PEARSON

GENERAL MATHEMATICS

QUEENSLAND

STUDENT BOOK

12

UNITS 3 & 4

Pearson

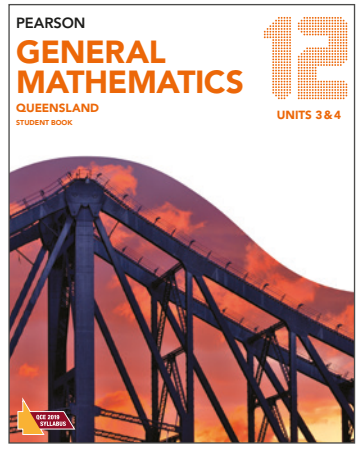
PEARSON

GENERAL MATHEMATICS

QUEENSLAND

12

UNITS 3&4



Writing and development team

We are grateful to the following people for their time and expertise in contributing to **Pearson General Mathematics 12**.

Nicola Silva
Retired teacher, QLD
Author

Geoff Phillips
Mathematics author, VIC
Contributing author

Daniel Hernandez Navas
Lead publisher
Content and learning specialist
Pearson Australia

Lindy Bayles
Mathematics teacher, VIC
Development Editor

Kylie Agnew
Mathematics teacher, QLD
Reviewer

Rodney Anderson
Mathematics teacher, QLD
Answer checker

Chris Brennan
Mathematics teacher, VIC
Teacher support author

Scott Brown
Leading teacher – STEAM Programs, VIC
Teacher support author
Reviewer

David Coffey
Mathematics author, VIC
Teacher support author

Heath Copeland
Mathematics teacher, SA
Answer checker

Andrew Duncan
Mathematics teacher and HoD, QLD
Teacher support author
Reviewer and answer checker

Emily Frazer
Mathematics teacher, QLD
Answer checker

Paul Grace
Mathematics teacher, VIC
Answer checker

Robyn Johnson
Mathematics teacher, QLD
Teacher support author

Sergei Pushkarev
Mathematics teacher, QLD
Reviewer and answer checker

Julie Watson-Hayward
Associate lecturer, Griffith University, QLD
Reviewer and answer checker

Contents

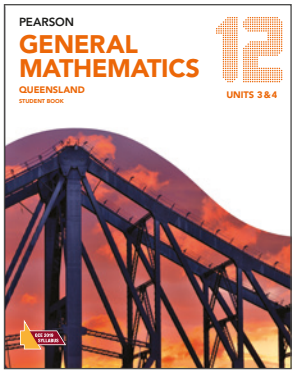
PEARSON

GENERAL MATHEMATICS

QUEENSLAND

12

UNITS 3&4



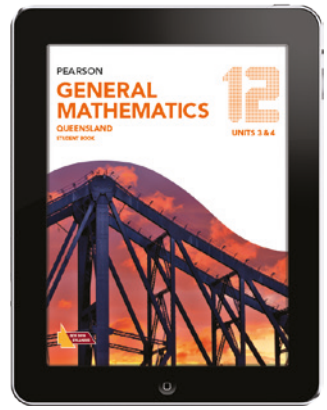
General Mathematics 12
Student book

Student book

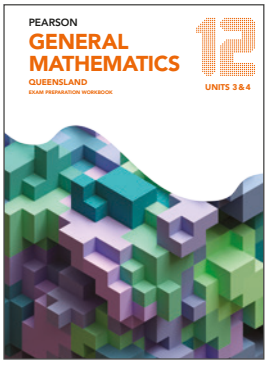
The student book has been written by local authors, ensuring quality content and complete curriculum coverage for Queensland, enabling students to prepare with ease and confidence. We have covered the breadth of the content within our exercise questions, from simpler skills-focused questions to those using unfamiliar contexts and application of the theory learnt. The theory, worked examples and question sets are written in line with the assessment objectives, with the aim of familiarising students with QCE cognitive verbs in the process of dependent and guided instruction. Additional interactives that help explain the theory and consolidate concepts have been included throughout all chapters

Pearson Reader+

Pearson Reader+ is our next-generation eBook. This is an electronic textbook that students can access on any device, online or offline, and is linked to features, interactives and visual media that will help consolidate their understanding of concepts and ideas, as well as other useful content developed specifically for senior mathematics. It supports students with appropriate online resources and tools for every section of the student book, providing access to exemplar worked solutions that demonstrate high levels of mathematical and everyday communication. Students will have the opportunity to learn independently through the Explore further tasks, which have been designed to engage and support conceptual understanding. Additionally, teachers have access to syllabus maps, a teaching program, sample exams, problem-solving and modelling tasks, and additional banks of questions for extra revision.



General Mathematics 12
eBook



General Mathematics 12
Exam preparation workbook

Exam preparation workbook

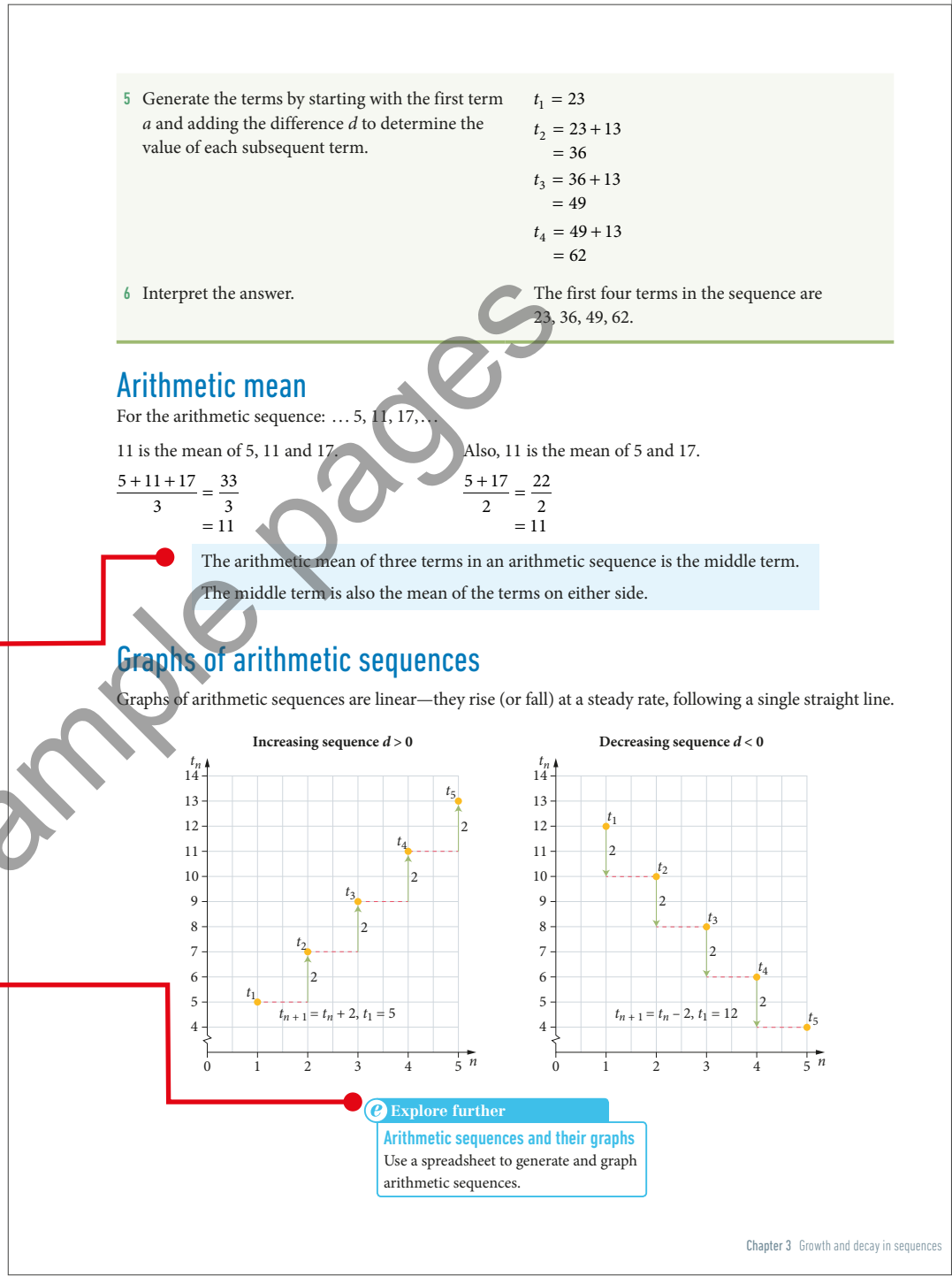
Additional component for Year 12 only

The exam preparation workbook provides additional support in preparing students for the external exam. It has been constructed to guide the students through a sequence of preparatory steps and build confidence leading up to the external exam.

How to use this book

Pearson General Mathematics 12 Queensland Units 3 & 4

This Queensland senior mathematics series has been written by a team of experienced Queensland teachers for the QCE 2019 syllabus. It offers complete curriculum coverage, rich content and comprehensive teacher support.



5 Generate the terms by starting with the first term a and adding the difference d to determine the value of each subsequent term.

$t_1 = 23$
$t_2 = 23 + 13 = 36$
$t_3 = 36 + 13 = 49$
$t_4 = 49 + 13 = 62$

6 Interpret the answer.

The first four terms in the sequence are 23, 36, 49, 62.

Arithmetic mean

For the arithmetic sequence: … 5, 11, 17, …

11 is the mean of 5, 11 and 17.

$\frac{5+11+17}{3} = \frac{33}{3} = 11$

Also, 11 is the mean of 5 and 17.

$\frac{5+17}{2} = \frac{22}{2} = 11$

The arithmetic mean of three terms in an arithmetic sequence is the middle term. The middle term is also the mean of the terms on either side.

Graphs of arithmetic sequences

Graphs of arithmetic sequences are linear—they rise (or fall) at a steady rate, following a single straight line.

Increasing sequence $d > 0$

Decreasing sequence $d < 0$

Explore further

Arithmetic sequences and their graphs

Use a spreadsheet to generate and graph arithmetic sequences.

Chapter 3 Growth and decay in sequences

Key information

Key information and rules are highlighted throughout the chapter.

Explore further

This eBook feature provides an opportunity for students to consolidate their understanding of concepts and ideas with the aid of technology, and answer a small number of questions to deepen their understanding and broaden their skill base. These activities should take approximately 5–15 minutes to complete.

6 Determining the value of the *n*th term

The number of visitors to a park during a fortnight of school holidays begins with 2000 on Monday and increases by 250 each day.

How many people visit the park on the 8th day (the second Monday) of the school holidays?

THINKING	WORKING
1 Recognise the arithmetic sequence and identify the first term a, common difference d and number of terms n.	Arithmetic sequence: $t_1 = 2000,\ d = 250,\ n = 8$
2 Recall the arithmetic term formula. Substitute values for a, d and n to determine the value of the 8th term.	$t_n = t_1 + (n-1)d$ $t_8 = 2000 + (8-1) \times 250$ $= 3750$
3 Interpret the answer.	On the 8th day, 3750 people visit the park.

Starting an arithmetic sequence with $n = 0$

In many situations, instead of beginning with $n = 1$ and 'term 1', t_1, it is appropriate to begin with $n = 0$, and designate an initial 'term 0', as this makes it easier to match term number to time elapsed after the initial term. For example, t_n could be the value after n weeks. In this case, the general term of the arithmetic sequence is given by $t_n = a + nd$.

If the first term of an arithmetic sequence is $t_0 = a$:
$t_n = a + nd$, and there are $n + 1$ terms in the sequence.

Additional information

Scatter plots

Try the activity, to practise identifying relationships between variables.

2 Ella bought a car for $14 000. She paid a cash deposit of $2200 and borrowed the balance at an interest rate of 8.48% p.a. compounding every 6 months. Ella intends to pay off the loan and all added interest in 2 or 3 years.

(a) Which recurrence relation models the value of Ella's loan balance at the end of each half year?

A $E_0 = 11\,800,\ E_{n+1} = 1.0424 \times E_n$
B $E_0 = 14\,000,\ E_{n+1} = 1.0424 \times E_n$
C $E_0 = 11\,800,\ E_{n+1} = 1.0848 \times E_n$
D $E_0 = 14\,000,\ E_{n+1} = 1.0848 \times E_n$

(b) Explain the common error made by a student who had a growth rate of 1.0848 in their solution.

WARNING
If the angular distance is entered using the degrees/minutes/seconds key, the value of D will appear in degrees/minutes/seconds format. To determine the distance in kilometres, convert the value to decimal format at the end.

Every worked example and question is graded

Every example and question is graded using the three levels of difficulty, as specified in the QCE syllabus:

- simple familiar (1 bar)
- complex familiar (2 bars)
- complex unfamiliar (3 bars).

The visibility of this grading helps ensure all levels of difficulty are well covered.

Meeting the needs of the QCE Syllabus

The authors have integrated both the cognitive verbs and the language of the syllabus objectives throughout the worked examples and questions.

Additional information

These interactives appear in the ebook in two forms, as videos explaining specific concepts or as interactive questions to check students' understanding.

Highlighting common errors

Throughout the exercises, authors have integrated questions designed to highlight common errors frequently made by students. Explanations are given in the worked solutions.

Worked solutions

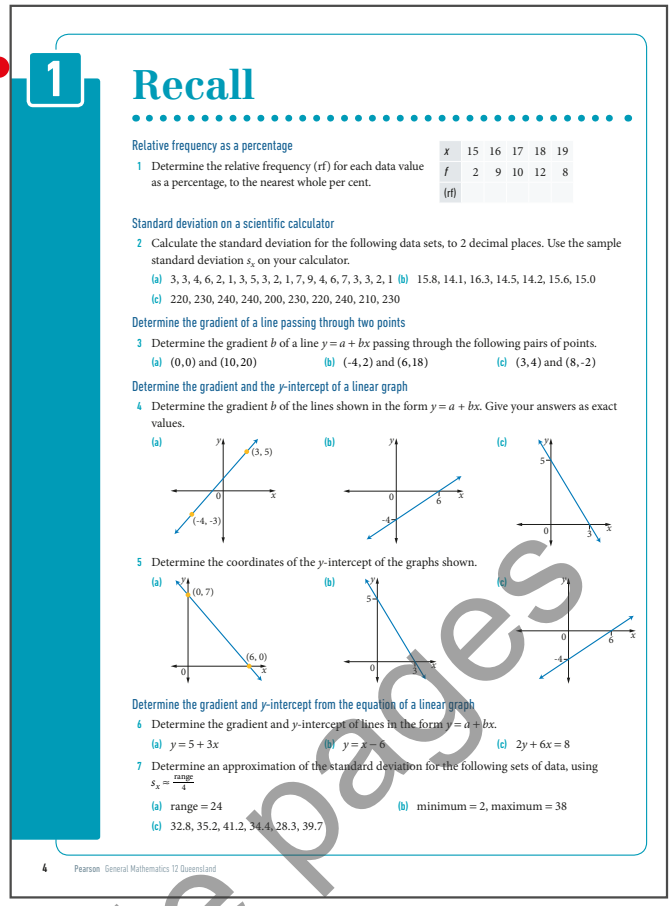
Fully worked solutions are provided for every question in the student textbook and can be accessed from the accompanying eBook.

Warning boxes

Warning boxes are located throughout the chapter to alert students to common errors and misconceptions.

Recall

Each chapter begins with a review of assumed knowledge for the chapter.



1

Recall

Relative frequency as a percentage

1 Determine the relative frequency (rf) for each data value as a percentage, to the nearest whole per cent.

x	15	16	17	18	19
f	2	9	10	12	8
(rf)					

Standard deviation on a scientific calculator

2 Calculate the standard deviation for the following data sets, to 2 decimal places. Use the sample standard deviation s_x on your calculator.

(a) 3, 3, 4, 6, 2, 1, 3, 5, 3, 2, 1, 7, 9, 4, 6, 7, 3, 3, 2, 1 (b) 15.8, 14.1, 16.3, 14.5, 14.2, 15.6, 15.0

(c) 220, 230, 240, 240, 200, 230, 220, 240, 210, 230

Determine the gradient of a line passing through two points

3 Determine the gradient b of a line $y = a + bx$ passing through the following pairs of points.

(a) (0, 0) and (10, 20) (b) (-4, 2) and (6, 18) (c) (3, 4) and (8, -2)

Determine the gradient and *y*-intercept of a linear graph

4 Determine the gradient b of the lines shown in the form $y = a + bx$. Give your answers as exact values.

(a) (3, 5), (-4, -3) (b) (c)

5 Determine the coordinates of the *y*-intercept of the graphs shown.

(a) (0, 7), (6, 0) (b) (c)

Determine the gradient and *y*-intercept from the equation of a linear graph

6 Determine the gradient and *y*-intercept of lines in the form $y = a + bx$.

(a) $y = 5 + 3x$ (b) $y = x - 6$ (c) $2y + 6x = 8$

7 Determine an approximation of the standard deviation for the following sets of data, using $s_x \approx \frac{\text{range}}{4}$

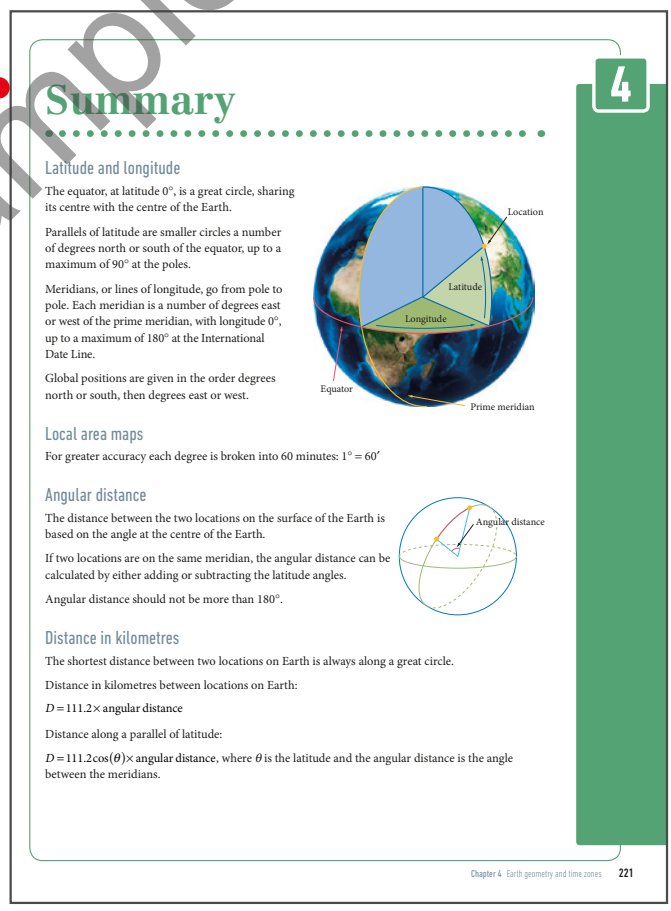
(a) range = 24 (b) minimum = 2, maximum = 38

(c) 32.8, 35.2, 41.2, 34.4, 28.3, 39.7

4 Pearson General Mathematics 12 Queensland

Summary

At the end of each chapter, there is a summary of the key facts and rules discussed in the chapter.



4

Summary

Latitude and longitude

The equator, at latitude 0°, is a great circle, sharing its centre with the centre of the Earth.

Parallels of latitude are smaller circles a number of degrees north or south of the equator, up to a maximum of 90° at the poles.

Meridians, or lines of longitude, go from pole to pole. Each meridian is a number of degrees east or west of the prime meridian, with longitude 0°, up to a maximum of 180° at the International Date Line.

Global positions are given in the order degrees north or south, then degrees east or west.

Local area maps

For greater accuracy each degree is broken into 60 minutes: 1° = 60′

Angular distance

The distance between the two locations on the surface of the Earth is based on the angle at the centre of the Earth.

If two locations are on the same meridian, the angular distance can be calculated by either adding or subtracting the latitude angles.

Angular distance should not be more than 180°.

Distance in kilometres

The shortest distance between two locations on Earth is always along a great circle.

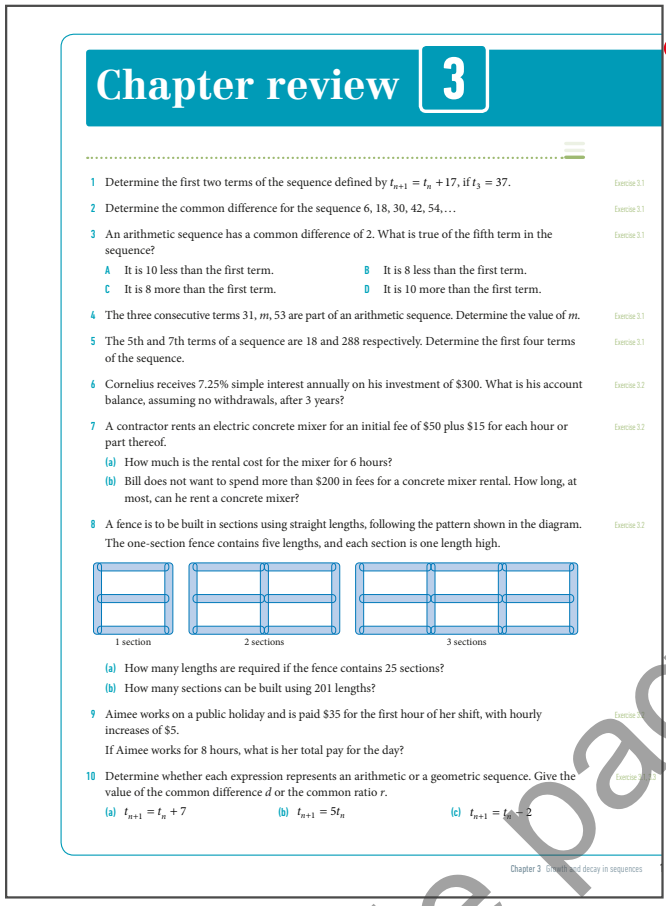
Distance in kilometres between locations on Earth:

$D = 111.2 \times$ angular distance

Distance along a parallel of latitude:

$D = 111.2\cos(\theta) \times$ angular distance, where θ is the latitude and the angular distance is the angle between the meridians.

Chapter 4 Earth geometry and time zones 221



Chapter review 3

1 Determine the first two terms of the sequence defined by $t_{n+1} = t_n + 17$, if $t_3 = 37$. Exercise 3.1

2 Determine the common difference for the sequence 6, 18, 30, 42, 54,… Exercise 3.1

3 An arithmetic sequence has a common difference of 2. What is true of the fifth term in the sequence? Exercise 3.1

A It is 10 less than the first term. B It is 8 less than the first term.
C It is 8 more than the first term. D It is 10 more than the first term.

4 The three consecutive terms 31, m, 53 are part of an arithmetic sequence. Determine the value of m. Exercise 3.1

5 The 5th and 7th terms of a sequence are 18 and 288 respectively. Determine the first four terms of the sequence. Exercise 3.1

6 Cornelius receives 7.25% simple interest annually on his investment of $300. What is his account balance, assuming no withdrawals, after 3 years? Exercise 3.2

7 A contractor rents an electric concrete mixer for an initial fee of $50 plus $15 for each hour or part thereof. Exercise 3.2

(a) How much is the rental cost for the mixer for 6 hours?
(b) Bill does not want to spend more than $200 in fees for a concrete mixer rental. How long, at most, can he rent a concrete mixer?

8 A fence is to be built in sections using straight lengths, following the pattern shown in the diagram. Exercise 3.2
The one-section fence contains five lengths, and each section is one length high.

1 section 2 sections 3 sections

(a) How many lengths are required if the fence contains 25 sections?
(b) How many sections can be built using 201 lengths?

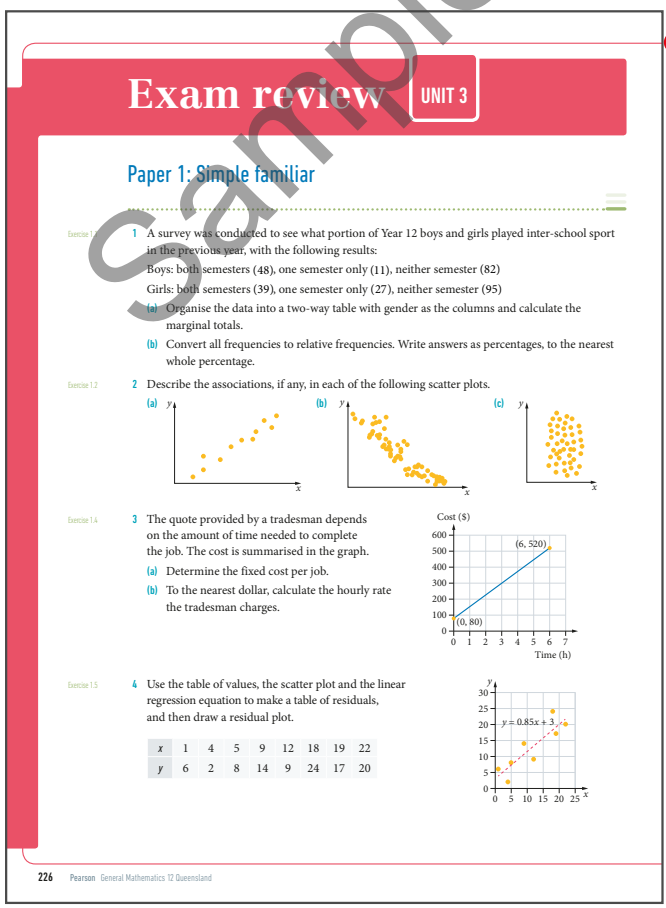
9 Aimee works on a public holiday and is paid $35 for the first hour of her shift, with hourly increases of $5.
If Aimee works for 8 hours, what is her total pay for the day?

10 Determine whether each expression represents an arithmetic or a geometric sequence. Give the value of the common difference d or the common ratio r.
(a) $t_{n+1} = t_n + 7$ (b) $t_{n+1} = 5t_n$ (c) $t_{n+1} = t_n - 2$

Chapter 3 Growth and decay in sequences

Chapter review

Every chapter review follows the QCAA examination proportions for levels of difficulty, which is 60% simple familiar, 20% complex familiar and 20% complex unfamiliar.



Exam review UNIT 3

Paper 1: Simple familiar

Exercise 1.1 1 A survey was conducted to see what portion of Year 12 boys and girls played inter-school sport in the previous year, with the following results:
Boys: both semesters (48), one semester only (11), neither semester (82)
Girls: both semesters (39), one semester only (27), neither semester (95)
(a) Organise the data into a two-way table with gender as the columns and calculate the marginal totals.
(b) Convert all frequencies to relative frequencies. Write answers as percentages, to the nearest whole percentage.

Exercise 1.2 2 Describe the associations, if any, in each of the following scatter plots.
(a) (b) (c)

Exercise 1.4 3 The quote provided by a tradesman depends on the amount of time needed to complete the job. The cost is summarised in the graph.
(a) Determine the fixed cost per job.
(b) To the nearest dollar, calculate the hourly rate the tradesman charges.

Cost ($) (6, 520) (0, 80) Time (h)

Exercise 1.5 4 Use the table of values, the scatter plot and the linear regression equation to make a table of residuals, and then draw a residual plot.

x	1	4	5	9	12	18	19	22
y	6	2	8	14	9	24	17	20

$y = 0.85x + 3$

226 Pearson General Mathematics 12 Queensland

Exam review

Exam reviews give cumulative practice of content covered, to prepare students for end-of-year exams. They have been strategically placed between the chapters.

1

Bivariate data analysis

1 Recall

Relative frequency as a percentage

1 Determine the relative frequency (rf) for each data value as a percentage, to the nearest whole per cent.

x	15	16	17	18	19
f	2	9	10	12	8
(rf)					

Standard deviation on a scientific calculator

2 Calculate the standard deviation for the following data sets, to 2 decimal places. Use the sample standard deviation s_x on your calculator.

(a) 3, 3, 4, 6, 2, 1, 3, 5, 3, 2, 1, 7, 9, 4, 6, 7, 3, 3, 2, 1 (b) 15.8, 14.1, 16.3, 14.5, 14.2, 15.6, 15.0

(c) 220, 230, 240, 240, 200, 230, 220, 240, 210, 230

Determine the gradient of a line passing through two points

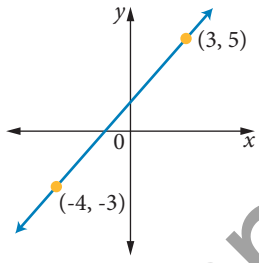
3 Determine the gradient b of a line $y = a + bx$ passing through the following pairs of points.

(a) (0,0) and (10,20) (b) (-4,2) and (6,18) (c) (3,4) and (8,-2)

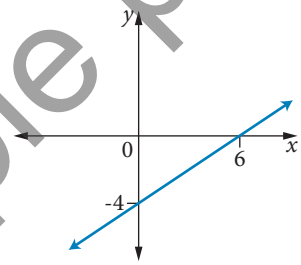
Determine the gradient and the y-intercept of a linear graph

4 Determine the gradient b of the lines shown in the form $y = a + bx$. Give your answers as exact values.

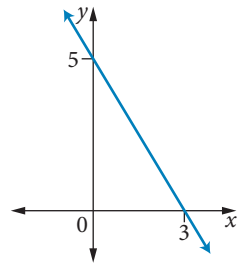
(a)



(b)

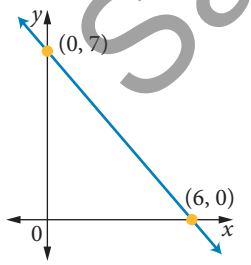


(c)

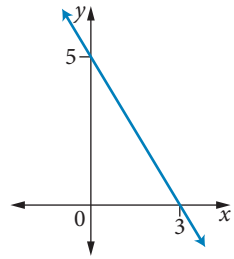


5 Determine the coordinates of the y-intercept of the graphs shown.

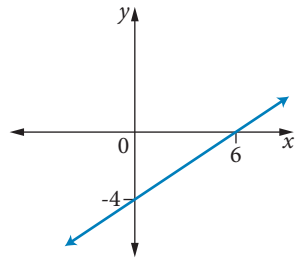
(a)



(b)



(c)



Determine the gradient and y-intercept from the equation of a linear graph

6 Determine the gradient and y-intercept of lines in the form $y = a + bx$.

(a) $y = 5 + 3x$ (b) $y = x - 6$ (c) $2y + 6x = 8$

7 Determine an approximation of the standard deviation for the following sets of data, using $s_x \approx \frac{\text{range}}{4}$

(a) range = 24 (b) minimum = 2, maximum = 38

(c) 32.8, 35.2, 41.2, 34.4, 28.3, 39.7

Association between categorical variables 1.1

Bivariate data

Bivariate data is data that comes in pairs. Analysis of bivariate data consists of collecting and pairing data on two variables and then using the data in comparative ways to determine whether a relationship exists between the variables. If a relationship does exist, then the type and strength of the relationship are determined.

Displaying categorical data

For categorical data to be bivariate, the data must be categorised in two ways. A two-way table is a convenient way to display the data. Totals for each row and column, sometimes called *marginal totals*, enable you to calculate the percentage, or proportion, of the whole represented in each subcategory.

1 Constructing a two-way table

A survey is conducted in which young people are asked to choose their favourite film genre from a given list. Each participant's gender is recorded with their response. The results are given below.

Males: Horror (5), Comedy (16), Action (22), Science fiction (16), Romantic comedy (2), Animation (0), Thriller (18)

Females: Horror (18), Comedy (20), Action (4), Science fiction (16), Romantic comedy (48), Animation (30), Thriller (32)

(a) Organise the data into a two-way table with gender as the columns.

THINKING

The categories for one variable go along the top. The categories for the other variable go in a column at the left.

Fill in the numbers.

WORKING

	Males	Females
Horror	5	18
Comedy	16	20
Action	22	4
Science fiction	16	16
Romantic comedy	2	48
Animation	0	30
Thriller	18	32

(b) Calculate the marginal totals.

1 Sum each row and column.

Write in the grand total, by adding either the row totals or the column totals.

	Males	Females	Total
Horror	5	18	23
Comedy	16	20	36
Action	22	4	26
Science fiction	16	16	32
Romantic comedy	2	48	50
Animation	0	30	30
Thriller	18	32	50
Total	79	168	247

2 Interpret the result.

The survey was completed by 247 individuals.

(c) Convert all frequencies to relative frequencies. Write answers as percentages to the nearest whole per cent.

1 Divide each value in the table by the total number of participants surveyed. To convert to a percentage, multiply by 100 and attach the percentage symbol. Round answers as directed.

Replace the numbers with the percentages.

As an example: $\frac{5}{247} \times 100\% = 2\% \text{ (nearest \%)}$

	Males	Females	Total
Horror	2%	7%	9%
Comedy	6%	8%	~15%
Action	9%	2%	11%
Science fiction	6%	6%	~13%
Romantic comedy	1%	19%	20%
Animation	0%	12%	12%
Thriller	7%	13%	20%
Total	~32%	~68%	100%

2 Interpret the results.

The table shows the percentage of total participants in each category. For example, 9% of respondents prefer horror movies, 15% prefer comedy etc.

32% of respondents are male and 68% are female.

Explore further

Two-way tables

Create a two-way table and then calculate and interpret values using a spreadsheet.

WARNING

The sum of the rows and columns within a percentage frequency table may not always give an accurate reading, due to rounding values throughout the table.

Percentaged two-way tables

The unequal number of males and females in the survey (the total number of entries in each male and female column) makes it difficult to see any difference or similarity that may exist in the answers for the two genders.

If each column is converted to percentages for just that column, so that the total of each column is 100%, a more meaningful analysis of association can be performed.

The differences between the percentages are more clearly seen here. You can see the effect of the variable – gender – on the relative frequencies for each of the movie genre choices. Male participants preferred 'Comedy', 'Action', 'Science fiction' and 'Thriller'. Females predominantly chose 'Romantic comedy', 'Animation' and 'Thriller' films.

	Males (%)	Females (%)
Horror	$\frac{5}{79} \times 100\% \approx 6$	$\frac{18}{168} \times 100\% \approx 11$
Comedy	20	12
Action	28	2
Science fiction	20	9
Romantic comedy	3	29
Animation	0	18
Thriller	23	19
Total	100	100

2 Using a spreadsheet to create a percentaged two-way table

The favourite superheroes of a group of 70 people are shown in the table.

	Spiderman	Batman	Green Lantern	The Hulk	Captain America	Thor	Iron Man
Males	1	7	2	4	2	8	15
Females	6	14	0	3	1	2	5

Using a spreadsheet, produce an appropriately percentaged two-way table so that an analysis of the effect of gender on the choice of superhero can be carried out.

THINKING

1 Hypothesise to determine which variable you should sum.

2 Enter the data with more rows than columns, for convenience.

Calculate the total number of each column. For example, in cell B9 insert the formula =SUM(B2:B8).

WORKING

The choice of superhero could be influenced by gender. Determine the total for each gender.

	A	B	C
1		Males	Females
2	Spiderman	1	6
3	Batman	7	14
4	Green Lantern	2	0
5	The Hulk	4	3
6	Captain America	2	1
7	Thor	8	2
8	Iron Man	15	5
9	Total	39	31

3 In columns D and E, create percentages, to the nearest whole percentage, by dividing each frequency value by the total for that category – here, males and females.

For the spreadsheet method, for the Spiderman, Males (%) cell, insert the formula =B2/B$9. The $ symbol ensures that each value is divided by the same total, B9. Then drag the cross-hair to fill the column.

In cell E2 insert the formula =C2/C$9.

	A	D	E
1		Males (%)	Females (%)
2	Spiderman	3	19
3	Batman	18	45
4	Green Lantern	5	0
5	The Hulk	10	10
6	Captain America	5	3
7	Thor	21	6
8	Iron Man	38	16
9	Total	100	100

4 Interpret the results.

Of the two most popular superheroes, Batman (21 people) and Iron Man (20), 38% of male respondents chose Iron Man as their favourite superhero, while 45% of female respondents chose Batman as a clear favourite superhero.

Describing an association

In any analysis of association:

STEP 1: Decide on the variable that may have influence over the other variable.

STEP 2: Calculate percentages of the variable with influence.

STEP 3: Describe the association in terms of differences in the percentages across the categories.

STEP 4: Interpret the observations in the context of the data.

3 Describing an association from a percentaged two-way table

The following data represents the results obtained on a particular test (marked out of 40), taken by two groups: Group A (students 16 years of age) and Group B (students 18 years of age). The pass mark for the test is 25.

Group A: 37, 26, 31, 23, 34, 38, 29, 17, 33, 26, 23, 21, 29, 30, 33

Group B: 24, 28, 29, 34, 18, 19, 32, 29, 37, 28

Determine whether there is an association between age and the likelihood of passing the test.

THINKING

1 Create a two-way table using the categories described.

Count the number of figures in each category and place the frequency in the relevant column.

Total the frequencies for each of the categories that might influence the outcome.

WORKING

	Pass (≥25 marks)	Fail (<25 marks)	Total
Group A	13	2	15
Group B	7	3	10

2 Convert each frequency to a percentage of the total in each row.

Redraw the table to show the percentages.

	Pass	Fail	Total
Group A	$\frac{13}{15} \times 100\% \approx 87\%$	$\frac{2}{15} \times 100\% \approx 13\%$	100%
Group B	$\frac{7}{10} \times 100\% = 70\%$	$\frac{3}{10} \times 100\% = 30\%$	100%

3 Describe any associations you can identify from the data, giving values and context.

87% of 16-year-olds passed, compared with 70% of 18-year-olds.

It is important to note that this does not lead to the conclusion that younger people are more likely to pass the test. There are likely to be a variety of factors that influence a student's chance of success, including class size, preparation time, test conditions and teacher input.

EXERCISE 1.1 Association between categorical variables

Worked Example 1

1 An excursion is arranged for students in Year 12 and Year 7 to see a movie together. The students have a choice of popcorn, fries or a choc top to take into the movie with them. The teachers keep a record of the orders, for a later comparison of movie snack choices at different ages:

Year 12: popcorn (35), fries (21), choc tops (11)

Year 7: popcorn (40), fries (26), choc tops (50)

(a) Organise the data into a two-way table and calculate the marginal totals.

(b) Convert all frequencies to relative frequencies. Write answers as percentages, to the nearest whole per cent.

2 The manager of a cafe offers a choice of chips, vegetables or salad with each meal. She asks her staff to keep a record of the choices made by different age groups. Here are their findings:

Children: chips (20), vegetables (2), salad (5)

Teenagers: chips (31), vegetables (1) , salad (20)

Adults: chips (10), vegetables (28), salad (63)

(a) Organise the data into a two-way table and calculate the marginal totals.

(b) Calculate each of the following, to the nearest per cent.

(i) the percentage of customers who ordered chips

(ii) the percentage of teenage customers

(iii) the percentage of teenagers who ordered salad

(iv) the difference between the percentage of children who ordered chips and the percentage of adults who ordered chips.

3 For which of the following sets of survey data would you expect very little (or no) association?

A Year 12 and Year 8 students, and amounts of study (small, moderate, large)

B mode of travel to the movies, and age group

C preference for video games or going to the movies, and gender

D ability in Drama, and preference for blue or red

Worked Example 2

4 A group of males and females were asked: 'Which of the following movie villains do you like the most?' Their responses are summarised in the table.

	Lex Luthor	Goldfinger	The Penguin	The Joker	KAOS
Males	46	10	80	40	12
Females	84	22	60	55	28

(a) Produce an appropriately percentaged two-way table.

(b) Which villain is most popular with each gender?

5 A class of Year 12 students spent their Friday Maths lesson breaking open coloured choc-coated peanuts and sultanas to determine whether the outside colour had some influence on whether the chocolate contained a peanut or a sultana. Their findings are shown in the table.

	Blue	Brown	Yellow	White
Peanut	20	24	15	27
Sultana	14	28	26	26

(a) Produce an appropriately percentaged two-way table. Round your answers to the nearest per cent each time.

(b) Which colour was most likely to contain a peanut?

6 An ice-cream vendor wanted to know whether people who chose to eat their ice-cream from a cone were more likely to choose vanilla than those ordering ice-cream in a cup. His findings are shown.

	Vanilla	Other flavour
Cone	165	206
Cup	43	75

(a) Produce an appropriately percentaged two-way table. Round your answers to the nearest per cent each time.

(b) Determine the percentage difference for choosing vanilla in a cone compared to a cup.

(c) The ice-cream vendor wishes to explore proportions among vanilla ice-cream eaters. Determine the proportions of vanilla ice-cream customers eating from cones and from cups.

7 For the data in the tables, assume that categories X and Y are expected to affect the frequencies within the categories A, B and C.

I

	A	B	C
X	50%	30%	20%
Y	40%	40%	20%

II

	A	B	C
X	80%	60%	30%
Y	20%	40%	70%

III

	A	B	C
X	2	15	20
Y	18	15	0

IV

	A	B	C
X	16	10	14
Y	12	19	9

(a) For which of the data can association be explored without further calculation?

A I only B II only C I and II D I and IV

(b) Explain the common error made by a student who chose an incorrect option in part (a).

8 The following data about travel time for Year 7 and Year 12 students was collected, to determine whether there is an association between the year level of a student and the time taken to travel to school.

Travel time (minutes)	Number of Year 7 students	Number of Year 12 students
0–<10	25	2
10–<20	40	8
20–<30	24	25
30–<40	12	55
40–<50	4	24
50–<60	2	6

(a) Convert the table to a percentage-based two-way frequency diagram using the travel time categories < 30 minutes and ≥ 30 minutes. Give percentages to the nearest whole per cent.

(b) Use percentage difference to describe the association between year level and travel time to school.

9 The table shows the number of people of various ages with high blood pressure (greater than 140/90 mmHg). The sample size for each age group is 100 males and 100 females.

Age group	Males	Females
18–24	7	5
25–34	13	4
35–44	19	11
45–54	29	22
55–64	33	27
65–74	37	41
>74	42	52

Source ABS: Australian Health Survey First results

(a) In which age group does the percentage of women with high blood pressure become higher than the percentage of men with high blood pressure?

(b) Is it reasonable to say that 12% of people aged 18–24 years old have high blood pressure?

10 Vitamin C is thought by some people to help ward off colds, or at least shorten the length of a cold. In order to test this theory, a Year 12 Biology class conducted a test with 250 Year 11 and 12 students at their school. A controlled experiment was conducted by giving half the students vitamin C tablets to take for the winter and the other half, a placebo, sugar tablets that looked the same as the vitamin C tablets. Students were not told which type of tablet they had been given.

The results of the study are shown in the two-way table.

	Cold	No cold
Sugar tablet	29	96
Vitamin C tablet	24	101

(a) Convert this table to an appropriate percentaged table for investigating whether there is an association between taking vitamin C and the likelihood of getting a cold. Give answers to the nearest whole per cent.

(b) Express any association between taking vitamin C and the likelihood of getting a cold in terms of percentage difference.

11 The table shows the results of a survey of 200 shoppers about a proposal to change the arrangements for childcare at a shopping centre.

	Males	Females	Total
Agree	25	88	113
Disagree	45	42	87
Total	70	130	200

(a) Use an appropriately percentaged table to determine whether a person's gender is associated with their likelihood of agreeing with the proposal. Round percentages to the nearest whole per cent.

(b) Express any association between gender and agreement with the proposal in terms of percentage difference.

12 The data shown in the table represents a group of men and women surveyed and categorised as overweight to obese, or normal to underweight. They were then put into the following subgroups: diabetic, pre-diabetic or within the healthy range (in relation to diabetes).

Data was collected for 800 males and females aged between 24 and 34.

	Overweight to obese		Normal to underweight		Total
	Males	Females	Males	Females	
Diabetic	14	8	20	10	52
Pre-diabetic	9	7	13	5	34
Healthy range	77	105	267	265	714
Total	100	120	300	280	800

(a) Construct a two-way frequency table for the 'Overweight to obese' data by adding a column showing the percentage for males, females and total. Give your answers to the nearest per cent.

(b) Construct a two-way frequency table for the 'Normal to underweight' data by adding a column showing the percentage for males, females and total. Give your answers to the nearest per cent.

(c) What association can you see between the risk of diabetes for the two groups of people?

13 A group of 100 tennis players were observed serving a tennis ball. The table shows the number of unreturnable serves (aces) per set for tennis players with a fast serve, compared to those with a slow serve.

	0 aces	1 ace	2 aces	3 aces	4 aces	Total
Slow serve	7		8	4	0	27
Fast serve	1	8				
Total			22	34		100

(a) Complete the table by using the data provided to determine the missing values.

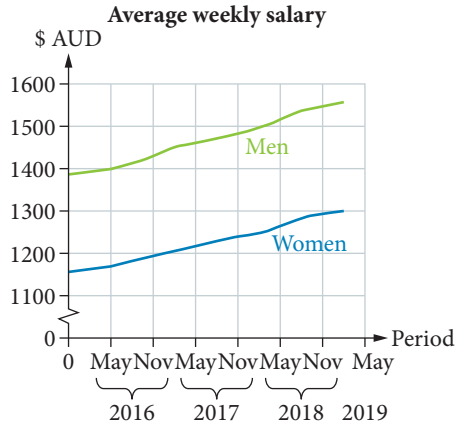
(b) How many players served three aces?

(c) Calculate the percentage of fast and slow servers who served more than two aces. Give your answers to the nearest per cent.

(d) Describe the association observed between the speed of a serve and the number of aces.

14 Information about the average weekly salaries ($AUD) of men and women measured every six months from May 2016 to May 2019 is shown in the graph.

Describe the association between the average weekly salaries of men and women in Australia. Support your conclusions with values, stating the strengths and limitations of your result.



15 The lengths of rivers longer than 100 km in the North and South Islands of New Zealand are listed below.

North Island (km): 172, 290, 425, 193, 105, 175, 241, 161, 158, 154, 182, 137, 143, 132, 119, 137

South Island (km): 209, 169, 288, 209, 138, 121, 177, 161, 322, 145, 121, 108

The approximate areas of the North and South Islands are 114 000 km^2 and 150 000 km^2 respectively.

Investigate whether there is an association between the size of the island and the number of rivers of a particular length.

(a) Organise the data and write your observations for rivers longer than 100 km, 150 km and 200 km.

(b) What do you conclude about there being an association between the area of the island and the number of rivers of particular lengths?

Association between numerical variables 1.2

Comparing numerical data sets

In this section you will examine how to compare two numerical data sets.

It is common for two data sets to have an association where one variable may influence the other. For example, if you were to measure a random group of people's heights and weights you would probably find a positive association between height and weight. The assumed association would be that taller people are more likely to be heavier (as opposed to heavier people being more likely to be tall), because a person's height is independent of how much they weigh.

Scatter plots

A **scatter plot** provides a visual representation of any trend or underlying pattern in the data. A scatter plot is drawn by treating the bivariate data as a series of coordinate pairs and plotting the pairs on a suitable set of axes.

The data used often comes from measurements in real-life situations, so the graph is quite often limited to the positive part of each scale or the first quadrant on the Cartesian plane.

To construct a scatter plot:

- Draw a suitable set of axes with a consistent scale {0, 1, 2, 3…} by noting the highest and lowest value from each set of figures.
- Plot each data pair as a coordinate point.

WARNING

For a scatter plot:

- Do not start at zero on any axis where the values are clustered away from zero.
- Do not join the plotted points.

Explore further

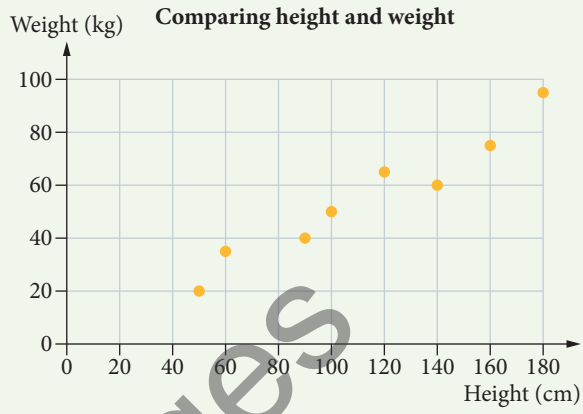
Scatter plots and the line of best fit

Use a spreadsheet to construct a scatter plot, and insert a line of best fit.

4 Constructing a scatter plot

Construct a scatter plot of the following bivariate data set and describe any trend you see.

Person	A	B	C	D	E	F	G	H
Height (cm)	140	160	180	90	100	50	60	120
Weight (kg)	60	75	95	40	50	20	35	65

THINKING	WORKING
1 Draw a suitable set of axes by noting the range of each set of figures. Plot the points.	In this case, height (cm) is from 50 to 180, so 9 divisions of 20 starting at zero is appropriate. Weight (kg) varies from 20 to 95, so 5 divisions of 20 starting at zero is appropriate. 
2 Comment on any pattern or trend observed.	From the graph, it appears that the taller the person is, the heavier they are.

Explanatory and response variables

Often one of the variables can explain the association, and this is known as the independent or *explanatory variable*. The other variable responds to a change in the explanatory variable and is known as the dependent or *response variable*. In the case of height versus weight, a person's weight is more likely explained by their height than the other way around.

The **explanatory (or independent) variable** is placed on the horizontal axis.

The **response (or dependent) variable** is placed on the vertical axis.

When a saucepan of water is placed on the stove to boil, the water's temperature responds to, or depends upon, the time spent on the hot plate.

When an ice-cream vendor sets up a booth near the beach, the number of ice creams sold will likely respond to the temperature of the day.

Sometimes two variables may both be dependent on other variables. For example, a student's study score in English and Music are related but both depend on other variables such as effort, hours of practice, teacher input and study.

Sometimes variables may have no association at all. For example, a student's study score in English is not related to the length of their hair.

WARNING

Not all bivariate data sets have clear explanatory and response variables.

5 Identifying the explanatory variable

Assuming an association exists, identify the explanatory variable between each of the following pairs:

(i) age and wealth

(ii) age and the number of offspring

(iii) temperature of the day and the number of people at the beach

(iv) number of cigarettes smoked and the chance of getting cancer

(v) volume of petrol remaining in a car's tank and the distance the car has been driven.

THINKING

Determine which one of the variables could cause or explain a change in the other.

WORKING

(i) Age and wealth: The explanatory variable is age. In general, a longer working life is likely to explain an increase in wealth.

(ii) Age and the number of offspring: The explanatory variable is age. It takes time to reproduce, so a longer life span is likely to explain an increase in the number of offspring produced by a female.

(iii) Temperature of the day and the number of people at the beach: The explanatory variable is temperature. In general, an increase in temperature is likely to explain an increasing beach population.

(iv) Number of cigarettes smoked and the chance of getting cancer: The explanatory variable is number of cigarettes smoked. In general, a greater number of cigarettes smoked is likely to explain a higher likelihood of cancer.

(v) Volume of petrol remaining in a car's tank and the distance the car has been driven: The explanatory variable is the distance driven. An increasing distance driven explains a decreasing volume of petrol in the tank.

Additional information

Scatter plots

Try the activity, to practise identifying relationships between variables.

Linear trend

A single line that best represents the general pattern of the data is called a *line of best fit*.

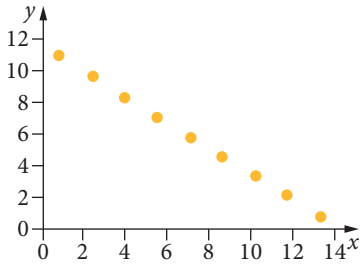
If the pattern is roughly a straight line, a linear association exists.

The trend indicates the general direction of the association. The pattern may show a positive, increasing trend (increasing as you move to the right ↗), meaning that as one variable increases the other also increases. Or it may show a negative, decreasing trend (decreasing as you move to the right ↘), meaning that as one variable increases the other decreases.

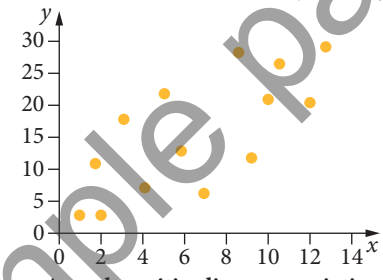
When analysing the scatter plot, look for the following characteristics:

- Is there is an observable pattern?
- Does the pattern show a linear form or non-linear form?
- Give the direction of the slope as positive or negative.
- Identify any outliers – that is, single points that seem to be well outside the general pattern of the rest of the data. If outliers are excluded from the data, the pattern will be easier to see.
- Consider the strength of the association, how the points are spread around and represent the data. A good fit represents a strong trend; a poor fit represents a weak trend.

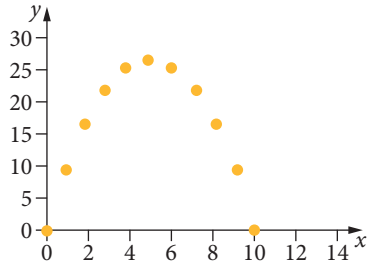
Examples are as follows:



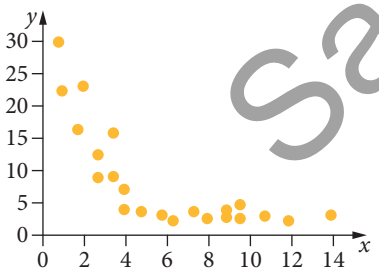
A strong negative linear association



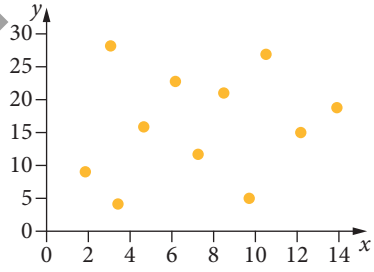
A weak positive linear association



A strong non-linear association



A moderate negative non-linear association

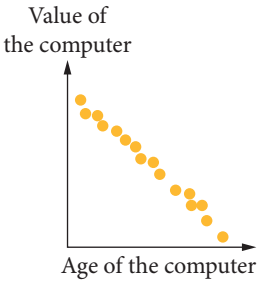


No association

6 Describing associations

Describe the associations, if any, between the variables represented in the following scatter plots.

(a)



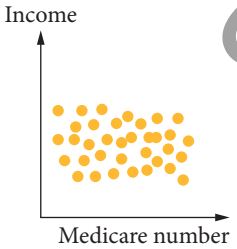
Additional information

Identifying correlation

Try the activity, to practise determining whether scatter plots show a positive or negative correlation.

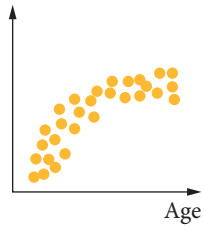
THINKING	WORKING
1 Identify whether a pattern exists. If there is a pattern, describe the strength of the association as strong, moderate or weak.	There is a pattern: the plotted points form a distinct line, so there is a strong association between the two variables.
2 Check for linearity.	The plotted points form a straight line, so the association is linear.
3 Determine the direction of the slope to be positive or negative.	The slope is negative, that is, a decreasing trend; as the explanatory variable increases, the response variable decreases.
4 Interpret the slope.	As a computer ages, the value of the computer decreases.
5 Describe the association.	The association between the age of the computer and the value of the computer is strong, negative and linear.

(b)



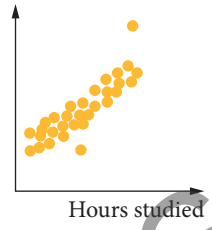
1 Identify whether a pattern exists.	No pattern is evident in the data, suggesting that no relationship exists between the two variables.
2 Describe the association.	There is no apparent connection between income and a person's Medicare number.

(c) Income



1 Identify whether a pattern exists. If there is a pattern, describe the strength of the association as strong, moderate or weak.	There is a pattern: the data shows a moderate association between the two variables.
2 Check for linearity.	The plotted points do not form a straight line, so the association is non-linear.
3 Determine the direction of the slope to be positive or negative.	The slope is positive and flattens out, an increasing trend. As the explanatory variable increases, the response variable also increases.
4 Interpret the slope.	As people age, personal income increases, plateauing out towards the end of the data.
5 Describe the association.	The association between age and income is moderate, positive and non-linear.

(d) Exam score



1 Identify whether a pattern exists. If there is a pattern, describe the strength of the association as strong, moderate or weak.	There is a pattern: there appears to be a moderate association between the two variables.
2 Check for linearity.	The plotted points form a reasonably straight line, so the association is linear.
3 Determine the direction of the slope to be positive or negative.	The slope is positive, an increasing trend: as the explanatory variable increases, the response variable also increases.
4 Interpret the slope.	As the time spent studying increases, the exam score achieved also increases.
5 Describe the association (noting the presence of outliers).	The association between the time spent studying and the exam score is moderate, positive and linear. There are two possible outliers in the data.

More than two variables

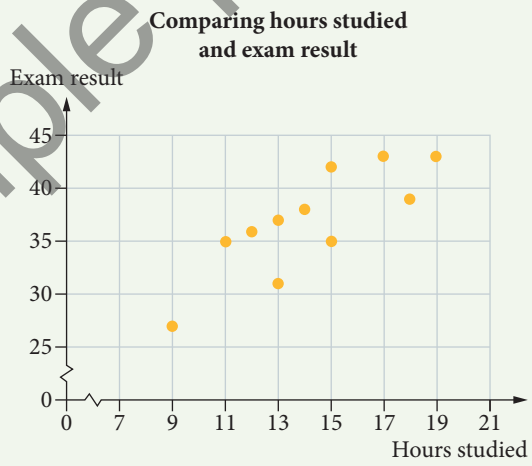
When more than two sets of numerical data are given for a group of individuals, you can use bivariate data analysis to explore the association between any pair of variables.

7 Bivariate data analysis when there are three variables

The data set shows the class marks for a Mathematics test, broken into three aspects: exam result out of 50 (ER), average number of hours the student works in a part-time job (PJ) and average time spent studying per week (S).

S	18	11	15	13	19	14	17	12	15	13	9
PJ	14	15	12	10	6	4	8	14	4	16	18
ER	39	35	35	37	43	38	43	36	42	31	27

(a) Describe the association between the exam result and time spent studying.

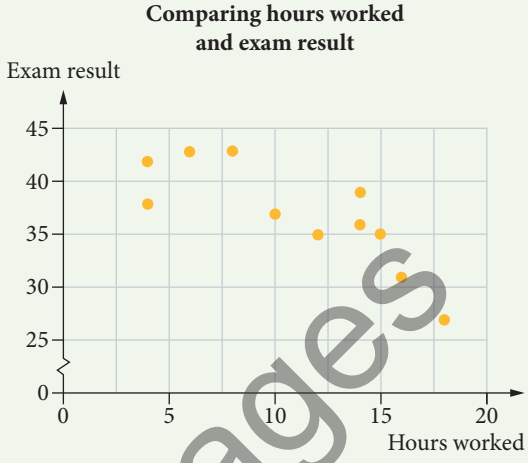
THINKING	WORKING
1 Consider the possible relationship between the exam result and time spent studying.	A positive association may be present, as it is likely that the more a student studies, the better their exam result will be. Exam preparation also relies on other factors, and effective study techniques are very important.
2 Construct a scatter plot with the explanatory variable on the horizontal axis and the response variable on the vertical axis.	
3 Describe the trend in terms of strength, direction and form.	There is a moderate positive linear relationship between the number of hours studied and the exam result achieved. The increasing trend indicates that, as the number of hours of study increases, the exam result also increases.

(b) Describe the association between the exam result and time spent working in a part-time job.

1 Consider the possible relationship between the exam result and time spent working in a part-time job.

A negative association may be present, as it is likely that if students take on too much work they will have less time to study and sleep, which are important aspects in learning and preparing well for assessment such as exams.

2 Construct a scatter plot with the explanatory variable on the horizontal axis and the response variable on the vertical axis.



3 Describe the trend in terms of strength, direction and form.

There is a moderate negative non-linear relationship between the number of hours worked and the exam result achieved.

The decreasing trend indicates that, as the number of hours of part-time work increases beyond approximately 10 hours, the exam result obtained decreases.

(c) Describe the association between the time spent working in a part-time job and the time spent studying.

1 Consider the possible relationship between time spent studying and time spent working in a part-time job.

A negative association may be present, as there are only a set number of hours in a week, so the more time spent working, the less time there is to study. However, there are many reasons for reduced study time including family, sporting or other extra-curricular activities, interest in the subject and the determination of the student. High working hours does not cause low study hours.

2 Construct a scatter plot with the explanatory variable on the horizontal axis and the response variable on the vertical axis.

3 Describe the trend in terms of strength, direction and form.

There is a weak negative non-linear relationship. The negative trend indicates that, as the number of hours of part-time work increases beyond approximately 10 hours, the time spent studying decreases.

EXERCISE 1.2 Association between numerical variables

Worked Example 4

1 Five friends compete with each other to do the best on their upcoming General Maths Examination. In the final month before the exam they record how much time they spend studying. The final results are given in the table.

Name	Alex	Bel	Cam	Darmi	Echo
Hours of study	30	15	5	10	35
General Maths result (%)	70	45	60	80	95

Construct a scatter plot and describe any pattern you see in terms of strength, direction and form.

5

2 Assuming an association exists, choose the explanatory variable in each of the following pairs.

(a) Years of employment; Value of superannuation

(b) Rainfall; Size of plants

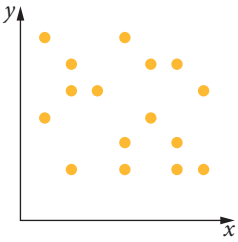
(c) Temperature; Number of ice-creams sold

(d) Waist measurement; Cans of soft drink consumed

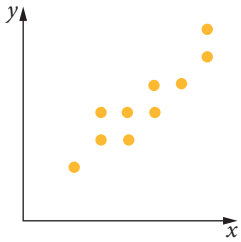
6

3 Describe the associations, if any, in the following scatter plots.

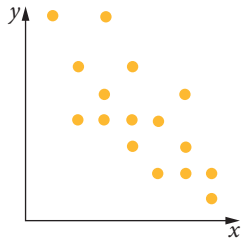
(a)



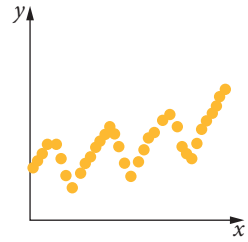
(b)



(c)



(d)



4 Consider the scatter plot comparing arm strength and grip strength.

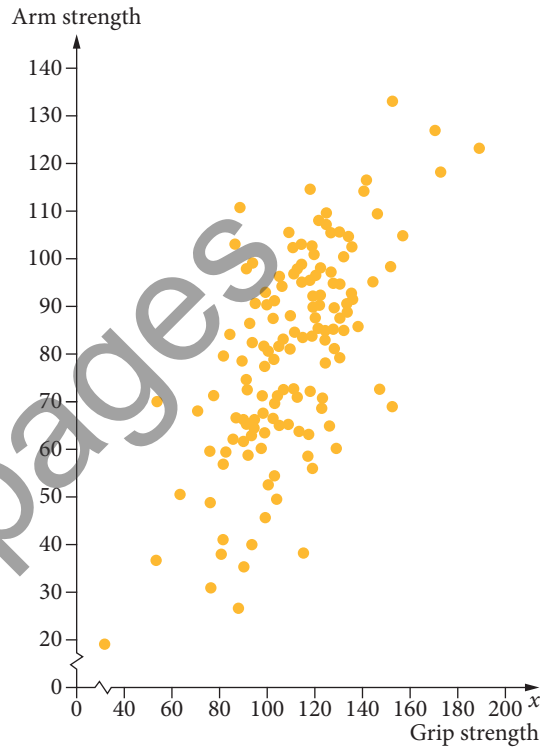
Which pattern describes the scatter plot?

A strong, positive, linear

B moderate, negative, linear

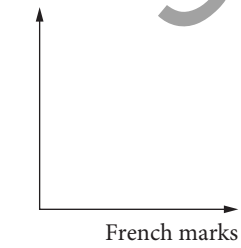
C no association

D moderate, positive, non-linear

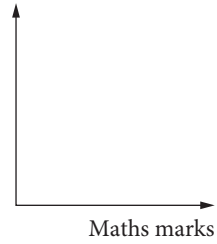


5 A student has started to draw four scatter plots.

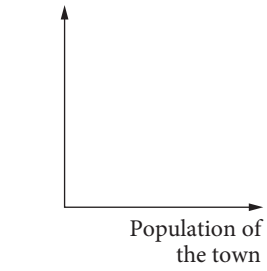
Scatter plot 1



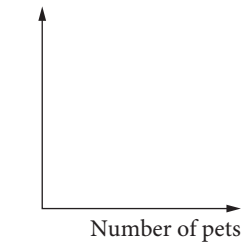
Scatter plot 2



Scatter plot 3



Scatter plot 4



(a) On which scatter plot has the student placed the axis titles incorrectly?

A Scatter plot 1 **B** Scatter plot 2 **C** Scatter plot 3 **D** Scatter plot 4

(b) Explain the common error made by a student who chose any of the incorrect options listed in **(a)**.

6 The following data gives the latitude of some Australian cities, along with the average maximum temperature for June for each city.

	Cairns	Townsville	Mackay	Rockhampton	Brisbane	Canberra
Latitude (°S)	16.9	19.3	21.1	23.4	27.5	35.3
Temperature (°C)	26	26	22	24	21	13

	Toowoomba	Sydney	Melbourne	Hobart	Perth	Adelaide
Latitude (°S)	27.6	33.9	37.8	42.9	32.0	34.9
Temperature (°C)	17	18	15	12	19	16

(a) Determine the explanatory variable.

(b) Describe the kind of linear association you would expect to find between the variables, if any.

(c) Graph the scatter plot and describe any association.

7 The following data gives the 1-year-old value and the 5-year-old value for a variety of used cars.

	Mazda 3	Mazda 6	Ford Focus	Subaru Outback	Toyota Camry	Toyota Prado	Kia Cerato
1-year-old value ($'000)	22.0	32.9	19.6	42.4	17.9	51.3	17.5
5-year-old value ($'000)	8.9	16.5	13.9	20.0	12.8	46.0	13.0

(a) Determine the explanatory variable.

(b) Describe the association you would expect to find between the variables, if any.

(c) Graph the scatter plot and describe any association that is revealed.

8 The following data from 2017 gives median house prices for Australian capital cities, along with the population of each city.

	Melbourne	Sydney	Brisbane	Perth	Hobart	Adelaide	Canberra	Darwin
Median house price ($'000)	880	1150	550	550	410	520	720	590
Population ($'000)	4500	5070	2300	2250	220	1700	360	150

(a) Determine the explanatory variable.

(b) Describe the kind of linear association you would expect to find between the variables, if any.

(c) Graph the scatter plot and describe the relationship.

Worked Example

7

9 The following data set shows the final percentage results for 10 students in the three subjects they have in common.

In each case put the first-named variable on the horizontal axis.

Maths	40	60	75	60	80	30	45	75	90	55
English	50	70	70	80	90	50	40	65	75	60
Drama	60	80	50	90	75	40	55	80	80	35

(a) Draw a scatter plot to show any potential association between the students' Maths and English scores.

(b) Draw a scatter plot to show any potential association between the students' Maths and Drama scores.

(c) Draw a scatter plot to show any potential association between the students' Drama and English scores.

(d) Which association appears to be the strongest?

10 The data shown in the table gives the value, age and kilometres travelled, for similar models of Volkswagen Golfs.

Value ($'000)	15.5	18.0	17.0	17.0	20.4	17.5	25.0	21.5	15.6	9.5	14.0	22.5	19.0
Age (years)	7	2	2	3	3	7	5	4	5	10	3	7	3
Km travelled ('000)	63	29	50	53	56	129	54	50	53	184	92	72	35

(a) Which of the three variables would be the response variable in any pairing with the other two variables?

(b) Produce the three possible scatter plots for the data, being mindful of the explanatory variable in each case.

(c) Describe any association that exists between the variables. Compare the strengths and limitations of the associations.

11 The data in the table gives the population, average life expectancy, GDP per capita (a measure of economic wealth) and average adult height, in several countries.

	Population (millions)	Life expectancy (years)	Per capita GDP ($US'000)	Adult height (cm)
Australia	24	82.5	50	168.7
Japan	127	85.0	39	165.0
Spain	49	81.7	27	168.5
New Zealand	4.5	81.2	39	170.5
UK	64	80.7	40	168.6
USA	324	79.8	57	168.8
China	1 474	75.5	8	161.5
Iran	83	71.4	5	163.8
Nigeria	186	53.4	2.2	160.8
Chad	12	50.2	0.7	165.6

(a) Some studies have suggested an association between the general health of a population and height.

(i) Which two variables would you use to investigate this association? Identify the explanatory variable.

(ii) Produce a scatter plot and describe any association it reveals between these two variables.

(b) There is thought to be an association between the general health of a population and its wealth.

(i) Which two variables would you use to investigate this association? Identify the explanatory variable.

(ii) Produce a scatter plot and describe any association it reveals between these two variables.

12 In the previous question, the variable 'population' was not used in any of the scatter plots. Draw three scatter plots to see if the size of a country's population affects any of the other aspects previously investigated. Describe any associations that are revealed.

13 The following experimental data was collected by scientists doing research on climate change. They measured the volume of carbon dioxide, CO_2, produced by vehicles running on petrol and LPG (liquid petroleum gas). (Note that LPG is up to 50% cheaper than petrol per litre, but is less fuel efficient than petrol.)

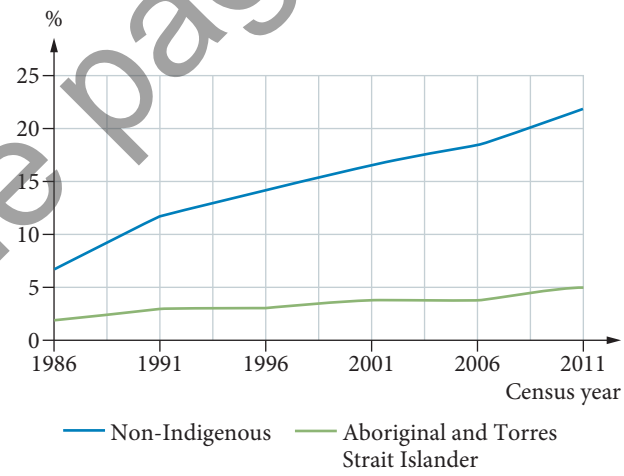
Petrol vehicles								
Petrol burned (L)	10	20	30	40	50	60	70	80
CO_2 produced (L)	13.5	25.6	38	53.5	67	74.2	25	108
LPG vehicles								
LPG burned (L)	10	20	30	40	50	60	70	80
CO_2 produced (L)	11.1	22	32.5	46	57.5	62.8	80.2	90

(a) Draw a scatter plot for each set of data.

(b) Describe the association evident in each scatter plot.

(c) Which fuel produces the least carbon dioxide per litre of fuel burned? Explain your answer.

14 The 2013 Census results published by the Australian Bureau of Statistics (ABS) contained information about the percentage of 15–24 year olds who were in higher education. The results comparing Indigenous and non-Indigenous students are shown in the graph.



(a) Draw a scatter plot of the percentage of Indigenous versus the percentage of non-Indigenous data, using the pair of values for each year.

(b) Describe the association shown by the data in the scatter plot.

(c) By examining the figures for 1986 and 2011, determine the rate of growth for each population and comment upon which population group is increasing its percentage more rapidly.

15 A class of Year 12 students were asked the age of each of their parents. The results are given in the table.

Mother's age	48	47	39	56	59	61	51	53	42	48	50	47	52
Father's age	45	50	42	59	62	61	52	52	44	53	49	53	56

(a) Write a general statement describing the association between the mothers' ages and the fathers' ages. Use a scatter plot to support your description.

(b) One of the students remembers participating in the same activity in Year 8. If the student was able to locate her analysis of the data, explain in what ways it would be the same and in what ways it would be different to the analysis in part (a).

(c) If the students were now asked to submit the age of each parent at the time they were born, explain in what ways the analysis would be the same and in what ways it would be different to the analysis in part (a).