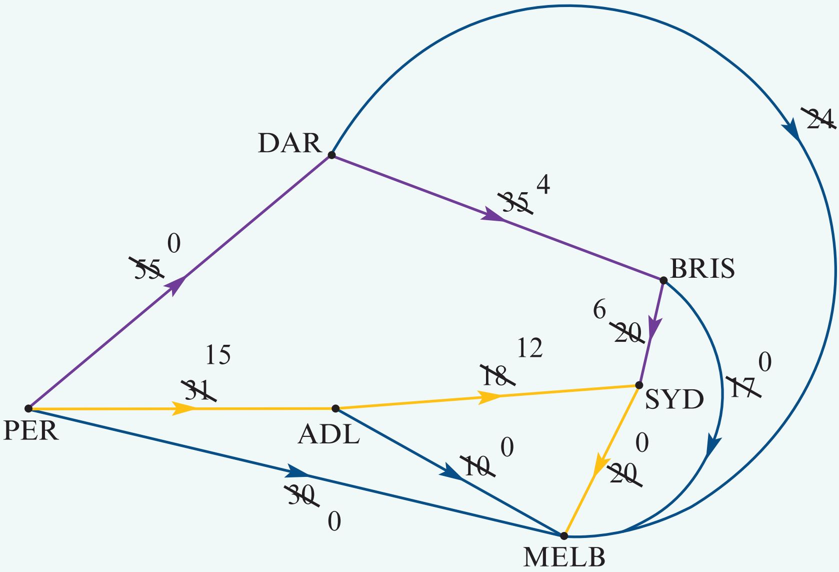
This new flight has no impact on overall maximum flow.

c Maximum flow can only be increased by increasing the flow of edges that are currently at capacity.
The maximum amount these can be increased is determined by the lowest amount of excess on adjacent edges.



Consider flow from ADL – MELB:
Flight from PER – ADL is the only incoming path and has an excess of 15. By increasing ADL – MELB by 15 this would now allow an extra 15 000 bags per day to flow from PER – MELB.

ISBN 978-1-009-25917-0



Consider flow from SYD – MELB:
Flights flow into SYD from both BRIS and ADL. PER – ADL has an excess of 15, but ADL – SYD only has 12. So overall increase could be up to 12 000 extra bags. BRIS – SYD has an excess of 6, however, PER – DAR – BRIS cannot take any more bags since PER – DAR is at capacity with no excess.

Best solution is to increase flow from ADL – MELB by 15 000 bags per day so that both PER – ADL and ADL – MELB will now be operating at capacity.

ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

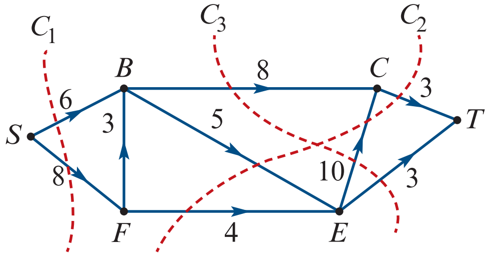
Skill-sheet

Exercise 9E

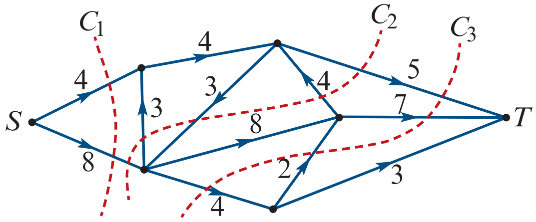
Calculating cut capacity

Example 12

1 Determine the capacity of each of the cuts in the digraph. The source is vertex S and the sink is vertex T.



2 Determine the capacity of each of the cuts in the digraph. The source is vertex S and the sink is vertex T.

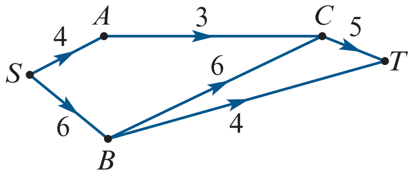


Calculating minimum-cut maximum-flow

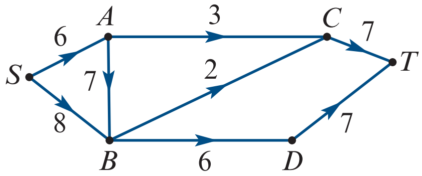
Example 13

3 Find the maximum flow for each of the following graphs. The source is vertex S and the sink is vertex T.

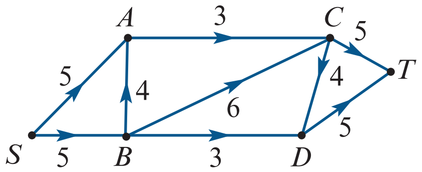
a



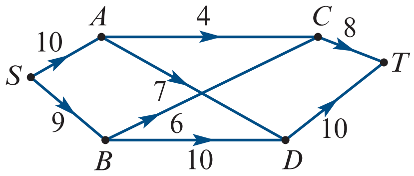
b



c

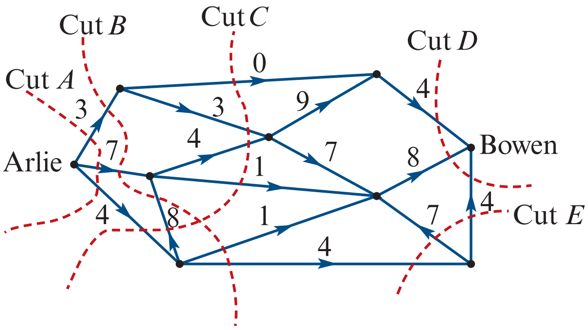


d



ISBN 978-1-009-25917-0

4 A train journey consists of a connected sequence of stages formed by arcs on the following directed network from Arlie to Bowen. The number of available seats for each stage is indicated beside the corresponding arc, as shown on the diagram on the right.



The five cuts, *A*, *B*, *C*, *D* and *E*, shown on the network are attempts to find the maximum number of available seats that can be booked for a journey from Arlie to Bowen.

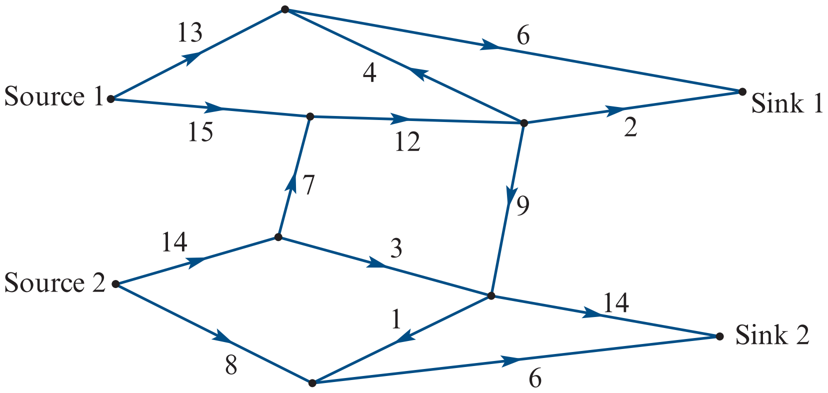
a Write the capacity of cut *A*, cut *B*, cut *C*, and cut *D*.

b Explain why cut *E* is not a valid cut when trying to find the minimum cut between Arlie and Bowen.

c Find the maximum number of available seats for a train journey from Arlie to Bowen.

Example 14

5 Water pipes of different capacities are connected to two water sources and two sinks. The network of water pipes is shown in the diagram below. The numbers on the arcs represent the capacities, in kilolitres per minute, of the pipes.



Find the maximum flow, in kilolitres per minute, to each of the sinks in this network.

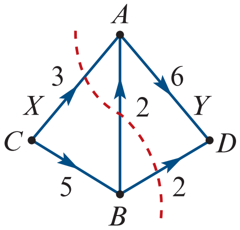
Key ideas and chapter summary

Assignment

Tree	A **tree** is a connected graph that contains no cycles, multiple edges or loops. A tree with n vertices has $n - 1$ edges.
Spanning tree	A **spanning tree** is a tree that connects every vertex of a graph. It is found by counting the number of vertices (n) and removing enough edges so that there are $n - 1$ edges left that connect all vertices.
Minimum spanning tree	A **minimum spanning tree** is a spanning tree for which the sum of the weights of the edges is as small as possible.
Prim's algorithm	**Prim's algorithm** is an algorithm for determining the minimum spanning tree of a network.
Connector problems	**Connector problems** involve situations where the cost of connecting vertices into a network needs to be kept as small as possible. Minimum spanning trees can be used to solve connector problems.
Directed graph (digraph)	A **directed graph** is a graph where direction is indicated for every edge. This is often abbreviated to **digraph**.
Arc	The edges of directed graphs are called **arcs**.
Bipartite graph	A **bipartite graph** has two distinct groups or categories for the vertices. Connections exist between a vertex or vertices from the other group. There are no connections between the vertices within a group.
Complete bipartite graph	If every vertex in one group of vertices of a bipartite graph is connected to every vertex in the other group of vertices, then the graph will be a **complete bipartite graph**.
Allocation	An **allocation** is made when each of the vertices in one group from a bipartite graph are matched with one of the vertices in the other group from that graph. An allocation is possible when both groups have exactly the same number of vertices. The vertices in each group are matched to only one vertex from the other group.
Cost matrix	A table that contains the costs of allocating objects from one group (such as people) to another (such as tasks). The 'cost' can be money, or other factors such as the time taken.
Hungarian algorithm	The **Hungarian algorithm** is an algorithm that is used to determine the best allocation to minimise the overall cost.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Flow	The transfer of material through a directed network. **Flow** can refer to the movement of water or traffic.
Capacity	The maximum flow of substance that an edge of a directed graph can allow during a particular time interval. The **capacity** of water pipes is the amount of water (usually in litres) that the pipe will allow through per time period (minutes, hours, etc.). Other examples of capacity are number of cars per minute or number of people per hour.
Source	The **source** is the origin of the material flowing through a network.
Sink	The **sink** is final destination of the material flowing through a network.
Cut	A **cut** is a line dividing a directed graph into two parts (shown as a broken line dividing the graph below into two sections, labelled *X* and *Y*). 
Cut capacity	The sum of the capacities (weights) of the edges directed from *X* to *Y* that the cut passes through. For the cut in the weighted digraph shown, the capacity of the cut is 7.
Minimum cut	The **minimum cut** is the cut with the minimum capacity. The cut must separate the source from the sink.
Maximum flow	The **maximum flow** through a directed graph is equal to the capacity of the minimum cut.
Flow problems	**Flow problems** involve the analysis of flow through a network, usually in terms of the maximum flow possible.

Skills check

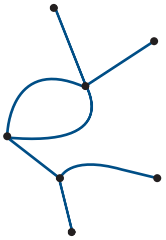
Having completed this chapter you should be able to:

- define tree, spanning tree, minimum spanning tree
- draw a minimum spanning tree using Prim's algorithm.
- solve connector problems using minimum spanning trees
- identify directed graphs
- write and interpret adjacency matrices for directed graphs
- identify and interpret bipartite graphs
- solve allocation problems using bipartite graphs and the Hungarian algorithm
- calculate the capacity of a cut
- calculate the maximum flow of a network using the minimum cut for that network
- solve flow problems.

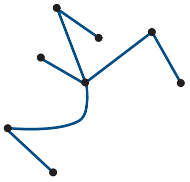
Short-answer questions

1 Which one of the following graphs is a tree?

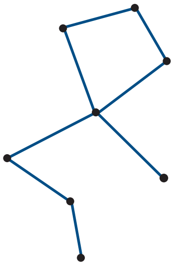
A



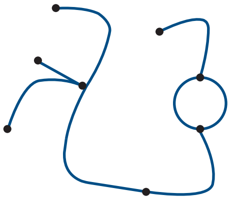
B



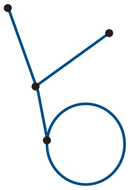
C



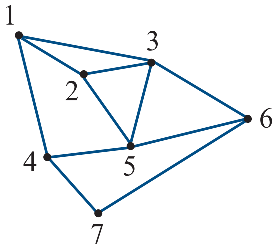
D



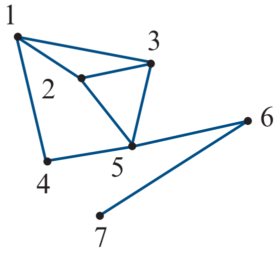
E



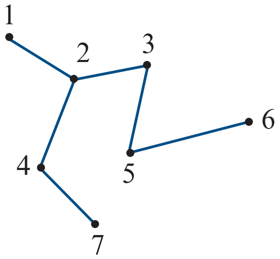
2 Which one of the following graphs is a spanning tree for the graph shown?



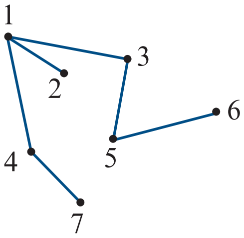
A



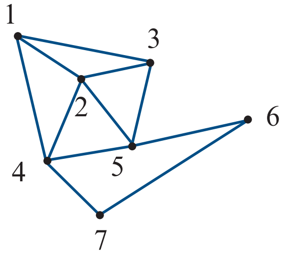
B



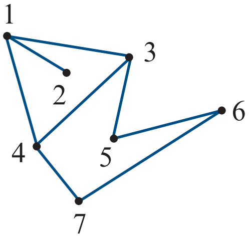
C



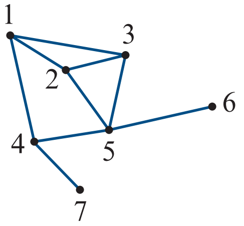
D



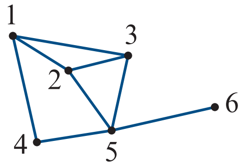
E



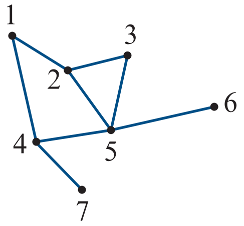
3 Which of the following graphs is a spanning tree for the network shown?



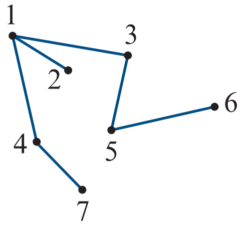
A



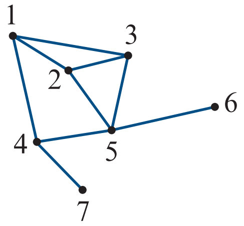
B



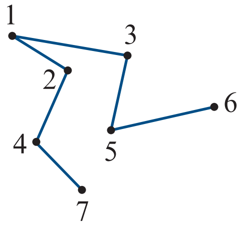
C



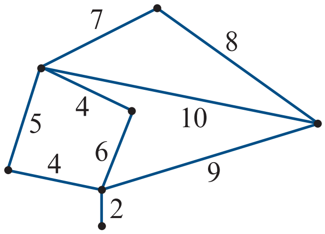
D



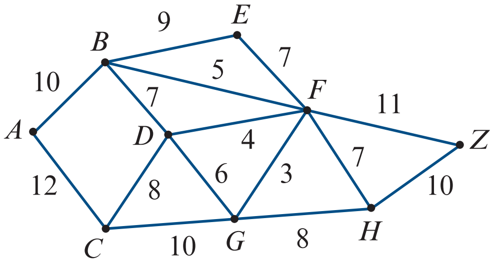
E



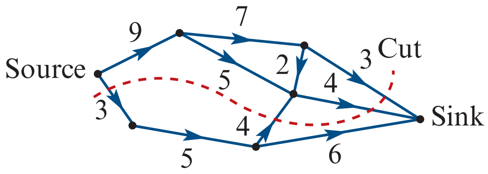
4 For the graph shown here, determine the length of the minimum spanning tree.



5 Determine the length of the minimum spanning tree for the network shown on the right.



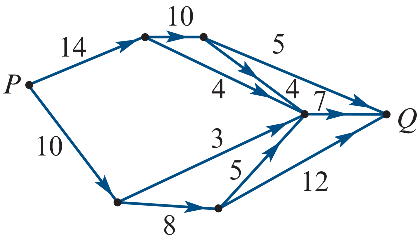
6 For the network shown on the right, determine the capacity of the cut.



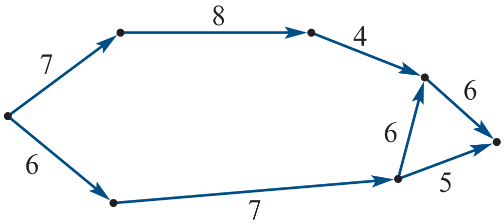
ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

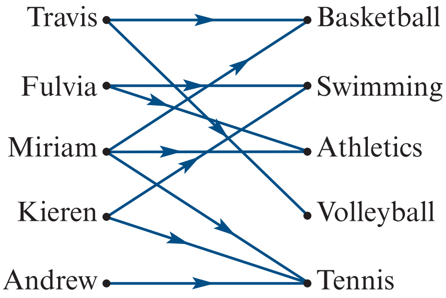
7 In the communications network shown, the numbers represent transmission capacities for information (data) in scaled units. What is the maximum flow of information from station P to station Q?



8 Determine the maximum flow in the network opposite, from source to sink.

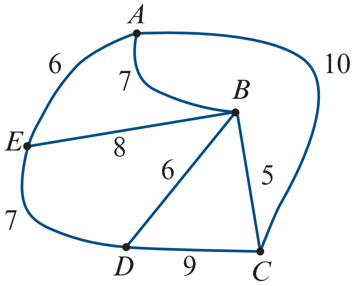


9 A group of five students represent their school in five different sports. Use the bipartite graph to allocate each student to a sport.



Extended-response questions

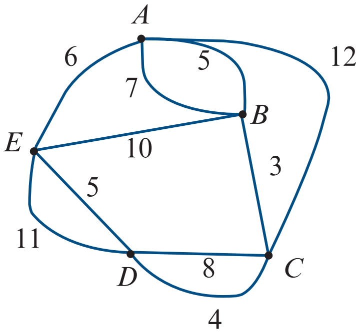
1 For the weighted graph shown, determine the length of the minimum spanning tree.



2 In the network below, the numbers on the edges represent distances in kilometres.
Determine the length of:

a the shortest path between vertex A and vertex D

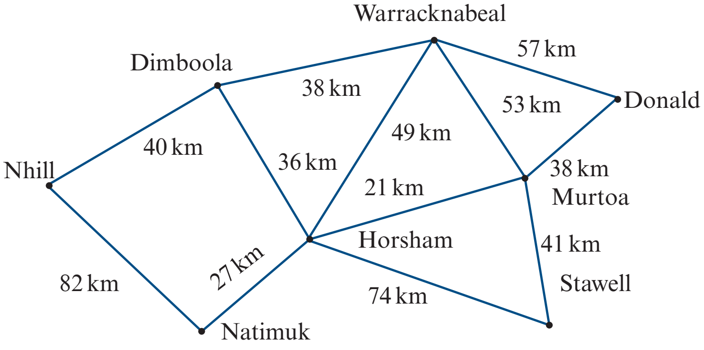
b the length of the minimum spanning tree.



ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

3 The network below shows the major roads connecting eight towns in Victoria and the distances between them in kilometres.



a Find the shortest distance between Nhill and Donald using these roads.

b Verify Euler's formula for this network.

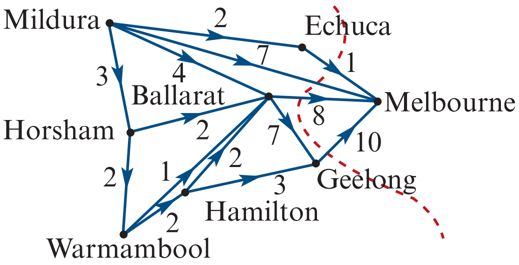
c An engineer based in Horsham needs to inspect each road in the network without travelling along any of the roads more than once. He would also like to finish his inspection at Horsham.

- **i** Explain why this cannot be done.
- **ii** The engineer can inspect each road in the network without travelling along any of the roads more than once if he starts at Horsham but finishes at a different town. Which town is that? How far will he have to travel in total?
- **iii** Identify one route, starting at Horsham, that the engineer can take to complete his inspection without travelling along any of the roads more than once.

d The engineer can complete his inspection in Horsham by only travelling along one of the roads twice. Which road is that?

e A telecommunications company wants to connect all of the towns to a central computer system located in Horsham. What is the minimum length of cable that they will need to complete this task?

4 WestAir Company flies routes in western Victoria. The network shows the layout of connecting flight paths for WestAir, which originate in Mildura and terminate in either Melbourne or on the way to Melbourne.



The available spaces for passengers flying out of various locations on one morning are shown.

The network has one cut shown.

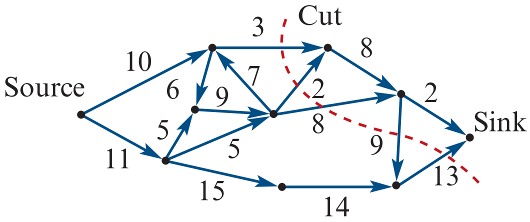
a What is the capacity of this cut?

b What is the maximum number of passengers who could travel from Mildura to Melbourne for the morning?

5 A school swimming team wants to select a 4 × 200 metre relay team. The fastest times of its four best swimmers in each of the strokes are shown in the table below. Which swimmer should swim each stroke to give the team the best chance of winning, and what would be their time to swim the relay? Assume each swimmer may only swim one stroke in the relay team.

Swimmer	*Backstroke*	*Breaststroke*	*Butterfly*	*Freestyle*
Rob	76	78	70	62
Joel	74	80	66	62
Henk	72	76	68	58
Sav	78	80	66	60

6 In the network opposite, the values on the edges give the maximum flow possible between each pair of vertices. The arrows show the direction of flow in the network. Also shown is a cut that separates the source from the sink.



a Determine the capacity of the cut shown.

b Determine the maximum flow through this network.

7 Blue Water Retreat has five cabins, A, B, C, D and E. There are paths between some of these cabins and the office. The lengths (in metres) of the paths are given in the following table:

	A	*B*	*C*	*D*	*E*	*Office*
A		60	65			
B	60		55		90	
C	65	55		35	75	
D			35			105
E		90	75			110
Office				105	110	

a Use Prim's algorithm to find the length of the minimum spanning tree.

b The owner of the retreat wishes to only upgrade the paths that are on the minimum spanning tree. The cost of upgrading is $37.50 per metre. What is the total cost of the upgrade?

c During the winter months the path between A and C becomes unusable, and so must be included as part of the upgrade. How does the inclusion of this path affect the cost of the upgrade?

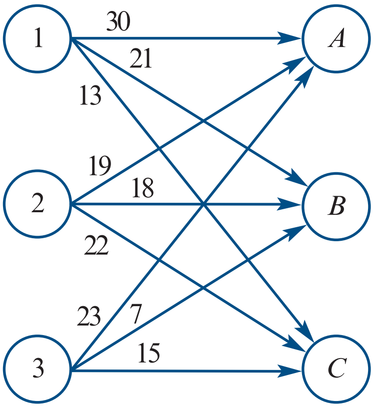
ISBN 978-1-009-25917-0

Cambridge University Press

Updated January 2024

8 Three students are working on a group project and decide to divide up the workload so that each person is working on a task on their own. The tasks to be completed are denoted by A, B and C.

Each person can complete the three tasks that are available in a set time. These times, in minutes, are shown in the bipartite graph on the right.



a Represent these costs in a matrix.

b Use the Hungarian algorithm to assign each person to a task so that the group project can be completed in the least time possible.

9 A store manager asks each of their staff for their availability during a particularly busy period. The table on the right shows the number of hours each staff member is available to work across the four days.

	Thursday	*Friday*	*Saturday*	*Sunday*
Ottilie	7	3	6	9
Percy	7	5	7	5
Quinn	3	1	6	5
Rachel	2	4	9	9

How should the manager allocate each of their staff to the days of the week, such that they can maximise the number of hours worked? State the total number of hours worked during the busy period. Assume that no staff member can be assigned to more than one day of work.

10 A concert venue offers different pay for different jobs at its stadium. Five new employees have been hired for a large concert at the weekend. The table right shows the different pay rates, in dollars per hour, that each of the new employees would be paid to complete each job at the stadium. Use the Hungarian algorithm to assign one employee to each of the jobs such that the venue can be most cost-effective.

	Usher	*Tickets*	*Cleaning*	*Food Service*
Asha	20	15	17	22
Brian	16	12	15	14
Charlie	19	19.5	19	30
Dani	18	17.5	16	19
Emma	17	14	18	20

10

Project planning

In this chapter

Syllabus references

Topic: Project planning and scheduling using critical path analysis (CPA)

Subtopics: 4.3.4 – 4.3.8

ISBN 978-1-009-25917-0

10A Precedence tables and activity networks

Drawing activity networks from precedence tables

Building a house, manufacturing a product, organising a wedding and other similar projects all require many individual **activities** to be completed before the project is finished. The individual activities often rely upon each other and some can't be performed until other activities are complete.

In the organisation of a wedding, invitations would be sent out to guests, but a plan for seating people at the tables during the reception can't be completed until the invitations are accepted. When building a house, the plastering of the walls can't begin until the house is sealed from the weather.

For any project, if activity A must be completed before activity B can begin then activity A is said to be an **immediate predecessor** of activity B. The activities within a project can have multiple immediate predecessors and these are usually recorded in a table called a **precedence table**.

This precedence table shows some of the activities involved in a project and their immediate predecessors.

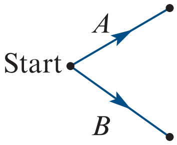
The information in the precedence table can be used to draw a network diagram called an **activity network**.

Activity networks do not have labelled vertices, other than the *start* and *finish* of the project. The activities in the project are represented by the edges of the diagram and *so it is the edges that must be labelled*, not the vertices.

Activity	*Immediate predecessors*
A	–
B	–
C	A
D	B
E	B
F	C, D
G	E, F

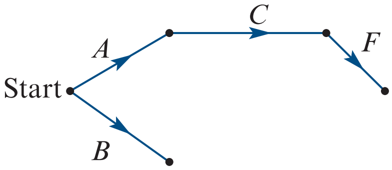
Activities A and B have no immediate predecessors.

These activities can start immediately and can be completed at the same time.



Activity A is an immediate predecessor of activity C, so activity C must follow immediately after activity A.

Activity C is an immediate predecessor of activity F, so activity F must follow immediately after activity C.

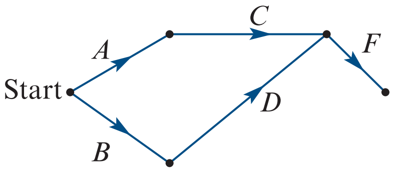


ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

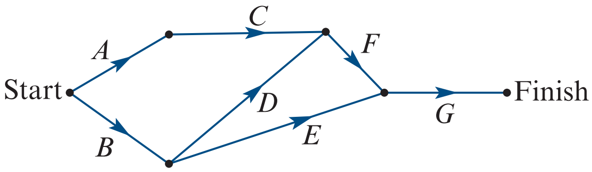
Activity D has immediate predecessor activity B so it follows immediately after activity B.

Activity D is also an immediate predecessor of activity F so activity F must follow immediately after activity D.



Activity E has immediate predecessor activity B so it will follows immediately after activity B.

Activity G has immediate predecessor activity F and activity E and so it must follow immediately after both of these activities.



Activity G is not an immediate predecessor for any activity and so the project is finished after this activity is complete.

Activity networks

When activity A must be completed before activity B can begin, activity A is called an *immediate predecessor* of activity B.

A table containing the activities of a project, and their immediate predecessors, is called a *precedence table*.

An *activity network* can be drawn from a precedence table. Activity networks have edges representing activities. The vertices are not labelled, other than the start and finish vertices.

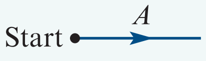
Example 1 Constructing an activity network from a precedence table

Draw an activity network from the precedence table shown on the right.

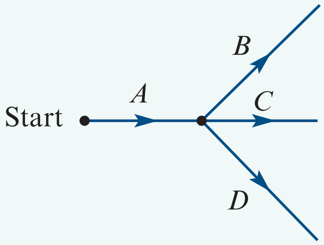
Activity	Immediate predecessors
A	–
B	A
C	A
D	A
E	B
F	C
G	D
H	E, F, G

Solution

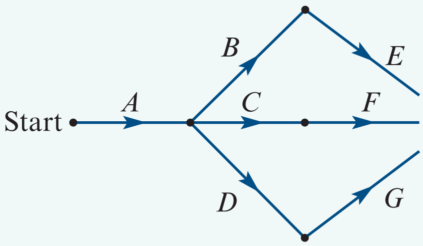
Activity A has no immediate predecessors, so it is the start of the project.



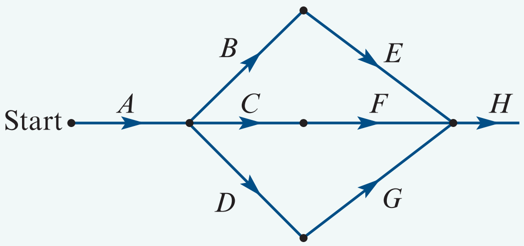
Activities B, C and D all have A as an immediate predecessor. So these three activities all will lead out of A.



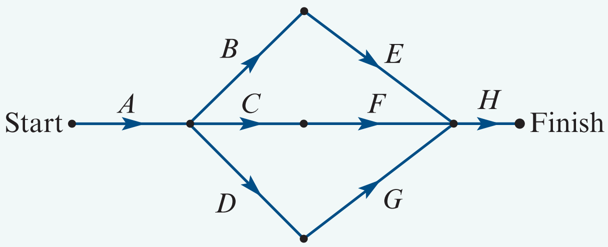
Activity B is an immediate predecessor of activity E. There will be a path from B to E. Activity C is an immediate predecessor of activity F. There will be a path from C to F. Activity D is an immediate predecessor of activity G. There will be a path from D to G.



H has immediate predecessors E, F and G and so these three activities will lead into activity H.



H is not an immediate predecessor for any other activity so it will lead to the finish of the project.



ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

Sketching activity networks

Activities that have no immediate predecessors follow from the start vertex.

Activities that are not immediate predecessors for other activities lead to the finish vertex.

For every other activity:

- look for which activities it is an immediate predecessor of
- look for which activities it has as immediate predecessors.

Construct the activity network from this information.

Dummy activities

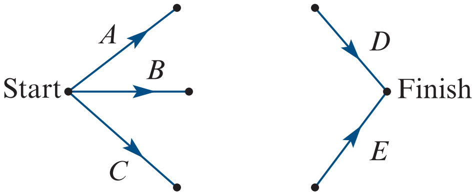
Activity	*Immediate predecessors*
A	–
B	–
C	–
D	*A*, *B*
E	*B*, *C*

Sometimes two activities will have some of the same immediate predecessors, but not all of them. In this simple precedence table, activity *D* and activity *E* share the immediate predecessor activity *B*, but they both have an immediate predecessor activity that the other does not.

This overlap of predecessors presents some difficulty when constructing the activity network, but this difficulty is easily overcome.

Activity *D* and activity *E* are not immediate predecessors for any other activity, so they will lead directly to the finish vertex of the project.

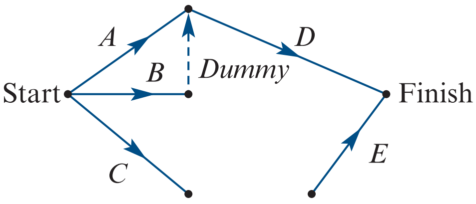
Activites *A*, *B* and *C* have no immediate predecessors, so they will follow directly from the start vertex of the project.



The start and finish of the activity network are shown in the diagram above. We need to use the precedence information for activity *D* and activity *E* to join these two parts together.

Activity *D* needs to follow directly from activity *A* and activity *B*, but we can only draw one edge for activity *D*. Activity *E* needs to follow directly from both activity *B* and activity *C*, but again we only have one edge for activity *E*, not two.

The solution is to draw the diagram with activity *D* starting after one of its immediate predecessors, and using a **dummy activity** for the other. The dummy activities are represented by dotted edges and are, in effect, imaginary. They are not real activities, but they allow all of the predecessors from the table to be correctly represented.

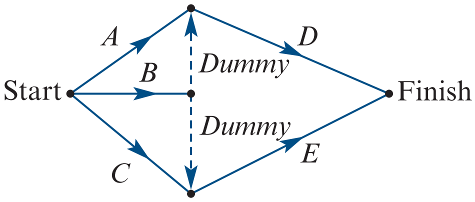


The dummy activity for *D* allows activity *D* to directly follow both activity *A* and *B*.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

A dummy activity is also needed for activity E because it, too, has to start after two different activities, activity B and C.



Dummy activities

A dummy activity is required if two activities share some, but not all, of their immediate predecessors.

A dummy activity will be required *from the end* of each shared immediate predecessor *to the start* of the activity that has additional immediate predecessors.

Dummy activities are represented in the activity network using *dotted lines*.

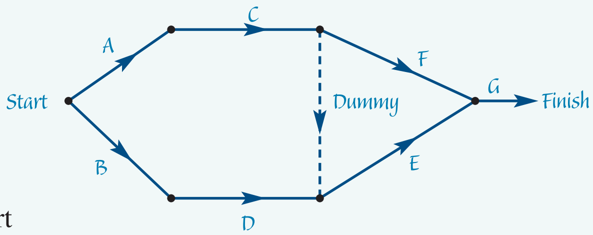
Example 2 Constructing an activity network requiring a dummy activity from a precedence table

Draw an activity network from the precedence table shown on the right.

Activity	*Immediate predecessors*
A	–
B	–
C	A
D	B
E	C, D
F	C
G	E, F

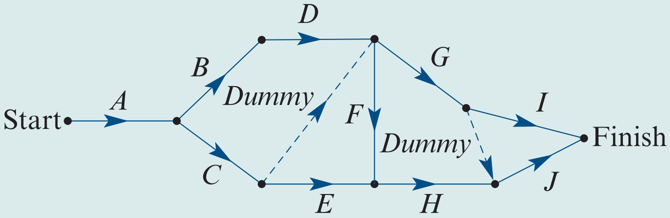
Solution

- A and B will lead from the start vertex.
- G will lead to the end vertex.
- A dummy will be required from the end of activity C (shared immediate predecessor) to the start of activity E (the activity with an additional immediate predecessor).



Example 3 Constructing a precedence table from an activity network

Write a precedence table for the activity network shown on the right.



Solution

1. Create a table with a row for each activity.
2. Look at the start of an activity. Write down all of the activities that lead directly to this activity in the immediate predecessor column.
3. The dummy activity makes activity C a predecessor of activities E, F and G.
4. The dummy activity makes activity G a predecessor of activities I and J.

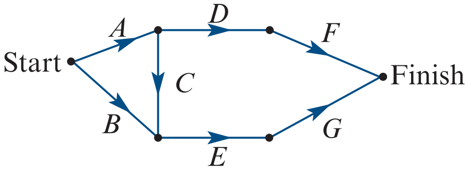
Activity	Immediate predecessors
A	–
B	A
C	A
D	B
E	C
F	D, C
G	D, C
H	E, F
I	G
J	G, H

Exercise 10A

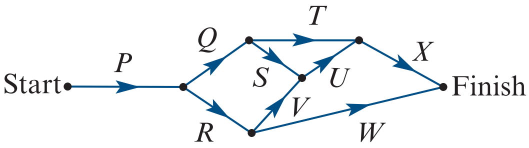
Constructing precedence tables from activity networks

1 Write a precedence table for the activity networks shown below.

a

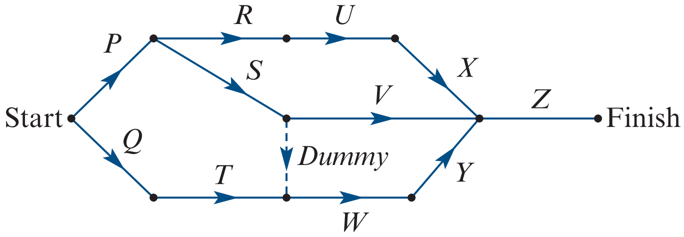


b



ISBN 978-1-009-25917-0

c



Constructing activity networks from precedence tables

Example 1

2 Draw an activity network for each of the precedence tables below.

a

Activity	*Immediate predecessors*
A	–
B	A
C	A
D	B
E	C

b

Activity	*Immediate predecessors*
P	–
Q	–
R	P
S	Q
T	R, S

c

Activity	*Immediate predecessors*
T	–
U	–
V	T
W	U
X	V, W
Y	X
Z	Y

d

Activity	*Immediate predecessors*
F	–
G	–
H	–
I	F
J	G, I
K	H, J
L	K

Constructing activity networks requiring dummy activities from precedence tables

Example 2

3 Draw an activity network for the following precedence tables. Dummy activities will need to be used.

a

Activity	*Immediate predecessors*
F	–
G	–
H	F
I	H, G
J	G

b

Activity	*Immediate predecessors*
A	–
B	A
C	A
D	B
E	B, C

ISBN 978-1-009-25917-0

10B Scheduling problems

Scheduling

Projects that involve multiple activities are usually completed against a time schedule. Knowing how long individual activities within a project are likely to take allows managers of such projects to hire staff, book equipment and also to estimate overall costs of the project. Allocating time to the completion of activities in a project is called *scheduling*. Scheduling problems involve analysis to determine the minimum overall time it would take to complete a project.

Weighted precedence tables

The estimated time to complete activities within a project can be recorded in a precedence table, alongside the immediate predecessor information.

A precedence table that contains the estimated duration, in days, of each activity is shown on the right.

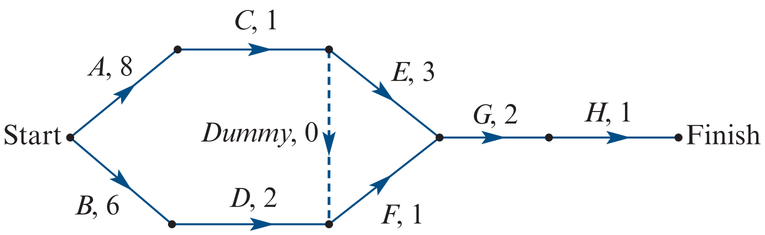
These estimated times are the weights for the edges of the activity network and need to be recorded alongside the name of the activity on the graph.

A dummy activity is required from the end of activity C. Both E and F have C as an immediate predecessor, but F also has an extra immediate predecessor, D. That is why it requires a dummy activity.

Activity	*Estimated completion time (days)*	*Immediate predecessors*
A	8	–
B	6	–
C	1	A
D	2	B
E	3	C
F	1	C, D
G	2	E, F
H	1	G

The weight (duration) of dummy activities is always zero.

The activity network for this project is shown below.

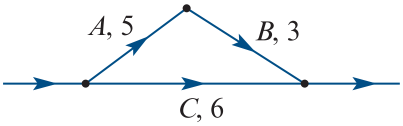


ISBN 978-1-009-25917-0

Float times

The diagram below shows a small section of an activity network. There are three activities shown, with their individual durations, in hours.

Activity B and activity C are both immediate predecessors to the next activity, so the project cannot continue until both of these tasks are finished. Activity B cannot begin until activity A is finished.



Activity C can be completed at the same time as activity A and activity B.

Activity A and B will take a total of $5 + 3 = 8$ hours, while activity C only requires 6 hours. There is some flexibility around when activity C needs to start. There are $8 - 6 = 2$ hours spare for the completion of activity C. This value is called the **float time** for activity C. Another name for float time is slack time.

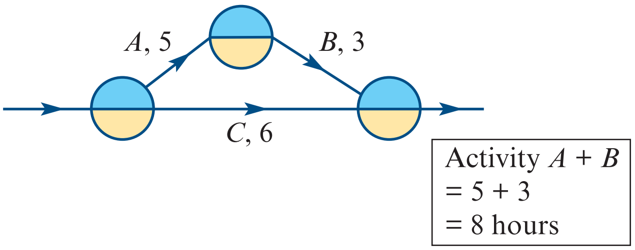
The flexibility around the starting time for activity C can be demonstrated with the following diagram.

	A	A	A	A	A	B	B	B
Start at same time	C	C	C	C	C	C	Slack	Slack
Delay C by 1 hour	Slack	C	C	C	C	C	C	Slack
Delay C by 2 hour	Slack	Slack	C	C	C	C	C	C

The five red squares represent the 5 hours it takes to complete activity A. The three green squares represent the 3 hours it takes to complete activity B.

The six yellow squares represent the 6 hours it takes to complete activity C. Activity C does not have to start at the same time as activity A because it has some slack time available (2 hours).

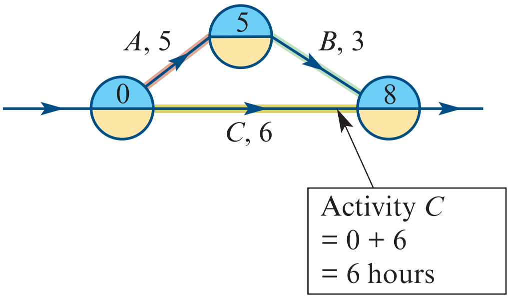
To be able to track the progression of a project network the vertices are drawn as circles, divided into two halves. The completion time is written in the circle as the project works towards completion.



Activity $A + B$
$= 5 + 3$
$= 8$ hours

The next activity will not start until A, B and C are all completed.

Activity C finishes in 6 hours so it must wait 2 hours before the next stage of the project because activities A and B take 8 hours total to complete.



Activity *C* should not be delayed by more than 2 hours because this would cause delays to the project. The next activity requires *B* and *C* to be complete before it can begin.

Critical path analysis

Critical path analysis can be used to identify information about each of the activities that make up a project and the project as a whole. The overall minimum completion time for the project is an important piece of information for project planning and scheduling.

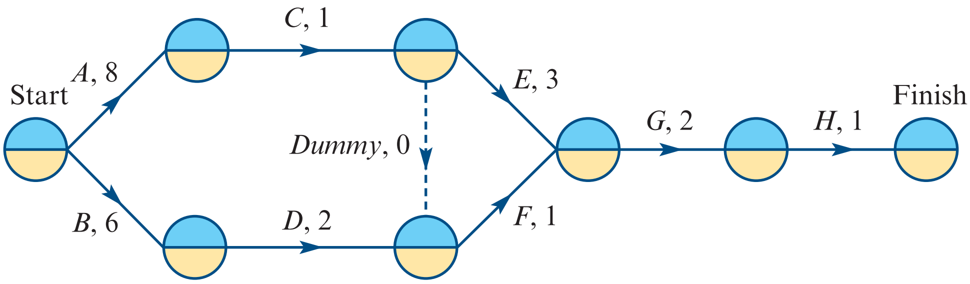
Information about individual activities, such as the **earliest starting time (EST)**, latest starting time (LST) and latest finishing time (LFT) can be identified from critical path analysis. These are the times that are required to ensure the project is completed in minimum completion time.

One of the most important uses of critical path analysis is to identify activities that have no float time and that, if delayed, will cause a delay to the completion of the entire project. These activities are critical to completing the project in the shortest time possible and together they make up what is called the **critical path** of the project.

The process of critical path analysis will be demonstrated using the activity network at the foot of the previous page, with the duration of each activity in days.

Forward scanning

1. Each vertex is drawn as a circle, split into two halves, as shown. The semi-circles in the diagram are coloured blue and orange to help identify which half we are using. The top half (blue) at any vertex will contain the earliest start time (EST) for any activity that begins at that vertex.

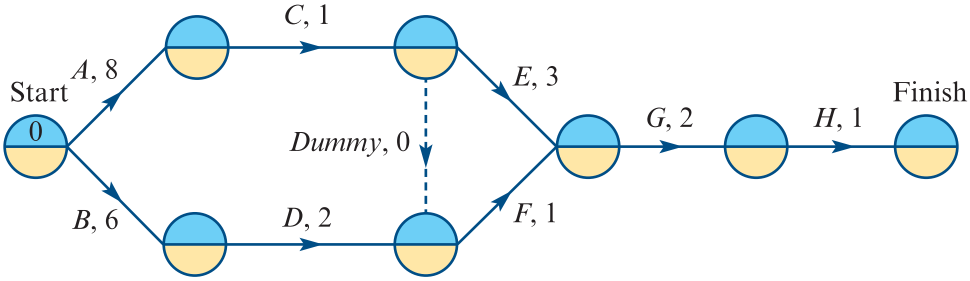


ISBN 978-1-009-25917-0

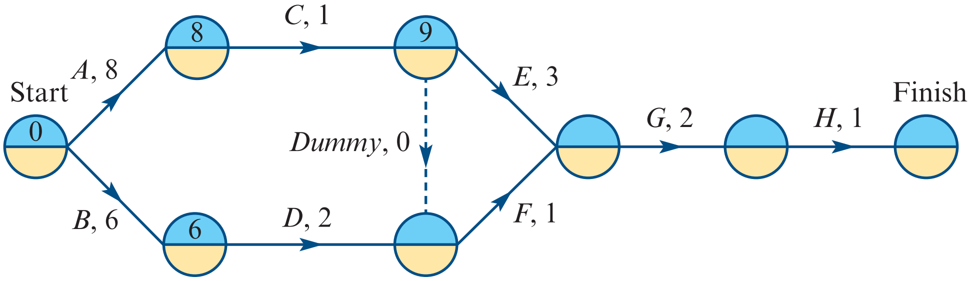
Cambridge University Press

Updated January 2024

2 Put a zero (0) in the top half of the first circle (blue) at the start vertex. This represents the start of the entire project. It also represents the EST for activities *A* and *B* because they start at this vertex.



3 Each activity will have a circle at the vertices at either end of the edge that represents that activity. Add the value in the top half (blue) of the vertex at the start of the activity to the duration of the activity and write the answer in the top half (blue) of the circle at the end of the activity. This is the EST for the activity or activities that follow.



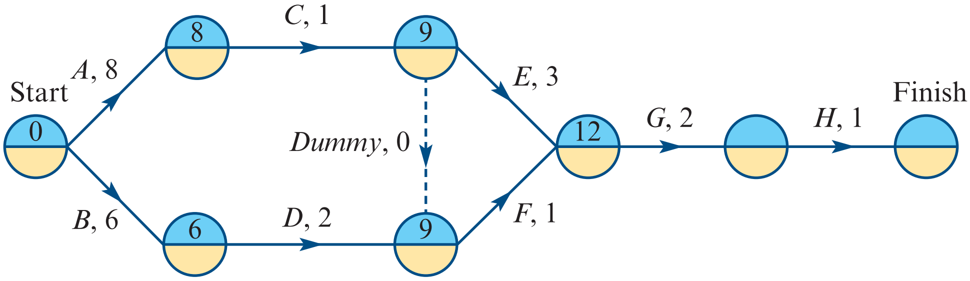
Notes:

1 The vertex at the end of activity *A* has value 0 + 8 = 8. This is the EST for activity *C*.

2 The vertex at the end of activity *B* has value 0 + 6 = 6. This is the EST for activity *D*.

3 The vertex at the end of activity *C* has value 8 + 1 = 9. This is the EST for activity *E* and dummy.

4 If the edges representing more than one activity end at the same vertex, the top half (blue) of the circle at this vertex must contain the *largest* of the possible values because this activity must wait for all predecessor activities to be completed before it can begin.



Notes:

1 The vertex at the end of activity *D* and dummy could be:

- from activity *D*: 6 + 2 = 8
- from dummy activity: 9 + 0 = 9

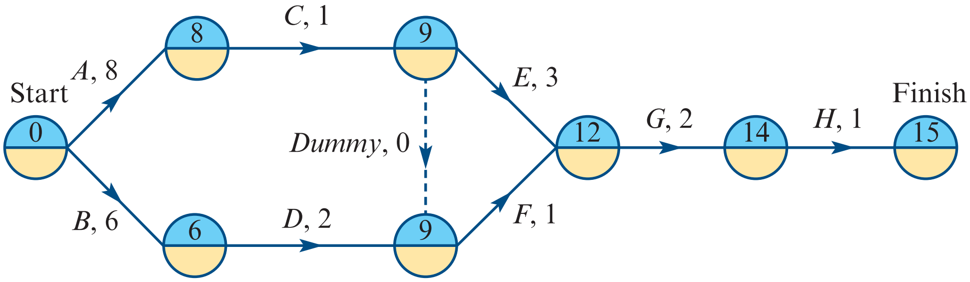
The largest of these options is 9. This is the EST for activity *F*.

2 The vertex at the end of activity *E* and *F* could be:

- from activity *E*: 9 + 3 = 12
- from activity *F*: 9 + 1 = 10

The largest of these options is 12. This is the EST for activity *G*.

5 Continue adding the previous EST value to the duration to calculate the following EST values until the finish vertex is reached.



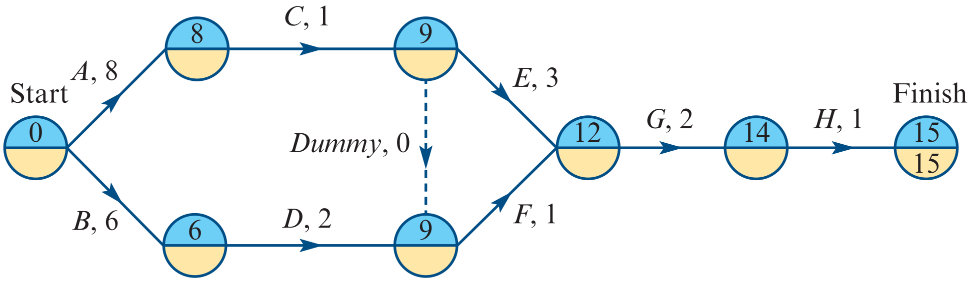
Identifying the minimum completion time for the project

The circle for the final vertex is unique. The top half of the circle (blue) usually contains the EST for the activities that follow, but there are no further activities in this project. The top half (blue) of the final vertex contains the minimum completion time for the project. This project can be completed in a minimum of 15 days.

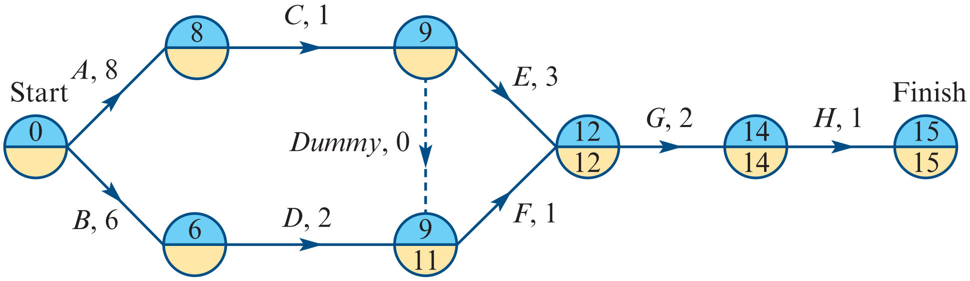
To find out more information about the activities in this project, we will need to complete the **backward scanning** process.

Backward scanning

1 Look at the circle at the finish vertex. Copy the value in the top half (blue) to the bottom half (yellow). The bottom semi-circle contains the latest finishing time (LFT) for any activity that ends at this vertex.
The LFT for activity *H* is 15.



2 For each activity, subtract the duration of the activity from the LFT on the bottom half (yellow) value of the circle at the end of the activity. Write the answer in the bottom half (yellow) of the circle at the start of the activity. This will give the LFT for any activity that ends at this vertex.



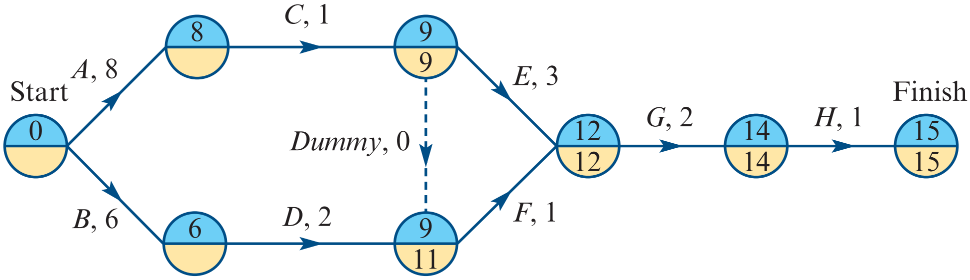
Notes:

1 The vertex at the start of activity H has value $15 - 1 = 14$. This is the LFT for activity G.

2 The vertex at the start of activity G has value $14 - 2 = 12$. This is the LFT for activity E and F.

3 The vertex at the start of activity F has value $12 - 1 = 11$. This is the LFT for activity D and dummy.

3 If the edges representing more than one activity start at the same vertex, the bottom half (yellow) of the circle at this vertex must contain the *smallest* of the possible values. This ensure that the longest of the activities that follow this vertex will have time to be completed.



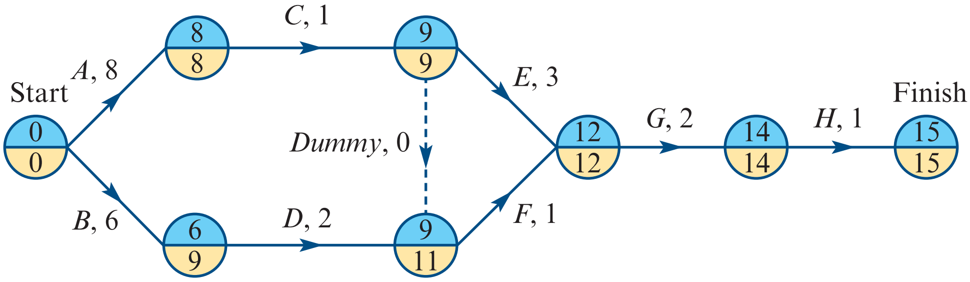
Note:

1 The vertex at the start of activity E and dummy could be:

- from activity E: $12 - 3 = 9$
- from dummy activity: $11 - 0 = 11$

The smallest of these options is 9. This is the LFT for activity C.

4 Complete the backward scanning for all remaining activities.



ISBN 978-1-009-25917-0

Notes:

1. The vertex at the start of activity C has value 9 − 1 = 8. This is the LFT for activity A.
2. The vertex at the start of activity D has value 11 − 2 = 9. This is the LFT for activity B.
3. The value at the start vertex has either:
    - from activity A : 8 − 8 = 0
    - from activity B : 9 − 6 = 3

    The smallest of these options is 0.
4. The circle at the start vertex will always contain a zero in both the top (blue) and bottom (yellow) halves.

Identifying earliest starting time (EST)

The earliest starting time (EST) for an activity is the earliest time, after the project has begun, that the activity can begin. It will always be the value in the top half (blue) of the circle *at the start of the activity*. EST values can be read directly from the forward scanning results.

The EST values for each activity in the example are shown in a table on the right. For example, activity *E* can start, at the earliest, 9 days after the project begins. This assumes that all the preceding activities also begin at their earliest time and that they are completed on time.

Activity	*EST*
A	0
B	0
C	8
D	6
E	9
F	9
G	12
H	14

Identifying latest starting time (LST)

The **latest starting time (LST)** for an activity is the latest time, after the project has begun, that the activity can begin, without extending the minimum overall completion time of the project.

LST values can be easily calculated from the backward scanning results:

LST = LFT − duration of activity

For example, the LFT for activity *D* is the bottom half (yellow) value in the circle at the vertex where activity *D* finishes. This value is 11. The LST for activity *D* is:

LFT − 2 = 11 − 2 = 9.

This means that activity *D*, at the very latest, must start 9 days after the beginning of the project if the project is to be completed in the minimum time possible. The LST calculations and values for each activity are shown in the table.

You might notice that these calculations are exactly the calculations you performed in the backward scanning process. However, even if two activities begin at the same vertex, they might not have the same LST.

Activity	*EST*	*LST calculation*	*LST*
A	0	8 − 8	0
B	0	9 − 6	3
C	8	9 − 1	8
D	6	11 − 2	9
E	9	12 − 3	9
F	9	12 − 1	11
G	12	14 − 2	12
H	14	15 − 1	14

ISBN 978-1-009-25917-0

Identifying float time

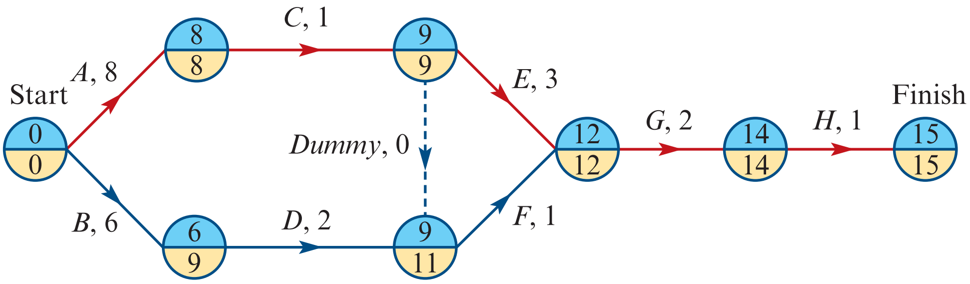
Notice that some of the activities in the table above have the same value for EST and LST. These activities are A, C, E, G, H, and are the activities that make up the critical path for the project that we identified earlier. These activities must start at the EST in order for the project to be completed in the minimum time possible. There is no float time available for these activities. Any delay starting them will extend the overall time for the project.

Other activities have different EST and LST, indicating that these activities have some float time. For example, activity D has an EST of 6 and an LST of 9. Activity D can begin, at the earliest 6 days after the start of the project, and at the latest 9 days after the start of the project, without affecting the minimum time to complete the project. Activity D can be delayed by up to 3 hours.

$$\begin{aligned}\text{For activity } D,\ \text{float} &= \text{LST} - \text{EST}\\ &= 3 \text{ days}\end{aligned}$$

Identifying the critical path

The critical path for the project can be identified by highlighting all the activities that have the same EST and LST. The critical path for this project is highlighted in red on the diagram below.



Critical path

The critical path is the sequence of activities that cannot be delayed without affecting the overall completion time of the project. It is also the largest path through the activity network.

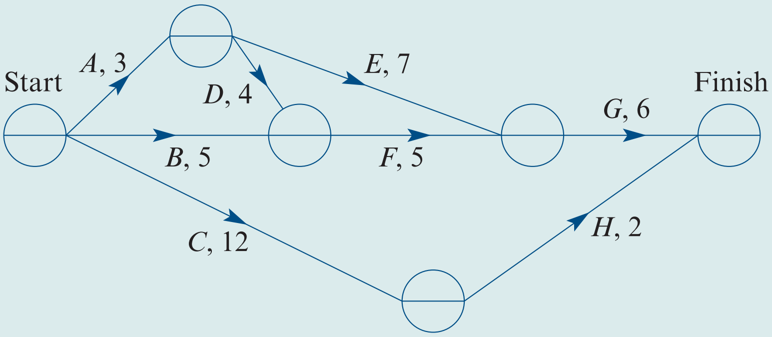
The process for determining the critical path is called **critical path analysis**. In most projects there will be a single critical path running from start to finish, but it is possible for a project to have a critical path that branches. For example, if the time for activity F was increased to 3 days in the project above, then LST for F would become $12 - 3 = 9$ and so F and the dummy activity would also be on the critical path.

Critical path analysis

- Draw a circle divided into two halves (top and bottom) at the vertex of the project network.
- Forward scan: add the value in the top half of the circle at the start of the activity to the duration to give the value in the top half of the circle at the end of the activity.
- Use the largest of the possibilities if there is more than one activity ending at the same vertex.
- Backward scan: subtract the duration of the activity from the value in the bottom half of the circle from the circle at the end of the activity.
- Use the smallest of the possibilities if there is more than one activity that begins at the same vertex.
- Identify EST values as the value in the top half of the circle at the start of each activity.
- Calculate LST values as the value in the bottom half of the circle at the end of the activity minus the duration of that activity.
- Float = LST – EST
- If EST = LST, float will equal zero and the activity is on the critical path/s.
- Highlight all activities that have zero float time to identify the critical path/s.

Example 4 Completing a critical path analysis

The activity network for a project consisting of eight activities, A, B, C, D, E, F, G and H is shown below. The number next to the activity name is the time it takes, in weeks, to complete that activity.



a Complete the forward scanning process to identify the minimum time it will take to complete this project.

b Complete the backward scanning process.

c What is the earliest starting time for activity E?

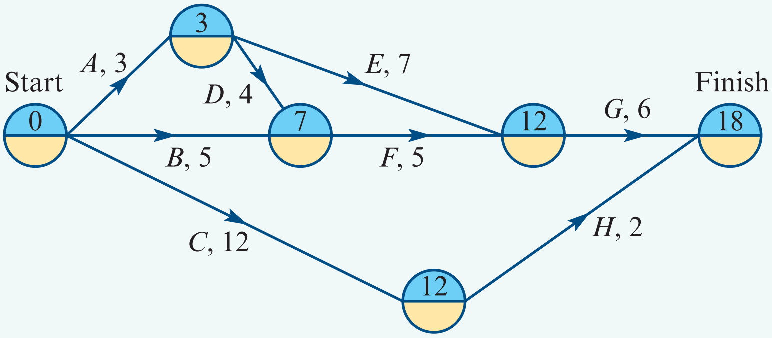
d What is the latest starting time for activity E?

e Identify the critical path for this project.

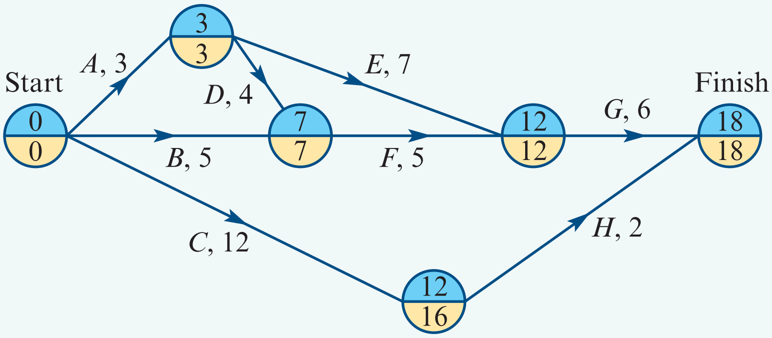
f The person responsible for completing activity E falls sick three weeks into the project. If he will be away from work for two weeks, will this cause the entire project to be delayed?

Solution

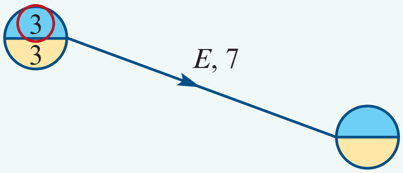
a The forward scanning process results are shown in the diagram.



b The backward scanning process results are shown in the diagram

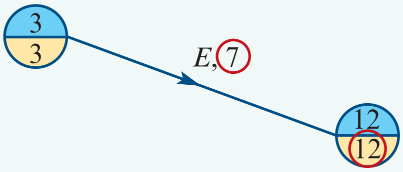


c Earliest starting time is the value in the top half of the circle at the beginning of the activity



The EST for activity E = 3

d Latest starting time is the value in the bottom half of the circle at the end of the activity, subtract the duration.



The LST for activity E = 12 − 7 = 5

e The critical path joins all of the activities that have the same EST and LST, and therefore which have zero float time.

The critical path for this project is A − D − F − G

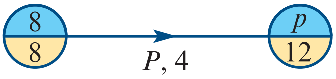
f **i** Calculate the float for activity E. This tells us how long the start of activity E can be delayed, without delaying the entire project.	Float E = LST − EST = 5 − 3 = 2weeks
ii If the float time is more or equal to the delay in the start of activity E, the project will not be affected.	The person will be away for two weeks, starting 3 weeks into the project. This is equal to the float time for activity E and so delaying the start of activity E until the person comes back to work will not affect the overall completion time of the project.

Skill-sheet

Exercise 10B

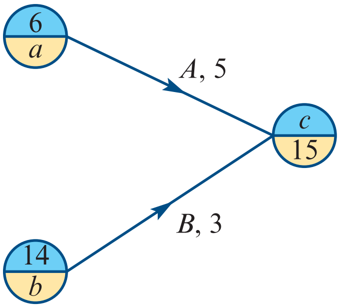
Calculations from elements of an activity network

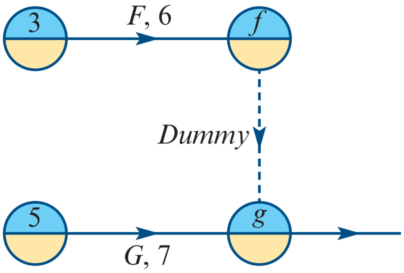
1 Write the value of each pronumeral in the sections of activity networks below.

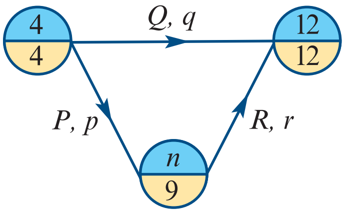
a 

b 

c 

d 

e 

f 

2 Consider the section of an activity network shown in the diagram below.

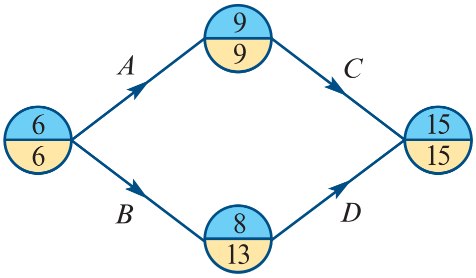
a What is the duration of activity A?

b What is the critical path through this section of the activity network?

c What is the float time of activity B?

d What is the latest time that activity D can start?

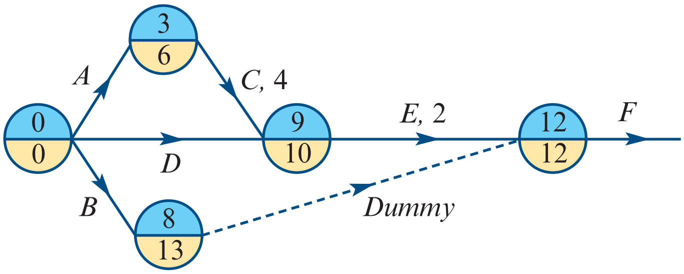
e What is the duration of activity D?



ISBN 978-1-009-25917-0

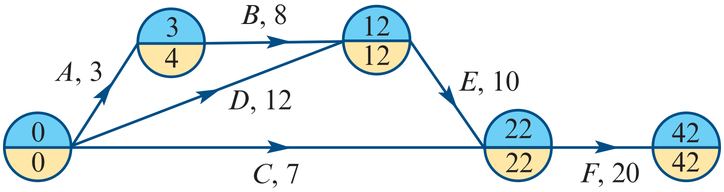
3 Consider the section of an activity network shown in the diagram below.

- **a** What is the duration of activity *B*?
- **b** What is the latest start time for activity *E*?
- **c** What is the earliest time that activity *E* can start?
- **d** What is the float time for activity *E*?
- **e** What is the duration of activity *A*?
- **f** What is the duration of activity *D*?



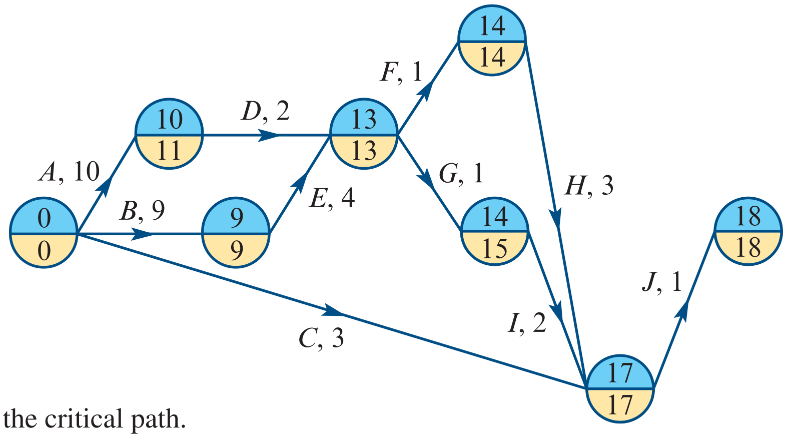
Completing a critical path analysis from a given activity network

4 An activity network is shown in the diagram below.



- **a** Write the critical path for this project.
- **b** Calculate the float times for non-critical activities.

5 An activity network is shown in the diagram below.



- **a** Write the critical path.
- **b** Write the float times for all non-critical activities.

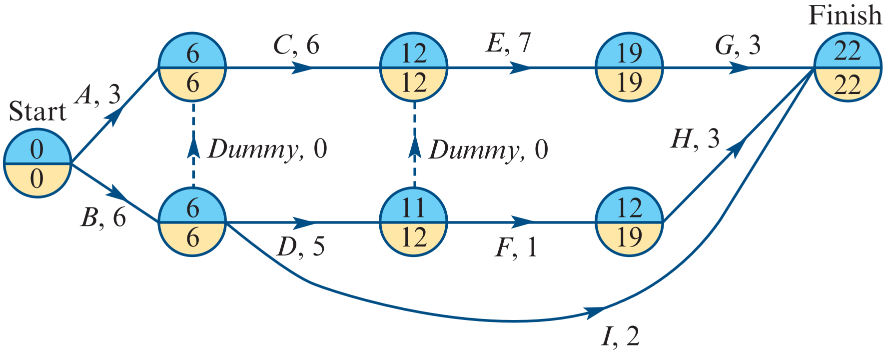
ISBN 978-1-009-25917-0

6 A precedence table and activity network for a project are shown below. The precedence table is incomplete.

a Complete the table on the right.

b Write the critical path for this project.

Activity	*Duration (weeks)*	*Immediate predecessors*
A	3	–
B	6	–
C	6	
D		*B*
E	7	
F	1	*D*
G		*E*
H	3	
I	2	*B*



Completing a critical path analysis from a precedence table only

Example 4

7 The precedence table for a project is shown.

a Draw an activity network for this project.

b Complete the critical path analysis to calculate the EST and LST for each activity.

c Write the critical path of this project.

d What is the minimum time required to complete the project?

Activity	*Duration (weeks)*	*Immediate predecessors*
P	4	–
Q	5	–
R	12	–
S	3	*P*
T	6	*Q*
U	3	*S*
V	4	*R*
W	8	*R*, *T*, *U*
X	13	*V*
Y	6	*W*, *X*

8 The precedence table for a project is shown.

a Draw an activity network for this project.

b Complete the critical path analysis to calculate the EST and LST for each activity.

c Write the critical path of this project.

d What is the minimum time required to complete the project?

Activity	*Duration (weeks)*	*Immediate predecessors*
I	2	–
J	3	–
K	5	–
L	4	I
M	8	J, N
N	1	K
O	6	L, M
P	6	J, N
Q	7	J, N
R	5	K
S	1	O
T	9	Q, R

10C Altering completion times

Altering completion times

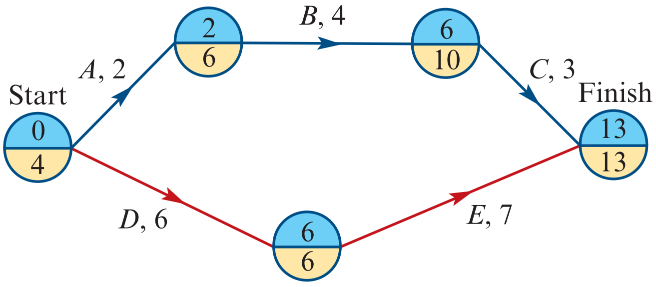
The minimum time it takes to complete a project depends upon the time it takes to complete the individual activities of the project, and upon the predecessors that each activity has. Critical path analysis can be completed to find the overall minimum completion time.

Sometimes, the managers of a project might arrange for one or more activities within the project to be completed in a shorter time than originally planned. Changing the conditions of an activity within a project, and recalculating the minimum completion time for the project, is called **crashing**.

An individual activity could be altered by employing more staff, sourcing alternative materials or simply because weather or other factors allow the activity to be completed in a shorter time than usual.

A simple example

A simple activity network is shown in the diagram on the right. The forwards and backwards scanning processes have been completed and the critical path has been determined. The critical path is shown in red on the diagram.



The minimum time for completion is currently 13 hours. In order to reduce this overall time, the manager of the project should try to complete one, or more, of the activities in a shorter time than normal. Reducing the time taken to complete activity A, B or C would not achieve this goal however. These activities are not on the critical path and so they already have slack time. Reducing their completion time will not shorten the overall time taken to complete the project.

Activity D and E, on the other hand, lie on the critical path. Reducing the duration of these activities will reduce the overall time for the project. If activity D was reduced in time to 4 hours instead, the project will be completed in 11, not 13, hours.

Completion times and cost

Shortening the completion time for any individual activity could result in an extra cost for the project. In the simple example above, the cost of reducing the completion time of activity D by 1 hour is \$150, while the cost of reducing the completion time of activity E by 1 hour is \$18.

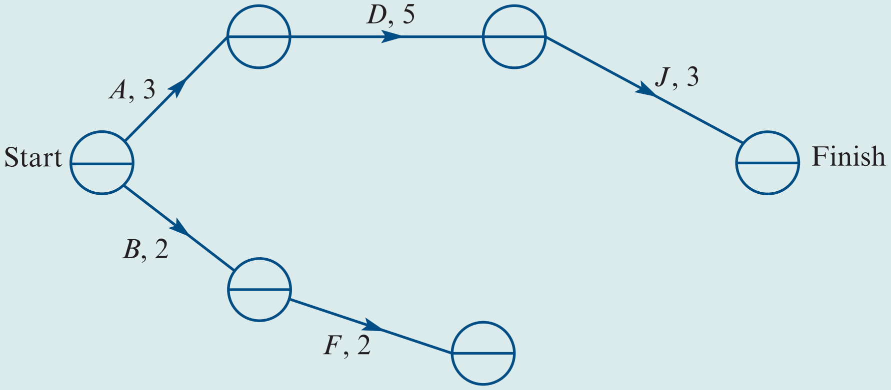
Clearly it is best to reduce the completion time of the activity that will cost the least.

Example 5 Completion times and cost

The project network below shows the proposed production schedule for the creation of a web series. The production schedule includes ten activities with their duration, in days, shown in the table below.

Task	*Task description*	*Duration (days)*	*Immediate predecessors*
A	Location Scouting	3	-
B	Script Breakdown	2	-
C	Costume Design	3	B
D	Set Design	5	A
E	Casting	6	A
F	Shot List	2	B
G	Shooting Schedule	3	F
H	Rehearsals	4	C, E
I	Filming	2	G, H
J	Photography	3	D

a Complete the project network below, showing all activities and durations.



b Determine the minimum completion time, in days, to produce the web series.

c Write the critical path for this project.

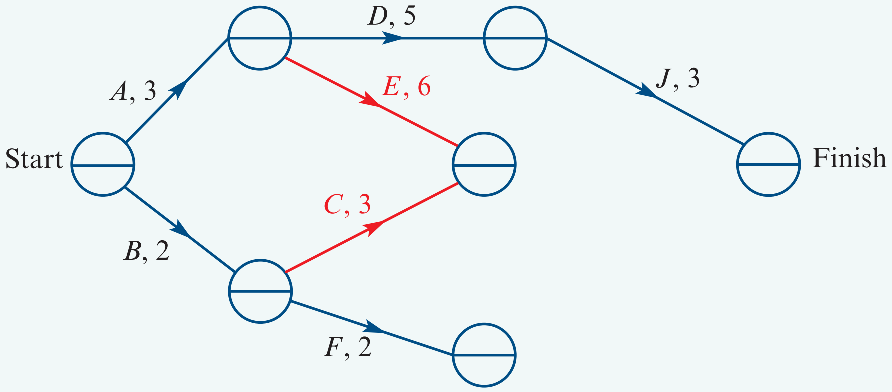
d What is the earliest start time that the script breakdown (Task B) may be completed, and what is the latest start time?

e Due to outside delays, the set design (Task D) is going to take longer than expected to complete. How long can this task be delayed without effecting the overall completion time of the project?

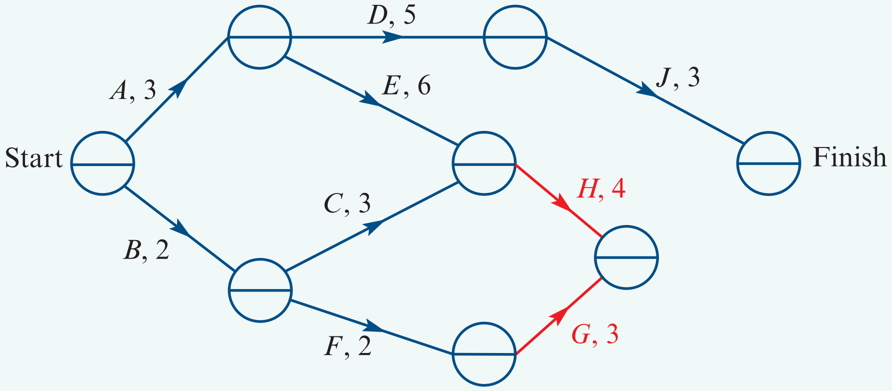
f More funding becomes available and can be used to hire more people to work on the production. These people will enable the location scouting (Task A) to be completed in one day less and casting (Task E) to be reduced by 4 days. Determine the minimum time for the project to be completed now that these activities can be reduced in time. Comment on any changes to the critical path.

Solution

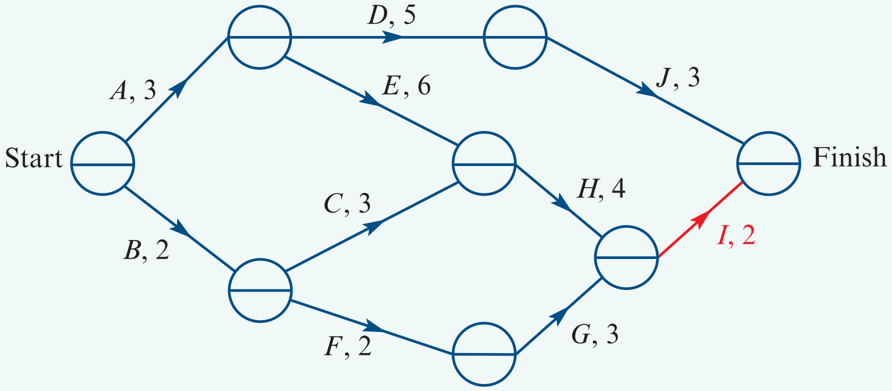
a **1** Activity C follows from B. Activity E follows from A. Activities C and E must finish together as both are predecessors for H



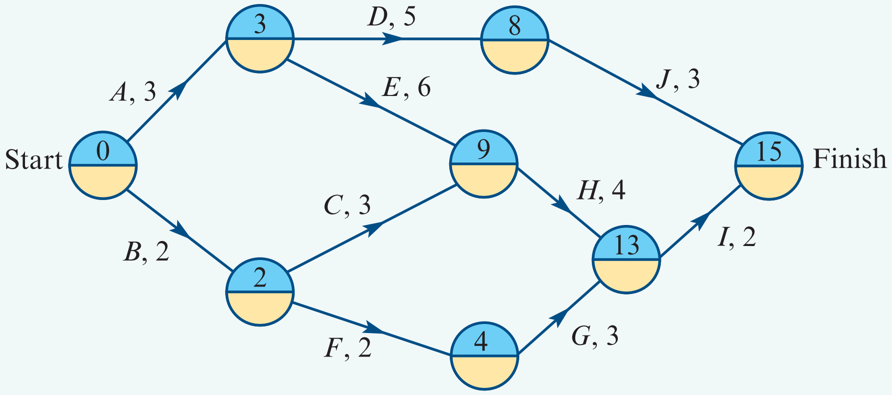
2 Activity H follows C and E. Activity G follows F. Activities G and H must finish together as both are predecessors for I



3 Activity I follows G and H. Activity I is not a predecessor for any activity so must go to the Finish



b **1** Perform forwards scanning to determine the EST for all activities and the minimum time to complete the project.



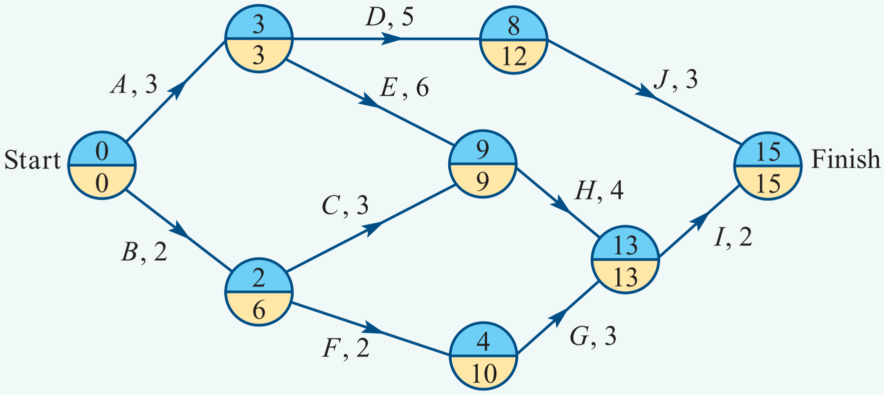
2 Write your answer, with correct units.

The minimum completing time is 15 days.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

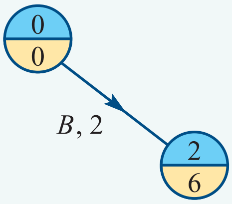
c **1** Perform backwards scanning to determine the LST for all activities.



2 Follow the activities that have circles with the same EST and LST.

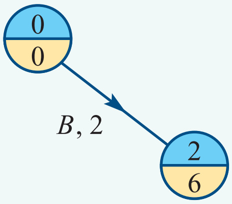
The critical path for this project is A – E – H – I.

d **1** The EST is the value in the top half of the circle at the beginning of the activity.



The EST for Activity B is Day 0.

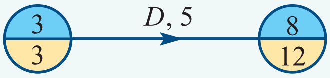
2 The LST is the value in the bottom half of the circle at the end of the activity, subtract the duration.



$6 - 2 = 4$

The LST for Activity B is Day 4.

e **1** Look at Activity *D*. Determine its EST and LST.



EST = 3

LST = 12 – 5 = 7

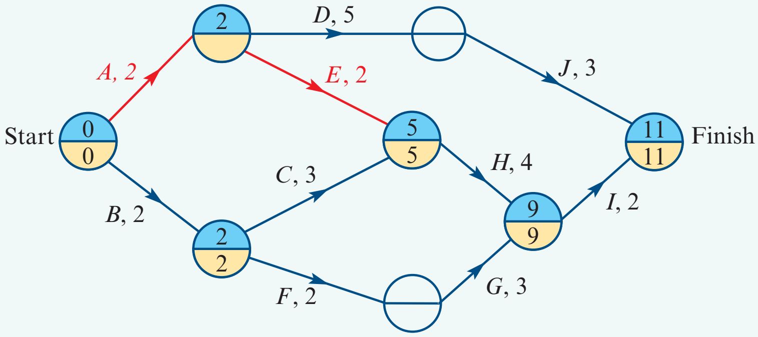
2 Calculate float time, to determine the amount of allowed delay.

Float = LST – EST

= 7 – 3

= 4

f **1** Reduce Task A by 1 day and Task E by 4 days.



2 Check for change in minimum completion time and critical path.

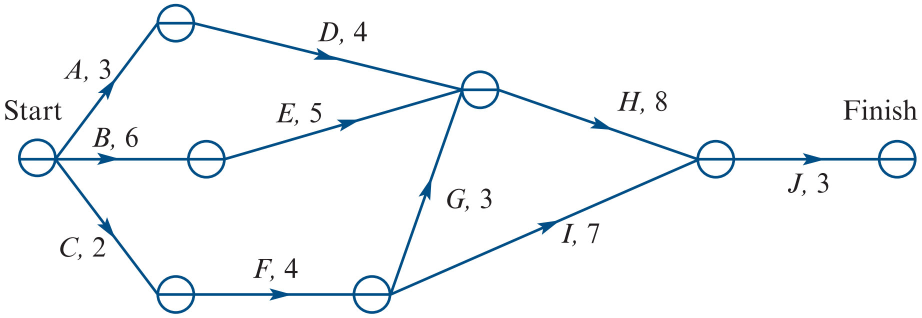
New minimum completion time = 11 days
Overall minimum completion time decreased by 4 days
New critical path $= B - C - H - I$

Skill-sheet

Exercise 10C

Altering completion times

1 The activity network for a project is shown in the diagram below. The duration for each activity is in hours.



a Identify the critical path for this project.

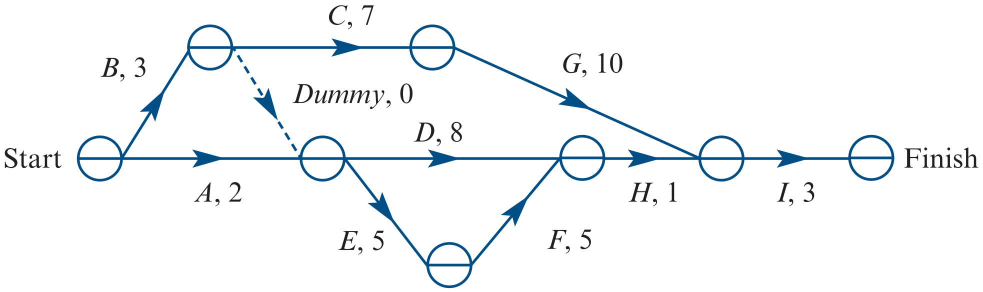
b What is the maximum number of hours that the completion time for activity E can be reduced without changing the critical path of the project?

c What is the maximum number of hours that the completion time for activity H can be reduced without affecting the critical path of the project?

d Every activity can be reduced in duration by a maximum of 2 hours. If every activity was reduced by the maximum amount possible, what is the minimum completion time for the project?

ISBN 978-1-009-25917-0

2 The diagram shows an activity network for a renovation project with the duration of each activity in hours.



a What is the critical path for the renovation?

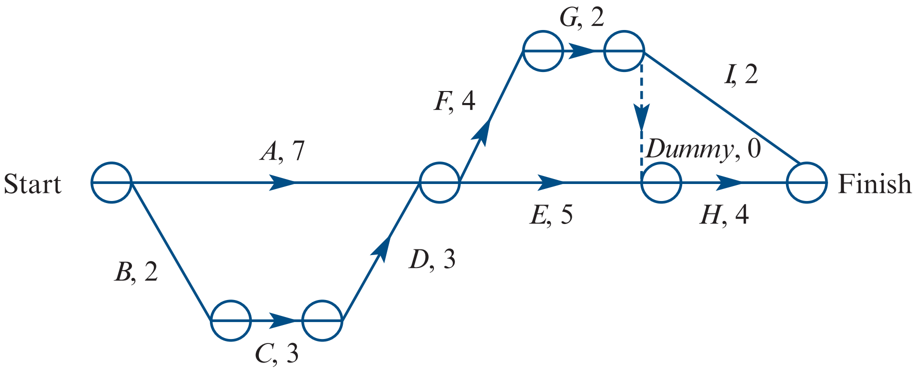
b If activity *A* is halved in duration, how much time is saved overall?

c If the duration of activities *C* and *G* can be reduced, what is the greatest possible reduction in the completion of the renovation?

Completion times and cost

Example 5

3 The diagram shows an activity network for a project, with the duration of the activities in hours.



a Identify the shortest time for completion of the project.

b Identify the project’s critical path.

c Activity *E* can be reduced by two hours. Why will this not affect the completion of the project?

d Activity *F* can be reduced at a cost of \$60 per hour. Activity *D* can be reduced at a cost of \$50 per hour. What is the cost of reducing the project’s completion time as much as possible? Activity *E* will have duration reduced to 3 hours.

Key ideas and chapter summary

Assign-ment

Activity network	An **activity network** is a directed graph that shows the required order of completing individual activities that make up a project.
Immediate predecessor	If activity *A* is an **immediate predecessor** to activity *B*, activity *A* must be completed before activity *B* can begin.
Precedence table	A **precedence table** is a table that records the activities of a project and their immediate predecessors. Precedence tables can also contain the duration of each activity.
Dummy activity	A **dummy activity** has zero cost. It is required if two activities share some, but not all, of the same immediate predecessors. It allows the network to show all precedence relationships in a project correctly.
Earliest starting time (EST)	**EST** is the earliest time an activity in a project can begin.
Latest starting time (LST)	**LST** is the latest time an activity in a project can begin, without affecting the overall completion time for the project.
Float (slack) time	**Float (slack) time** is the difference between the latest starting time and the earliest starting time. Float = LST − EST The float time is sometimes called the slack time. It is the largest amount of time that an activity can be delayed without affecting the overall completion time for the project.
Forward scanning	**Forward scanning** is a process of determining the EST for each activity in an activity network. The EST of an activity is added to the duration of that activity to determine the EST of the next activity. The EST of any activity is equal to the largest forward scanning value determined from all immediate predecessors.
Backward scanning	**Backward scanning** is a process of determining the LST for each activity in an activity network. The LST of an activity is equal to the LST of the activity that follows, minus the duration of the activity.
Critical path	The **critical path** is the series of activities that cannot be delayed without affecting the overall completion time of the project. Activities on the critical path have no slack time. Their EST and LST are equal.
Critical path analysis	**Critical path analysis** is a project planning method in which activity durations are known with certainty.

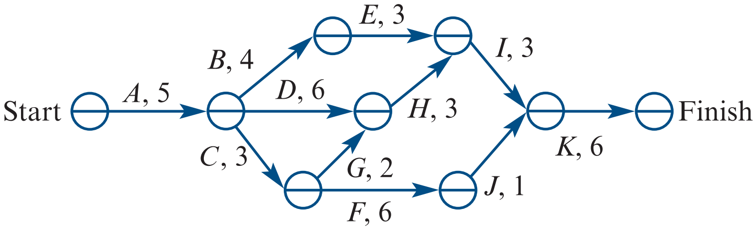
Skills check

Having completed this chapter you should be able to:

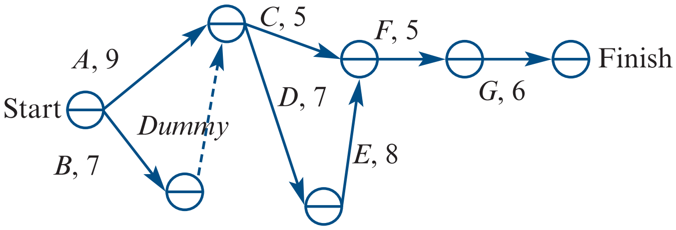
- create an activity network from a precedence table
- write a precedence table from an activity network
- decide when to use dummy activities in an activity network
- use forward scanning to determine the earliest starting time of activities in an activity network
- use backward scanning to determine the latest starting time of activities in an activity network
- determine the float time for activities in an activity network
- determine the overall minimum completion time for a project using critical path analysis
- determine the critical path for an activity network

Short-answer questions

1 This activity network is for a project where the component times in days are shown. Determine the critical path for the project network.

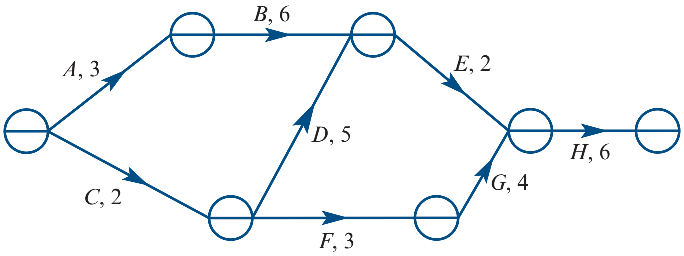


2 The activity network shown represents a project development with activities and their durations (in days) listed on the edges of the graph. Note that the dummy activity takes zero time.



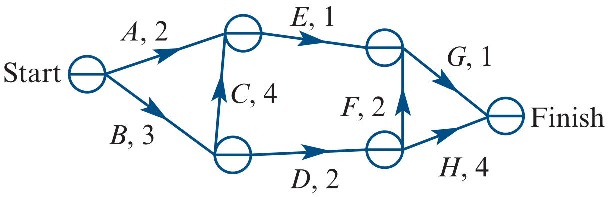
What is the earliest time (in days) that activity F can begin?

3 The edges in this activity network correspond to the tasks involved in the preparation of an examination.



The numbers indicate the time, in weeks, needed for each task. Determine the total number of weeks needed for the preparation of the examination.

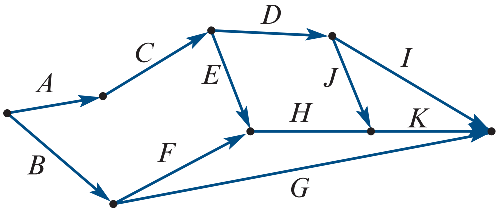
4 The activity network represents a manufacturing process with activities and their duration (in hours) listed on the edges of the graph.



What is the earliest time (in hours) after the start that activity G can begin?

5 This network represents a project development with activities listed on the edges of the graph.

Which of the following statements must be true?



A A must be completed before B can start.

B A must be completed before F can start.

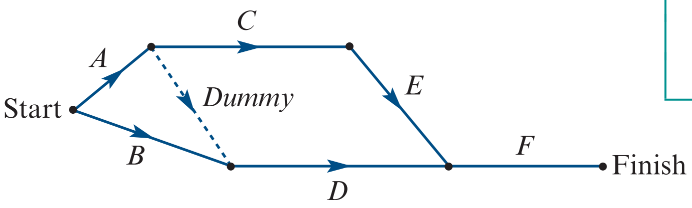
C E and F must start at the same time.

D E and F must finish at the same time.

E E cannot start until A is finished.

6 Complete the table that shows the immediate predecessors of the activities in the project network shown below.

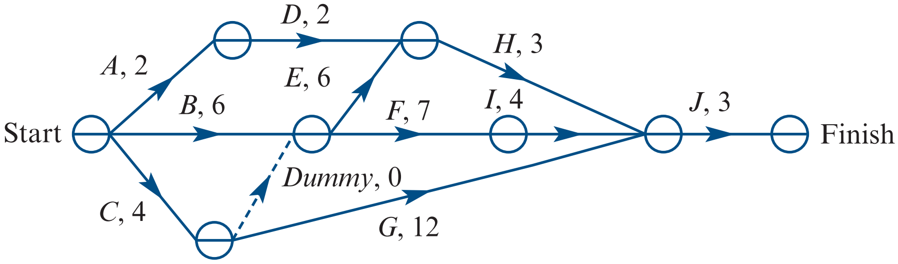
Activity	*Immediate predecessors*
A	
B	
C	
D	
E	
F	



ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

7 The activity network below shows the activities required to complete a particular project and the durations, in hours, of those activities.



a Determine the earliest starting time for activity I.

b Determine the latest starting time for activity E.

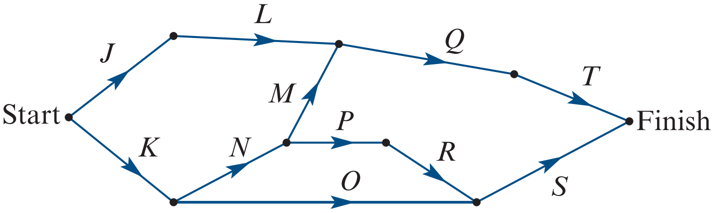
c What is the float time of activity H?

d Write the critical path for the project.

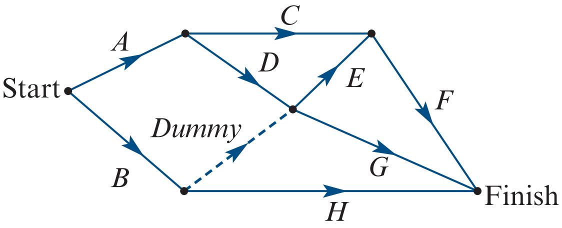
Extended-response questions

1 Write a precedence table for the activity networks shown below.

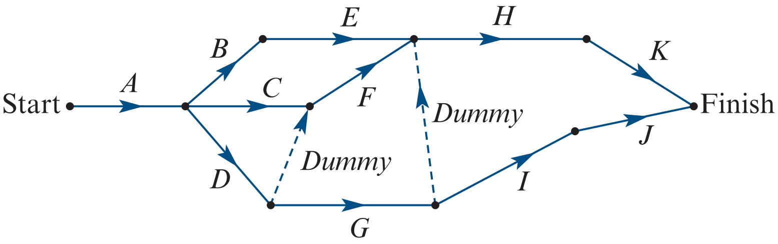
a



b



c



ISBN 978-1-009-25917-0

2 Draw an activity network for each of the precedence tables below.

a

Activity	*Immediate predecessors*
K	–
L	–
M	K
N	M
O	N, L
P	O
Q	P
R	M
S	R, Q

b

Activity	*Immediate predecessors*
A	–
B	–
C	–
D	B
E	A, D
F	E, C
G	F
H	G
I	E, C
J	G
K	H, I

c

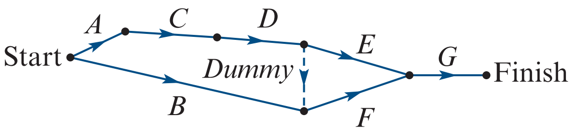
Activity	*Immediate predecessors*
P	–
Q	–
R	P
S	Q
T	Q
U	R, S
V	R, S, T

d

Activity	*Immediate predecessors*
A	–
B	A
C	A
D	B, C
E	C
F	E
G	D
H	F, G
I	H
J	I

3 The following activity network shows the activities in a project to repair a dent in a car panel. The activities are listed in the table on the right.

Activity	*Description*
A	Remove panel
B	Order component
C	Remove broken component
D	Pound out dent
E	Repaint
F	Install new component
G	Replace panel



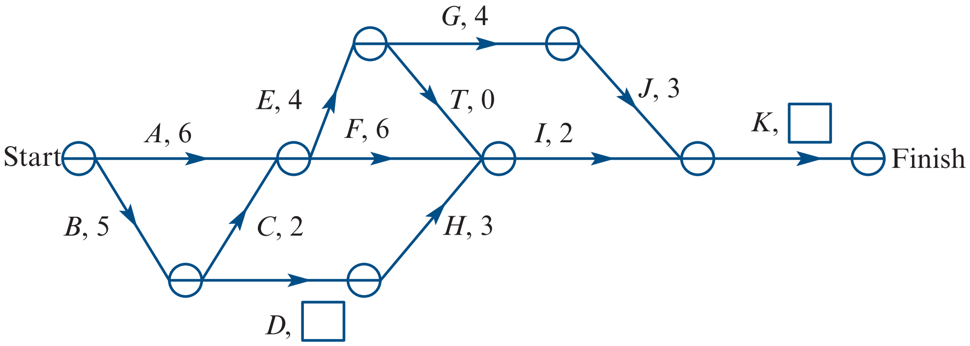
a Which activity or activities are the immediate predecessors of the event 'remove broken component'?

b Which activities are the immediate predecessors of the activity 'install new component'?

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

4 All the activities and their durations (in hours) in a project at the quarry are shown in the network diagram below. The least time required for completing this entire project is 30 hours.



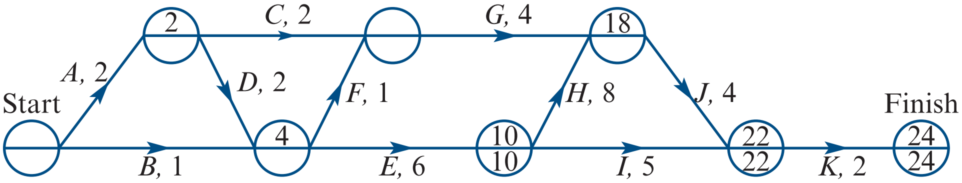
For each activity in this project, the table below shows the completion time, the earliest starting time and the latest starting time.

Activity	*Completion time (hours)*	*Earliest starting time (hours)*	*Latest starting time (hours)*
A	6	0	
B	5	0	0
C	2	5	5
D		5	9
E	4	7	7
F	6	7	
G	4	11	11
H	3	9	13
I	2	13	16
J	3	15	15
K		18	18

a Complete the missing times in the table.

b Write the critical path for this project.

5 The assembly of machined parts in a manufacturing process can be represented by the following network. The activities are represented by the letters on the arcs and the numbers represent the time taken (in hours) for the activities scheduled.



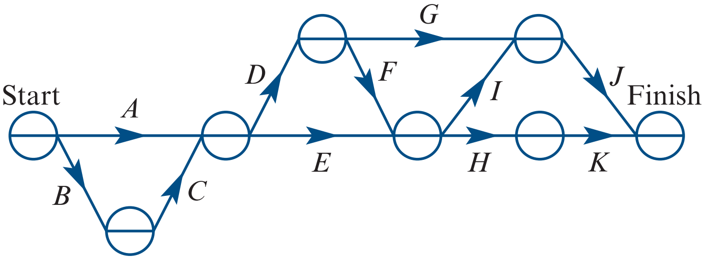
Activity	A	B	C	D	E	F	G	H	I	J	K
EST	0	0	2	2	4	4		10	10	18	22

a The earliest start times (EST) for each activity except G are given in the table. Complete the table by finding the EST for G.

b What is the shortest time required to assemble the product?

c What is the float (slack time) for activity I?

6 In laying a pipeline, the various jobs involved have been grouped into a set of specific tasks A–K, which are performed in the precedence described in the network below.



a List all the task(s) that must be completed before task E is started.

ISBN 978-1-009-25917-0

Cambridge University Press

Updated January 2024

The durations of the tasks are given in Table A below.

b Use the information in Table A to complete Table B.

Table A Task durations

Task	*Normal completion time (months)*
A	10
B	6
C	3
D	4
E	7
F	4
G	5
H	4
I	5
J	4
K	3

Table B Starting times for tasks

Task	*EST*	*LST*
A	0	0
B	0	
C	6	7
D	10	10
E		11
F	14	14
G	14	18
H	18	20
I	18	
J	23	23
K	22	24

c For this project:

i write the critical path

ii determine the length of the critical path (that is, the earliest time the project can be completed).

d If the project managers are prepared to pay more for additional labour and machinery, the time taken to complete task *A* can be reduced to 8 months, task *E* can be reduced to 5 months and task *I* can be reduced to 4 months.

Under these circumstances:

i what would be the critical path(s)?

ii how long would it take to complete the project?

Glossary

A

Activity (CPA) [p. 418] A task to be completed as part of a project. Activities are represented by the edges in the project diagram.

Activity network [p. 418] A weighted directed graph that shows the required order of completion of the activities that make up a project. The weights indicate the durations of the activities they represent.

Adjacency matrix [p. 206] A square matrix showing the number of edges joining each pair of vertices in a graph.

Algorithm [p. 236] A step-by-step procedure for solving a particular problem that involves applying the same process repeatedly. Examples include Prim's algorithm and the Hungarian algorithm.

Allocation [p. 381] The process of assigning a series of tasks to different members of a group in a way that enables the tasks to be completed for the minimum time or cost.

Amortisation [p. 314] The repayment of a loan or an investment with regular payments made over a period of time.

Amortisation table [p. 315] A table that charts the amortisation (repayment) of a reducing balance loan or annuity on a step-by-step (payment-by payment) basis.

Annuity [p. 330] A compound interest investment from which regular payments are made.

Arc [p. 377] An edge in a directed graph (digraph) can also be called an arc.

Arithmetic sequence [p. 111] A sequence where each term is generated by adding or subtracting a fixed amount to the term before it.

Association [p. 6] Association is a general term used to describe the relationship between variables.

B

Backward scanning [p. 429] The process of determining the LST for each activity in a project activity network.

Balance [p. 312] The amount owed or accrued after a period of time.

Bipartite graph (bigraph) [p. 379] A graph whose set of vertices can be split into two subsets, X and Y, in such a way that each edge of the graph joins a vertex in X and a vertex in Y.

Bivariate data [p. 3] Data in which each observation involves recording information about two variables for the same person or thing. An example would be data recording the height and weight of the children in a preschool.

Bridge [p. 206] A single edge in a connected graph that, if removed, leaves the graph disconnected.

C

Capacities (flow network) [p. 393] The weights of the directed edges in a flow network. They give the maximum amount that can move between the two points in the flow network represented by these vertices in a particular time interval. This could be, for example, the maximum amount of water in litres per minute or the maximum number of cars per hour.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Categorical variable [p. 6] Categorical variables are used to represent characteristics of individuals, for example, place of birth, house number.

Causality [p. 46] The existence of an association where change in one variable directly influences the change in another variable.

Centring [p. 271] If smoothing takes place over an even number of data values, the smoothed values do not align with an original data value. A second stage of smoothing is carried out to centre the smoothed values at an original data value.

Coefficient of determination (r^2) [p. 42] A coefficient which gives a measure of the predictive power of a regression line. It gives the percentage of variation in the response variable that can be explained by the variation in the explanatory variable.

Common difference [p. 112] The fixed amount added to each term in an arithmetic sequence to generate the next term of that sequence.

Common ratio [p. 126] The fixed amount multiplied to each term of a geometric sequence.

Complete graph [p. 201] A graph with edges connecting all pairs of vertices.

Compound interest [p. 161] Where the interest paid on a loan or investment is added to the principal and subsequent interest is calculated on the total.

Compound interest investments with periodic repayments [p. 343] Investments to which additions are made to the principal on a regular basis, also known as 'adding to the principal'.

Compounding period [p. 299] The time period for the calculation of interest for an investment or loan. Typical compounding periods are yearly, quarterly, monthly, fortnightly, weekly or daily.

Connected graph [p. 205] A graph in which there is a path between every pair of vertices.

Convergence [p. 181] The situation where the terms of a sequence decrease or increase towards a constant value.

Correlation coefficient r [p. 34] A statistical measure of the strength of the linear association between two numerical variables.

Cost matrix [p. 381] A table that contains the cost of allocating objects from one group, such as people, to objects from another group, such as tasks. The cost can be money, or other factors such as the time taken to complete the project.

Critical path [p. 427] The project path that has the longest completion time.

Critical path analysis [pp. 427, 432] A project planning method in which activity durations are known with certainty.

Cut [p. 397] A line dividing a directed (flow) graph into two parts in a way that separates all 'sinks' from their 'sources'.

Cut capacity [p. 398] The sum of the capacities of the arcs passing through the cut that represents flow from the source to the sink. Edges that represent flow from the sink to the source do not contribute to the capacity of the cut.

Cycle (graphs) [p. 221] A walk with no repeated vertices that starts and ends at the same vertex.

Cycle (time series) [p. 261] Periodic movement in a time series but over a period greater than a year.

D

Data [p. 2] Information collected about a situation. One particular piece of information is called a datum.

Degree of a vertex (deg(A)) [p. 195] The number of edges attached to the vertex. The degree of vertex A is written as deg(A).

Depreciation [p. 151] The reduction in value of an item over time.

Deseasonalise [p. 276] The process of removing seasonality in time series data.

Dijkstra's algorithm [p. 236] An algorithm for finding the shortest path between two vertices in a weighted graph. Pronounced 'Di-stra': 'Di' as in 'die' and 'stra' as in 'car'.

Directed graph (digraph) [p. 377] A graph or network in which directions are associated with each of the edges.

Disconnected graphs [p. 201] A graph is disconnected if it has at least one pair of vertices between which there is no path.

ISBN 978-1-009-25917-0

Cambridge University Press

Updated January 2024

Dummy activity [p. 421] An artificial activity of zero time duration added to a project diagram to ensure that all predecessor activities are properly accounted for.

E

Earliest starting time (EST) [p. 427] The earliest time an activity in a project can be started.

Edge [p. 195] A line joining one vertex in a graph or network to another vertex or itself (a loop).

Effective interest rate [p. 301] Used to compare the interest paid on loans (or investments) with the same annual nominal interest rate r but with different compounding periods (daily, monthly, quarterly, annually, other).

Eulerian graph [p. 226] A graph that has a closed trail (starts and finishes at the same vertex) that includes every edge only once. To be Eulerian, a graph must be connected and have all vertices of an even degree.

Eulerian trail [p. 226] A walk in a graph or network that includes every edge just once (but does not start and finish at the same vertex).

To have an Eulerian trail, a network must be connected and have exactly two vertices of odd degree, with the remaining vertices having even degree.

Euler's formula [p. 215] The formula $v - e + f = 2$, which relates the number of vertices, edges and faces in a connected graph.

Explanatory variable [p. 3] When investigating associations in bivariate data, the explanatory variable (EV) is the variable used to explain or predict the value of the response variable (RV).

Extrapolation [p. 67] Using a mathematical model to make a prediction *outside* the range of data used to construct a model.

F

Face [p. 215] An area in a graph or network that can only be reached by crossing an edge. One such area is always the area surrounding a graph.

Finance solver [p. 305] A computer/ calculator application that automates the computations associated with analysing a reducing balance loan, an annuity or an annuity investment.

Flat-rate depreciation [p. 151] Depreciation where the value of an item is reduced by the same amount each year. Flat-rate depreciation is equivalent, but opposite, to simple interest.

Float (slack) time [p. 426] The amount of time available to complete a particular activity that does not increase the total time taken to complete the project.

Flow [p. 393] The movement of something from a source to a sink.

Forward scanning [p. 427] The process of determining the EST for each activity in a project activity network.

G

Geometric decay [p. 170] When a recurrence rule involves multiplying by a factor less than one, the terms in the resulting sequence are said to decay geometrically.

Geometric growth [p. 170] When a recurrence rule involves multiplying by a factor greater than one, the terms in the resulting sequence are said to grow geometrically.

Geometric sequence [p. 126] A sequence where each term is generated by multiplying or dividing a term by a fixed amount to generate the next.

Grand sum [p. 6] The sum of the row totals, or the column totals, in a two-way frequency table. The grand sum is also the number of data points contained in the frequency table.

Graph or network [p. 194] A collection of points called vertices and a set of connecting lines called edges.

H

Hamiltonian cycle [p. 226] A Hamiltonian path that starts and finishes at the same vertex.

Hamiltonian path [p. 226] A path through a graph or network that passes through each vertex exactly once. It may or may not start and finish at the same vertex.

Hungarian algorithm [p. 381] An algorithm for solving allocation (assignment) problems.

I

Immediate predecessor [p. 418] An activity that must be completed immediately before another one can start.

Intercept *see* ***y*-intercept**

Interest [p. 148] The amount of money paid (earned) for borrowing (lending) money over a period of time.

Interest rate [p. 148] The rate at which interest is charged or paid. Usually expressed as a percentage of the money owed or lent.

Interpolation [p. 67] Using a regression line to make a prediction *within* the range of values of the explanatory variable.

Irregular (random) fluctuations [p. 263] Unpredictable fluctuations in a time series. Always present in any real world time series plot.

L

Latest start time (LST) [p. 431] The latest time an activity in a project can begin, without affecting the overall completion time for the project.

Least squares method [p. 57] One way of finding the equation of a regression line. It minimises the sum of the squares of the residuals. It works best when there are no outliers.

Limit [p. 178] The value towards which the terms of a converging sequence approach.

Linear decay [p. 170] When a recurrence rule involves subtracting a fixed amount, the terms in the resulting sequence are said to decay linearly.

Linear growth [p. 170] When a recurrence rule involves adding a fixed amount, the terms in the resulting sequence are said to grow linearly.

Linear regression [p. 57] The process of fitting a straight line to bivariate data.

Long-term steady state [p. 178] If a sequence is generated from a situation involving combined geometric decay and arithmetic growth, the terms of the sequence, in the long term, will be constant.

Loop [p. 200] An edge in a graph or network that joins a vertex to itself.

M

Matrix [p. 206] A rectangular array of numbers or symbols set out in rows and columns within square brackets (plural: matrices).

Maximum flow (graph) [p. 393] The capacity of the ‘minimum’ cut.

Minimum cut (graph) [p. 399] The cut through a graph or network with the minimum capacity.

Minimum spanning tree [p. 359] The spanning tree of minimum length. For a given connected graph, there may be more than one minimum spanning tree.

Modelling [p. 146] The use of a mathematical rule or formula to represent real-life situations.

Moving averages [p. 268] In a time series, a moving average is a method used to smooth the data, whereby each original data value is replaced by the mean of itself and a number of data values on either side.

Multiple edge [p. 200] Where more than one edge connects the same two vertices in a graph.

N

Network [p. 233] A set of points called vertices and connecting lines called edges, enclosing and surrounded by areas called faces.

Nominal interest rate [p. 299] The annual interest rate for a loan or investment that assumes the compounding period is 1 year. If the compounding period is less than a year, for example monthly, the actual or effective interest rate will be greater than r.

Numerical variable [p. 15] A variable used to represent quantities that are counted or measured. For example, the number of people in a queue, the heights of these people in cm.

O

Outliers [p. 35] Data values that appear to stand out from the main body of a data set.

P

Path [p. 221] A walk with no repeated vertices. *See also* **trail**.

Perpetuity [p. 339] An investment where an equal amount is paid out on a regular basis forever.

Planar graph [p. 213] A graph that can be drawn in such a way that no two edges intersect, except at the vertices.

Precedence table [p. 418] A table that records the activities of a project, their immediate predecessors and often the duration of each activity.

Prim's algorithm [p. 361] An algorithm for determining a minimum spanning tree in a connected graph.

Principal (P) [p. 148] The initial amount borrowed, lent or invested.

R

Recurrence relation [p. 106] A relation that enables the value of the next term in a sequence to be obtained by one or more current terms. Examples include 'to find the next term, add two to the current term' and 'to find the next term, multiply the current term by three and subtract five'.

Recursion [p. 103] The repetitive application of the same rule or procedure.

Reducing-balance depreciation [p. 165] When the value of an item is reduced by the same percentage each year. Reducing-balance depreciation is equivalent to, but opposite to, compound interest.

Reducing-balance loan [p. 312] A loan that attracts compound interest, but where regular repayments are also made. In most instances the repayments are calculated so that the amount of the loan and the interest are eventually repaid in full.

Reseasonalise [p. 277] The process of converting seasonal data back into its original form.

Residual [p. 57] The vertical distance from a data point to a straight line fitted to a scatterplot is called a residual:

residual = actual value − predicted value

Residuals are sometimes called *errors of prediction*.

Residual plot [p. 75] A plot of the residuals against the explanatory variable. Residual plots can be used to investigate the linearity assumption.

Response variable [p. 3] The variable of primary interest in a statistical investigation.

S

Scatterplot [p. 15] A statistical graph used for displaying bivariate data. Data pairs are represented by points on a coordinate plane, the explanatory variable is plotted on the horizontal axis and the response variable is plotted on the vertical axis.

Scrap value [p. 151] The value at which an item is no longer of use to a business.

Seasonal indices [p. 276] Indices calculated when the data shows seasonal variation. Seasonal indices quantify seasonal variation.

Seasonality [p. 261] The tendency for values in the time series to follow a seasonal pattern, increasing or decreasing predictably according to time periods such as time of day, day of the week, month, or quarter.

Segmented bar chart [p. 8] A statistical graph used to display the information contained in a two-way frequency table. It is a useful tool for identifying associations between two categorical variables.

Semi-Eulerian graph [p. 226] A graph that has an open trail that includes every edge in that graph only once.

Sequence [p. 101] A list of numbers or symbols written down in succession, for example, 5, 15, 25, . . .

Shortest path [p. 234] The path through a graph or network with minimum length.

Simple graph [p. 201] A graph with no loops or multiple edges.

Simple interest [p. 148] Interest that is calculated for an agreed period and paid only on the original amount invested or borrowed.

Sink [p. 393] *See* **sink and source**.

Sink and source [p. 393] In a flow network, a source generates flow while a sink absorbs the flow.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Slope (of a straight line) [p. 58] The slope of a straight line is defined to be: slope $= \frac{\text{rise}}{\text{run}}$. The slope is also known as the gradient.

Smoothing [p. 268] A technique used to eliminate some of the variation in a time series plot so that features such as seasonality or trend are more easily identified.

Source [p. 393] *See* **sink and source**.

Spanning tree [p. 358] A subgraph of a connected graph that contains all the vertices of the original graph, but without any multiple edges, circuits or loops.

Statistical investigation process [p. 2] The process of identifying and understanding a problem, collecting data related to that problem, analysing the data, interpreting the analysis and communicating the results.

Statistician [p. 2] A mathematician with particular skills in the collection and analysis of data, and the interpretation and communication of the results of this analysis.

Strength of an association [p. 32] Classified as weak, moderate or strong. Determined by observing the degree of scatter in a scatterplot or calculating a correlation coefficient.

Structural change (time series) [p. 262] A sudden change in the established pattern of a time series plot.

Subgraph [p. 202] Part of a graph that is also a graph in its own right.

T

Time series data [p. 258] A collection of data values along with the times (in order) at which they were recorded.

Time series plot [p. 258] A line graph where the values of the response variable are plotted in time order.

Trail [p. 220] A walk with no repeated edges. *See also* **path**.

Traversable graph [p. 224] A trail that includes every edge in the graph.

Tree [p. 357] A connected graph with no circuits, multiple edges or loops.

Trend [p. 259] The tendency for values in the time series to generally increase or decrease over a significant period of time.

Trend line [p. 259] A line fitted to an increasing or decreasing time series.

Trend line forecasting [p. 286] A line fitted to an increasing or decreasing time series to predict future values.

Two-way frequency table [p. 6] A frequency table in which subjects are classified according to two categorical variables. Two-way frequency tables are commonly used to investigate the associations between two categorical variables.

U

Unit-cost depreciation [p. 151] Depreciation based on how many units have been produced or consumed by the object being depreciated. For example, a machine filling bottles of drink may be depreciated by 0.001 cents per bottle it fills.

V

Variable [p. 2] Used to represent a quantity that can have many different values in a given situation.

Vertex (graph) [p. 195] The points in a graph or network (*pl* vertices).

W

Walk [p. 220] Any continuous sequence of edges, linking successive vertices, that connects two different vertices in a graph. *See also* **trail** and **path**.

Weighted graph [p. 233] A graph in which a number representing the size of some quantity is associated with each edge. These numbers are called weights.

Y

***y*-intercept** [p. 58] The point at which a graph cuts the y-axis.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Answers

Chapter 1

Exercise 1A

1 a EV: *age* RV: *diameter*
  b EV: *weeks* RV: *weight loss*
  c EV: *age* RV: *price*
  d EV: *hours* RV: *amount*
  e EV: *balls bowled* RV: *runs*
  f EV: *colour* RV: *toxicity*
  g EV: *diet* RV: *weight loss*
  h EV: *fuel* RV: *cost*
  i EV: *location* RV: *price*

2 a *Age*
  b *Years of education*
  c *Temperature*
  d *Time of year*
  e *Age group*
  f *State of residence*

Exercise 1B

1 a *Enrolment status*
  b No. The percentage of full-time and part-time students who drank alcohol is similar: 80.5% to 81.8%. This indicates that drinking behaviour is not related to enrolment status.

2 a *Handedness*
  b

	Sex (%)	
Handedness	*Male*	*Female*
Left	9.0	9.8
Right	91.0	90.2
Total	100.0	100.0

  c No. There is little difference in the percentage of males and females who are left handed, 9.0% compared to 9.8%.

3 a *Sex* **b** 54.9%
  c

	Sex (%)	
Exercised	*Male*	*Female*
Rarely	28.8	39.2
Sometimes	52.5	54.9
Regularly	18.6	5.9
Total	99.9	100.0

  d There are several ways you can answer the question.

Focusing on the category 'rarely'.
Yes. The percentage of males who rarely exercised (28.8%) was significantly less than the percentage of females who rarely exercised (39.2%).
or
Yes. The percentage of males who exercised regularly (18.6%) was significantly higher than the percentage of females who exercised regularly (5.9%).
Note: For the category 'sometimes', there is no association between level of exercise and sex.

4 a 11.9% **b** 52.3% **c** *Marital status*
  d Yes. There are several ways that this can be seen. For example, by comparing the married and widowed groups, we can see that a smaller percentage of those widowed found life exciting (33.8%) compared to those who were married (47.6%). Or: a bigger percentage of widowed people found life pretty routine (54.3% to 48.7%) and dull (11.9% to 3.7%) compared to those who were married.

5 a

	Group		
Engagement	*Primary*	*Secondary*	*Total*
Play	125	36	161
Do not play	28	11	39
Total	153	47	200

b 200 **c** 161

d

	Group	
Engagement	*Primary*	*Secondary*
Play	81.7%	76.6%
Do not play	18.3%	23.4%
Total	100%	100%

e Primary students appear to play more sport.

6 a The attitude to flying is likely to be dependent on the sex of the person.

b

	Sex		
Attitude	*Female*	*Male*	*Total*
Enjoy	52	164	216
Do not enjoy	30	86	116
Have never	4	14	18
Total	86	264	350

c

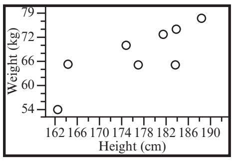
	Sex	
Attitude	*Female*	*Male*
Enjoy	60.5%	62.1%
Do not enjoy	34.9%	32.6%
Have never	4.7%	5.3%
Total	100.1%	100%

d 32.6% of males do not enjoy flying.

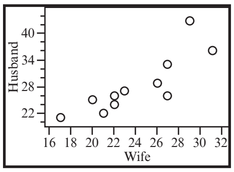
e There does not seem to be an association between sex and attitude to flying because the percentages in all categories are very close for men and women.

Exercise 1C

1

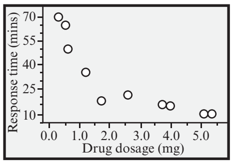


2



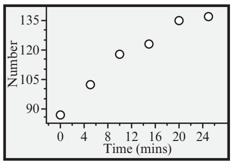
3 a *Drug dosage*

b

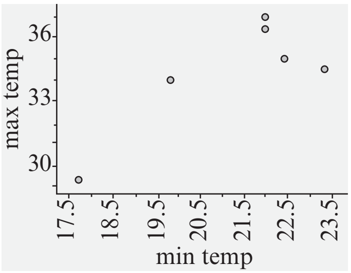


4 a *Time*

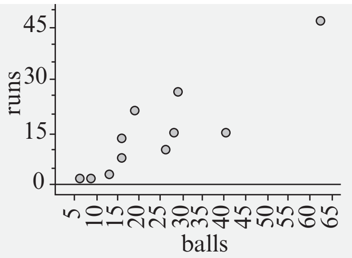
b



5



6



7 D

8 a EV: *hours of sleep*, RV: *mistakes*

b 12 people **c** 9 hours, 6 mistakes

Chapter 1 review

Short-answer questions

1 a Plays sport and sex; both categorical variables

b 96 **c** 33.3%

d Yes, there is an association between sex and likelihood to play sport.
A higher percentage of males (66.7%) play sport compared to females (45.1%).

2 a 6 **b** 100

c Yes, there is an association between the variables attends football matches and sex. The percentage of males and females who attend football matches occasionally or never is significantly higher.

3 D

Extended-response questions

1 a *number of accidents* and *age*; both categorical variables

b RV: *number of accidents*, EV: *age*

c 470

d

Number of accidents	*Age* < 30	*Age* ≥ 30
At most one accident	21.7%	42.5%
More than one accident	78.3%	57.5%
Total	100%	100%

e The statement is correct. Of drivers aged less than 30, 78.3% had more than one accident compared to only 57.5% of drivers in the older category.

2 a 4 **b** socialising

c EV: Time of observation **d** 12.7%

e

	Time of observation			
Activity	*Morning*	*Noon*	*Afternoon*	*Evening*
Travelling	8.3%	40%	60.9%	16.5%
Feeding	38.9%	26.7%	0%	70.9%
Socialising	52.8%	33.3%	39.1%	12.7%
Total	100%	100%	100%	100%

f There does seem to be an association between the dolphin activity and time of observation. The percentage of dolphins in each activity is different for every time of observation. For example, 70.9% of dolphins observed in the evening are feeding, while 52.8% of dolphins observed in the morning are socialising.

3 a

	Type of student		
Attitude	*Day*	*Evening*	*Total*
Satisfied	90	22	112
Neutral	18	5	23
Dissatisfied	12	3	15
Total	120	30	150

b 30

c

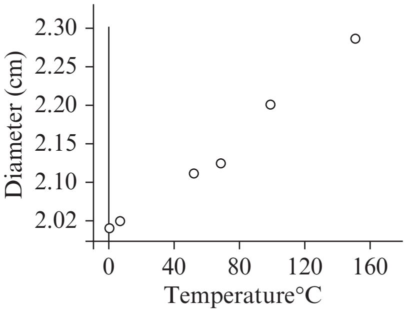
	Type of student	
Attitude	*Day*	*Evening*
Satisfied	75%	73.3%
Neutral	15%	16.7%
Dissatisfied	10%	10%
Total	100%	100%

d 75%

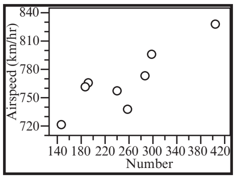
e There does not seem to be a relationship between satisfaction with the course and the type of student. Similar percentages of day and evening students are satisfied, neutral and dissatisfied with the course.

f The statement refers to actual numbers of students rather than percentages. There were more day students surveyed so the number of students satisfied with the course, neutral or dissatisfied with the course is understandably larger. The percentage of students, however, is almost the same.

4



5



6 a *Number of seats* **b** 8 aircraft

c Around 800 km/h

Chapter 2

Exercise 2A

1 a Positively associated

b Positively associated

c Negatively associated

d No association

e Positively associated

f Positively associated

2 Note: Estimates of strength can vary by one level and still be appropriate

a Strong positive linear association

b Strong negative linear association

c Moderate positive linear association

d Strong positive non-linear association

e Moderate negative linear association

f Strong positive linear association with possible outlier

Exercise 2B

1 1: variables numerical; 2: association linear; 3: no clear outliers

2 a A: strong, positive, non-linear relationship with no outliers

B: strong, negative, linear relationship with an outlier

C: weak, negative, linear relationship with no outliers

b A: non-linear

B: outlier

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

3 a None **b** Weak negative
c Strong negative **d** Weak positive
e Strong positive **f** Moderate negative
g Moderate positive **h** None
i Weak negative **j** Weak positive
k Perfect positive or strong positive
l Perfect negative or strong negative

4 a-c Answers given in question.

Exercise 2C

1 a 45.6% **b** 11.9% **c** 32.1% **d** 45.3%
e 1.5%

2 a $r = 0.906$ **b** $r = -0.353$

3 a The coefficient of determination is $r^2 = (-0.611)^2 = 0.373$ or 37.3%; that is, 37.3% of the variation observed in hearing test scores can be explained by variation in age.

b The coefficient of determination is $r^2 = (0.716)^2 = 0.513$ or 51.3%; that is, 51.3% of the variation observed in mortality rates can be explained by variation in smoking rates.

c The coefficient of determination is $r^2 = (-0.807)^2 = 0.651$ or 65.1%; that is, 65.1% of the variation observed in life expectancies can be explained by variation in birth rates.

d The coefficient of determination is $r^2 = (0.818)^2 = 0.669$ or 66.9%; that is, 66.9% of the variation observed in daily maximum temperature is explained by the variability in daily minimum temperatures.

e The coefficient of determination is $r^2 = (0.8782)^2 = 0.771$ or 77.1%; that is, 77.1% of the variation in the runs scored by a batsman is explained by the variability in the number of balls they face.

Exercise 2D

Note: These answers are for guidance only. Alternative explanations for the source of an association may be equally acceptable as the variables suggested.

1 Not necessarily. In general, older children are taller and have been learning mathematics longer. Therefore, they tend to do better on mathematics tests. Age is the probable common cause for this association.

2 Not necessarily. While one possible explanation is that religion is encouraging people to drink, a better explanation might be that towns with large numbers of churches also have large populations, thus explaining the larger amount of alcohol consumed. Town size is the probable common cause for this association.

3 Probably not. The amount of ice-cream consumed and the number of drownings would both be affected by weather conditions. Weather conditions are the probable common cause.

4 Maybe but not necessarily. Bigger hospitals tend to treat more people with serious illnesses and these require longer hospital stays. A common cause could be the type of patients treated at the hospital.

5 Not necessarily. Possible confounding variables include age and diet.

6 There is no logical link between eating cheese and becoming tangled in bed sheets and dying. The correlation is probably spurious and the result of coincidence.

7 Not necessarily. For example, the more serious the fire, the more fire trucks in attendance and the greater the fire damage. A possible common cause is the severity of the fire.

Chapter 2 review

Short-answer questions

1 a Strong negative linear
b Weak positive linear
c Strong positive linear
d No association
e Strong negative non-linear

2 Infant mortality tends to decrease as birth weight increases

3 a Correlation coefficient not appropriate; non-linear association
b Correlation coefficient appropriate; linear association
c Correlation coefficient not appropriate; contains outlier
d Correlation coefficient not appropriate; non-linear association
e Correlation coefficient appropriate; linear association

4 D

5 A

6 Weak negative

7 0.8

8 Strong positive linear

9 a Response times decrease with increased drug dosage

b $r^2 = 0.81$

c 81% of the variation in response time can be explained by the variation in drug dosage.

10 a $r^2 = 0.5852$. 58.52% of the variation in height weight can be explained by the variation in body weight.

b Heavier mice tend to have heavier hearts.

11 a Positive association. This means that people on high salaries tend to recycle more rubbish.

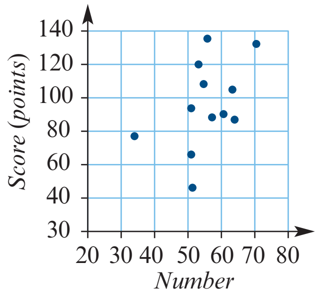
b This is not a valid conclusion since correlation does not imply causation. It cannot be concluded that increasing salaries will cause an increase in recycling.

12 This is not a valid conclusion as this correlation is just coincidence and changing the marriage rate will not affect the number of people drowning in any way. The population of Kentucky increasing most likely explains the increased marriage and drowning numbers.

Extended-response questions

1 a *Score*

b



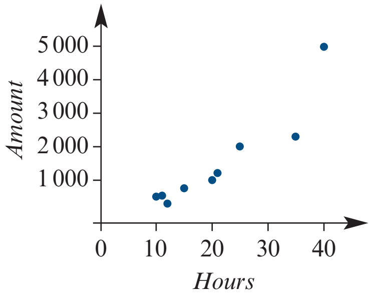
c Weak linear positive association

2 $r = 0.9272$

3 a EV: *hours*

RV: *amount*

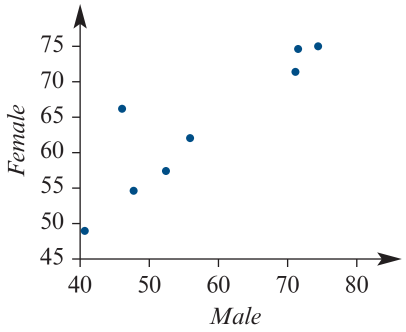
b



c $r = 0.922$

d Strong positive linear association: Those who gambled for longer tended to spend more on gambling.

4 a



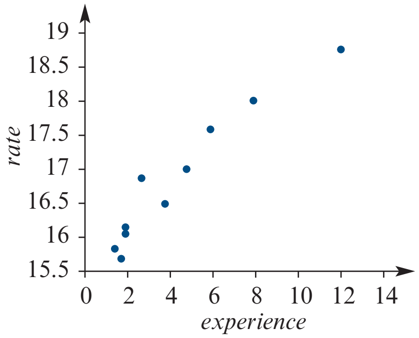
b $r = 0.894$

c Strong positive linear association with an outlier: Those countries with high percentages of males with eye disease also tended to have a high percentage of females with eye disease.

d $r^2 = 79.92\%$

79.92% of the variation in the percentage of females with eye disease can be explained by the percentage of males with eye disease.

5 a



'Rate' is the response variable.

b There is a strong positive linear relationship; that is, people with more experience are generally being paid a higher starting pay rate. There are no apparent outliers. $r \approx +0.97$.

c Coefficient of determination = 0.935; that is, 93.5% of variation in pay rate is explained by the variation in experience.

d 0.967

ISBN 978-1-009-25917-0

Chapter 3

Exercise 3A

1 a r is also negative.

b Slope is zero: regression line is horizontal.

2–3 Answers given in question.

4 a Answers given in question.

b $r = 1.06b + 8.30$

5 a RV: *number of TVs*

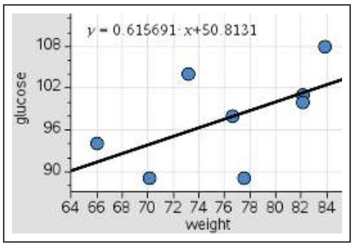
b Answer given in question.

c $t = 0.930c + 61.2$

Exercise 3B

1 $m = -4d + 80$

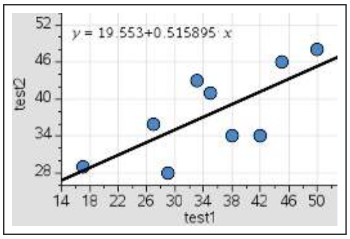
2 a & b



c $g = 0.616\,w + 50.8$

d $r = 0.570$

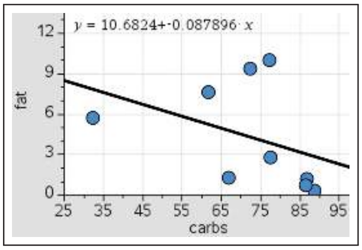
3 a & b



c $B = 0.516A + 19.6$

d $r = 0.722$

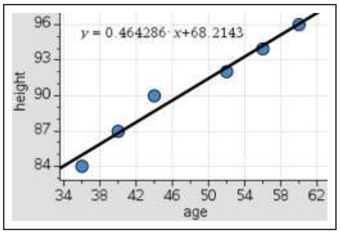
4 a & b



c $f = -0.0879\,c + 10.7$

d $r = -0.396$

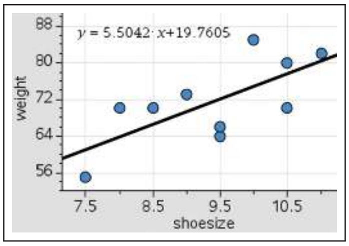
5 a & b



c $height = 0.464 \times age + 68.2$

d $r = 0.985$

6 a & b



c $weight = 5.50 \times shoe\ size + 19.8$

d $r = 0.702$

7 a interpolation **b** extrapolation

8 a 88 cm, interpolation

b 94 cm, interpolation

c 100 cm, extrapolation

9 a 59 kg, extrapolation **b** 66 kg, interpolation

c 72 kg, interpolation

10 a \$175, extrapolation **b** \$523, interpolation

c \$691, interpolation

11 a 171 cm, interpolation

b 197 cm, extrapolation

c 159 cm, interpolation

12 a height **b** 0.33, 2.9 **c** 55.7

13 a fuel consumption **b** 0.01, −0.1

c 9.7

14 a 14.7, 27.8 **b** 14.7 **c** 0.87

d 75.69, fat content **e** 145.4

15 a −0.278: the slope predicts that success rate decreases by 0.278 for each additional metre the golfer is from the hole.

b 73.5% **c** 3.54 m **d** −0.705

e 49.7%: 49.7% of the variation in success rate in putting is explained by the variation in the distance the golfer is from the hole.

Exercise 3C

1 a Linear model not appropriate, since clear curved pattern in the residuals (not random)

b Linear model appropriate, since no pattern in the residuals (random)

c Linear model not appropriate, since clear curved pattern in the residuals (not random)

ISBN 978-1-009-25917-0

2 **a** Positive **b** Negative **c** Negative

3 **a** True – the residual for $x = 3$ is positive which means that the actual value is above the least squares regression line.

b True – there is no pattern to the residual plot.

4 **a** $y = -1.34x + 18.04$

b **i** $\hat{y} = -1.34 \times 6 + 18.04 = 10$

ii Residual $= y - \hat{y} = 8 - 10 = -2$

iii The residual is negative, so the predicted value is greater than the actual value. The workmen will take less time to lay the bricks than what is predicted using the least squares regression line.

c There is a pattern to the residuals, so a linear model is not appropriate for the data.

5 **a** $y = 0.010x + 0.404$

b $A = -0.51$

$B = 12.104$

$C = 3.096$

c There is no pattern to the residuals, so a linear model is appropriate for the data.

6 **a** Yes, linear relationship

b 0.9351 or 93.5%

c 93.5%

d *pay rate* = 8.56 + 0.289 × *experience*

e The pay rate for a worker with no experience

f On average, the pay rate increases by 29 cents per hour for each additional year of experience.

g **i** \$10.87 **ii** \$0.33

h Yes; no clear pattern in the residual plot

7 **a** $r = -0.608$

b 37% of the variation in the hearing test score is explained by the variation in age.

c *score* = 4.9 – 0.043 × *age*

d –0.043; the hearing test score, on average, decreases by 0.043 as age increases by 1.

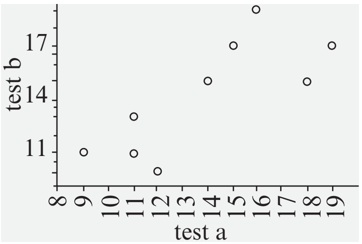
e **i** 4.04 **ii** –2.04

f **i** 0.3 **ii** –0.4

g Yes; no clear pattern in the residual plot

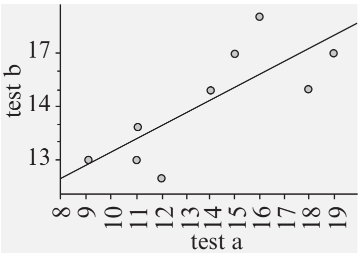
Exercise 3D

1 **a**

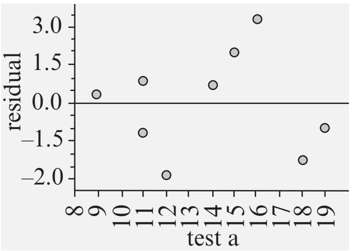


b $y = 0.72\,x + 4.2,\ r = 0.78,\ r^2 = 0.61$

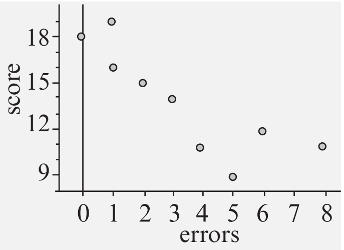
c



d

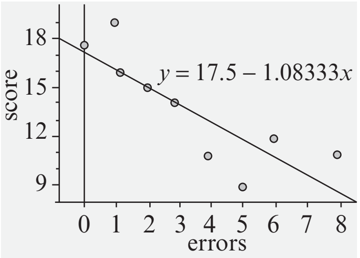


2 **a**

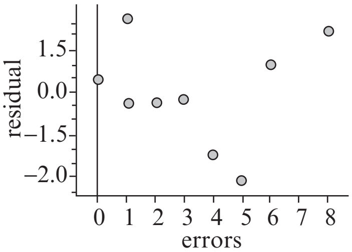


b *score* = 17.5 – 1.08 × *errors*, $r = -0.841,\ r^2 = 0.708$

c

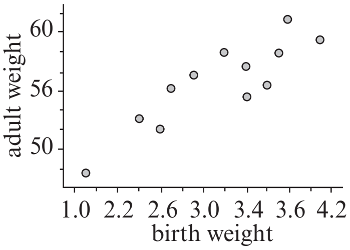


d



3 **a** RV: *adult weight*; EV: *birth weight*

b



c **i** Strong positive linear association with no outliers

ii Your judgement

d *adult weight* = 38.4 + 5.87 × *birth weight*, $r^2 = 0.765$, $r = 0.875$

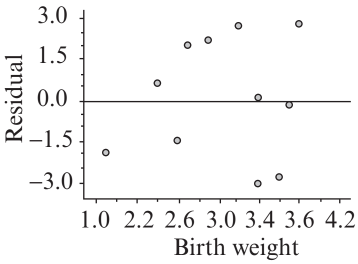
e 76.5% of the variation in the adult weight is explained by the variation in birth weight.

f On average, adult weight increases by 5.9 kg for each additional kilogram of birth weight.

g **i** 56.0 **ii** 53.1 **iii** 61.3

h Yes. 76.5% of the variation in the adult weight is explained by the variation in birth weight.

i



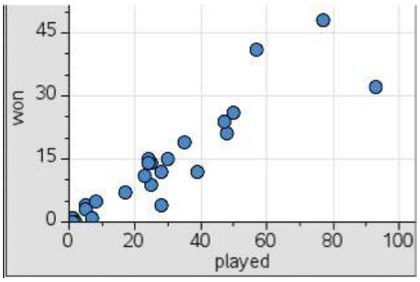
The lack of a clear pattern in the residual plot supports the assumption that the relationship between adult weight and birth weight is linear.

Exercise 3E

1 negative, drug dose, −0.9492; −9.3; 55.9; decreases, 9.3; 55.9; 90.1, response time, drug dose; clear pattern

2 The scatterplot shows that there is a strong positive linear relationship between radial length and femur length: $r = 0.9876$. There are no outliers.
The equation of the least squares line is:
radial length = 0.74 × *femur length* – 7.2
The slope of the least squares line predicts that, on average, radial length increases by 0.74 cm for each centimetre increases in femur length.
The coefficient of determination indicates that 97.5% of the variation in radial lengths can be explained by the variation in femur lengths.
The residual plot shows no clear pattern, supporting the assumption that the relationship between radial and femur length is linear.

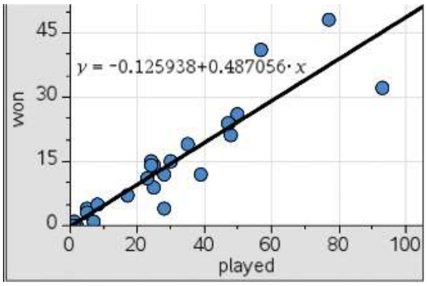
3 **a**



The estimated value of the correlation coefficient r is (*your estimate*).

b There is a strong positive linear association between the number of tests played as captain and the number of matches won. There are no clear outliers.

c



d The equation of the least squares line is: *won* = −0.13 + 0.49 × *played* (to 2 sig. figs.) slope = 0.49: on average, the number of tests won by a captain increased by 0.49 for each additional match played as captain.

e $r = 0.9$ (to 1 d.p.) – use this answer to check your estimate in **a**.

f Statistical report will vary; for style, see example containing a statistical report in the chapter.

Chapter 3 review

Short-answer questions

1 slope = 0.52 and y-intercept = −1.2

2 $r = -0.5$

3 $y = -37$

4 $y = -0.69x + 24.4$

5 $y = 0.5x + 7.5$

6 84.0

7 A

8 A

ISBN 978-1-009-25917-0

9 **a** EV: height

b Weight increases by 0.95 kg for each 1 centimetre increase in height

c $r^2 = 0.6241$; 62.41% of the variation in weight can be explained by the variation in height

d 8 kg

10 $r = -0.7$

11 $e = 0.482n - 42.864$

12 **a** \$100

b 10 cents

Extended-response questions

1 **a** *Magnesium content*

b On average, taste scores increases by 7.3 points for each one mg/litre increase in magnesium content.

c 94.8 milligrams/litre

2 **a** *errors* = 14.92 – 0.53 × *time*

b $r = -0.94$

3 **a** days of rain

b –6.88, 2850

c 2024

d decrease, 6.88

e –0.696

f 48.4, days of rain

g **i** 1873 **ii** –483

h interpolation

4 **a** *Cost*

b Answer given in question.

c **i** \$182.30, interpolating

ii \$125.60, extrapolating

d **i** 81.5. The fixed cost of preparing meals is \$81.50.

ii \$2.10. The slope of the least squares line predicts that, on average, meal preparation costs increase by \$2.10 for each additional meal produced.

e 0.956; 95.6% of the variation in the cost of preparing meals is explained by the variation in the number of meals produced.

5 **a** RV: *height*; EV: *age*

b Answer given in question.

c 83 cm, extrapolation

d On average, height increases by 6.4 cm for each extra year.

e $r^2 = 0.995$; that is, 99.5% of the variation in height is explained by the variation in age.

f **i** 140.3 cm **ii** –0.7 cm

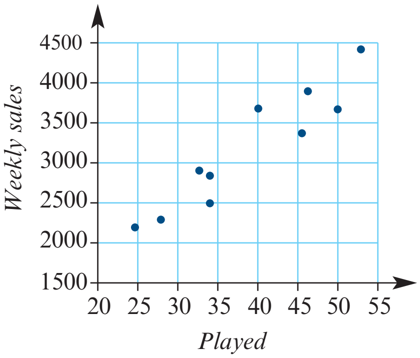
g Answer given in question.

6 Statistical report will vary; for style, see example containing a statistical report in the chapter.

7 **a** EV**:** *played*

RV**:** *weekly sales*

b



c $r = 0.9458$

d Strong, positive, linear relationship

e *weekly sales* = 293 + 74.3 × *played*

f Slope: on average, the number of down loads increases by 74.3 for each additional time the song is played on the radio in the previous week.

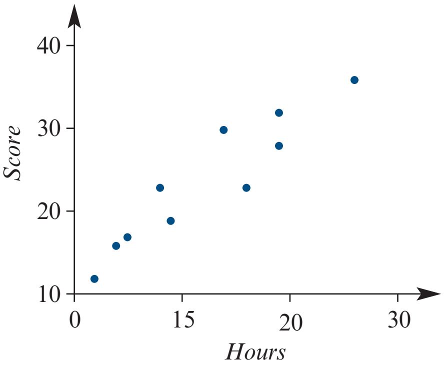
Intercept: predicts 293 downloads of the song if it is not played on radio in the previous week.

g 7723

h Extrapolating

8 **a** EV: *hours* RV: score

b



c $r = 0.9375$

d Positive strong, linear

e *score* = 12.3 + 0.930 × *hours*

f Slope: on average, test scores of learner drivers increases by 0.93 marks when instruction time increases by one hour.

Intercept: on average, test scores of learner drivers who received no instruction prior to taking the test is 12.3 marks.

g 22

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

9 a The statement is questionable because it implies causality. The existence of even a strong relationship between two variables is not, by itself, sufficient information to conclude that one variable causes a change in the other.

b *exam mark* = 25 + 0.70 × *assignment mark* (correct to 2 sig. figures)

c Intercept: on average those who score 0 on the assignment scored 25 marks on the final exam (or equivalent).
Slope: on average, exam marks increase by 0.7 for each additional mark obtained on the assignment.

d 60

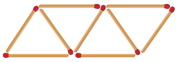
e Reliable: The prediction made in part **d** falls within the range of the data (interpolating).

Chapter 4

Exercise 4A

1 a 19 **b** 6
c 64 **d** March
e 1 **f** 3
g 3 **h** 125
i i **j** ♣
k 81 **l** ⇒
m 162 **n** ↑ **o** F (week days)
p F (counting words)
q I (letters written using straight lines)

2 a



b 3, 5, 7, . . .
c 9, 11
d Start at 3 and add 2 to make each new term.

3 a

b

Pattern number	1st	2nd	3rd	4th	5th
Number of dots	1	4	7	10	13

c Start with 1 and add 3 to make each new term.

4 a Add 3; 17, 20 **b** Add 9; 55, 64
c Subtract 4; 22, 18 **d** Subtract 8; 34, 26
e Multiply by 2; 48, 96
f Multiply by 3; 324, 972
g Divide by 2; 8, 4
h Multiply by −2; 48, −96
i Add the previous two terms; 8, 13

5 a 2, 8, 14, 20, 26, . . .
b 5, 2, −1, −4, −7, . . .
c 1, 4, 16, 64, 256, . . .
d 10, 5, 2.5, 1.25, 0.625, . . .
e 6, 14, 30, 62, 126, . . .
f 12, 9, 7.5, 6.75, 6.375, . . .

6 a 4, 6, 8, 10, 12, . . .
b 24, 20, 16, 12, 8, . . .
c 2, 6, 18, 54, 162, . . .
d 50, 10, 2, 0.4, 0.08, . . .
e 5, 13, 29, 61, 125, . . .
f 18, 16.4, 15.12, 14.096, 13.2768, . . .

7 a 6 **b** 16
c 11 **d** 21
e 26 **f** 31

8 a i 6 **ii** 18 **iii** 22
b i 2 **ii** 128 **iii** 512
c i 29 **ii** 8 **iii** 1
d i 96 **ii** 12 **iii** 6

Exercise 4B

1 a T_6 **b** T_{13}
c T_{11} **d** T_9

2 a 11 **b** 13
c 19 **d** 17

3 a 2, 5, 8, 11, 14, . . .
b 50, 45, 40, 35, 30, . . .
c 1, 3, 9, 27, 81, . . .
d 3, -6, 12, -24, 48, . . .
e 5, 9, 17, 33, 65, . . .
f 2, 7, 17, 37, 77, . . .
g −2, −1, 2, 11, 38, . . .
h −10, 35, −100, 305, −910, . . .

4 a 12, 57, 327, 1947, 11667, . . .
b 20, 85, 280, 865, 2620, . . .
c 2, 11, 47, 191, 767, . . .
d 64, 15, 2.75, −0.3125, −1.078125, . . .
e 48 000, 45 000, 42 000, 39 000, 36 000, . . .
f 25 000, 21 950, 19 205, 16 734.5, 14 511.05, . . .

5 6

6 7

Exercise 4C

1 a 3 **c** −4
d −3 **f** 0
b, e Are not arithmetic

2 a 3, 5, 7, . . . **b** 2 **c** 9, 11

3 a 7, 12, 17, . . .
b Yes, $d = 5$ **c** 22, 27 matchsticks

4 **a** $d = 6, T_5 = 29$ **b** $d = -4, T_5 = 1$
c $d = 4, T_5 = 27$ **d** $d = -4, T_5 = -8$
e $d = -5, T_5 = 15$ **f** $d = 0.5, T_5 = 3.5$

5 **a** 41, 47 **b** 2, −1 **c** 0, −0.5 **d** 59, 67
e −15, −27 **f** 2, 2.3

6 **a** 28, 33 **b** −10, −16 **c** 24, 33
d 13, 8 **e** 11, 19 **f** 13, 21 **g** 29, 18
h 29, 15 **i** 23, 39, 55

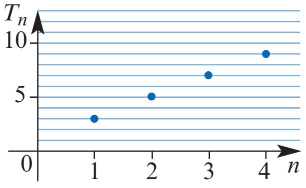
7 **a** **i** 1 **ii** 5 **iii** 16
b **i** 45 **ii** −2 **iii** 39
c **i** 34 **ii** 6 **iii** 52
d **i** 0 **ii** 3.5 **iii** 10.5

8 **a** $T_1 = 3, T_{n+1} = T_n + 5$
b $T_1 = 16, T_{n+1} = T_n - 7$
c $T_1 = 1.6, T_{n+1} = T_n + 2.3$
d $T_1 = 8.7, T_{n+1} = T_n - 3.1$
e $T_1 = 293, T_{n+1} = T_n - 67$

9 **a** **i** 9 **ii**

n	1	2	3	4
T_n	3	5	7	9

iii

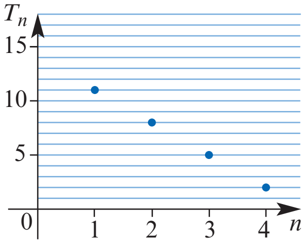


iv Points lie on a line with positive slope.

b **i** 2 **ii**

n	1	2	3	4
T_n	11	8	5	2

iii



iv Points lie on a line with negative slope.

Exercise 4D

1 **a** $a = 7, d = 4$ **b** $a = 8, d = -3$
c $a = 14, d = 9$ **d** $a = 62, d = -27$
e $a = -9, d = 5$ **f** $a = -13, d = -6$

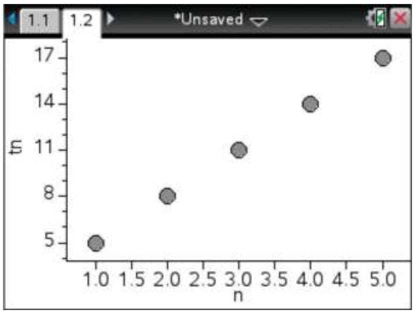
2 **a** $T_n = 12 + (n - 1) \times 9$
b $T_n = 146 + (n - 1) \times 50$
c $T_n = 27 + (n - 1) \times -5$
d $T_n = 8 + (n - 1) \times -5$

3 **a** $T_n = 5 + (n - 1) \times 9$
b $T_n = -8 + (n - 1) \times 3$
c $T_n = 3 + (n - 1) \times -7$
d $T_n = -1 + (n - 1) \times -6$

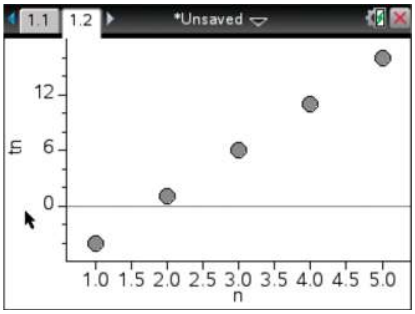
4 **a** 120 **b** 306 **c** −441 **d** 436
e −119 **f** 25.3 **g** −198 **h** −7.4

5 323 **6** 1908 **7** −33

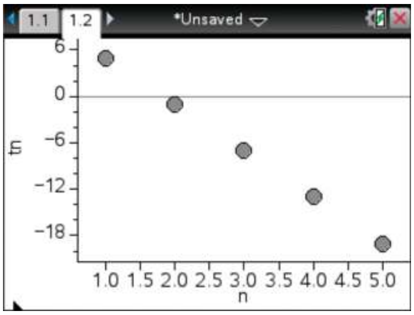
8 **a**



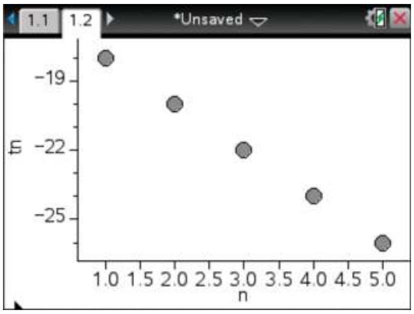
b



c



d



Exercise 4E

1 **a** Yes, 2 **b** Yes, 3 **c** No **d** Yes, 3
e Yes, $\frac{1}{2}$ **f** No **g** No **h** Yes, $\frac{1}{3}$
i Yes, 2

2 **a** 2 **b** $\frac{1}{4}$ **c** 5 **d** 4
e $\frac{1}{2}$ **f** 6 **g** 10 **h** 7
i 8

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

3 a 56, 112 **b** 375, 1875 **c** 36, 108
d 5, 10 **e** 8, 512 **f** 9, 81

4 a i 7 **ii** $r = 5$ **iii** 875
b i 3 **ii** $r = 6$ **iii** 648
c i 96 **ii** $r = 0.5$ **iii** 12
d i 160 **ii** $r = 0.25$ **iii** 2.5

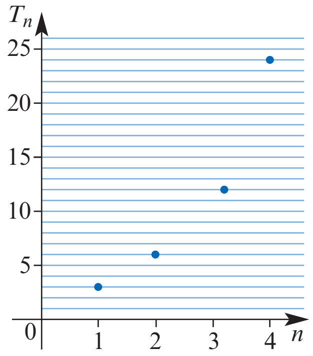
5 a $T_1 = 2, T_{n+1} = 7T_n$
b $T_1 = 15, T_{n+1} = \frac{1}{5}T_n$
c $T_1 = 24, T_{n+1} = \frac{1}{8}T_n$
d $T_1 = 9, T_{n+1} = 3T_n$

6 a i 24

ii

n	1	2	3	4
T_n	3	6	12	24

iii



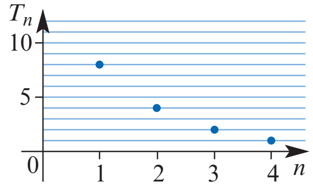
iv The graph is a curve with increasing values.

b i 1

ii

n	1	2	3	4
T_n	8	4	2	1

iii



iv The graph is a curve with values decreasing and approaching zero.

Exercise 4F

1 a $a = 12, r = 2$ **b** $a = 6, r = 3$
c $a = 2, r = 4$ **d** $a = 56, r = \frac{1}{2}$
e $a = 36, r = \frac{1}{3}$ **f** $a = 8, r = 7$
g $a = 1, r = 10$ **h** $a = 100, r = 0.1$
i $a = 17, r = 13$

2 a $a = 3, r = 4$ **b** $a = 5, r = 2$
c $a = 8, r = 7$ **d** $a = 200, r = 1.1$
e $a = 600, r = 0.9$ **f** $a = 3, r = 2$

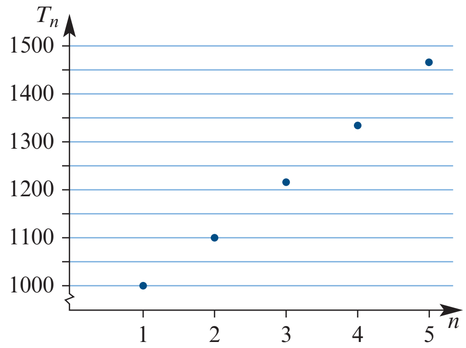
3 a $T_n = 9 \times 2^{n-1}$ **b** $T_n = 54 \times (\frac{1}{3})^{n-1}$
c $T_n = 4 \times 5^{n-1}$ **d** $T_n = 6 \times 7^{n-1}$
e $T_n = 5 \times 4^{n-1}$ **f** $T_n = 8 \times 3^{n-1}$

4 a 78 732 **b** 1536 **c** 39 366
d 262 144 **e** 196 830 **f** 1

5 a 1 **b** 1152 **c** 524 288
d 531 441 **e** $\frac{1}{3}$ **f** 0.0001

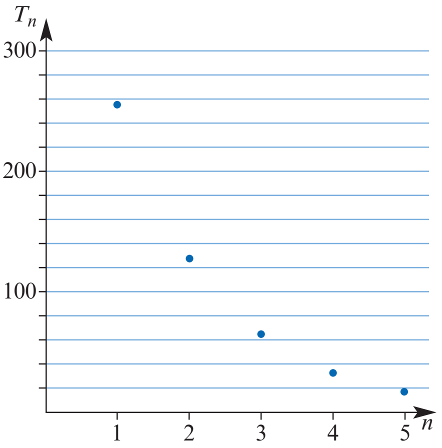
6 3072 **7** 3172.62 **8** 5.73

9 a 1000, 1100, 1210, 1331, 1464.1



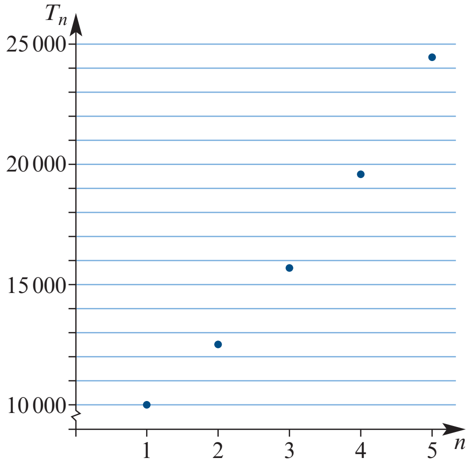
b 3138.43

10 a 256, 128, 64, 32, 16



b 0.5

11 a 10 000, 12 500, 15 625, 19 531, 24 414



b 2 117 582 to nearest whole number

ISBN 978-1-009-25917-0

Chapter 4 review

Short-answer questions

1 10

2 C

3 −8

4 −35

5 41

6 75

7 13

8 C

9 $T_1 = 3, T_{n+1} = 2T_n$

10 1 572 864

11 5, 2, −1, −4, −7

12 132

13 37

14 $T_n = 40 - 16(n - 1)$

15 83

16 1536

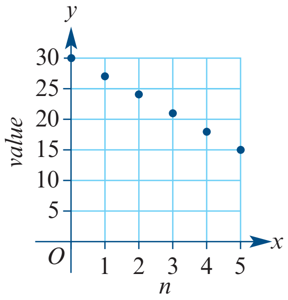
17 112

Extended-response questions

1 **a** B **b** D **c** A **d** C

2 **a** 800, 630, 494, 385.2, 298.16
b T_{14}

3 **a** $T_1 = 30,\ T_{n+1} = T_n - 3$
b



c T_{11}

4 **a** **i** $a = 2, d = 4$
ii $T_1 = 2, T_{n+1} = T_n + 4$
iii $T_n = 2 + (n - 1) \times 4$
iv $T_{15} = 58$
b **i** $a = 140, d = -9$
ii $T_1 = 140, T_{n+1} = T_n - 9$
iii $T_n = 140 + (n - 1) \times -9$
iv $T_{15} = 14$
c **i** $a = 1.5, d = 2.2$
ii $T_1 = 1.5, T_{n+1} = T_n + 2.2$
iii $T_n = 1.5 + (n - 1) \times 2.2$
iv $T_{15} = 32.3$

5 **a** **i** $a = 3, r = 4$
ii $T_1 = 3, T_{n+1} = 4 \times T_n$
iii $T_n = 3 \times 4^{n-1}$
iv $T_7 = 12\,288$
b **i** $a = 1400, r = 0.5$
ii $T_1 = 1400, T_{n+1} = 0.5 \times T_n$
iii $T_n = 1400 \times 0.5^{n-1}$
iv $T_7 = 21.875$
c **i** $a = 7200, r = 0.6$
ii $T_1 = 7200, T_{n+1} = 0.6 \times T_n$
iii $T_n = 7200 \times 0.6^{n-1}$
iv $T_7 = 335.9232$

Chapter 5

Exercise 5A

1 \$123 **2** 15.25 m

3 68.4, 68.1, 67.8, 67.5 seconds

4 \$43

5 **a** 10 logs **b** 61 logs

6 **a** 7, 11, 15 **b** 7, 11, 15, 19, 23
c 43

7 **a** 16, 22, 28 **b** 16, 22, 28, 34
c Yes **d** 70

8 **a** \$2076, \$2152, \$2228 **b** 14
c $V_0 = 1500, V_{n+1} = V_n + 90$

9 **a** \$7518, \$8036, \$8554 **b** 6
c $V_0 = 12\,000, V_{n+1} = V_n + 984$

10 **a** **i** \$15 000 **ii** \$525 **iii** 3.5%
b 29

11 **a** \$2100, \$1700, \$1300 **b** 4
c $V_0 = 1800, V_{n+1} = V_n - 350$

12 **a** \$22 195, \$21 390, \$20 585 **b** 17
c $V_0 = 37\,000, V_{n+1} = V_n - 700$
d $V_0 = 12\,000, V_{n+1} = V_n - 540$

13 **a** **i** \$1500 **ii** \$102 **iii** 6.8%
b \$684

14 **a** \$449 **b** $V_0 = 300, V_{n+1} = V_n - 0.08$

15 **a** \$47 800, \$47 600, \$47 400 **b** \$45 000
c 25 000

Exercise 5B

1 1 cm, 3 cm, 9 cm, 27 cm

2 **a** \$50, \$100, \$200 **b** \$350

3 **a** 5, 25, 125, 625 letters
b 780 letters

4 **a** 2, 4, 8, 16 **b** See **a** **c** 1024
d See **c**

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

5 a 100, 95, 90.25, 85.74
  b 100, 120, 144, 172.8
  c 5000, 5150, 5304.50, 5463.64
  d 7000, 6720, 6451.20, 6193.15

6 \$4267.25

7 \$102 829.46

8 a \$6252, \$6514.58, \$6788.20 **b** 7 years
  c $V_0 = 5000, V_{n+1} = 1.068V_n$

9 a \$21 260, \$22 599.38, \$24 023.14
  b 7 years **c** $V_0 = 18\,000, V_{n+1} = 1.094V_n$

10 a $V_0 = 7600, V_{n+1} = 1.005V_n$
  b \$7791.91

11 a $V_0 = 3500, V_{n+1} = 1.02V_n$
  b \$3788.51

12 a $V_0 = 9800, V_{n+1} = 0.965V_n$
  b \$9800, \$9457, \$9126, \$8806.59, \$8498.36, \$8200.92
  c \$8200.92 **d** \$319.41

13 a $V_0 = 18\,000, V_{n+1} = 0.955V_n$
  b \$18 000, \$17 190, \$16 416.45, \$15 677.71, \$14 972.21, \$14298.46
  c \$15 677.71 **d** \$3701.54

Exercise 5C

1 a Arithmetic decay **b** Geometric growth
  c Geometric decay **d** Arithmetic growth

2 a Arithmetic growth **b** Arithmetic decay
  c Geometric growth **d** Geometric decay

3 a Geometric growth **b** Arithmetic growth
  c Combined growth **d** Geometric growth
  e Combined growth **f** Combined decay

4 a 8600 **b** 15% **c** 750 **d** Growing
  e 12 241

5 a 2600 **b** 5% **c** 200
  d Decaying **e** 1107 **f** 10 years

6 a $A = 3000, B = 1.04, C = 40$ **b** 4061
  c 11 months

7 a $H_0 = 260, H_{n+1} = 0.955\ H_n + 8$
  b i 246 **ii** 29 months

8 a $S_0 = 825, S_{n+1} = 1.03\ S_n - 20$
  b 840 **c** 14

9 a $C_0 = 360, C_{n+1} = 0.88\ C_n + 3$ **b** 226
  c 12 months

Exercise 5D

1 a Steady state **b** Increasing
  c Steady state **d** Steady state
  e Steady state **f** Steady state
  g Steady state

2 a 100 **b** 93.75 **c** 20 **d** 10

3 40

4 0.6

5 a 0.98
  b $W_0 = 250, W_{n+1} = 0.98\ W_n + P$
  c i 5 **ii** 4

6 a $H_0 = 100, H_{n+1} = 0.8\ H_n + 80$
  b The geometric decay factor, R, is less than 1 and the difference in height between each year decreases, suggesting it will eventually reach a steady state.
  c 400 cm (4 m)
  d 32%

Chapter 5 review

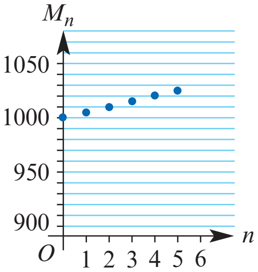
Short-answer questions

1 1.07

2 $T_0 = 2, T_{n+1} = T_n + 3$

3

n	0	1	2	3	4	5
M_n	1000	1005	1010	1015	1020	1025



4 $V_0 = 2000,\ V_{n+1} = V_n + 102$

5 flat-rate depreciation at 6.1% of its value per annum

6 \$896

7 \$0.15

8 $P_0 = 2700,\ P_{n+1} = 0.92 \times P_n$

9 B

10 $V_n = 0.86^n \times 4500$

11 1.0105

12 a $V_0 = 6000,\ V_{n+1} = 1.0457V_n$
  b 7 years

13 $W_0 = 490,\ W_{n+1} = 1.08W_n - 60$

14 C

15 600

16 a 500
  b 12.5

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Extended-response questions

1 45, 75, 105, 135 seconds

2 48, 24, 12, 6 metres

3 659.08

4 **a** $V_0 = 20\,000,\ V_{n+1} = V_n + 1880$

b \$29 400 **c** 0.18% **d** 17

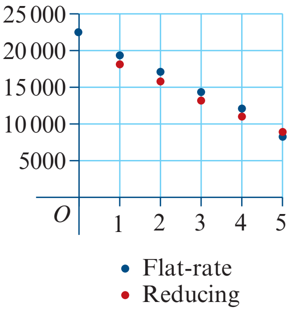
5 **a** **i** $V_0 = 22\,500,\ V_{n+1} = V_n - 2700$

ii \$9000

b **i** $V_0 = 22\,500,\ V_{n+1} = 0.84V_n$

ii \$9409.77

c



6 **a** \$0.20

b Let V_n be the value of the vacuum cleaner after cleaning n offices.

$V_0 = 650,\ V_{n+1} = V_n - 0.20$

c \$250

7 **a** \$6575 **b** \$6722.75 **c** 6.9%

8 **a** $a = 300,\ r = 2$ **b** 38 400 bacteria

c Day 12

9 **a** $a = 65\,000,\ r = 1.09$

b \$77 226.50 **c** 6th year

10 **a** $\dfrac{2}{3}$ **b** 144 cm **c** 1170 cm

11 **a** $V_0 = 30\,000,\ V_{n+1} = (1.006) \times V_n$

b \$30 000, \$30 180, \$30 361.08, \$30 543.25, \$30 726.51, \$30 910.86

c \$32 232.73 **d** \$33 410.65

12 \$328.03

13 **a** $F_0 = 840,\ F_{n+1} = 1.06\ F_n - 80,\ n \geq 0$

b 673

c 18

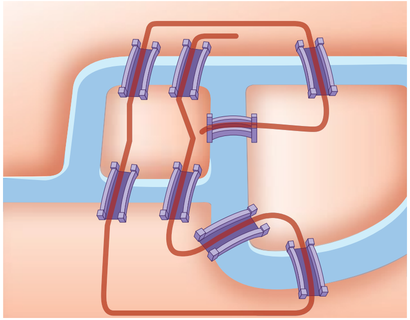
14 **a** $B_0 = 20,\ B_{n+1} = 0.95\ B_n + 3$ **b** 60

c 1 **d** 10%

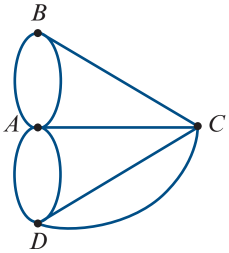
Chapter 6

Exercise 6A

1 **a** **i** Two solutions exist. One example is:



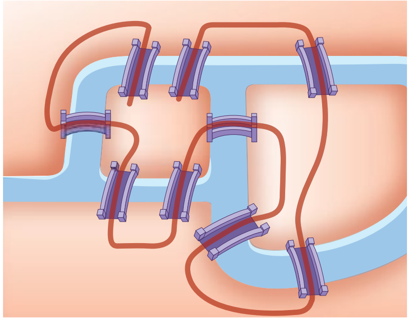
ii



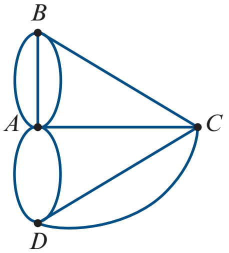
iii See graph above – the odd degree vertices are B and A

iv No answers – exploration

b **i** Many solutions exist. One example is:



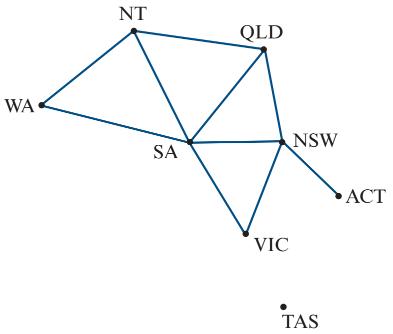
ii



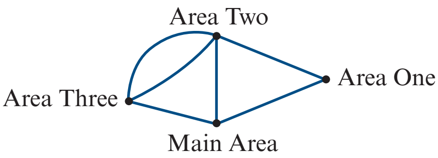
iii See graph above

iv No answers – exploration

2

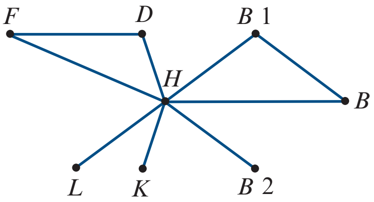


3 a



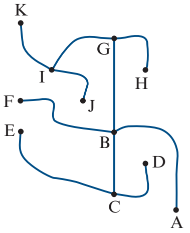
b 4 paths

4 a

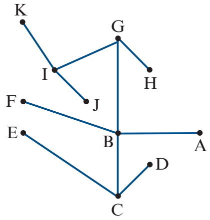


b No, it is not possible to start in the hallway and walk through each room only once because some rooms (e.g. Laundry, Kitchen, Bedroom 2) has only one door into and out of the room. This means that to walk through the house the real estate agent must go back to the hallway multiple times.

5 a

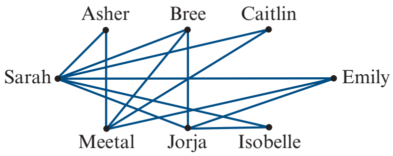


OR



b A – B – G – I – K

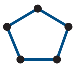
6



Exercise 6B

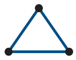
1 a i 5 **ii** 6 **iii** 0 **iv** 2
v 3 **vi** 4 **vii** 1
b i 4 **ii** 7 **iii** 1 **iv** 6
v 2 **vi** 2 **vii** 2
c i 4 **ii** 7 **iii** 0 **iv** 3
v 4 **vi** 2 **vii** 2
d i 8 **ii** 14 **iii** 2 **iv** 5
v 3 **vi** 8 **vii** 0

2 a 10;



many graphs are possible

b 6;



many graphs are possible

c 2;

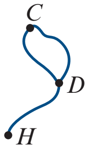
3 Because each edge must start and end at a vertex. It is a bit like shaking hands, there must be two hands at the end of each shake, even if you are shaking hands with yourself (a loop).

4 a Increase by two **b** Increase by one

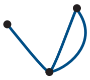
5 a i 3 **ii** 2 **iii** 1

b 14

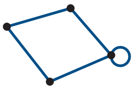
c



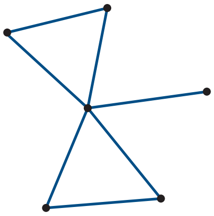
6 a



b



c

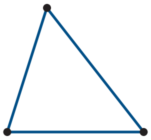


Exercise 6C

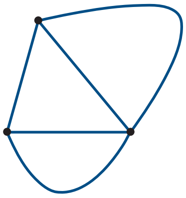
1 A, D, F

2 Many graphs are possible.
Examples include:

a



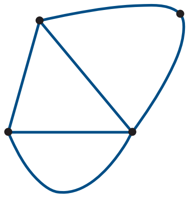
b



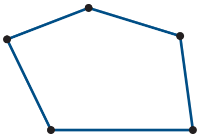
ISBN 978-1-009-25917-0

Cambridge University Press Updated January 2024

c

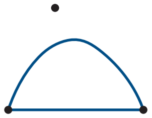


d

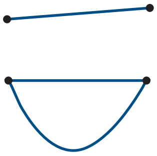


3 Many graphs are possible.
Examples include:

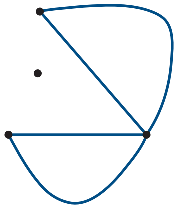
a



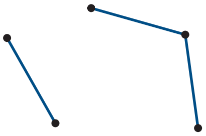
b



c



d



4 3 edges – try it and see

5 a BD **b** CB and AB

c XW and WV

6

7 a

$$\begin{array}{cc} & \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{array}{c} A \\ B \\ C \\ D \end{array} & \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \end{array}$$

b

$$\begin{array}{cc} & \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{array}{c} A \\ B \\ C \\ D \end{array} & \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \end{array}$$

c

$$\begin{array}{cc} & \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{array}{c} A \\ B \\ C \\ D \end{array} & \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \end{array}$$

d

$$\begin{array}{cc} & \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{array}{c} A \\ B \\ C \\ D \end{array} & \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \end{array}$$

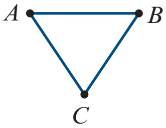
e

$$\begin{array}{cc} & \begin{array}{cccccc} A & B & C & D & E & F \end{array} \\ \begin{array}{c} A \\ B \\ C \\ D \\ E \\ F \end{array} & \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \end{array}$$

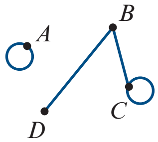
f

$$\begin{array}{cc} & \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{array}{c} A \\ B \\ C \\ D \end{array} & \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 2 \\ 0 & 1 & 2 & 0 \end{bmatrix} \end{array}$$

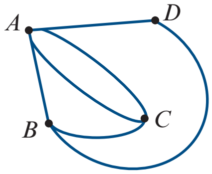
8 a



b



c



9 C is an isolated vertex.

10 Leading diagonals will all be '1'.

11

$$\begin{array}{cc} & \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} & \begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \end{array}$$

12 a

$$\begin{array}{cc} & \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{array}{c} A \\ B \\ C \\ D \end{array} & \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

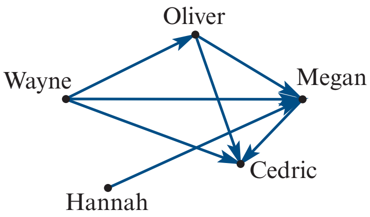
b 3 ways

c All buses end at route D and do not return to starting vertex A. For example, it is impossible for the buses to go from D to C, C to A, D to B, B to A, B to C. The direction of the buses is one way between each of the cities, whereas usually go in both directions between cities.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

13 a

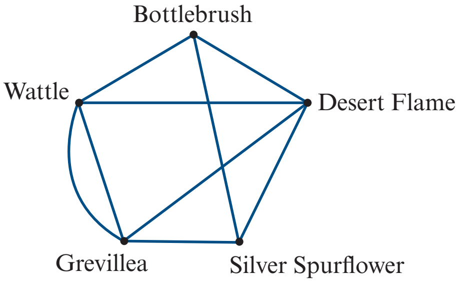


b 1 way; Hannah > Megan > Cedric

14 a

$$\begin{array}{c} \\ W \\ B \\ DF \\ SS \\ G \end{array}\begin{array}{c} \begin{array}{ccccc} W & B & DF & SS & G \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \end{array}$$

b



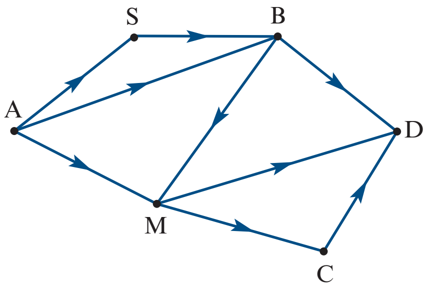
$$\begin{array}{c} \\ W \\ B \\ DF \\ SS \\ G \end{array}\begin{array}{c} \begin{array}{ccccc} W & B & DF & SS & G \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 0 & 2 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 2 & 0 & 1 & 1 & 0 \end{bmatrix} \end{array}$$

15 a

$$\begin{array}{c} \\ V \\ T \\ F \end{array}\begin{array}{c} \begin{array}{ccc} V & T & F \end{array} \\ \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix} \end{array}$$

b 4 routes

16



17 a

$$\begin{bmatrix} 0 & 4 & 0 & 0 & 0 & 1 & 0 & 1 \\ 4 & 0 & 2 & 1 & 0 & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 2 & 1 & 0 & 1 & 0 \end{bmatrix}$$

b **i** 4 routes

ii 23 routes

c

$$\begin{bmatrix} 0 & 4 & 0 & 0 & 0 & 1 & 0 & 1 \\ 4 & 0 & 2 & 1 & 0 & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 2 & 1 & 0 & 1 & 0 \end{bmatrix}^2$$

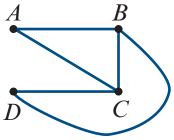
$$= \begin{bmatrix} 18 & 0 & 9 & 6 & 1 & 0 & 6 & 0 \\ 0 & 22 & 0 & 1 & 2 & 5 & 1 & 9 \\ 9 & 0 & 6 & 4 & 1 & 0 & 3 & 1 \\ 6 & 1 & 4 & 6 & 2 & 1 & 3 & 1 \\ 1 & 2 & 1 & 2 & 2 & 0 & 1 & 1 \\ 0 & 5 & 0 & 1 & 0 & 2 & 0 & 2 \\ 6 & 1 & 3 & 3 & 1 & 0 & 4 & 2 \\ 0 & 9 & 1 & 1 & 1 & 2 & 2 & 8 \end{bmatrix}$$

Exercise 6D

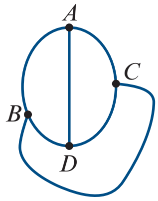
1 A, B, D, F

2 Many solutions are possible.
Examples include:

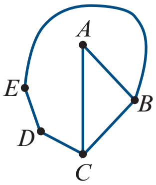
a



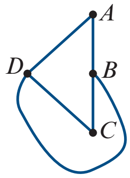
b



c



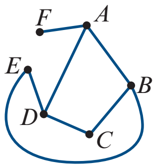
d



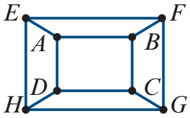
ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

e



f



3 **a** **i** $v = 4, e = 4, f = 2$
ii $v - e + f = 4 - 4 + 2 = 2$
b **i** $v = 7, e = 9, f = 4$
ii $v - e + f = 7 - 9 + 4 = 2$
c **i** $v = 7, e = 12, f = 7$
ii $v - e + f = 7 - 12 + 7 = 2$
d **i** $v = 7, e = 10, f = 5$
ii $v - e + f = 7 - 10 + 5 = 2$

4 **a** $f = 2$ **b** $v = 3$ **c** $e = 4$ **d** $v = 4$
e $f = 4$ **f** $f = 7$ **g** $e = 19$

5 Graph 1: $v = 4, e = 6, f = 4$;
$v - e + f = 4 - 6 + 4 = 2$
Graph 2: $v = 8, e = 12, f = 6$;
$v - e + f = 8 - 12 + 6 = 2$
Graph 3: $v = 6, e = 12, f = 8$;
$v - e + f = 6 - 12 + 8 = 2$
Graph 4: $v = 20, e = 30, f = 12$;
$v - e + f = 20 - 30 + 12 = 2$
Graph 5: $v = 12, e = 30, f = 20$;
$v - e + f = 12 - 30 + 20 = 2$

Exercise 6E

1 **a** Path **b** Open Trail **c** Path
d Walk **e** Open Trail **f** Path

2 **a** ii **b** i & ii **c** ii **d** i & iii

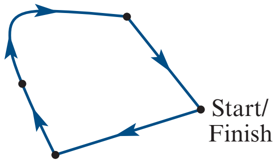
3 **a** Walk **b** Cycle **c** Path **d** Walk
e Path **f** Walk

Exercise 6F

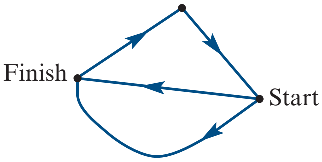
1 Other trails or cycles are possible in each case.

a Not traversable; more than two vertices odd

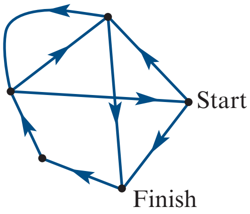
b Traversable; all vertices even.



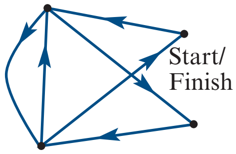
c Traversable; two vertices odd, the other even



d Traversable; two vertices odd, the rest even

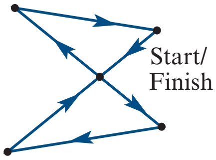


e Traversable; all vertices even

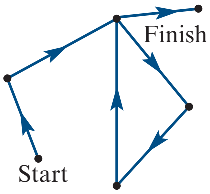


f Not traversable; more than two vertices odd

g Traversable; all vertices even



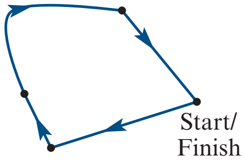
h Traversable; two vertices odd, the rest even



i Not traversable; more than two vertices odd

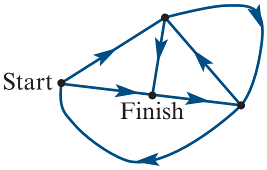
2 Other trails are possible in each case.

a Eulerian: all even vertices

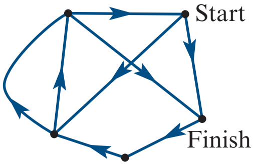


b Neither: more than two odd vertices

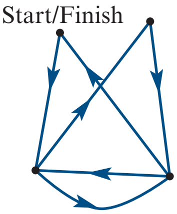
c Semi-Eulerian: two odd vertices, rest even



d Semi-Eulerian: two odd vertices, the rest even

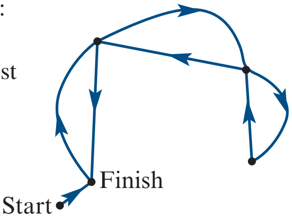


e Eulerian: all even vertices

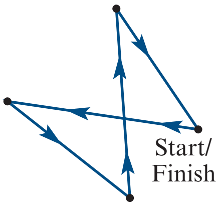


ISBN 978-1-009-25917-0

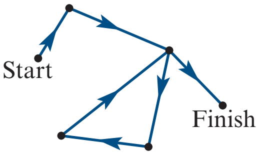
f Semi-Eulerian: two odd vertices, the rest even



g Eulerian: all even vertices



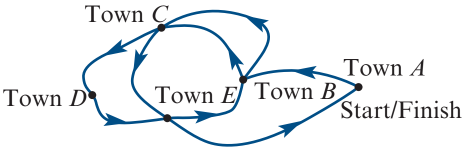
h Semi-Eulerian: two odd vertices, the rest even



i Neither: more than two odd vertices

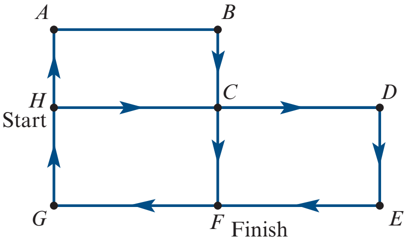
3 a Yes, all vertices even.

b Other routes are possible.

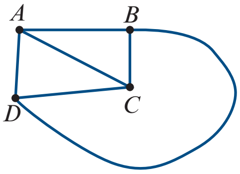


4 a No, not all even vertices.

b Several routes are possible. One is shown below.

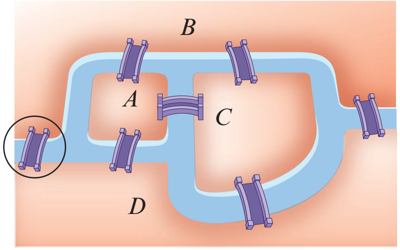


5 a

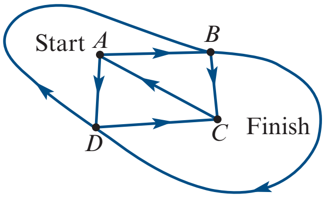


b An Eulerian trail does not exist. The graph has more than two odd vertices.

c **i** One possible solution circled is shown.



ii



iii The bridges can now be crossed only once in a single walk because an Eulerian trail now exists. The graph has two odd vertices and the rest are even. See the graph above for a possible route.

6 a Yes; it is an Eulerian graph because it has only even vertices.

b *K-M-G-D-E-G-K-E-S-K*

7 Other answers are possible.

a *A-F-G-B-C-H-E-D* **b** *F-A-B-C-D-E-H-G*

8 Other answers are possible.

a *A-B-C-D-E-F-A* **b** *A-B-C-D-E-A*

c *A-F-E-D-C-B-G-A*

d *A-B-C-F-I-H-E-G-D-A*

e No Hamiltonian cycle exists.

f *A-E-F-G-H-D-C-B-A*

9 a No **b** Yes: *C-D-E-B-A*, Hamiltonian path

c Yes: *E-A-B-C-D-E*, Hamiltonian cycle

10 a Yes: *K-M-T-L-S-E-D-G-K*, Hamiltonian cycle

b Yes: *D-E-S-L-T-M-G-K*, Hamiltonian path

Exercise 6G

1 a *D–E* **b** 17 minutes

c 8 minutes **d** 34 minutes

2 *A-C-D-E*; 11 hours

3 *A-B-D*; 35 m

4 *B-A-D*; $6

5 *B-G-A-F*; 7 minutes

6 *A-B-C-E-F-D-A;* 63 minutes

7 a *A* **b** *D*, *E* **c** *C* **d** 6

8 R| $\boxed{3}$ $\boxed{1}$ 7 4 4

9 a *P* **b** *U* **c** 1

10 a 10 **b** 16 **c** *A–B–C–E*

ISBN 978-1-009-25917-0

Cambridge University Press

Updated January 2024

11 **a** 7 **b** $A–E–G$ **c** 8
d $A–E–G–I$

12 **a** $S–B–D–F$, 12 **b** $S–A–C–D–F$, 10
c $S–B–D–F$, 15 **d** $S–A–E–G–F$, 19

13 19 km

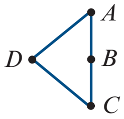
Chapter 6 review

Short-answer questions

1 **a** 6 **b** 9 **c** 4 **d** 2

2 14

3



4

$$\begin{array}{cc} & \begin{matrix} A & B & C & D \end{matrix} \\ \begin{matrix} A \\ B \\ C \\ D \end{matrix} & \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \end{array}$$

5 **a** walk **b** closed trail
c open trail **d** cycle

6 **a** 26 **b** $v = 9,\ e = 13,\ f = 6$
c $v - e + f = 2 \quad 9 - 13 + 6 = 2$

7 13

8 C

9 B

10 **a** B
b C

11 E

12 Many possible solutions: *AC, AD, AE, CD, CE* or *DE*

13 18

14 *F-E-D-A-B-C-F* or *F-C-B-A-D-E-F* (clock-wise or anti-clockwise)

15 A

16 Eulerian trail

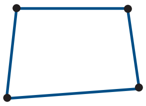
17 Hamiltonian cycle

18 6 **19** 25 **20** A **21** 7

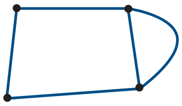
Extended-response questions

1 Many answers are possible. Examples:

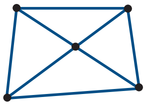
a



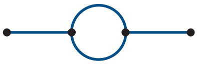
b



c

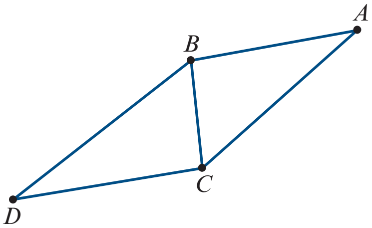


d

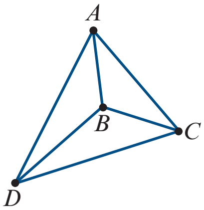


2 Other answers are possible in each case.

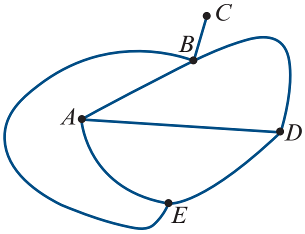
a



b



c



3 **a** $\deg(C) = 3$
b 2 odd, 2 even
c Other answers are possible, starting at *B* and ending at either *C* and tracing each edge once only.
Example: *B-A-C-B-D-C*

4

$$\begin{array}{cc} & \begin{matrix} A & B & C & D \end{matrix} \\ \begin{matrix} A \\ B \\ C \\ D \end{matrix} & \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \end{array}$$

5 **a** $\deg(C) = 4$
b No odd vertices, five even vertices.
c Other answers are possible, ending at *A* and tracing each edge once only.
Example: *A-B-C-D-E-B-A-E-C-D-A*

6 **a** No edges intersect, except at vertices.
b $v = 9, e = 14, f = 7; 9 - 14 + 7 = 2$
c 750 m
d No odd, 9 even
e
  i Yes, all vertices are even.
  ii Many answers are possible. Example: *P-C1-C8-C2-C1-C4-C2-C3-C4-C5-C7-C8-C6-C5-P*

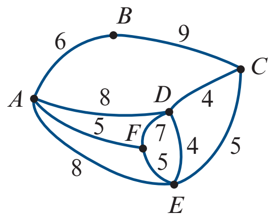
f 1270 m

g **i** Hamiltonian cycle

ii *C7* Park Office; other answers possible.

iii *P-C1-C2-C3-C4-C5-C6-C8-C7-P*, or the same route in reverse; other answers possible.

7 **a** **i**



ii 24

iii

$$\begin{array}{c|cccccc} & A & B & C & D & E & F \\ \hline A & 0 & 1 & 0 & 1 & 1 & 1 \\ B & 1 & 0 & 1 & 0 & 0 & 0 \\ C & 0 & 1 & 0 & 1 & 1 & 0 \\ D & 1 & 0 & 1 & 0 & 1 & 1 \\ E & 1 & 0 & 1 & 1 & 0 & 1 \\ F & 1 & 0 & 0 & 1 & 1 & 0 \end{array}$$

b **i** 45 km (at minimum)

ii Some vertices are visited more than once.

iii *F–E–D–C–B–A–F*

iv 33 km (for route above; other answers possible)

c *C* and *F*

8 **a** Pemberton and Manjimup

b 438 km

c Two of the vertices have an odd degree.

d **i** Eulerian trail. The graph is semi-Eulerian.

ii Bridgetown and Manjimup

Chapter 7

Exercise 7A

1

Feature	*Plot A*	*Plot B*	*Plot C*
Irregular fluctuations	✓	✓	✓
Increasing trend			✓
Decreasing trend	✓		
Cycles			
Outlier			✓

2

Feature	*Plot A*	*Plot B*	*Plot C*
Irregular fluctuations	✓	✓	✓
Increasing trend			✓
Decreasing trend	✓		
Cycles		✓	
Seasonality	✓		✓

3

Feature	*Plot A*	*Plot B*	*Plot C*
Irregular fluctuations	✓		✓
Structural change	✓		
Increasing trend	✓		
Decreasing trend	✓		
Seasonality			✓

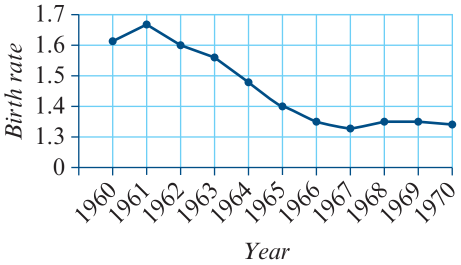
4 The demand for accommodation appears to be seasonal, at its lowest in the June quarter and peaking in the March quarter each year. The graph does not show any clear trend.

5 **a** The percentage of males who smoke has consistently decreased since 1945, while the percentage of females who smoke increased from 1945 to 1975 but then decreased at a similar rate to males over the period 1975–1992.

b Decrease

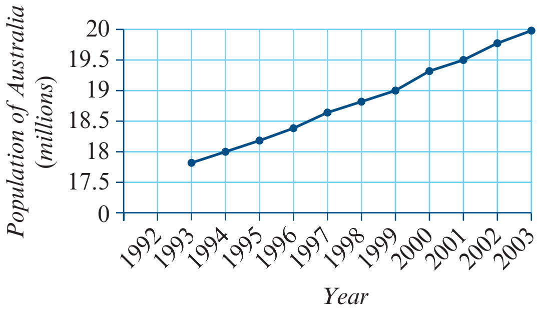
6 The number of whales caught increased rapidly between 1920 and 1930 but levelled off during the 1930s. In the period 1940–1945 there was a rapid decrease in the number of whales caught and numbers fell to below the 1920 catch. In the period 1945–1965 the numbers increased again but then fell again until 1985 when numbers were back to around the 1920 level.

7

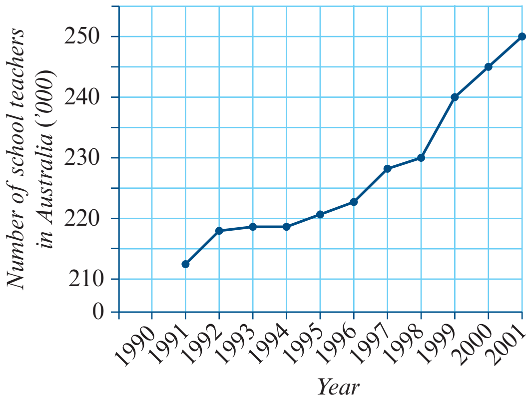


ISBN 978-1-009-25917-0

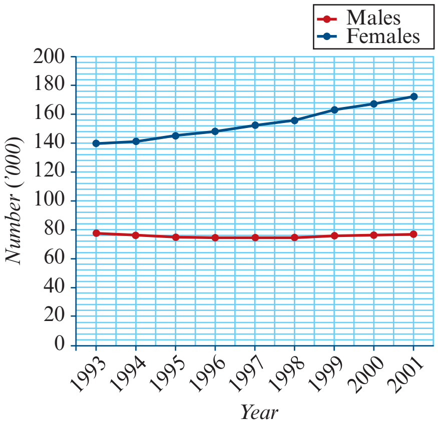
8



9



10 a



b The number of male school teachers has remained relatively constant over the years 1993–2001, whereas the number of female school teachers has increased over this time.

Exercise 7B

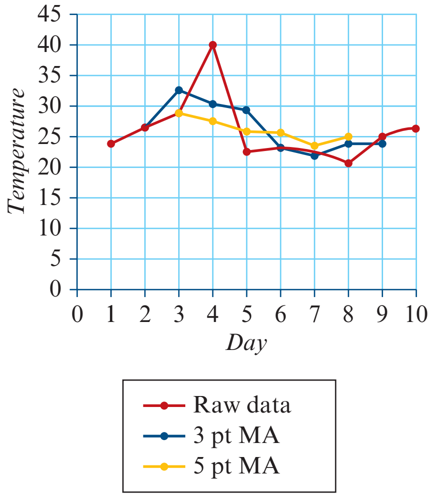
1 **a** 24.4 **b** 20.0 **c** 23.2 **d** 29.4
e 19.1

2 **a** 3 **b** 1 **c** 4 **d** 3.2
e 1.2 **f** 2.2 **g** 3.75 **h** 2
i 3.25 **j** 1.5

3

t	1	2	3	4	5	6	7	8	9
y	10	12	8	4	12	8	10	18	2
3pt MA	–	10	8	8	8	10	12	10	–
5pt MA	–	–	9.2	8.8	8.4	10.4	10	–	–

4 **a, c**

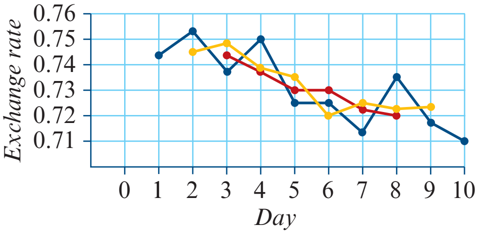


The smoothed plots show that the 'average' maximum temperature changes relatively slowly over the 10-day period (the 5-day average varies by only 5°) when compared to the daily maximum, which can vary quite widely (for example, nearly 20° between the fourth and fifth day) over the same period of time.

b

Day	Temperature (C°)	3-moving mean	5-moving mean
1	24	–	–
2	27	26.3	–
3	28	31.7	28.2
4	40	30.0	28.0
5	22	28.3	27.0
6	23	22.3	25.6
7	22	22.0	22.6
8	21	22.7	23.4
9	25	24.0	–
10	26	–	–

5 **a, c**



ISBN 978-1-009-25917-0

Cambridge University Press

Updated January 2024

● exchange rate
● 3-point MA
● 5-point MA

The exchange rate has a downward trend over the 10-day period. This is most obvious from the smoothed plots, particularly the 5-point moving average plot.

b

Day	*Exchange rate*	*3-point MA*	*5-point MA*
1	0.743	–	–
2	0.754	0.745	–
3	0.737	0.747	0.742
4	0.751	0.737	0.738
5	0.724	0.733	0.730
6	0.724	0.720	0.729
7	0.712	0.724	0.722
8	0.735	0.721	0.720
9	0.716	0.721	–
10	0.711	–	–

6

Month	*Number of births*	*2-point MA*	*2-point CMA*
January	10	–	–
February	12	11	10
March	6	9	7.25
April	5	5.5	9.5
May	22	13.5	16.75
June	18	20	17.75
July	13	15.5	12.75
August	7	10	9
September	9	8	8.75
October	10	9.5	9.25
November	8	9	10.25
December	15	11.5	

7

Month	*Internet usage*	*4-point MA*	*4-point CMA*
April	21	–	–
May	40	–	–
June	52	38.75	43.375
July	42	48	52.875
August	58	57.5	61.375
September	79	65	66.5
October	81	68	67
November	54	66	–
December	50		

Exercise 7C

1 a 1.0 **b** 7.8 **c** 6.7 **d** 3.9
e 6.9 **f** 30% **g** 10% **h** C

2 a 1.2 **b** 1514 **c** 1437 **d** 1005

3

Number of students:	56	125	126	96
Deseasonalised numbers:	112	125	97	80
Seasonal index:	0.5	1.0	1.3	1.2

4 a, c

Deseasonalised:	152	142	148	153
Seasonal index:	1.30	1.02	0.58	1.1

b In quarter 1 the restaurant chain employs 30% more waiters than the number employed in an average quarter.

5

Q1	*Q2*	*Q3*	*Q4*
0.89	0.83	1.12	1.16

6

Summer	*Autumn*	*Winter*	*Spring*
1.09	0.99	0.90	1.02

7 a $A = 136.6$, $B = 102.3$, $C = 94.5$

b Tues = 92.5, Wed = 109.4

c $D = 135.5$, $E = 137.0$

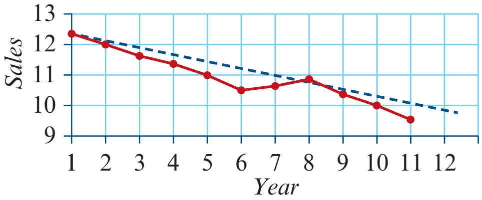
Exercise 7D

1 a There was a general increasing trend in the number of university students in Australia during the period 1992–2001.

b *students ('000s)* = 10.1 × *year* + 520.4
On average, the number of university students in Australia has increased by 10 100 per year.

c 2020 is Year 29. Predicted number in 2020 = 813 000 (to nearest thousand)

2 a, d



b General decreasing trend in the percentage of retail sales made in department stores

c *sales* = 12.5 − 0.258 × *year* (to 3 sig. figs)
The percentage of total retail sales that are made in department stores is decreasing by approximately 0.3% per year.

e 8.6%

3 a *age* = 27.14 + 0.187 × *year*
The average age of mothers having their first child increased by around 0.2 years each year between 1989 and 2002.

b 32.48 years. We are extrapolating 16 years beyond the period in which the data were collected.

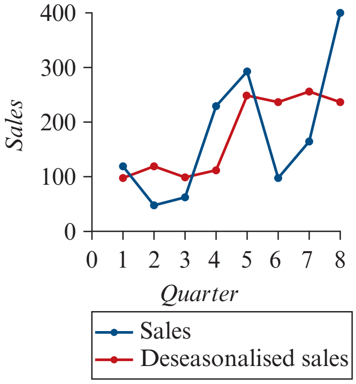
4 a *deseasonalised number* $= 50.9 + 1.59 \times$ *quarter number*

b deseasonalised number = 76.34
reseasonalised (actual) number = 90 (to the nearest whole number)

5 a

Year	*Quarter 1*	*Quarter 2*	*Quarter 3*	*Quarter 4*
1	122	128	118	130
2	250	245	263	236

b



The deseasonalised sales appear to show an increasing trend over time.

c *deseasonalised sales* $= 80.8 + 23.5 \times$ *quarter*

d forecasted actual sales $= 386.3 \times 1.13 = 437$ (to the nearest whole number)

6 a Deseasonalised values
See table in bottom of the page.

b *deseasonalised* $= 0.056 \times t + 148$

c 136.6 cents/litre

7 a

	Summer	*Autumn*	*Winter*	*Spring*
Seasonal index	1.17	1.03	0.83	0.97

b $d = 4.38t + 1007.25$

c 905 enrolments

Chapter 7 review

Short-answer questions

1 The time series graph shows an increasing trend with four points per season. The peaks occur at the fourth point in the cycle and the troughs occur at the first point in the cycle.

2 a 3.4 **b** 3.6 **c** 4.2 **d** 3.9

3 0.8

4 a 240

b 114

c 50% more than the seasonal average

5 0.82

6 a 29.5

b On average, the age of marriage for males is increasing by about 3 months per year

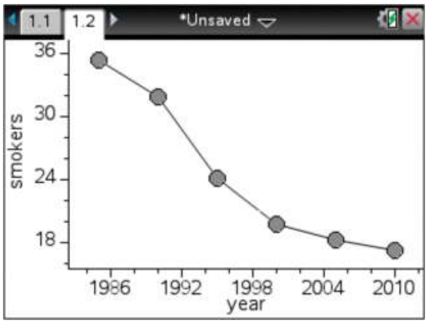
7 a *price* $= 1.25 \times$ *day* $+ 84.3$

b 89.3

c 63.1

Extended-response questions

1 a



b Decreasing trend

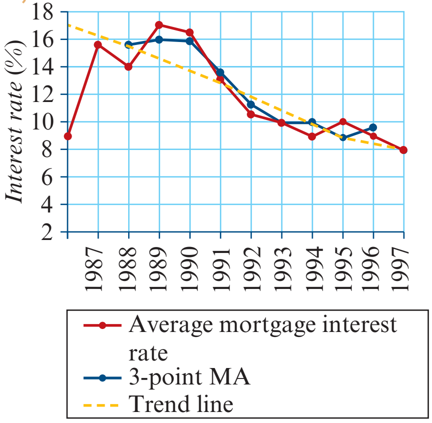
c $s = -0.7754\,t + 34.11$ **d** 13.17%

e 0.8488

2 a 3-point MA for 1992

$$= \frac{13.0 + 10.5 + 9.5}{3} = \frac{33}{3} = 11\%$$

b, c



c Both the raw data and the smoothed data reveal a steadily decreasing trend.

	Sun	*Mon*	*Tue*	*Wed*	*Thu*	*Fri*	*Sat*
Week 1	149.0	148.3	147.6	149.0	150.2	146.5	149.0
Week 2	151.2	150.5	149.8	149.0	148.4	148.4	146.9
Week 3	146.9	148.3	149.8	149.0	148.4	152.3	151.1

ISBN 978-1-009-25917-0 Cambridge University Press Updated January 2024

3 a $A = 184.25,\ B = 1.22,\ C = 182$

b

	Jan- Mar	Apr-Jun	Jul -Sep	Oct -Dec
Seasonal index	1.21	1.04	0.99	0.76

c $D = 163.46,\ E = 160.61$

d Deseasonalised = $-2.024t + 188.389$

e **i** 177

ii Not a valid prediction, since it is extrapolation

4 a 1.4830 **b** 835 **c** 1115

Chapter 8

Exercise 8A

1 a 0.413% **b** 2.08% **c** 0.238% **d** 0.142%
e 0.0348%

2 a 6.48% **b** 5.8% **c** 14.82% **d** 9.88%
e 8.03%

3 a 6.38% **b** 8.76% **c** 4.91% **d** 13.10%
e 7.64%

4 a More frequent compounds earn more interest.
b 4.68% **c** 4.70%
d The higher effective interest rate occurs with more frequent compounds.

5 a Less frequent compounds mean less interest is paid.
b 8.25% **c** 8.24%
d The lowest effective interest rate occurs with less frequent compounds.

6 a A – 8.62%, B – 8.11%
b A – \$3018, B – \$2837
c B – this loan will be charged less interest

7 a A – 5.43%, B – 5.61%
b A – \$7603, B – \$7860
c B – this investment will earn more interest

Exercise 8B

Note: The answers to questions in this exercise have been verified on TI-Nspire and Casio Classpad CAS calculators. Other finance solvers may use rounding differently and so produce slightly different answers.

1 a \$21 424.50 **b** \$9539.97
c \$6498.61 **d** \$29 545.73
e \$4884.39

2 a \$32 829.15 **b** \$12 384.40
c \$10 173.61 **d** \$2300.37
e \$11 114.02

3 142 months

4 313 weeks

5 5.9%

6 4.6%

7 \$44 884.73

Exercise 8C

1 a 1% **b** \$16.75 **c** 1025.37 **d** 342.17
e \$0.59 **f** \$2070.59
g \$70.60 or \$70.59 depending on method used. Difference due to accumulation of rounding errors.

2 a **i** \$4000 **ii** \$557.85 **iii** \$100
iv \$3072.85 **v** \$2591.83 **vi** \$558.02
b **i** 2.5% **ii** 10%
c A = \$64.80, B = \$1075.37, C = \$26.88

3 a \$1117.03
b no – \$626.83 (83 cents off)

4 a \$3366.47
b 7 repayments, final repayment \$0.26

5 a $V_0 = 2000,\ V_{n+1} = 1.005V_n - 339$
b \$674.10 **c** 7 months, final repayment \$1.17

6 a $B_0 = 10\,000,\ B_{n+1} = 1.03V_n - 2690.27$
b \$5147.75
c Yes – balance rounded to the nearest cent is zero.

7 a $V_0 = 3500,\ V_{n+1} = 1.004V_n - 280$
b $V_0 = 150\,000,\ V_{n+1} = 1.0014V_n - 650$

Exercise 8D

Note: The answers to questions in this exercise have been verified on TI-Nspire and Casio Classpad CAS calculators. Other finance solvers may use rounding differently and so produce slightly different answers.

1 a \$6061.91 **b** \$12 095.13
c \$168 519.40 **d** \$45 196.78
e \$33 735.99

2 a \$617.80 **b** \$413.38 **c** \$506.64
d \$4175.11 **e** \$687.65

3 a **i** 20 **ii** \$818.40
b **i** 70 **ii** \$179.99
c **i** 48 **ii** \$76.63
d **i** 118 **ii** \$1182.26
e **i** 243 **ii** \$1236.55

ISBN 978-1-009-25917-0

4 a \$857.09
b \$860.05
c \$308 555.36
d \$218 555.36

5 a \$197 793.85
b i \$2902.96 **ii** \$2901.92
iii \$418 025.25 **iv** \$178 025.25

6 a 55 **b** \$4746.24

7 a \$349.43 **b** \$48 865.60
c \$32 437.90 **d** \$418.66

8 a \$132 119.82 **b** \$1196.29

9 a \$229 994.24 **b** \$318 245.73
c \$169 798.80 **d** \$111 567.70
e \$1487.29

Exercise 8E

1 a \$2030.50 **b** no; \$1031.02

2 a \$3016.56 **b** no; \$1521.21

3 a $V_0 = 40\,000,\ V_{n+1} = 1.015V_n - 10\,380$
b \$20 293.30

4 a 0.25% **b** \$15.00 **c** \$5025.30
d \$3522.64
e A = 8.81, B = 3023.45, C = 1.97
f \$509.97 **g** \$6097.97 **h** \$97.97

5 59 months

6 22 quarters

7 \$474.81

8 a \$692.50 **b** \$73 212.95
c 153 months **d** \$684.73

9 a \$278 394.49 **b** 6.57%

Exercise 8F

1 a \$100 000 **b** 3.125%

2 a \$18 519 **b** 4.17%

3 \$200 000 **4** \$4791.67

5 a \$9790.50 **b** \$642 000
c \$642 000

Exercise 8G

1 a \$8805.26 **b** \$2000 **c** \$1000
d 8%

2 a \$27 689.06 **b** \$20 000 **c** \$2000
d 2.5% quarterly, 10% annually

3 $V_0 = 1500,\ V_{n+1} = 1.0075V_n + 40$

4 a i \$5000 **ii** \$100 **iii** \$50
iv 1% **v** 12%
vi \$5201.50 **vii** \$6242.85
b i $A = 5301.50,\ B = 53.02,\ C = 5454.52$
ii \$442.85

5 \$15 136.46

6 9 months

7 a \$42 378.59 **b** \$25 569.07

8 a \$81 939.67 **b** \$67 141.09

Chapter 8 review

Short-answer questions

1 8.4%

2 $V_0 = 18\,000,\ V_{n+1} = 1.068V_n + 2500$

3 a \$22 690.33
b \$1181.47

4 \$11 765

5 \$584.92

6 \$35 125

7 \$8040

8 a \$40 000 **b** \$400 **c** \$39 190.36
d 38.79%

Extended-response questions

1 a $V_0 = 250\,000$
$V_{n+1} = (1.004)\,V_n - 1800$
b \$240 185.91 **c** 56

2 a \$781.25 **b** \$147 298.48
c 38 months
d i 42 months **ii** \$3323.07

3 a \$656.72 **b** \$13 134
c \$3134.39

4 a \$2580.21 **b** \$2184.63
c \$2897.92 **d** \$2255.11

5 a 40 **b** \$320.78
c Yes, paying monthly will reduce the balance more often so less interest is paid.

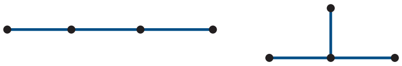
6 a \$247.04 **b** \$83 713.37

7 a \$1 175 244.58 **b** 291 months
c \$3427.80

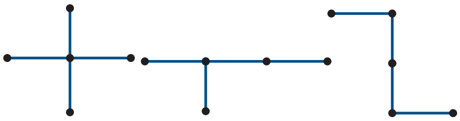
Chapter 9

Exercise 9A

1 a 14 **b** 6
c



d



2 8, 7

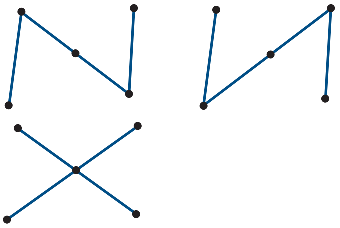
3 A, B, D

ISBN 978-1-009-25917-0

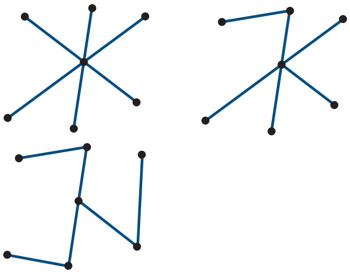
Cambridge University Press
Updated January 2024

4 Other answers are possible.

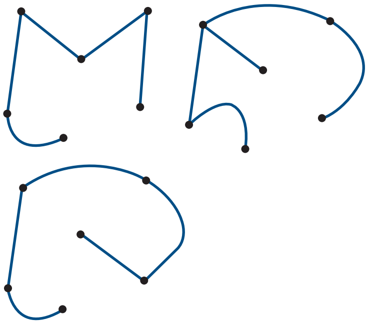
a



b

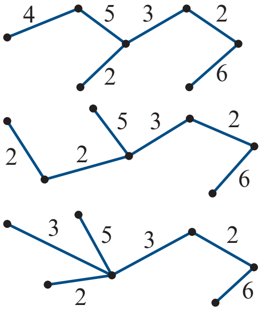


c



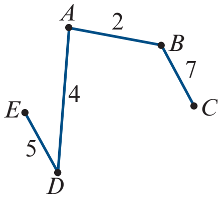
5 a 6

b



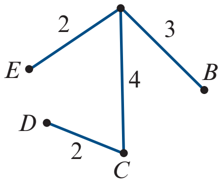
c 22, 20, 21

6 a



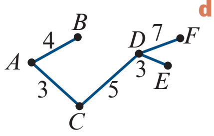
Length = 18

b



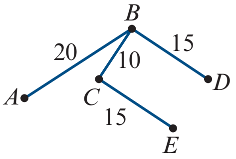
Length = 11

c



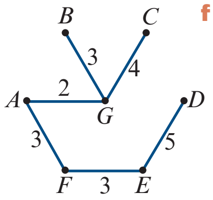
Length = 22

d



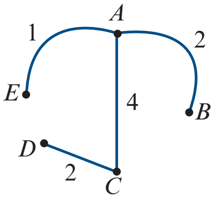
Length = 60

e



Length = 20

f



Length = 9

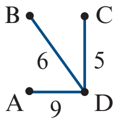
7 Length = 44 m

8 Length = 94 km

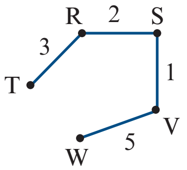
9 490 m

Exercise 9B

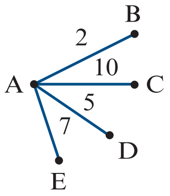
1 a Length = 20



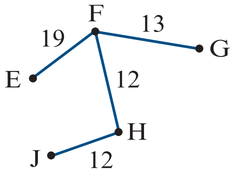
b Length = 11



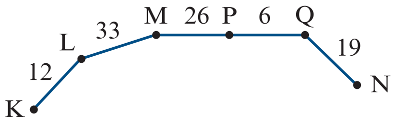
c Length = 24



d Length = 56



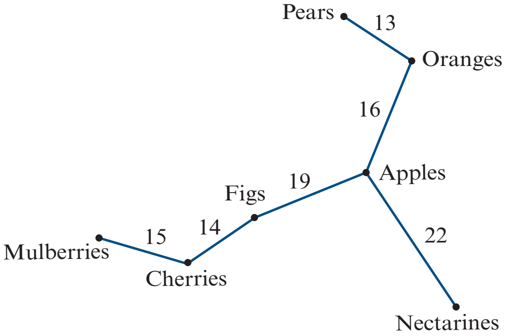
e Length = 96



ISBN 978-1-009-25917-0

2 a Length = 15 + 14 + 19 + 16 + 13 + 22 = 99 metres

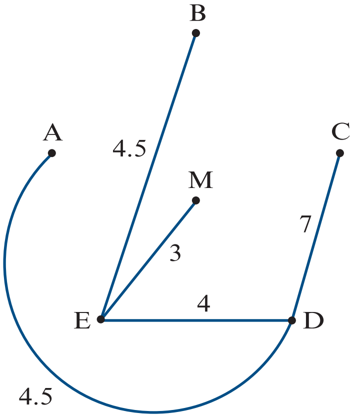
b



c Cost = $940.50

3 a Length = 23 metres

b



4 a 6 paths

b 612.5 m

c Minimum spanning tree must now include Entrance – Main Lawn (6.5) instead of edge Entrance – Wetlands (5). Length of minimum spanning tree increased by 1.5 cm, i.e. increased length by 37.5 metres.

Exercise 9C

1 a
$$\begin{array}{cc}\begin{matrix}A\\B\\C\end{matrix} & \begin{bmatrix}0&1&1\\1&0&1\\0&0&0\end{bmatrix}\end{array}$$

b
$$\begin{array}{cc}\begin{matrix}A\\B\\C\\D\end{matrix} & \begin{bmatrix}0&0&1&0\\1&0&0&1\\0&1&0&0\\0&1&1&0\end{bmatrix}\end{array}$$

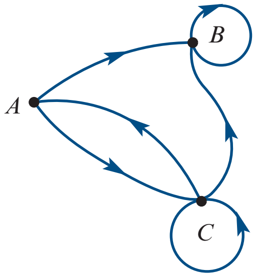
c
$$\begin{array}{cc}\begin{matrix}M\\N\\P\\Q\end{matrix} & \begin{bmatrix}0&1&0&1\\1&0&0&0\\0&1&0&0\\0&0&1&1\end{bmatrix}\end{array}$$

d
$$\begin{array}{cc}\begin{matrix}E\\F\\G\\H\end{matrix} & \begin{bmatrix}0&1&0&1\\1&0&1&0\\0&1&0&1\\0&0&1&0\end{bmatrix}\end{array}$$

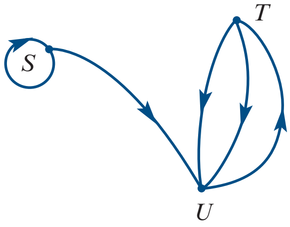
e
$$\begin{array}{cc}\begin{matrix}A\\B\\C\\D\\E\end{matrix} & \begin{bmatrix}0&1&1&0&0\\0&1&0&1&0\\1&1&0&1&0\\0&0&0&0&1\\0&0&1&1&0\end{bmatrix}\end{array}$$

f
$$\begin{array}{cc}\begin{matrix}A\\B\\C\\D\\E\\F\end{matrix} & \begin{bmatrix}0&1&0&0&0&0\\0&0&1&1&0&0\\1&0&0&1&0&0\\0&0&0&0&0&1\\0&0&1&1&0&0\\0&0&0&0&1&0\end{bmatrix}\end{array}$$

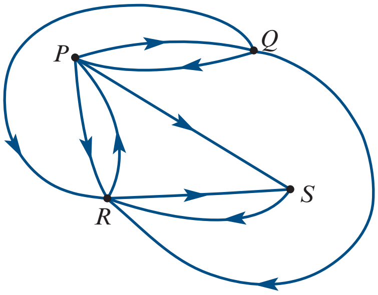
2 a



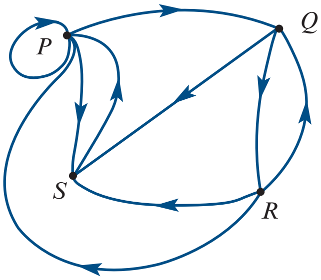
b



c



d

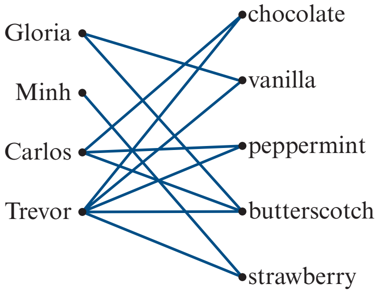


ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

3 **a** two distinct groups of vertices (people and flavours)

b

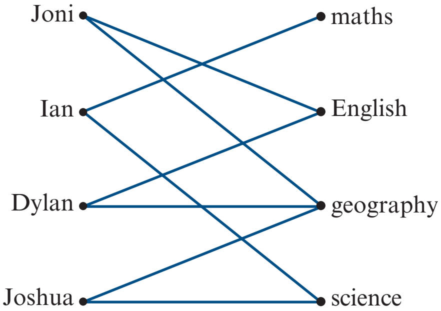


c 5

d Gloria

e Trevor – chocolate, Minh – strawberry.

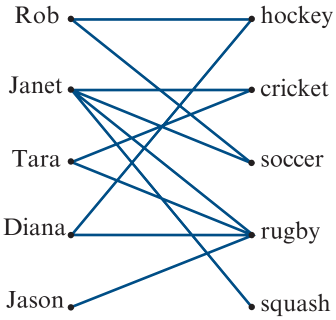
4 **a**



b Ian is the only teacher who can take maths, which means he cannot take science. Joshua is the only other teacher who can take science.

c Ian – maths, Joshua – science,
Joni – English, Dylan – geography
OR
Ian – maths, Joshua – science,
Joni – geography, Dylan – English

5 **a**



b Jason must coach rugby as this is the only sport he can coach. This means that Diana cannot coach rugby. The only other sport she can coach is hockey.

c Jason – rugby, Diana – hockey, Rob – soccer, Tara – cricket, Janet – squash

6 **a** $W–D$, $X–A$, $Y–B$, $Z–C$

b E.g., minimum cost is 11; $W–A$, $X–B$, $Y–D$, $Z–C$

7 Dimitri 800 m, John 400 m, Carol 100 m, Elizabeth 1500 m

8 Joe C, Meg A, Ali B

9 $A–Y$, $B–Z$, $C–X$, $D–W$

10 Champs home, Stars away, Wests neutral; or Champs neutral, Stars away, Wests home.
Cost = $20 000

11 A Mark, B Karla, C Raj, D Jess; or A Karla, B Raj, C Mark, D Jess
The distance will come to 55 km.

Exercise 9D

1 Subcontractor A: I
Subcontractor B: III
Subcontractor C: II
Subcontractor D: No allocation

2 **a** Swim: Kim
Cycle: Lauren
Run: Jasmine

b Time = 19 + 54 + 26 = 99 minutes

3 Driver 1 – Parcel B
Driver 2 – Parcel D
Driver 3 – No parcel
Driver 4 – Parcel A
Driver 5 – Parcel C

4 Adams – 1
Brown – 2
Cooper – 3

5 **a** Fred: Product 4
Gigi: Product 3
Hazel: Product 1
Ian: Product 2

b Fred = 37 × 8 = 296
Gigi = 25 × 8 = 200
Hazel = 21 × 8 = 168
Ian = 37 × 8 = 296

6 Clothing – Floor 1
Toys – Floor 3
Homewares – Floor 2
Electrical – Floor 4
Beauty – Ground floor
Profit = $327 000

Exercise 9E

1 $C_1 = 14,\ C_2 = 12,\ C_3 = 21$

2 $C_1 = 12,\ C_2 = 16,\ C_3 = 16$

ISBN 978-1-009-25917-0

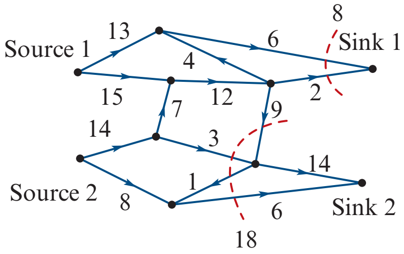
3 **a** 9 **b** 11 **c** 8 **d** 18

4 **a** A, 14; B, 23; C, 12; D, 16; E, not a cut that can be used to determine the maximum number of available seats.

b It does not prevent flow from source (Arlie) to sink (Bowen).

c 12

5 Sink 1 = 8, sink 2 = 18



Chapter 9 review

Short-answer questions

1 B

2 C

3 C

4 30

5 54

6 10

7 20

8 10

9 Travis – Volleyball; Fulvia – Athletics; Miriam – Basketball; Kieren – Swimming; Andrew – Tennis;

Extended-response questions

1 24

2 **a** 11 km **b** 17 km

3 **a** 135 km (there are two shortest paths).

b $v = 8, e = 12, f = 6 : 8 - 12 + 6 = 2$

c **i** This network does not have an Eulerian circuit as it contains two odd vertices

ii Dimboola, 556 km

iii H-S-M-H-W-Don-M-W-Dim-H-Nat-Nhill-Dim

d The Dimboola/Horsham road

e 241 km

4 **a** 26 **b** 15

5 Rob – breaststroke, Joel – backstroke, Henk – freestyle, Sav – butterfly or: Rob – breaststroke, Joel – butterfly, Henk – backstroke, Sav – freestyle. Time = 276 seconds

6 **a** 26 **b** 15

7 **a** Length = 330 metres

b Cost = $12 375

c Cost of the upgrade increases by $187.50 (additional 5 metres added to minimum spanning tree).

8 **a** $\begin{bmatrix} 30 & 21 & 13 \\ 19 & 18 & 22 \\ 23 & 7 & 15 \end{bmatrix}$

b Student 1 – C; Student 2 – A; Student 3 – B

9 Thursday – Ottilie; Friday – Percy; Saturday – Quinn; Sunday – Rachel

Total hours worked = 7 + 5 + 6 + 9

= 27 hours

10 Usher – Emma; Tickets – Asha; Cleaning – Dani; Food Service – Brian

Chapter 10

Exercise 10A

1 **a**

Activity	*Immediate predecessors*
A	–
B	–
C	*A*
D	*A*
E	*B*, *C*
F	*D*
G	*E*

b

Activity	*Immediate predecessors*
P	–
Q	*P*
R	*P*
S	*Q*
T	*Q*
U	*S*, *V*
V	*R*
W	*R*
X	*T*, *U*

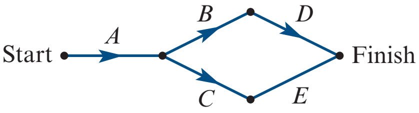
c

Activity	*Immediate predecessors*
P	–
Q	–
R	*P*
S	*P*
T	*Q*
U	*R*
V	*S*
W	*S*, *T*
X	*U*
Y	*W*
Z	*V*, *X*, *Y*

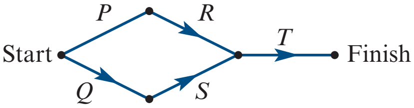
ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

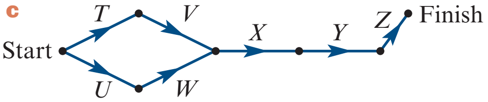
2 a



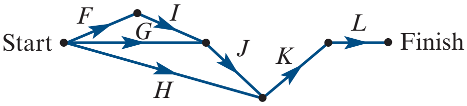
b



c

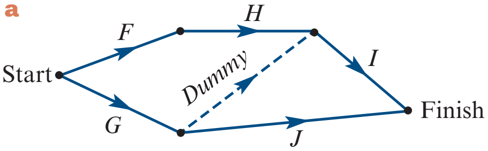


d

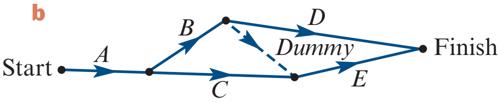


3

a



b



Exercise 10B

1 a $p = 12$ **b** $w = 10$ **c** $m = 8,\ n = 8$

d $a = 10,\ b = 12,\ c = 17$

e $f = 9,\ g = 12$

f $q = 8,\ r = 3,\ p = 2,\ n = 6$

2 a 3 **b** $A–C$ **c** 5 **d** 13

e 2

3 a 8 **b** 10 **c** 9 **d** 1

e 3 **f** 9

4 a $D–E–F$ **b** A: 1, B: 1, C: 15

5 a $B–E–F–H–J$

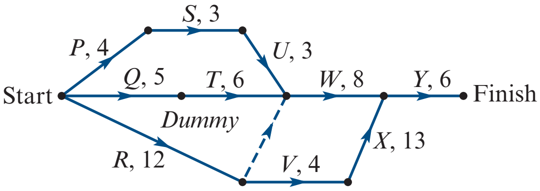
b A: 1, C: 14, D: 1, G: 1, I: 1

6 a

Activity	*Duration (weeks)*	*Immediate predecessors*
A	3	–
B	6	–
C	6	$A,\ B$
D	5	B
E	7	$C,\ D$
F	1	D
G	3	E
H	3	F
I	2	B

b $B–C–E–G$

7 a

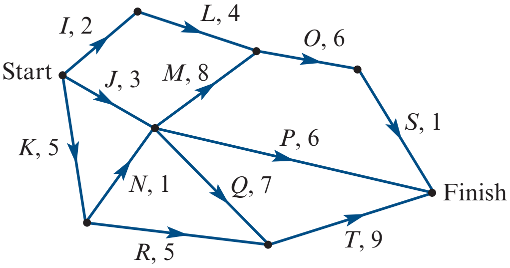


b

Activity	*EST*	*LST*
P	0	11
Q	0	10
R	0	0
S	4	15
T	5	15
U	7	18
V	12	12
W	12	21
X	16	16
Y	29	29

c $R–V–X–Y$ **d** 35 weeks

8 a



b

Activity	*EST*	*LST*
I	0	9
J	0	3
K	0	0
L	2	11
M	6	7
N	5	5
O	14	15
P	6	16
Q	6	6
R	5	8
S	20	21
T	13	13

c $K–N–Q–T$ **d** 22 weeks

Exercise 10C

1 a $B–E–H–J$ **b** 2 hours

c 6 hours **d** 14 hours

2 a $B–C–G–I$

b None, A is not on the critical path.

c 6 hours

ISBN 978-1-009-25917-0

3 **a** 18 hours

b $B–C–D–F–G–H$

c E is not on the critical path. It has slack time and reducing it will have no further effect.

d $110

Chapter 10 review

Short-answer questions

1 $A – D – H – I – K$

2 24 days

3 17 weeks

4 8 hours

5 E

6

Activity	Immediate predecessors
A	–
B	–
C	*A*
D	*A,B*
E	*C*
F	*D,E*

7 **a** 13 hours **b** 8 hours **c** 2 hours

d $B – F – I – J$

Extended-response questions

1 **a**

Activity	Immediate predecessors
J	–
K	–
L	*J*
M	*N*
N	*K*
O	*K*
P	*N*
Q	*L, M*
R	*P*
S	*O, R*
T	*Q*

b

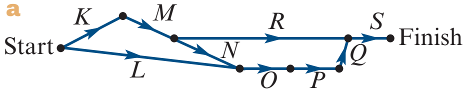
Activity	Immediate predecessors
A	–
B	–
C	*A*
D	*A*
E	*D, B*
F	*C, E*
G	*D, B*
H	*B*

c

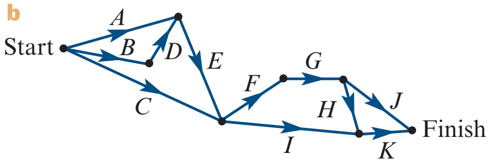
Activity	Immediate predecessors
A	–
B	*A*
C	*A*
D	*A*
E	*B*
F	*C, D*
G	*D*
H	*E, F, G*
I	*G*
J	*I*
K	*H*

2

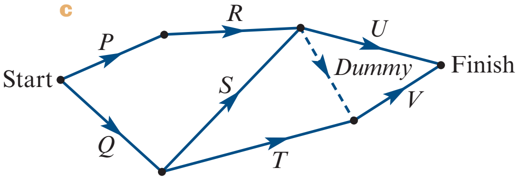
a



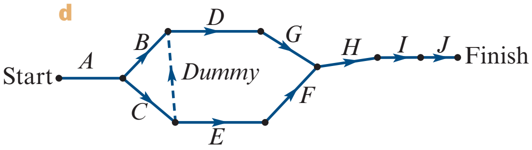
b



c



d



3 **a** Remove panel.

b ‘Order component’ and ‘Pound out dent’

4 **a** $A \rightarrow 1$, $D \rightarrow 4$, $F \rightarrow 10$, $K \rightarrow 12$,

b $B–C–E–G–J–K$

5 **a** 5 **b** 24 hours **c** 7 hours

6 **a** A, B, C

b LST for B is 1, EST for E is 10, LST for I is 18

c **i** $A–D–F–I–J$ **ii** 27 months

d **i** $B–C–D–F–I–J$ **ii** 25 months

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Appendix

A

TI-Nspire CAS CX with OS4.0 (digital only)

Keystroke actions and short cuts for the TI-Nspire CAS CX

esc : removes menus and dialogue boxes

ctrl + esc : undo last move

⇧shift + esc : redo last move

tab : move to next entry box (field)

ctrl + tab : switch applications in split screen

Navpad (Touchpad)

ctrl : accesses secondary (blue) commands

ctrl + ▲: displays page sorter

ctrl + ◀: displays previous page

ctrl + ▶: displays next page

⌂on :displays icon page to select applications, mode, My Documents and start a new document

menu :options for each application

ctrl + menu :contextual menus (same as right mouse click)

☝ : mouse pointer (cursor). Selects items.

ctrl + ☝ : grab

del : backspace, deletes a character

📖 : catalogue

⊞ : 2D maths template

ctrl + ÷ : adds fraction template

enter : completes commands and displays results

ISBN 978-1-009-25917-0

Mode Settings

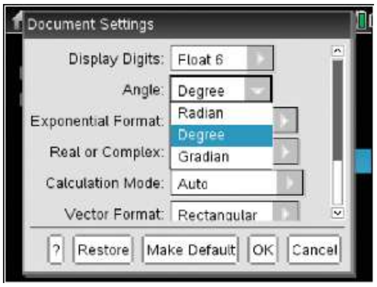
How to set in Degree mode

For this subject it is necessary to set the calculator to **Degree** mode right from the start. The calculator will remain in this mode unless you change the setting again or update the operating system.

Press ⌂on and move to **Settings>Document Settings**, arrow down to the down to the **Angle** field, press ▶ and select **Degree** from the list, and then arrow down to the **Make Default** tab. Select **OK** to accept the change.

Note that there is a separate settings menu for the **Graphs Geometry** pages. These settings are accessed from the relevant pages. For Mathematics it is not necessary for you to change these settings.

Note: When you start your new document you will see **DEG** in the top status line.


Document Settings

Display Digits: Float 6

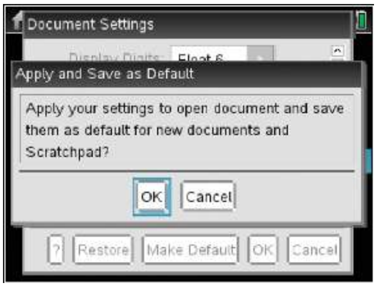
Angle: Degree

Exponential Format:

Real or Complex:

Calculation Mode: Auto

Vector Format: Rectangular


Document Settings

Apply and Save as Default

Apply your settings to open document and save them as default for new documents and Scratchpad?

How to set in Approximate (Decimal) mode

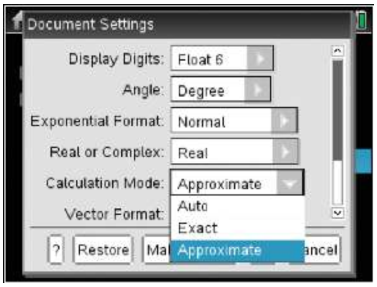
For this subject it is useful to set the calculator to **Approximate (Decimal)** mode right from the start. The calculator will remain in this mode unless you change the setting again or update the operating system.

Press ⌂on and move to **Settings>Document Settings**, arrow down to the **Calculation Mode** field, press ▶ and select **Approximate** from the list, and then arrow down to the **Make Default** tab. Select **OK** to accept the change.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Note that you can make both the **Degree** and **Approximate Mode** selections at the same time if desired.


Document Settings

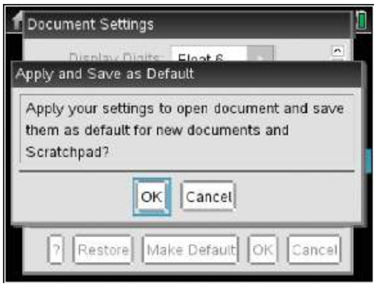
Display Digits: Float 6

Angle: Degree

Exponential Format: Normal

Real or Complex: Real

Calculation Mode: Approximate


Apply and Save as Default

Apply your settings to open document and save them as default for new documents and Scratchpad?

The home screen is divided into two main areas: **Scratchpad** and **Documents**.

All instructions given in the text, and in Appendix A, are based on the **Documents** platform.

Documents

Documents can be used to access all the functionality required for this subject including all calculations, graphing, statistics and geometry.

Starting a new document

1 To start a new document, press ⌂on and select **New Document**.

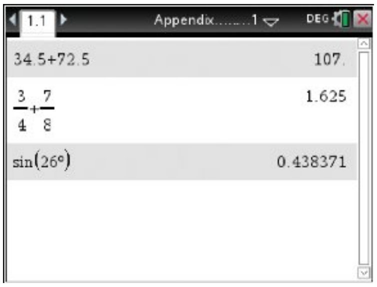
2 If prompted to save an existing document move the cursor to **No** and press enter.

Note: Pressing ctrl+N will also start a new document.

A: Calculator page – this is a fully functional CAS calculation platform that can be used for calculations such as arithmetic, algebra, finance, trigonometry and matrices. When you open a new document select **Add Calculator** from the list.

ISBN 978-1-009-25917-0

1 You can enter fractions using the fraction template if you prefer. Press ctrl ÷ to paste the fraction template and enter the values. Use the tab key or arrows to move between boxes. Press enter to display the answer. Note that all answers will be either whole numbers or decimals because the mode was set to approximate (decimal).



34.5+72.5	107.
$\frac{3}{4}+\frac{7}{8}$	1.625
$\sin(26°)$	0.438371

2 For problems that involve angles (e.g. evaluate sin(26°)) it is good practice to include the degree symbol even if the mode is set to degree (DEG) as recommended.

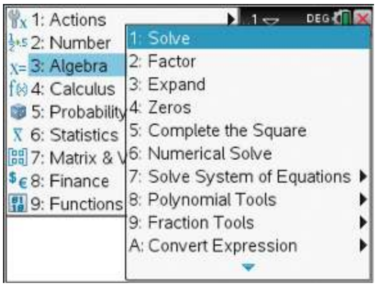
Note: If the calculator is accidentally left in radian (RAD) mode the degree symbol will override this and compute using degree values.

The degree symbol can be accessed using ?!▸. Alternatively select from the **Symbols** palette ctrl.To enter trigonometry functions such as *sin*, *cos*, etc., press the trig key or just type them in with an opening parenthesis.

Solving equations

Using the **Solve command**, solve $2y + 3 = 7$ for y.

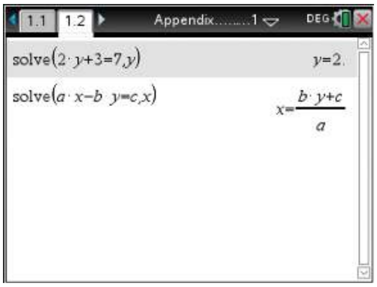
In a **Calculator** page press menu**>Algebra>Solve** and complete the **Solve** command as shown opposite. You must include the variable you are making the subject at the end of the command line.



Hint: You can also type in **solve(** directly from the keypad but make sure you include the opening bracket.

Literal equations such as $ax - by = c$ can be solved in a similar way.

Note that you must use a multiplication sign between two letters.



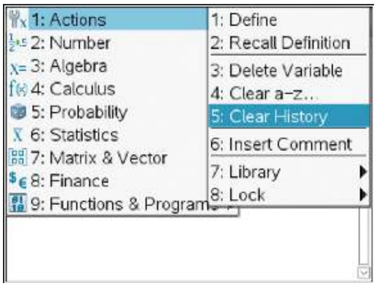
$\text{solve}(2\cdot y+3=7,y)$	$y=2.$
$\text{solve}(a\cdot x-b\cdot y=c,x)$	$x=\frac{b\cdot y+c}{a}$

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

Clearing the history area

Once you have pressed `enter` the computation becomes part of the **History** area. To clear a line from the history area, press ▲ repeatedly until the expression is highlighted and press `enter`. To completely clear the History Area, press `menu`**>Actions>Clear History** and press `enter` again.

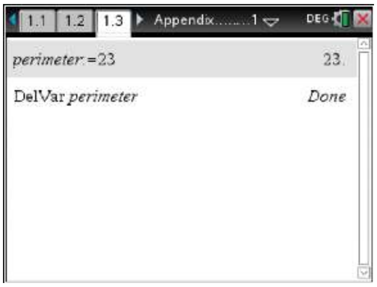


Alternatively press `ctrl`+`menu` to access the contextual menu.

It is also useful occasionally to clear any previously stored values. Clearing **History** does not clear stored variables.

Pressing `menu`**>Actions>Clear a-z...** will clear any stored values for single letter variables that have been used.

Use `menu`**>Actions>Delete Variable** if the variable name is more than one letter. For example, to delete the variable perimeter, then use **DelVar *perimeter***.



```
perimeter:=23        23
DelVar perimeter     Done
```

Note: When you start a new document any previously stored variables are deleted.

How to construct parallel box plots from two data lists

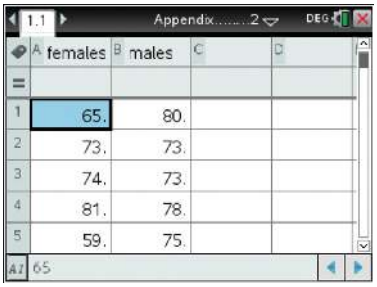
Construct parallel box plots to display the pulse rates of 23 adult females and 23 adult males.

Pulse rate (beats per minute)	
Females	*Males*
65 73 74 81 59 64 76 83 95 70 73 79 64 77 80 82 77 87 66 89 68 78 74	80 73 73 78 75 65 69 70 70 78 58 77 64 76 67 69 72 71 68 72 67 77 73

Steps

1 Start a new document: ctrl + N

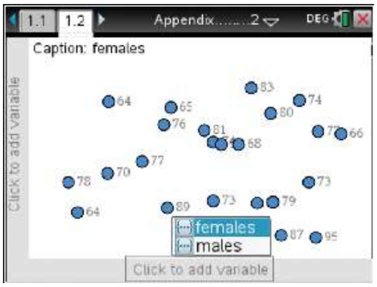
2 Select **Add Lists & Spreadsheet**. Enter the data into lists called ***females*** and ***males*** as shown.



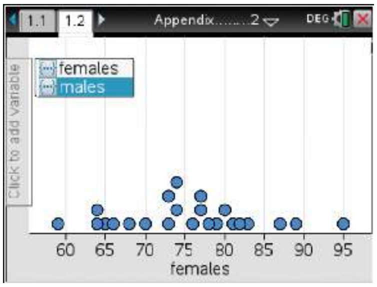
	A females	B males
1	65.	80.
2	73.	73.
3	74.	73.
4	81.	78.
5	59.	75.

3 Statistical graphing is done through the **Data & Statistics** application.
Press ctrl + I and select **Add Data & Statistics**.
(or press on and arrow 📊 to and press enter).

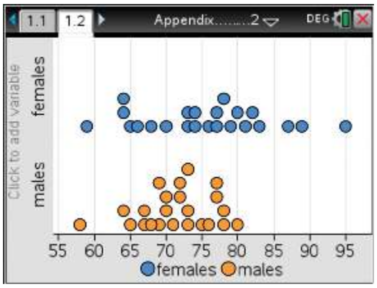
Note: A random display of dots will appear – this indicates list data are available for plotting. It is not a statistical plot.



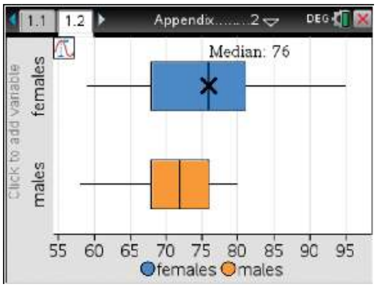
a Press tab, or navigate and click on the 'Click to add variable' box to show the list of variables. Select the variable, ***females***. Press enter or ↖ to paste the variable to the x-axis. A dot plot is displayed by default as shown.



b To add another variable to the x-axis press menu**>Plot Properties>Add X Variable**, then enter. Select the variable ***males***. Parallel dot plots are displayed by default.



c To change the plots to box plots press menu**>Plot Type>Box Plot**, and then press enter. Your screen should now look like that shown opposite.



Use ▼ to trace the other plot.

Press esc to exit the **Graph Trace** tool.

ISBN 978-1-009-25917-0

Cambridge University Press
Updated January 2024

4 Data analysis

Use menu**>Analyze>Graph Trace** and use the cursor arrows to navigate through the key points. Alternatively just move the cursor over the key points. Starting at the far left of the plots, we see that, for females, the:

- minimum value is 59: **MinX = 59**
- first quartile is 68: **Q1= 68**
- median is 76: **Median = 76**
- third quartile is 81: **Q3 = 81**
- maximum value is 95: **MaxX = 95**

and for males, the:

- minimum value is 58: **MinX = 58**
- first quartile is 68: **Q1 = 68**
- median is 72: **Median = 72**
- third quartile is 76: **Q3 = 76**
- maximum value is 80: **MaxX = 80**

Appendix

B Casio ClassPad II (digital only)

Operating system

Written for operating system 2.0 or above.

Terminology

Some of the common terms used with the ClassPad are:

The menu bar

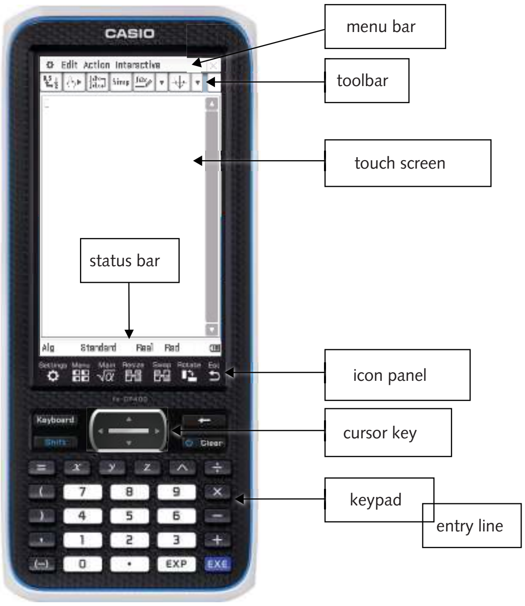
The toolbar

The touch screen contains the work area where the input is displayed on the left and the output is displayed on the right. Use your finger or stylus to tap and perform calculations.

The icon panel contains seven permanent icons that access settings, applications and different view settings. Press **escape** to cancel a calculation that causes the calculator to freeze.

The cursor key works in a similar way to a computer cursor keys

The keypad refers to the hard keyboard



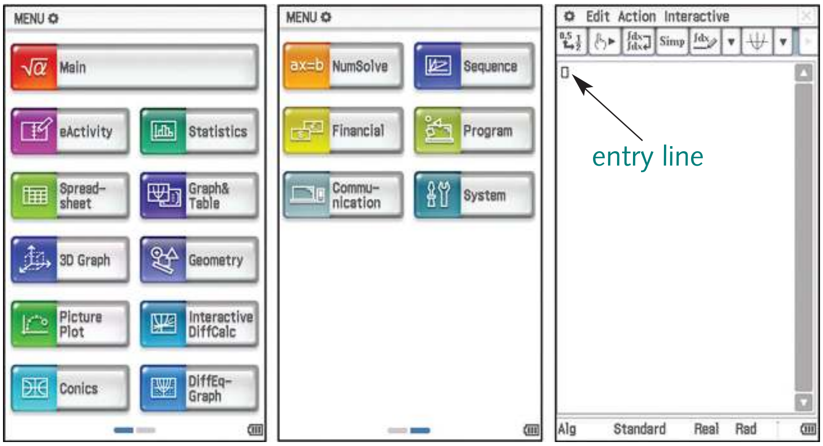
Calculating

Tap Menu from the **icon panel** to display the application menu if it is not already visible.

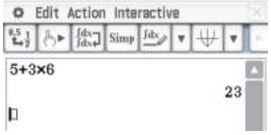
Tap √α Main to open the **Main** application.

Note: There are two application menus. Alternate between the two by tapping on the screen selector at the bottom of the screen.

1 The main screen consists of an entry line (which is recognised by a flashing vertical line (cursor) inside a small square). The history area, showing previous calculations, is above the entry line.



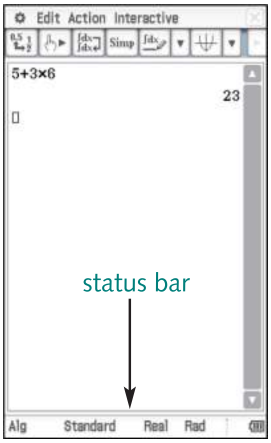
2 To calculate, enter the required expression in the entry line and press EXE. For example, if we wish to evaluate 5 + 3 × 6, type the expression in the entry line and press EXE.
You can move between the entry line and the history area by tapping or using the cursor keys ▬ (i.e. ▼, ◀, ▲, ▶).



5+3×6

23

3 The ClassPad gives answers in either exact form or as a decimal approximation. Tapping settings in the **status bar** will toggle between the available options.



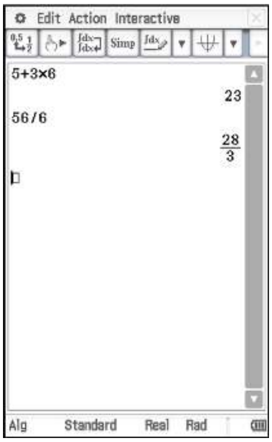
5+3×6

23

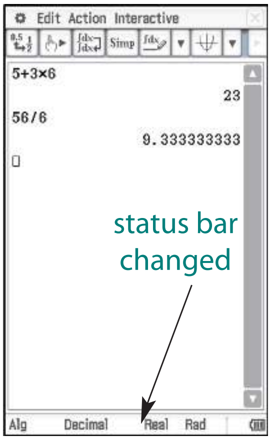
ISBN 978-1-009-25917-0

4 For example, if an exact answer is required for the calculation 56 ÷ 6, the **Standard** setting must be selected.

5 If a decimal approximation is required, change the **Standard** setting to **Decimal** by tapping it and press EXE



```
5+3×6
                    23
56/6
                  28/3
```



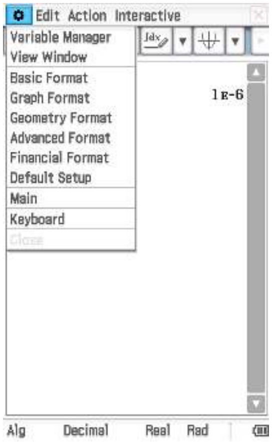
```
5+3×6
                    23
56/6
           9.333333333
```

status bar changed

Extremely large and extremely small numbers

When solving problems that involve large or small numbers the calculator's default setting will give answers in scientific form.

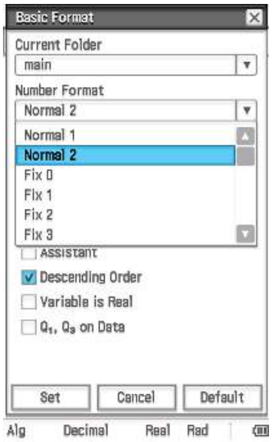
For example, one millionth, or $\frac{1}{1000000}$, in scientific form is written as 1×10^{-6} and the calculator will present this as 1E-6.



```
                  1E-6
```

To change this setting, tap on the settings icon and select **Basic Format.**

Under the Number Format select **Normal 2** and tap SET.



Basic Format

Current Folder: main

Number Format: Normal 2

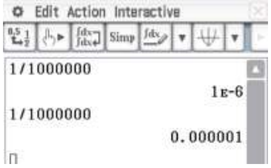
☐ Assistant

☑ Descending Order

☐ Variable is Real

☐ Q_1, Q_3 on Data

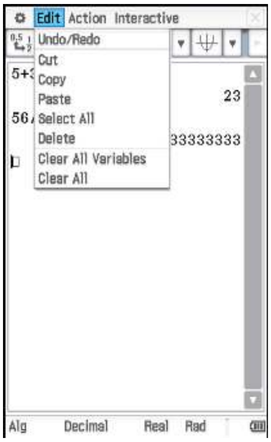
In the Main screen type $\frac{1}{1000000}$ and press EXE.
The answer will now be presented in decimal form 0.000001.
This setting will remain until the calculator is reset.



```
1/1000000
                 1E-6
1/1000000
             0.000001
```

Clearing the history screen

To clear the **Main** application screen, select **Edit** from the menu bar and then tap **Clear All**. Confirm your selection by tapping **OK**. The entire screen is now cleared. To clear the entry line only, press Clear on the front calculator.



```
5+3
                 23
56/
           33333333
```

```
5+3×6
                 23
56/6
```

Clear All

Are you sure?

Clearing variables

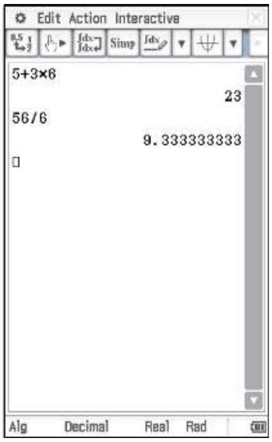
To clear stored variable values, select **Edit** from the menu bar and then tap **Clear All Variables**. Confirm your selection by tapping OK.

The variables are cleared but the history created on the main screen is kept.

```
5+3×6
                 23
```

Clear All Variables

Are you sure you want to clear all variables?



```
5+3×6
                 23
56/6
         9.333333333
```

Degree mode

When solving problems in trigonometry, your calculator should be kept in **Degree** mode.

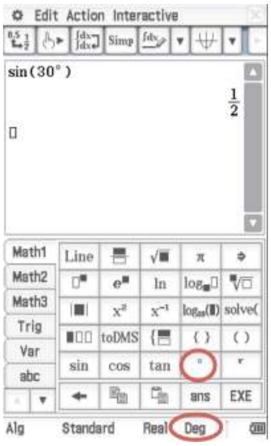
In the main screen, the status bar displays the angle mode.

Alg Decimal Real Rad

To change the angle mode, tap on the angle unit in the status bar until **Deg** is displayed.

In addition, it is recommended that you always insert the degree symbol after any angle. This overrides any mode changes and reminds you that you should be entering an angle, not a length.

The degree symbol is found in the Math1 keyboard.



sin(30°)

$\frac{1}{2}$

Alg Standard Real Deg

ISBN 978-1-009-25917-0

Cambridge University Press

Updated January 2024