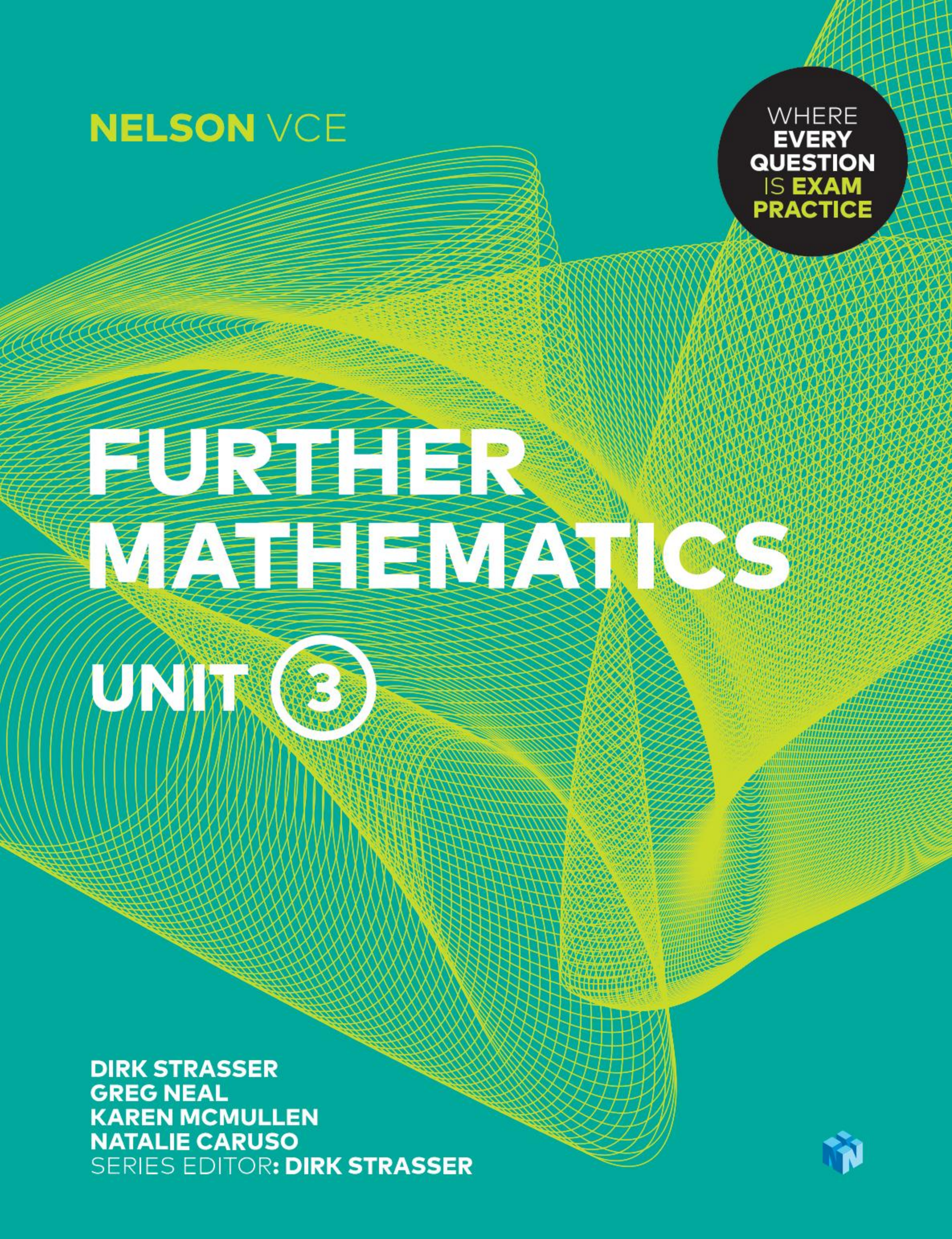


NELSON VCE

WHERE
EVERY
QUESTION
IS **EXAM**
PRACTICE

FURTHER MATHEMATICS

UNIT ③

DIRK STRASSER
GREG NEAL
KAREN MCMULLEN
NATALIE CARUSO
SERIES EDITOR: **DIRK STRASSER**

CENGAGE
Learning™

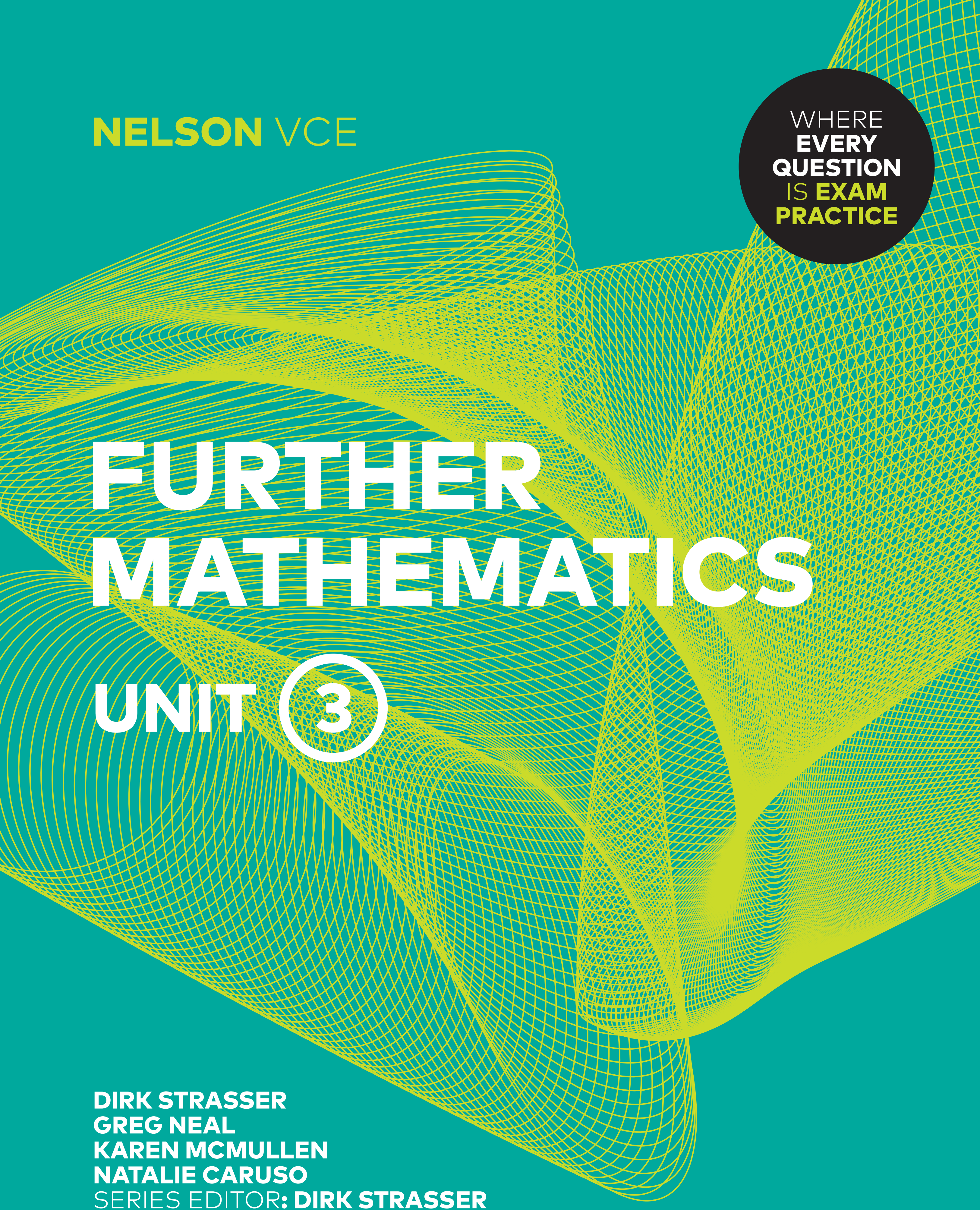
This is a legal agreement between the you, (the “Customer”) and Cengage Learning Australia Pty Limited (ABN 14 058 280 149) (the “Licensor”) which provides the terms and conditions of this non-exclusive licence and the limited warranty for the Product. Use of the Product indicates an acknowledgement that the Customer has read and agreed to be bound by the terms and conditions of this Agreement. If you do not agree to these terms and conditions, return the Product to the place of purchase within 15 days of the date of purchase (with proof of purchase) for a full refund

1. **Licence Grant**
   You do not receive title to the Product. Copyright in the Product (which includes all images, photographs, video, animations, audio, music and text incorporated in the Product, including all of the accompanying printed material) is owned by the Licensor and/or its suppliers and is protected by Australian copyright laws. The Licensor grants you a non-exclusive licence to use the Product subject to the restrictions and terms set out in this Agreement.

2. **A Licence allows you to:**
   Use the Product on your computer. The Customer represents that they shall in no way place the Product in the public domain or in any way compromise our copyright in the Material. You agree to take reasonable steps to protect our copyright.
3. **You may not:**
   Alter, modify, translate, reverse engineer, decompile, or adapt the software or create derivative works based on the Product. Make further copies by any means technological, electronic, digital whatsoever without the written permission of the Licensor. Rent or transfer all or any part of your rights under this Agreement. Remove or alter any copyright or other proprietary notice or label attached to the software.

4. **Termination**
   Any failure to comply with the terms and conditions of this agreement will result in the automatic termination of this licence. Upon termination of this licence for any reason, the Customer must destroy or return to the Licensor all copies of the software and accompanying documentation.

5. **Warranties**
   To the extent permitted by law, the Licensor’s liability for any breach of the warranty or any term implied by law into this licence is limited to the lowest cost of replacing the goods, acquiring equivalent goods or having the goods repaired.



NELSON VCE

WHERE
EVERY
QUESTION
IS EXAM
PRACTICE

FURTHER MATHEMATICS

UNIT 3

DIRK STRASSER
GREG NEAL
KAREN MCMULLEN
NATALIE CARUSO
SERIES EDITOR: **DIRK STRASSER**

NELSON
CENGAGE Learning

Nelson VCE Further Mathematics Unit 3
1st Edition
Dirk Strasser, Greg Neal, Karen McMullen, Natalie Caruso

Publisher: Dirk Strasser
Project editors: Anna Pang and Alan Stewart
Production controller: Erin Dowling
Permissions researcher: Kaitlin Jordan
Cover image: Corbis/© David Lidbetter/Galeries
Cover design: Chris Starr (Makework)
Text design: Nicole Melbourne
Typeset by: Cenveo Publisher Services

Selected VCE Examination questions (2002-2014) are copyright Victorian Curriculum and Assessment Authority (VCAA), reproduced by permission. VCE® is a registered trademark of the VCAA. The VCAA does not endorse this product and makes no warranties regarding the correctness or accuracy of this study resource. To the extent permitted by law, the VCAA excludes all liability for any loss or damage suffered or incurred as a result of accessing, using or relying on the content.
Current and past VCAA exams and related content can be accessed directly at www.vcaa.vic.edu.au

For product information and technology assistance,
in Australia call **1300 790 853**;
in New Zealand call **0800 449 725**

For permission to use material from this text or product, please email
aust.permissions@cengage.com

National Library of Australia Cataloguing-in-Publication Data
Strasser, Dirk, author.
Nelson VCE Further Mathematics Unit 3 / Dirk Strasser.

9780170371100 (paperback)
For secondary school age.

Mathematics--Study and teaching (Secondary)--Australia.
Mathematics.

510

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

Cengage Learning New Zealand
Unit 4B Rosedale Office Park
331 Rosedale Road, Albany, North Shore 0632, NZ

For learning solutions, visit **cengage.com.au**

Printed in China by China Translation & Printing Services.
1 2 3 4 5 6 7 19 18 17 16 15

About this series

Have you ever wondered why VCE maths series are filled with questions that will never be asked on the exam?

Nelson VCE Maths is the only VCE maths series where every question is exam practice.

Student unit books

- have a unique structure based on actual exam questions in a learning context
- use student performance data on exams to grade questions
- integrate continuous revision
- are half the weight of other VCE maths textbooks
- eliminate the need to buy a separate book of past exam questions

Nelson VCE Maths is the only VCE maths series where every support product is FREE*.

NelsonNetBooks

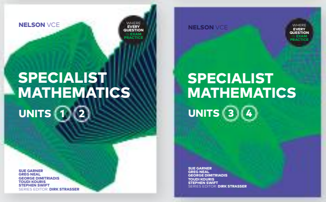
- are web-based ebooks accompanying each printed textbook
- can be customised with voice recordings, highlighting, drawings, and links to websites
- allow fast easy-to-use navigation

NelsonNet student website

- contains worksheets that are write-in enabled PDFs and prior learning exercises

NelsonNet teacher website*

- provides worked solutions to every question, for teachers to download and distribute to students
- includes curriculum grids, ExamView chapter quizzes and chapter PDFs of the student book



For more information, visit
www.nelsonsecondary.com.au/vcemaths

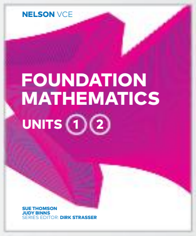
ExamView*

- is exam and assessment software that can downloaded from the NelsonNet teacher website
- includes a question bank of multiple choice and written response questions with answers
- allows you to quickly generate your own online or paper tests and practice exams
- allows you to make changes and add your own questions

*Please note that complimentary access to the NelsonNet teacher website and ExamView is only available to teachers who use the accompanying student textbook as a core educational resource in their classroom. Contact your sales representative for information about access codes and conditions.

And for the non-exam based VCE maths subject: ***Nelson VCE Foundation Mathematics Units 1 and 2***

- written by Sue Thomson and Judy Binns, experts on teaching and writing for senior applied mathematics courses
- engaging students with interesting and relevant applications of mathematics, with an emphasis on numeracy and literacy
- featuring optional extension topics for students who wish to advance to the Further Mathematics pathway



Contents

CORE: DATA ANALYSIS

1

Review of data distributions 1

2

Further data distributions 43

3

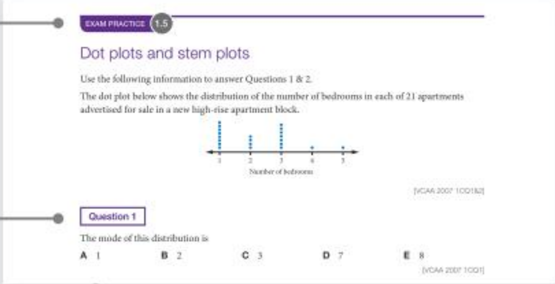
Associations between two variables 81

4

Linear associations 137

9780170371100

About this book



EXAM PRACTICE 1.5

Dot plots and stem plots

Use the following information to answer Questions 1 & 2.

The dot plot below shows the distribution of the number of bedrooms in each of 21 apartments advertised for sale in a new high-rise apartment block.

[VCAA 2007 1CQ1&2]

Question 1

The mode of this distribution is

A 1 **B** 2 **C** 3 **D** 7 **E** 8

[VCAA 2007 1CQ1]

Exam practice for every topic featuring carefully selected past exam questions

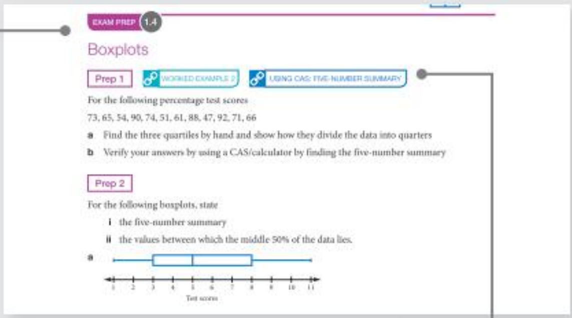
Questions graded in difficulty according to VCAA data on student exam performance

Questions requiring recall from previous textbook sections to support continuous revision

Exam practice

Are you under pressure to rush through course content to leave time to do sufficient exam practice at the end of the year?

Nelson VCE Maths eases the stress by integrating past exam and practice exam questions throughout the year.



EXAM PREP 1.4

Boxplots

Prep 1 WORKED EXAMPLE 2 USING CAS: FIVE-NUMBER SUMMARY

For the following percentage test scores

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

a Find the three quartiles by hand and show how they divide the data into quarters

b Verify your answers by using a CAS/calculator by finding the five-number summary

Prep 2

For the following boxplots, state

  i the five-number summary

  ii the values between which the middle 50% of the data lies.

a

Exam Prep focusing specifically on what is required to answer exam questions on a particular topic

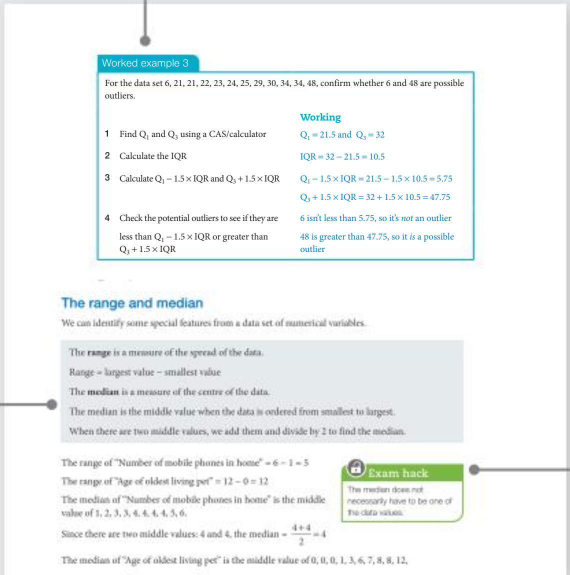
Links to matching worked examples and Using CAS

Exam preparation

Do your students struggle with the maths workload and complain you are ‘going too fast’?

Nelson VCE Maths filters out the unnecessary busy work for students that usually fills VCE mathematics textbooks and targets the requirements of the exams, so students can concentrate on what matters.

Worked Examples with clear instructions and working



Worked example 3

For the data set 6, 21, 21, 22, 23, 24, 25, 29, 30, 34, 34, 48, confirm whether 6 and 48 are possible outliers.

		Working
1	Find Q_1 and Q_3 using a CAS/calculator	$Q_1 = 21.5$ and $Q_3 = 32$
2	Calculate the IQR	$IQR = 32 - 21.5 = 10.5$
3	Calculate $Q_1 - 1.5 \times IQR$ and $Q_3 + 1.5 \times IQR$	$Q_1 - 1.5 \times IQR = 21.5 - 1.5 \times 10.5 = 5.75$ $Q_3 + 1.5 \times IQR = 32 + 1.5 \times 10.5 = 47.75$
4	Check the potential outliers to see if they are less than $Q_1 - 1.5 \times IQR$ or greater than $Q_3 + 1.5 \times IQR$	6 isn’t less than 5.75, so it’s *not* an outlier 48 is greater than 47.75, so it *is* a possible outlier

The range and median

We can identify some special features from a data set of numerical variables.

The **range** is a measure of the spread of the data.

Range = largest value − smallest value

The **median** is a measure of the centre of the data.

The median is the middle value when the data is ordered from smallest to largest.

When there are two middle values, we add them and divide by 2 to find the median.

The range of “Number of mobile phones in home” = 6 − 1 = 5

The range of “Age of oldest living pet” = 12 − 0 = 12

The median of “Number of mobile phones in home” is the middle value of 1, 2, 3, 3, 4, 4, 4, 4, 5, 6.

Since there are two middle values: 4 and 4, the median = $\frac{4+4}{2} = 4$

The median of “Age of oldest living pet” is the middle value of 0, 0, 0, 1, 3, 6, 7, 8, 8, 12,

Exam hack

The median does not necessarily have to be one of the data values.

Short, sharp summary boxes

Exam hacks with ‘insider’ information based on difficulties students have encountered in past exams

Targeted theory

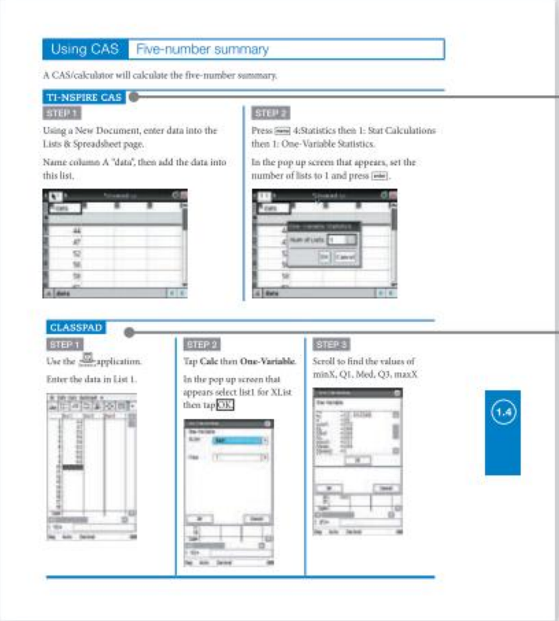
Do your students find it difficult to make sense of overly detailed and convoluted theory in textbooks?

Nelson VCE Maths provides clear and concise theory, so your students don’t get bogged down with wordy explanations.

CAS focus

Do your students struggle to find relevant instructions on how to use their CAS at the point of need?

Nelson VCE Maths has CAS instructions clearly labelled and located where students need them.



Using CAS Five-number summary

A CAS/calculator will calculate the five-number summary.

TI-NSPIRE CAS

STEP 1

Using a New Document, enter data into the Lists & Spreadsheet page.

Name column A "data", then add the data into this list.

STEP 2

Press menu 4:Statistics then 1: Stat Calculations then 1: One-Variable Statistics.

In the pop up screen that appears, set the number of lists to 1 and press enter.

CLASSPAD

STEP 1

Use the application.

Enter the data in List 1.

STEP 2

Tap **Calc** then **One-Variable**.

In the pop up screen that appears select list1 for XList then tap OK.

STEP 3

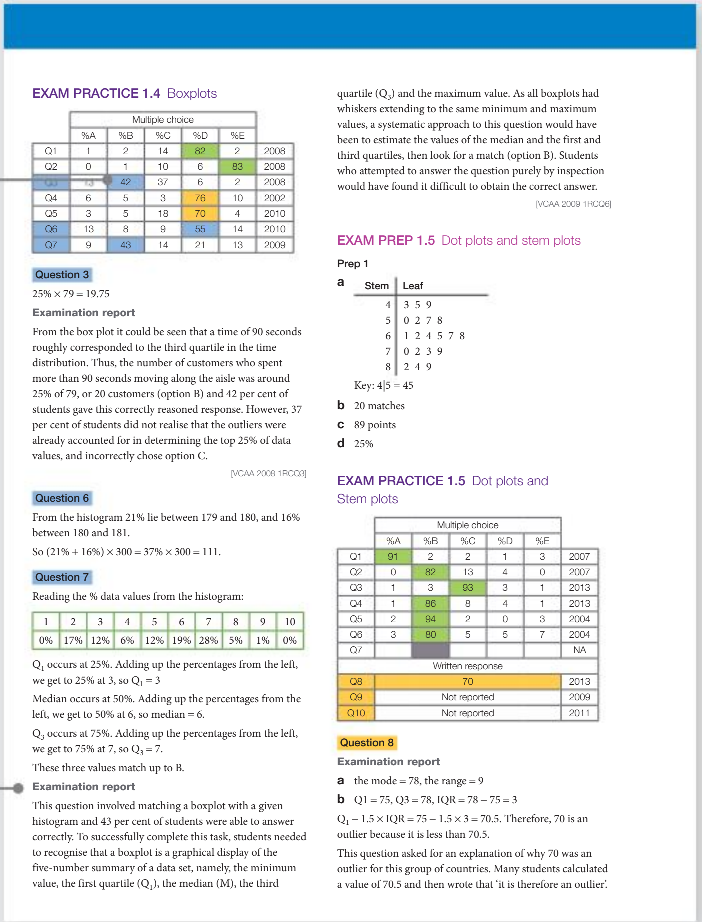
Scroll to find the values of minX, Q1, Med, Q3, maxX

Both TI-Nspire CAS and ClassPad steps and screen shots

Detailed solutions

Do your students sometimes need more than just answers at the back of the book?

For selected questions, ***Nelson VCE Maths*** also provides worked solutions and identification of common errors from the VCAA examination reports (worked solutions to *all* questions can be downloaded from the NelsonNet teacher website).



EXAM PRACTICE 1.4 Boxplots

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1	1	2	14	82	2	2008
Q2	0	1	10	6	83	2008
Q3	13	42	37	6	2	2008
Q4	6	5	3	76	10	2002
Q5	3	5	18	70	4	2010
Q6	13	8	9	55	14	2010
Q7	9	43	14	21	13	2009

Question 3

25% × 79 = 19.75

Examination report

From the box plot it could be seen that a time of 90 seconds roughly corresponded to the third quartile in the time distribution. Thus, the number of customers who spent more than 90 seconds moving along the aisle was around 25% of 79, or 20 customers (option B) and 42 per cent of students gave this correctly reasoned response. However, 37 per cent of students did not realise that the outliers were already accounted for in determining the top 25% of data values, and incorrectly chose option C.

[VCAA 2008 1RCQ3]

Question 6

From the histogram 21% lie between 179 and 180, and 16% between 180 and 181.

So (21% + 16%) × 300 = 37% × 300 = 111.

Question 7

Reading the % data values from the histogram:

1	2	3	4	5	6	7	8	9	10
0%	17%	12%	6%	12%	19%	28%	5%	1%	0%

Q_1 occurs at 25%. Adding up the percentages from the left, we get to 25% at 3, so $Q_1 = 3$

Median occurs at 50%. Adding up the percentages from the left, we get to 50% at 6, so median = 6.

Q_3 occurs at 75%. Adding up the percentages from the left, we get to 75% at 7, so $Q_3 = 7$.

These three values match up to B.

Examination report

This question involved matching a boxplot with a given histogram and 43 per cent of students were able to answer correctly. To successfully complete this task, students needed to recognise that a boxplot is a graphical display of the five-number summary of a data set, namely, the minimum value, the first quartile (Q_1), the median (M), the third quartile (Q_3) and the maximum value. As all boxplots had whiskers extending to the same minimum and maximum values, a systematic approach to this question would have been to estimate the values of the median and the first and third quartiles, then look for a match (option B). Students who attempted to answer the question purely by inspection would have found it difficult to obtain the correct answer.

[VCAA 2009 1RCQ6]

EXAM PREP 1.5 Dot plots and stem plots

Prep 1

a

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

Key: 4|5 = 45

b 20 matches

c 89 points

d 25%

EXAM PRACTICE 1.5 Dot plots and Stem plots

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1	91	2	2	1	3	2007
Q2	0	82	13	4	0	2007
Q3	1	3	93	3	1	2013
Q4	1	86	8	4	1	2013
Q5	2	94	2	0	3	2004
Q6	3	80	5	5	7	2004
Q7						NA
	Written response					
Q8	70					2013
Q9	Not reported					2009
Q10	Not reported					2011

Question 8

Examination report

a the mode = 78, the range = 9

b Q1 = 75, Q3 = 78, IQR = 78 − 75 = 3

$Q_1 - 1.5 \times IQR = 75 - 1.5 \times 3 = 70.5$. Therefore, 70 is an outlier because it is less than 70.5.

This question asked for an explanation of why 70 was an outlier for this group of countries. Many students calculated a value of 70.5 and then wrote that 'it is therefore an outlier'.

Colour-coded percentages based on VCAA data indicating the difficulty level of exam questions

Question-specific extracts from the VCAA Examination report

About the authors

Series editor and lead author Dirk Strasser taught VCE Mathematics for 12 years and has been a senior educational publisher for 18 years. He has published and co-written eight best-selling mathematics series, including Heinemann Maths Zone 7–10 and Pearson Mathematics 7–10.

Greg Neal is the Mathematics Learning Area Leader at Ballarat High School and has been teaching for over 30 years. He has been a VCAA examination assessor, is an experienced Further Mathematics author, and extremely proficient with CAS technology.

Karen McMullen is the Mathematics Learning Leader at Killester College, Springvale. She has completed a Master's degree in School Leadership (Numeracy), taught VCE maths for ten years, and has written Further Mathematics SACs for MAV.

Natalie Caruso is a former Head of Department, currently teaching mathematics at Loreto Mandeville Hall, Toorak, in Victoria. She has authored a significant number of mathematics texts and study guides, and has many years of experience working with CAS calculators. She has presented regularly at MAV, and has been a VCE exam vetter and marker.

CHAPTER 1

REVIEW OF DATA DISTRIBUTIONS

Prior learning

1.1 Data and variables

Data needs to be organised so that we can make sense of it. Without organisation, data is simply a meaningless collection of information.

Here's some data:

Responses by a group of ten students
1, 1, 1, 0, 3, 3, 4, 2, 1, 0
white, white, grey, white, red, blue, red, red, red, silver
4, 3, 5, 2, 1, 4, 6, 3, 4, 4
8, 3, 6, 1, 0, 0, 12, 7, 8, 0
3149, 3149, 3148, 3149, 3149, 3149, 3166, 3147, 3149, 3148

There's not much we can do with this. What are missing are the **variables**, the things about which we are recording information:

Reponses by a group of ten students

Variables	Data
Rating given to latest Star Wars movie where 1 = great, 2 = okay, 3 = awful, 4 = didn't see it	1, 1, 1, 0, 3, 3, 4, 2, 1, 0
Colour of the first car seen this morning	white, white, grey, white, red, blue, red, red, red, silver
Number of mobile phones in home	4, 3, 5, 2, 1, 4, 6, 3, 4, 4
Age of oldest living pet (in years)	8, 3, 6, 1, 0, 0, 12, 7, 8, 0
Postcode of home address	3149, 3149, 3148, 3149, 3149, 3149, 3166, 3147, 3149, 3148

How you organise this data depends on what sort of variables are involved.

Getty Images/Anadolu Agency

Types of data

Statistical data match-up

Categorical data
Data that can be sorted into categories or groups
Ask: Is the number being used like a name?
YES

Nominal data
Categorical data that doesn't have a natural order, even when numbers are involved
e.g. Colour of first car this morning
Postcodes of home address
Numbers on basketball player uniforms
Ask: Does it make sense to order them?
NO

Ordinal data
Categorical data that has a natural order, but doesn't involve counting anything and can't be measured to an increasing level of accuracy
e.g. Ratings given to the latest Star Wars movie (1 = great, 2 = okay, 3 = awful, 4 = didn't see it)
House numbers
Shoe sizes
Grades in a test (A, B, C, D, E)
Clothing sizes (small, medium, large, extra large)
Ask: Does it make sense to order them?
YES

Numerical data
Data that can be counted or measured to an increasing level of accuracy
Ask: Is the number being used like a name?
NO

Discrete data
Numerical data that can be counted
e.g. The number of mobile phones in home
The number of people who like Vegemite
The amount of money spent in a shop
Ask: Does it involve counting something?
YES

Continuous data
Numerical data that can be measured to an increasing level of accuracy
e.g. The age of oldest living pet
Time taken to run 100 metres
The height of players in a basketball team
Ask: Could I measure this more accurately if I wanted to?
YES

The range and median

We can identify some special features of a data set of numerical variables.

The **range** is a measure of the spread of the data.

Range = largest value – smallest value

The **median** is a measure of the centre of the data.

The median is the middle value when the data is ordered from smallest to largest.

When there are two middle values, we add them and divide by 2 to find the median.

The range of 'Number of mobile phones in home' = 6 – 1 = 5

The range of 'Age of oldest living pet' = 12 – 0 = 12

The median of 'Number of mobile phones in home' is the middle value of 1, 2, 3, 3, 4, 4, 4, 4, 5, 6.

Since there are two middle values: 4 and 4, the median $= \dfrac{4+4}{2} = 4$

Exam hack

The median does not necessarily have to be one of the data values.

The median of 'Age of oldest living pet' is the middle value of 0, 0, 0, 1, 3, 6, 7, 8, 8, 12.

Since there are two middle values: 3 and 6, the median $= \dfrac{3+6}{2} = 4.5$

EXAM PREP 1.1

Data and variables

Prep 1

State whether the following variables are nominal categorical, ordinal categorical, discrete numerical, or continuous numerical.

Recording information on

a number of television sets in the house

b how much time spent doing homework last night

c the name of the suburb in home address

d the distance travelled to school

e the speed of cars at a certain point on a freeway

f candidate intending to vote for in the next election

g numbers on the Socceroos' uniforms

h opinion of AFL football on a scale of 1 to 5 where 1 is hate and 5 is love

i mother's salary in dollars

j mother's salary classified as high, medium or low

Prep 2

Find the range and median of each of the following.

a the ages of people (in years) working in a restaurant: 42, 21, 18, 35, 19, 18, 27

b the number of people buying coffees at a café on 6 consecutive mornings: 39, 33, 35, 38, 56, 41

c the lowest maximum monthly temperature (°C) at a ski resort from April to November:
3.0, −2.0, −1.6, −2.0, −1.8, −2.8, 0.8, 2.1

EXAM PRACTICE 1.1

Data and variables

Question 1

The variables

region (city, urban, rural)
population density (number of people per square kilometre)

A are both categorical.

B are both numerical.

C are categorical and numerical respectively.

D are numerical and categorical respectively.

E are neither categorical nor numerical.

[VCAA 2011 1CQ4]

Use the following information to answer Questions 2 & 3.

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

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

Question 2

The median investment return is

A –4.7% **B** –4.6% **C** –4.5% **D** –4.4% **E** 0.3%

[VCAA 2003 1CQ1]

Question 3

The range of investment returns is

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

[VCAA 2003 1CQ2]

Question 4

Researchers conducted a survey of 403 school leavers who had recently entered the workforce. The aim was to determine whether the type of work they undertook was gender related. Work type was classified as 'trade', 'clerical', 'manual' or 'professional'.

In this survey, the variables

work type (trade, clerical, manual or professional)

and

gender (male or female)

are

A both categorical variables.

B both numerical variables.

C categorical and numerical variables respectively.

D numerical and categorical variables respectively.

E neither categorical nor numerical variables.

[VCAA 2001 1CQ1]

Question 5

The level of water usage of 250 houses was rated in a survey as low, medium or high, and the size of the houses as small, standard or large.

The variables **level of water usage** and **size of house**, as recorded in this survey, are

A both numerical variables.

B both categorical variables.

C neither numerical nor categorical variables.

D numerical and categorical variables respectively.

E categorical and numerical variables respectively.

[VCAA 2003 1CQ7]

1.2 Tables and charts

When you are dealing with a large number of data values, to see patterns or draw conclusions you need to organise and display the data in a manageable form. When choosing a display for your data, you must decide which one best shows what you wish to communicate.

WS

Frequency distribution tables

Frequency tables

Frequency tables can be used to display both categorical and numerical data. The data values are listed in one column and the corresponding frequencies are displayed in a frequency column.

Variable name	Tally	Frequency (*f*)
Response 1		
Response 2		
Response 3		
Response 4		
	Total	

Place a line in the tally column each time a response occurs. This helps us to keep track. Remember that 卌 means 5.

Shutterstock.com/Tyler Olson

Worked example 1

Set up a frequency table for a coffee shop owner who wants to record the first 20 types of hot drinks he sold one morning.

Working

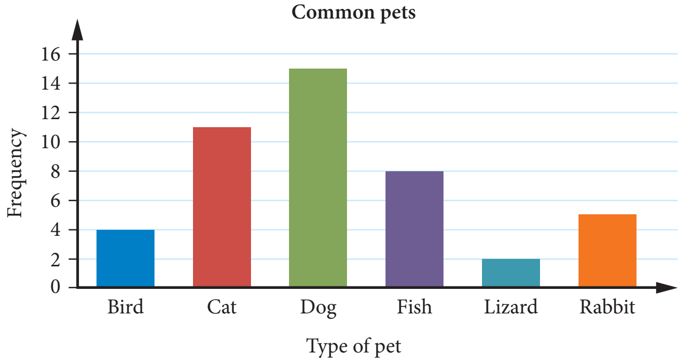
Set up a three column table. Tally and record the frequency.

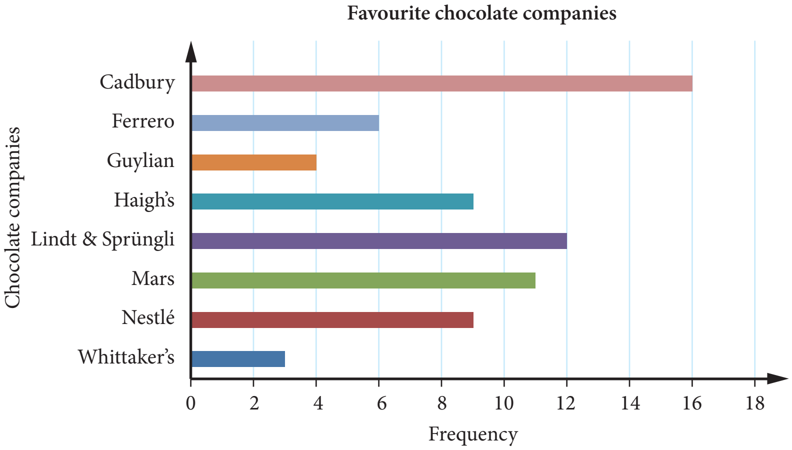
Hot drink	Tally	Frequency (*f*)
Cappuccino	卌 |||	8
Latte	卌 |||	8
Tea	||||	4
Total		20

Always add up the frequency column to make sure that you have included all the data.

Bar charts

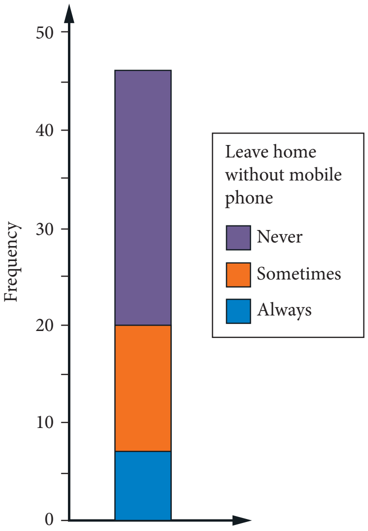
Bar charts help us see patterns when dealing with categorical variables. Categories can be represented on the horizontal or vertical axis, with their corresponding frequency on the other axis:





Segmented bar charts

In a **segmented bar chart**, the bars are stacked on one another to give one bar with several segments. The length of each segment tells us the frequency, and the height of the bar gives the total frequency. In a **percentage segmented bar chart** the height of the bar is 100. A segmented bar chart should have no more than five segments or it becomes difficult to read. A legend needs to be included to explain what the segments represent.



EXAM PREP 1.2

Tables and charts

Prep 1

WORKED EXAMPLE 1

For the following frequency table, construct a bar chart with the categories on the horizontal axis. Write an explanation of the data displayed on the graph.

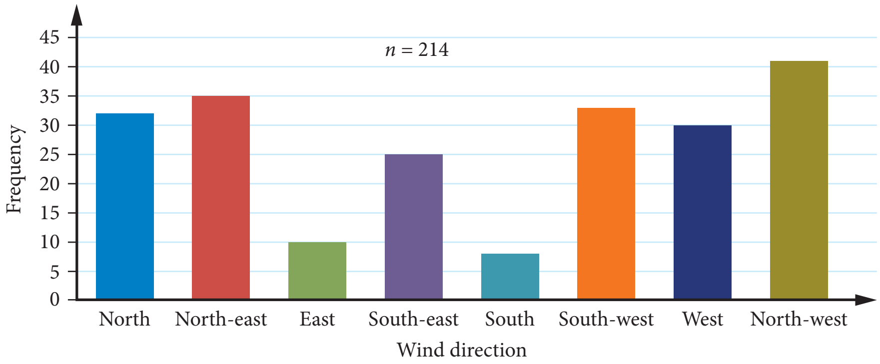
Hot drink	Tally	Frequency (*f*)
Cappuccino	卌 |||	8
Latte	卌 |||	8
Tea	||||	4
Total		20

EXAM PRACTICE 1.2

Tables and charts

Use the following information to answer Questions 1 & 2.

The following bar chart shows the distribution of wind directions recorded at a weather station at 9.00 a.m. on each of 214 days in 2011.



Question 1

According to the bar chart, the most frequently observed wind direction was

A south-east.
B south.
C south-west.
D west.
E north-west.

[VCAA 2012 1CQ1]

Question 2

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

A 10%
B 16%
C 25%
D 33%
E 35%

[VCAA 2012 1CQ2]

Question 3

The passengers on a train were asked why they travelled by train. Each reason, along with the percentage of passengers who gave that reason, is displayed in the segmented bar chart below.

The percentage of passengers who gave the reason 'no car' is closest to

A 14%

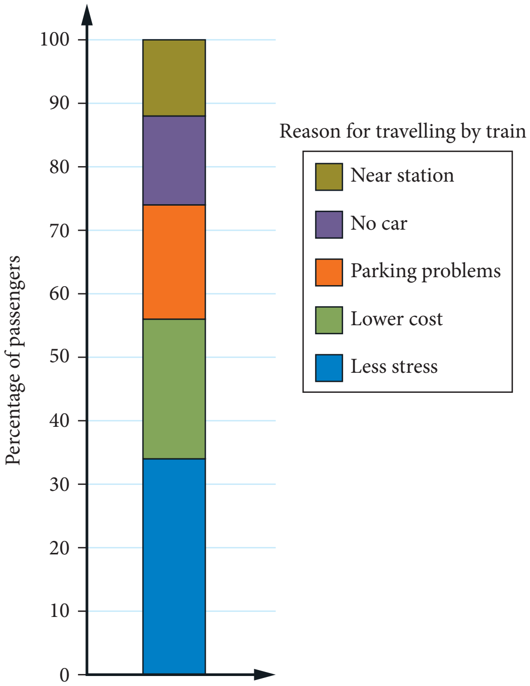
B 18%

C 26%

D 74%

E 88%

[VCAA 2010 1CQ4]



Question 4

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

sat	stood	sat	ran	sat
walked	walked	sat	walked	ran
sat	walked	walked	walked	ran
walked	ran	walked	ran	walked
ran	sat	ran	ran	walked

Use the data to

a copy and complete the following frequency table

Activity	Frequency
walked	
sat or stood	
ran	
Total	25

1 mark

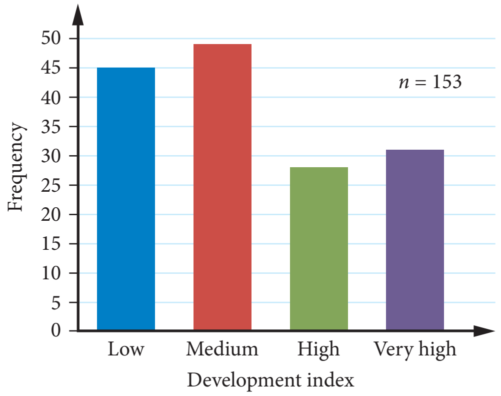
b determine the percentage of Year 8 girls who ran for most of the time during a typical school lunch time. 1 mark

[VCAA 2008 2CQ1]

Question 5

A development index is used as a measure of the standard of living in a country.

The bar chart displays the development index for 153 countries in four categories: low, medium, high and very high.



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

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

[VCAA 2013 2CQ1]

Histograms

1.3

A **histogram** is a graphical way of displaying numerical data from a frequency table. It is effective when dealing with data that has been grouped into a small number (usually between 5 and 15) of intervals.

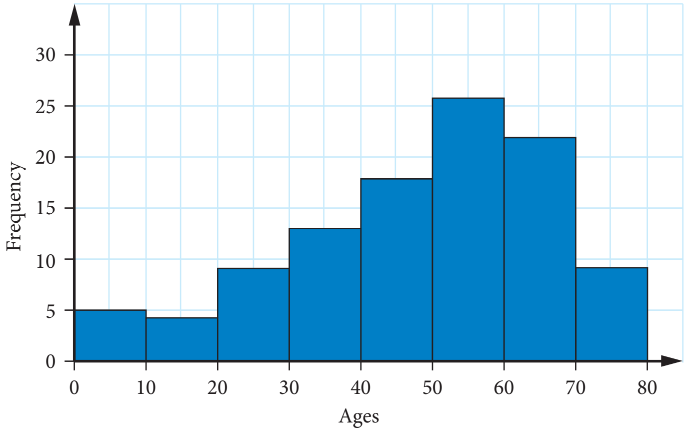
Continuous data

Histograms for continuous numerical data have bars that:

- correspond to data intervals and
- sit in between the interval values on the horizontal axis.

The following is a histogram for the ages of people in a small town where the intervals are < 10 years, 10–< 20 years, 20–< 30 years etc.

The interval 10–<20, for example, includes all ages starting at 10 up to but not including 20.



Grouped discrete data

Histograms for discrete numerical data that has been grouped into intervals work much the same way as for continuous numerical data. For example, the previous histogram could be showing the number of occasions families have eaten takeaway in the last year, where the intervals are 0–9 occasions, 10–19 times etc.

Exam hack

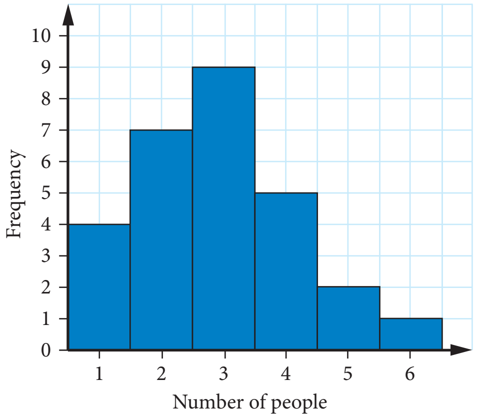
Although at first glance a histogram looks similar to a bar chart, there are a number of differences:

- Histograms are used for displaying numerical data, while bar charts display categorical data
- Histograms don't have any spaces between the columns, while bar charts do
- Histograms always have the frequency on the vertical axis, while bar charts can have the frequency on either axis.

Ungrouped discrete data

Histograms for discrete numerical data that *hasn't* been grouped into intervals have bars starting and ending halfway between scale marks on the horizontal axis.

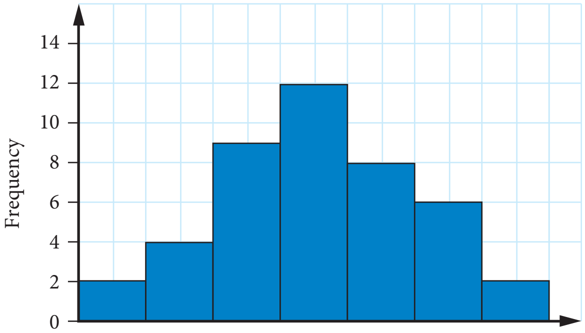
The following is a histogram of the number of people living in each house in a suburban block.



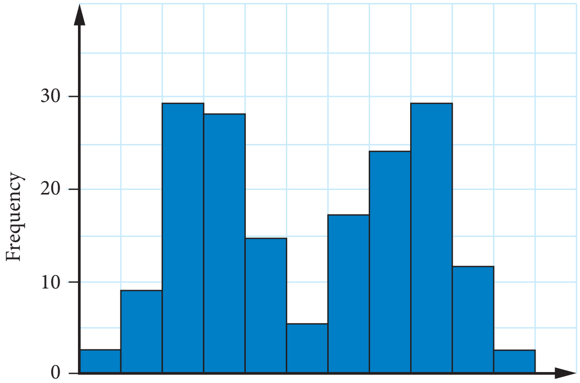
Symmetric and skewed distributions

The shape of a histogram tells us about the underlying frequency distribution. The following two histograms show **symmetric distributions**.

This is a common single-peaked approximately symmetric distribution:

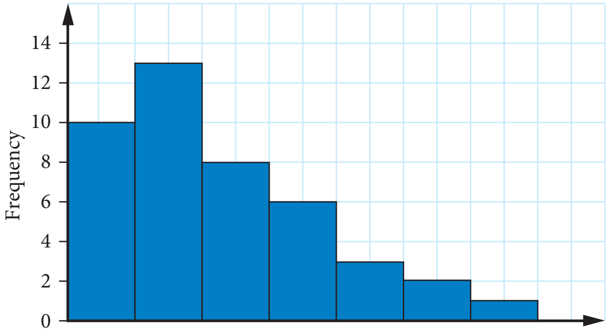


This is a double-peaked approximately symmetric distribution:

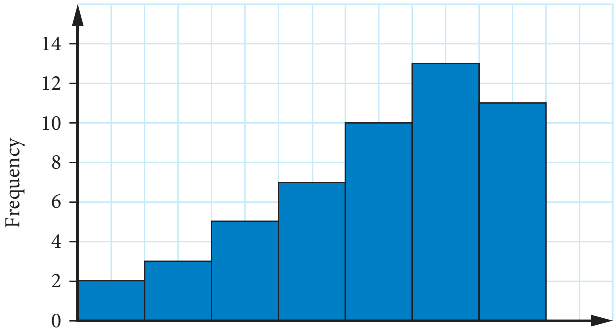


The following two histograms show **skewed distributions**.

This histogram is **positively skewed**.



This histogram is **negatively skewed**.



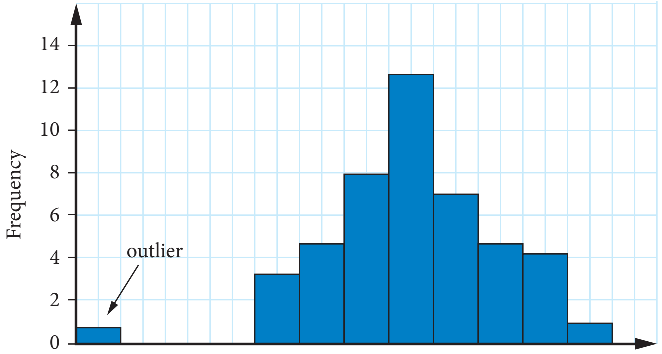
Exam hack

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

iStock.com/Julie Birch

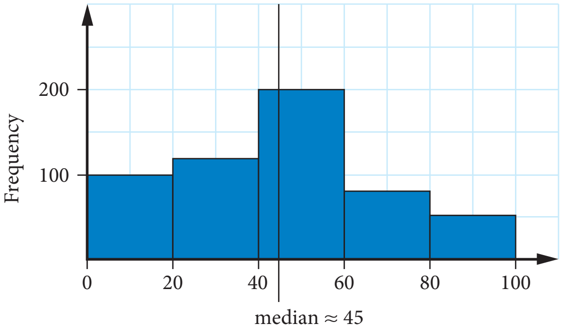
Outliers

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



Medians

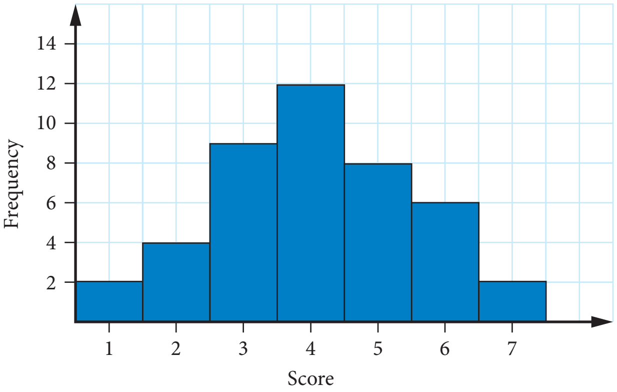
An estimate of the **median**, or middle, of a distribution can often be found by looking at the histogram. The median occurs at the vertical line that splits the histogram in half with equal areas on either side.



Range

The **range** can be read from a histogram only when we have discrete values increasing by ones. For the following histogram,

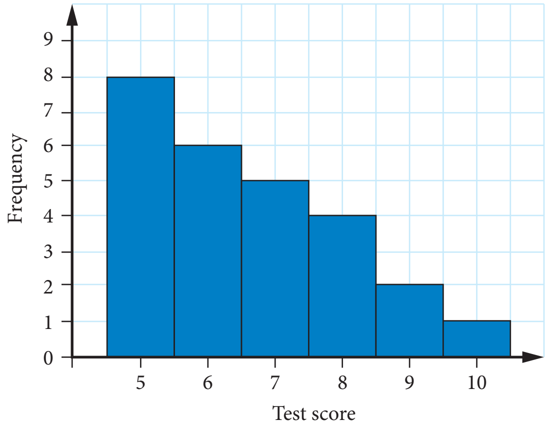
the range = largest value − smallest value = 7 − 1 = 6



Histograms

Prep 1

The following histogram displays the results of a Year 12 History quiz.



a How many students did the History quiz?

b Describe the shape of the data distribution.

c Using the histogram, copy and complete the frequency table.

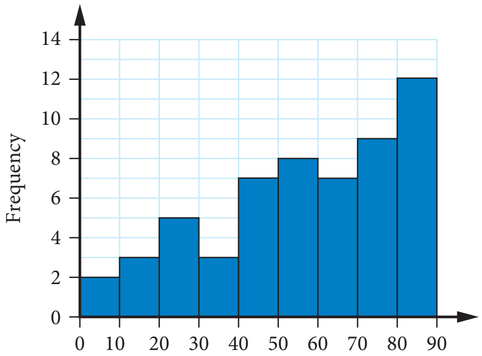
d What percentage of students achieved a score greater than 8 for the History quiz? Give your answer to 1 decimal place.

e What is the range of the data?

Score	Frequency
5	
6	
7	
8	
9	
10	
Total	

Prep 2

Find the median of the data in this histogram.



EXAM PRACTICE **1.3**

Histograms

Question 1

1526 students sat for an examination. The histogram shows the distribution of marks.

The median examination mark of these students is closest to

A 50

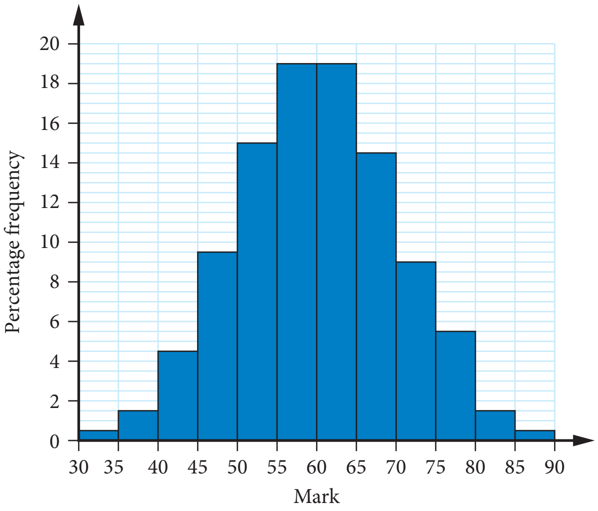
B 55

C 60

D 65

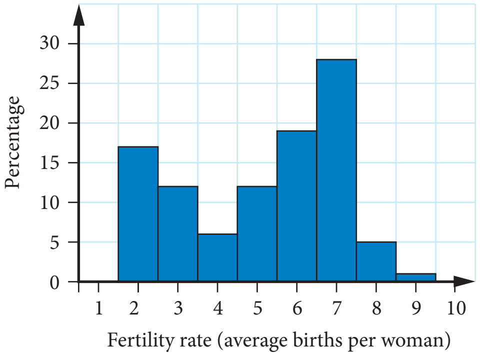
E 70

[VCAA 2002 1CQ5]



Use the following information to answer Questions 2 & 3.

The percentage histogram shows the distribution of the fertility rates (in average births per woman) for 173 countries in 1975.



Question 2

In 1975, for these 173 countries, fertility rates were most frequently

A less than 2.5

B between 1.5 and 2.5

C between 2.5 and 4.5

D between 6.5 and 7.5

E greater than 7.5

[VCAA 2009 1CQ5]

Question 3

In 1975, the percentage of these 173 countries with fertility rates of 4.5 or greater was closest to

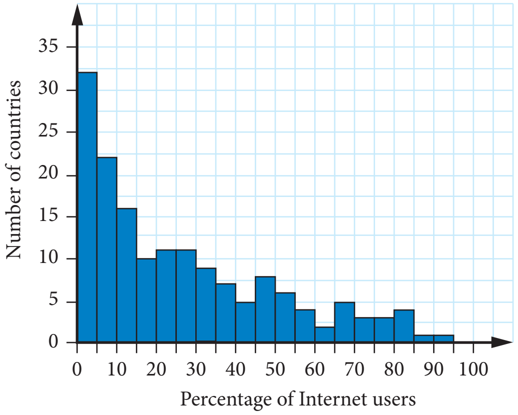
A 12% **B** 35% **C** 47% **D** 53% **E** 65%

[VCAA 2009 1CQ4]

9780170371100

Use the following information to answer Questions 4–6.

The histogram below displays the distribution of the percentage of Internet users in 160 countries in 2007.



Question 4

The shape of the histogram is best described as

A approximately symmetric.

B bell shaped.

C positively skewed.

D negatively skewed.

E bi-modal.

[VCAA 2011 1CQ1]

Question 5

The number of countries in which fewer than 10% of people are Internet users is closest to

A 10 **B** 16 **C** 22 **D** 32 **E** 54

[VCAA 2011 1CQ2]

Question 6

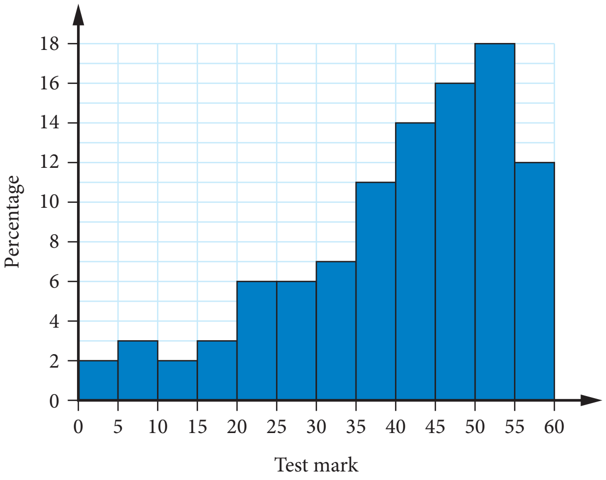
From the histogram, the median percentage of Internet users is closest to

A 10% **B** 15% **C** 20% **D** 30% **E** 40%

[VCAA 2011 1CQ3]

Use the following information to answer Questions 7 & 8.

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



Question 7

The pass mark on the test was 30 marks.

The percentage of students who passed the test is

A 7% **B** 22%

C 50% **D** 78%

E 87%

[VCAA 2006 1CQ5]

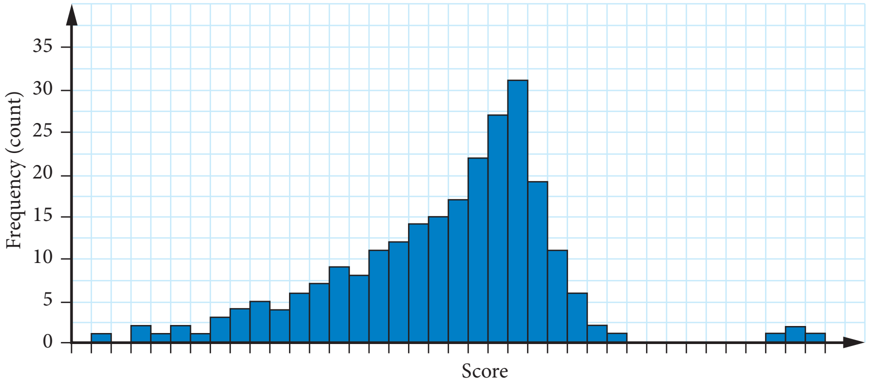
Question 8

The median mark lies between

A 35 and 40 **B** 40 and 45 **C** 45 and 50 **D** 50 and 55 **E** 55 and 60

[VCAA 2006 1CQ6]

Question 9



The histogram above is best described as

A negatively skewed.

B positively skewed.

C symmetric.

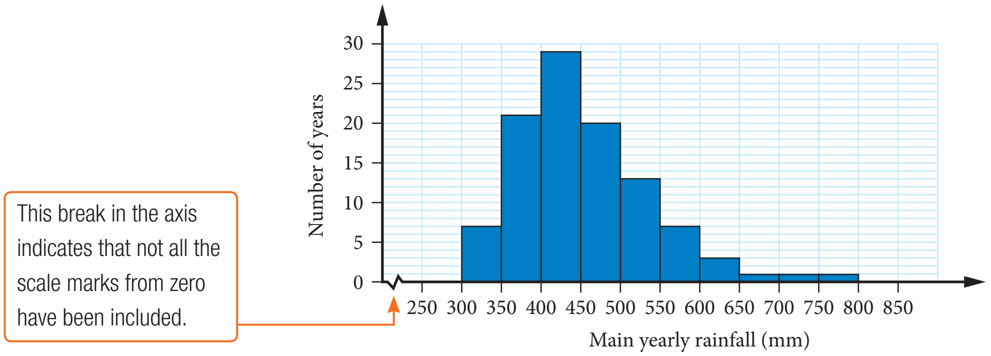
D negatively skewed with outliers.

E positively skewed with outliers.

[VCAA 2005 1CQ3]

Question 10

The histogram shows the distribution of mean yearly rainfall (in mm) for Australia over 103 years.



This break in the axis indicates that not all the scale marks from zero have been included.

Data source: ABS 2007

a Describe the shape of the histogram. 1 mark

b Use the histogram to determine

- **i** the number of years in which the mean yearly rainfall was 500 mm or more 1 mark
- **ii** the percentage of years in which the mean yearly rainfall was between 500 mm and 600 mm. Write your answer correct to 1 decimal place. 1 mark

[VCAA 2007 2CQ1]

Boxplots

1.4

Quartiles

WS Boxplots 1

WS Boxplots 2

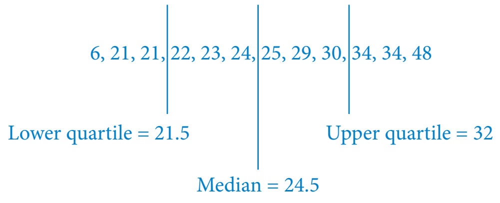
Quartiles are the three points that divide a set of data into quarters.

- The first or **lower quartile** has 25% of the data below it.
- The second quartile (which is the same as the median) has 50% of the data below it.
- The third or **upper quartile** has 75% of the data below it.

Worked example 2

Find the three quartiles for the following data and show how they divide the data into quarters:

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

	Working
1 Order the data from smallest to largest.	6, 21, 21, 22, 23, 24, 25, 29, 30, 34, 34, 48
2 Find the median. There is an even number of data values, so there are two middle points (24 and 25). Add them and divide by 2 to find the median.	median $= \frac{24+25}{2} = 24.5$
3 Find the median of the lower half of the data. This is the **lower quartile**.	The lower half of the data is 6, 21, 21, 22, 23, 24. lower quartile $= \frac{21+22}{2} = 21.5.$
4 Find the median of the upper half of the data. This is the **upper quartile**.	The upper half of the data is 25, 29, 30, 34, 34, 48 upper quartile $= \frac{30+34}{2} = 32.$
5 Show how the three quartiles divide the data into quarters.	

We use the following symbols when dealing with quartiles:

Q_1 is the lower quartile (the median of the lower half of the data)

Q_2 is the median

Q_3 is the upper quartile (the median of the upper half of the data)

Exam hack

If there is an odd number of data values (which means the median will be one of the actual data values), then you can't split the data into equal lower and upper halves. When this happens, leave out the median value when you do the quartile calculations.

The five-number summary

The **five-number summary** provides a good overview of a distribution.

It consists of

1. The minimum data value
2. Q_1
3. The median
4. Q_3
5. The maximum data value

So, for worked example 2, the five-number summary is:

minimum = 6, $Q_1 = 21.5$, median = 24.5, $Q_3 = 32$, maximum = 48

Shutterstock.com/alysta

Using CAS Five-number summary

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

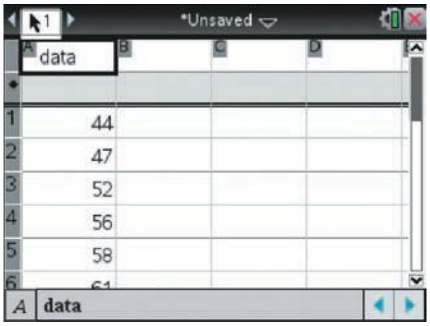
65, 47, 61, 44, 63, 56, 65, 52, 58

TI-NSPIRE CAS

STEP 1

Using a New Document, enter data into the Lists & Spreadsheet page.

Name column A 'data', then add the data into this list.

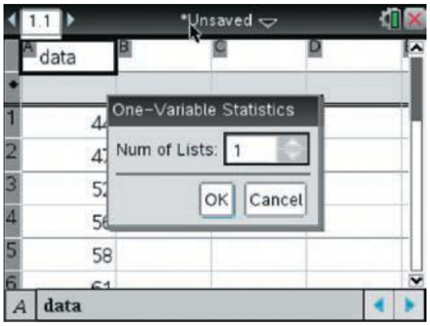


	data
1	44
2	47
3	52
4	56
5	58

STEP 2

Press menu 4: Statistics then 1: Stat Calculations then 1: One-Variable Statistics.

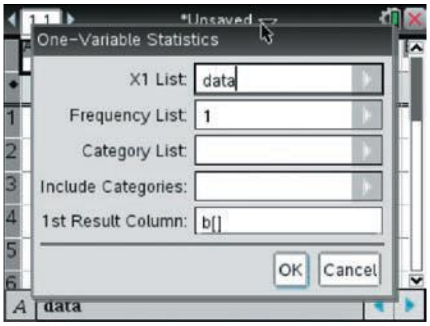
In the pop up screen that appears, set the number of lists to 1 and press enter.



Num of Lists: 1

STEP 3

Select 'data' for the X1 List, press ▶ then ▼ until 'data is selected then press enter. Leave the frequency list set as 1 then press tab down to OK and press enter.



X1 List: data

Frequency List: 1

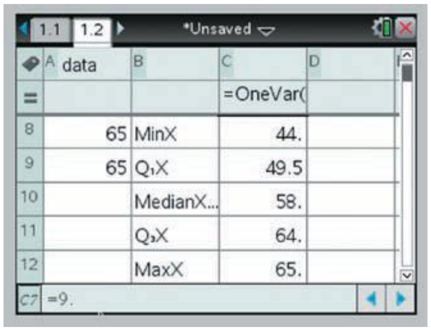
Category List:

Include Categories:

1st Result Column: b[]

STEP 4

Scroll to find the values of MinX, Q1X, MedianX, Q3X, MaxX



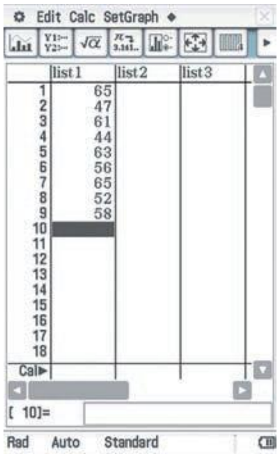
	A data	B	C	D
=			=OneVar(	
8	65	MinX	44.	
9	65	Q_1X	49.5	
10		MedianX...	58.	
11		Q_3X	64.	
12		MaxX	65.	

CLASSPAD

STEP 1

Use the Statistics application.

Enter the data in List 1.



```
list1
1  65
2  47
3  61
4  44
5  63
6  56
7  65
8  52
9  58
```

STEP 2

Tap **Calc** then **One-Variable**.

In the pop-up screen that appears select list1 for XList then tap OK.

Set Calculation

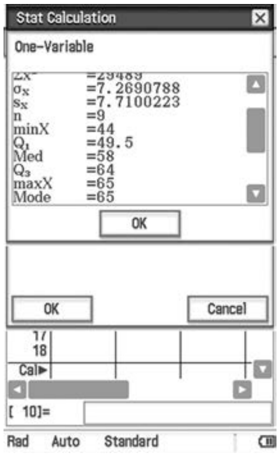
One-Variable

XList: list1

Freq: 1

STEP 3

Scroll to find the values of minX, Q1, Med, Q3, maxX.



Stat Calculation

One-Variable

```
Σx²    =29489
σx     =7.2690788
sx     =7.7100223
n      =9
minX   =44
Q1     =49.5
Med    =58
Q3     =64
maxX   =65
Mode   =65
```

IQR and outliers

Interquartile range

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

IQR = $Q_3 - Q_1$

The IQR is often a better measure of spread than the range because, by looking at only the middle 50% of data, we avoid taking outliers into account.

The IQR is also used to define possible outliers, so we don't have to rely on imprecise methods (i.e. simply saying something 'looks like an outlier').

IQR = $Q_3 - Q_1$

A data value is a possible outlier if it is

less than **the lower fence** $Q_1 - 1.5 \times \text{IQR}$ or

greater than **the upper fence** $Q_3 + 1.5 \times \text{IQR}$

Worked example 3

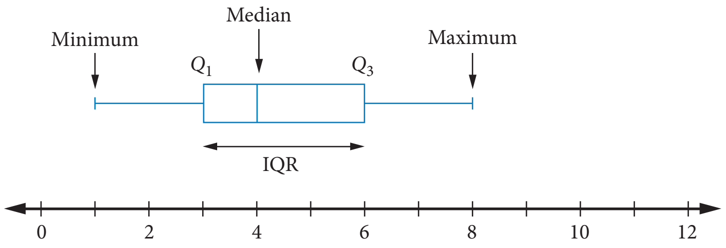
For the data set 6, 21, 21, 22, 23, 24, 25, 29, 30, 34, 34, 50, confirm whether 6 and 50 are possible outliers.

	Working
1 Find Q_1 and Q_3 using a CAS/calculator.	$Q_1 = 21.5$ and $Q_3 = 32$
2 Calculate the IQR.	$\text{IQR} = 32 - 21.5 = 10.5$
3 Calculate the lower and upper fences, $Q_1 - 1.5 \times \text{IQR}$ and $Q_3 + 1.5 \times \text{IQR}$	$Q_1 - 1.5 \times \text{IQR} = 21.5 - 1.5 \times 10.5 = 5.75$ $Q_3 + 1.5 \times \text{IQR} = 32 + 1.5 \times 10.5 = 47.75$
4 Check the potential outliers to see if they are less than $Q_1 - 1.5 \times \text{IQR}$ or greater than $Q_3 + 1.5 \times \text{IQR}$.	6 isn't less than 5.75, so it's *not* an outlier. 50 is greater than 47.75, so it *is* a possible outlier.

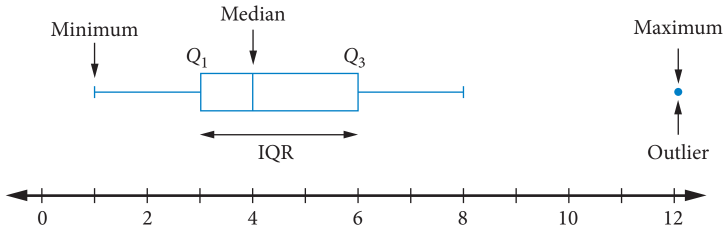
Boxplots

Boxplots, also known as **box-and-whisker plots**, display numerical data based on the five-number summary, IQR and outliers.

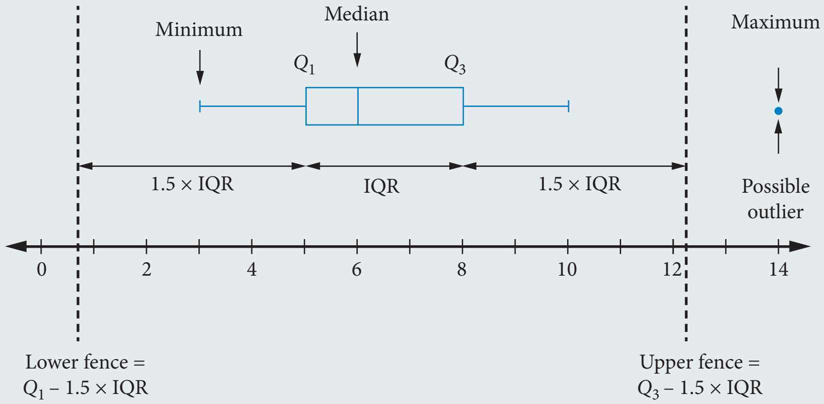
If there are no outliers, the **whiskers** show the minimum and maximum values:



When there are outliers, the whiskers show the lowest or highest values that are not outliers. Outliers are shown as points.



Boxplots provide the following information:



Boxplots can also be displayed vertically.

Using CAS Boxplots

The following information represents the weekly wages ($) of university students who work part-time.

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

Use a CAS/calculator to display a boxplot of this data.

TI-NSPIRE CAS

STEP 1

Using a New Document, enter data into the Lists & Spreadsheet page.

Name column A 'data' or something more specific (e.g. 'wages') then add the data into this list.

	A wages	B	C	D
=				
1	304			
2	285			
3	331			
4	270			
5	306			

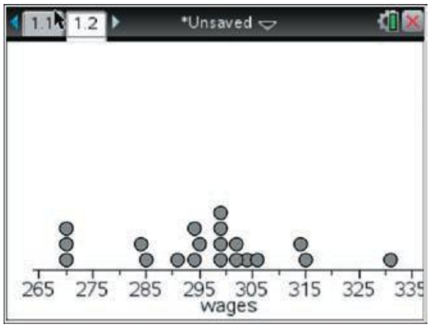
A20 302

STEP 2

Press CTRL Doc

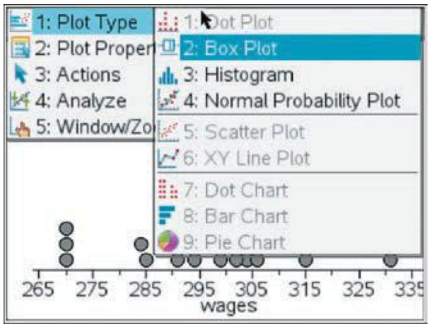
Then add a Data & Statistics page, then press `tab` and select 'data' (or whatever you've called it) as the *x* variable.

A dot plot will appear. This is the default display.



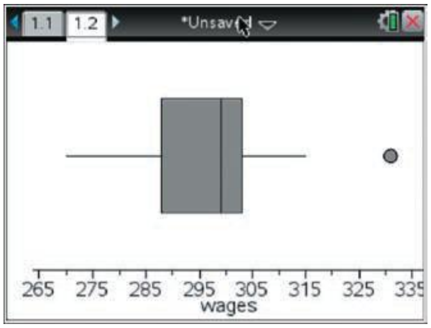
STEP 3

To change it to a boxplot, press menu 1: Plot Type, 2: Boxplot.



STEP 4

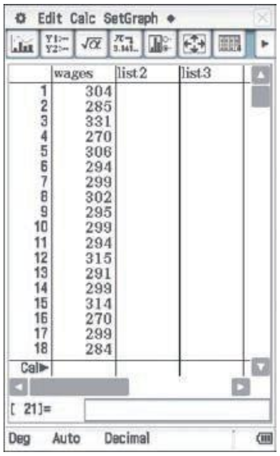
As you move the cursor over the graph the important values will appear. Dots indicate outliers.



CLASSPAD

STEP 1

Using the Statistics application, enter the data. Rename list1 as 'data' or something more specific (e.g. 'wages') by first tapping the list1 cell. To access the letter keyboard, press Keyboard and tap abc.



Edit Calc SetGraph ◆

	wages	list2	list3
1	304		
2	285		
3	331		
4	270		
5	306		
6	294		
7	299		
8	302		
9	295		
10	299		
11	294		
12	315		
13	291		
14	299		
15	314		
16	270		
17	299		
18	284		
Cal▶			

[21]=

Deg Auto Decimal

STEP 2

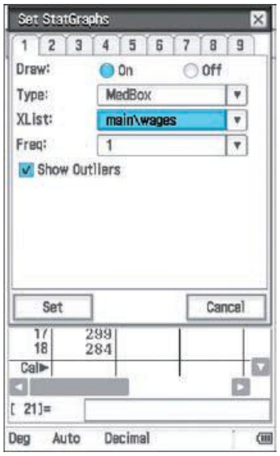
Tap **SetGraph**, making sure **StatGraph1** only is ticked and **Stat WindowAuto** under the ◆ menu is on.

Tap **Setting**.

For **Type** select MedBox.

For **Xlist** select main\wages (or whatever you renamed list1 as).

Make sure that the box for Show Outliers is checked.



Set StatGraphs

1 2 3 4 5 6 7 8 9

Draw: ● On ○ Off

Type: MedBox

XList: main\wages

Freq: 1

☑ Show Outliers

Set Cancel

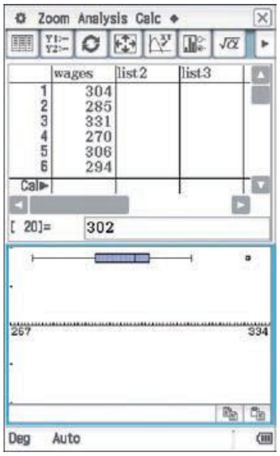
17	299		
18	284		
Cal▶			

[21]=

Deg Auto Decimal

STEP 3

Tap **Set** then tap .

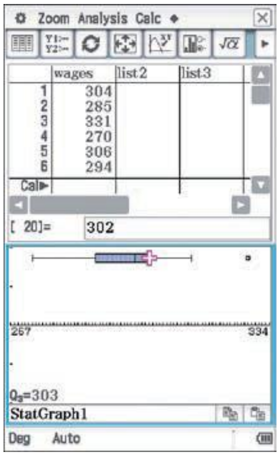


	wages	list2	list3
1	304		
2	285		
3	331		
4	270		
5	306		
6	294		

[20]= 302

STEP 4

Tap on the graph screen and then tap **Analysis** followed by **Trace**. Using the right and left arrow keys enables you to read minX, Q1, Med, Q3, maxX from the graph. Dots indicate outliers.



	wages	list2	list3
1	304		
2	285		
3	331		
4	270		
5	306		
6	294		

[20]= 302

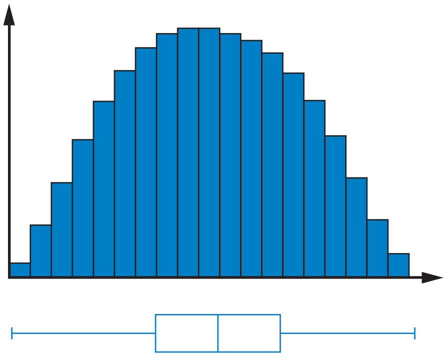
Q_3=303

Comparing boxplots and histograms

If you know what the histogram of a distribution looks like, you can often have some idea of what the boxplot looks like.

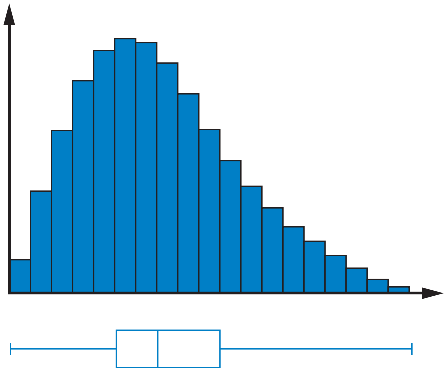
Symmetric distributions

- Median approximately in the middle
- Boxes and whiskers about the same length



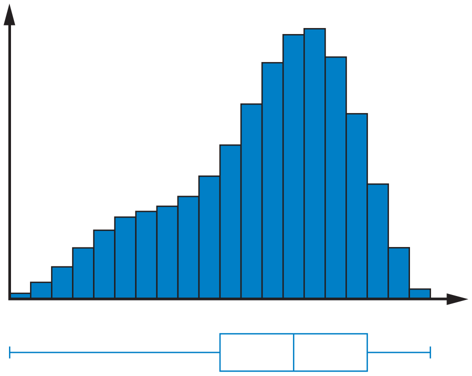
Positively skewed distributions

- Median usually to the left of centre
- Long right-hand whisker



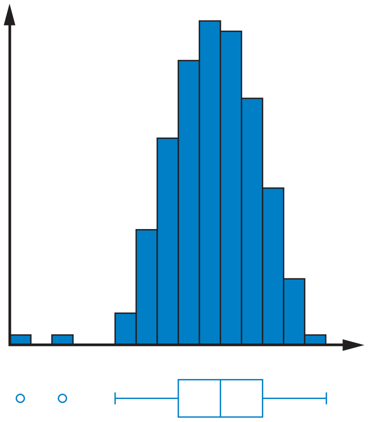
Negatively skewed distributions

- Median usually to the right of centre
- Long left-hand whisker



Distributions with outliers

- Boxplot matches the histogram, ignoring outliers
- Outliers shown by dots



EXAM PREP 1.4

Boxplots

Prep 1 WORKED EXAMPLE 2 USING CAS: FIVE-NUMBER SUMMARY

For the following test scores 73, 65, 54, 90, 74, 51, 61, 88, 47, 92, 71, 66

a find the three quartiles by hand and show how they divide the data into quarters

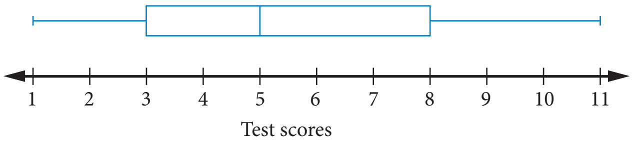
b verify your answers by using a CAS/calculator by finding the five-number summary.

Prep 2

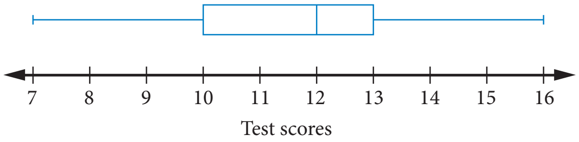
For the following boxplots, state

i the five-number summary **ii** the values between which the middle 50% of the data lies.

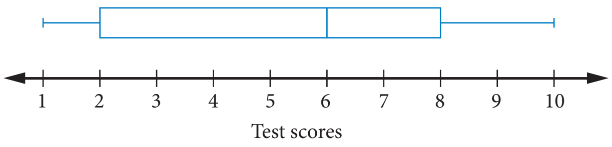
a



b



c



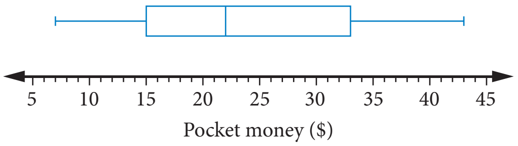
Prep 3 WORKED EXAMPLE 3 USING CAS: BOXPLOTS

For each of the following data sets use a CAS/calculator to construct a boxplot and do a calculation to confirm possible outliers.

a	45	50	44	54	37	44	15	50	41	52
	38	26	37	42	48	39	46	49	43	64
b	104	88	110	99	40	156	96	97	86	
	105	79	77	89	93	95	110	99	115	
c	15	22	25	25	25	27	17	27		
	23	20	24	26	29	27	21	25		

Prep 4

This boxplot represents the amount of pocket money in dollars earned by a sample of 48 children.



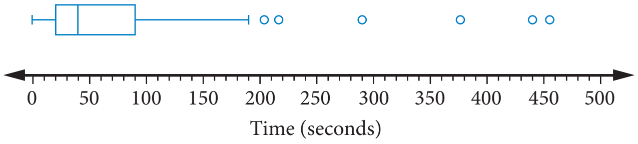
a What percentage of children earned less than $15?

b How many children earned $22 or more of pocket money?

Boxplots

Use the following information to answer Questions 1–3.

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



data source: www.stars.ac.uk

Question 1

The longest time, in seconds, spent moving along this aisle is closest to

A 40 **B** 60 **C** 190 **D** 450 **E** 500

[VCAA 2008 1CQ1]

Question 2

The shape of the distribution is best described as

A symmetric.

B negatively skewed.

C negatively skewed with outliers.

D positively skewed.

E positively skewed with outliers.

[VCAA 2008 1CQ2]

Question 3

The number of customers who spent more than 90 seconds moving along this aisle is closest to

A 7 **B** 20 **C** 26 **D** 75 **E** 79

[VCAA 2008 1CQ3]

Question 4

The following data was recorded from measurements made on 12 men.

Age (years)	Mass (kg)	Waist (cm)
26	84	84
29	72	74
32	67	89
32	59	75
34	97	106
37	112	114
39	67	80
40	91	101
41	98	101
43	89	94
45	117	126
51	62	82

For these men, the median age (M) and the interquartile range (IQR), in years, are respectively

A $M = 37$ and IQR $= 10$

B $M = 37.4$ and IQR $= 7.2$

C $M = 37.4$ and IQR $= 6.9$

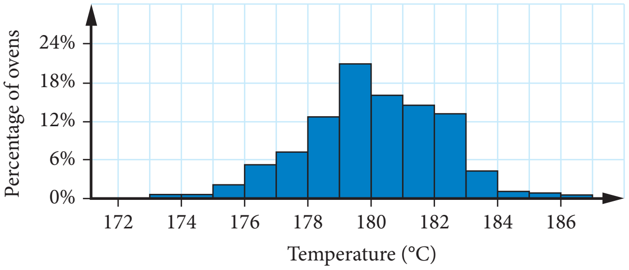
D $M = 38$ and IQR $= 10$

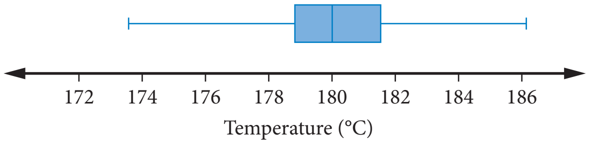
E $M = 38$ and IQR $= 25$

[VCAA 2002 1CQ7]

Use the following information to answer Questions 5 & 6.

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





Question 5

The interquartile range for temperature is closest to

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

[VCAA 2010 1CQ2]

Question 6

A total of 300 ovens were tested and their temperatures were recorded.

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

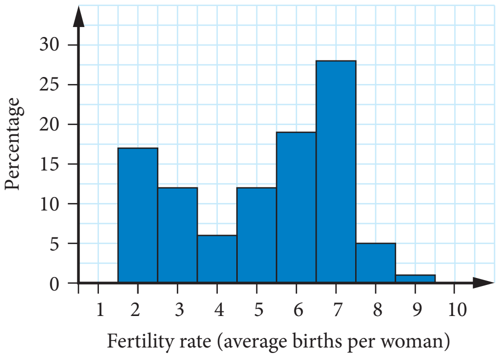
A 40 **B** 50 **C** 70 **D** 110 **E** 150

[VCAA 2010 1CQ1]

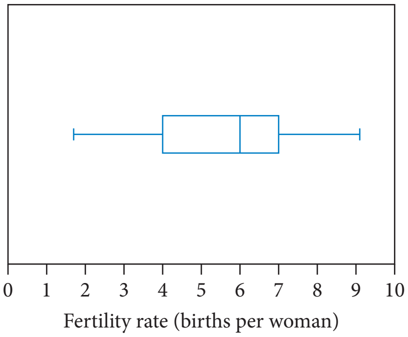
Question 7

The percentage histogram shows the distribution of the fertility rates (in average births per woman) for 173 countries in 1975.

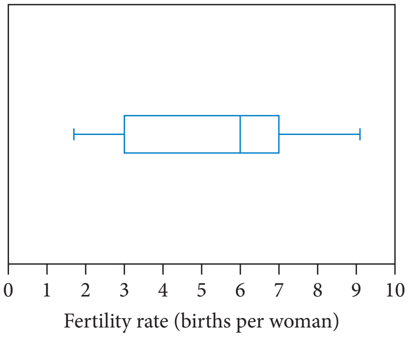
Which one of the boxplots below could best be used to represent the same fertility rate data as displayed in the percentage histogram?



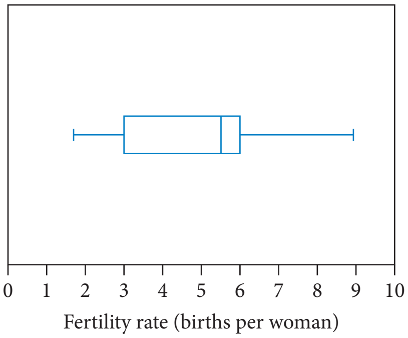
A



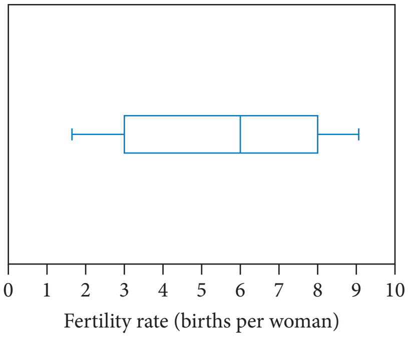
B



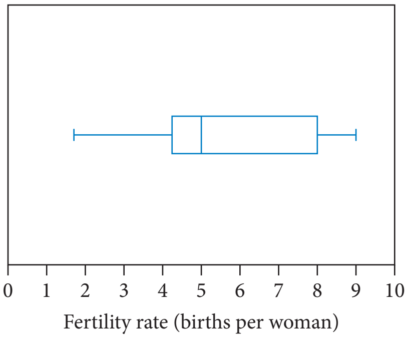
C



D



E



[VCAA 2009 1CQ6]

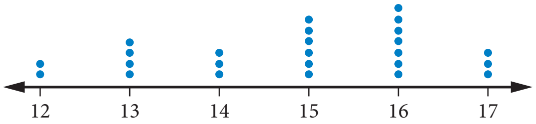
1.5

Dot plots and stem plots

Dot plots

Dot plots are the simplest way to display numerical data. They are best used for a maximum of 50 data values and when the data values are not too spread out.

Alamy/Megapress



Ages of students at a school function

Looking at the dot plot of the ages of students at a school function, we can easily read the following from the plot:

The minimum data value = 12

The maximum data value = 17

The range = largest value – smallest value = 17 – 12 = 5

The **mode**, or the most frequently occurring value in the data set, is 16 years.

To get more information from a dot plot, it's necessary to count dots:

To calculate the median, you need to

1. count the total number of dots (25 in this case)
2. then count from the left to find where the middle dot occurs (in this case it's the 13th dot because it has 12 dots before it and 12 dots after it)
3. write down the value where this occurs. (In this example the median is 15 years.)

The **lower quartile** is the median of the lower 12 values, so we need to find the 6th and 7th value, then add them together and divide by 2. $Q_1 = \frac{13+14}{2} = 13.5$.

The **upper quartile** is the median of the upper 12 values, so we need to find the 6th and 7th value from the right, then add them together and divide by 2. $Q_3 = \frac{16+16}{2} = 16$.

The **interquartile range**, $\text{IQR} = Q_3 - Q_1 = 16 - 13.5 = 2.5$.

Exam hack

To decide whether a dot plot is symmetric, positively skewed or negatively skewed, look at the shape the dots are making and try to picture it as a histogram.

Stem plots

WS Stem-and-leaf plots

Stem plots, also known as **stem-and-leaf plots**, are an alternative to histograms whose main advantage is that the actual data values appear. A stem plot is best used with up to a maximum of 50 data values, otherwise it becomes unwieldy. We would normally use **ordered stem plots** where the leaves are ordered from smallest to largest. A key is always required. The stem plot here represents test scores out of 60 for a class of 33 students.

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

Key: 2|0 = 20

Because you can see all the data values, it is relatively straightforward to find the five-number summary and other information directly from the plot.

Looking at the example:

- The minimum data value = 1
- The maximum data value = 57
- There are 33 ordered values, so the median will be the 17th value (i.e. it will have 16 values either side). Counting up to the 17th value from the start, the median = 35.
- The lower quartile is the median of the lower 16 values, so we need to find the 8th and 9th value, then add them together and divide by 2. $Q_1 = \frac{23+23}{2} = 23$.
- The upper quartile is the median of the upper 16 values, so we need to find the 8th and 9th value from the bottom right corner of the plot, then add them together and divide by 2. $Q_3 = \frac{48+49}{2} = 48.5$.
- The interquartile range $\text{IQR} = Q_3 - Q_1 = 48.5 - 23 = 25.5$.
- The range = largest value − smallest value = 57 − 1 = 56
- The 1 score may be an outlier. Check using the lower fence $Q_1 - 1.5 \times \text{IQR} = 23 - 1.5 \times 25.5 = -15.25$. Since 1 isn't less than −15.25, it's not an outlier. (Note that negative test scores would be impossible anyway.)
- There are two modes, 21 and 49, so we say this data set is bi-modal.

Exam hack

To decide whether a stem plot is symmetric, positively skewed or negatively skewed, rotate the page so that the stem forms the horizontal axis, look at the shape the data values are making, and try to picture it as a histogram.

Which display do you use?

Often there is more than one suitable display, but here are some guidelines on which statistical display is the best one to use:

Display	Type of data	Guidelines
Bar chart	Categorical data	Categories can be represented on the horizontal or vertical axis
Segmented bar chart	Categorical data	Should have no more than five segments
Histogram	Numerical data	Best if data has been grouped into between 5 and 15 intervals
Boxplot	Numerical data	Best if you want to read the five-number summary easily
Dot plot	Numerical data	Best used with a maximum of 50 data values and when the data values are not too spread out
Stem plot	Numerical data	Best used with a maximum of 50 data values

EXAM PREP 1.5

Dot plots and stem plots

Prep 1

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

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

Key: 4|5 = 45

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

b How many matches were played in the season?

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

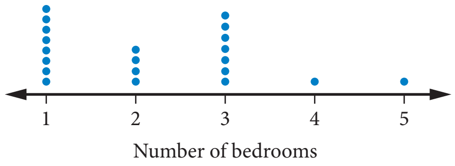
d For what percentage of matches did the Blues score below 56 points?

EXAM PRACTICE 1.5

Dot plots and stem plots

Use the following information to answer Questions 1 & 2.

The dot plot shows the distribution of the number of bedrooms in each of 21 apartments advertised for sale in a new high-rise apartment block.



Question 1

The mode of this distribution is

A 1 **B** 2 **C** 3 **D** 7 **E** 8

[VCAA 2007 1CQ1]

Question 2

The median of this distribution is

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

[VCAA 2007 1CQ2]

Use the following information to answer Questions 3 & 4.

The ordered stem plot shows the percentage of homes connected to broadband Internet for 24 countries in 2007.

Key: 1|6 = 16%

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

Question 3

The number of these countries with more than 22% of homes connected to broadband Internet in 2007 is

A 4 **B** 5 **C** 19 **D** 20 **E** 22

[VCAA 2013 1CQ1]

Question 4

Which one of the following statements relating to the data in the ordered stem plot is **not** true?

A The minimum is 16%.

B The median is 30%.

C The first quartile is 23.5%.

D The third quartile is 32%.

E The maximum is 38%.

[VCAA 2013 1CQ2]

Use the following information to answer Questions 5 & 6.

The marks obtained by students who sat for a test are displayed as an ordered stem plot as shown.

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

Question 5

The number of students who sat the test is

A 25 **B** 26 **C** 27

D 32 **E** 50

[VCAA 2004 1CQ1]

Question 6

The interquartile range of these test marks is closest to

A 9 **B** 13 **C** 30 **D** 36 **E** 41

[VCAA 2004 1CQ2]

Question 7

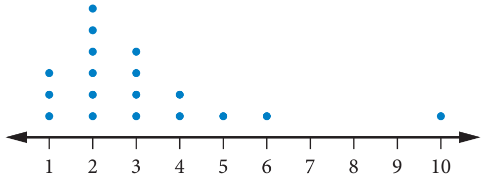
Which of these displays could be described as positively skewed with a possible outlier?

A

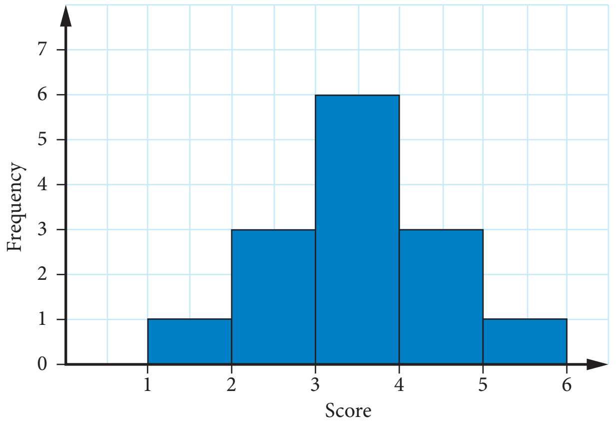
Stem	Leaf
7	1 5 6 8 9 9
8	0 0 2 3 4 6 8 9
9	0 2 5 7 9
10	2 3 5 6 7 7 8

Key: 7|9 = 79

B



C

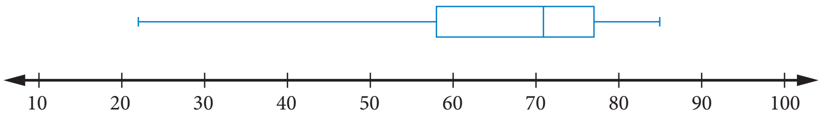


D

Stem	Leaf
12	3
13	
14	
15	4 7 9 9 9
16	0 2 3 7 8 8
17	0 0 2 4 4 5 5 7 7 7 9
18	3 9
19	1 1 2

Key: 12|3 = 123

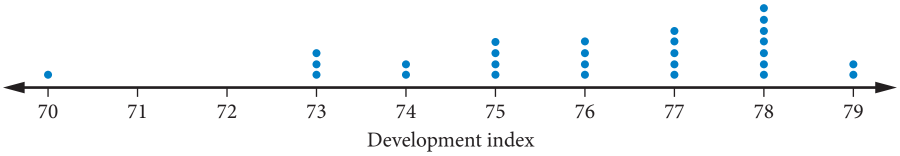
E



Question 8

The development index for each country is a whole number between 0 and 100.

The dot plot below displays the values of the development index for each of the 28 countries that has a high development index.



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

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

[VCAA 2013 2CQ2]

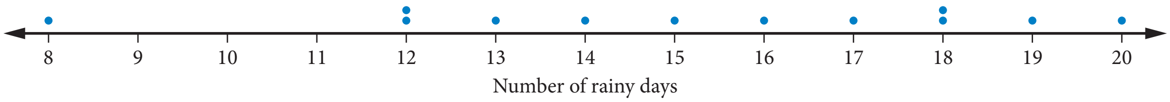
Question 9

Table 1 shows the number of rainy days recorded in a high rainfall area for each month during 2008.

Table 1

Month	Number of rainy days
January	12
February	8
March	12
April	14
May	18
June	18
July	20
August	19
September	17
October	16
November	15
December	13

The dot plot below displays the distribution of the number of rainy days for the 12 months of 2008.



a Copy the dot plot and **circle** the dot that represents the number of rainy days in April 2008. 1 mark

b For the year 2008, determine

- **i** the median number of rainy days per month 1 mark
- **ii** the percentage of months that have more than 10 rainy days. Write your answer correct to the nearest per cent. 1 mark

[VCAA 2009 2CQ1]

Question 10

The stemplot in Figure 1 shows the distribution of the average age, in years, at which women first marry in 17 countries.

Average age, in years, of women at first marriage

Key: 27|3 represents 27.3 years

24	
25	0
26	6
27	1 1 3 4 7
28	2 2 2 3 3 6
29	1 1
30	1 4
31	

Figure 1

a For these countries, determine

- **i** the lowest average age of women at first marriage 1 mark
- **ii** the median average age of women at first marriage. 1 mark

The stem plot in Figure 2 shows the distribution of the average age, in years, at which men first marry in 17 countries.

Average age, in years, of men at first marriage

Key: 32|5 represents 32.5 years

25	
26	0
27	
28	9
29	0 9 9
30	0 0 3 5 6 7 9
31	0 0 2
32	5 9
33	

Figure 2

b For these countries, determine the interquartile range (IQR) for the average age of men at first marriage. 1 mark

c If the data values displayed in Figure 2 were used to construct a boxplot with outliers, then the country for which the average age of men at first marriage is 26.0 years would be shown as an outlier.

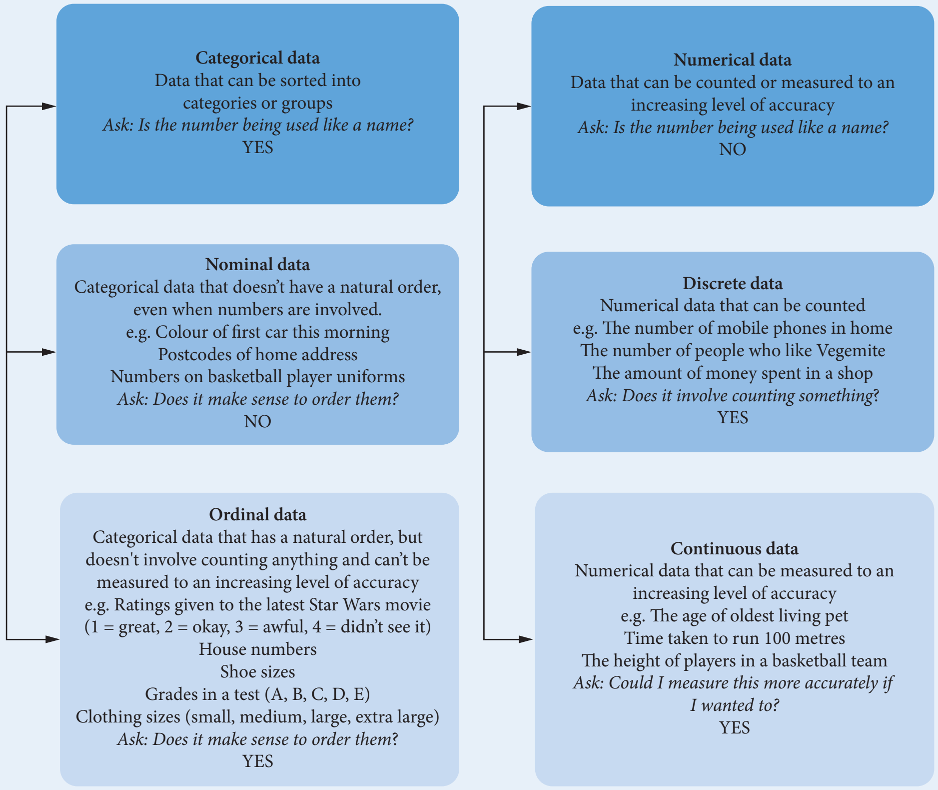
Explain why this is so. Show an appropriate calculation to support your explanation. 2 marks

[VCAA 2011 2CQ1]

Review of data distributions

Practice quiz

Types of data



Categorical data
Data that can be sorted into categories or groups
Ask: Is the number being used like a name?
YES

Nominal data
Categorical data that doesn't have a natural order, even when numbers are involved.
e.g. Colour of first car this morning
Postcodes of home address
Numbers on basketball player uniforms
Ask: Does it make sense to order them?
NO

Ordinal data
Categorical data that has a natural order, but doesn't involve counting anything and can't be measured to an increasing level of accuracy
e.g. Ratings given to the latest Star Wars movie (1 = great, 2 = okay, 3 = awful, 4 = didn't see it)
House numbers
Shoe sizes
Grades in a test (A, B, C, D, E)
Clothing sizes (small, medium, large, extra large)
Ask: Does it make sense to order them?
YES

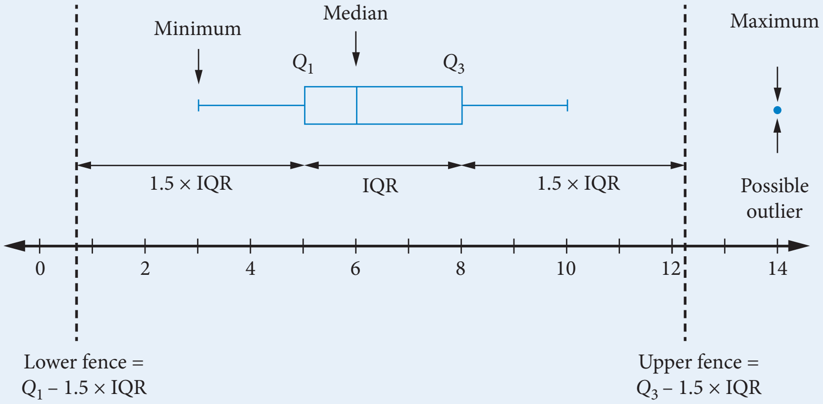
Numerical data
Data that can be counted or measured to an increasing level of accuracy
Ask: Is the number being used like a name?
NO

Discrete data
Numerical data that can be counted
e.g. The number of mobile phones in home
The number of people who like Vegemite
The amount of money spent in a shop
Ask: Does it involve counting something?
YES

Continuous data
Numerical data that can be measured to an increasing level of accuracy
e.g. The age of oldest living pet
Time taken to run 100 metres
The height of players in a basketball team
Ask: Could I measure this more accurately if I wanted to?
YES

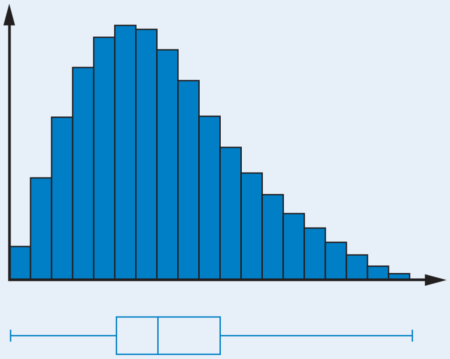
Tables, charts and plots

- **Frequency tables** list data values or categories in one column and the corresponding frequencies in a frequency column.
- **Bar charts** have categories on the horizontal or vertical axis, with their corresponding frequency on the other axis.
- In a **segmented bar chart**, the bars are stacked on one another to give one bar with several segments. A legend needs to be included to explain what the segments represent.
- A **histogram** is a graphical way of displaying numerical data from a frequency table. It is effective when dealing with data that has been grouped into a small number of intervals.
  - Histograms for continuous numerical data and grouped discrete numerical data have bars that sit in between the interval values on the horizontal axis.
  - Histograms for discrete numerical data that hasn't been grouped into intervals have bars starting and ending halfway between scale marks on the horizontal axis.
- **Boxplots** display numerical data based on the five-number summary, IQR and outliers.

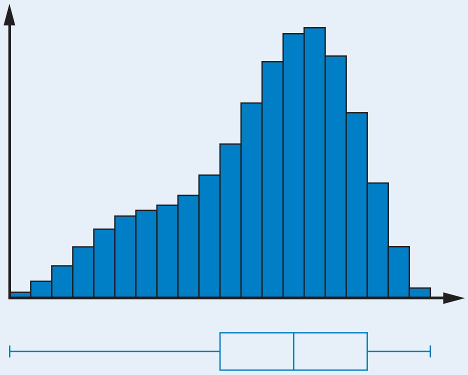


- **Dot plots** are the simplest way to display numerical data. They are best used for a maximum of 50 data values and when the data values are not too spread out.
- **Stem plots** are an alternative to histograms whose main advantage is that the actual data values appear. A key is always required.

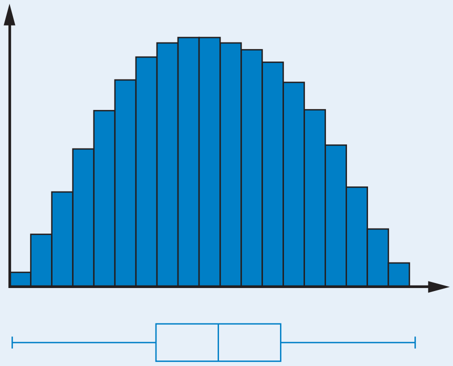
Distribution shapes



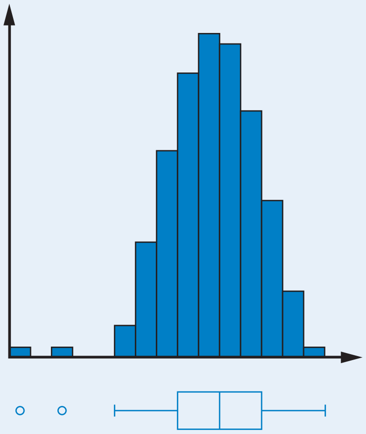
Positively skewed distribution



Negatively skewed distribution



Symmetric distribution



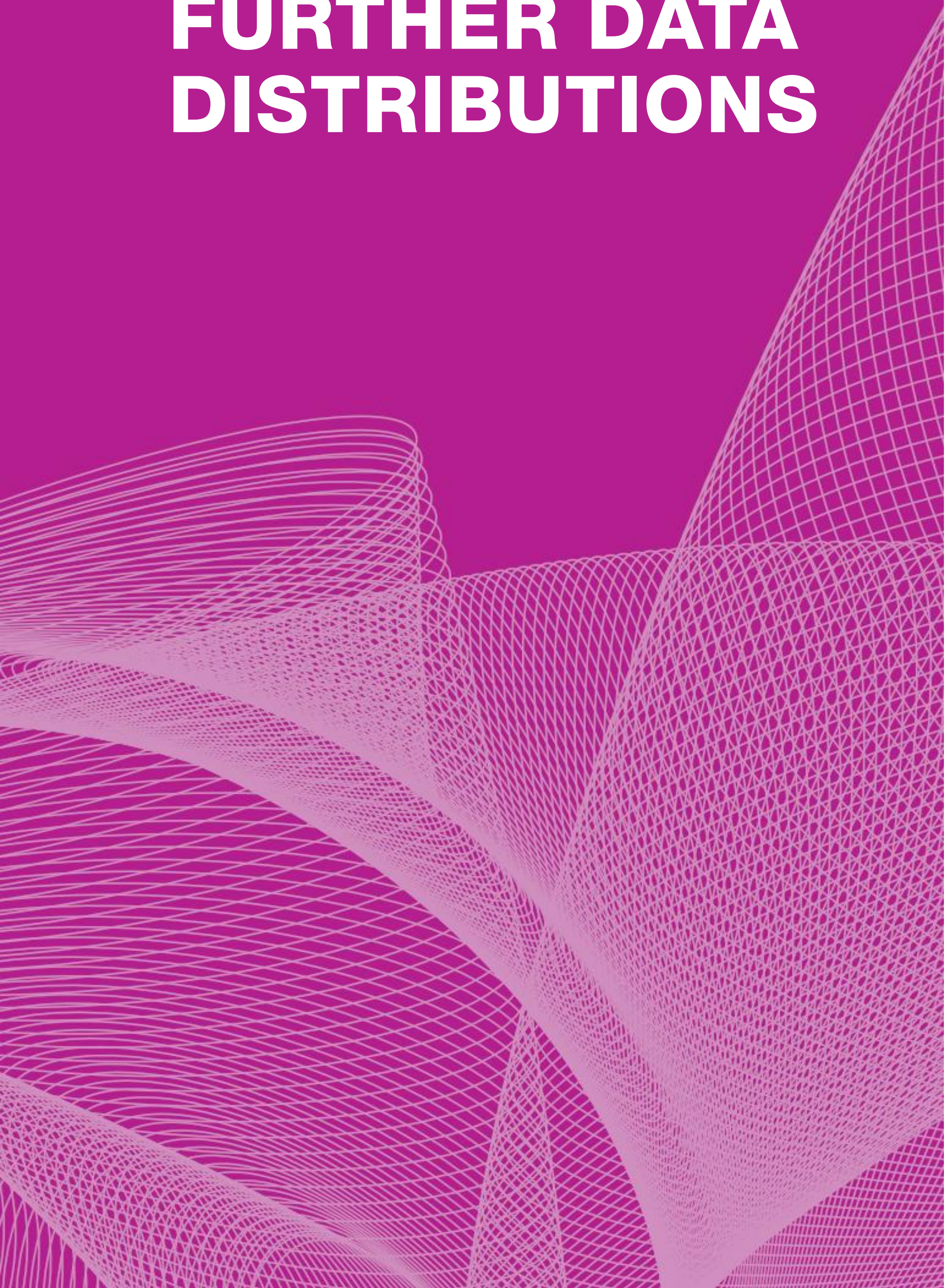
Distribution with outliers

Statistical values

- The **range** is a measure of the spread of the data.
  Range = largest value − smallest value
- The **median** is a measure of the centre of the data.
  - The median is the middle value when the data is ordered from smallest to largest.
  - When there are two middle values, we add them and divide by 2 to find the median.
  - The median on a histogram occurs at the vertical line that splits the histogram in half with equal areas on either side.
- **Quartiles** are the three points that divide a set of data into quarters.
  - The first or lower quartile has 25% of the data below it.
  - The second quartile (which is the same as the median) has 50% of the data below it.
  - The third or upper quartile has 75% of the data below it.
- The **five-number summary** consists of
  1. The minimum data value
  2. Q_1 (lower quartile)
  3. The median
  4. Q_3 (upper quartile)
  5. The maximum data value
- The **interquartile range** (or IQR) is the measure of the spread of the middle 50% of the data values. $\text{IQR} = Q_3 - Q_1$
- A data value is a possible **outlier** if it is either
  - less than $Q_1 - 1.5 \times \text{IQR}$ or
  - greater than $Q_3 + 1.5 \times \text{IQR}$

CHAPTER 2

FURTHER DATA DISTRIBUTIONS

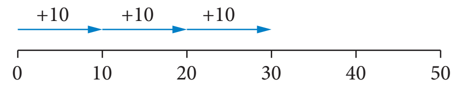


2.1 Log scales

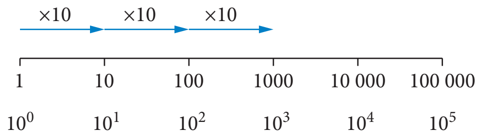
Linear and log scales

Amana Images/U Aung/Xinhua Press

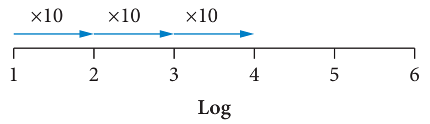
All the plots and graphs we've used so far have had **linear scales**. On a linear scale, you *add* the same number to get from one scale mark to the next. In this example the linear scale involves adding 10 each time.



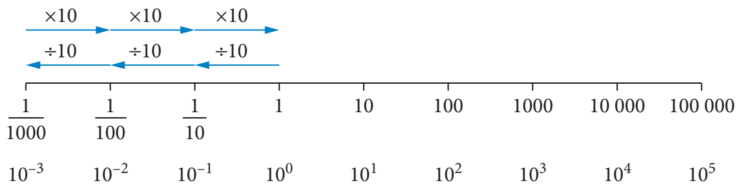
There are some situations where it is better to use a **log scale**. On a log (short for logarithmic) scale, you *multiply* the same number to get from one scale mark to the next. In this example the log scale involves multiplying by 10 each time.



This means log base 10 scales can be used to show very large values effectively. Log scales are sometimes written with powers of 10 (which are also called logarithms or logs) rather than the number in full.



By extending the scale to the left, log scales can also be used to show very small values:



Notice that although a log scale can get extremely close to zero, it can never actually have the value zero.

A log scale doesn't have to be of base 10, just as a linear scale doesn't have to add 10 each time. However, log scales of base 10 are the most common log scales and we will only be looking at these.

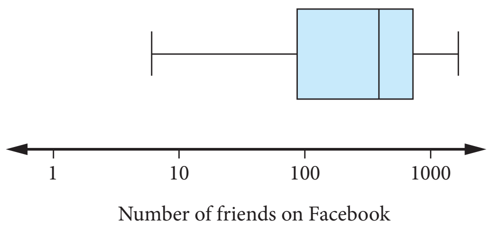
Why use a log scale?

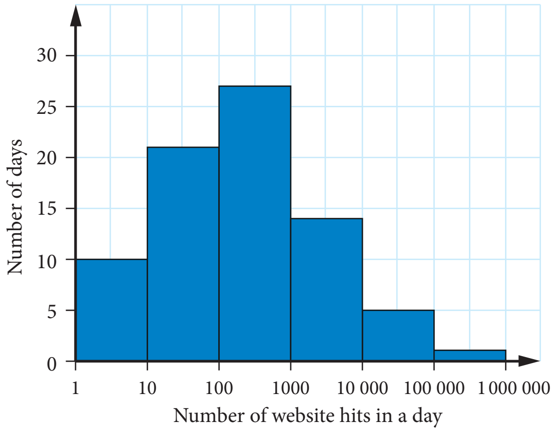
A logarithmic scale is used when the range of the data is very large. It would be impossible, for example, to find an effective linear scale to plot the following data:

30, 5764, 12, 468 832, 152, 1 004 453

A linear scale for this data would either be impractically long or would not provide any useful information.

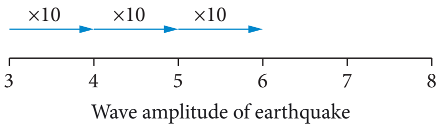
Histograms and boxplots with large ranges can benefit from having a log base 10 scale:





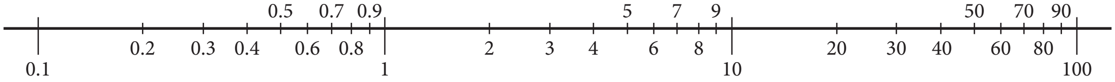
The Richter magnitude scale

An example of a log base 10 scale is the **Richter magnitude scale**, which involves measuring the amplitude of the largest wave of an earthquake. On the Richter scale the wave amplitude for a level 5 earthquake is 10 times greater than for a level 4 earthquake, and 100 times greater than for a level 3 earthquake.



Locating values on a log base 10 scale

When locating values on a log scale, note that the values aren't evenly spaced between the intervals and that they bunch up near the end of each interval:



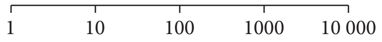
You can use the log function on your CAS/calculator to locate values on a log base 10 scale.

Exam hack

Watch out for log scales that are indicated just by the powers (i.e. 1, 2, 3, etc. instead of $10^1, 10^2, 10^3$ etc.).

Worked example 1

Copy the log base 10 scale below and mark where 50 is.



Working

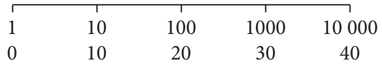
1 Find log 50 using your CAS/calculator.

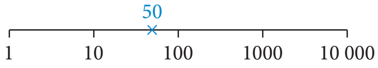
$\log 50 \approx 1.7$

2 Multiply the value by 10.

$1.7 \times 10 = 17$

3 Think of the scale as linear with 10 added each time and mark the value on the scale.





EXAM PREP 2.1

Log scales

Prep 1

For each of the following, state whether it would be better to display the data involved using a log base 10 scale or a linear scale.

a a timeline showing the end of the Mesozoic period (the time of the dinosaurs), the fall of the Roman Empire, the end of the Second World War, and the first person to walk on the Moon

b the growth of Internet usage from when it was invented

c the number of people to attend a particular cinema over a month

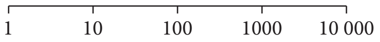
d prices of cars in Australia

e the increase in mobile phone subscriptions in Australia from the first year they came on the market to the present

f the percentage of the Australian population from one year to the next who get a cold during winter

Prep 2 WORKED EXAMPLE 1

Copy the log base 10 scale below and mark where the following values are.



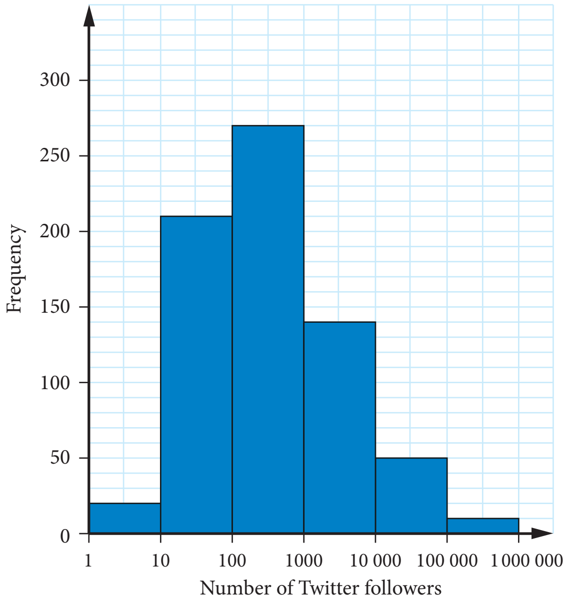
a 20 **b** 150 **c** 2000

EXAM PRACTICE 2.1

Log scales

Use the following information to answer Questions 1–4.

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



Question 1

Which one of these statements is true about this histogram?

A It has a vertical log base 10 scale.

B It has a horizontal linear scale.

C It has a horizontal log base 10 scale and a vertical log base 10 scale.

D It has a horizontal log base 10 scale and a vertical linear scale.

E It has two linear scales.

Question 2

The number of the blogger's followers who have between 1000 and 10000 followers is closest to

A 1000 **B** 480 **C** 270 **D** 210 **E** 140

Question 3

Which one of these statements is *not* true?

A The number of the blogger's followers who have more than 10 000 but fewer than 100 000 followers is 50.

B The number of the blogger's followers who have fewer than 100 followers is 210.

C The number of the blogger's followers who have more than 10 000 followers is 60.

D None of the blogger's followers has more than 1 000 000 followers.

E The number of the blogger's followers with fewer than 10 followers is twice the number of those with more than 100 000 followers.

Question 4

All of the following except one are reasons why a log base 10 scale is a better choice than a linear scale for the histogram of Twitter followers. Which of the options isn't a reason?

A The range of the data is too large.

B There is a huge variability in the number of followers.

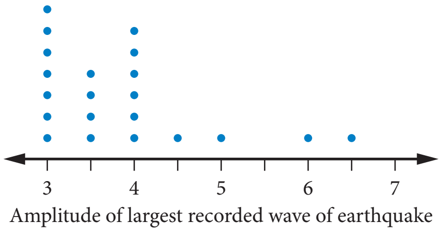
C A linear scale of intervals of 100 000 would provide little information.

D The median is too low.

E A linear scale of intervals of 10 would be impossibly long.

Use the following information to answer Questions 5–7.

The amplitude of the largest recorded wave of 21 earthquakes in a certain region has been recorded in the dot plot with a log base 10 scale.



Question 5

What percentage of the earthquakes have wave amplitudes less than 5?

A 18% **B** 19% **C** 85% **D** 86% **E** 90%

Question 6

What are the median and range of this distribution?

A Median = 3.5 and Range = 3.5

B Median = 5 and Range = 3.5

C Median = 5 and Range = 4

D Median = 4 and Range = 3.5

E Median = 3.5 and Range = 4

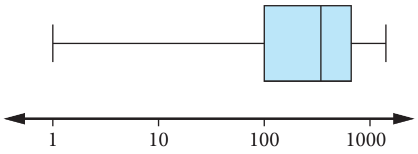
Question 7

How many times greater is the wave amplitude of the earthquake recorded as 6.5 than the one recorded as 4.5?

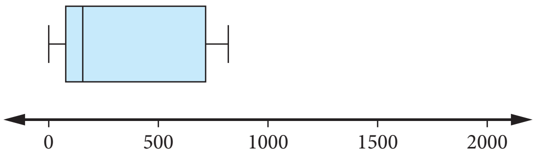
A 200 **B** 100 **C** 2 **D** 4 **E** 100 000

Question 8

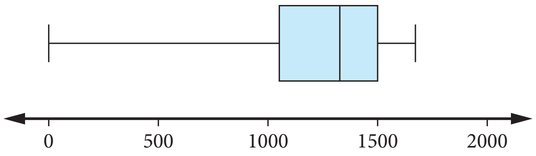
The boxplot to the right has a log base 10 scale. Which one of the following boxplots would best match its linear scale equivalent?



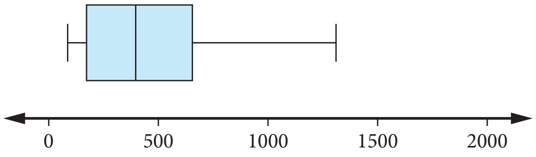
A



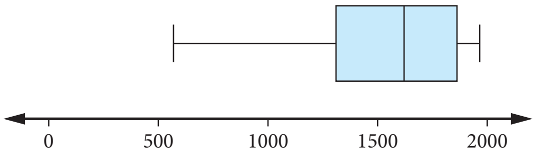
B



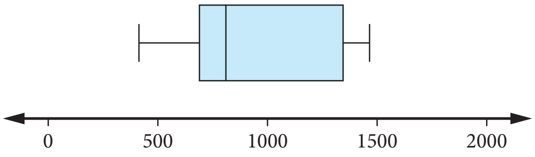
C



D



E



2.2 The sample mean and sample standard deviation

Amana Images/Karl Weatherly

The sample mean

The **sample mean** is what is often referred to in everyday life as the average. The symbol we will be using for the sample mean of a set of data is $\bar{x}$ (called 'x bar').

The method used for calculating the sample mean depends on whether the data is a list of values or organised in a frequency table.

The sample mean for a list of data values

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

$$= \frac{\Sigma x}{n}$$

where Σ means 'sum of'

The sample mean for data in a frequency table:

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

$$= \frac{\Sigma xf}{\Sigma f}$$

9780170371100

The median vs the sample mean

Both the median and the sample mean are measures of the centre of a distribution. While the median is the mid-point of a distribution, the sample mean is the balance-point of the distribution.

When looking at the median vs the sample mean, be aware of the following:

- For symmetric distributions, the median = the sample mean.
- For distributions that are approximately symmetric, the median and the sample mean will be approximately equal.
- The sample mean is greater than the median for positively skewed distributions.
- The sample mean is less than the median for negatively skewed distributions.
- Outliers don't generally affect the median, but they can significantly affect the mean.

Exam hack

If you add the same number to every value in the data set, the sample mean will increase by that number.

Exam hack

How do you choose whether to use the median or sample mean as the measure of the centre of a distribution?

Approximately symmetric distributions with no outliers	Sample mean or median
Approximately symmetric distributions with outliers	Median
Skewed distributions	Median

The sample standard deviation

The **sample standard deviation**, like the range and the interquartile range, is a measure of the spread of data. Smaller values of the standard deviation indicate less spread. Larger values indicate more spread. While the interquartile range measures the spread of data around the median, the sample standard deviation measures the spread of data around the sample mean.

The formula for the sample standard deviation is

$$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$$

although calculations are done quickly through a CAS/calculator.

Exam hack

A quick way of estimating the sample standard deviation (and checking CAS/calculator results) for small sets of data is:

$$s \approx \frac{\text{range}}{4}$$

Exam hack

If you add the same number to every value in the data set, the sample standard deviation (and other measures of spread) will stay the same.

Using CAS Sample mean and sample standard deviation

To find the sample mean $\bar{x}$ and the sample standard deviation s for a list of data values, follow the same steps as for 'Using CAS: Five-number summary' (Chapter 1 p. 21) and locate $\bar{x}$ and s_x from the choices at the end.

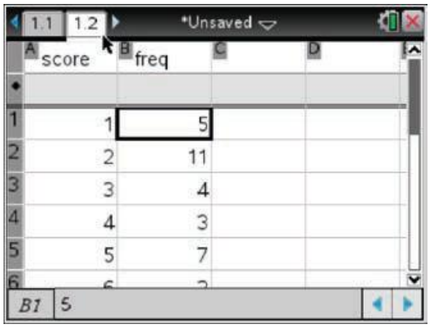
To find the sample mean $\bar{x}$ and the sample standard deviation s (and any of the values for the five-number summary) for data in a frequency table, follow the steps for this frequency table example:

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

TI-NSPIRE CAS

STEP 1

Open a New Document with a Lists & Spreadsheet page. Name column A 'score' and column B 'freq', then enter the values into the list called 'score' and the corresponding frequencies into the list called 'freq'.



score	freq
1	5
2	11
3	4
4	3
5	7

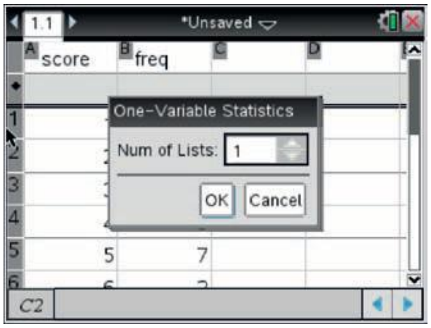
B1 5

STEP 2

Press menu 4: Statistics then 1: Stat Calculations then 1: One-Variable Statistics.

In the pop-up screen that appears, set the number of lists to 1 and press enter.

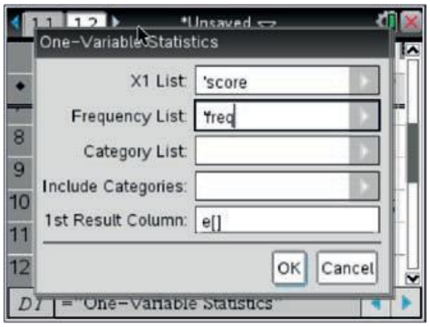
Although you have 2 columns in the table, only one list contains the values of the variable.



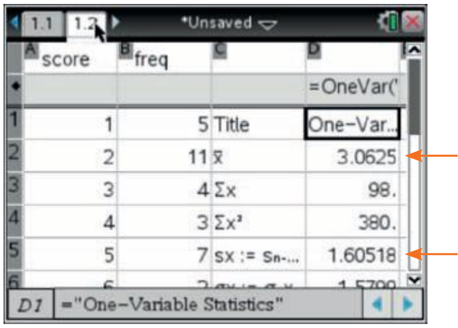
Num of Lists: 1

STEP 3

Select variable name 'score' for X1 List and select 'freq' for Frequency List, then press tab, down to OK and press enter.



X1 List: 'score
Frequency List: 'freq
Category List:
Include Categories:
1st Result Column: e[]



score	freq		=OneVar('
1	5	Title	One-Var...
2	11	$\bar{x}$	3.0625
3	4	Σx	98.
4	3	Σx^2	380.
5	7	sx := Sn-...	1.60518

D1 ="One-Variable Statistics"

mean $\bar{x}$

standard deviation (s)

CLASSPAD

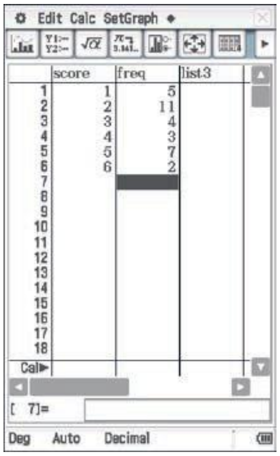
STEP 1

Use the Statistics application. Rename list1 as 'score' by first tapping the list1 cell. To access the letter keyboard, press `Keyboard` and tap `abc`.

Follow a similar process to rename the list2 column as 'freq' and enter values.

If the frequency table has class intervals, then the midpoint of each interval is entered into list 1.

Tap **Calc** then **One-Variable**.



	score	freq	list3
1	1	5	
2	2	11	
3	3	4	
4	4	3	
5	5	7	
6	6	2	
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			

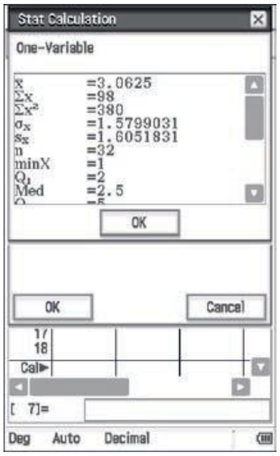
STEP 2

In the pop-up window that appears, set the **XList** to main\score and **Freq** to main\freq then tap `OK`.

One-Variable

XList: main\score

Freq: main\freq



One-Variable

```
x̄       =3.0625
Σx      =98
Σx²     =380
σx      =1.5799031
sx      =1.6051831
n       =32
minX    =1
Q1      =2
Med     =2.5
```

EXAM PREP 2.2

The sample mean and sample standard deviation

Prep 1 USING CAS: SAMPLE MEAN AND SAMPLE STANDARD DEVIATION

Find the sample mean and sample standard deviation of each of the following data sets. Answer correct to 2 decimal places.

a 26 27 23 24 27 25 23

19 25 21 22 20 27 21

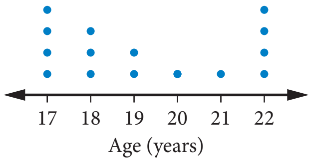
b

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

c

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

d



EXAM PRACTICE 2.2

The sample mean and sample standard deviation

Question 1

The total mass of nine oranges is 1.53 kg.

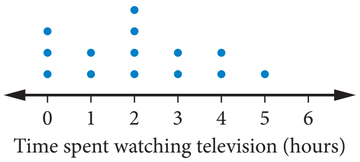
Using this information, the mean mass of an orange would be calculated to be closest to

A 115 g **B** 138 g **C** 153 g **D** 162 g **E** 170 g

[VCAA 2012 1CQ3]

Question 2

A sample of 14 people were asked to indicate the time (in hours) they had spent watching television on the previous night. The results are displayed in the dot plot.



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

A $\bar{x} = 2.0$ $s = 1.5$

B $\bar{x} = 2.1$ $s = 1.5$

C $\bar{x} = 2.1$ $s = 1.6$

D $\bar{x} = 2.6$ $s = 1.2$

E $\bar{x} = 2.6$ $s = 1.3$

[VCAA 2008 1CQ5]

Use the following information to answer Questions 3 & 4.

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

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

Question 3

For this sample of households, the percentage of households with **at least** one DVD player is

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

[VCAA 2004 1CQ4]

Question 4

For this sample of households, the mean number of DVD players in these 20 households is

A 0.75 **B** 1.00 **C** 1.15 **D** 1.64 **E** 2.00

[VCAA 2004 1CQ5]

Use the following information to answer Questions 5 & 6.

The mean mass of twelve people is 72 kg; the standard deviation of the masses of these twelve people is 5 kg.

Question 5

The total mass of the twelve people is

A 77 kg **B** 360 kg **C** 864 kg **D** 924 kg **E** 4320 kg

[VCAA 2003 1CQ4]

Question 6

These twelve people are about to go on a rafting adventure. Before boarding the raft, they are all required to put on a life-saving vest that weighs 2 kg. The effective mass of each person is now their mass plus the mass of the life-saving vest.

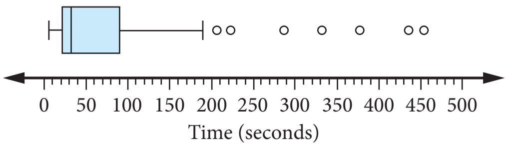
The effective masses of the twelve people have

A a mean of 72 kg with a standard deviation of 5 kg.
B a mean of 72 kg with a standard deviation of 7 kg.
C a mean of 74 kg with a standard deviation of 5 kg.
D a mean of 74 kg with a standard deviation of 7 kg.
E a mean of 74 kg with a standard deviation of 10 kg.

[VCAA 2003 1CQ5]

Question 7

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



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

A less than the mean time.

B equal to the mean time.

C greater than the mean time.

D half of the interquartile range.

E one quarter of the range.

[VCAA 2008 1CQ4]

Question 8

This table shows the percentage of women ministers in the parliaments of 22 countries in 2008.

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

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

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

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

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

c Copy and complete the stem plot by adding the value for Canada. 1 mark

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

[VCAA 2010 2CQ1]

Bell-shaped distributions

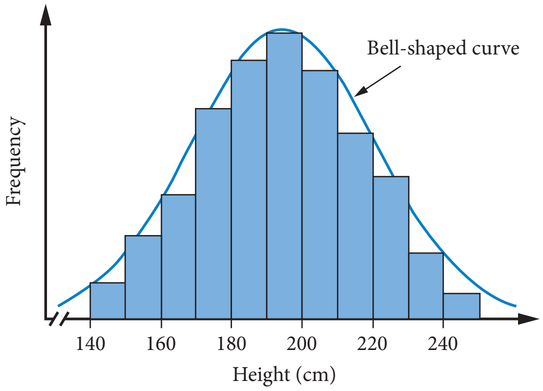
2.3

The normal or bell-shaped distribution

WS

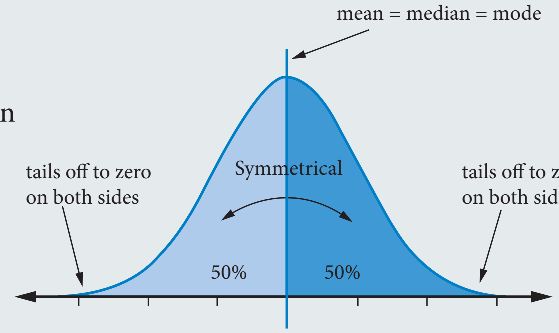
The normal curve

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



Bell-shaped distributions

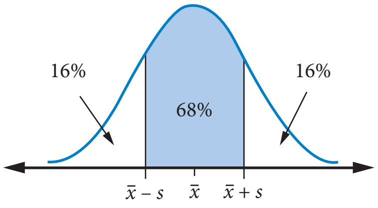
- are symmetrical about the mean
- have 50% of the data either side of the mean
- have mean = median = mode
- have a peak in the centre and tail off to zero on both sides



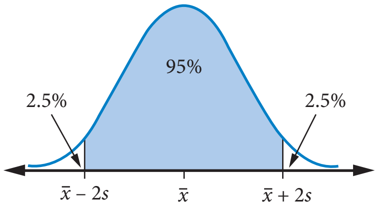
The 68–95–99.7% rule

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

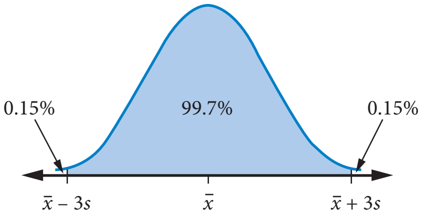
Around 68% of the data values lie within *one* standard deviation of the mean



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



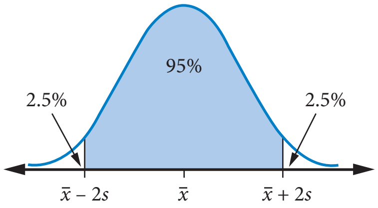
Around 99.7% of the data values lie within *three* standard deviations of the mean



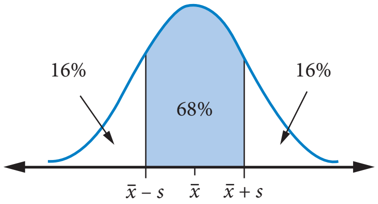
Worked example 2

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

a Find the percentage of packets that have between 36 and 48 chockbits.

	Working
1 Find whether the data values involved are 1, 2 or 3 standard deviations from the mean.	$36 = 42 - 6 = 42 - 2 \times 3 = \bar{x} - 2 \times s$ $48 = 42 + 6 = 42 + 2 \times 3 = \bar{x} + 2 \times s$
2 Sketch the relevant 68–95–99.7% rule diagram.	
3 Read the percentage from the diagram.	95% of packets have between 36 and 48 chockbits.

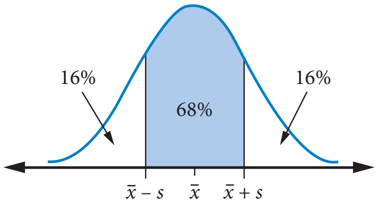
b Find the percentage of packets with more than 45 chockbits.

	Working
1 Find whether the data value involved is 1, 2 or 3 standard deviations from the mean.	$45 = 42 + 3 = 45 + 1 \times 3 = \bar{x} + 1 \times s$
2 Sketch the relevant 68–95–99.7% rule diagram.	
3 Read the percentage from the diagram.	16% of packets have more than 45 chockbits.

c Find the percentage of packets with more than 39 chockbits.

	Working
1 Find whether the data value involved is 1, 2 or 3 standard deviations from the mean.	$39 = 42 - 3 = 45 - 1 \times 3 = \bar{x} - 1 \times s$

2 Sketch the relevant 68–95–99.7% rule diagram.



3 Read the percentage from the diagram.

68% + 16% = 84% of packets have more than 45 chockbits.

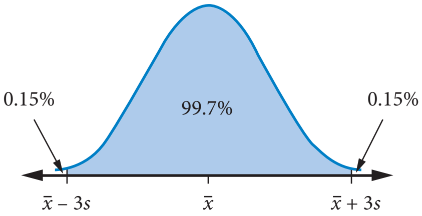
d A supermarket has bought 4000 chockbits packets. How many of these would they expect to have fewer than 33 chockbits?

Working

1 Find whether the number involved is 1, 2 or 3 standard deviations from the mean.

$33 = 42 - 9 = 42 - 3 \times 3 = \bar{x} - 3 \times s$

2 Sketch the relevant 68–95–99.7% rule diagram.



3 Read the percentage from the diagram.

0.15% of packets have fewer than 33 chockbits.

4 Find this percentage of the total given.

0.15% of 4000 = 6

The supermarket would expect 6 packets to have fewer than 33 chockbits.

Exam hack

If we are given the two data values that are one standard deviation from the mean, and we know we are dealing with an approximately normal distribution, then

$$\text{mean} \approx \frac{\text{smaller data value} + \text{larger data value}}{2}$$

$$\text{standard deviation} \approx \frac{\text{larger data value} - \text{smaller data value}}{2}$$

Shutterstock.com/bonchan

EXAM PREP 2.3

Bell-shaped distributions

Prep 1

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

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

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

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

d 95% of students are late to class between 16 and 25 days in a year.

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

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

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

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

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

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

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

Prep 2

WORKED EXAMPLE 2

A restaurant chain records the total time diners spend in the restaurant from when they enter to when they exit. It was found that the times approximated a bell-shaped distribution with the mean amount of time spent in the restaurant being 73 minutes and the standard deviation 8 minutes.

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

b Find the percentage of times spent in the restaurant that are between 49 and 97 minutes.

c Find the percentage of times spent in the restaurant that are less than 57 minutes.

d Find the percentage of times spent in the restaurant that are more than 49 minutes.

e One of the restaurants in the chain had 426 customers one day. How many of these would they expect to spend more than 81 minutes in the restaurant?

Bell-shaped distributions

Use the following information to answer Questions 1 & 2.

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

Question 1

The time that 95% of these students spent sleeping on a school night could be

A less than 6.0 hours.

B between 6.0 and 8.8 hours.

C between 6.7 and 8.8 hours.

D less than 6.0 hours or more than 8.8 hours.

E less than 6.7 hours or more than 9.5 hours.

[VCAA 2013 1CQ5]

Question 2

The number of these students who spent more than 8.1 hours sleeping on a school night was closest to

A 16 **B** 248 **C** 1302 **D** 1510 **E** 1545

[VCAA 2013 1CQ6]

Question 3

The level of oil use in certain countries is approximately normally distributed with a mean of 42.2 units and a standard deviation of 10.2 units.

The percentage of these countries in which the level of oil use is greater than 32 units is closest to

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

[VCAA 2009 1CQ7]

Use the following information to answer Questions 4 & 5.

The pulse rates of a large group of 18-year-old students are approximately normally distributed with a mean of 75 beats/minute and a standard deviation of 11 beats/minute.

Question 4

The percentage of 18-year-old students with pulse rates less than 75 beats/minute is closest to

A 32% **B** 50% **C** 68% **D** 84% **E** 97.5%

[VCAA 2008 1CQ6]

Question 5

The percentage of 18-year-old students with pulse rates less than 53 beats/minute or more than 86 beats/minute is closest to

A 2.5% **B** 5% **C** 16% **D** 18.5% **E** 21%

[VCAA 2008 1CQ7]

Question 6

The head circumference (in cm) of a population of infant boys is normally distributed with a mean of 49.5 cm and a standard deviation of 1.5 cm.

Four hundred of these boys are selected at random and each boy's head circumference is measured. The number of these boys with a head circumference of less than 48.0 cm is closest to

A 3 **B** 10 **C** 64 **D** 272 **E** 336

[VCAA 2006 1CQ4]

Question 7

The length of 3-month-old baby boys is approximately normally distributed with a mean of 61.1 cm and a standard deviation of 1.6 cm.

The percentage of 3-month-old baby boys with length greater than 59.5 cm is closest to

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

[VCAA 2007 1CQ4]

Question 8

The distribution of the weights of eggs produced by a chicken farm is approximately bell-shaped with a mean of 85 g and a standard deviation of 5 g.

Eggs weighing 95 g or more are classified as Extra Large.

The percentage of eggs that would be classified as Extra Large is closest to

A 0.15% **B** 0.35% **C** 2.5% **D** 5% **E** 16%

[VCAA 2004 1CQ3]

Question 9

The distribution of fuel consumption of a particular model of car is approximately bell-shaped with a mean of 8.8 km per litre and a standard deviation of 2.2 km per litre.

The percentage of this model of car that has a fuel consumption less than 6.6 km per litre is closest to

A 2.5% **B** 5% **C** 16% **D** 32% **E** 68%

[VCAA 2005 1CQ6]

Question 10

The distribution of test scores obtained when 2500 students sit for an examination is bell-shaped with a mean of 64 and a standard deviation of 12.

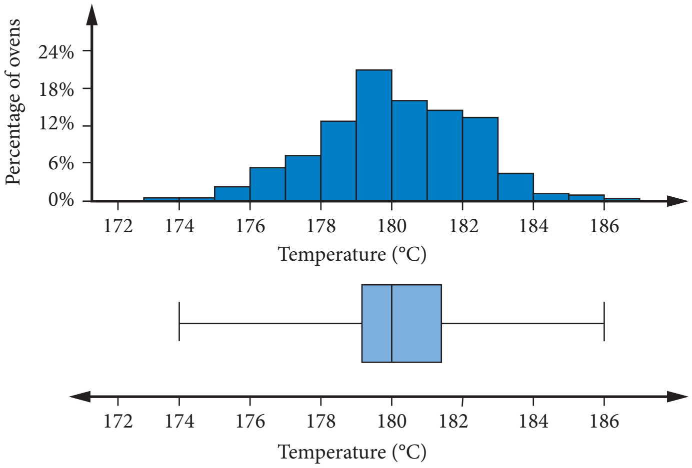
From this information we can conclude that the number of these students who obtained marks between 52 and 76 is closest to

A 68 **B** 95 **C** 850 **D** 1700 **E** 2375

[VCAA 2003 1CQ3]

Question 11

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



Using the 68–95–99.7% rule, the standard deviation for temperature is closest to

A 1°C **B** 2°C **C** 3°C **D** 4°C **E** 6°C

[VCAA 2010 1CQ3]

Question 12

The following data was recorded from measurements made on 12 men.

Age (years)	Mass (kg)	Waist (cm)
26	84	84
29	72	74
32	67	89
32	59	75
34	97	106
37	112	114
39	67	80
40	91	101
41	98	101
43	89	94
45	117	126
51	62	82

The sample of men has been drawn from a population in which the distribution of masses is bell-shaped, with a mean of 81.1 kg and a standard deviation of 17.9 kg.

The percentage of men in this population with a mass greater than that of the heaviest man in this sample is closest to

A 0.05%

B 2.5%

C 5%

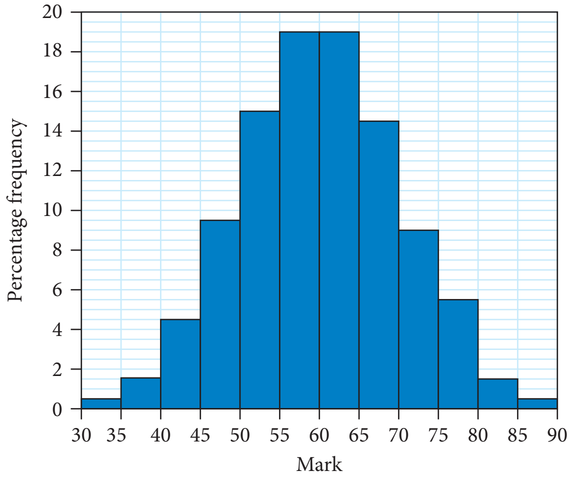
D 50%

E 95%

[VCAA 2002 1CQ8]

Use the following information to answer Questions 13 & 14.

1526 students sat for an examination. The histogram shows the distribution of marks.



Question 13

The median examination mark of these students is closest to

A 50 **B** 55 **C** 60 **D** 65 **E** 70

[VCAA 2002 1CQ5]

Question 14

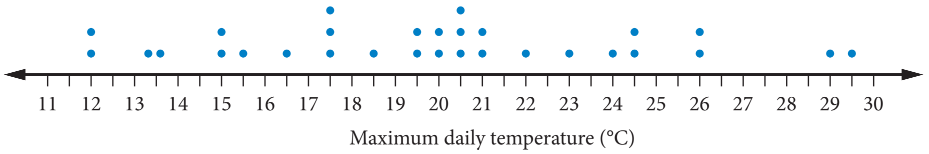
Using the 68–95–99.7% rule, the standard deviation of the marks is closest to

A 5 **B** 10 **C** 15 **D** 20 **E** 30

[VCAA 2002 1CQ6]

Question 15

The dot plot below displays the maximum daily temperature (in °C) recorded at a weather station on each of the 30 days in November 2011.



a From this dot plot, determine

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

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

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

b Determine the percentage of days in November that are expected to have a minimum daily temperature less than 14°C at this weather station.

Write your answer correct to 1 decimal place. 1 mark

[VCAA 2012 2CQ1]

Standardised values

2.4

Z-scores

z-scores

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

The standardised values

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

To calculate standardised values, use the formula:

$$\text{standardised value} = \frac{\text{data value} - \text{mean}}{\text{standard deviation}}$$

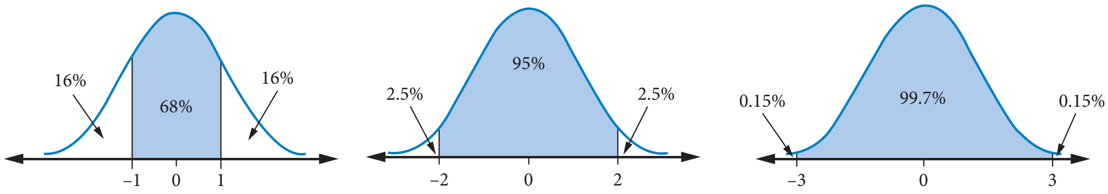
$$\text{or } z = \frac{x - \bar{x}}{s}$$

You can say the following about the original data value if you know its *z*-score:

Z-score	What does it mean?
z-score is positive	The data value is above the mean.
z-score is negative	The data value is below the mean.
z-score = 0	The data value equals the mean.
z-score = 1	The data value is 1 standard deviation *above* the mean.
z-score = –2	The data value is 2 standard deviations *below* the mean.

Exam hack

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



iStock.com/MsSponge

Worked example 3

The length of a handspan of adults is known to be normally distributed with a mean of 21.5 cm and a standard deviation of 1.5 cm. A person has a handspan of 20 cm.

a Calculate the standardised value for a this data value.

Working

1 Write the z-score formula.

$$z = \frac{x - \bar{x}}{s}$$

2 Substitute in the data value, mean and standard deviation.

$$z = \frac{20 - 21.5}{1.5} = \frac{-1.5}{1.5} = -1$$

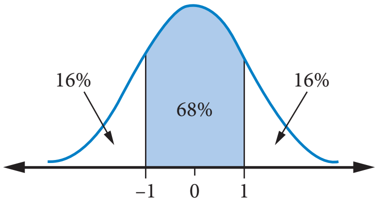
b Show that this standardised value means that 84% of people have handspans wider than this person.

Working

1 State how the z-score relates to the number of standard deviations.

$z = -1$ means the person's handspan is one standard deviation below the mean.

2 Sketch the relevant standardised 68–95–99.7% rule diagram.



3 Read the percentage from the diagram.

From the diagram, 68% + 16% = 84%.

So 84% of people have a handspan greater than this person.

Using *z*-scores to compare

Suppose you had a mark in the Further Mathematics mid-year exam of 91% and a mark in the Specialist Mathematics mid-year exam of 49%. You couldn't necessarily say you performed better in Further Mathematics because Specialist Mathematics is a more difficult subject.

The way to compare the two results, assuming both distributions are bell-shaped, is to **standardise** them.

Worked example 4

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

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

Assuming the results for each of the three subjects approximates a bell-shaped distribution:

a In which mathematics subject did the student perform the best?

Working

1 Write the z-score formula.

$$z = \frac{x - \bar{x}}{s}$$

2 Substitute the values for each subject.

Further Mathematics

$$z = \frac{91 - 73}{6} = \frac{18}{6} = 3$$

Mathematical Methods

$$z = \frac{67 - 51}{4} = \frac{16}{4} = 4$$

Specialist Mathematics

$$z = \frac{49 - 31}{5} = \frac{18}{5} = 3.6$$

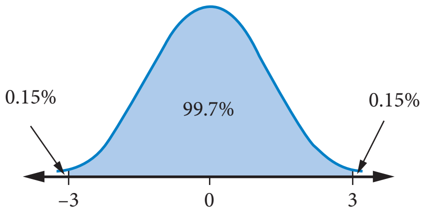
3 State which z-score is the largest.

The student performed best in Mathematical Methods.

b In which of these subjects was she in the top 0.15% of students?

Working

1 Sketch the relevant standardised 68–95–99.7% rule diagram.



2 Read the answer from the diagram.

From the diagram, the top 0.15% of students had a z-score of 3 or more. The student had z-scores of 3, 4 and 3.6, so she was in the top 0.15% of students in all three mathematics subject.

EXAM PREP **2.4**

Standardised values

Prep 1

WORKED EXAMPLE 3

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

a Calculate the standardised value for a this data value.

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

Prep 2

WORKED EXAMPLE 4

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

	Marks	Mean	Standard deviation
Further Mathematics	78	72	5
German	47	48	2
Hospitality	77	71	2
Psychology	50	54	4
Systems Engineering	62	68	5

Assuming the results for each of the subjects approximate a bell-shaped distribution:

a In which of the subjects did the student perform the best?

b In which of these subjects was the student in the bottom 16%?

EXAM PRACTICE **2.4**

Standardised values

Use the following information to answer Questions 1 & 2.

The lengths of the left feet of a large sample of Year 12 students were measured and recorded. These foot lengths are approximately normally distributed with a mean of 24.2 cm and a standard deviation of 4.2 cm.

Question 1

A Year 12 student has a foot length of 23 cm.

The student's standardised foot length (standard z-score) is closest to

A −1.2 **B** −0.9 **C** −0.3 **D** 0.3 **E** 1.2

[VCAA 2010 1CQ5]

Question 2

The percentage of students with foot lengths between 20.0 and 24.2 cm is closest to

A 16% **B** 32% **C** 34% **D** 52% **E** 68%

[VCAA 2010 1CQ6]

Use the following information to answer Questions 3 & 4.

The length of a type of ant is approximately normally distributed with a mean of 4.8 mm and a standard deviation of 1.2 mm.

Question 3

From this information it can be concluded that around 95% of the lengths of these ants should lie between

A 2.4 mm and 6.0 mm **B** 2.4 mm and 7.2 mm **C** 3.6 mm and 6.0 mm

D 3.6 mm and 7.2 mm **E** 4.8 mm and 7.2 mm

[VCAA 2011 1CQ9]

Question 4

A standardised ant length of $z = -0.5$ corresponds to an actual ant length of

A 2.4 mm **B** 3.6 mm **C** 4.2 mm **D** 5.4 mm **E** 7.0 mm

[VCAA 2011 1CQ10]

Question 5

A student obtains a mark of 56 on a test for which the mean mark is 67 and the standard deviation is 10.2. The student's standardised mark (standard z-score) is closest to

A –1.08 **B** –1/01 **C** 1.01 **D** 1.08 **E** 49.4

[VCAA 2007 1CQ3]

Question 6

A class of students sat for a Biology test and a Legal Studies test. Each test had a possible maximum score of 100 marks. The table shows the mean and standard deviation of the marks obtained in these tests.

	Subject	
	Biology	Legal Studies
Class mean	54	78
Class standard deviation	15	5

The class marks in each subject are approximately normally distributed. Sashi obtained a mark of 81 in the Biology test.

The mark that Sashi would need to obtain on the Legal Studies test to achieve the same standard score for both Legal Studies and Biology is

A 81 **B** 82 **C** 83 **D** 87 **E** 95

[VCAA 2012 1CQ4]

2.5 Populations and samples

Shutterstock.com/Kzenon

Populations vs samples

Data can be collected from a **population** or **sample**.

A population includes all the items in the group being studied. These items can be people, things or or events.

A sample is a section of the population that is selected so that it is representative of the population and can be used to draw accurate conclusions about the population.

Here are some examples of populations:

- If you are investigating the height of Year 12 students in Victoria, the population would be *every* Year 12 student in Victoria.
- If you are investigating how many defective components a factory produced in 2016, the population would be *every* component produced in the factory in 2016.
- If you are investigating the number of road accidents in Victoria in 2017, the population would be *every* road accident in Victoria in 2017.

Data is usually collected from a sample rather than a population unless the population is small. It would be very time-consuming and costly, for example, to measure the height of *every* Year 12 student in Victoria. Two exceptions involving a large population are:

- the Australian census undertaken every five years, where *every* person in Australia is required to answer a series of questions.
- Australian elections where *all* Australian citizens are required to register a vote by a nominated date.

Population parameters vs sample statistics

A **population parameter** is a number that describes a characteristic of a population.

A **sample statistic** is a number calculated from sample data that approximates a population parameter.

It's often possible to tell from a statistical statement whether a population parameter or a sample statistic is most likely involved:

The 'median mark in a particular class test' is most likely a population parameter.

(The median calculation would have taken into account *all* the marks given for that particular class test.)

The 'median number of televisions in Victorian households' is most likely a sample statistic.

(The median calculation here would have involved surveying a sample rather than asking every Victorian household.)

Population parameters are fixed numbers (even if we don't always know what those numbers are). No one will probably ever measure the height of every Year 12 student in Victoria, but we know that the population mean height is a fixed number.

Sample statistics, on the other hand, vary depending on the sample taken.

To make clear distinction between the sample mean and the population mean and between the sample standard deviation and population standard deviation, we use different notation.

$\bar{x}$ is used to represent the sample mean, and s is used to represent the sample standard deviation.

μ (Greek letter mu) is used to represent the **population mean**, and σ (Greek letter sigma) is used to represent the **population standard deviation**.

The formulas for calculating $\bar{x}$ and μ are the same.

The formulas for s and σ are slightly different.

Sample standard deviation $s = \sqrt{\dfrac{\Sigma(x-\bar{x})^2}{n-1}}$

Population standard deviation $\sigma = \sqrt{\dfrac{\Sigma(x-\bar{x})^2}{n}}$

Exam hack

Make sure you choose the correct standard deviation formula on your CAS/calculator. In most cases it will be the sample standard deviation (sx on the TI-Nspire CAS and s_x on the ClassPad) rather than the population standard deviation (σx on the TI-Nspire CAS and σ_x on the ClassPad).

WS

Tables of random numbers

Random sampling

The purpose of taking a sample is to be able to use it to generalise for the population. To do this the sample needs to be representative of the population. The simplest way to get a representative sample from a population is to select a **simple random sample**.

In a simple random sample

- everything or everyone in the population has an equal chance of being chosen
- no member of the population is excluded
- each member of the population is selected independently (i.e. choosing a member doesn't affect the selection of another member in any way).

Random numbers can be generated on a CAS/calculator to help in selecting a simple random sample. Each member of the population is assigned a number starting at 1. If the numbers 2, 32, 54 are generated, it means select person or item 2, 32 and 54 from the population.

A variation of random sampling involves randomly allocating subjects to groups. For example, a medical study may require people to be randomly divided into 5 groups that get 5 different treatments. A CAS/calculator can be used to randomly generate the numbers 1 to 5 for each person to ensure the group allocation is random.

Using CAS Generating random numbers for sampling

Generate 15 random numbers ranging from 1 to 200.

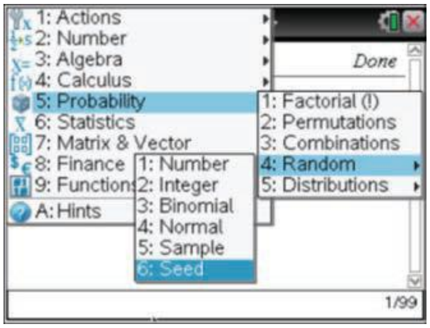
TI-NSPIRE CAS

STEP 1

Set seed. (This only needs to be done once.) Each student should enter a different seed value, possibly their birth day and month, otherwise they will generate the same set of random numbers.

Press `menu` 5: Probability, 4: Random, 6: Seed

Then enter the chosen value; we have chosen 12.



RandSeed 12 Done

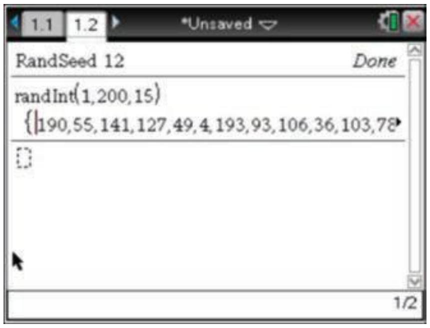
STEP 2

Generate random integers.

Press `menu` 5: Probability, 4: Random, 2: Integer

randint(starting number, end number, how many numbers to generate) `enter`

Depending on how many numbers you generate, you may need to Press ▸ to see them all.



RandSeed 12 Done

randInt(1,200,15)

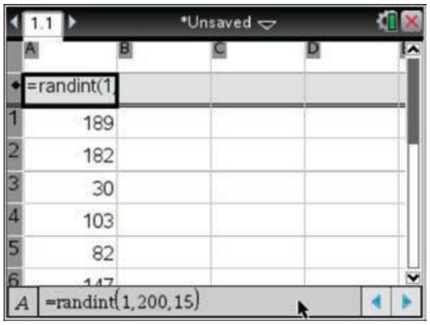
{190,55,141,127,49,4,193,93,106,36,103,78

Alternative

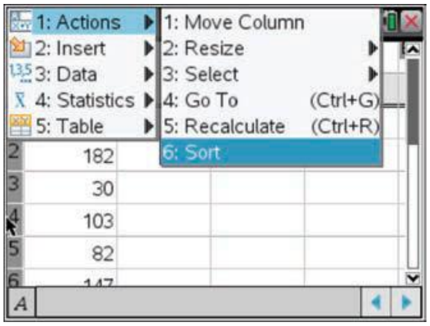
STEP 1

This same method can be used in a Lists & Spreadsheet page.

Press menu 3: Data, 5: Random, 2: Integer then **randint(starting number, end number, how many numbers to generate)** enter into the dark formula cell of column A. The benefit is that you can then sort the data from smallest to largest. Press menu 1: Actions, 6: Sort

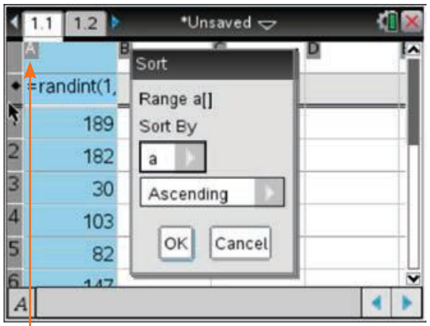


A =randint(1,200,15)



STEP 2

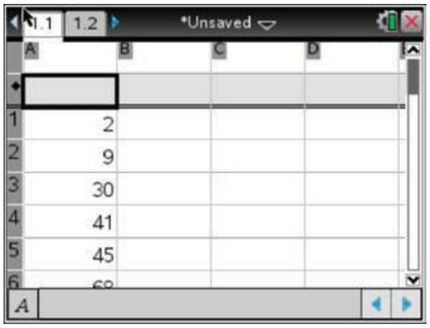
Select the column to be sorted and ascending for smallest to largest. Click OK



The cursor must be in the top cell of column A to sort the data.

STEP 3

The random numbers are now sorted.



CLASSPAD

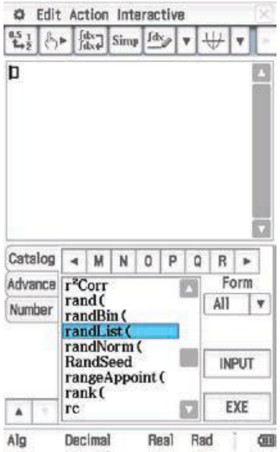
STEP 1

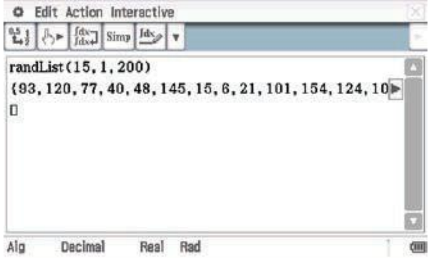
Tap √α. If necessary, use the **Edit** menu to **Clear All**.

Press Keyboard and tap ▼ to access the Catalog of functions. Use ▶ to scroll along as necessary to see R. Tap R then tap **randList** twice on the list of functions.

Add to **randList(** displayed on the screen by typing 15 , 1 , 200) and press EXE.

The first number states how many numbers to generate (15); the next two are the lower and upper limits (1 to 200).





randList(15, 1, 200)
{93, 120, 77, 40, 48, 145, 15, 6, 21, 101, 154, 124, 10▶

STEP 2

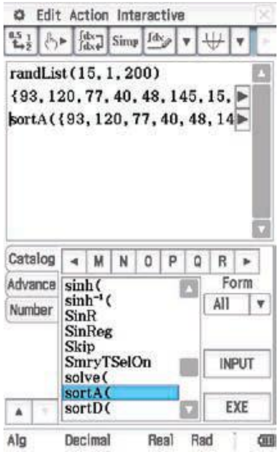
Sort the list of numbers.

Tap **sortA(** twice from the Catalog.

Tap the list of numbers. They should all be highlighted.

Drag the highlighted list into the **sortA** bracket, then tap EXE.

To see the numbers not displayed on the screen, tap the arrows at the end of the list, ▶ to scroll right or ◀ to scroll left as necessary, or rotate screen.

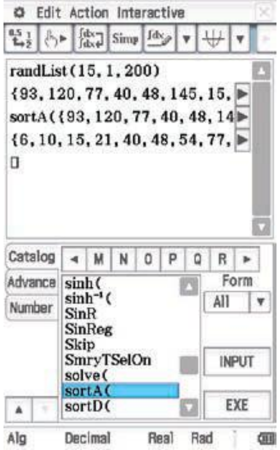


randList(15, 1, 200)
{93, 120, 77, 40, 48, 145, 15, ▶
sortA({93, 120, 77, 40, 48, 14▶

STEP 3

Pressing EXE will result in the list being sorted in ascending order.

Write down the numbers in order.

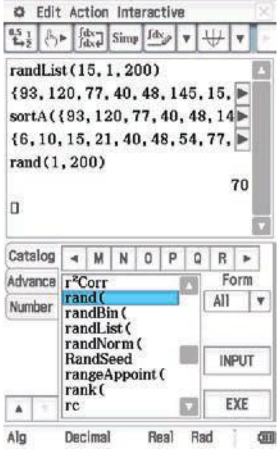

```
randList(15, 1, 200)
{93, 120, 77, 40, 48, 145, 15, ▶
sortA({93, 120, 77, 40, 48, 14▶
{6, 10, 15, 21, 40, 48, 54, 77, ▶
```

STEP 4

Sometimes you will notice that the same number appears twice. Simply use **rand(** to select a new random number from 1 to 200.

rand(1, 200) generates a random number from 1 to 200.

To repeat: If you need another random number, press EXE without entering **rand(1, 200)** again, as many times as required.


```
randList(15, 1, 200)
{93, 120, 77, 40, 48, 145, 15, ▶
sortA({93, 120, 77, 40, 48, 14▶
{6, 10, 15, 21, 40, 48, 54, 77, ▶
rand(1, 200)
                              70
```

EXAM PREP 2.5

Populations and samples

Prep 1

State whether the following statistical statements are most likely dealing with data collected from a population or a sample.

a 35% of Australians have travelled overseas in the last 5 years.

b 35% of my class bought their lunch at the canteen yesterday.

c The range of ages in my basketball team is 21 months.

d The range of ages of the players in my state-wide basketball competition is 21 months.

e The median mark on the chapter test is 72.

f The median mark on the Drivers Learner Permit Knowledge Test is 72.

g The mean number of people living in houses in a particular suburb is 2.3.

h The mean number of people living in houses adjacent to your school is 2.3.

Prep 2

Use your CAS/calculator to calculate the standard deviation, using the correct notation, for the following ATAR Subject Scores correct to 1 decimal place:

34.22, 38.23, 41.48, 29.95, 33.67, 41.2, 37.16

a if this is a population

b if this is a sample

Prep 3 USING CAS: GENERATING RANDOM NUMBERS FOR SAMPLING

Generate the following lists of random numbers and then sort them from smallest to largest.

a 5 different numbers ranging from 10 to 50

b 15 different numbers ranging from 0 to 150

c 30 different numbers ranging from 1 to 250

EXAM PRACTICE 2.5

Populations and samples

Question 1

Which one of the following statistical statements is most likely dealing with data collected from a population?

A 28% of Victorians drive a white car.

B 28% of AFL fans prefer a night Grand Final.

C 28% of the Year 12 students at a particular school had a job.

D 28% of Australians visited the beach at least once last summer.

E 28% of car accidents on freeways involve serious injury.

Question 2

Which one of the following statistical statements is most likely dealing with data collected from a sample?

A 20% of players at a particular netball club are aged over 35.

B 32% of Australians rate the Prime Minister as 'doing a good job'.

C 11% of United Nations members supported military action.

D The mean raw score on the last Further Mathematics Examination was 54%.

E The median price of houses sold by a real estate company in Melbourne on the weekend was $945 543.

Question 3

A farmer decides to work out which of 6 different varieties of tomatoes grows best in the conditions on his farm. Each variety was planted on 5 randomly selected plots and their quality was measured when they were harvested. Which one of these statements is incorrect?

A The populations involved are the 6 different varieties of tomatoes.

B The samples involved are the 5 plots of each variety of tomato.

C The results calculated are population parameters.

D This investigation involves random sampling.

E The results calculated are sample statistics.

Question 4

A polling company was commissioned to research how many minutes of television Australians spend watching on Sundays. 500 adults were surveyed and the number of minutes watched was recorded. The mean was calculated as 128 minutes and the standard deviation as 36 minutes. Which one the following statements is correct?

A $\bar{x} = 128$ and $\sigma = 36$.

B The population mean is 128.

C $\mu = 128$

D The population standard deviation is 36.

E $s = 36$

Question 5

Which one of the following is not an example of simple random sampling?

A A polling company wanting to measure the popularity of the prime minister across Australia ensures that its sample of 2000 Australians has specific numbers randomly selected from each of the states that match the population proportions of each state.

B A shop wanting to measure customer satisfaction randomly selects 100 people over six months and asks them a series of questions.

C A company checks the quality of its products by randomly selecting 200 products and testing their quality.

D The RACV wants to know how many of their members are using their mobile phone while driving. It develops a questionnaire, randomly selects people to survey from their membership list, and then phones them to ask the questions.

E A school wants to find out whether parents would support a change in uniform. It randomly selects families from its data base and sends them a questionnaire.

SUMMARY

②

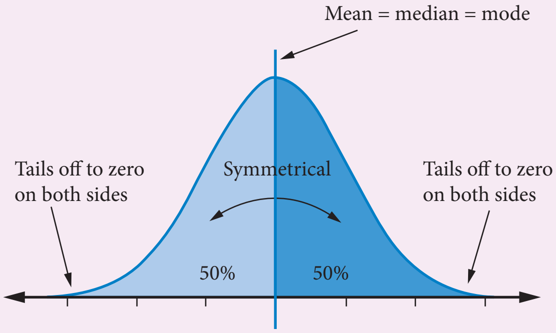
Further data distributions

Practice quiz

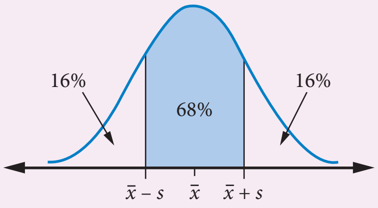
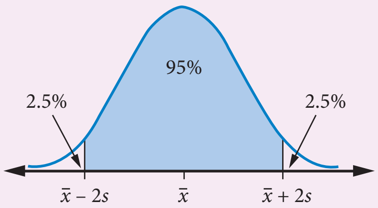
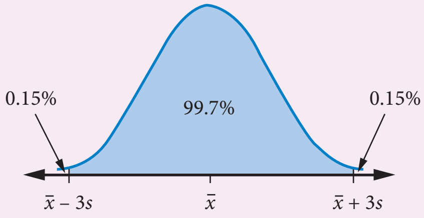
Scales

- On a **linear scale**, you *add* the same number to get from one scale mark to the next.
- On a **log scale**, you *multiply* the same number to get from one scale mark to the next.

Normal or bell-shaped distributions

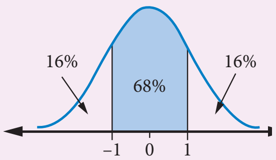


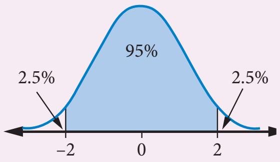
- The **68–95–99.7% rule** for approximate **bell-shaped distributions** says:

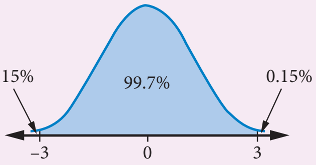
Around 68% of the data values lie within *one* standard deviation of the mean.	
Around 95% of the data values lie within *two* standard deviations of the mean.	
Around 99.7% of the data values lie within *three* standard deviations of the mean.	

Standardised values (*z*-scores)

- **standardised value** $= \dfrac{\text{data value} - \text{mean}}{\text{standard deviation}}$ or $z = \dfrac{x - \bar{x}}{s}$
- 68–95–99.7% rule for z-**scores**:







Populations and samples

- A **population** includes all the items in the group being studied. These items can be people, things or events.
- A **sample** is a section of the population that is selected so that it is representative of the population and can be used to draw accurate conclusions about the population.
- A **population parameter** is a number that describes a characteristic of a population.
- A **sample statistic** is a number calculated from sample data that approximates a population parameter.
- In a **simple random sample** everyone or everything in the population has an equal chance of being chosen.

Mean and standard deviation

- The **sample mean** is $\bar{x}$ and the **population mean** is μ. The formulas for calculating $\bar{x}$ and μ are the same.

The sample mean for a list of data values: $\bar{x} = \dfrac{\text{sum of all the values}}{\text{number of values}}$

$$= \frac{\Sigma x}{n}$$

The sample mean for data in a frequency table: $\bar{x} = \dfrac{\text{sum of (each value} \times \text{its corresponding frequency)}}{\text{sum of frequencies}}$

$$= \frac{\Sigma xf}{\Sigma f}$$

- The standard deviation is a measure of the spread of data. The **sample standard deviation** is s and **population standard deviation** is σ. The formulas for calculating s and σ are slightly different:

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}} \qquad \sigma = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n}}$$

Median vs the sample mean

- For symmetric distributions, the median = the sample mean.
- For distributions that are approximately symmetric, the median and the sample mean will be approximately equal.
- The sample mean is greater than the median for positively skewed distributions.
- The sample mean is less than the median for negatively skewed distributions.
- Outliers don't generally affect the median, but they can significantly affect the mean.

CHAPTER 3

ASSOCIATIONS BETWEEN TWO VARIABLES

Prior learning

3.1 Explanatory and response variables

iStock.com/joannawnuk

Dealing with two variables

All the examples we've looked at so far have involved only one variable. From now on we will be exploring associations between two variables. Data associated with two related variables is called **bivariate data**.

Here are some examples of questions that involve analysing the association between two variables.

- Does smoking cause lung cancer?
- Does human activity cause global warming?
- Does the number of books in a home affect academic success?
- Does the amount of time spent on social media affect VCE results?
- Does the consumption of mozzarella cheese in a region have an effect on the number of civil engineering doctorates awarded in that region?

Identifying explanatory and response variables

It's important to decide what type of variable you are dealing with.

An **explanatory variable** (or **independent variable**) is a variable that we expect to affect another variable. A **response variable** (or **dependent variable**) is a variable that we expect to be affected by another variable.

The question to ask to decide whether a variable is explanatory or response is 'Which variable do we expect to affect the other?'

For example, if researchers wish to investigate the association between age and the amount of time you sleep, they would be looking at how age affects the amount of time you sleep (not how the amount of time you sleep affects your age). So 'age' is the explanatory variable and 'the amount of time you sleep' is the response variable.

EXAM PREP 3.1

Explanatory and response variables

Prep 1

Identify the explanatory variable in the following cases.

a A researcher wishes to find out whether the amount of time spent in front of a screen affects school results.

b Research is undertaken to see whether eating certain amounts of chocolate gives you headaches.

c A study is done to establish whether there is an association between the time taken to complete a Fun Run and the age of a person.

d A statistical analysis is undertaken to establish whether or not males are more likely than females to be left handed.

e A researcher wants to determine whether there is a connection between success at school and the number of television sets in the home.

f An experiment was conducted to test the association between driving response times and levels of sleep deprivation.

EXAM PRACTICE 3.1

Explanatory and response variables

Question 1

A consumer group conducts research to determine whether the price of a laptop is affected by its mass. They weigh 50 laptops and note each of their prices on a spreadsheet. The explanatory and response variables are, respectively:

A the 50 laptops and their prices

B the price of the laptop and the mass of the laptop

C the mass of the laptop and the price of the laptop

D the prices of the laptops and the spreadsheet

E the consumer group and the study

Question 2

The Psychology department of a university undertakes a study where 5 groups of mice are timed on how long they take to go through a series of mazes. Each mouse in the first group is given one food pellet every time they make it through a maze, each mouse in the second group is given two pellets, and so on. What is the explanatory variable?

A the mice

B the average amount of time it takes the mice to make it through the maze

C the amount of time each individual mouse takes to make it through the maze

D the maze

E the number of food pellets

Question 3

A research company sets up a study where job recruiters are shown one of two CVs. The two CVs are identical except for the name. On one CV the name is a short simple Anglo-Saxon name and on the other it's a long, complex, difficult-to-pronounce name. The job recruiters are asked to rate how likely it would be that the candidate would be successful in getting the job. What is the response variable?

A The ethnic background of the job recruiters

B The two CVs

C The job recruiters' ratings of the candidates' expected success

D The name on the CV

E The job recruiters

Associations between two categorical variables

3.2

iStock.com/Lisay

WS

Two-way tables

Two-way frequency tables

An effective way to look at the association between two categorical variables is by using a two-way **frequency table** (or **contingency table**).

Here's an example of a two-way frequency table which looks at the association between gender and exercise habits.

Explanatory variable

Response variable

	Gender		
Exercise habits	**Male**	**Female**	**Total**
Exercise	3028	1084	4112
No exercise	1532	946	2478
Total	4560	2030	6590

Row totals

Column totals

Number of males who exercise

Exam hack

The variable appearing in the columns of a two-way frequency table is usually, *but not always*, the explanatory variable.

It's possible to have more than two categories for each variable.

Worked example 1

A market researcher interviewed 50 people about their preferences of ice-cream flavours. Of the 30 women, 20 said they preferred chocolate, while 14 of the men preferred other flavours.

Construct a two-way frequency table showing this information.

Working

1 Create a table using gender as the explanatory variable for the columns and ice-cream flavour as the response variable for the rows.

	Gender		
Ice-cream flavour	**Women**	**Men**	**Total**
Chocolate			
Other			
Total			

2 Fill in information from the question.

- 50 people
- 30 women
- 20 women preferred chocolate
- 14 men preferred other flavours

	Gender		
Ice-cream flavour	**Women**	**Men**	**Total**
Chocolate	20		
Other		14	
Total	30		50

3 Complete the table using column and row totals.

20 + 10 = 30

30 + 20 = 50

10 + 14 = 24

26 + 24 = 50 and

20 + 6 = 26

	Gender		
Ice-cream flavour	**Women**	**Men**	**Total**
Chocolate	20	6	26
Other	10	14	24
Total	30	20	50

Percentaging two-way frequency tables

To get more information from a two-way frequency table, we **percentage** the data values. There are a number of ways to calculate the percentages in a two-way frequency table, depending on what we want to find out about, but usually we percentage the explanatory variables.

Percentaging gender in our two-way frequency table of gender and exercise habits gives:

	Gender	
Exercise habits	**Male**	**Female**
Exercise	66.4%	53.4%
No exercise	33.6%	46.6%
Total	100.0%	100.0%

$$\frac{\text{data value}}{\text{column total}} \times 100\% = \frac{3028}{4560} \times 100\%$$

From this **percentage two-way frequency table** we can see that there is an association between gender and exercise habits: 66.4% of males exercise while only 53.4% of females exercise. This indicates that there is a difference in exercise habits between males and females. If there was no association, the male and female percentages would have been approximately equal (at most a few percentage points apart).

Worked example 2

Convert the two-way frequency table from worked example 1 into a percentage two-way frequency table by percentaging the explanatory variable. What does the table suggest about the association between gender and chocolate ice-cream preferences?

Working

1 The explanatory variable forms the columns, so redraw the two-way table using only the column totals.

	Gender	
Ice-cream flavour	**Women**	**Men**
Chocolate	20	6
Other	10	14
Total	30	20

2 Calculate the percentage of women who prefer chocolate.

$\frac{20}{30} \times 100 \approx 67\%$

3 Calculate all other required percentages.

Women Other: $\frac{10}{30} \times 100 \approx 33\%$

Men Chocolate: $\frac{6}{20} \times 100 = 30\%$

Men Other: $\frac{14}{20} \times 100 = 70\%$

4 Put the percentages into the two-way table.

	Gender	
Ice-cream flavour	**Women**	**Men**
Chocolate	67%	30%
Other	33%	70%
Total	100%	100%

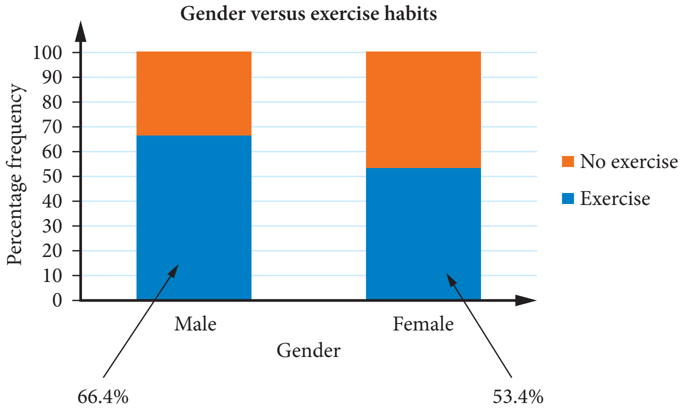
5 Check whether there are significant differences between the categories of the explanatory variable.

The table suggests there is an association between gender and ice-cream preferences, and that women have a greater preference for chocolate ice-cream than do men.

Two-way frequency tables and segmented bar charts

It is often easier to see patterns in data when it is displayed as a chart rather than a table. Information from a percentage two-way frequency table can be displayed as a **percentage segmented bar chart** where each cell in the table corresponds to a segment in the bar chart.

For example, the percentage two-way frequency table of gender and exercise habits can be shown as the following percentage segmented bar chart:

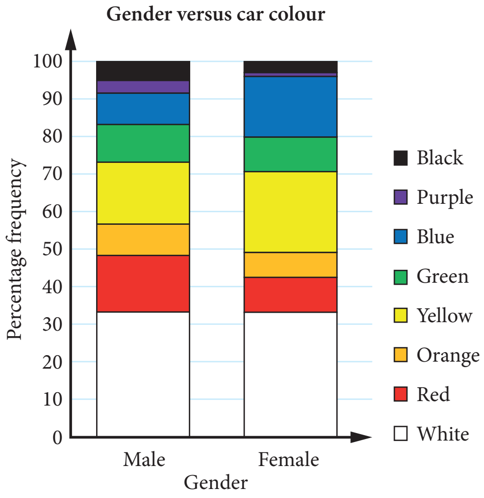


Exam hack

Percentage segmented bar charts are particularly useful when we are dealing with variables with more than two categories.

Worked example 3

The following percentage segmented bar chart shows the results of a survey of preferred car colours of males and females. Discuss if this suggests there is an association between car colour and gender.



Working

Look at how many of the segments are similar and how many are different.

There are considerable differences in the numbers of males and females who choose blue, yellow and red cars, which by itself would suggest that there may be an association between car colour and gender. However, the five other colours were very similar for males and females. This suggests that there may be an association, but only with certain colours.

EXAM PREP 3.2

Associations between two categorical variables

Prep 1

WORKED EXAMPLE 1

80 people were asked about whether they use sunscreen at the beach. Of the 40 women, 30 said they do, with 35 of the men also stating that they use sunscreen at the beach. Construct a two-way frequency table using gender for the columns.

Prep 2

WORKED EXAMPLE 2

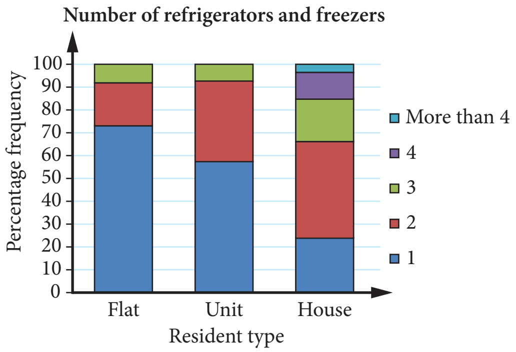
Convert the following two-way frequency table into a percentage two-way frequency table (to the nearest percentage) by percentaging the explanatory variable. What does the table suggest about the association between gender and library membership?

	Gender		
Liberary membership	Male	Female	Total
Library member	20	32	52
Not a library member	8	15	23
Total	28	47	75

Prep 3

WORKED EXAMPLE 3

The following percentage segmented bar chart shows the results of a survey about the type of residence and the number of refrigerators and freezers in it. Discuss if this suggests there is an association between the type of residence and the number of refrigerators and freezers.



EXAM PRACTICE 3.2

Associations between two categorical variables

Use the following information to answer Questions 1 & 2.

The table lists the speed (in km/h) of ten cars recorded in a 60 km/h zone. Also recorded are the ages (in years) of the drivers.

Speed	Age
71.8	27
68.3	38
65.1	22
63.2	64
62.8	57
62.6	37
62.5	21
61.3	19
60.1	57
59.8	61

Question 1

The median speed (in km/h) of the ten cars is

A 62.6 **B** 62.7 **C** 62.8

D 63.0 **E** 63.5

[VCAA 2005 1CQ1]

Question 2

The percentage of the drivers over the age of 25 years is

A 30% **B** 40% **C** 50%

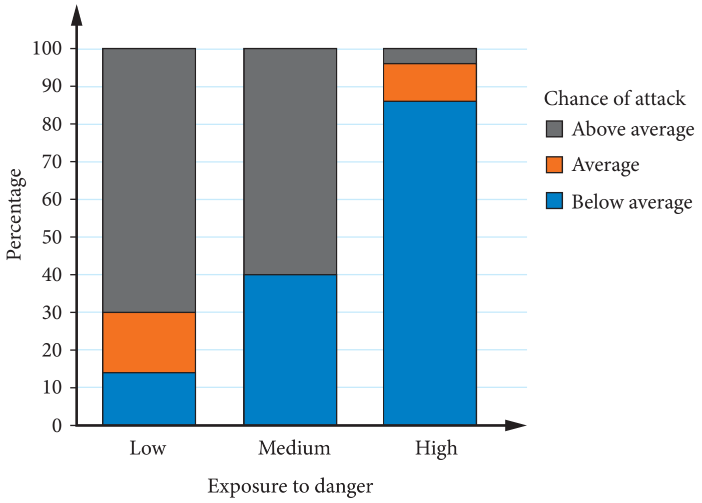
D 60% **E** 70%

[VCAA 2005 1CQ2]

Question 3

An animal study was conducted to investigate the association between *exposure to danger* during sleep (high, medium, low) and *chance of attack* (above average, average, below average). The results are summarised in the percentage segmented bar chart.

The percentage of animals whose *exposure to danger* during sleep is high, and whose *chance of attack* is below average, is closest to



A 4% **B** 12% **C** 28% **D** 72% **E** 86%

[VCAA 2009 1CQ8]

9780170371100

Question 4

The table below shows the percentage of households with and without a computer at home for the years 2007, 2009 and 2011.

	Year		
	2007	2009	2011
Households with a computer	66.4%	77.7%	84.5%
Households without a computer	33.6%	22.3%	15.5%

In the year 2009, a total of 5 170 000 households were surveyed.

The number of households **without** a computer at home in 2009 was closest to

A 801 000 **B** 1 153 000 **C** 1 737 000

D 3 433 000 **E** 4 017 000

[VCAA 2012 1CQ6]

Use the following information to answer Questions 5 & 6.

The level of Internet usage (never used, sometimes used, often used) for 217 school students sampled from Years 3 to 12 is indicated in the table below. Some of the entries in the table are missing.

	Year group			
Level of Internet usage	3–6	7–10	11–12	Total
Never used	44	9	8	
Sometimes used	16			58
Often used	10		47	
Total			73	217

Question 5

For this sample of students, the total number of students who never used the Internet is

A 44 **B** 51 **C** 61 **D** 70 **E** 217

[VCAA 2004 1CQ6]

Question 6

The percentage of Year 7–10 students who sometimes used the Internet is closest to

A 11% **B** 24% **C** 27% **D** 28% **E** 32%

[VCAA 2004 1CQ7]

Use the following information to answer Questions 7–9.

Researchers conducted a survey of 403 school leavers who had recently entered the workforce. The aim was to determine whether the type of work they undertook was gender related. Work type was classified as 'trade', 'clerical', 'manual' or 'professional'.

The data in the table comes from the survey.

	Gender	
Work type	Male	Female
Trade	104	18
Clerical	21	143
Manual	72	31
Professional	8	6
Total	205	198

Question 7

Of the females surveyed, the percentage who became clerical workers is closest to

A 10% **B** 14% **C** 35%

D 72% **E** 87%

[VCAA 2002 1CQ2]

Question 8

Of the school leavers surveyed, the percentage who became clerical or manual workers is closest to

A 26% **B** 31% **C** 41% **D** 50% **E** 66%

[VCAA 2002 1CQ3]

Question 9

Which statement supports the suggestion that the type of work undertaken by school leavers is associated with gender?

A 50.9% of those surveyed are male compared to 49.1% female.

B 30.3% of those surveyed entered a trade while only 3.5% entered professions.

C 35.1% of male school leavers entered manual work compared to 15.7% of female school leavers.

D 403 school leavers participated in the survey.

E A similar but small percentage of male and female school leavers undertook professional work.

[VCAA 2002 1CQ4]

Use the following information to answer Questions 10 & 11.

Text messaging use (never, sometimes, every day) and the number of mobile phones in the household were recorded for a sample of 154 households. The results are shown in the table below.

	Number of mobile phones in household				
Text messaging use	**0**	**1**	**2**	**3**	**Total**
Never	34	10	3	0	47
Sometimes	0	23	12	2	37
Every day	0	45	15	10	70
Total	34	78	30	12	154

Question 10

Of the households with two mobile phones in the sample, the percentage that never used text messaging is

A 0% **B** 6% **C** 10% **D** 20% **E** 30%

[VCAA 2005 1CQ4]

Question 11

The mean number of mobile phones in these 154 households is closest to

A 1.13 **B** 1.45 **C** 1.50 **D** 1.54 **E** 2.00

[VCAA 2005 1CQ5]

Use the following information to answer Questions 12 & 13.

The heights of 82 mothers and their eldest daughters are classified as 'short', 'medium' or 'tall'. The results are displayed in the frequency table below.

		Mother		
		Short	**Medium**	**Tall**
Eldest daughter	**Short**	16	10	3
	Medium	8	14	11
	Tall	5	7	8

Question 12

The number of mothers whose height is classified as 'medium' is

A 7 **B** 10 **C** 14 **D** 31 **E** 33

[VCAA 2013 1CQ3]

Question 13

Of the mothers whose height is classified as 'tall', the percentage who have eldest daughters whose height is classified as 'short' is closest to

A 3% **B** 4% **C** 14% **D** 17% **E** 27%

[VCAA 2013 1CQ4]

Question 14

The table below shows the percentage of students in two age groups (15–19 years and 20–24 years) who regularly use the Internet at one or more of three locations.

- at home
- at an educational institution
- at work

	Age group	
Location of Internet use	**15–19 years**	**20–24 years**
At home	95%	95%
At an educational institution	85%	18%
At work	38%	74%

For the students surveyed, which one of the following statements, by itself, supports the contention that the location of Internet use is associated with the age group of the Internet user?

A 85% of students aged 15–19 years used the Internet at an educational institution.

B 95% of students aged 15–19 years used the Internet at home, but only 38% of 15–19 year olds used it at work.

C 95% of students aged 15–19 years used the Internet at home and 18% of 20–24 year olds used the Internet at an educational institution.

D The percentage of students who used the Internet at an educational institution decreased from 85% for those aged 15–19 years to 18% for those aged 20–24 years.

E The percentage of students who used the Internet at home was 95% for those aged 15–19 years and 95% for those aged 20–24 years.

[VCAA 2012 1CQ7]

Question 15

The level of water usage of 250 houses was rated in a survey as low, medium or high, and the size of the houses as small, standard or large. The results of the survey are displayed in the table below.

	Size of house		
Level of water usage	**Small**	**Standard**	**Large**
Low	15	14	9
Medium	22	71	11
High	15	47	46

The percentage of standard-sized houses rated as having a high level of water usage is

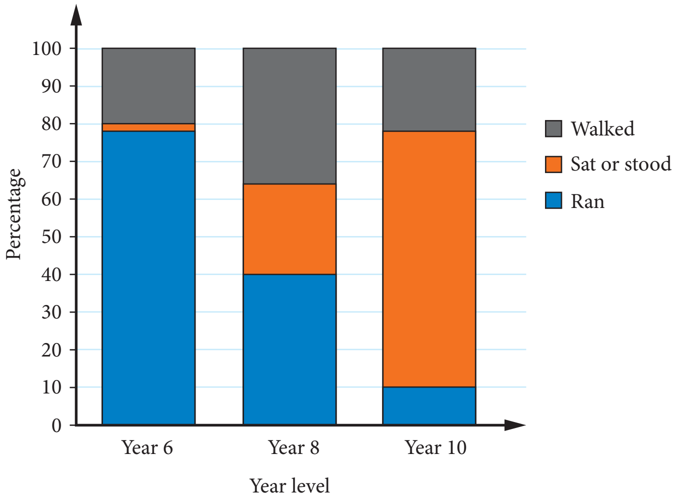
A 18.8% **B** 35.6% **C** 43.5% **D** 47.0% **E** 53.8%

[VCAA 2003 1CQ6]

Question 16

In a large survey, Years 6, 8 and 10 girls were asked what they did (walked, sat, stood, ran) for most of the time during a typical school lunch time. The results are displayed in the percentage segmented bar chart.

Does the percentage segmented bar chart support the opinion that, for these girls, the lunch time activity (walked, sat or stood, ran) undertaken is associated with Year level? Justify your answer by quoting appropriate percentages.



2 marks

[VCAA 2008 2CQ2]

Question 17

Table 1 shows information about a particular country. It shows the percentage of women, by age at first marriage, for the years 1986, 1996 and 2006.

Table 1

	Year of marriage		
Age of women at first marriage	1986	1996	2006
19 years and under	8.5%	3.7%	2.0%
20 to 24 years	42.1%	31.3%	21.5%
25 to 29 years	23.4%	31.7%	34.5%
30 years and over	26.0%	33.3%	42.0%

a Of the women who first married in 1986, what percentage were aged 20 to 29 years inclusive? 1 marks

b Does the information in Table 1 support the opinion that, for the years 1986, 1996 and 2006, the age of women at first marriage was associated with year of marriage?

Justify your answer by quoting appropriate percentages. It is sufficient to consider one age group only when justifying your answer. 2 marks

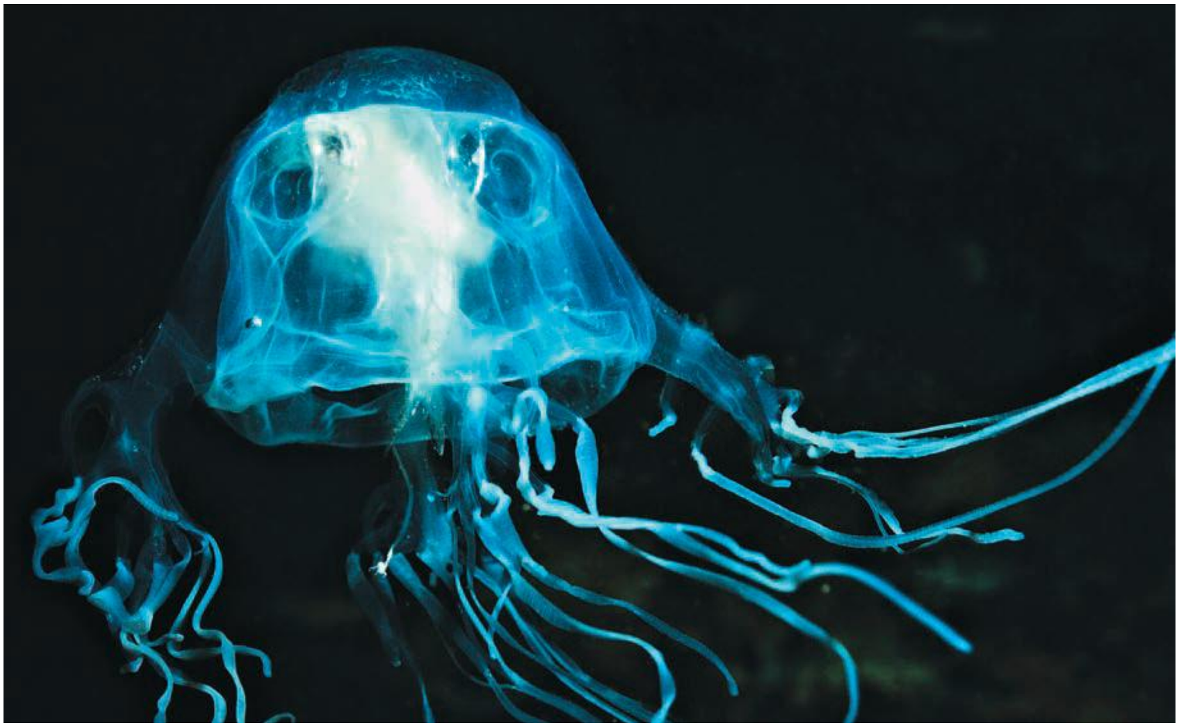
[VCAA 2011 2CQ2]

3.3 Associations between numerical and categorical variables

Comparing data

Comparing group measures

Calculating and interpreting statistics



Amana Images/Nature Connect

There are times when when we want to investigate the association between a numerical variable and a categorical variable by looking at how the two variables affect one another. To do this, we can use a back-to-back stem plot, parallel dot plot or parallel boxplot.

Back-to-back stem plots

Back-to-back stem plots are a good choice if you are dealing with just two sets of data values for the same variable and want to see the original data values. A back-to-back stem plot has two sets of leaves, one on the left of the stem and one on the right. This allows us to display the data for the two groups being compared, as in the example below.

Test results for two classes

Key: 5|6 = 56%

Class A	Stem	Class B
4 3 2	5	
8 5 4 3 1	6	0 0 2 4
9 6 4 4 3 1 0	7	3 5 6 7 8 8 8 9
8 8 7 5 3 2 0	8	0 1 4 6 9
1 0	9	2 3 5 6 7 7 8

largest value in this leaf

smallest value in this leaf

Exam hack

To comment on the shape of the data on the left side of a back-to-back stem plot, it may help to picture what it would look like if it was on the right side of a back-to-back stem plot.

Worked example 4

Liz and George deliver pamphlets to letterboxes. The number of pamphlets delivered per hour over 12 hours is shown below.

Liz: 32 24 27 35 28 26 32 28 31 35 32 25

George: 21 18 24 31 25 38 32 15 45 29 31 38

a Display the data with a back-to-back stem plot.

b Comment on the shape of the data for each person.

c Calculate the median, range and IQR for the number of pamphlets delivered by each person.

d Who would you say is the better delivery person and why?

Working

a As the data range is reasonably small, split the stems so that the shape of the data will be easier to interpret. Remember to order the values of the leaves.

George	Stem	Liz
8 5	1	
4 1	2	4
9 5	2	5 6 7 8 8
2 1 1	3	1 2 2 2
8 8	3	5 5
	4	
5	4	

Key: 4|5 = 45

b Comment on symmetry and skewness for each distribution.

Liz's data is approximately symmetrical whereas George's looks more negatively skewed.

c Calculate the median, range and IQR from the back-to-back stem plot.

Liz: median = 29.5, range = 11, IQR = 32 – 26.5 = 5.5

George: median = 30, range = 30, IQR = 35 – 22.5 = 12.5

d Use the results to decide who is the better delivery person.

The medians are about the same, but George's range and IQR are considerably higher than Liz's. This means Liz's deliveries have less variability and are more consistent than George's, which indicates that Liz is the better delivery person.

Parallel dot plots

Parallel dot plots can be used if you dealing with *two or more* sets of data values for the same variable and want to be able to read the original data values from the plot.

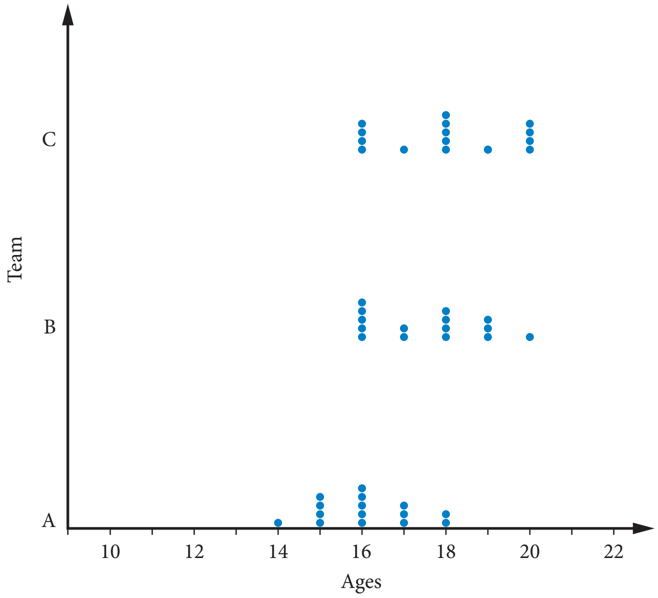
Worked example 5

The following parallel dot plots show the ages (in years) of three sporting teams in a competition.

a Describe the shape of each distribution.

b Calculate the median, range and IQR of each team.

c Do you agree with the statement 'Team A is at a disadvantage when playing the other two teams'? Justify your answer.



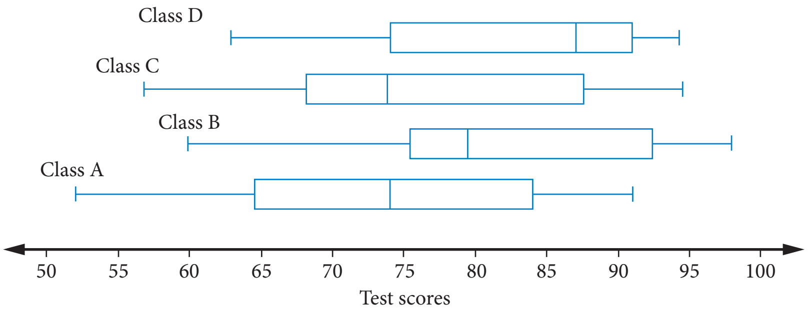
	Working
1 Comment on symmetry and skewness for each distribution.	Team A is approximately symmetrical. Team B appears positively skewed. Team C is symmetrical.
2 Calculate the median, range and IQR of each team from the parallel dot plots.	Team A: median = 16, range = 4, IQR = 2 Team B: median = 18, range = 4, IQR = 3 Team C: median = 18, range = 4, IQR = 4
3 Use the results to decide whether Team A may be at a disadvantage when playing the other two teams	The median age for Team A is lower than for Teams B and C. Performance in sport generally increases with age from 14 to 20, so a team with a median age of 16 may be at a disadvantage against teams with a median age of 18. Team A also has the lowest IQR of the three teams. This indicates that the middle 50% of its players' ages are concentrated around the low median, which also supports the view that Team A may be at a disadvantage against the other two teams.

Note that constructing parallel dot plots are straightforward on the TI-Nspire because dot plots are the default statistical graph. It is not possible to construct parallel dot plots on the current version of the ClassPad.

Parallel boxplots

Parallel boxplots are a good choice if you dealing with *two or more* sets of data values for the same variable, particularly when the data set is large. It is also easier to compare medians and quartiles from parallel boxplots than from back-to-back stem plots and parallel dot plots.

Here's an example based on the test results for four classes. It's relatively easy to find which class has the highest median, lowest Q_1 or highest maximum value etc.



Using CAS Constructing parallel boxplots

The test results of two Year 12 classes in Mathematical Methods are shown below.

Class A: 58 46 53 52 67 36 61 49 47 59 66 53 94 69 46 44 57

Class B: 60 50 70 69 86 43 60 60 44 56 49 50 56 56 42 65 47 67 25 46

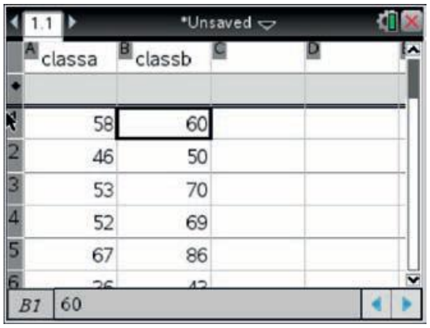
Construct parallel boxplots for the data.

TI-NSPIRE CAS

STEP 1

Open a New Document with a Lists & Spreadsheet page.

Type each set of data into a column.



	A classa	B classb	C	D
1	58	60		
2	46	50		
3	53	70		
4	52	69		
5	67	86		

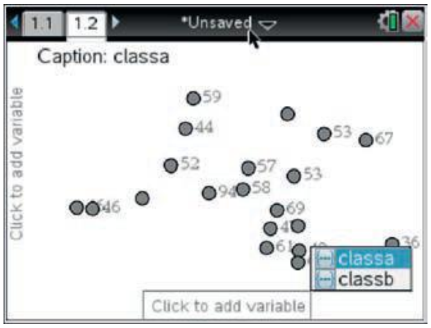
B1 60

STEP 2

Press ctrl Doc to add a Data & Statistics page.

Click in the 'Click to add variable' space at the bottom of the page.

Select classa and a dot plot will appear.

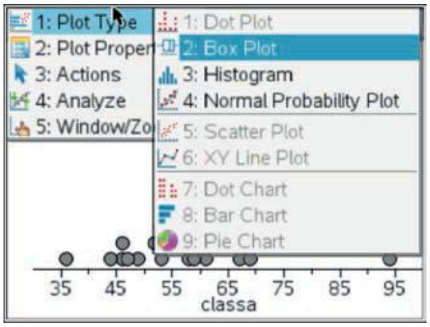


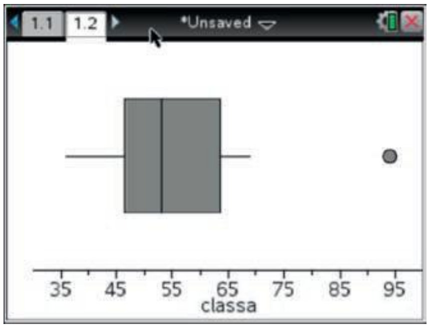
STEP 3

Press menu

1: Plot Type

2: Boxplot





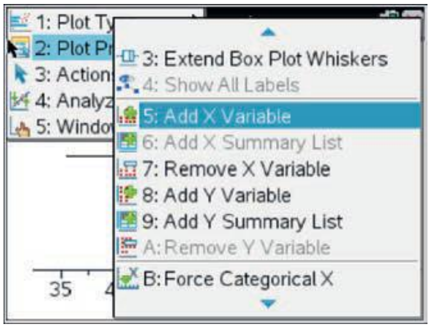
STEP 4

Press menu

2: Plot Properties

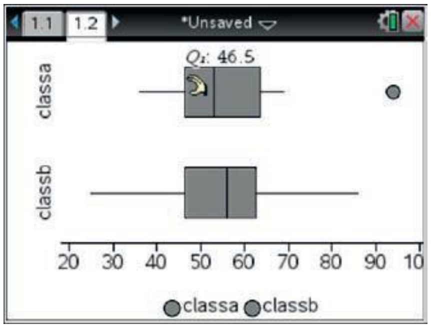
5: Add X Variable

Select classb.



STEP 5

Use the touch pad to move around the screen and the five-number summary values will appear.

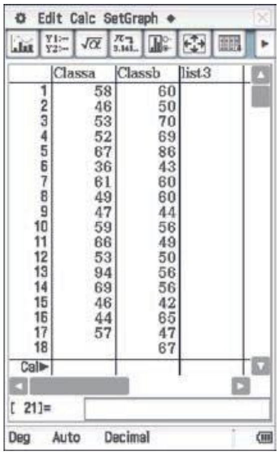


CLASSPAD

STEP 1

Use the mode.

Type each set of data into columns named Classa and Classb.



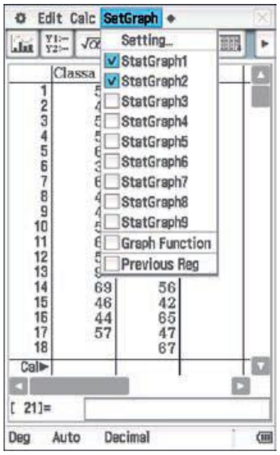
	Classa	Classb	list3
1	58	60	
2	46	50	
3	53	70	
4	52	69	
5	67	86	
6	36	43	
7	61	60	
8	49	60	
9	47	44	
10	59	56	
11	66	49	
12	53	50	
13	94	56	
14	69	56	
15	46	42	
16	44	65	
17	57	47	
18		67	

STEP 2

Tap **SetGraph**.

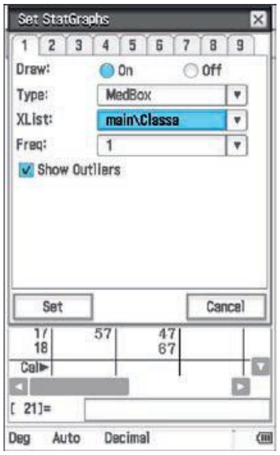
Make sure that StatGraph1 is checked and StatGraph2 is checked.

Tap **Setting**.



STEP 3

Complete the window as shown, then tap the tab numbered 2.



Set StatGraphs

Draw: ● On ○ Off

Type: MedBox

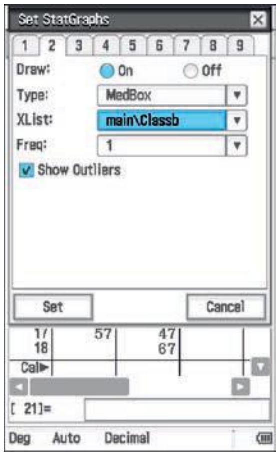
XList: main\Classa

Freq: 1

☑ Show Outliers

STEP 4

Complete the window as shown, then tap Set.



Set StatGraphs

Draw: ● On ○ Off

Type: MedBox

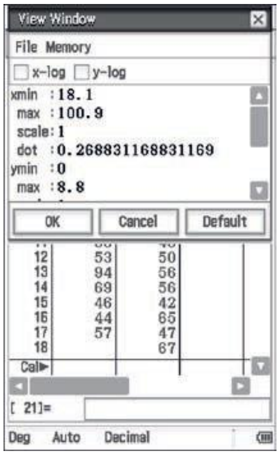
XList: main\Classb

Freq: 1

☑ Show Outliers

STEP 5

Tap ⚙ then **View Window** and set **xmin** and **max** below it and **ymin** and **max** below that as shown on the screen. Tap **OK**.



View Window

☐ x-log ☐ y-log

xmin : 18.1

max : 100.9

scale : 1

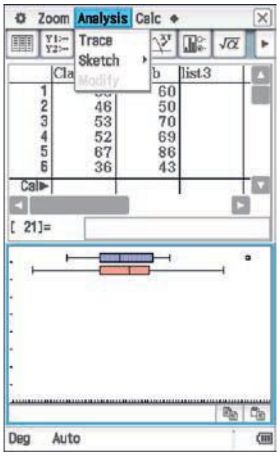
dot : 0.268831168831169

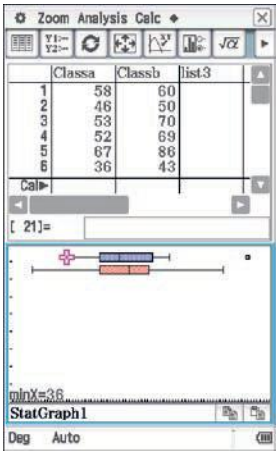
ymin : 0

max : 8.8

STEP 6

Tap 📊, then Analysis and Trace, use the arrow keys to move from one plot to the other and read the five-number summary from each boxplot, which appears at the bottom of the screen.





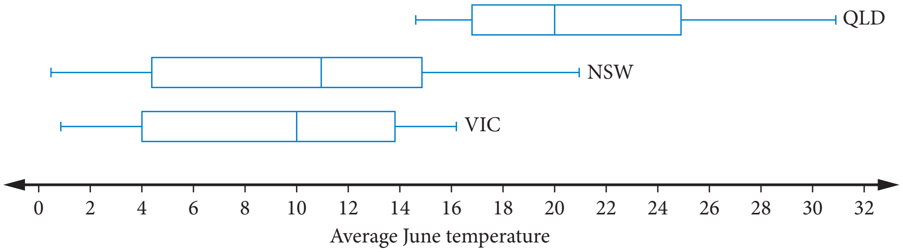
	Classa	Classb	list3
1	58	60	
2	46	50	
3	53	70	
4	52	69	
5	67	86	
6	36	43	

minX=36

StatGraph1

Worked example 6

The parallel boxplots below represent the average temperature in June for Victoria, New South Wales and Queensland over a number of years.



a Which state has the highest median average June temperature?

b Which state has the largest range of average June temperatures?

c Which state's data is best described as positively skewed?

d Which state had the lowest average June temperature?

e Write a brief summary comparing the average June temperatures for each state.

f Do you think there is an association between the states and their average June temperatures? Provide some evidence for your view.

		Working
a	Look for the state whose median line is furthest along the scale.	Queensland
b	Look for the longest boxplot, including whiskers.	New South Wales
c	Look for the boxplot with its median to the left of the box and a right whisker being longer than the left one.	Queensland
d	Look for lowest left endpoint.	New South Wales
e	Compare the states, noting similarities and differences.	Victoria and New South Wales have similar average June temperatures, whereas Queensland has higher average June temperatures than the other two states.
f	Are there differences?	There may be an association between the state and average June temperature as Queensland's data is significantly different from the other two states'. The middle 50% of the Queensland data has no overlap at all with the middle 50% of the NSW or Victorian data.

EXAM PREP 3.3

Associations between numerical and categorical variables

Prep 1

WORKED EXAMPLE 4

Two radar cameras that were positioned on different roads recorded car speeds (in km/h) as follows.

Camera 1:	78	63	75	69	71	83	80	67	74	72	73	74
	90	83	65	73	69	89	76	102	83	78	69	71
Camera 2:	112	139	120	116	116	136	140	123	135	131	120	117
	138	131	127	119	125	130	130	134	123	148	169	130

a Display the data with a back-to-back stem plot.

b Comment on the shape of the data for each road.

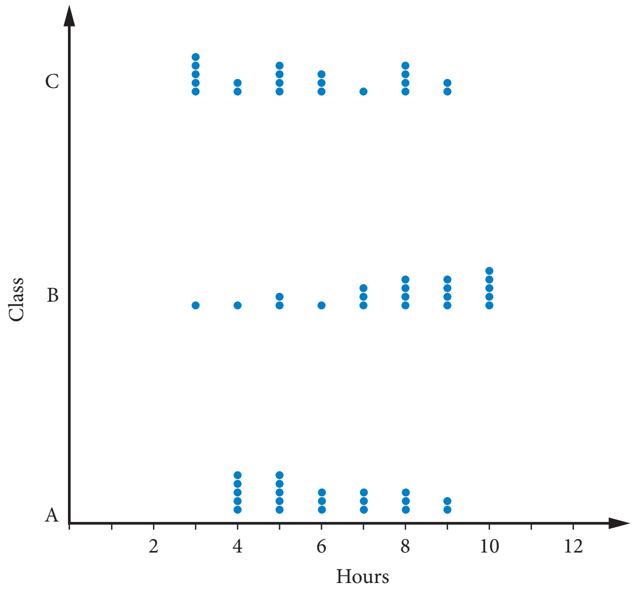
c Calculate the median, range and IQR for the car speeds recorded by each of the two radar cameras.

d Is there an association between the position of the cameras and the speed recorded? Justify your answer.

Prep 2

WORKED EXAMPLE 5

Three Year 12 classes were asked to record the number of hours they spent watching television during a school week. The results were presented as the following parallel dot plots.



a Describe the shape of each distribution.

b Calculate the median, range and IQR of each class.

c Do you agree with the statement ‘Class B watched more television during the week than Class A’? Justify your answer.

d Do you agree with the statement ‘Of the three classes, Class B has the greatest variation among the middle 50% of students’? Justify your answer.

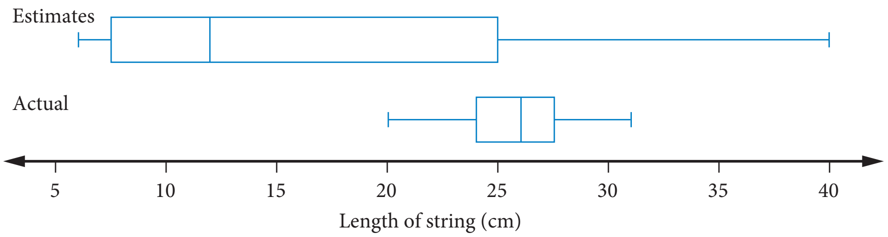
Prep 3

USING CAS: CONSTRUCTING PARALLEL BOXPLOTS

Construct parallel boxplots for the data in **Prep 1**.

Prep 4 WORKED EXAMPLE 6

Yashneel decided to investigate his theory that ‘People always underestimate the length of a piece of string’. He asked students to estimate the lengths of several pieces of string, measured the actual lengths and displayed the results in parallel boxplots.



a Write down the median of:

- **i** the estimated lengths
- **ii** the actual lengths

b Compare the range of each data set.

c Compare the interquartile range of each data set.

d Classify the shape of each data set.

e Do you agree with Yashneel’s theory? Justify your answer.

EXAM PRACTICE 3.3

Associations between numerical and categorical variables

Use the following information to answer Questions 1–3.

Smoking rates (%)

Female		**Male**
9 9 9 7 7 6 5	1	7 9
8 6 5 5 5 5 5 3 2 1 0	2	2 4 4 4 5 6 7 7 7
	3	0 0 1 1 6 9
	4	7

The back-to-back ordered stem plot shows the female and male smoking rates, expressed as a percentage, in 18 countries.

Question 1

For these 18 countries, the lowest female smoking rate is

A 5% **B** 7% **C** 9% **D** 15% **E** 19%

[VCAA 2009 1CQ1]

Question 2

For these 18 countries, the interquartile range (IQR) of the female smoking rates is

A 4 **B** 6 **C** 19 **D** 22 **E** 23

[VCAA 2009 1CQ2]

Question 3

For these 18 countries, the smoking rates for females are generally

A lower and less variable than the smoking rates for males.

B lower and more variable than the smoking rates for males.

C higher and less variable than the smoking rates for males.

D higher and more variable than the smoking rates for males.

E about the same as the smoking rates for males.

[VCAA 2009 1CQ3]

Use the following information to answer Questions 4–6.

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

The back-to-back **ordered** stem plot shows the distribution of maximum temperatures (in °Celsius) of two towns, Beachside and Flattown, over 21 days in January.

Question 4

The variables

temperature (°Celsius)

and

town (Beachside or Flattown)

are

A both categorical variables.

B both numerical variables.

C categorical and numerical variables respectively.

D numerical and categorical variables respectively.

E neither categorical nor numerical variables.

[VCAA 2006 1CQ1]

Question 5

For **Beachside**, the range of maximum temperatures is

A 3°C **B** 23°C **C** 32°C **D** 33°C **E** 38°C

[VCAA 2006 1CQ2]

Question 6

The distribution of maximum temperatures for **Flattown** is best described as

A negatively skewed.

B positively skewed.

C positively skewed with outliers.

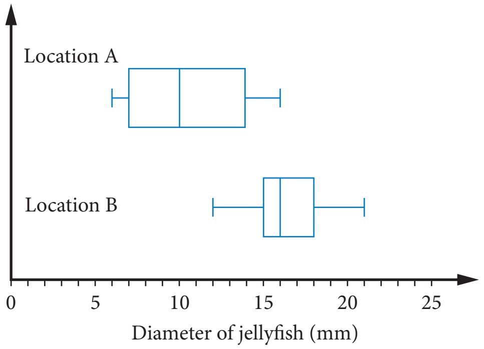
D approximately symmetric.

E approximately symmetric with outliers.

[VCAA 2006 1CQ3]

Use the following information to answer Questions 7 & 8.

Samples of jellyfish were selected from two different locations, A and B. The diameter (in mm) of each jellyfish was recorded and the resulting data is summarised in the boxplots shown.



Question 7

The percentage of jellyfish taken from location A with a diameter greater than 14 mm is closest to

A 2% **B** 5% **C** 25% **D** 50% **E** 75%

[VCAA 2007 1CQ5]

Question 8

From the boxplots, it can be concluded that the diameters of the jellyfish taken from location A are generally

A similar to the diameters of the jellyfish taken from location B.

B less than the diameters of the jellyfish taken from location B and less variable.

C less than the diameters of the jellyfish taken from location B and more variable.

D greater than the diameters of the jellyfish taken from location B and less variable.

E greater than the diameters of the jellyfish taken from location B and more variable.

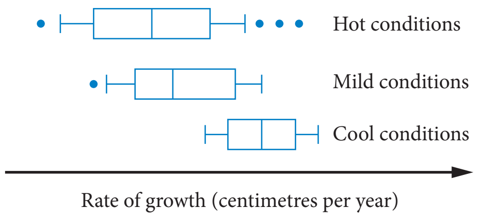
[VCAA 2007 1CQ6]

Question 9

As part of an experiment, three samples of pine trees were planted. Each sample contained 50 trees.

One sample was grown under hot conditions, one sample was grown under mild conditions and one sample was grown under cool conditions.

The parallel boxplots show the rate of growth (in centimetres per year) of these three samples.



From the parallel boxplots it can be concluded that, as conditions change from hot to mild to cool, the rate of growth for these trees

A decreases on average and becomes less variable.

B decreases on average and becomes more variable.

C does not change on average but becomes more variable.

D increases on average and becomes less variable.

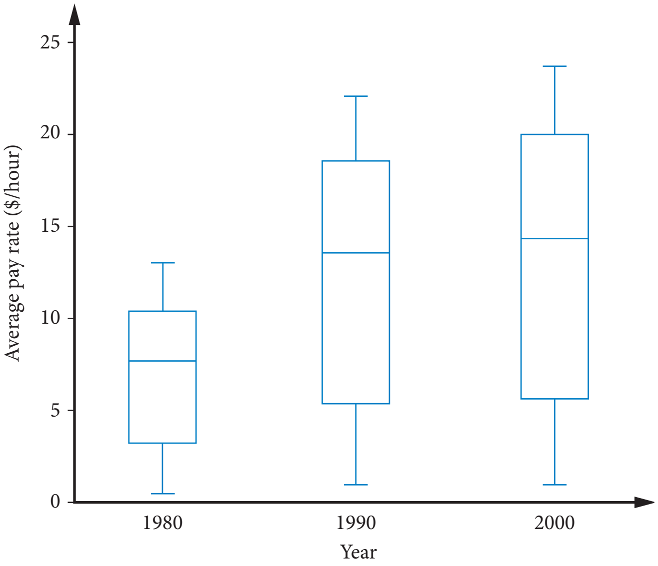
E increases on average and becomes more variable.

[VCAA 2005 1CQ7]

Question 10

The boxplots display the distribution of average pay rates, in dollars per hour, earned by workers in 35 countries for the years 1980, 1990 and 2000.

Based on the information contained in the boxplots, which one of the following statements is **not** true?



A In 1980, over 50% of the countries had an average pay rate less than $8.00 per hour.

B In 1990, over 75% of the countries had an average pay rate greater than $5.00 per hour.

C In 1990, the average pay rate in the top 50% of the countries was higher than the average pay rate for any of the countries in 1980.

D In 1990, over 50% of the countries had an average pay rate less than the median average pay rate in 2000.

E In 2000, over 75% of the countries had an average pay rate greater than the median average pay rate in 1980.

[VCAA 2011 1CQ5]

Question 11

Body mass index (BMI) is defined as

$$BMI = \frac{\text{mass}}{(\text{height})^2}$$ where *mass* is measured in kilograms and *height* in metres

a Determine the body mass index of a person who weighs 66 kg and who is 1.69 m tall. Write your answer correct to 1 decimal place. 1 mark

The BMI for each person in a sample of 17 males and 21 females is recorded in Table 1.

Table 1 BMI of males and females

Body mass index			
Males	Females	Males	Females
31.4	27.0	22.0	21.8
30.1	26.9	21.8	21.5
26.8	25.2	21.6	21.4
25.7	24.6	21.1	20.9
25.5	24.2	20.9	20.6
25.5	24.2	20.6	20.3
23.6	23.4	20.1	
23.3	23.4	19.9	
22.5	22.8	18.8	
22.4	22.5	17.5	
22.3	22.4		

b Write the range of the BMI data for **males** in the sample. 1 mark

c A BMI greater than 25 is sometimes taken as an indication that a person is overweight.

Use this criterion and the data in Table 1 to copy and complete Table 2, the two-way frequency table.

Table 2 Weight rating by gender

	Gender	
Weight rating	Male	Female
overweight		
not overweight		
Total	17	21

2 marks

d Does the data support the contention that, for this sample, weight rating is associated with gender? Justify your answer by quoting appropriate percentages. 2 marks

e The parallel boxplots in **Figure 1** have been constructed to compare the distribution of BMI for males and females in this sample.

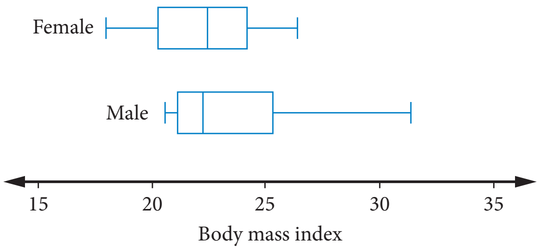


Figure 1 Parallel boxplots comparing BMI for males and females

i Use the parallel boxplots to identify and name two **similar** properties of the BMI distributions for males and females. 2 marks

ii Use the information in Table 1 to determine the mean BMI for the males in this sample. Write your answer correct to 1 decimal place. 1 mark

iii The median BMI for males is 22.5. Of the **mean** or **median**, which measure gives a better indication of the typical BMI for males? Explain your answer. 2 marks

[VCAA 2004 2CQ2]

Question 12

Table 1 below shows the heights (in cm) of three groups of randomly chosen boys aged 18 months, 27 months and 36 months respectively.

Table 1 Height (cm)

18 months	27 months	36 months
76.0	82.0	88.0
78.5	83.1	88.8
78.6	84.0	90.0
80.0	86.8	92.3
80.5	87.2	93.0
81.2	87.6	94.1
82.8	88.3	94.2
83.2	90.7	95.8
83.4	91.0	96.9
83.7	92.3	97.1
85.8	92.5	97.8
86.6	93.1	99.2
87.3	94.8	100.6
89.8	97.2	103.8

a Complete **Table 2** below by calculating the standard deviation of the heights of the **18-month-old** boys. Write your answer correct to 1 decimal place.

Table 2

Age	18 months	27 months	36 months
Mean	82.7	89.3	95.1
Standard deviation		4.5	4.5

1 mark

b A **27-month-old** boy has a height of 83.1 cm.

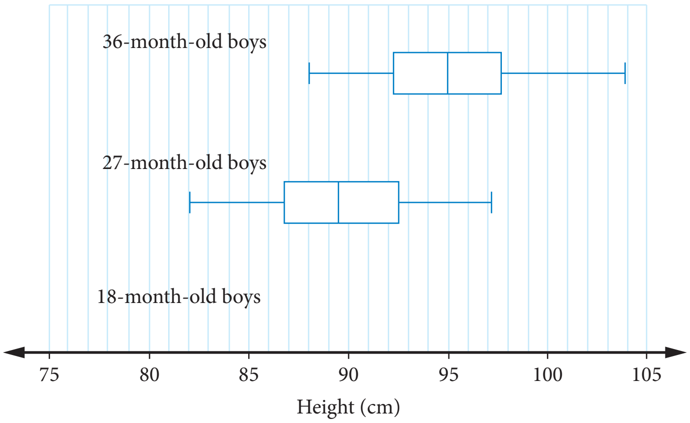
Calculate his standardised height (z-score) relative to this sample of 27-month-old boys. Write your answer correct to 1 decimal place. 1 mark

c The heights of the **36-month-old** boys are normally distributed. A 36-month-old boy has a standardised height of 2.

Approximately what percentage of 36-month-old boys will be shorter than this child? 1 mark

d Using the data from **Table 1**, boxplots have been constructed to display the distributions of heights of 36-month-old and 27-month-old boys as shown.

Copy and complete the display by constructing and drawing a boxplot that shows the distribution of heights for the **18-month-old** boys.



2 marks

e Use the appropriate boxplot to determine the median height (in centimetres) of the **27-month-old** boys. 1 mark

f The three parallel boxplots suggest that *height* and *age* (18 months, 27 months, 36 months) are **positively** related. 1 mark

Explain why, giving reference to an appropriate statistic.

[VCAA 2006 2CQ1]

Question 13

A weather station records the wind speed and the wind direction each day at 9:00 a.m. The wind speed is recorded, correct to the nearest whole number.

The parallel boxplots have been constructed from data that was collected on the 214 days from June to December in 2011.

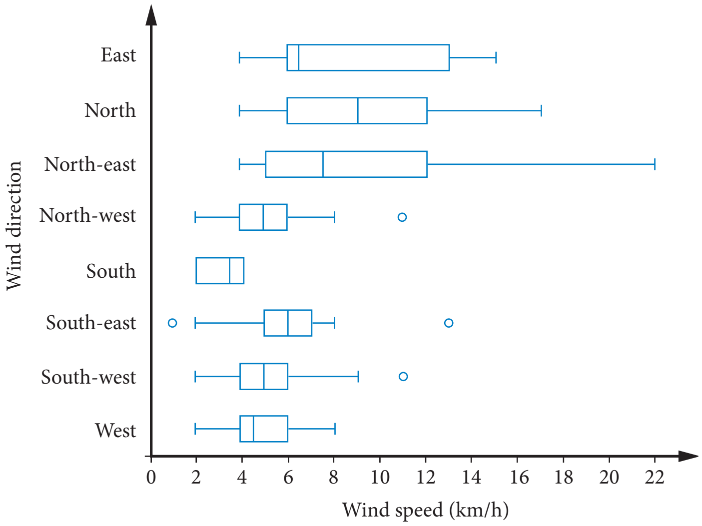
a Complete the following statements.

The wind direction with the lowest recorded wind speed

was ☐.

The wind direction with the largest range of recorded wind

speeds was ☐.



1 mark

b The wind blew from the south on eight days. Reading from the parallel boxplots we know that, for these eight wind speeds, the

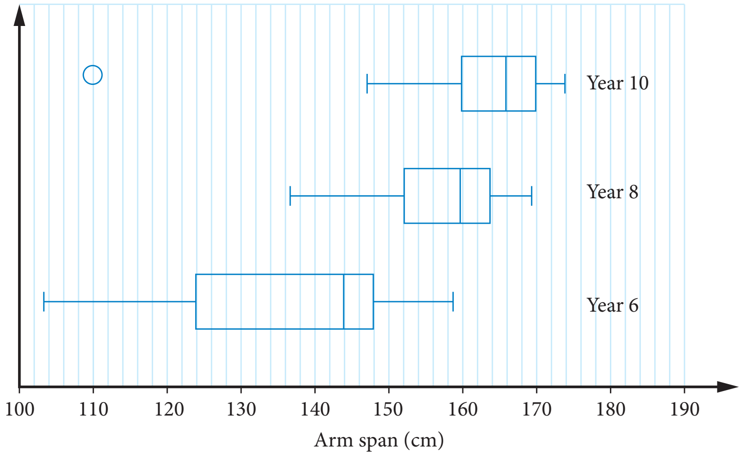
first quartile $Q_1 = 2$ km/h median $M = 3.5$ km/h third quartile $Q_3 = 4$ km/h

Given that the eight wind speeds were recorded to the nearest whole number, write down the eight wind speeds. 1 mark

[VCAA 2012 2CQ3]

Question 14

The arm spans (in cm) of Year 6, 8 and 10 girls were recorded in a survey. The results are summarised in the three parallel boxplots displayed.



a Complete the following sentence. The middle 50% of Year 6 students have an arm span

between ☐ and ☐ cm. 1 mark

b The three parallel boxplots suggest that arm span and Year level are associated. Explain why. 1 mark

c The arm span of 110 cm of a Year 10 girl is shown as an outlier on the boxplot. This value is an error. Her real arm span is 140 cm. If the error is corrected, would this girl's arm span still show as an outlier on the boxplot? Give reasons for your answer, showing an appropriate calculation.

2 marks

[VCAA 2008 2CQ3]

3.4 Associations between two numerical variables

WS A page of scatterplots

WS Height vs shoe size


iStock.com/AndreasG

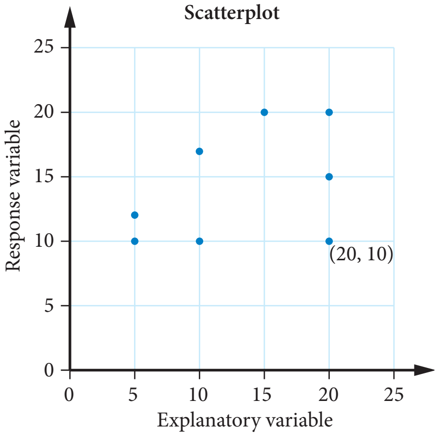
Scatterplots

A **scatterplot** is used when we wish to investigate the association between two numerical variables. It's constructed by plotting points onto a Cartesian plane where the horizontal or *x*-axis is used for the explanatory variable and the vertical or *y*-axis is used for the response variable.

To construct a scatterplot from a table

Explanatory variable (x)	5	5	10	10	15	20	20	20
Response variable (y)	10	12	10	17	20	10	15	20

1. Decide which variable is the explanatory variable (x) and which is the response variable (y).
2. Set up an appropriate scale for each axis and label them clearly with the variable names.
3. Plot each pair of data points from the table using a dot (•)



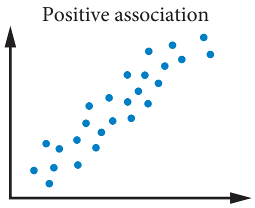
Exam hack

Don't assume that the top row of a table is the explanatory variable. The question to ask is 'Which variable do we expect to affect the other?'

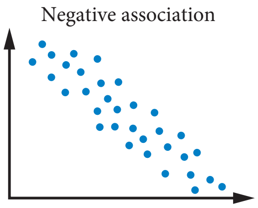
Interpreting scatterplots

There are three ways that scatterplots can be used to describe the association between two numerical variables.

1 Direction

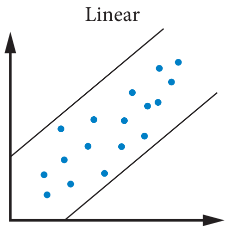


The direction of a positive association *rises* from left to right. This indicates that the *y* (response) variable tends to *increase* as the *x* (explanatory) variable increases.

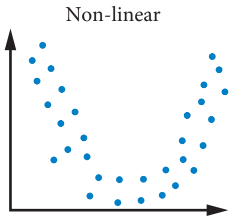


The direction of a negative association *falls* from left to right. This indicates that the *y* (response) variable tends to *decrease* as the *x* (explanatory) variable increases.

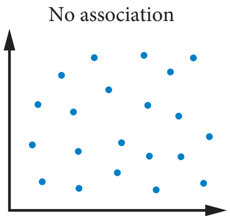
2 Form



Data in general follows a straight line pattern.

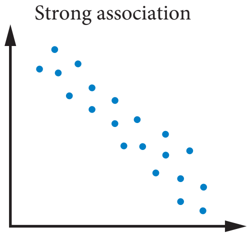


Data does not occur in a straight line pattern but does follow a curved pattern.

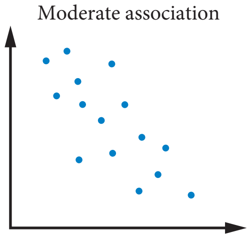


Data is randomly scattered and shows no pattern at all.

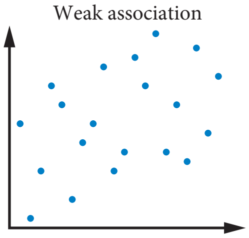
3 Strength



Data points are all reasonably close together.



Data points are more spread out than a strong association.



Data points are widely spread out.

Using CAS Constructing a scatterplot

The number of hot drinks sold at a milkbar per day and the day's maximum temperature were recorded over a period of 2 weeks.

Temperature	35	33	27	22	18	29	38	36	24	25	29	34	21	19
No. of drinks sold	25	28	29	54	76	48	18	39	42	61	36	49	68	53

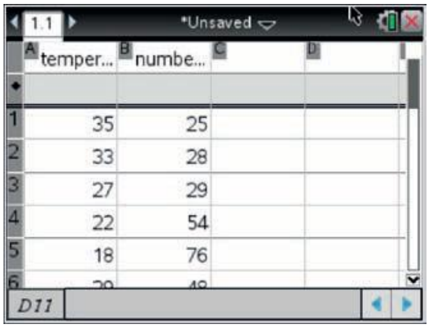
Use a CAS/calculator to construct a scatterplot for this information and describe the association shown in the scatterplot.

Decide first which is the explanatory variable and which is the response variable. Temperature affects the number of drinks sold, so temperature is the explanatory variable (*x*) and the number of drinks sold is the response variable (*y*).

TI-NSPIRE CAS

STEP 1

Enter data into the Lists & Spreadsheet page, remembering to name each column. It is common practice to enter the values for the explanatory variable in the first column and the response variable in the second column.



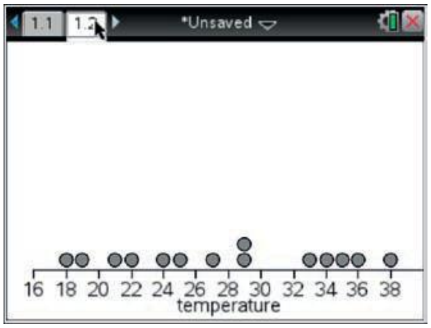
	A temper...	B numbe...	C	D
1	35	25		
2	33	28		
3	27	29		
4	22	54		
5	18	76		

Remember, ctrl 7 will take you back to the top cell

STEP 2

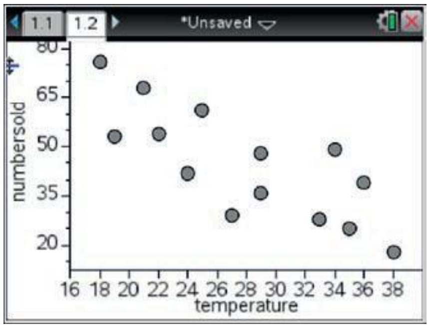
Press ctrl doc▾ and select 5: Add Data & Statistics.

Scroll down to Click to add variable, then click and select explanatory variable 'temperature' name from the menu.



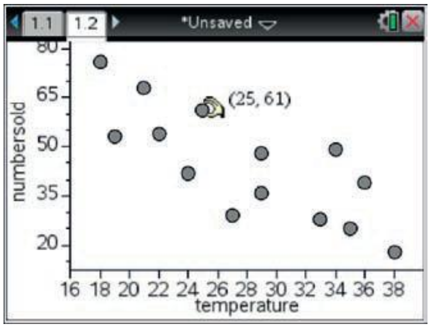
STEP 3

Scroll across to left side to Click to add variable, then click and select response variable 'numbersold' name from the menu.



STEP 4

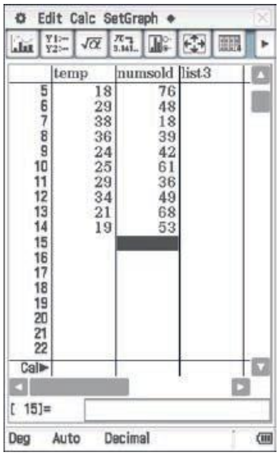
Move the hand around the screen to each dot and the coordinates of a point will appear.



CLASSPAD

STEP 1

Using the Statistics mode, rename columns and type in data. It is common practice to enter the values for the explanatory variable in the first column and the response variable in the second column.

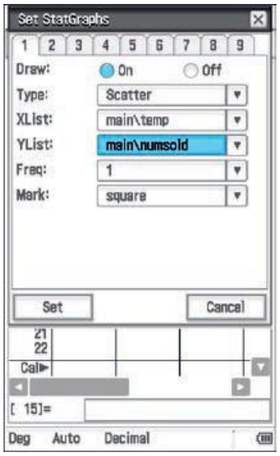


	temp	numsold	list3
5	18	76	
6	29	48	
7	38	18	
8	36	39	
9	24	42	
10	25	61	
11	29	36	
12	34	49	
13	21	68	
14	19	53	
15			
16			
17			
18			
19			
20			
21			
22			

[15]=

STEP 2

Tap **SetGraph**, making sure only one StatGraph is ticked, then tap **Setting…** Complete the screen as shown and tap Set.

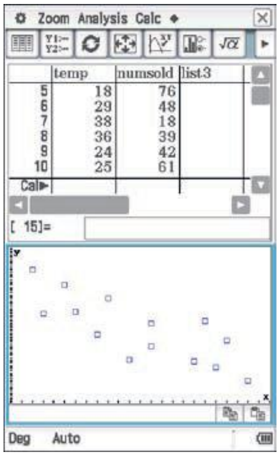


Draw: ● On ○ Off
Type: Scatter
XList: main\temp
YList: main\numsold
Freq: 1
Mark: square

[15]=

STEP 3

Tap to insert the graph.

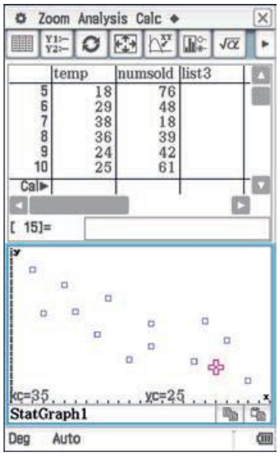


	temp	numsold	list3
5	18	76	
6	29	48	
7	38	18	
8	36	39	
9	24	42	
10	25	61	

STEP 4

If you tap **Analysis** then **Trace**, you can move from point to point using the left and right arrow buttons and the coordinates will appear at the bottom of the screen.

Note that trace goes through points in the order of points in the table.



	temp	numsold	list3
5	18	76	
6	29	48	
7	38	18	
8	36	39	
9	24	42	
10	25	61	

xc=35 yc=25

StatGraph1

The association between the day's maximum temperature and the number of hot drinks sold at a milkbar per day can be described as negative, linear and moderate.

Associations between two numerical variables

Prep 1

For each pair of variables, state whether they would generally have a positive association, a negative association or no association.

a height and weight of a person

b salary level and heart rate of an employee

c price of a brand of car and the number of that brand sold

d amount of effort studying for an exam and the exam score

e number of cigarettes smoked and incidence of lung cancer

f number of police on roads and number of speeding cars

g number of chocolate bars eaten and how many pets a person owns

h kilojoules consumed and weight gained

i alcohol consumption and reaction time

j height and the number of hours sleep a person has per night

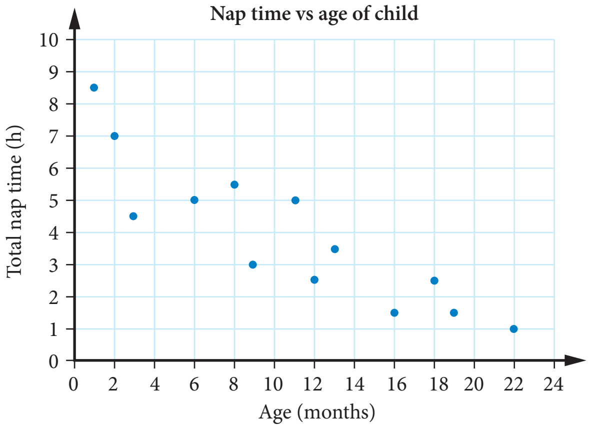
k temperature and number of people at a beach

l time spent travelling home and distance from home

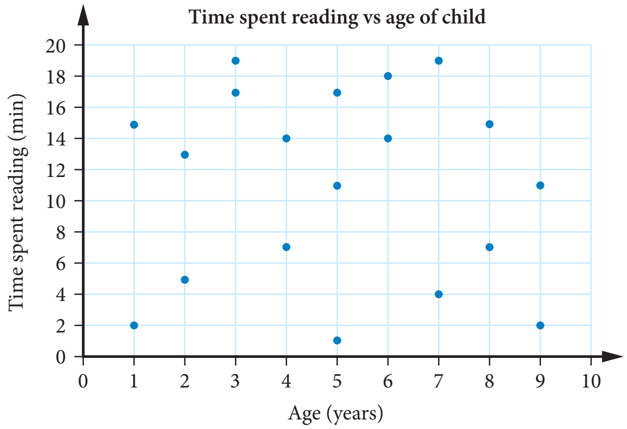
Prep 2

Describe each of the following scatterplots with reference to their direction, form and strength.

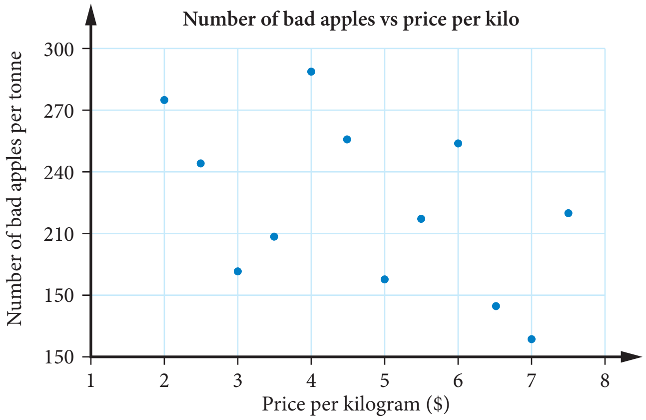
a



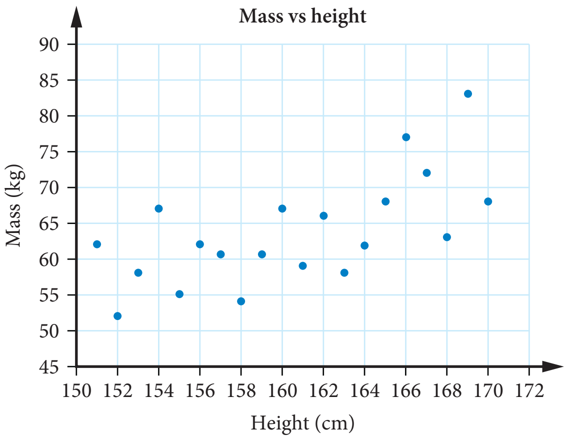
b


Time spent reading vs age of child
Time spent reading (min)
20
18
16
14
12
10
8
6
4
2
0
0 1 2 3 4 5 6 7 8 9 10
Age (years)

c


Number of bad apples vs price per kilo
Number of bad apples per tonne
300
270
240
210
150
150
1 2 3 4 5 6 7 8
Price per kilogram ($)

d


Mass vs height
Mass (kg)
90
85
80
75
70
65
60
55
50
45
150 152 154 156 158 160 162 164 166 168 170 172
Height (cm)

Prep 3 USING CAS: CONSTRUCTING A SCATTERPLOT

For each of the following use a CAS/calculator to construct a scatterplot for the information and describe the association shown in the scatterplot.

a The number of customers in a store over a two-week period was recorded and the results are shown below.

Time (days)	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Number of customers	76	92	63	79	81	64	102	71	86	119	86	78	111	92

b The hearing sensitivities (highest audible frequency measured in kilohertz) of different people from the same extended family were tested and the results are shown below.

Age	5	12	15	21	46	50	62	70	75
Frequency (kHz)	30	25	23	22	19	18	16	17	15

c People at a milkbar were asked how many chocolate bars they ate on average a month.

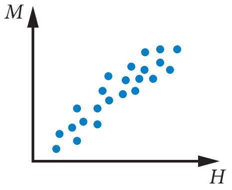
Age (years)	8	16	23	87	46	38	24	76	11	33	19	26	31	59	65	50
Number of chocolate bars	4	13	15	4	13	11	11	6	12	9	16	8	12	6	3	8

EXAMPLE PRACTICE 3.4

Associations between two numerical variables

Use the following information to answer Questions 1 & 2.

For the following scatterplot



Question 1

Which statement best describes the association between M and H?

A no association
B non-linear association
C negative, linear and weak
D positive, linear and weak
E positive, linear and strong

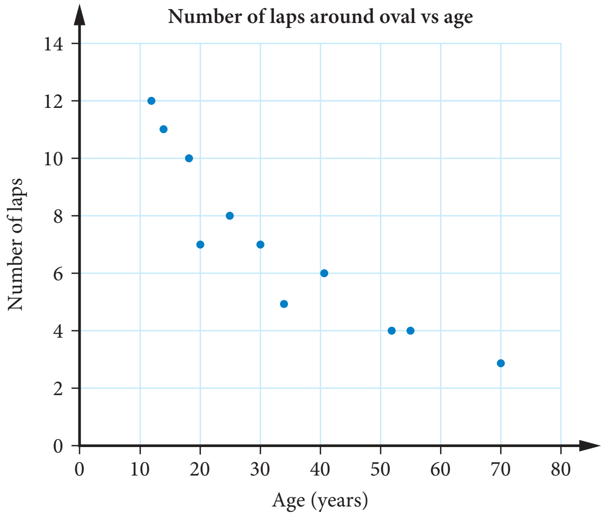
Question 2

Which of the following sentences is **correct**?

A It can be concluded that M should decrease as H increases.
B There is little evidence to show that H should increase as M increases.
C It can be concluded that M should increase as H increases.
D There is limited evidence to support that H will increase as M increases.
E There is some evidence to support that M will increase as H increases.

Question 3

Eleven people of different ages ran around an oval for 30 minutes. The number of laps they ran was recorded and the results are displayed in the scatterplot.



Which one of the following statements is **false**?

A The youngest person in the group ran the most laps.

B Age is the explanatory variable and number of laps is the response variable.

C The association between age and number of laps can be described as strong.

D The association between age and number of laps can be described as positive.

E The association between age and number of laps can be described as non-linear.

Question 4

The ages of both parents for students in a Year 12 class were recorded. Which of the following is true?

A The association between the two variables can be displayed using a scatterplot.

B The association between the two variables can be displayed using a histogram.

C The association between the two variables can be displayed using a percentaged segmented bar chart.

D This involves looking at the association between a numerical and a categorical variable.

E This involves looking at the association between two numerical variables.

Question 5

The association between the variables

size of car (1 = small, 2 = medium, 3 = large)

and

salary level (1 = low, 2 = medium, 3 = high)

is best displayed using

A a scatterplot. **B** a histogram. **C** parallel boxplots.

D a back-to-back stem plot. **E** a percentaged segmented bar chart.

[VCAA 2007 1CQ10]

3.5 The Pearson correlation coefficient

Scatterplots and correlation

Correlations matching game

iStock.com/Doctor_bass

Scatterplots and the Pearson correlation coefficient

The scale of a scatterplot can make a big difference to what conclusions we come to, so we need a systematic way of judging the association between variables. **The Pearson correlation coefficient**, r, (also called the **product-moment correlation coefficient**) is a number on a scale from –1 to 1 that measures the strength and direction of *linear* associations. The formula for calculating r is given below (although a CAS/calculator is usually used to find r).

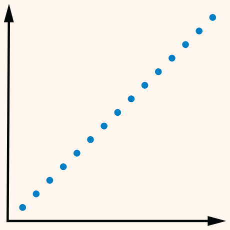
$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{(n-1)s_x s_y}$$

where $\bar{x}$ and s_x are the sample mean and standard deviation of the x values, and $\bar{y}$ and s_y are the sample mean and standard deviation of the y values.

It’s possible to estimate the value of *r* from the shape of the scatterplot using these guidelines:

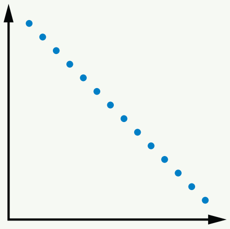
Perfect positive association

$r = 1$



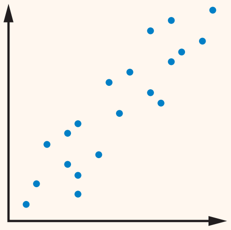
Perfect negative association

$r = -1$



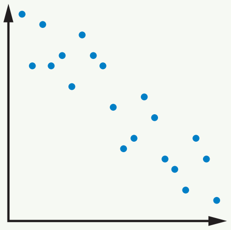
Strong positive association

$0.75 \le r < 1$



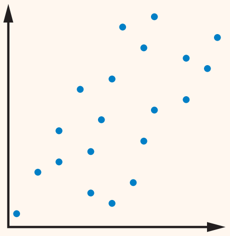
Strong negative association

$-1 < r \le -0.75$



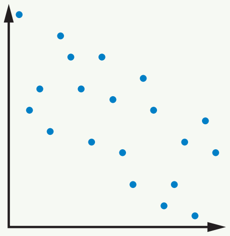
Moderate positive association

$0.5 \le r < 0.75$



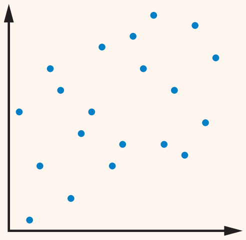
Moderate negative association

$-0.75 < r \le -0.5$



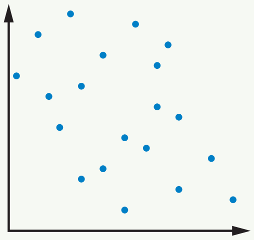
Weak positive association

$0.25 \le r < 0.5$



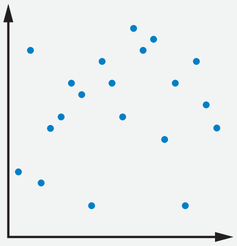
Weak negative association

$-0.5 < r \le -0.25$



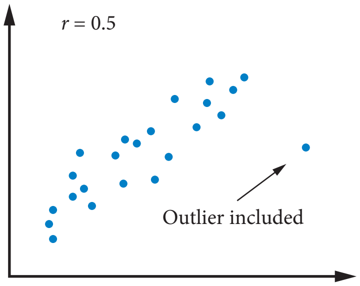
No association

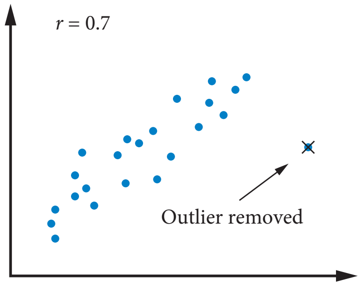
$-0.25 < r < 0.25$



Outliers and the Pearson correlation coefficient

Outliers can have a significant effect on the Pearson correlation coefficient, so it’s important to decide whether the outliers are wayward results that can be ignored, or whether they should be included in the calculation.





Using CAS Calculating the Pearson correlation coefficient

Calculate the correlation coefficient, correct to 2 decimal places, for the data in the table showing the number of umbrellas sold at a shopping centre and the amount of rain in the area recorded over 8 weeks. Discuss if there is an association between the number of umbrellas sold and the amount of rain.

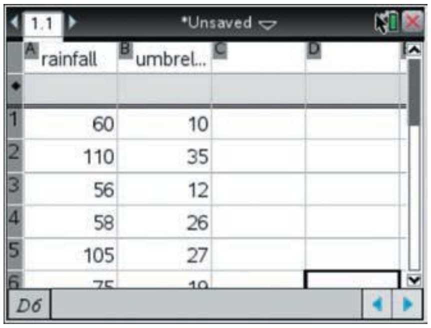
Umbrellas sold	10	35	12	26	27	19	17	18
Rainfall (mm)	60	110	56	58	105	75	48	90

Decide first which variable is the explanatory (*x*) and response variable (*y*). Rainfall affects the number of umbrellas sold, so rainfall is the explanatory variable and umbrellas sold is the response variable.

TI-NSPIRE CAS

STEP 1

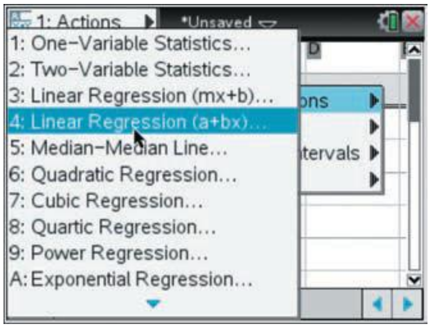
Enter the data into the Lists & Spreadsheet page, remembering to name each column.



	A rainfall	B umbrel...	C	D
1	60	10		
2	110	35		
3	56	12		
4	58	26		
5	105	27		

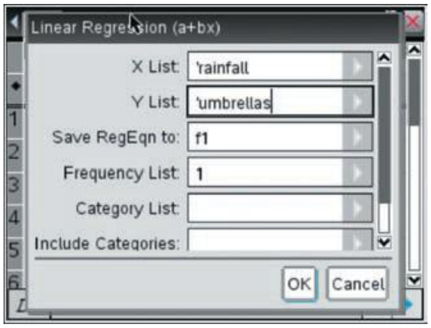
STEP 2

Press menu 4: Statistics, 1: Stat Calculations, 4: Linear Regression (a + bx).



STEP 3

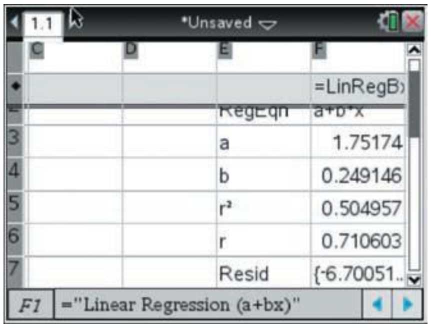
Select the X List Press ▶▶ and select 'rainfall'
Press TAB
Repeat for the Y list
Press ▶ and select 'umbrellas', then click on OK.



X List: 'rainfall
Y List: 'umbrellas
Save RegEqn to: f1
Frequency List: 1
Category List:
Include Categories:

STEP 4

Using the screen that appears, scroll down to find the *r* value. Give the answer correct to the specified number of places.



	E	F
		=LinRegB)
	RegEqn	a+b*x
3	a	1.75174
4	b	0.249146
5	r^2	0.504957
6	r	0.710603
7	Resid	{-6.70051...

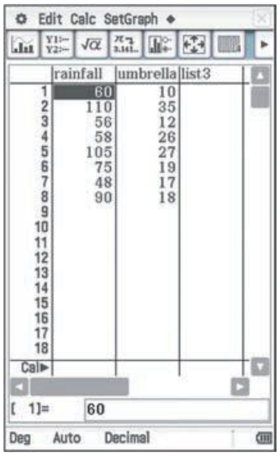
F1 ="Linear Regression (a+bx)"

$r \approx 0.71$

CLASSPAD

STEP 1

Using the Statistics mode, rename two columns and type in the data.



	rainfall	umbrella	list3
1	60	10	
2	110	35	
3	56	12	
4	58	26	
5	105	27	
6	75	19	
7	48	17	
8	90	18	

STEP 2

Tap Calc, then select regression, then **Linear Reg**.

Complete the screen as shown and tap OK.

Set Calculation

Linear Reg

XList: main\rainfall

YList: main\umbrella

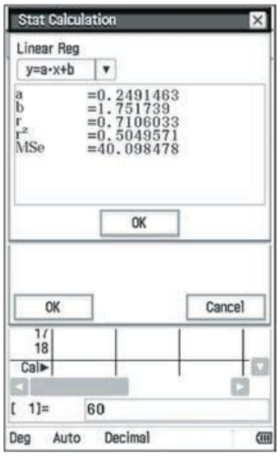
Freq: 1

Copy Formula: Off

Copy Residual: Off

STEP 3

A screen appears with the *r* value being the third value down. Give the answer correct to the specified number of places.



Stat Calculation

Linear Reg

y=a·x+b

```
a      =0.2491463
b      =1.751739
r      =0.7106033
r²     =0.5049571
MSe    =40.098478
```

$r \approx 0.71$

An *r* value of 0.71 indicates that there is a moderate, positive linear association between rainfall and the numbers of umbrellas sold. There is some evidence to suggest that the number of umbrellas sold should increase as rainfall increases.

Exam hack

Some sample sentences that could be used to interpret the correlation coefficient are shown below with the words ***y* variable** and ***x* variable**. When writing these sentences, insert the actual variable names being investigated.

Linear, positive and strong	**Linear, positive and moderate**	**Linear, positive and weak**
It can be concluded that the ***y* variable** should increase as the ***x* variable** increases.	There is some evidence to suggest that the ***y* variable** should increase as the ***x* variable** increases.	There is limited evidence to suggest that the ***y* variable** should increase as the ***x* variable** increases.
Linear, negative and strong	**Linear, negative and moderate**	**Linear, negative and weak**
It can be concluded that the ***y* variable** should decrease as the ***x* variable** increases.	There is some evidence to suggest that the ***y* variable** should decrease as the ***x* variable** increases.	There is limited evidence to suggest that the ***y* variable** should decrease as the ***x* variable** increases.

EXAM PREP 3.5

The Pearson correlation coefficient

Prep 1

USING CAS: CALCULATING THE PEARSON CORRELATION COEFFICIENT

At a local cinema, people were asked their yearly income and the number of times they had been to the movies in the past year. Calculate the correlation coefficient, correct to 2 decimal places, for the data in the table. Discuss if there is an association between income and the number of cinema movies seen in the past year.

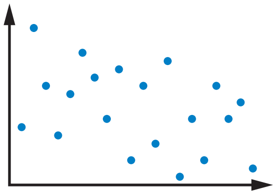
Income ($000 per year)	63	125	32	19	136	49	102	67	82	25	91	42
Number of times been to the movies in past year	4	12	1	8	16	13	13	6	7	6	12	10

Prep 2

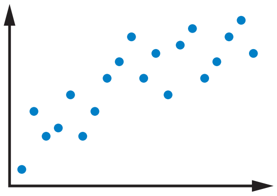
Match the following scatterplots with their correlation coefficients:

0.2, −0.9, −0.4, −0.5, 0.8, 0.6

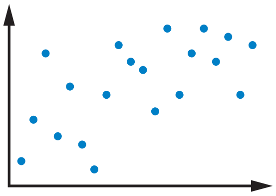
a



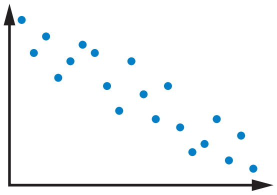
b



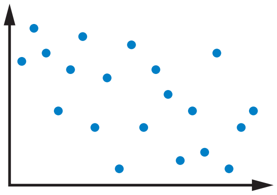
c



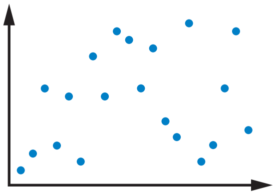
d



e



f



Prep 3

Interpret the correlation coefficient values and write a brief sentence describing the association for the studies described below.

a A study is investigating whether the amount of exercise a person does depends on their height. The correlation coefficient was found to be $r = 0.2946$.

b A study is investigating whether mass depends on the number of hours a person spends sitting down. The correlation coefficient was found to be $r = 0.5283$.

c A study is investigating whether the percentage of good peaches in a container depends on their number of weeks in storage. The correlation coefficient was found to be $r = -0.9108$.

d A study is investigating whether the number of sales at a local store depends on the temperature over a year. The correlation coefficient was found to be $r = 0.7241$.

e A study is investigating whether the number of pets owned depends on the average temperature over a year. The correlation coefficient was found to be $r = -0.0628$.

f A study is investigating whether the number of home-cooked meals per week depends on income. The correlation coefficient was found to be $r = -0.4637$.

The Pearson correlation coefficient

Use the following information to answer Questions 1 & 2.

Hours per week spent doing paid work	26	37	49	19	0	11	40	38	29	32	78	39	62
Hours per week spent doing unpaid work	4	10	11	8	13	6	7	9	7	6	2	11	9

Question 1

The correlation coefficient for this data is

A −0.4174 **B** −0.3402 **C** −0.0421 **D** 0.0421 **E** 0.3402

Question 2

Which of the following sentences best describes the association between the number of hours of paid work and the number of hours of unpaid work per week?

A There is limited evidence to suggest that the number of hours of unpaid work per week should increase as the number of hours of paid work increases.

B The number of hours of unpaid work per week should increase as the number of hours of paid work increases.

C There is some evidence to suggest that the number of hours of unpaid work per week should decrease as the number of hours of paid work increases.

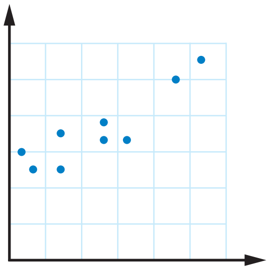
D There is limited evidence that to suggest that the number of hours of unpaid work per week should decrease as the number of hours of paid work increases.

E There is no association between the number of hours of paid work per week and the number of hours of unpaid work.

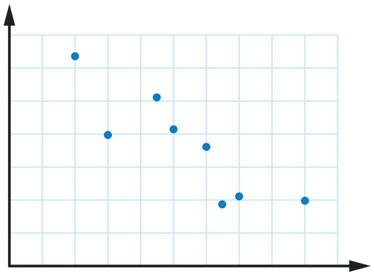
Question 3

Which of the following scatterplots have a correlation coefficient $r = 0.9218$?

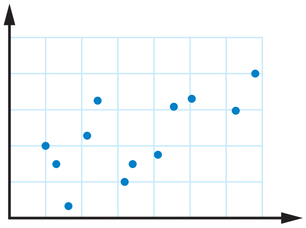
A



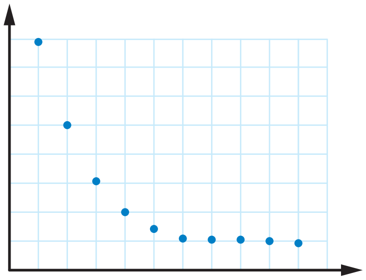
B



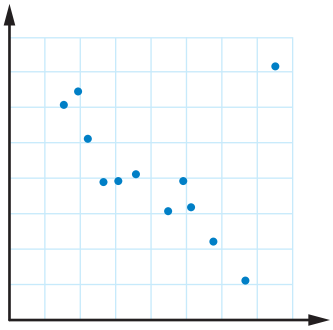
C



D



E



Question 4

The table below shows the hourly rate of pay earned by 10 employees in a company in 1990 and in 2010.

	Hourly rate of pay ($)	
Employee	**1990**	**2010**
Ben	9.53	17.02
Lani	9.15	16.71
Freya	8.88	15.10
Jill	8.60	15.93
David	7.67	14.40
Hong	7.96	13.32
Stuart	6.42	15.40
Mei Lien	11.86	19.79
Tim	14.64	23.38
Simon	15.31	25.11

The value of the correlation coefficient, *r*, for this set of data is closest to

A 0.74 **B** 0.86 **C** 0.92 **D** 0.93 **E** 0.96

[VCAA 2013 1CQ8]

Question 5

The following data was recorded from measurements made on 12 men.

Age (years)	Mass (kg)	Waist (cm)
26	84	84
29	72	74
32	67	89
32	59	75
34	97	106
37	112	114
39	67	80
40	91	101
41	98	101
43	89	94
45	117	126
51	62	82

The value of the correlation coefficient, r, for mass against waist measurement, is closest to

A 0.6061 **B** –0.7785 **C** 0.8675 **D** 0.9314 **E** 0.9651

[VCAA 2002 1CQ10]

Question 6

A large study of Year 12 students shows that there is a negative association between the time spent doing homework each week and the time spent watching television. The correlation coefficient is $r = –0.6$.

From this information it can be concluded that

A the time spent doing homework is 60% lower than the time spent watching television.

B 36% of students spend more time watching television than doing homework.

C the slope of the line of best fit is 0.6.

D if a student spends less time watching television, they will do more homework.

E an increased time spent watching television is associated with a decreased time doing homework.

[VCAA 2008 1CQ10]

3.6 Cause and effect

Shutterstock.com/Elpis Ionnidis

Correlation and causation

If two variables have a high correlation or association, it doesn't necessarily mean that one *causes* the other. Here are some examples.

The number of people each year from 2000 to 2009 who died by becoming tangled in their bedsheets in the USA and the total revenue generated by skiing facilities in the USA over the same years have a correlation coefficient of $r = 0.969\ 724$.

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Number of people who died by becoming tangled in their bedsheets Deaths (US) (CDC)	327	456	509	497	596	573	661	741	809	717
Total revenue generated by skiing facilities (US) Dollars in millions (US Census)	1551	1635	1801	1827	1956	1989	2178	2257	2476	2438

Sources: CDC & US Census http://www.tylervigen.com/

The per capita consumption of mozzarella cheese each year in the USA from 2000 to 2009 and the civil engineering doctorates awarded in the USA over the same years have a correlation coefficient of $r = 0.958\ 648$.

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Per capita consumption of mozzarella cheese (US) Pounds (USDA)	9.3	9.7	9.7	9.7	9.9	10.2	10.5	11	10.6	10.6
Civil engineering doctorates awarded (US) Degrees awarded (National Science Foundation)	480	501	540	552	547	622	655	701	712	708

Sources: USDA & National Science Foundation http://www.tylervigen.com/

Often when two variables have a high correlation, the cause is a third variable which is contributing to the changes in both of them. In the first example, it's likely that population increase in the USA is causing similar increases in both the number of people who died by becoming tangled in their bedsheets and the total revenue generated by skiing facilities.

In the second example, the consumption of mozzarella cheese is per capita (i.e. by each individual), so we can't explain it by the increase in the US population. Sometimes high levels of correlation occur by **coincidence** without any underlying cause.

Non-causal explanations

There are four possible explanations for an association between two variables:

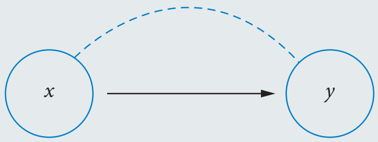
Explanation for association

- - - - - - - - - = association

⟶ = causation

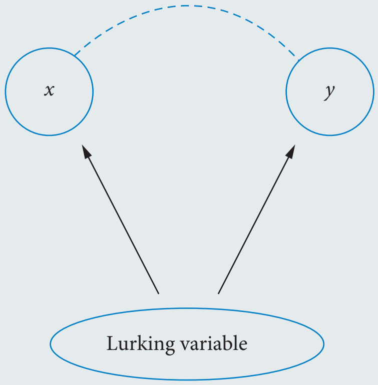
1. Causation

One variable is actually *causing* a change in the other.



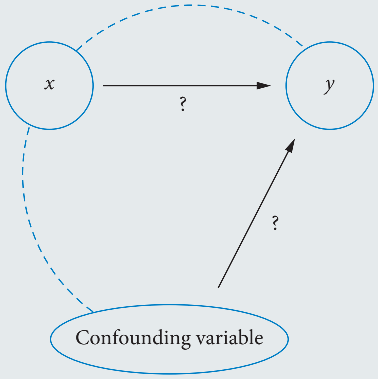
2. Common response

Some other factor (called a **lurking variable**) is causing both variables to change.



3. Confounding

It's unclear how the variables are related and we aren't able to draw conclusions about causation or common response. There may be an unknown factor (called a **confounding variable**) involved.

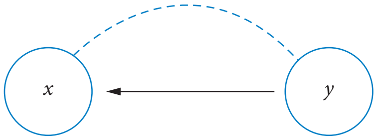


4. Coincidence

The association between the two variables is occurring completely by chance.

Exam hack

It is possible in an association to make the mistake of having the explanatory and response variables around the wrong way, so that what we thought at first was the response variable is actually causing the explanatory variable to change.



Observation and experimentation

Data can be gathered by either **observation** or **experimentation**. In an observational study the researcher passively observes an existing situation, while in an experimental study, the researcher actively manipulates a situation to eliminate possible confounding variables *before* observing it.

Suppose a researcher wanted to find out whether taking longer strides increases running speed.

An observational study could be to look at YouTube footage of Olympic 100 m runners, count the number of strides it takes them to complete the race, calculate the average stride length and compare this with their recorded speed.

An experimental study could be to randomly select 50 people, video and time them running 100 m races on a number of occasions under different conditions over a year, then count strides to calculate their average stride length and compare this with their speed.

The main reason for gathering data from experimentation rather than observation is to try to establish causation, i.e. to establish that the explanatory variable isn't just *associated* with the response variable, it's *causing* it.

Only an experimental study can definitively prove cause and effect. In the proposed stride length vs speed observational study, the results may only apply specifically to Olympic level runners who, unlike the average person, would have worked hard on stride technique over years. The results could also be influenced by variables such as training techniques and how wind moves in large stadiums.

Experimental studies eliminate variables that contribute to confounding by:

- controlling the factors that can be controlled
- randomisation
- repeating the experiment

Despite the advantages of experimental studies, we often have to use observational studies because doing experimental studies is impractical, unethical or too expensive. In the stride length vs speed example, it would clearly be far easier and more practical to simply look at YouTube videos. If we wanted to establish whether women who smoked through pregnancy gave birth to underweight babies, as another example, it would be unethical to set up an experiment that involved getting non-smoking pregnant women to smoke.

EXAM PREP 3.6

Cause and effect

Prep 1

State whether the most likely explanation for the association in each case is causation, common response, confounding, or coincidence.

a money spent on advertising and sales

b car speed and severity of car accident

c number of weddings in Victoria and the consumption of apples in Victoria

d number of weddings in Victoria and the per capita consumption of apples in Victoria

e number of books in home and academic success

f number people who drowned by falling into a swimming pool and number of films Jennifer Lawrence appeared in each year

g the number of firefighters called out to fight a blaze and the level of fire damage

h ice-cream sales and the number of people who drown

i the number of police and the number of violent crimes

Prep 2

A study shows there is an association between secondary school results and university results. Discuss in 50 words or less what sort of association this is likely to be.

EXAM PRACTICE 3.6

Cause and effect

Question 1

For a group of 15-year-old students who regularly played computer games, the correlation between the time spent playing computer games and fitness level was found to be $r = -0.56$.

On the basis of this information it can be concluded that

A 56% of these students were not very fit.

B these students would become fitter if they spent less time playing computer games.

C these students would become fitter if they spent more time playing computer games.

D the students in the group who spent a short amount of time playing computer games tended to be fitter.

E the students in the group who spent a large amount of time playing computer games tended to be fitter.

[VCAA 011 1CQ11]

Question 2

There is a high positive association between the number of laptops sold in a country and the life expectancy of people living in that country. Which of the following explanations is the most likely?

A causation

B common response

C confounding

D coincidence

E chance

Question 3

A worldwide study has shown that there is a high positive correlation between the number of cars a person owns and the age they live to. Which of the following lurking variables would best explain this common response association?

A the wealth a person has

B the number of cars a person owns

C the number of times a person lurks

D the age a person lives to

E the number of cars in a country

Question 4

A study has shown that there is a high negative association between the hours a primary student spends playing computer games and how well they do at school. Based on the study, which one of the following statements could be made?

A The level of parental supervision could be a confounding variable.

B Playing computer games will result in low school marks.

C Students with low marks hate school so they play computer games instead of doing homework.

D A common response lurking variable could be the number of hours the student plays computer games.

E This study indicates the association between the hours spent playing computer games and school performance is a coincidence.

Question 5

A controlled experiment involving a random selection of subjects has shown that people who use calorie-free sweeteners in place of sugar tend to gain weight. Which of the following is the most plausible explanation for this association?

A Calorie-free sweeteners are not actually calorie free.

B Calorie-free sweeteners cause people to gain weight.

C Calories don’t have anything to do with weight gain.

D There is another variable causing the association.

E The people who use calorie-free sweeteners are lying.

Question 6

Which of the following studies can best be described as an experimental study?

A The number of accidents at train level crossing across Victoria over a year are recorded to identify which crossings should be replaced by overpasses.

B The main meals ordered by people in a restaurant are recorded over a month to see which dishes could be cut from the menu to make room for new ones.

C The effect of a particular fertiliser on a fruit tree is tested by randomly selecting sections of orchard and using the fertiliser on those sections.

D To establish whether VCE students in Victoria would be happy for the school day to start and finish an hour later each day, all the VCE students at a school are surveyed.

E To establish whether eating breakfast affects school marks, a particular school asks its students if they eat breakfast and correlates this against their marks.

Question 7

Centuries ago, the people of the Hebrides, a chain of islands north of Scotland where head lice were common, were convinced that the head lice cured people who were sick with fever. They had noticed that while healthy people nearly always had head lice, sick people didn't have any.

a What is the causal link that the people of the Hebrides were making?

b Is this association based on observation or experimentation?

c What do you think was really happening to cause this association?

d Explain the mistake the people of the Hebrides were making in terms of explanatory and response variables.

e Briefly describe an experiment that you could set up to prove whether or not head lice cause people to be healthy. Comment on the ethics of your experiment.

SUMMARY

3

Associations between two variables

Practice quiz

Explanatory and response variables

- An explanatory variable is a variable that we expect to affect another variable.
- A response variable is a variable that we expect to be affected by another variable.

Two-way tables

- **Two-way frequency tables** can be used to look at the association between two categorical variables.
- **Percentaging** two-way frequency tables gives us more information. We usually percentage the explanatory variables.
- Information from a percentage two-way frequency table can be displayed as a **percentage segmented bar chart** where each cell in the table corresponds to a segment in the bar chart.

Graphs showing association between two variables

Explanatory variable	Response variable	Graph	Comments
Categorical	Categorical	Segmented bar chart	
Categorical	Numerical	Back-to-back stem plot	Lets you see original data, but only 2 categories are possible.
Categorical	Numerical	Parallel dot plot	Allows you to read original data from graph, but comparisons are not as easily made as with a parallel boxplot.
Categorical	Numerical	Parallel boxplot	Makes it easier to compare medians and quartiles.
Numerical	Numerical	Scatterplot	

Scatterplots

Three ways **scatterplots** can be used to describe the association between two numerical variables:

1. **Direction**: positive or negative
2. **Form**: linear, non-linear or no association
3. **Strength**: strong association, moderate association, weak association

9780170371100

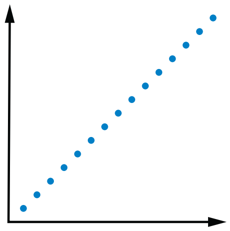
The Pearson correlation coefficient

- **The Pearson correlation coefficient**, r, is a number on the scale from −1 to 1 that measures the strength and direction of *linear* associations.

  It's possible to estimate the value of r from the shape of the scatterplot using these guidelines:

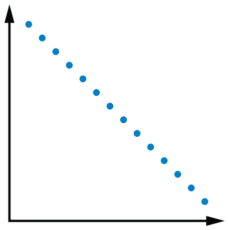
Perfect positive association

$r = 1$



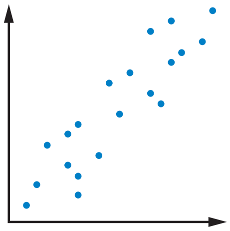
Perfect negative association

$r = -1$



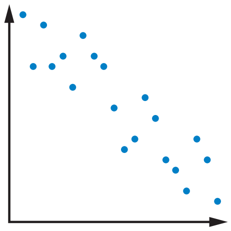
Strong positive association

$0.75 \leq r < 1$



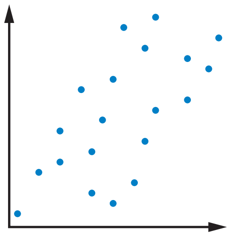
Strong negative association

$-1 < r \leq -0.75$



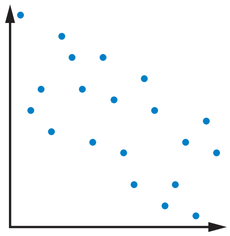
Moderate positive association

$0.5 \leq r < 0.75$



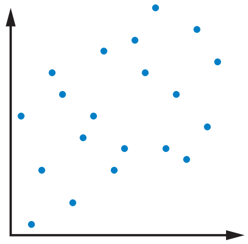
Moderate negative association

$-0.75 < r \leq -0.5$



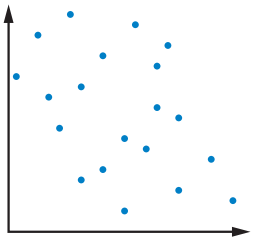
Weak positive association

$0.25 \leq r < 0.5$



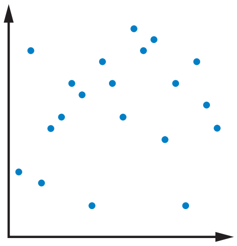
Weak negative association

$-0.5 < r \leq -0.25$



No association

$-0.25 < r < 0.25$



- Outliers can have a significant effect on the Pearson correlation coefficient.

Cause and effect

- Just because two variables have a high correlation or association, doesn't necessarily mean that one *causes* the other.

  The four possible explanations for an association between two variables are

  1. **Causation**: One variable is actually *causing* a change in the other.
  2. **Common response:** Some other factor (called a **lurking variable**) is causing both variables to change.
  3. **Confounding:** It's unclear how the variables are related and we aren't able to draw conclusions about causation or common response. There may be an unknown factor (called a **confounding variable**) involved.
  4. **Coincidence:** The association between the two variables is occurring completely by chance.

Observation or experimentation

- Data can be gathered by either **observation** or **experimentation**. In an observational study the researcher passively observes an existing situation, while in an experimental study the researcher actively manipulates a situation to eliminate possible confounding variables before observing it.
- Only an experimental study can definitively prove cause and effect, but sometimes an experimental study isn't possible for practical, expense or ethical reasons.

CHAPTER 4

LINEAR ASSOCIATIONS

Prior learning

4.1 The least squares line of best fit

Lines of fit

Least-squares regression line

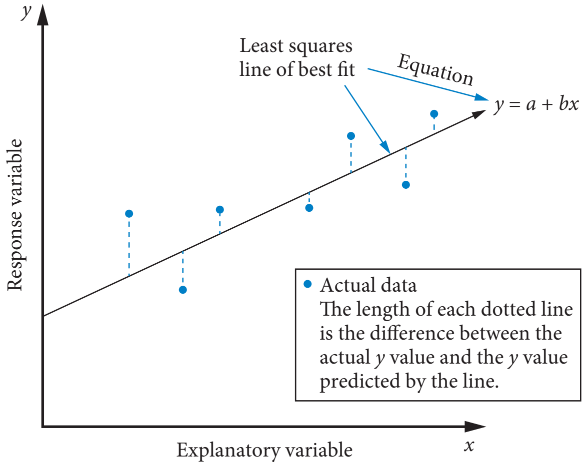
Getty Images/Torsten Blackwood

Line of best fit

A **line of best fit** is a straight line that is the best approximation for a set of data. It can be used to model the association between two numerical variables. The equation of the line functions as a simple formula which helps us understand the association and make predictions.

Least squares line of best fit

The **least squares line of best fit** (also known as a **least squares regression line**) is the most common method of finding a line of best fit to data. It involves finding the line that minimises the sum of the squares of the vertical distances (the dotted lines in diagram) between the line and each data point in a scatterplot. These vertical distances are the differences between the predictions made by the line and the actual data points, so it makes sense that we would want to minimise these prediction errors.



The *x*-axis variable is the explanatory variable and the *y*-axis variable is the response variable. When a least squares line of best fit equation is found, the equation should be written in terms of the variable names rather than using *x* and *y*.

The general form of the equation for the least squares line of best fit is $y = a + bx$ where

the **slope** or **gradient** of the line is $b = r\dfrac{s_y}{s_x}$

the ***y*-intercept** of the line is $a = \bar{y} - b\bar{x}$

And

- r is the Pearson correlation coefficient
- s_x and s_y are the sample standard deviations of x and y
- $\bar{x}$ and $\bar{y}$ are the sample means of x and y

We always calculate b first because b is needed in the calculation for a.

Exam hack

The values *a* and *b* in the least squares line of best fit equation are often called the '**coefficients**' of the equation. This is a general term for these values in an equation. Don't confuse this with the use of the same word in the term 'Pearson correlation coefficient'.

Rounding to significant figures

WS Significant figures

In problems involving least squares lines of best fit, you are often asked to round to a number of significant figures.

In the following examples, the significant figures are bolded:

170.0043	has seven significant figures
0.00**43**	has two significant figures
0.00**430**	has three significant figures
170	has two significant figures

Significant figures:

- Any non-zero digit
- Zeros between non-zero digits
  e.g. 21.005 has five significant figures and 3007 has four significant figures.
- Trailing zeros in decimals
  e.g. 62.78200 has seven significant figures and 3.00 has three significant figures.

Not significant figures:

- Leading zeros in decimals
  e.g. 0.0004 has one significant figure and 0.0305 has three significant figures.
- Trailing zeros in whole numbers
  e.g. 6340 has three significant figures and 50 000 has one significant figure.

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

- '0–4 round down' and '5–9 round up'
  e.g. 24 501 rounded to two significant figures is 25 000
- If 9s need to be rounded up, round the 9 to 0 and carry the rounding over to the next digit on the left.
  e.g. 3.9822 rounded to two significant figures is 4.0
  976.1 rounded to one significant figure is 1000

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

10 618.028 rounded to two decimal places is 10 618.03
10 618.028 rounded to two significant figures is 11 000.

In many real-life situations it makes more sense to round to significant figures than decimal places, particularly when a combination of large whole numbers and decimals is involved.

Exam hack

Occasionally the real-life context of a question requires a deliberate breaking of the rules about significant figures. For example, an odometer on a car could say you have travelled 73 999 kilometres, which obviously has 5 significant figures. However, if you travel one more kilometre, it would read 74 000 kilometres, which according to the rules has only 2 significant figures. Since the odometer usually gives 5 significant figures, we would say that 74 000 has 5 significant figures.

Worked example 1

Round each number to

i 2 decimal places

ii 2 significant figures

a 1.333

b 6.268

c 0.5563

d 14 700

e 10.882

f 76 008.037

g 12.8989

			Working
a	**i**	Focus on the first 2 decimal places.	**1.33**3 rounded to 2 decimal places is 1.33.
	ii	Focus on the first 2 significant figures.	**1.3**33 rounded to 2 significant figures is 1.3.
b	**i**		**6.26**8 rounded to 2 decimal places is 6.27.
	ii		**6.2**68 rounded to 2 significant figures is 6.3.
c	**i**		**0.55**63 rounded to 2 decimal places is 0.56.
	ii		**0.55**63 rounded to 2 significant figures is 0.56.
d	**i**	The rounded number must have 2 decimal places.	**14 700** rounded to 2 decimal places is 14 700.00.
	ii	Write zeros after the first 2 significant figures.	**14** 700 rounded to 2 significant figures is 15 000.
e	**i**		**10.88**2 rounded to 2 decimal places is 10.88.
	ii		**10**.882 rounded to 2 significant figures is 11.
f	**i**		**76 008.03**7 rounded to 2 decimal places is 76 008.04.
	ii		**76** 008.037 rounded to 2 significant figures is 76 000.
g	**i**		**12.89**89 rounded to 2 decimal places is 12.90.
	ii		**12**.8989 rounded to 2 significant figures is 13.

Worked example 2

The total runs scored (y) and the total balls faced (x) by a batsman over a cricket season were recorded and the values of the following statistics were determined:

Sample mean of the balls faced was 22.45

Sample standard deviation of the balls faced was 15.81

Sample mean of the runs scored was 14.91

Sample standard deviation of the runs scored was 11.74

Pearson correlation coefficient was 0.89

Calculate the least squares line of best fit that models this data, rounding the coefficients a and b to 2 significant figures.

	Working
1 Write out the values in terms of the formula.	$\bar{x} = 22.45, s_x = 15.81, \bar{y} = 14.91, s_y = 11.74, r = 0.89$
2 Calculate the slope (b) using the formula and round to 2 significant figures.	$b = r \times \frac{s_y}{s_x}$ $= 0.89 \times \frac{11.74}{15.81}$ ≈ 0.66
3 Calculate the y-intercept (a) using the formula and round to 2 significant figures.	$a = \bar{y} - b\bar{x}$ $= 14.91 - 0.66088 \times 22.45$ ≈ 0.073
4 Write the equation for the least squares line of best fit.	$a = 0.073$ and $b = 0.66$ $y = 0.073 + 0.66x$
5 Replace y and x in the equation with the correct variable names.	runs scored = 0.073 + 0.66 × balls faced

It is important to use the unrounded answer for b when calculating the value of a.

Exam hack

When finding the equation of the least squares line of best fit, you usually need to decide which is the explanatory (i.e. x) variable and which is the response (i.e. y) variable. **Don't** assume that first variable listed is always the explanatory variable.

Using CAS Finding the least squares line of best fit equation

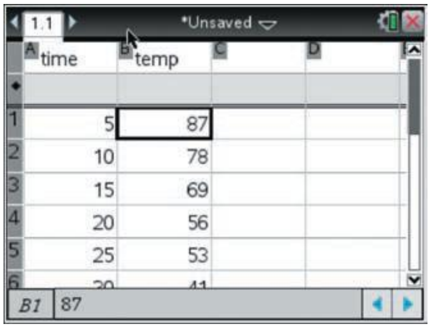
The following table shows the results of an experiment that measures the temperature of a liquid as it cools down after the source of heat is removed. Find the equation of the least squares line of best fit, rounding the coefficients *a* and *b* correct to 2 significant figures.

Time (minutes)	5	10	15	20	25	30
Temperature (°C)	87	78	69	56	53	41

TI-NSPIRE CAS

STEP 1

Time is the explanatory variable and Temperature is the response variable. Enter the data into a Lists & Spreadsheet page, remembering to name each column.

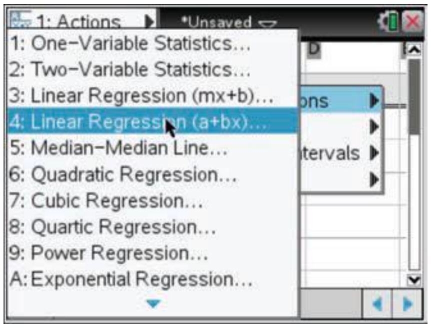


	A time	B temp	C	D
1	5	87		
2	10	78		
3	15	69		
4	20	56		
5	25	53		

B1 87

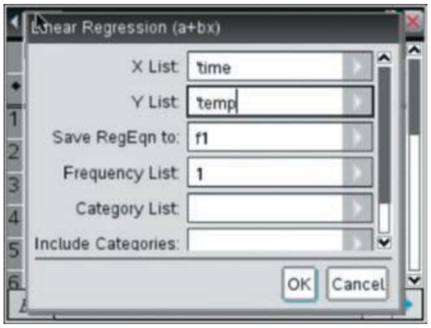
STEP 2

Press menu, then 4: Statistics, 1: Stat Calculations, 4: Linear Regression (a + bx).



STEP 3

Select X List, Press ▶▶, Select 'time', Press tab, select Y List, Press ▶, Select 'temp' then click on OK.



X List: time

Y List: temp

Save RegEqn to: f1

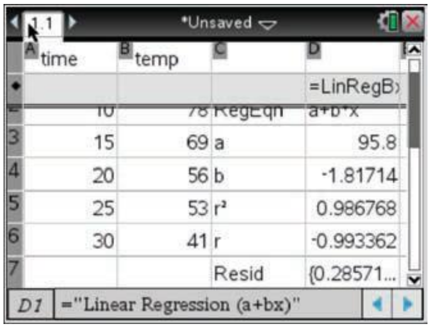
Frequency List: 1

Category List:

Include Categories:

STEP 4

A screen appears. Scroll down to find the *a* and *b* values.



	A time	B temp	C	D
◆				=LinRegBx
3	15	69	a	95.8
4	20	56	b	-1.81714
5	25	53	r²	0.986768
6	30	41	r	-0.993362
7			Resid	{0.28571...

D1 ="Linear Regression (a+bx)"

STEP 5

Round to 2 significant figures. $y = a + bx$

Write the equation. $a = 96$ and $b = -1.8$

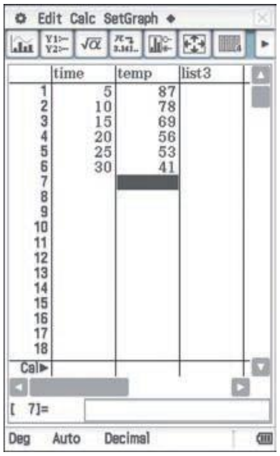
Replace *y* and *x* in the equation with the correct variable names. $y = 96 - 1.8x$

temperature = 96 – 1.8 × time

CLASSPAD

STEP 1

Time is the explanatory variable and Temperature is the response variable. Using the **Statistics** application, rename two columns and type in the data.

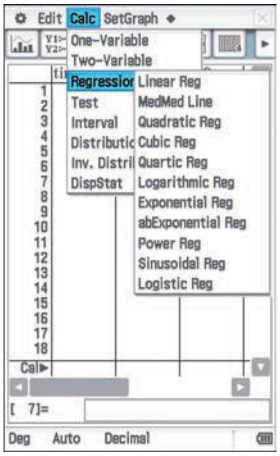


	time	temp	list3
1	5	87	
2	10	78	
3	15	69	
4	20	56	
5	25	53	
6	30	41	

[7]=

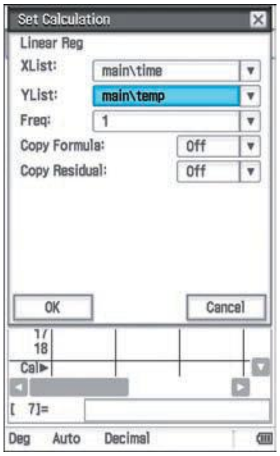
STEP 2

Tap **Calc**, then **Regression**, then select **Linear Reg**.



STEP 3

Complete the screen as shown and tap OK.



Linear Reg

XList: main\time

YList: main\temp

Freq: 1

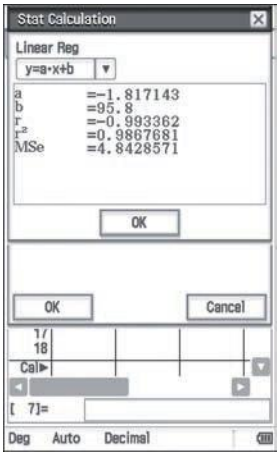
Copy Formula: Off

Copy Residual: Off

STEP 4

A screen appears showing the *a* and *b* values.

IMPORTANT NOTE: the ClassPad gives the least squares line of best fit in the form $y = ax + b$, so the gradient is *a* and the y-intercept is *b*.



Linear Reg

y=a·x+b

a =−1.817143
b =95.8
r =−0.993362
r² =0.9867681
MSe =4.8428571

STEP 5

Round to 2 significant figures.	$y = ax + b$
Write the equation.	$a = -1.8$ and $b = 96$
Replace *y* and *x* in the equation with the correct variable names.	$y = -1.8x + 96$ temperature = −1.8 × time + 96

Using CAS Drawing the least squares line of best fit

Draw the least squares line of best fit for the data below which measures the temperature of a liquid as it cools down over time after the source of heat is removed.

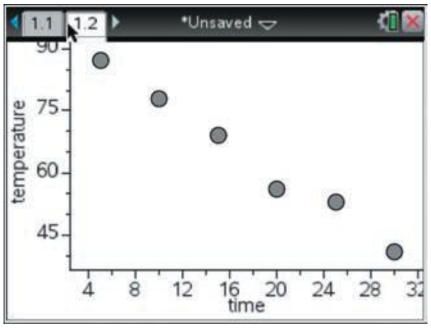
Time (minutes)	5	10	15	20	25	30
Temperature (°C)	87	78	69	56	53	41

Time affects temperature, so time is the explanatory (i.e. *x*) variable.

TI-NSPIRE CAS

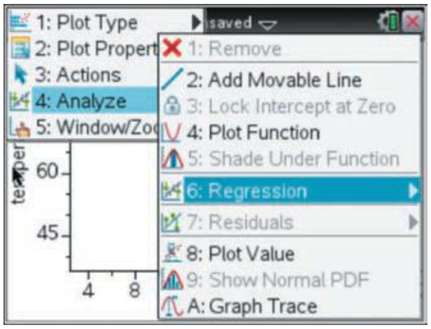
STEP 1

Create a scatterplot.



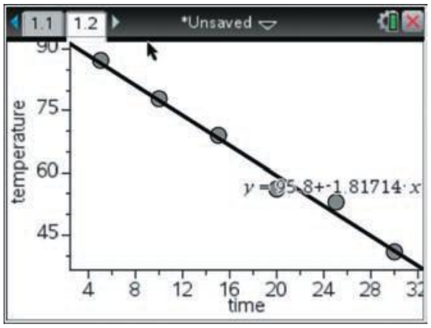
STEP 2

To add the least squares line of best fit, press menu 4: Analyze, 6: Regression, 2: Show linear(a + bx).



STEP 3

The least squares line of best fit is now plotted on the scatterplot and the equation is shown.



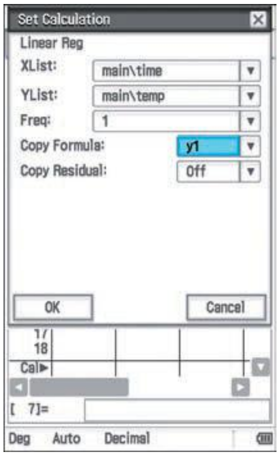
CLASSPAD

STEP 1

Using the **Statistics** application, enter the data into two lists.

Tap **Calc**, then select **Regression** then **Linear Reg** and complete the screen as shown.

Make sure that you select y1 in the Copy Formula menu, then tap OK.



Set Calculation

Linear Reg

XList: main\time

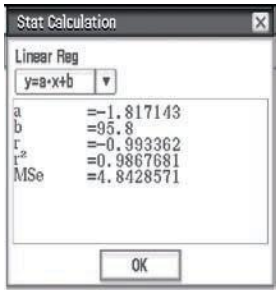
YList: main\temp

Freq: 1

Copy Formula: y1

Copy Residual: Off

The line of best fit with its equation will appear.



Stat Calculation

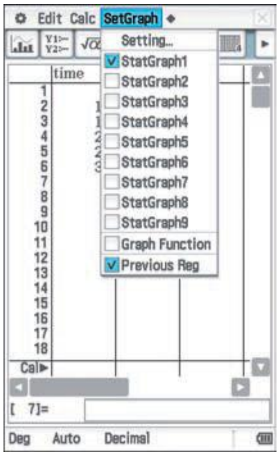
Linear Reg

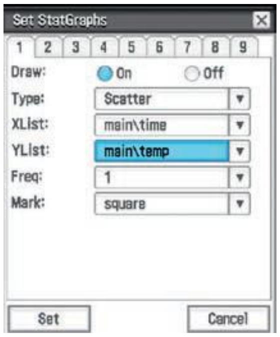
y=a·x+b

a =−1.817143
b =95.8
r =−0.993362
r^2 =0.9867681
MSe =4.8428571

STEP 2

Create a scatterplot using **SetGraph**. The drop-down menu should look like the one below. Tap **Setting** set Type to Scatter and fill in the box to plot temp against time, then tap Set and to draw the graph.





Set StatGraphs

Draw: ● On ○ Off

Type: Scatter

XList: main\time

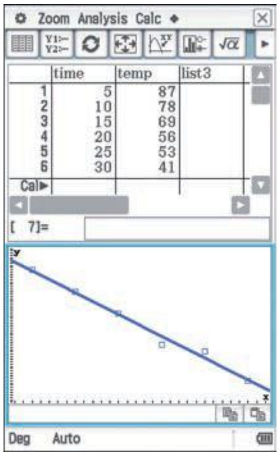
YList: main\temp

Freq: 1

Mark: square

STEP 3

The least squares line of best fit will be plotted on the scatterplot. Remember to tap Resize for a full page if needed.



	time	temp	list3
1	5	87	
2	10	78	
3	15	69	
4	20	56	
5	25	53	
6	30	41	

Interpreting the least squares line of best fit

For the previous example the equation of the least squares line of best fit line was:

temperature = –1.8 × time + 96

The y-intercept is 96. This means the temperature was 96°C when the source of heat was removed.

The slope is –1.8. This tells us that the temperature *decreases* by 1.8°C for every 1 minute increase in time.

We interpret the equation of a least squares line of best fit $y = a + bx$ by saying:

- The y-intercept is a. This means the **y variable** is a **units** when the **x variable** is zero **units**.
- The slope is b. This means the **y variable** increases/decreases by b **units** for every 1 **unit** increase in the **x variable**.

Use the word 'increases' when b is positive and 'decreases' when b is negative.

Replace the words in bold with the appropriate variable names or units of measure for each variable.

Worked example 3

For the following least squares line of best fit equation

hand span = 2.1 + 0.2 × height

where hand span and height are measured in cm

a identify and interpret the slope

b identify and interpret the y-intercept and comment on your result.

	Working
a The slope is the value that height is multiplied by in the equation. Interpret the slope in terms of the variables and units.	The slope is 0.2. This means hand span increases by 0.2 cm for every 1 cm increase in height.
b Identify and interpret the y-intercept and comment.	The y-intercept is 2.1. This means that hand span is 2.1 cm when height is 0 cm. A person of zero height can't have a 2.1 cm hand span. This least squares line of best fit only applies from a certain minimum height.

EXAM PREP 4.1

The least squares line of best fit

Prep 1 WORKED EXAMPLE 1

Round each number to

i 2 decimal places **ii** 2 significant figures

a 7.421 **b** 12.919 **c** 0.363 **d** 1800

e 20.666 **f** 72.0037 **g** 9.7962

Prep 2 WORKED EXAMPLE 2

The annual rainfall (x) and annual number of bushfires (y) were recorded for a region and the values of the following statistics were determined.

Sample mean of the annual rainfall was 19.88
Sample standard deviation of the annual rainfall was 6.31
Sample mean of the annual number of bushfires was 18.38
Sample standard deviation of the annual number of bushfires was 8.16
Pearson correlation coefficient was −0.88

Calculate the equation for the least squares line of best fit that models this data, rounding the coefficients a and b to 2 significant figures.

Prep 3 USING CAS: FINDING THE LEAST SQUARES LINE OF BEST FIT EQUATION

USING CAS: DRAWING THE LEAST SQUARES LINE OF BEST FIT

The following table shows the results of an investigation by a home loan company into the association between the interest rate and the number of loan applications they had in eight consecutive years.

Interest rate (p.a.)	8.3	9.7	10.4	9.5	8.1	9.1	10.8	10.0
No. of applications	55	46	29	36	47	45	26	32

a Find the equation of the least squares line of best fit, rounding the coefficients a and b to 2 significant figures.

b Sketch the least squares line of best fit from your CAS/calculator.

Prep 4 WORKED EXAMPLE 3

For the following least squares line of best fit equation

$$\text{mass} = -44 + 0.7 \times \text{height}$$

where height is measured in cm and mass is measured in kg

a identify and interpret the slope

b identify and interpret the y-intercept and comment on your result.

EXAM PRACTICE 4.1

The least squares line of best fit

Question 1

An investigation into the association between two data sets involving x and y gives the following values:

$$\bar{x} = 11.3, \quad s_x = 7.2, \quad \bar{y} = 12.5, \quad s_y = 9.4, \quad r = 0.86$$

The slope of the least squares line of best fit rounded to 2 significant figures is

A 0.67 **B** 0.78 **C** 0.95 **D** 1 **E** 1.1

Question 2

The effect of a dose of medicine on a child's temperature over time was measured, with the following results.

Time (minutes)	2	4	6	8	10
Temperature (°C)	39.5	39.1	38.9	38.2	37.7

The least squares line of best fit that models this data rounded to 1 significant figure is

A temperature = –0.2 + 40 × time

B temperature = 40 – 0.2 × time

C time = 40.0 – 0.2 × temperature

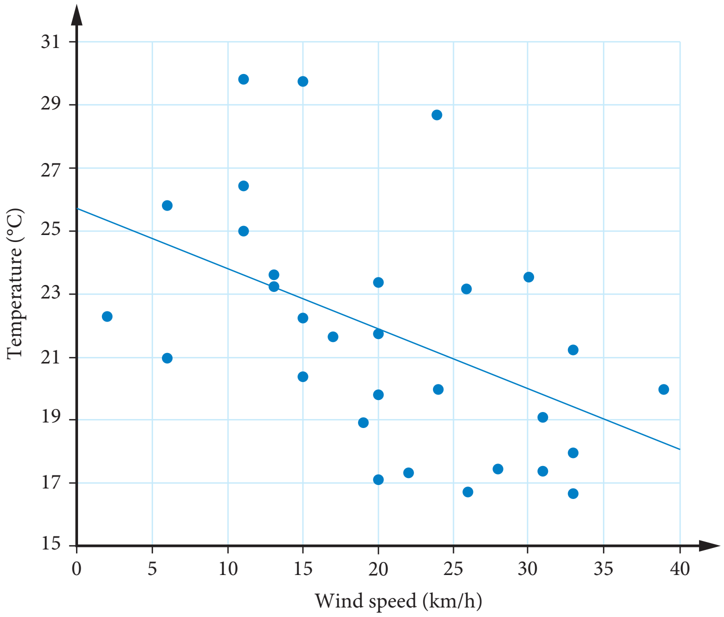
D temperature = 40.03 – 0.2 × time

E temperature = 40.03 – 0.23 × time

Question 3

The maximum wind speed and maximum temperature were recorded each day for a month. The data is displayed in the scatterplot and a least squares line of best fit has been fitted. The response variable is *temperature*. The explanatory variable is *wind speed*.

The equation of the least squares line of best fit is closest to



A *temperature* = 25.7 – 0.191 × *wind speed*

B *wind speed* = 25.7 – 0.191 × *temperature*

C *temperature* = 0.191 + 25.7 × *wind speed*

D *wind speed* = 25.7 + 0.191 × *temperature*

E *temperature* = 25.7 + 0.191 × *wind speed*

[VCAA 2012 1CQ8]

Use the following information to answer Questions 4 & 5.

The length (in metres) and wingspan (in metres) of eight commercial aeroplanes are displayed in the table below.

Length	70.7	70.7	63.7	58.4	54.9	39.4	36.4	33.4
Wingspan	64.4	59.6	60.3	60.3	47.6	35.8	28.9	28.9

Question 4

Correct to 4 decimal places, the value of the Pearson correlation coefficient for this data is

A 0.9371 **B** 0.9583 **C** 0.9681 **D** 0.9793 **E** 0.9839

[VCAA 2005 1CQ8]

Question 5

The equation of the least squares line of best fit for this data is

$$\text{wingspan} = -2.99 + 0.96 \times \text{length}$$

From this equation it can be concluded that, on average, for these aeroplanes, wingspan

A decreases by 2.03 metres with each one metre increase in length.

B increases by 0.96 metres with each one metre increase in length.

C decreases by 0.96 metres with each one metre increase in length.

D increases by 2.99 metres with each one metre increase in length.

E decreases by 2.99 metres with each one metre increase in length.

[VCAA 2005 1CQ9]

Use the following information to answer Questions 6 & 7.

The table below lists the average life span (in years) and average sleeping time (in hours/day) of 12 animal species.

Species	Life span (years)	Sleeping time (hours/day)
Baboon	27	10
Cow	30	4
Goat	20	4
Guinea pig	8	8
Horse	46	3
Mouse	3	13
Pig	27	8
Rabbit	18	8
Rat	5	13
Red fox	10	10
Rhesus monkey	29	10
Sheep	20	4

Question 6

Using *sleeping time* as the explanatory variable, a least squares line of best fit is fitted to the data. The equation of the least squares line of best fit is closest to

A *life span* = 38.9 – 2.36 × *sleeping time*.
B *life span* = 11.7 – 0.185 × *sleeping time*.
C *life span* = –0.185 – 11.7 × *sleeping time*.
D *sleeping time* = 11.7 – 0.185 × *life span*.
E *sleeping time* = 38.9 – 2.36 × *life span*.

[VCAA 2009 1CQ9]

Question 7

The value of the Pearson correlation coefficient for *life span* and *sleeping time* is closest to

A –0.6603
B –0.4360
C –0.1901
D 0.4360
E 0.6603

[VCAA 2009 1CQ10]

Question 8

For a set of bivariate data involving the variables x and y,

$$r = -0.5675,\ \bar{x} = 4.56,\ s_x = 2.61,\ \bar{y} = 23.93 \text{ and } s_y = 6.98$$

The equation of the least squares line of best fit $y = a + bx$ is closest to

A $y = 30.9 - 1.52x$
B $y = 17.0 - 1.52x$
C $y = -17.0 + 1.52x$
D $y = 30.9 - 0.2x$
E $y = 24.9 - 0.2x$

[VCAA 2006 1CQ7]

Question 9

For a set of bivariate data that involves the variables x and y, with y as the response variable

$$r = -0.644,\ \bar{x} = 5.30,\ \bar{y} = 5.60,\ s_x = 3.06,\ s_y = 3.20$$

The equation of the least squares line of best fit is closest to

A $y = 9.2 - 0.7x$
B $y = 9.2 + 0.7x$
C $y = 2.0 - 0.6x$
D $y = 2.0 - 0.7x$
E $y = 2.0 + 0.7x$

[VCAA 2010 1CQ10]

Question 10

The following data was recorded from measurements made on 12 men.

A scatterplot of mass against waist measurement for this sample of men shows that there is a strong linear association between mass and waist measurement.

The least squares line of best fit equation that would enable mass to be predicted from waist measurement is closest to

A mass = –25 + 1.11 × waist.

B mass = –20 + 1.11 × waist.

C mass = 1.11 – 20 × waist.

D mass = 28 + 0.78 × waist.

E mass = 0.78 + 28 × waist.

Age (years)	Mass (kg)	Waist (cm)
26	84	84
29	72	74
32	67	89
32	59	75
34	97	106
37	112	114
39	67	80
40	91	101
41	98	101
43	89	94
45	117	126
51	62	82

[VCAA 2002 1CQ9]

Question 11

The arm spans (in cm) and heights (in cm) for a group of 13 boys have been measured. The results are displayed in the table.

The aim is to find a linear equation that allows arm span to be predicted from height.

a What will be the explanatory variable in the equation? 1 mark

b Assuming a linear association, determine the equation of the least squares line of best fit that enables *arm span* to be predicted from *height*. Write this equation in terms of the variables *arm span* and *height*. Give the coefficients rounded to 2 significant figures. 2 marks

c Using the equation that you have determined in **part b**, interpret the slope of the least squares line of best fit in terms of the variables *height* and arm span. 1 mark

[VCAA 2008 2CQ4]

Arm span (cm)	Height (cm)
152	152
153	155
174	168
141	149
170	172
165	168
163	163
155	157
165	165
152	150
143	146
156	153
174	174

9780170371100

The coefficient of determination

4.2


Alamy/Rafael Ben-Ari

WS Coefficient of determination

Calculating the coefficient of determination

Now that we have used the least squares line of best fit to model sets of data, we should ask ourselves 'How well does our line of best fit actually represent our set of data?' To answer this question, we use the **coefficient of determination** or r^2.

The coefficient of determination

- can be calculated by squaring the Pearson correlation coefficient
- is a value between 0 and 1
- is a measure of how useful a line of best fit is as a linear model for a particular set of data (0 means it's a totally useless measure and 1 means it's a perfect measure).

The higher the coefficient of determination

- the stronger the association between the variables
- the better the line of best fit is as a model for the data.

Exam hack

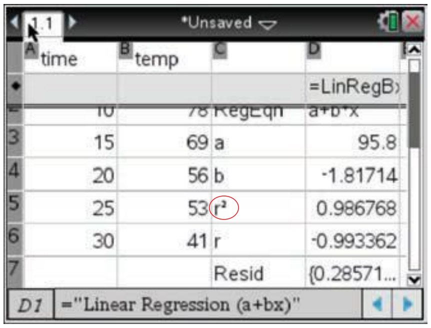
As pointed out previously, the values *a* and *b* in the least squares line of best fit equation are often called the 'coefficients' of the equation. This is a general term for these values in an equation. Don't confuse this with the use of the same word in the term 'coefficient of determination'.

9780170371100

Using CAS Finding the coefficient of determination

To calculate the coefficient of determination from a set of data, see the steps for ‘Using CAS: Finding the least squares line of best fit equation’ in the previous section. The r^2 value is one of the values that appears on the screen along with the values *a* and *b* for the least squares line of best fit equation.

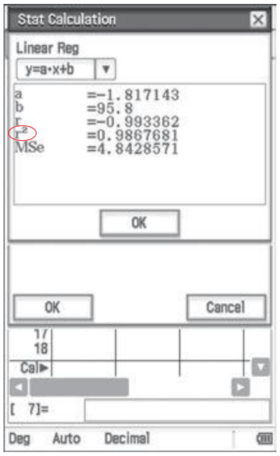
TI-NSPIRE CAS



	A time	B temp	C	D
◆				=LinRegB
2	10	78	RegEqn	a+b*x
3	15	69	a	95.8
4	20	56	b	-1.81714
5	25	53	r²	0.986768
6	30	41	r	-0.993362
7			Resid	{0.28571...

D1 ="Linear Regression (a+bx)"

CLASSPAD



```
a        =-1.817143
b        =95.8
r        =-0.993362
r²       =0.9867681
MSe      =4.8428571
```

Interpreting the coefficient of determination

The coefficient of determination is usually converted to a percentage, so that it is a value between 0 and 100%, and this value tells us the percentage of the variation in the response variable that is explained by the explanatory variable.

For example, if $r^2 = 0.85$ for a data set comparing height (the explanatory variable) and hand span (the response variable), then we can say that 85% of the variation in hand span can be explained by the variation in height (or, alternatively, that 15% of the variation in the hand span is *not* explained by the variation in height).

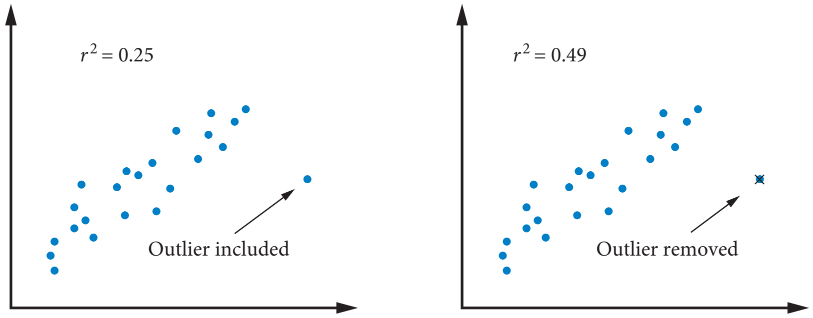
When interpreting the coefficient of determination, the following general sentence can be used:

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

where the **bold text** is replaced with the appropriate percentage and variable names of the set of data being investigated.

Exam hack

The coefficient of determination measures the predictive power of the association and it is affected by outliers. Removing an outlier increases the coefficient of determination. See the example below.



Worked example 4

Data was collected to investigate the association between the minimum daily temperature and the maximum daily temperature and is displayed in the table below.

Minimum temperature (°C)	12	15	13.5	14.1	10.2	11.7	13.2	11.3
Maximum temperature (°C)	23.4	25.6	23.1	25.3	20	21.1	22.6	21

a Assuming that minimum temperature is the explanatory variable, calculate the coefficient of determination, correct to 3 decimal places.

b Interpret the coefficient of determination.

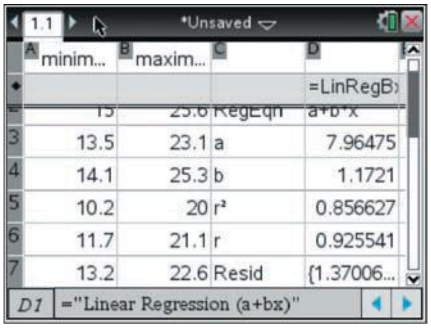
c What is the least squares line of best fit equation that models this data? Round the coefficients *a* and *b* to 2 decimal places.

d Do you think that the model is appropriate? Justify your answer.

Working

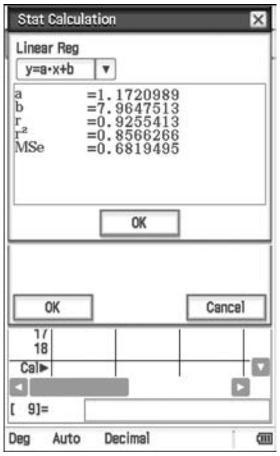
a We are told the minimum temperature is the explanatory variable. If we weren't told, we would have to determine this.

Use a CAS/calculator to find r^2. The *a* and *b* values for the least squares line of best fit equation will also be calculated.



	A minim...	B maxim...	C	D
◆				=LinRegB›
2	15	25.6	RegEqn	a+b*x
3	13.5	23.1	a	7.96475
4	14.1	25.3	b	1.1721
5	10.2	20	r^2	0.856627
6	11.7	21.1	r	0.925541
7	13.2	22.6	Resid	{1.37006...

D1 ="Linear Regression (a+bx)"



```
a              =1.1720989
b              =7.9647513
r              =0.9255413
r²             =0.8566266
MSe            =0.6819495
```

$r^2 = 0.857$

b	Calculate $r^2 \times 100\%$, then interpret the result using the general sentence.	85.7% of the variation in the maximum temperature can be explained by the variation in the minimum temperature.
c	Read the a and b values for the line of best fit equation from the screen and round to 2 decimal places. Write the equation using these values.	$y = 7.96 + 1.17x$
	Replace y and x in the equation with the correct variable names.	Maximum temperature = 7.96 + 1.17 × minimum temperature
d	Determine the appropriateness of the model, using r^2 to support your decision.	Yes, this is an appropriate model due to the high r^2 value. Only 14.3% (100 – 85.7) of the variation between the variables is unexplained.

Positive or negative association

The coefficient of determination is a squared number, so it will always be positive. It doesn't tell us about the direction (positive or negative) of the association. For example, if

$r = 0.54$, then $r^2 = (0.54)^2 = 0.2916$ and also if

$r = -0.54$, then $r^2 = (-0.54)^2 = 0.2916$

So if you know, for example, that $r^2 = 0.2916$ then the Pearson correlation coefficient could be either

$r = \sqrt{0.2916} = 0.54$ or

$r = -\sqrt{0.2916} = -0.54$

If you know the coefficient of determination and you need to work out whether the association is positive or negative, you need to look at either the:

- scatterplot to see if the direction is positive or negative, or
- the line of best fit equation to see if the slope (i.e. b) is positive or negative.

Worked example 5

The least squares line of best fit that enables the percentage mark on a test to be determined from the number of hours spent studying is

% mark = 15 + 9.5 × study hours

The coefficient of determination for this data is 0.92. Find the value of the Pearson correlation coefficient correct to 2 decimal places.

		Working
1	Calculate the square root of the coefficient of determination. Remember that this value could be positive or negative.	$r = \pm\sqrt{0.92}$ $\approx \pm 0.96$
2	Determine the slope of the least squares line of best fit. It is the number multiplied by 'study hours' in the equation. The sign of the slope enables us to determine if *r* is positive or negative.	The slope of the least squares line of best fit is 9.5 (which is positive). As the line of best fit has a positive slope, the *r* value is also positive. So $r = 0.96$

Exam hack

When more than one variable is involved in a question, always check that you are applying the formula to the right variable. You need to establish which is the explanatory variable and which is the response variable.

EXAM PREP 4.2

The coefficient of determination

Prep 1 WORKED EXAMPLE 4 USING CAS: FINDING THE COEFFICIENT OF DETERMINATION

The shoe sizes and heights of some students were measured and the results are shown below.

Shoe size	7	5.5	11	6	7	7.5	7	7	10	8	6.5
Height (cm)	168	168	182	162	162	157	168	171	178	171	168

a Calculate the coefficient of determination, correct to 3 decimal places.

b Interpret the coefficient of determination.

c What is the least squares line of best fit equation that models this data? Round the coefficients *a* and *b* to 2 significant figures.

d Do you think the model is appropriate? Justify your answer.

9780170371100

Prep 2 WORKED EXAMPLE 5

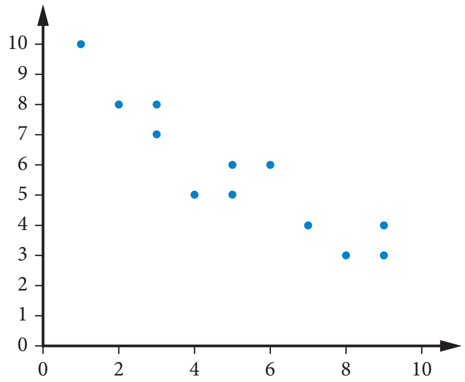
The least squares line of best fit that enables the distance travelled to be determined from the time spent travelling is

distance travelled = 31 + 1.5 × time.

The coefficient of determination for this data is 0.88. Find the value of the Pearson correlation coefficient correct to 2 decimal places.

Prep 3

The coefficient of determination for the data displayed in the scatterplot is 0.72. Find the value of the correlation coefficient correct to 2 decimal places.



EXAM PRACTICE 4.2

The coefficient of determination

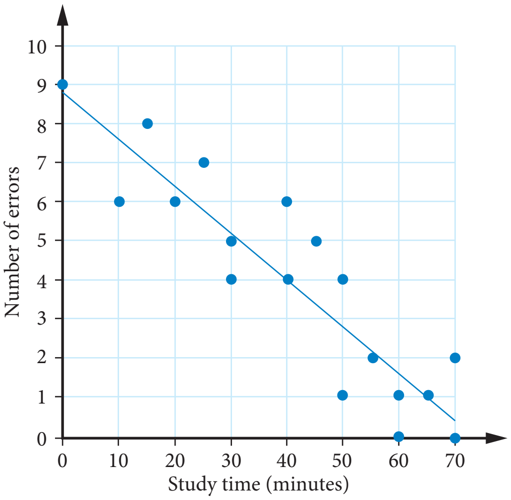
Use the following information to answer Questions 1 & 2.

Eighteen students sat for a 15 question multiple-choice test. In the scatterplot, the number of errors made by each student on the test is plotted against the time they reported studying for the test. A least squares line of best fit has been determined for this data and is also displayed on the scatterplot.

The equation for the least squares line of best fit is

number of errors = 8.8 – 0.120 × study time

and the coefficient of determination is 0.8198.



Question 1

Using the least squares line of best fit, it can be estimated that, on average, a student reporting a study time of 35 minutes would make

A 4.3 errors. **B** 4.6 errors. **C** 4.8 errors. **D** 5.0 errors. **E** 13.0 errors.

[VCAA 2003 1CQ8]

Question 2

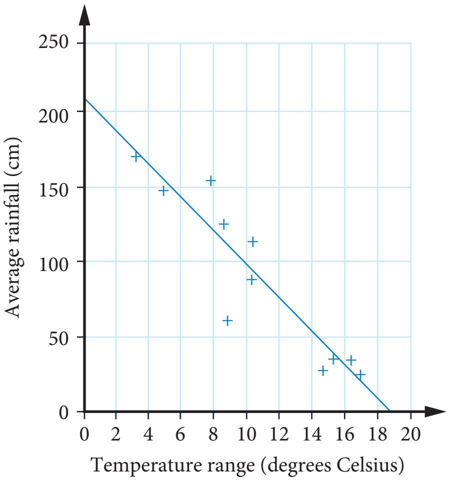
The value of the Pearson correlation coefficient, *r*, for this data, correct to 2 decimal places, is

A –0.91 **B** –0.82 **C** 0.67 **D** 0.82 **E** 0.91

[VCAA 2003 1CQ9]

Use the following information to answer Questions 3 & 4.

The average rainfall and temperature range at several different locations in the South Pacific region are displayed in the scatterplot.



Question 3

A least squares line of best fit has been fitted to the data as shown. The equation of this line is closest to

A average rainfall = 210 – 11 × temperature range.

B average rainfall = 210 + 11 × temperature range.

C average rainfall = 18 – 0.08 × temperature range.

D average rainfall = 18 + 0.08 × temperature range.

E average rainfall = 250 – 13 × temperature range.

[VCAA 2004 1CQ8]

Question 4

The value of the the Pearson correlation coefficient, *r*, for the data, is $r = -0.9260$. The value of the coefficient of determination is

A –0.9260 **B** –0.8575 **C** 0.8575 **D** 0.9260 **E** 0.9623

[VCAA 2004 1CQ9]

Use the following information to answer Questions 5–7.

When blood pressure is measured, both the systolic (or maximum) pressure and the diastolic (or minimum) pressure are recorded.

Table 1 displays the blood pressure readings, in mmHg, that result from fifteen successive measurements of the same person's blood pressure.

Table 1

	Blood pressure	
Reading number	Systolic	Diastolic
1	121	73
2	126	75
3	141	73
4	125	73
5	122	67
6	126	74
7	129	70
8	130	72
9	125	69
10	121	65
11	118	66
12	134	77
13	125	70
14	127	64
15	119	69

Question 5

Correct to 1 decimal place, the mean and standard deviation of this person's **systolic** blood pressure measurements are respectively

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

[VCAA 2011 1CQ6]

Question 6

Using systolic blood pressure (*systolic*) as the response variable, and diastolic blood pressure (*diastolic*) as the explanatory variable, a least squares line of best fit is fitted to the data in **Table 1**.

The equation of the least squares line of best fit is closest to

A *systolic* = 70.3 + 0.790 × *diastolic*

B *diastolic* = 70.3 + 0.790 × *systolic*

C *systolic* = 29.3 + 0.330 × *diastolic*

D *diastolic* = 0.330 + 29.3 × *systolic*

E *systolic* = 0.790 + 70.3 × *diastolic*

[VCAA 2011 1CQ7]

Question 7

From the fifteen blood pressure measurements for this person, it can be concluded that the percentage of the variation in systolic blood pressure that is explained by the variation in diastolic blood pressure is closest to

A 25.8% **B** 50.8% **C** 55.4% **D** 71.9% **E** 79.0%

[VCAA 2011 1CQ8]

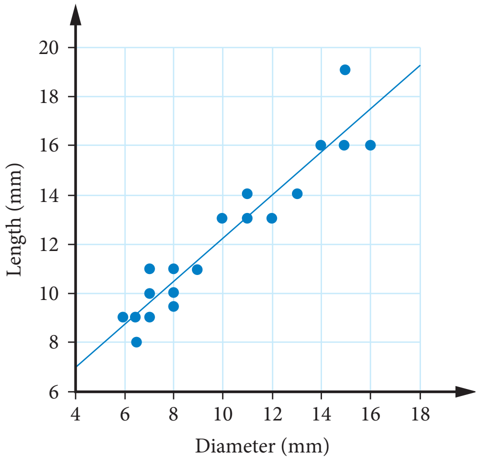
Use the following information to answer Questions 8 & 9.

The lengths and diameters (in mm) of a sample of jellyfish were recorded and displayed in the scatterplot. The least squares line of best fit for this data is shown.

The equation of the least squares line of best fit is

$$\text{length} = 3.5 + 0.87 \times \text{diameter}$$

The correlation coefficient is $r = 0.9034$



Question 8

Written as a percentage, the coefficient of determination is closest to

A 0.816% **B** 0.903% **C** 81.6% **D** 90.3% **E** 95.0%

[VCAA 2007 1CQ7]

Question 9

From the equation of the least squares line of best fit, it can be concluded that for these jellyfish, on average

A there is a 3.5 mm increase in *diameter* for each 1 mm increase in *length*.

B there is a 3.5 mm increase in *length* for each 1 mm increase in *diameter*.

C there is a 0.87 mm increase in *diameter* for each 1 mm increase in *length*.

D there is a 0.87 mm increase in *length* for each 1 mm increase in *diameter*.

E there is a 4.37 mm increase in *diameter* for each 1 mm increase in *length*.

[VCAA 2007 1CQ8]

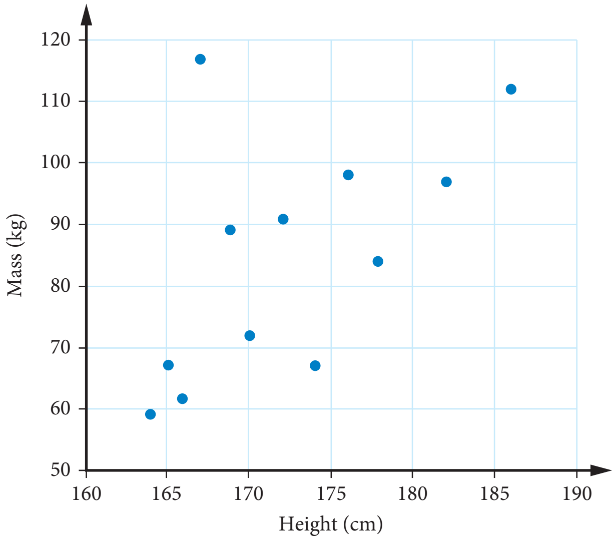
Question 10

A person's mass is known to be positively associated with their height. To investigate this association for 12 men, a scatterplot is constructed as shown.

While there is a moderately strong positive linear association between mass and height, there is a clear outlier.

When a least squares line of best fit is used to model this data, the coefficient of determination is found to be 0.3146.

If the outlier is removed from the data, and a least squares line of best fit refitted to the data of the remaining 11 men, the value of the coefficient of determination will



A remain the same.

B increase.

C decrease.

D be halved.

E not be able to be determined.

[VCAA 2002 1CQ12]

Question 11

For a city, the correlation coefficient between

- population density and distance from the centre of the city is $r = -0.563$
- house size and distance from the centre of the city is $r = 0.357$.

Given this information, which one of the following statements is true?

A Around 31.7% of the variation observed in house size in the city can be explained by the variation in distance from the centre of the city.

B Population density tends to increase as the distance from the centre of the city increases.

C House sizes tend to be larger as the distance from the centre of the city decreases.

D The slope of a least squares line of best fit relating population density to distance from the centre of the city is positive.

E Population density is more strongly associated with distance from the centre of the city than is house size.

[VCAA 2013 1CQ7]

Question 12

Table 1 Heights and masses of nine people

Height (m)	Mass (kg)
1.65	68
1.68	**63**
1.72	79
1.73	65
1.74	70
1.77	79
1.78	**81**
1.86	77
1.92	88

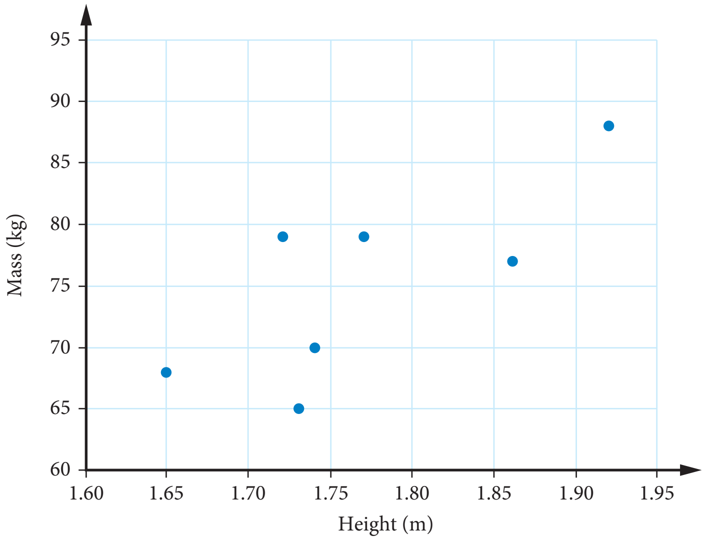


Table 1 gives the heights (m) and masses (kg) of a sample of nine people.

a On the scatterplot, the points representing the data for seven of these people has been plotted with *height* on the horizontal axis and *mass* on the vertical axis.

Copy the scatterplot and plot points representing the data for the remaining two people (shown in bold in **Table 1**) on it. 1 mark

b Determine the equation of the least squares line of best fit that fits the data in **Table 1**. Use *height* as the explanatory variable and *mass* as the response variable. Complete the line of best fit equation by writing the appropriate values in the boxes, correct to 1 decimal place.

mass = ☐ + ☐ × *height* 2 marks

c The coefficient of determination for this data is 0.61

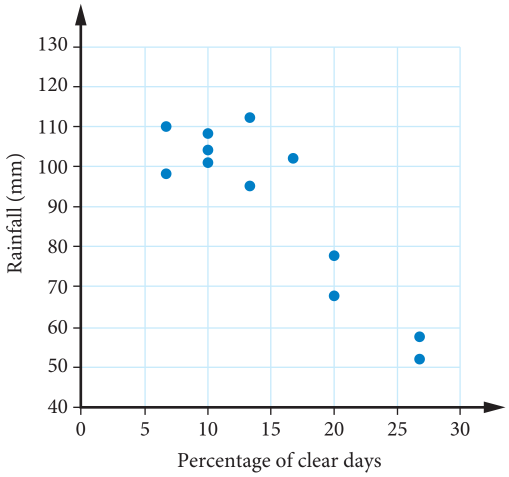
Complete the following sentence by filling in the missing information.

For this sample, 61% of the variation in the ☐ of the people can be explained by the variation in their ☐. 1 mark

[VCAA 2004 2CQ1]

Question 13

The scatterplot shows the *rainfall* (in mm) and the *percentage of clear days* for each month of 2008.



An equation of the least squares line of best fit for this data set is

rainfall = 131 − 2.68 × *percentage of clear days*

a Copy the scatterplot and draw this line on it. 1 mark

b Use the equation of the least squares line of best fit to predict the rainfall for a month with 35% of clear days. Write your answer in mm correct to 1 decimal place. 1 mark

c The coefficient of determination for this data set is 0.8081.

i Interpret the coefficient of determination in terms of the variables *rainfall* and *percentage of clear days*. 1 mark

ii Determine the value of the Pearson correlation coefficient. Write your answer correct to 3 decimal places. 2 marks

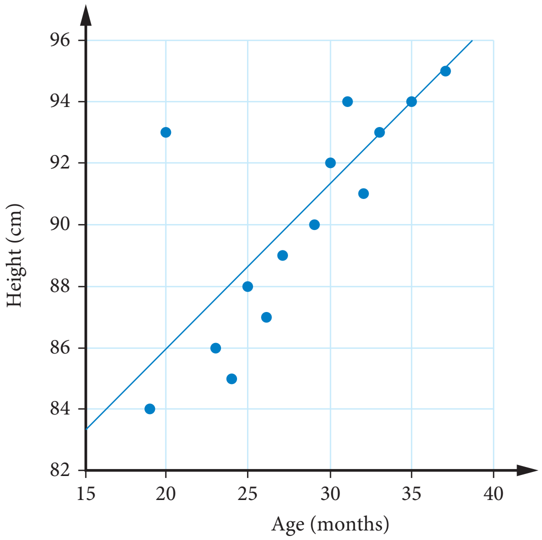
[VCAA 2009 2CQ3]

Question 14

The heights (in cm) and ages (in months) of a random sample of 15 boys have been plotted in the scatterplot. The least squares line of best fit has been fitted to the data.

The equation of the least squares line of best fit is *height* = 75.4 + 0.53 × *age*

The correlation coefficient is $r = 0.7541$



a Complete the following sentence.

On average, the height of a boy increases by ☐ cm for each one-month increase in age. 1 mark

b i Evaluate the coefficient of determination.

Write your answer, as a percentage, correct to 1 decimal place. 1 mark

ii Interpret the coefficient of determination in terms of the variables *height* and *age*. 1 mark

[VCAA 2006 2CQ2]

Making predictions

4.3

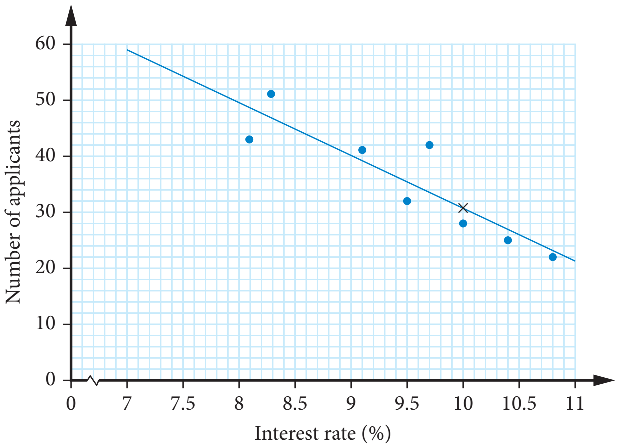
Shutterstock.com/Africa Studio

Interpolation and extrapolation

It is possible to make predictions directly from a least squares line of best fit graph. For example it is possible to predict from the graph below that an interest rate of 10% will have between 30 and 31 applicants. However, this method isn't accurate. It's difficult to determine by eye whether the prediction is closer to 30 or 31.

A better way to predict values is to use the equation of the least squares line of best fit by substituting values into it and solving. The equation can be used to predict *within* the original data range, which is called **interpolation**, as well as *outside* the original data range, which is called **extrapolation**.

Predictions based on extrapolation are not as reliable as those based on interpolation because we cannot be certain that the equation applies to values outside the range of data values we have.



Using CAS Solving the least squares line of best fit equation

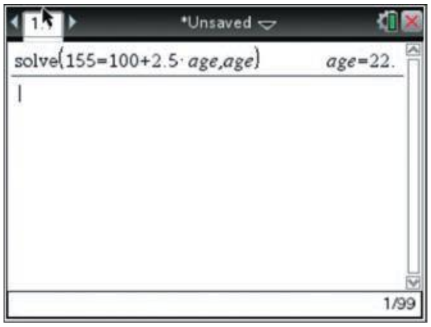
The following least squares line of best fit can be used to predict the age of people of different heights:

height (cm) = 100 + 2.5 × age (years)

Use the solve function on your CAS/calculator to find the predicted age for a person who is 155 cm tall.

TI-NSPIRE CAS

Add a Calculator page, then menu, 3: Algebra, 1: Solve, then enter as per the screen and press enter.



solve(155=100+2.5·age,age)   age=22.

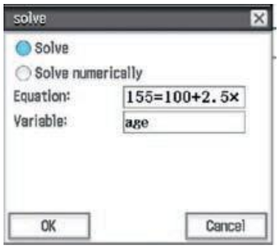
CLASSPAD

Open the main menu Main

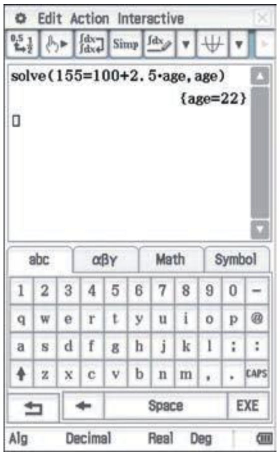
Type in required equation.

Highlight, then tap Interactive, Equation/Inequality, Solve (equation will already be there) and change variable to age.

Tap OK.



● Solve
○ Solve numerically
Equation: 155=100+2.5×
Variable: age



solve(155=100+2.5·age, age)
{age=22}

Note: always use a by hand method if it is quicker.

Worked example 6

Data was collected from people aged between 7 and 19 years of age and a least squares line of best fit was found to have the equation

height (cm) = 90 + 5 × age (years)

a Predict the height of a 15-year-old. Does this involve interpolation or extrapolation?

b Predict the height of a 24-year-old. Does this involve interpolation or extrapolation?

c Which of the predictions in parts **a** and **b** is more reliable? Justify your answer.

d Predict the age of a person of height 145 cm.

	Working
a Substitute 15 into the equation in place of age and solve for height by hand or CAS/calculator. Was the value used within or outside the original data range?	$\text{height} = 90 + 5 \times \text{age}$ $= 90 + 5 \times 15$ $= 165$ cm As 15 years old is inside the original data range of 7 to 19 years, this involves interpolation.
b Substitute 24 into the equation in place of age and solve for height by hand or CAS/calculator. Was the value used within or outside the original data range?	$\text{height} = 90 + 5 \times \text{age}$ $= 90 + 5 \times 24$ $= 210$ cm As 24 years old is outside the original data range of 7 to 19 years, this involves extrapolation.
c Decide which of the two is the more reliable prediction and justify your decision.	The prediction for the height of a 15-year-old involves interpolation, so it is more reliable than that for a 24-year-old, which involves extrapolation.
d Substitute 145 into the equation in place of height and solve for age by hand or CAS/calculator.	$145 = 90 + 5 \times \text{age}$ $5 \times \text{age} = 145 - 90$ (subtracting 90 from both sides) $5 \times \text{age} = 55$ $\text{age} = 55 \div 5$ (dividing both sides by 5) $\text{age} = 11$ years

EXAM PREP 4.3

Making predictions

Prep 1

Exam hack

When stating the predicted value, make sure that you include the units of measure as well.

The least squares line of best fit with equation

$$\text{number of laps} = 10.87 - 0.16 \times \text{age}$$

models the association between the number of laps run around an oval in 30 minutes and the age of the runner (in years). Use the model to predict, correct to 2 decimal places, the number of laps run in 30 minutes by a runner aged

a 12 **b** 34 **c** 55

Prep 2 USING CAS: SOLVING THE LEAST SQUARES LINE OF BEST FIT EQUATION

The least squares line of best fit with the equation given below models data relating the outside temperature (°C) to the amount of gas consumed (kWh) by a household over a three-month period.

gas consumed = 2212 – 125 × outside temperature

Use the model to predict the outside temperature, to the nearest degree, when the gas consumed is

a 2000 kWh **b** 1200 kWh **c** 0 kWh

Prep 3 WORKED EXAMPLE 6 USING CAS: SOLVING THE LEAST SQUARES LINE OF BEST FIT EQUATION

The masses (in grams) of boxes of chocolates containing between 15 and 50 chocolates were recorded. The least squares line of best fit for the data was found to have the equation

mass = 20 + 5 × number of chocolates

a Predict the mass of a box with 25 chocolates. Does this involve interpolation or extrapolation?

b Predict the mass of a box with 5 chocolates. Does this involve interpolation or extrapolation?

c Predict the mass of a box with 70 chocolates. Does this involve interpolation or extrapolation?

d Which of these predictions is the most reliable? Justify your answer.

e Predict the number of chocolates that would be in a box weighing 125 grams.

EXAM PRACTICE 4.3

Making predictions

Use the following information to answer Questions 1–3.

A least squares line of best fit modelling the height (cm) of 20 children aged between 5 and 15 years was found to have the equation

height = 99.97 + 2.59 × age

Question 1

The predicted height for a 14-year-old is

A 136 cm **B** 136.23 cm **C** 137.5 cm **D** 146 cm **E** 146.34 cm

Question 2

The predicted age of a child who is 120.69 cm is

A 5 **B** 6 **C** 7 **D** 8 **E** 9

Question 3

Which one of the following statements is *false*?

A Predicting the height of a child aged 9 years is an example of interpolation.

B For every one year of increase in age, the model predicts a child's height increases by 2.59 cm.

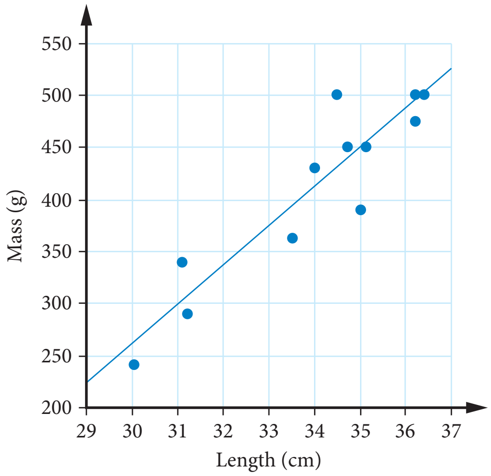
C Predicting the height of a 6-year-old will be more reliable than predicting the height of a 12-year-old.

D A child's height will increase as their age increases.

E Predicting the height of a 10-year-old is more likely to be reliable than predicting the height of an 18-year-old.

Use the following information to answer Questions 4 & 5.

The masses (in g) and lengths (in cm) of 12 fish were recorded and plotted in the scatterplot. The least squares line of best fit that enables the mass of these fish to be predicted from their length has also been plotted.



[VCAA 2008 1CQ8]

Question 4

The least squares line of best fit predicts that the mass (in g) of a fish of length 30 cm would be closest to

A 240 **B** 252 **C** 262

D 274 **E** 310

Question 5

The median mass (in g) of the 12 fish is closest to

A 346 **B** 375 **C** 440 **D** 450 **E** 475

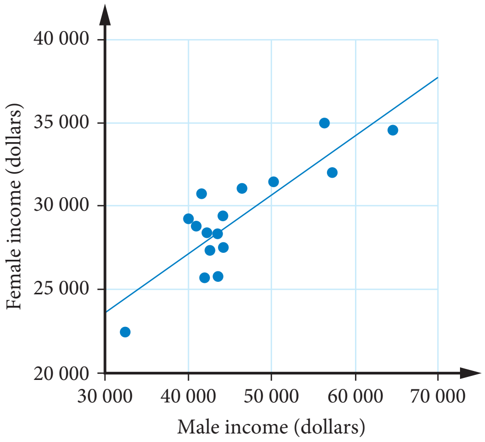
[VCAA 2008 1CQ9]

Question 6

In the scatterplot, average annual *female income*, in dollars, is plotted against average annual *male income*, in dollars, for 16 countries. A least squares line of best fit is fitted to the data.

The equation of the least squares line of best fit for predicting female income from male income is

$$\textit{female income} = 13\,000 + 0.35 \times \textit{male income}$$



a What is the explanatory variable? 1 mark

b Complete the following statement by filling in the missing information.

From the least squares line of best fit equation it can be concluded that, for these countries, on average, female income increases by $____ for each $1000 increase in male income. 1 mark

c **i** Use the least squares line of best fit equation to predict the average annual female income (in dollars) in a country where the average annual male income is $15 000. 1 mark

ii The prediction made in **part c i** is not likely to be reliable. Explain why. 1 mark

[VCAA 2010 2CQ2]

Residual analysis

4.4

Shutterstock.com/pisaphotography

WS Predictions and residuals

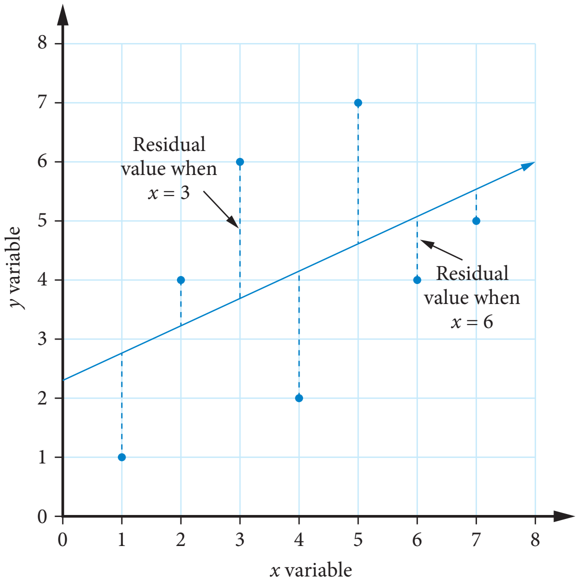
Residual values

As we've seen, it's important to know whether the association between the variables is linear or not. We've used the following to help determine whether the association is linear:

- the form of the scatterplot
- fitting a least squares line of best fit
- seeing how close the correlation coefficient, r, is to 1 and −1
- the value of the coefficient of determination, r^2.

However, one of the best ways to test the assumption that the association is linear is to calculate the **residual values**.

A residual is the vertical distance between each data point and the least squares line of best fit. In the graph, the residual value for the data value $x = 3$ looks to be around 2.25 and the residual value for the data value $x = 6$ looks to be around −1.1.



Residual value = data value – predicted value

The data value can be read from a scatterplot or a table.

The predicted value needs to be calculated from the least squares line of best fit.

Data values that lie

- *above* the least squares line of best fit will have a positive residual value
- *below* the least squares line of best fit will have a negative residual value
- *on* the least squares line of best fit will have a residual value of zero.

The further the data value is from the least squares line of best fit, the larger the residual value.

Worked example 7

The least squares line of best fit connecting height to age was found to have the equation

height (cm) = 100 + 2.5 × age (years)

Calculate the residual value for

a John who is 12 years old and 142 cm tall

b Jane who is 10 years old and 118 cm tall.

Working

a **1** Calculate the predicted height for John by substituting 12 for age into the equation.

$$\begin{aligned}\text{height} &= 100 + 2.5 \times \text{age}\\ &= 100 + 2.5 \times 12\\ &= 130\text{ cm}\end{aligned}$$

2 Find the residual value using the formula.

The actual y value was stated in the question.

Remember to place units relating to the y variable in the final answer.

$$\begin{aligned}\text{Residual} &= \text{actual } y \text{ value} - \text{predicted } y \text{ value}\\ &= 142 - 130\\ &= 12\text{ cm}\end{aligned}$$

b **1** Calculate the predicted height for Jane by substituting 10 for age into the equation.

$$\begin{aligned}\text{height} &= 100 + 2.5 \times \text{age}\\ &= 100 + 2.5 \times 10\\ &= 125\text{ cm}\end{aligned}$$

2 Find the residual value using the formula.

$$\begin{aligned}\text{Residual} &= \text{actual } y \text{ value} - \text{predicted } y \text{ value}\\ &= 118 - 125\\ &= -7\text{ cm}\end{aligned}$$

Using CAS Calculating residual values

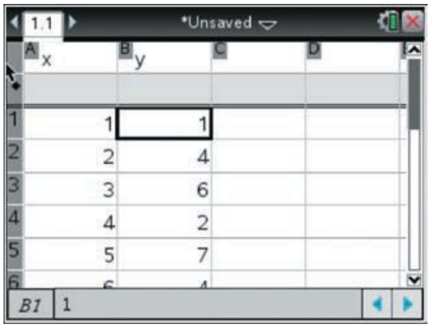
Create a table of residual values for the following set of data, stating all answers correct to 1 decimal place.

x	1	2	3	4	5	6	7
y	1	4	6	2	7	4	5

TI-NSPIRE CAS

STEP 1

Enter the data into a Lists & Spreadsheet page, remembering to name each column.



	A x	B y	C	D
1	1	1		
2	2	4		
3	3	6		
4	4	2		
5	5	7		
6	6	4		

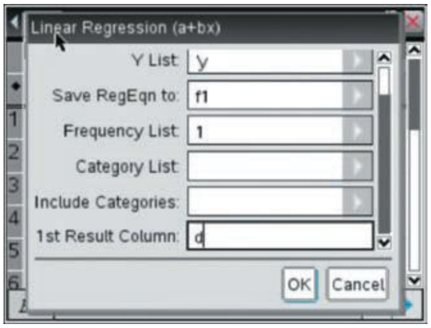
B1 1

STEP 2

Use the touchpad to move the cursor to column d

Press `menu`, then 4: Statistics, 1: Stat Calculations, 4: Linear Regression (a + bx).

Select X list enter "x, select Y List enter "y
The 1st result column should show d[]
indicating the result will appear in column D

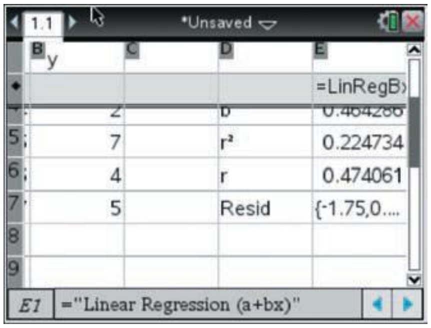

Linear Regression (a+bx)

Y List: y
Save RegEqn to: f1
Frequency List: 1
Category List:
Include Categories:
1st Result Column: d

Make sure that you select column D for answers so that column C can be used for the residual values.

STEP 3

Scroll down the page to find the residuals. Notice that you can only see the first one until you move over the cell and the others appear.



	B y	C	D	E
◆				=LinRegBx
4	2		b	0.464286
5	7		r²	0.224734
6	4		r	0.474061
7	5		Resid	{-1.75,0.…
8				
9				

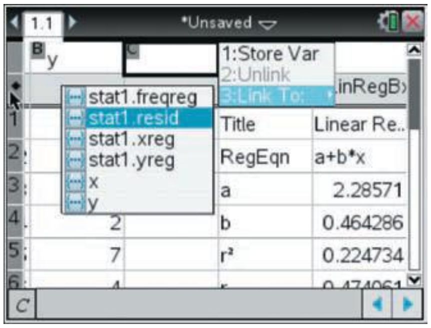
E1 ="Linear Regression (a+bx)"

STEP 4

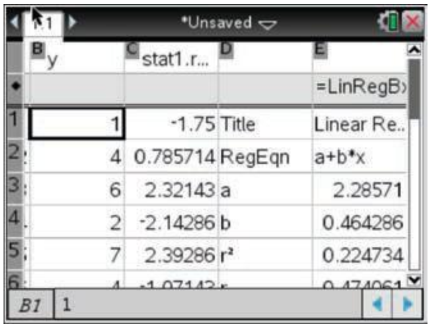
To see residuals more clearly add them to Column C of the spreadsheet.

The cursor must be in the formula cell for column C before you press var 3: Link to: stat1.resid

If you find another line of best fit without clearing your pages, then you would select stat2.resid.



	B y	C	D	E
◆				=LinRegBx
1			Title	Linear Re..
2			RegEqn	a+b*x
3			a	2.28571
4	2		b	0.464286
5	7		r²	0.224734



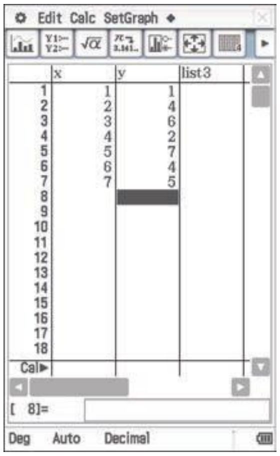
	B y	C stat1.r...	D	E
◆				=LinRegBx
1	1	-1.75	Title	Linear Re..
2	4	0.785714	RegEqn	a+b*x
3	6	2.32143	a	2.28571
4	2	-2.14286	b	0.464286
5	7	2.39286	r²	0.224734

B1 1

CLASSPAD

STEP 1

Using the Statistics application, rename the columns and type in the data.

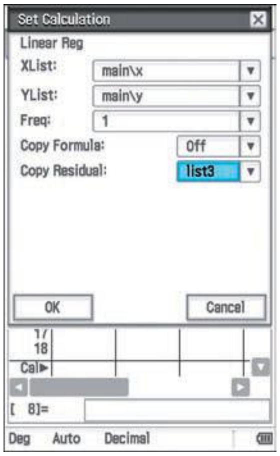


	x	y	list3
1	1	1	
2	2	4	
3	3	6	
4	4	2	
5	5	7	
6	6	4	
7	7	5	

STEP 2

Tap **Calc** then select **Linear Reg**.

Complete the screen as shown and click on OK.



Set Calculation

Linear Reg

XList: main\x

YList: main\y

Freq: 1

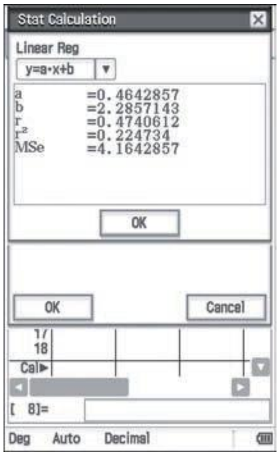
Copy Formula: Off

Copy Residual: list3

Make sure that you change Copy Residual from off to List 3. This means values will be entered into List 3.

STEP 3

Tap OK and the regression screen will be displayed. The split screen below will appear.

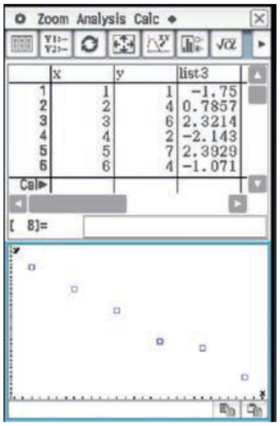


Stat Calculation

Linear Reg

y=a·x+b

```
a       =0.4642857
b       =2.2857143
r       =0.4740612
r²      =0.224734
MSe     =4.1642857
```

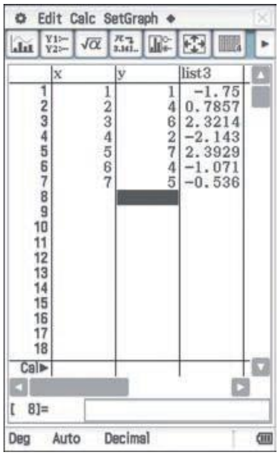


	x	y	list3
1	1	1	−1.75
2	2	4	0.7857
3	3	6	2.3214
4	4	2	−2.143
5	5	7	2.3929
6	6	4	−1.071

STEP 4

Tap Resize to get the full Table.

Tap OK and the residual values will appear in List 3.



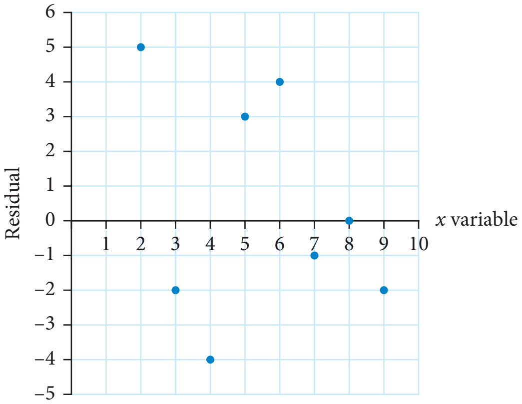
	x	y	list3
1	1	1	−1.75
2	2	4	0.7857
3	3	6	2.3214
4	4	2	−2.143
5	5	7	2.3929
6	6	4	−1.071
7	7	5	−0.536
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			

Complete a table with the residual values added from the calculator screen.

x	1	2	3	4	5	6	7
y	1	4	6	2	7	4	5
Residual	−1.8	0.8	2.3	−2.1	2.4	−1.1	−0.5

Residual plots

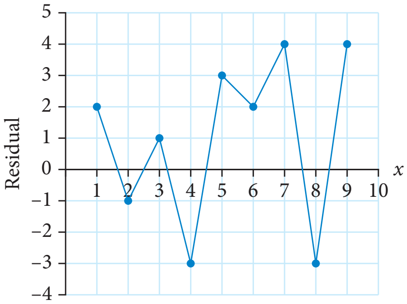
Once all of the residual values have been found, a **residual plot** can be constructed. A residual plot is like a scatterplot with the explanatory variable on the *x*-axis and the residual values on the *y*-axis. Since residual values are negative as well as positive, the residual plot will always have both positive and negative *y*-axes.



Most residual plots will fall into one of the following three types:

Type 1: Randomly scattered

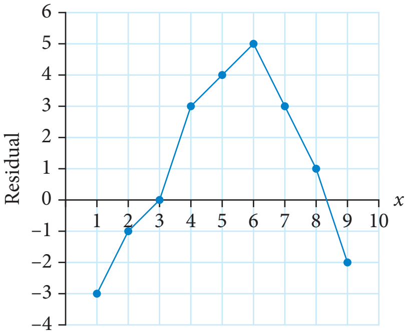
The residual values are randomly scattered above and below the *x*-axis. This indicates that the association between the original two variables, *x* and *y*, is probably linear.



It is not common practice to join the dots on a residual plot; it has only been done here to show the general pattern or lack of pattern in the residuals.

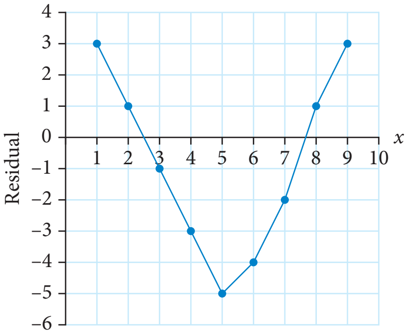
Type 2: Hill

The residual values show a hill or mountain shape. This indicates that the association between the original two variables, *x* and *y*, is probably non-linear.



Type 3: Valley

The residual values show a ‘u’ or ‘v’ shape. This also indicates that the association between the original two variables, *x* and *y*, is probably non-linear.



Using CAS Creating a residual plot

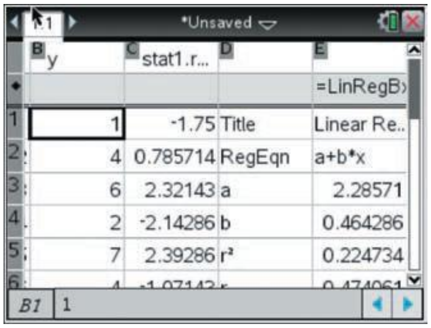
Construct a residual plot for the following data and use it to decide if the data being investigated is linear or non-linear.

x	1	2	3	4	5	6	7
y	1	4	6	2	7	4	5

TI-NSPIRE CAS

STEP 1

First calculate the residuals.



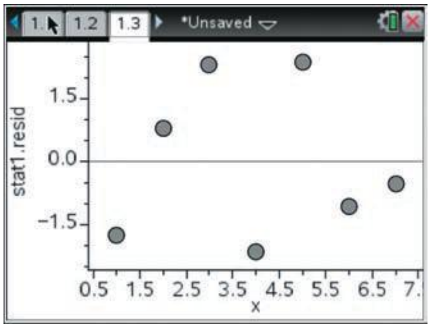
	B y	C stat1.r...	D	E
◆				=LinRegB›
1	1	-1.75	Title	Linear Re..
2	4	0.785714	RegEqn	a+b*x
3	6	2.32143	a	2.28571
4	2	-2.14286	b	0.464286
5	7	2.39286	r²	0.224734

STEP 2

Press ctrl doc▾ and select 5: Data and Statistics. Move to the bottom of the screen and select *x* as the *x* variable.

Move to the left side of the screen and select stat1.resid as the *y* variable.

The resulting graph is the residual plot.

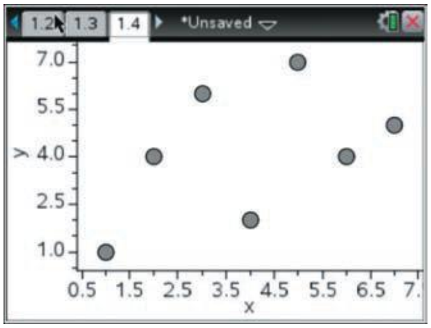


Using the TI-Nspire CAS, you can produce a scatterplot and residual plot on the one page. The steps are as follows:

TI-NSPIRE CAS

STEP 1

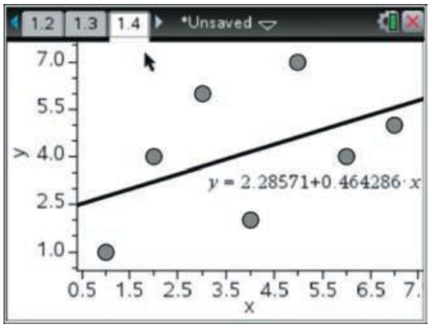
Create a scatterplot.



STEP 2

Add the line of best fit.

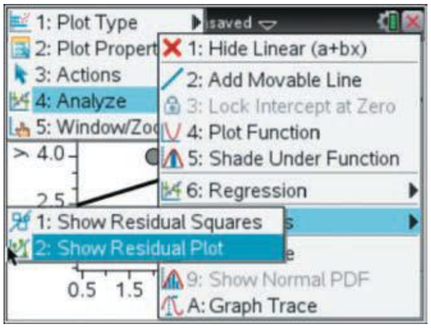
menu 4: Analyze, 6: Regression, 2: Show Linear(a + bx)



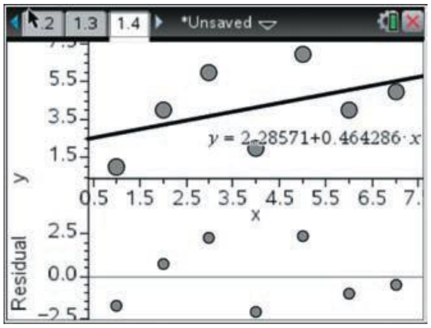
STEP 3

Add residual plot.

menu 4: Analyze, 7: Residuals, 2: Show Residual Plot



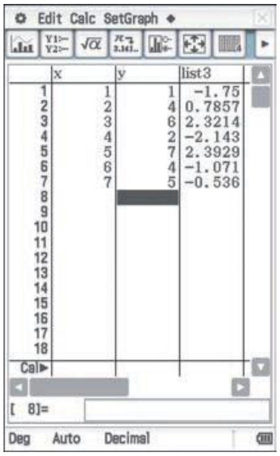
The screen will appear with the scatterplot at the top of the screen and the residual plot directly below it.



CLASSPAD

STEP 1

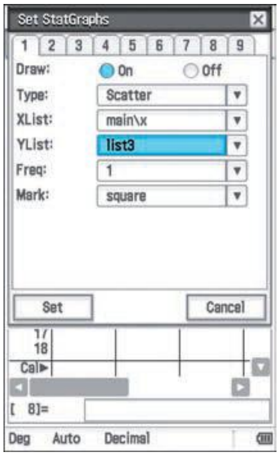
First calculate the residuals.



	x	y	list3
1	1	1	−1.75
2	2	4	0.7857
3	3	6	2.3214
4	4	2	−2.143
5	5	7	2.3929
6	6	4	−1.071
7	7	5	−0.536

STEP 2

Tap **Setgraph** and ensure that only StatGraph1 is checked. Tap **Setting** and fill in the screen as shown. Tap Set.

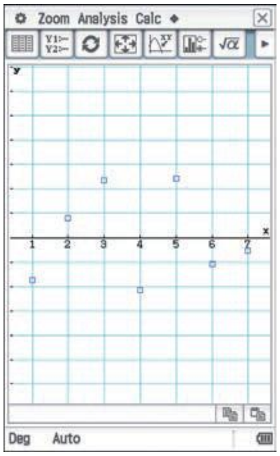


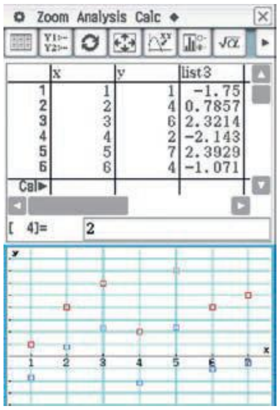
Draw: ● On ○ Off
Type: Scatter
XList: main\x
YList: list3
Freq: 1
Mark: square

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

STEP 3

Tap to show the graph then tap graph and resize to get a full screen plot.

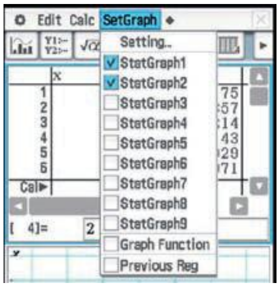




	x	y	list3
1	1	1	−1.75
2	2	4	0.7857
3	3	6	2.3214
4	4	2	−2.143
5	5	7	2.3929
6	6	4	−1.071

Tap to change the scale if points are not clear.

Note: You can also show both the scatterplot and the residuals on the same screen with the ClassPad. You simply make sure that StatGraph1 and StatGraph2 are BOTH checked, then when you go to Settings, set graph 1 to plot x against list3 (residuals) and set graph 2 to plot x against y. The graphs appear on the same screen in different colours.



☑ StatGraph1
☑ StatGraph2
☐ StatGraph3
☐ StatGraph4
☐ StatGraph5
☐ StatGraph6
☐ StatGraph7
☐ StatGraph8
☐ StatGraph9
☐ Graph Function
☐ Previous Reg

The data is probably linear because the residual values appear randomly scattered above and below the *x*-axis.

EXAM PREP 4.4

Residual analysis

Prep 1 WORKED EXAMPLE 7

The least squares line of best fit connecting the height of a building to the number of levels it was found to have the equation

height = 5 + 0.57 × number of levels

Correct to 1 decimal place, calculate the residual value for:

a a building with 12 levels that is 12.3 metres tall

b a building with 5 levels that is 7.2 metres tall

Exam hack

Don't lose easy marks. *Always* check whether the question has asked for a specific number of decimal places or significant figures.

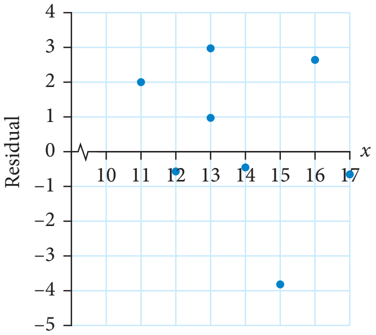
Prep 2 USING CAS: CALCULATING RESIDUAL VALUES USING CAS: CREATING A RESIDUAL PLOT

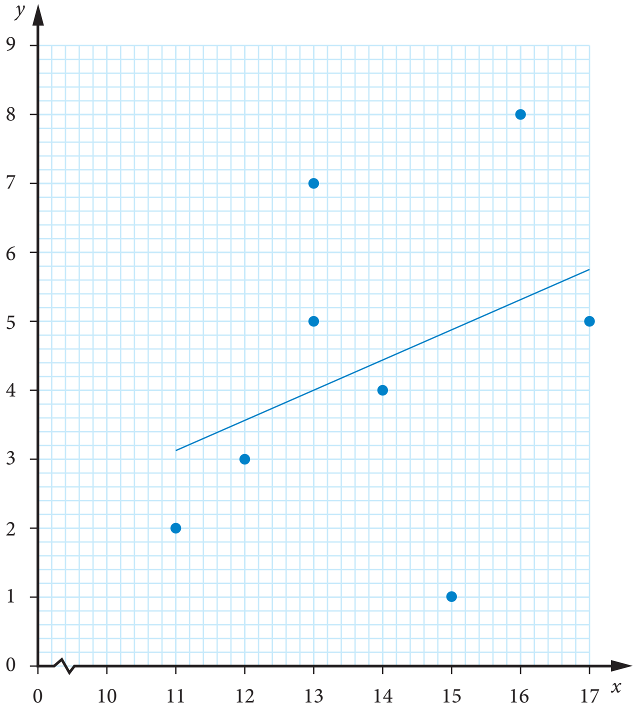
Create a table of residual values for the following set of data (assuming that height is the explanatory variable), giving all answers correct to 1 decimal place. Construct a residual plot and use it to decide if the data being investigated is linear or non-linear.

Height (cm)	Femur length (cm)
178	50.2
173	48.4
165	45.1
164	44.6
168	45
165	42.6
155	39.9
155	38

Prep 3

Explain why the residual plot below doesn't match the scatterplot and least squares line of best fit, referring specifically to the data value at $x = 11$.



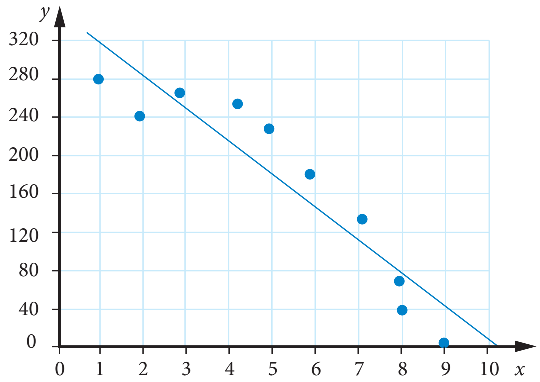


EXAM PRACTICE 4.4

Residual analysis

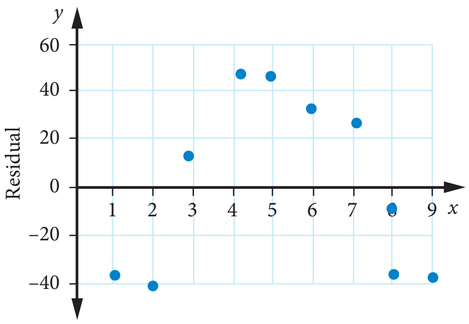
Question 1

A least squares line of best fit is fitted to data in a scatterplot, as shown.

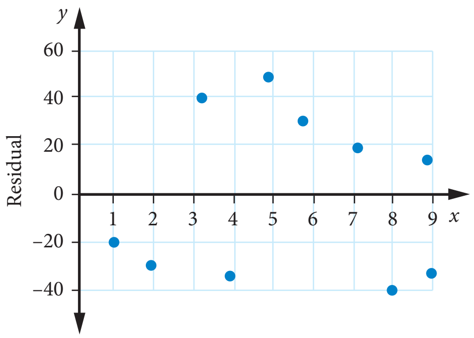


The corresponding residual plot is closest to

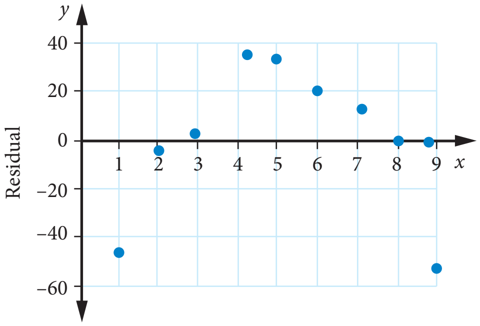
A



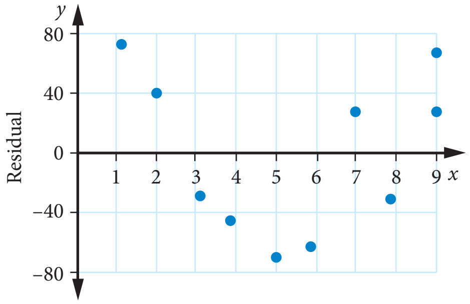
B



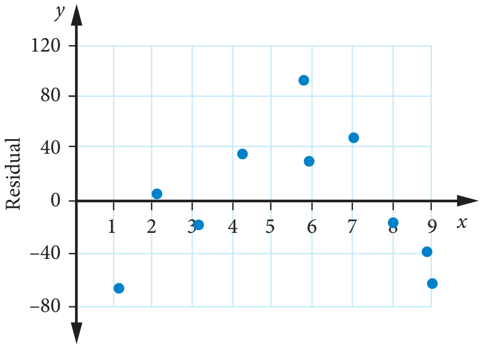
C



D



E

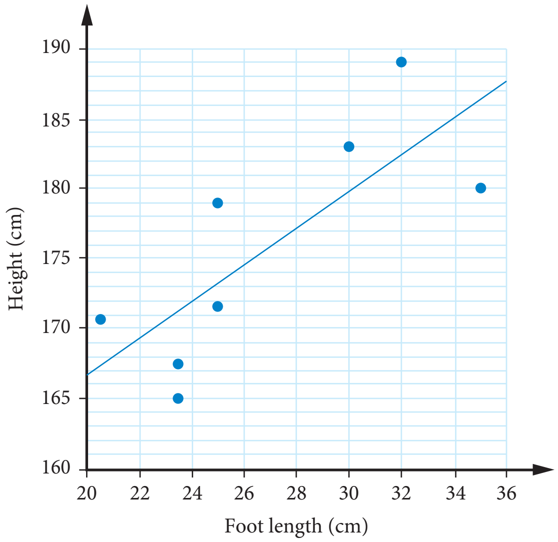


[VCAA 2013 1CQ11]

Use the following information to answer Questions 2–4.

The *height* (in cm) and *foot length* (in cm) for each of eight Year 12 students were recorded and displayed in the scatterplot.

A least squares line of best fit has been fitted to the data as shown.



Question 2

By inspection, the value of the correlation coefficient (r) for this data is closest to

A 0.98 **B** 0.78

C 0.23 **D** −0.44

E −0.67

[VCAA 2010 1CQ7]

Question 3

The explanatory variable is *foot length*.

The equation of the least squares line of best fit is closest to

A *height* = −110 + 0.78 × *foot length*.

B *height* = 141 + 1.3 × *foot length*.

C *height* = 167 + 1.3 × *foot length*.

D *height* = 167 + 0.67 × *foot length*.

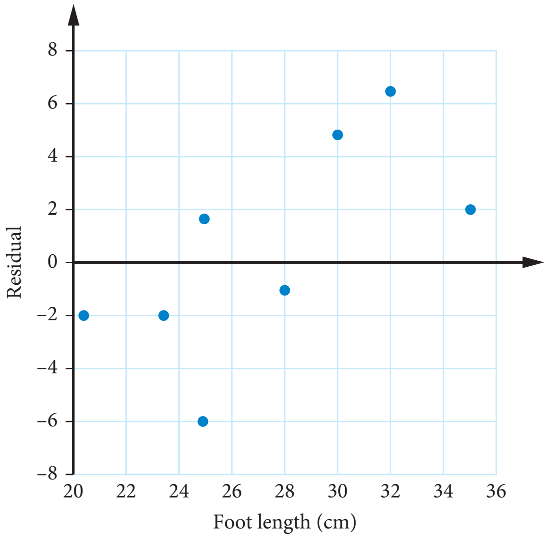
E *foot length* = 167 + 1.3 × *height*.

[VCAA 2010 1CQ8]

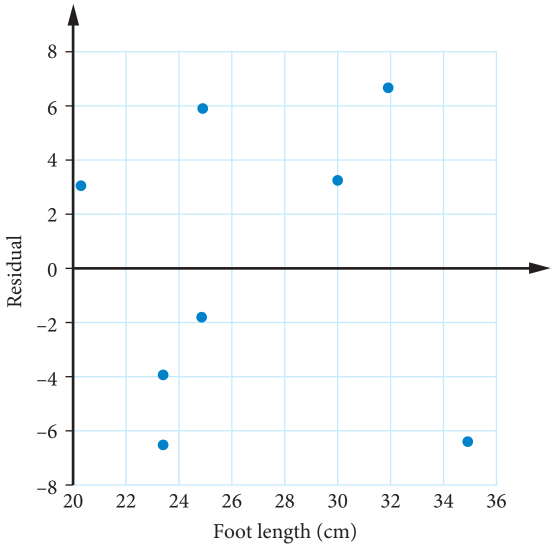
Question 4

The plot of the *residuals* against *foot length* is closest to

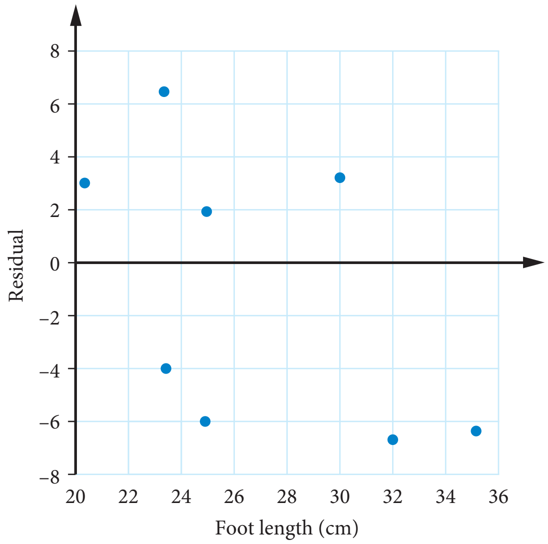
A



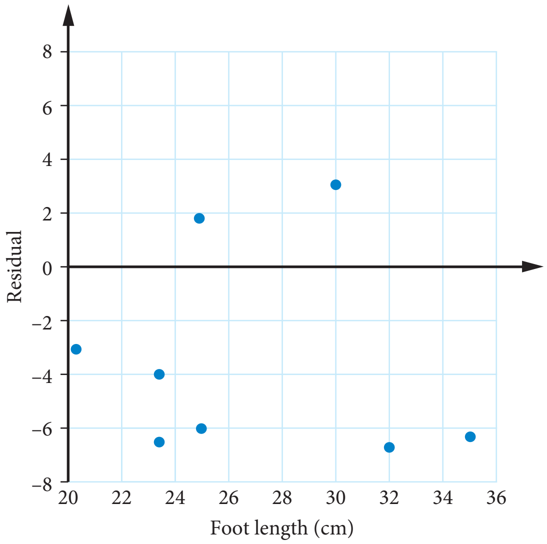
B



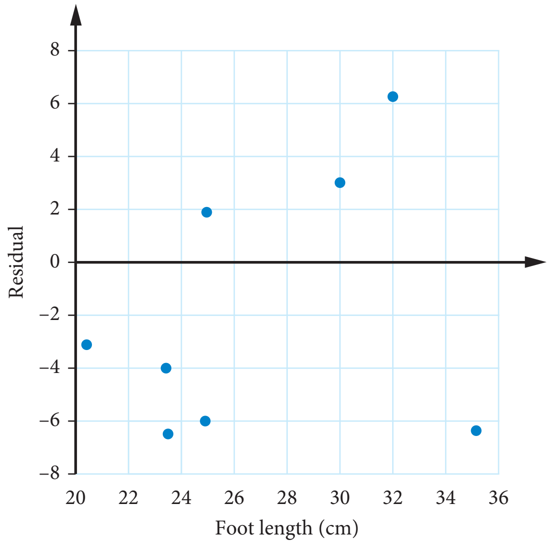
C



D



E



[VCAA 2010 1CQ9]

Question 5

Which one of the following statistics is never negative?

A a median

B a residual

C a standardised score

D an interquartile range

E a correlation coefficient

[VCAA 2012 1CQ10]

Question 6

The table lists the average *body mass* (in kg) and average *brain mass* (in g) of nine animal species.

Species	Body mass (kg)	Brain weight (g)
Baboon	10.55	179.5
Cat	3.30	25.6
Goat	27.70	115.0
Guinea pig	1.04	5.5
Rabbit	2.50	12.1
Rat	0.28	1.9
Red fox	4.24	50.4
Rhesus monkey	6.80	179.0
Sheep	55.50	175.0

A least squares line of best fit is fitted to the data using *body mass* as the explanatory variable.

The equation of the least squares line of best fit is

$$\textit{brain mass} = 49.4 + 2.68 \times \textit{body mass}$$

This equation is then used to predict the *brain mass* (in g) of the baboon. The residual value (in g) for this prediction will be closest to

A –351 **B** –102 **C** –78 **D** 78 **E** 102

[VCAA 2009 1CQ11]

Question 7

The waist measurement (cm) and mass (kg) of 12 men are displayed in the table below.

Waist (cm)	84	74	89	75	106	114	80	101	101	94	126	82
Mass (kg)	84	72	67	59	97	112	67	91	98	89	117	62

Using this data, the equation of the least squares line of best fit that enables mass to be predicted from waist measurement is

$$\text{mass} = -20 + 1.11 \times \text{waist}$$

When this equation is used to predict the mass of the man with a waist measurement of 80 cm, the residual value is closest to

A –11 kg **B** 11 kg **C** –2 kg **D** 2 kg **E** 69 kg

[VCAA 2006 1CQ8]

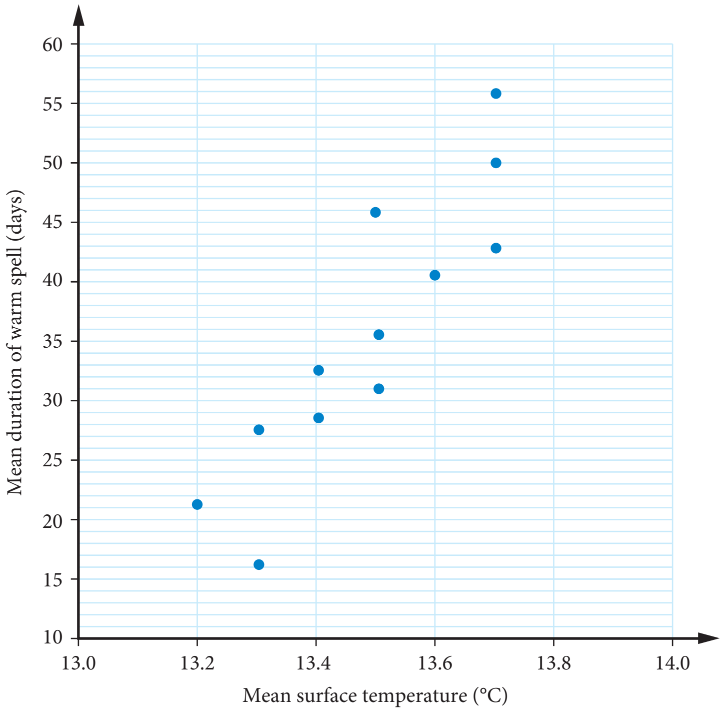
Question 8

The table displays the mean surface temperature (in °C) and the mean duration of warm spell (in days) in Australia for 13 years selected at random from the period 1960 to 2005.

Mean surface temperature (°C)	Mean duration of warm spell (days)	Mean surface temperature (°C)	Mean duration of warm spell (days)
13.2	21.4	13.5	35.5
13.3	16.3	13.6	40.6
13.3	27.6	13.7	42.8
13.4	32.6	13.7	49.9
13.4	28.7	13.7	55.8
13.5	30.9	**13.8**	**53.1**
13.5	45.9		

This data set has been used to construct the scatterplot. The scatterplot is incomplete.

a Copy and complete the scatterplot by plotting the **bold** data values given in the table above. Mark the point with a cross (×). 1 mark

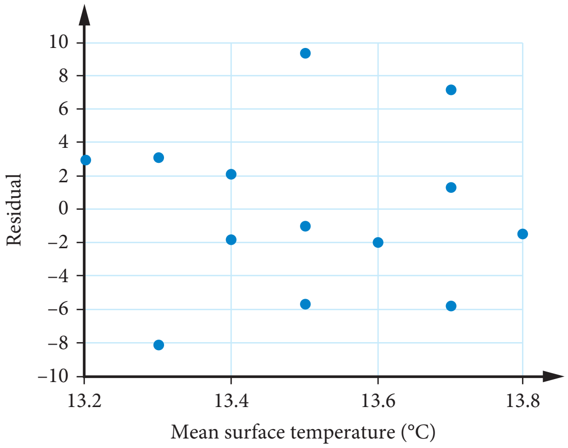


b Mean surface temperature is the explanatory variable.

- **i** Determine the equation of the least squares line of best fit for this set of data. Write the equation in terms of the variables *mean duration of warm spell* and *mean surface temperature*. Write the values of the coefficients correct to one decimal place. 2 marks
- **ii** Plot the least squares line of best fit on the scatterplot. 1 mark

c The residual plot was constructed to test the assumption of linearity for the association between the variables *mean duration of warm spell* and the *mean surface temperature*.

Explain why this residual plot supports the assumption of linearity for this association. 1 mark



d Write down the percentage of variation in the mean duration of a warm spell that is explained by the variation in mean surface temperature. Write your answer correct to the nearest per cent. 1 mark

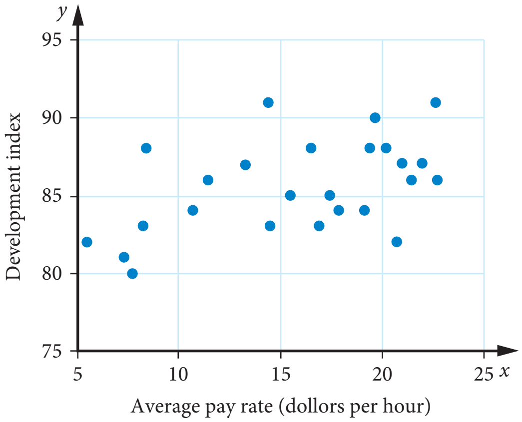
e Describe the association between the mean duration of a warm spell and the mean surface temperature in terms of strength, direction and form. 2 marks

[VCAA 2007 2CQ3]

Question 9

The development index and the average pay rate for workers, in dollars per hour, for a selection of 25 countries are displayed in the scatterplot.

The table contains the values of some statistics that have been calculated for this data.



Statistic	Average pay rate (x)	Development index (y)
Mean	$\bar{x} = 15.7$	$\bar{y} = 85.6$
Standard deviation	$s_x = 5.37$	$s_y = 2.99$
Correlation coefficient	$r = 0.488$	

a Determine the standardised value of the development index (z-score) for a country with a development index of 91. Write your answer, correct to 1 decimal place. 1 mark

b Use the information in the table to show that the equation of the least squares line of best fit for a country's development index, y, in terms of its average pay rate, x, is given by

$y = 81.3 + 0.272x$ 2 marks

c The country with an average pay rate of $14.30 per hour has a development index of 83.

Determine the residual value when the least squares line of best fit given in **part b** is used to predict this country's development index.

Write your answer, correct to 1 decimal place. 2 marks

[VCAA 2013 2CQ3]

Question 10

Table 1 shows the number of telephone calls (both internal calls and external calls) made on a given day by a sample of 12 people working in a large company. Also given is the cost of each person's calls for the day.

Table 1

Person	Number of calls	Cost (dollars)
A	33	4.54
B	15	1.00
C	22	5.96
D	27	4.47
E	52	8.87
F	34	8.50
G	55	11.09
H	47	8.51
I	11	3.98
J	18	2.42
K	36	11.30
L	27	7.48

a Determine the mean and standard deviation of the **cost** of the calls. Write your answers in dollars, correct to 2 decimal places.

2 marks

b For all workers in this company, the mean **number** of telephone calls made per day is 21.8.

For the sample of 12 workers in Table 1, determine the percentage of these workers who made more calls than the company mean of 21.8 calls. 1 mark

c Use the data in **Table 1** to

i determine the equation of the least squares line of best fit that will enable call costs per person to be predicted from the number of calls they make. Write the missing coefficient correct to 2 decimal places in the space provided.

cost = 0.66 + ☐ × number of calls 1 mark

ii determine the value of the Pearson correlation coefficient, *r*. Write your answer correct to 4 decimal places. 1 mark

d The value of the Pearson correlation coefficient (calculated in part **c ii**) measures the strength and direction of the ☐ association between call cost and number of calls.

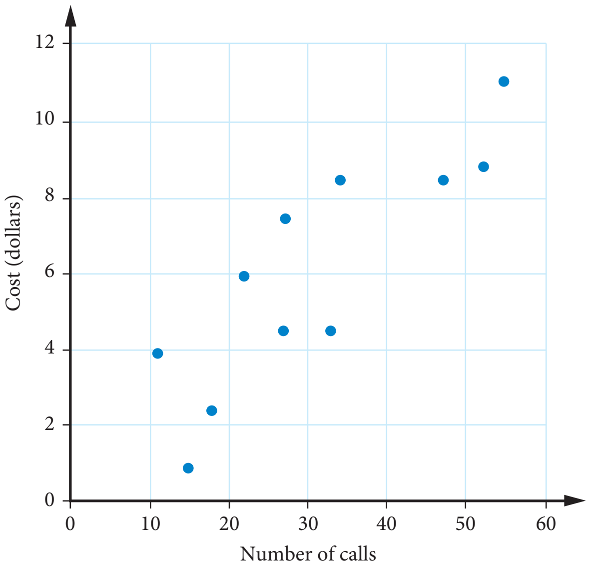
1 mark

e The scatterplot shown in **Figure 1** was constructed from the data displayed in **Table 1**. The point corresponding to person *K* has not been included. Copy and complete the scatterplot by adding in the data point for person *K*, marking the point with a cross (×). 1 mark

f We wish to predict the cost of calls from the number of calls made.

The response variable is

☐. 1 mark



g Complete the following sentences by filling in the boxes.

i On average, for each extra call made, a worker's call costs increase by $\boxed{}$ cents per call. 1 mark

ii To the nearest whole per cent, $\boxed{}$ % of the variation in call costs can be explained by the variation in the number of calls. 1 mark

h If we use the least squares line of best fit to estimate the call costs of a person who makes 34 calls, this would leave a residual value of $\boxed{}$ dollars. 1 mark

[VCAA 2003 2CQ1]

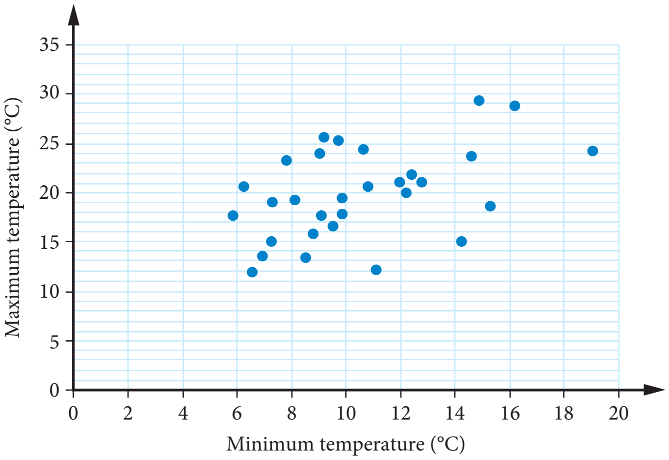
Question 11

The maximum temperature and the minimum temperature at this weather station on each of the 30 days in November 2011 are displayed in the scatterplot.

The correlation coefficient for this data set is $r = 0.630$.

The equation of the least squares line of best fit for this data set is

maximum temperature =
13 + 0.67 × *minimum temperature*



a Copy the scatterplot and on it draw this least squares line of best fit. 1 mark

b Interpret the vertical intercept of the least squares line of best fit in terms of maximum temperature and minimum temperature. 1 mark

c Describe the association between the maximum temperature and the minimum temperature in terms of strength and direction. 1 mark

d Interpret the slope of the least squares line of best fit in terms of maximum temperature and minimum temperature. 1 mark

e Determine the percentage of variation in the maximum temperature that may be explained by the variation in the minimum temperature. Write your answer, correct to the nearest percentage. 1 mark

On the day that the minimum temperature was 11.1 °C, the actual maximum temperature was 12.2 °C.

f Determine the residual value for this day if the least squares line of best fit is used to predict the maximum temperature. Write your answer, correct to the nearest degree. 2 marks

[VCAA 2012 2CQ2]

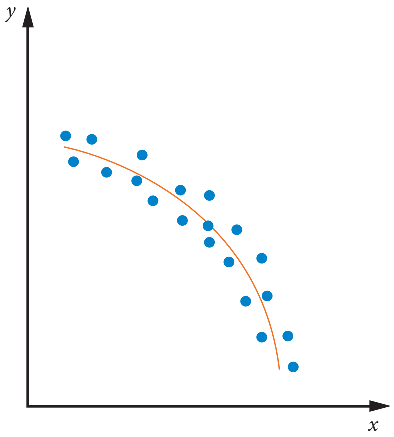
Data transformation

Types of data transformation

We now have a number of ways of dealing with linear associations. Not all associations are linear, however. The way that we deal with non-linear associations is to apply a **transformation** to one of the variables so that the association between the two variables becomes closer to a straight line.

Using a data transformation is similar to changing the units of measurement you work with (for example from inches to centimetres). It is simply another way of working with the same data.

When deciding which transformation to use it's important to look at what happens to the large x or large y values of the data set.

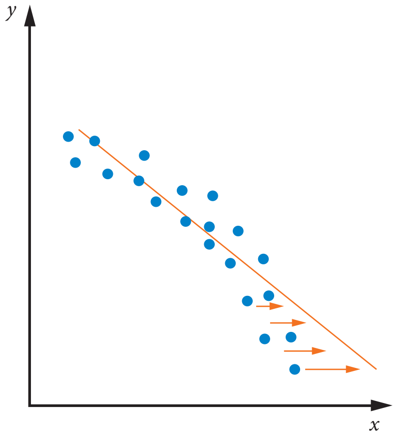


There are three types of data transformations.

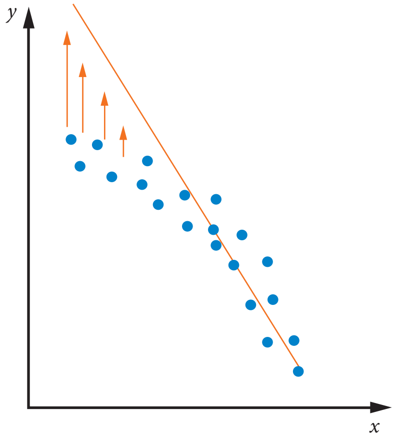
1. Squared transformation

This involves squaring either the x values or the y values. So 5 becomes 25, 9 becomes 81, 100 becomes 10 000 and so on. Large values increase more than small values, and the effect of the transformation is that the data is *stretched* either horizontally (for an x^2 transformation) or vertically (for a y^2 transformation).

To stretch large x values, use x^2



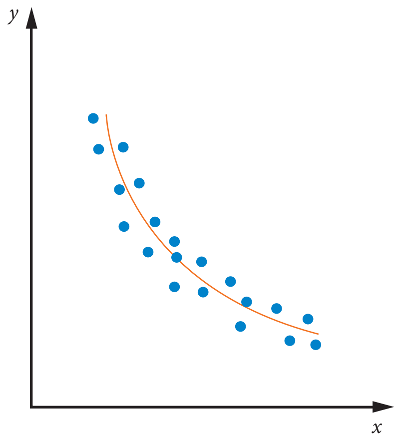
To stretch large y values, use y^2



2. Log transformation

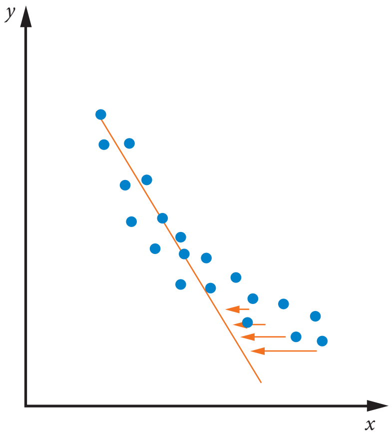
This involves finding the log base 10 of either the x values or the y values. So 5 becomes log (5) = 0.7, 9 becomes log (9) = 0.95, 100 becomes log (100) = 2, and so on. Large values are reduced more than small values, and the effect of the transformation is that the data is *compressed* either horizontally (for a log x transformation) or vertically (for a log y transformation).

3. Reciprocal transformation

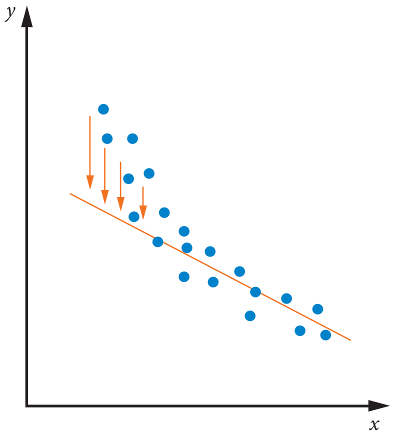


This involves taking the reciprocal of either the x values or the y values. So 5 becomes $\frac{1}{5} = 0.2$, 9 becomes $\frac{1}{9} = 0.11$, 100 becomes $\frac{1}{100} = 0.01$ and so on. As with the log transformation, large values are reduced more than small values, and the effect of the transformation is that the data is *compressed* either horizontally (for a $\frac{1}{x}$ transformation) or vertically (for a $\frac{1}{y}$ transformation). The compression of large values for a reciprocal transformation is greater than for a log transformation.

To compress large x values, use $\frac{1}{x}$ or log (x)



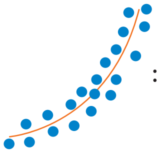
To compress large y values, use $\frac{1}{y}$ or log (y)



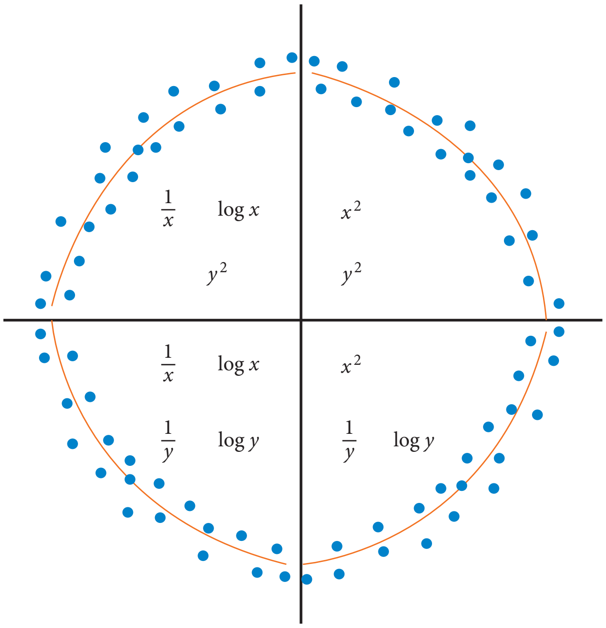
iStock.com/egal

The transformation circle

The transformation circle below summarises when the transformations that we have looked at can be used to linearise non-linear data.

For example, there are three transformation options for data shaped like  : x^2, $\frac{1}{y}$ or $\log(y)$.

It is difficult to tell in advance which of the options would give you the most linear result, so if there is no information about which is the best option, calculate the coefficient of determination of each of them to see which is closest to 1.



Worked example 8

The data for the association between the maximum daily temperature (°C) and the daily number of boxes of hot pies sold at the school canteen has been linearised by applying a reciprocal transformation to the temperature, giving the following least squares line of best fit.

$$\text{No. of boxes of pies} = -0.7 + \frac{90}{\text{temperature}}$$

Use this equation to predict the number of boxes of pies sold when the temperature was

a 17°C **b** 25°C **c** 30°C

	Working
a Substitute the value into the equation, solve, and round the answer to the nearest whole number.	No. of boxes of pies $= -0.7 + \frac{90}{17} = 4.6$ 5 boxes of pies
b Substitute the value into the equation, solve, and round the answer to the nearest whole number.	No. of boxes of pies $= -0.7 + \frac{90}{25} = 2.9$ 3 boxes of pies
c Substitute the value into the equation, solve, and round the answer to the nearest whole number.	No. of boxes of pies $= -0.7 + \frac{90}{30} = 2.3$ 2 boxes of pies

Exam hack

Sometimes you need to round to the nearest whole number, even though you are not specifically told to, because of the context of the question.

Using CAS Transforming non-linear data

Use the coefficient of determination to decide which transformation is the best choice for linearising the following data, and write down the equation of the least squares line of best fit in terms of the transformed variables, with the slope and intercept correct to 2 decimal places.

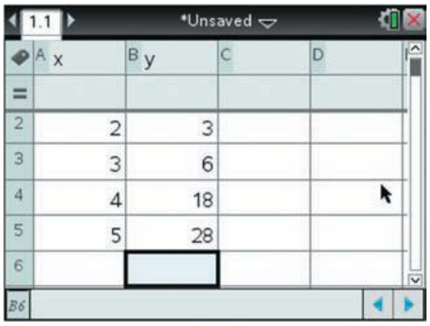
x	1	2	3	4	5
y	2	3	6	18	28

TI-NSPIRE CAS

STEP 1

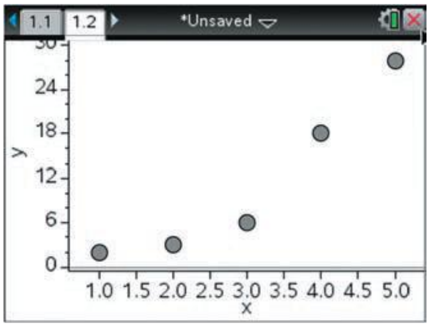
Open a New Document with a Lists & Spreadsheet page. Enter the data and construct the scatterplot.

Name column A x and column B y



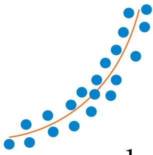
	A x	B y	C	D
=				
2	2	3		
3	3	6		
4	4	18		
5	5	28		
6				

B6



STEP 2

Using the transformation circle, decide which of the transformations to use to linearise the data.

The shape of the data is , so the options to linearise the data are x^2, $\frac{1}{y}$ or log (y).

STEP 3

Enter the transformed data for the first transformation option, x^2.

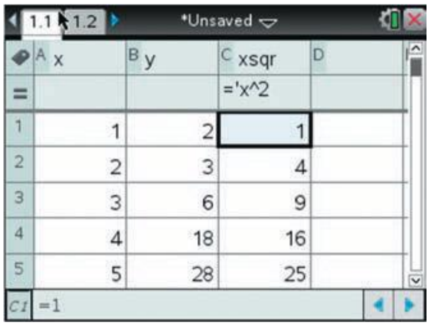
Name column C xsqr

In the formula cell for column C enter =x

Press x^2

If you name the column first the display will be xsqr:=

The display will require a variable reference for *x*



	A x	B y	C xsqr	D
=			='x^2	
1	1	2	1	
2	2	3	4	
3	3	6	9	
4	4	18	16	
5	5	28	25	

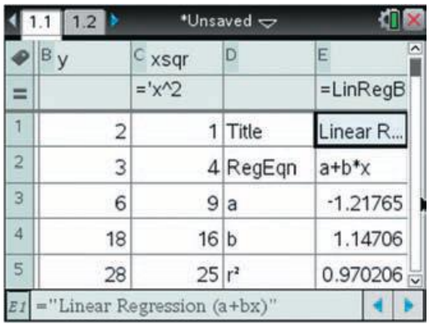
C1 =1

STEP 4

Press menu, then 4: Statistics followed by 1: Stat Calculations, then 4: Linear Regression (a + bx). Complete the screen as shown and press menu.

Write down the values of values of r^2, a and b correct to 2 decimal places.

X List: 'xsqr
Y List: y
Save RegEqn to: f1
Frequency List: 1
Category List:
Include Categories:



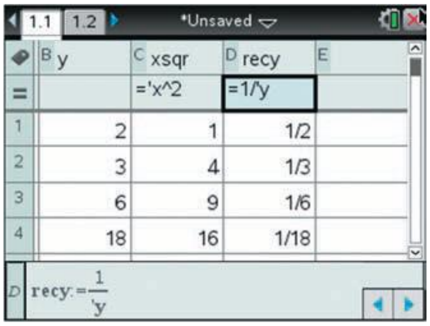
	B y	C xsqr	D	E
=		='x^2		=LinRegB
1	2	1	Title	Linear R...
2	3	4	RegEqn	a+b*x
3	6	9	a	-1.21765
4	18	16	b	1.14706
5	28	25	r²	0.970206

E1 ="Linear Regression (a+bx)"

$r^2 = 0.97$
$a = -1.22$
$b = 1.15$

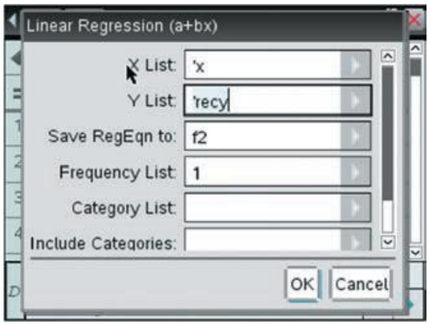
STEP 5

Repeat for the other transformation option, $\frac{1}{y}$.

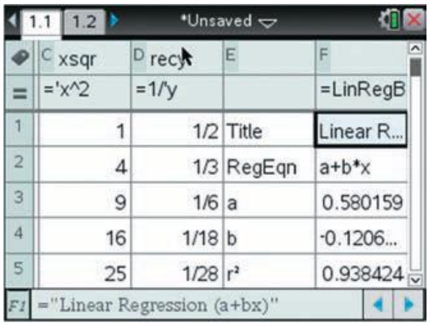


	B y	C xsqr	D recy	E
=		='x^2	=1/'y	
1	2	1	1/2	
2	3	4	1/3	
3	6	9	1/6	
4	18	16	1/18	

D recy:= $\frac{1}{'y}$


Linear Regression (a+bx)

X List: 'x
Y List: 'recy
Save RegEqn to: f2
Frequency List: 1
Category List:
Include Categories:



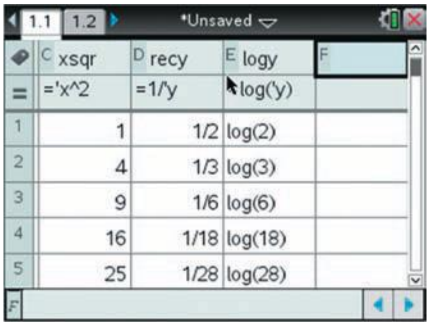
	C xsqr	D recy	E	F
=	='x^2	=1/'y		=LinRegB
1	1	1/2	Title	Linear R...
2	4	1/3	RegEqn	a+b*x
3	9	1/6	a	0.580159
4	16	1/18	b	-0.1206...
5	25	1/28	r²	0.938424

F1 ="Linear Regression (a+bx)"

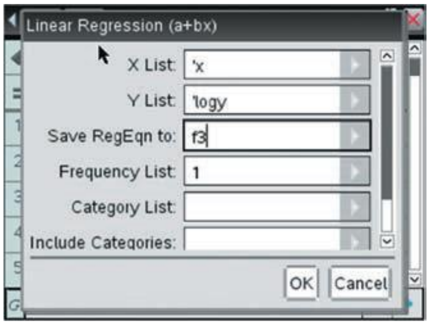
$r^2 = 0.94$
$a = 0.58$
$b = -0.12$

STEP 6

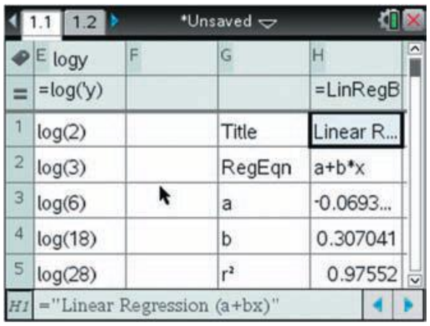
Repeat for the other transformation option, log (*y*).



	C xsqr	D recy	E logy	F
=	='x^2	=1/'y	log('y)	
1	1	1/2	log(2)	
2	4	1/3	log(3)	
3	9	1/6	log(6)	
4	16	1/18	log(18)	
5	25	1/28	log(28)	


Linear Regression (a+bx)

X List: 'x
Y List: 'logy
Save RegEqn to: f3
Frequency List: 1
Category List:
Include Categories:



	E logy	F	G	H
=	=log('y)			=LinRegB
1	log(2)		Title	Linear R...
2	log(3)		RegEqn	a+b*x
3	log(6)		a	-0.0693...
4	log(18)		b	0.307041
5	log(28)		r²	0.97552

H1 ="Linear Regression (a+bx)"

$r^2 = 0.98$
$a =-0.07$
$b = 0.31$

STEP 7

Choose the option with the coefficient of determination closest to 1.

The coefficient of determination of the log (*y*) transformation is closest to 1, so the log (*y*) transformation is the best choice for linearising the data.

STEP 8

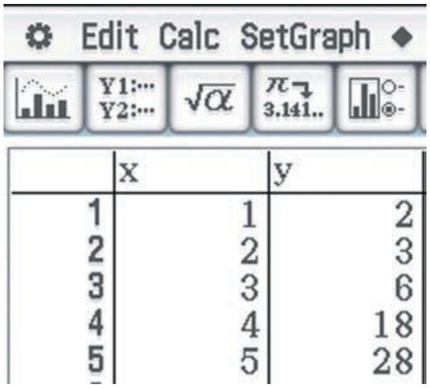
Use the *a* and *b* values to write the equation of the least squares line of best fit for this option.

log (*y*) = –0.07 + 0.31*x*

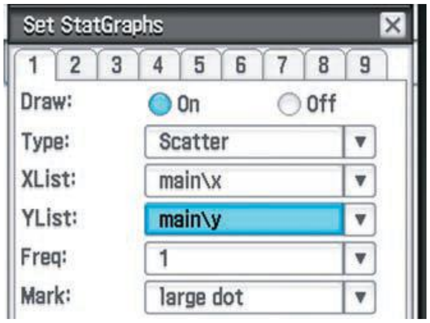
CLASSPAD

STEP 1

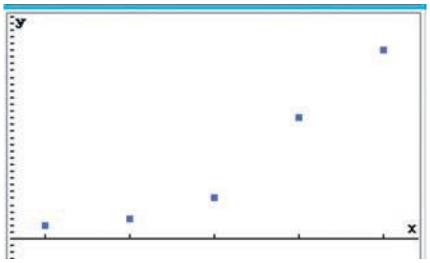
Using the **Statistics** application to enter the data and construct the scatterplot.



	x	y
1	1	2
2	2	3
3	3	6
4	4	18
5	5	28

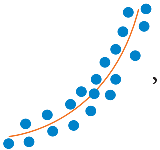


Draw: ● On ○ Off
Type: Scatter
XList: main\x
YList: main\y
Freq: 1
Mark: large dot



STEP 2

Using the transformation circle, decide which of the transformations to use to linearise the data.

The shape of the data is , so the options to linearise the data are x^2, $\frac{1}{y}$ or log (y).

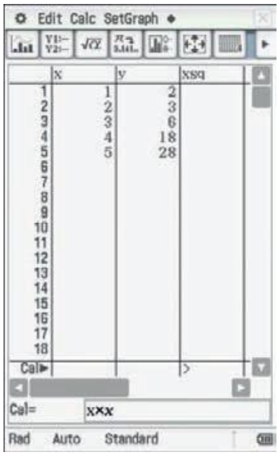
STEP 3

Name list3 as *x*sq (or something similar).

Tap the Cal cell at the bottom of the xsq column.

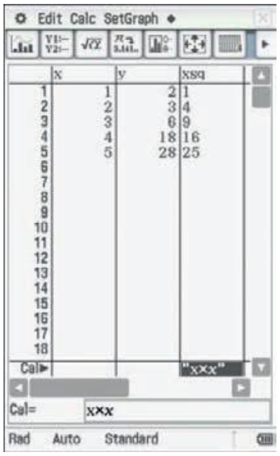
Tap the 'Cal=' box and enter x×x. See left screen below. Do not enter '=x×x' (no equals sign) and don't tap the heading to enter x. You must type it.

Press EXE. The third column will fill with the squares of the first column.



	x	y	xsq
1	1	2	
2	2	3	
3	3	6	
4	4	18	
5	5	28	

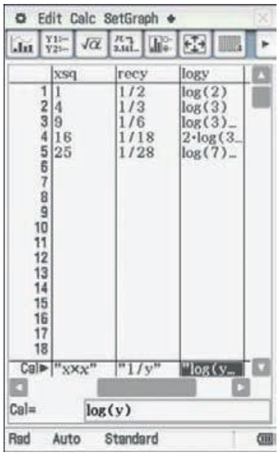
Cal= x×x



	x	y	xsq
1	1	2	1
2	2	3	4
3	3	6	9
4	4	18	16
5	5	28	25
Cal▸			"x×x"

Cal= x×x

Repeat for 1/*y* and log (*y*), using the fourth and fifth columns. Use the slider near the bottom of the screen to move the columns.



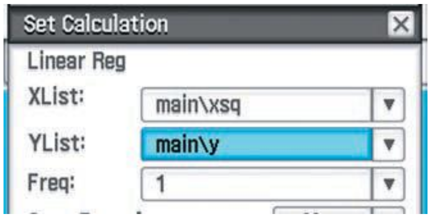
	xsq	recy	logy
1	1	1/2	log(2)
2	4	1/3	log(3)
3	9	1/6	log(3)…
4	16	1/18	2·log(3…
5	25	1/28	log(7)…
Cal▸	"x×x"	"1/y"	"log(y…

Cal= log(y)

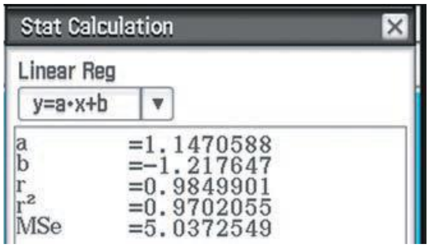
STEP 4

Tap **Calc**, then select Regression, then **Linear Reg**. Complete the screen as shown and tap OK.

Write down the values of values of r^2, a and b correct to 2 decimal places.



Linear Reg
XList: main\xsq
YList: main\y
Freq: 1



Linear Reg
y=a·x+b

```
a      =1.1470588
b      =-1.217647
r      =0.9849901
r²     =0.9702055
MSe    =5.0372549
```

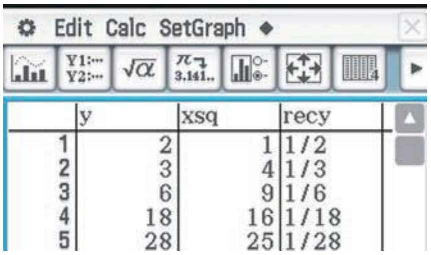
$r^2 = 0.97$

$a = 1.15$

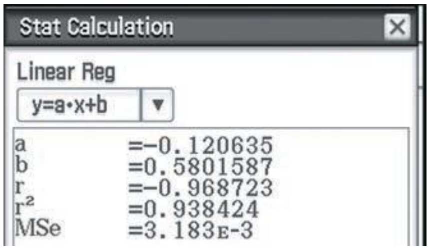
$b = -1.22$

STEP 5

Repeat for the other transformation option, $\frac{1}{y}$.



	y	xsq	recy
1	2	1	1/2
2	3	4	1/3
3	6	9	1/6
4	18	16	1/18
5	28	25	1/28



Linear Reg

y=a·x+b

a =−0.120635
b =0.5801587
r =−0.968723
r^2 =0.938424
MSe =3.183E-3

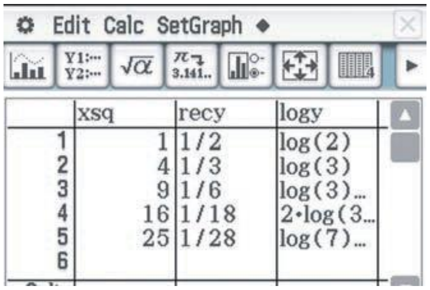
$r^2 = 0.94$

$a = -1.12$

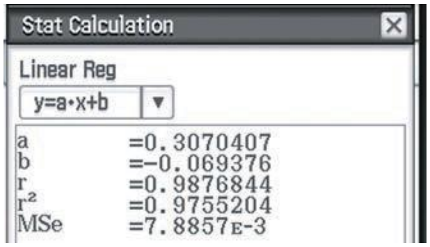
$b = 0.58$

STEP 6

Repeat for the other transformation option, log (y).



	xsq	recy	logy
1	1	1/2	log(2)
2	4	1/3	log(3)
3	9	1/6	log(3)…
4	16	1/18	2·log(3…
5	25	1/28	log(7)…
6			



Linear Reg

y=a·x+b

a =0.3070407
b =−0.069376
r =0.9876844
r^2 =0.9755204
MSe =7.8857E-3

$r^2 = 0.98$

$a = 0.31$

$b = -0.07$

STEP 7

Choose the option with the coefficient of determination closest to 1.

The coefficient of determination of the log (y) transformation is closest to 1, so the log (y) transformation is the best choice for linearising the data.

STEP 8

Use the a and b values to write the equation of the least squares line of best fit for this option.

log (y) = −0.07 + 0.31x

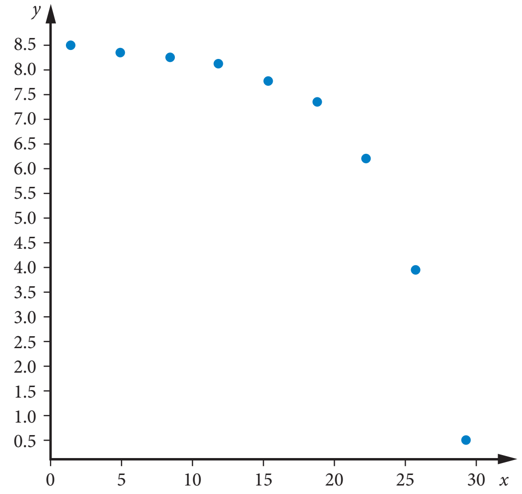
EXAM PREP 4.5

Data transformation

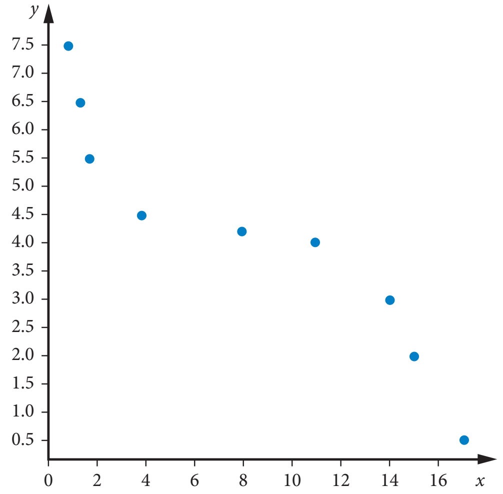
Prep 1

For each of the following scatterplots, state which of the transformations x^2, $\frac{1}{x}$, log (x), y^2, $\frac{1}{x}$ or log (y) (if any) you would use to linearise the association.

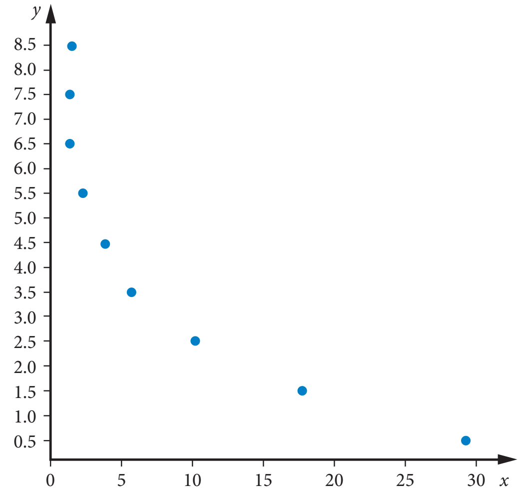
a



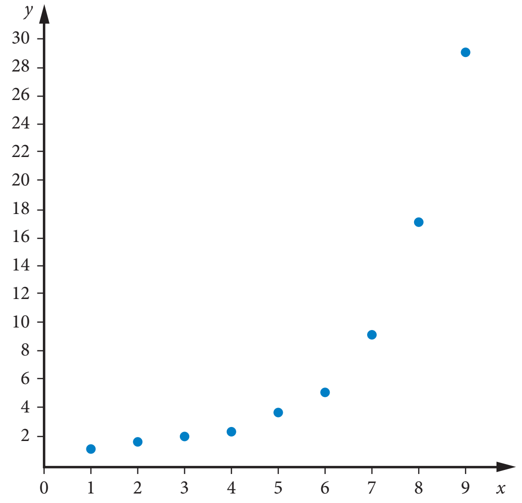
b



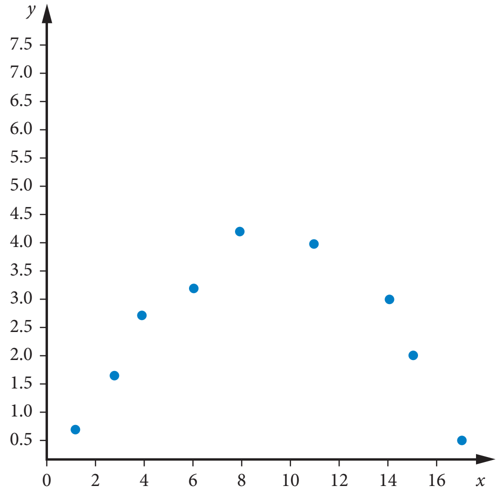
c



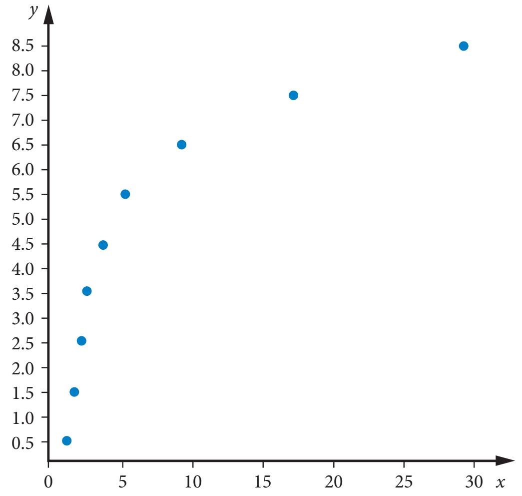
d



e



f



Prep 2 WORKED EXAMPLE 8

The data for the association between the diameter of a particular fruit (cm) and the number of seeds it contains has been linearised by applying a squared transformation to the number of seeds, giving the following least squares line of best fit.

$(\text{number of seeds})^2 = 36 + 8 \times \text{diameter}$

Use this equation to predict the number of seeds in a fruit of diameter

a 8 cm **b** 4 cm **c** 1 cm

Prep 3 USING CAS: TRANSFORMING NON-LINEAR DATA

Use the coefficient of determination to decide which transformation is the best choice for linearising the following data, and write down the equation of the least squares line of best fit in terms of the transformed variables, with the slope and intercept correct to 2 decimal places.

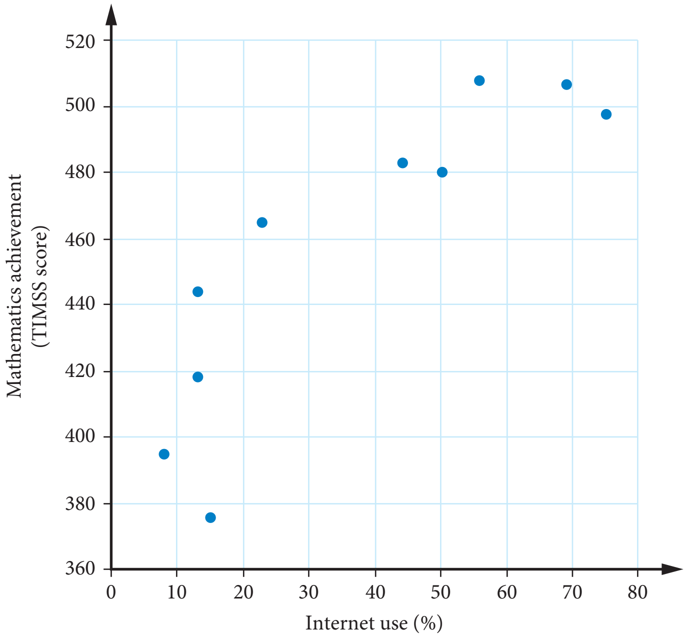
x	1	2	3	4	5
y	28	18	6	3	2

EXAM PRACTICE 4.5

Data transformation

Question 1

The *mathematics achievement* level (TIMSS score) for grade 8 students and the general rate of *Internet use* (%) for 10 countries are displayed in the scatterplot.



To linearise the data, it would be best to plot

A *mathematics achievement* against *Internet use*.

B log (*mathematics achievement*) against *Internet use*.

C *mathematics achievement* against log (*Internet use*).

D *mathematics achievement* against log $(\textit{Internet use})^2$.

E $\dfrac{1}{\textit{mathematics achievement}}$ against *Internet use*

[VCAA 2009 1CQ12]

Question 2

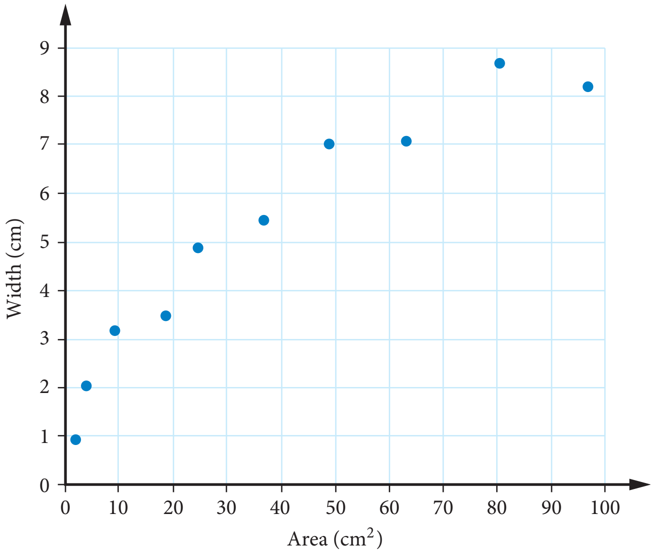
The data in the scatterplot shows the *width*, in cm, and the surface *area*, in cm^2, of leaves sampled from 10 different trees. The scatterplot is non-linear.

To linearise the scatterplot, $(width)^2$ is plotted against *area* and a least squares line of best fit is then fitted to the linearised plot.

The equation of this least squares line of best fit is

$(width)^2 = 1.8 + 0.8 \times area$

Using this equation, a leaf with a surface area of 120 cm^2 is predicted to have a width, in cm, closest to



A 9.2 **B** 9.9 **C** 10.6 **D** 84.6 **E** 97.8

[VCAA 2013 1CQ10]

Question 3

The association between the two variables y and x, as shown in the scatterplot, is non-linear.

Which one of the following transformations, by itself, is most likely to linearise this data?

A a $\frac{1}{x}$ transformation

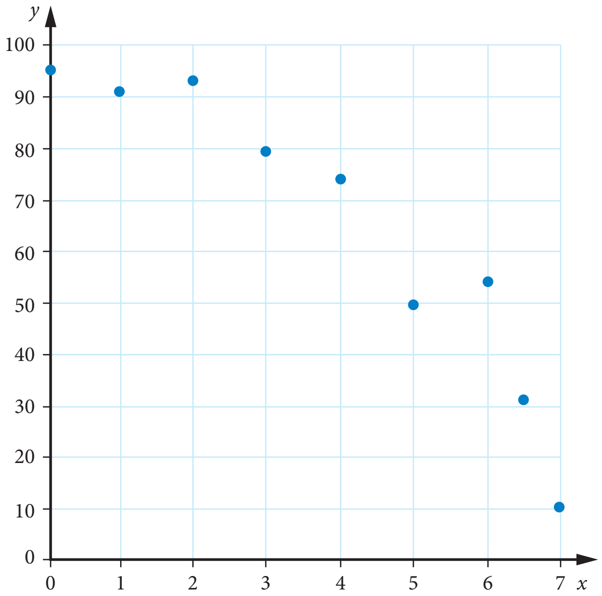
B a $\frac{1}{y}$ transformation

C an x^2 transformation

D a log (x) transformation

E a log (y) transformation

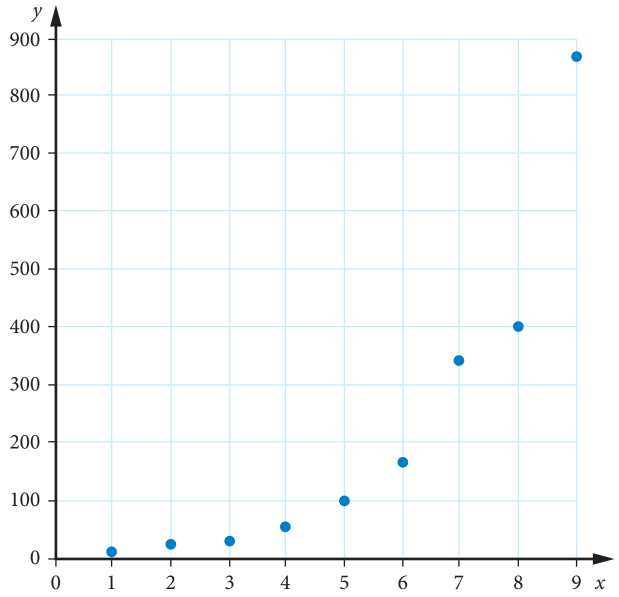
[VCAA 2003 1CQ10]



Question 4

A student uses the following data to construct the scatterplot shown.

x	1	2	3	4	5	6	7	8	9
y	12	25	33	58	98	168	345	397	869



To linearise the scatterplot, she applies a **log (y)** transformation; that is, a log transformation is applied to the y-axis scale.

She then fits a least squares line of best fit to the transformed data.

With x as the explanatory variable, the equation of this least squares line of best fit is closest to

A $\log(y) = -217 + 88.0x$

B $\log(y) = -3.8 + 4.4x$

C $\log(y) = 3.1 + 0.008x$

D $\log(y) = 0.88 + 0.23x$

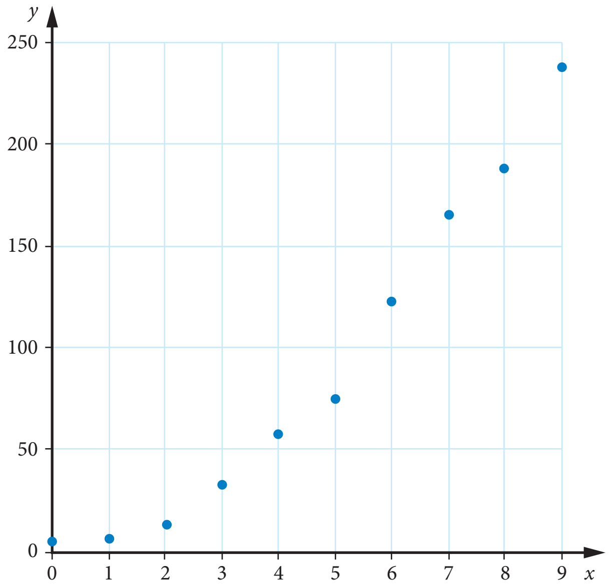
E $\log(y) = 1.58 + 0.002x$

[VCAA 2007 1CQ9]

Question 5

A student uses the following data to construct the scatterplot shown.

x	0	1	2	3	4	5	6	7	8	9
y	5	7	14	33	58	76	124	166	188	238



To linearise the scatterplot, she applies an x-squared transformation.

She then fits a least squares line of best fit to the **transformed data** with y as the response variable. The equation of this least squares line of best fit is closest to

A $y = 7.1 + 2.9x^2$

B $y = -29.5 + 6.8x^2$

C $y = 26.8 - 29.5x^2$

D $y = 1.3 + 0.04x^2$

E $y = -2.2 + 0.3x^2$

[VCAA 2006 1CQ9]

Question 6

A student uses the following data to construct the scatterplot shown.

x	1	2	3	4	5	6	7	8
y	245	130	84	58	52	36	33	30

A reciprocal transformation is applied to the x-axis and is used to linearise the scatterplot.

With y as the response variable, the slope of the least squares line of best fit that is fitted to the **linearised** plot is closest to

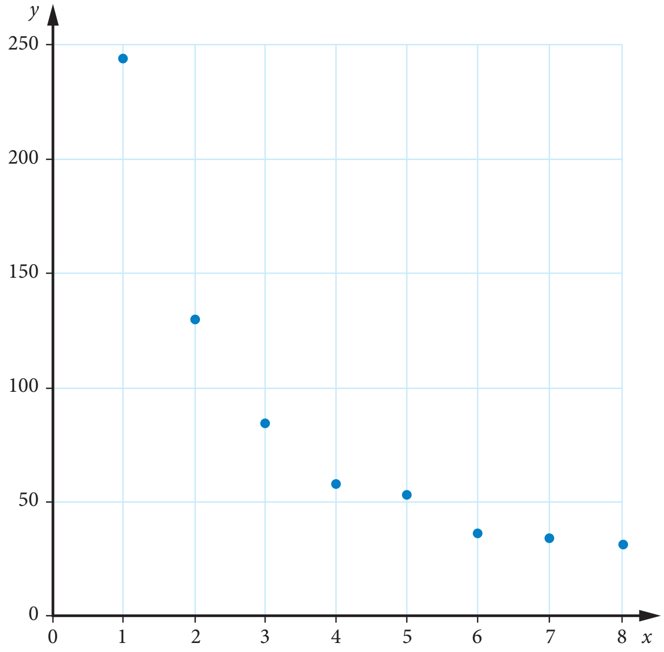
A –249

B –25

C 0.004

D 25

E 249



[VCAA 2010 1CQ11]

Question 7

Cars depreciate in value over time. **Table 1** gives the average value of a car (of the same brand and model) at different ages.

Table 1.

Age (years)	1	2	3	4	5	6	7	8	9
Value (dollars)	18 100	15 050	13 900	11 900	10 400	9 600	8 900	8 500	8 400

a The data is to be used to build a mathematical association that will enable the average value of this brand and model of car to be predicted from its age.

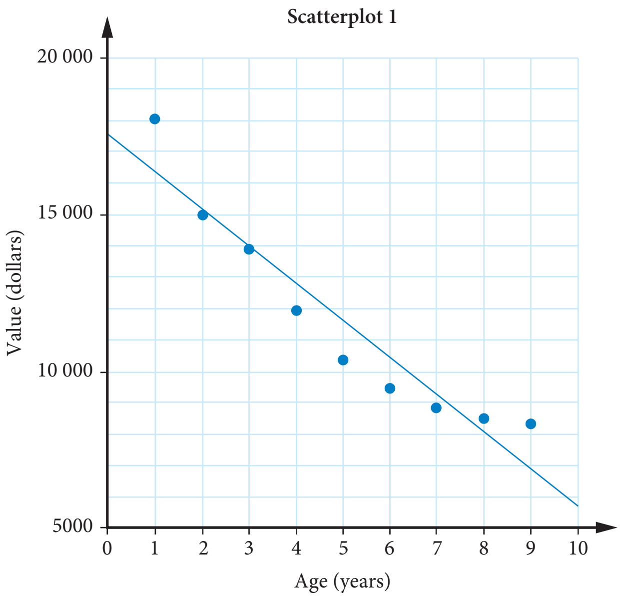
In this situation, the **response** variable is ☐. 1 mark

Scatterplot 1 is constructed from the data and a least squares line of best fit is fitted as shown.

b The coefficient of determination for this data is 0.9058.

- **i** Find the value of the correlation coefficient, *r*, correct to 3 decimal places. 1 mark
- **ii** Write down the percentage of the variation in the value of a car that can be accounted for by the variation in its age. 1 mark

c Using the line shown in **Scatterplot 1**, or otherwise, determine the equation of the least squares line of best fit. Write the coefficients correct to the nearest hundred.

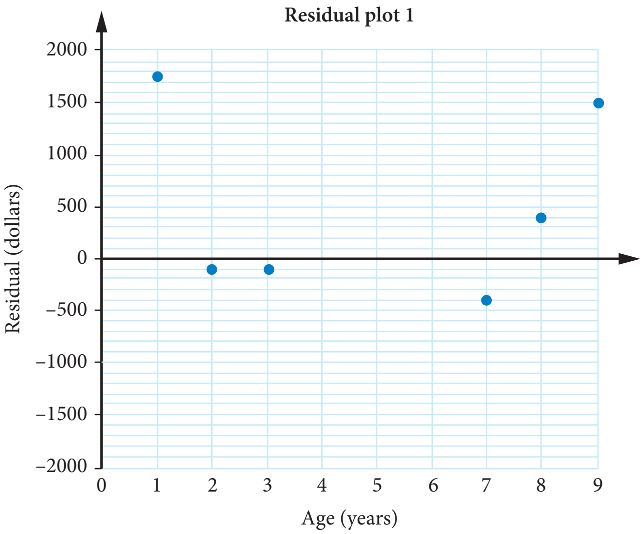


value = ☐ + ☐ × *age* 2 marks

Scatterplot 1 suggests that the association may be non-linear. To investigate this idea, **Residual plot 1** is constructed. It is incomplete.

d Using the information in **Scatterplot 1**, or otherwise, copy and complete **Residual plot 1** by marking in the missing residual values for cars aged 4, 5 and 6 years. 2 marks

e When complete, does **Residual plot 1** suggest that a non-linear association will provide a better fit for the data? Justify your response. 1 mark



Scatterplot 1 indicates that a logarithmic transformation of the horizontal (*age*) axis may linearise the data. The original data has been reproduced in **Table 2**. An extra row has been added for the transformed variable, log (*age*). The table is incomplete.

Table 2.

Age (years)	1	2	3	4	5	6	7	8	9
Log (age)	0	0.30	0.48	0.60	0.70	0.78	0.85	0.90	▨
Value (dollars)	18 00	15 050	13 900	11 900	10 400	9600	8900	8500	8400

f What is the missing value? Write your answer correct to 2 decimal places. 1 mark

g In **Scatterplot 2**, *value* is plotted against log (*age*).

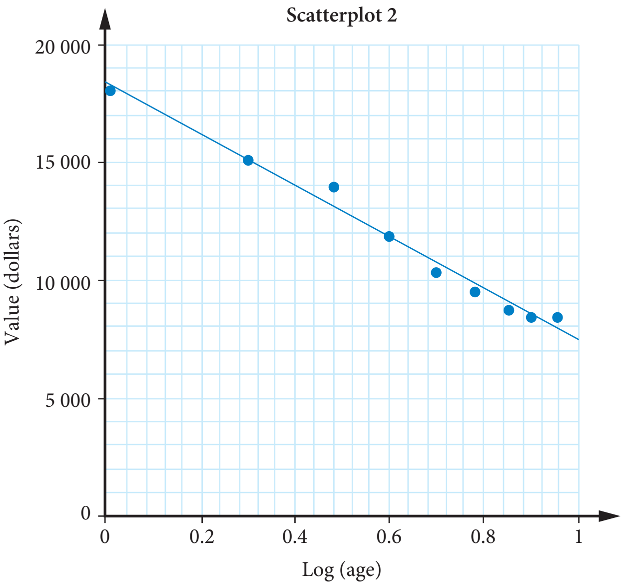
A least squares line of best fit fitted to the transformed data is also drawn.

Use the information in **Scatterplot 2** to describe the association between *value* and log (*age*) in terms of **direction**, **form** and **strength**. 3 marks

h The equation of this least squares line of best fit is

value = 18 300 – 10 800 × log (*age*)

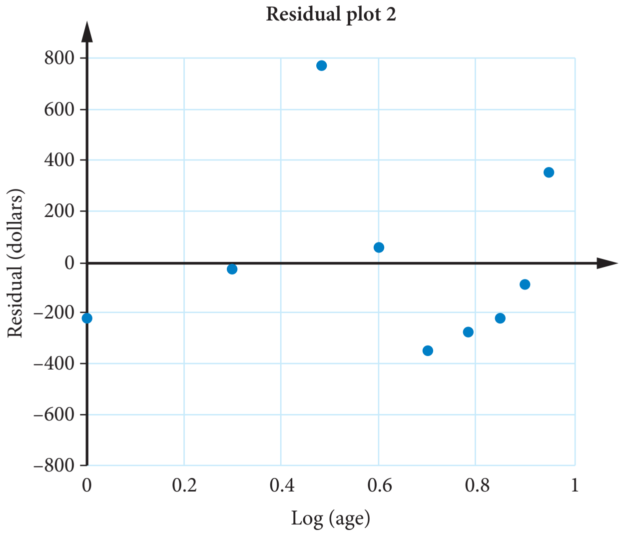
Use this equation to predict the value of a car that is three years old. Write your answer correct to the nearest hundred dollars. 1 mark



The residual plot for this association is shown in **Residual plot 2**.

i **Residual plot 2** suggests that the log (*age*) transformation has been successful in linearising the data. What feature of this residual plot shows that it has been successful? 1 mark

j A transformation applied to the *value* axis can also linearise the original data displayed in **Scatterplot 1**. Suggest a suitable transformation and **explain** why it will work. 1 mark

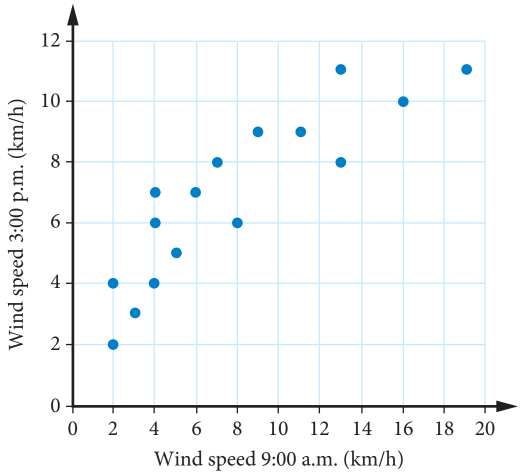


[VCAA 2005 2CQ1]

Question 8

The wind speeds (in km/h) that were recorded at the weather station at 9:00 a.m. and 3:00 p.m. respectively on 18 days in November are given in the table. A scatterplot has been constructed from this data set.

Wind speed (km/h)			
9:00 a.m.	3:00 p.m.	9:00 a.m.	3:00 p.m.
2	2	13	8
4	6	11	9
4	7	2	4
4	4	7	8
13	11	5	5
6	7	8	6
3	3	6	7
16	10	19	11
6	7	9	9



Let the wind speed at 9:00 a.m. be represented by the variable *ws9:00am* and the wind speed at 3:00 p.m. be represented by the variable *ws3:00pm*.

The association between *ws9:00am* and *ws3:00pm* shown in the scatterplot is non-linear.

A **squared transformation** can be applied to the variable *ws3:00pm* to linearise the data in the scatterplot.

a Apply the squared transformation to the variable *ws3:00pm* and determine the equation of the least squares line of best fit that allows $(ws3{:}00pm)^2$ to be predicted from *ws9:00am*.

In the boxes provided, write the coefficients for this equation correct to 1 decimal place.

$$(ws3{:}00pm)^2 = \boxed{} + \boxed{} \times ws9{:}00am$$ 2 marks

b Use this equation to predict the wind speed at 3:00 p.m. on a day when the wind speed at 9:00 a.m. is 24 km/h. Write your answer correct to the nearest whole number. 1 mark

[VCAA 2012 2CQ4]

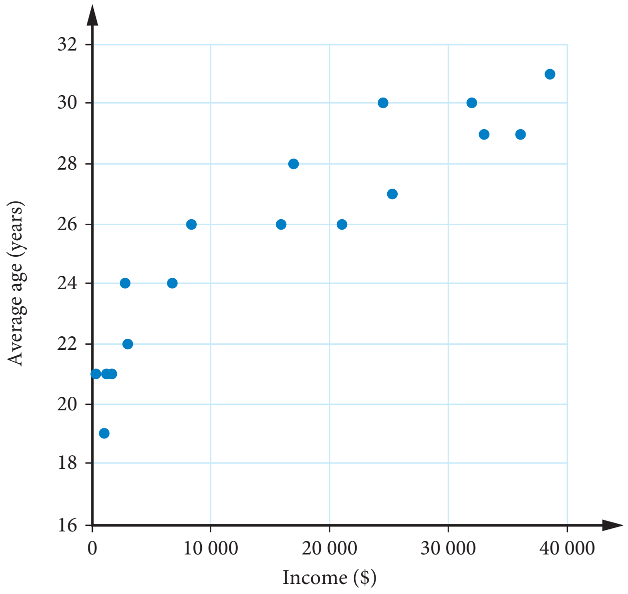
Question 9

The average age of women at first marriage in years (*average age*) and average yearly income in dollars per person (*income*) were recorded for a group of 17 countries.

The results are displayed in the table. A scatterplot of the data is also shown.

Average age (years)	Income ($)
21	1750
22	3200
26	8600
26	16 000
28	17 000
26	21 000
30	24 500
30	32 000
31	38 500

Average age (years)	Income ($)
29	33 000
27	25 500
29	36 000
19	1300
21	600
24	3050
24	6900
21	1400



The association between *average age* and *income* is non-linear.

A **log transformation** can be applied to the variable *income* and used to linearise the scatterplot.

a Apply this log transformation to the data and determine the equation of the least squares line of best fit that allows *average age* to be predicted from log (*income*).

Write the coefficients for this equation, correct to 2 decimal places, in the spaces provided.

average age = ☐ + ☐ × log (*income*) 2 marks

b Use this equation to predict the average age of women at first marriage in a country with an average yearly income of $20 000 per person.

Write your answer correct to 1 decimal place. 1 mark

[VCAA 2011 2CQ4]

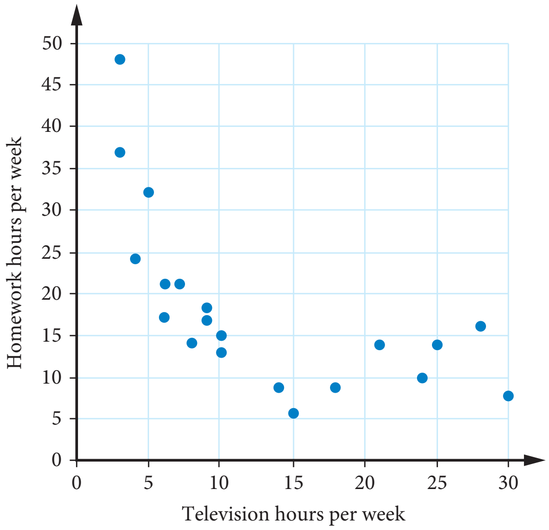
Question 10

The number of hours spent doing homework each week (*homework hours*) and the number of hours spent watching television each week (*television hours*) were recorded for a group of 20 Year 12 students.

The results are displayed in the table and a scatterplot constructed as shown.

Television hours	Homework hours
6	17
28	16
14	9
6	21
9	17
30	8
10	15
3	48
3	37
18	9

Television hours	Homework hours
7	21
9	18
4	24
25	14
8	14
24	10
21	14
5	32
15	6
10	13



The association between *homework hours* per week and *television hours* per week is clearly non-linear.

A **reciprocal** transformation applied to the variable, *television hours*, can be used to linearise the scatterplot.

a Apply this reciprocal transformation to the data and determine the equation of the least squares line of best fit that allows *homework hours* to be predicted from the reciprocal of *television hours*.

Write the coefficients correct to 2 decimal places. 2 marks

b If a student spends 12 hours per week watching television, use the least squares line of best fit to predict the number of hours that the student spends doing homework. Give your answer correct to 1 decimal place. 1 mark

[VCAA 2008 2CQ5]

SUMMARY

4

Linear associations

Practice quiz

Line of best fit

- A **line of best fit** is a straight line that is the best approximation for a set of data.
- The equation for the **least squares line of best fit** is $y = a + bx$ where
  - the **slope** or **gradient** of the line is $b = r\frac{s_y}{s_x}$
  - the ***y*-intercept** of the line is $a = \bar{y} - b\bar{x}$
  - r is the Pearson correlation coefficient
  - s_x and s_y are the sample standard deviations of x and y respectively
  - $\bar{x}$ and $\bar{y}$ are the sample means of x and y respectively
- The y-intercept is a. This means the ***y* variable** is a **units** when the ***x* variable** is zero **units**.
- The slope is b. This means the ***y* variable** increases/decreases by b **units** for every 1 **unit** increase in the ***x* variable**.

The coefficient of determination (r^2)

- can be calculated by squaring the Pearson correlation coefficient
- is a value between 0 and 1
- is a measure of how useful a line of best fit is as a linear model for a particular set of data (0 means it's a totally useless measure and 1 means it's a perfect measure).
- is usually converted to a percentage
- r^2 × **100%** of the variation in the **response variable** can be explained by the variation in the **explanatory variable**.

Making predictions

- We predict values by using the equation of the least squares line of best fit and substituting values into it and solving.
- Predicting *within* the original data range is called **interpolation**.
- Predicting *outside* the original data range is called **extrapolation**.
- Predictions based on extrapolation are not as reliable as those based on interpolation.

Residual analysis

- A **residual** is the vertical distance between each data point and the least squares line of best fit.
- residual value = data value – predicted value
    - The data value can be read from a scatterplot or a table.
    - The predicted value needs to be calculated from the least squares line of best fit.
- Data values that lie
    - above the least squares line of best fit will have a positive residual value
    - below the least squares line of best fit will have a negative residual value
    - on the least squares line of best fit will have a residual value of zero
- The further the data value is from the least squares line of best fit, the larger the residual value.

Residual plot

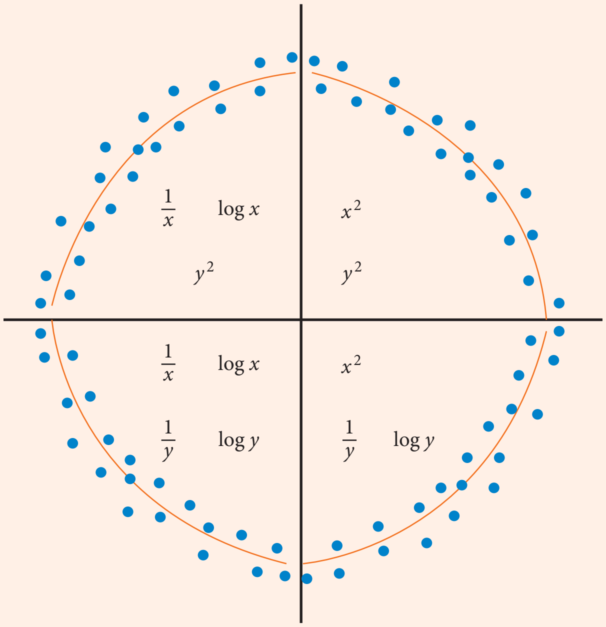
- A **residual plot** is like a scatterplot with the explanatory variable on the x-axis and the residual values on the y-axis.
- When residual values
    - are randomly scattered above and below the x-axis, the association between the original two variables, x and y, is probably *linear*.
    - show a hill or a valley shape, the association between the original two variables, x and y, is probably *non-linear*.

Data transformation

- **Data transformations** are used to linearise data that isn't linear.
- The three data transformations we use are:
  - **Squared transformation**
  - **Log transformation**
  - **Reciprocal transformation**

The transformation circle

- Calculate the coefficient of determination of each option to see which is the closest to 1.



CHAPTER 5

TIME SERIES

Prior learning

5.1 Time series plots

Alamy/Tim Slater/Stockimo

Constructing time series plots

Time series data represents an association between two variables where the explanatory variable is time. The aim of a time series is to help us investigate how a particular response variable changes over time. We have already looked at some of these associations involving time, but they need to be examined more closely because the data in time series, unlike the data in other associations, has a natural order.

A **time series plot** is a scatterplot where time is shown on the horizontal axis in regular intervals such as hours, weeks, months or years. The difference between a time series plot and a normal scatterplot is that the data points are joined and there is only one point for every value of time.

Using CAS Constructing time series plots

Create a time series plot using this table showing births by year in Texas from 1910 to 1930.

Year	Total live births
1910	41 104
1911	40 623
1912	39 200
1913	29 239
1914	41 300
1915	45 601
1916	45 122
1917	40 123
1918	47 563
1919	48 323
1920	48 124
1921	39 576
1922	45 623
1923	48 789
1924	49 792
1925	50 124
1926	50 678
1927	51 237
1928	53 456
1929	54 789
1930	53 627

TI-NSPIRE CAS

STEP 1

Open a New Document with a Lists & Spreadsheet page.

Enter the data as shown.

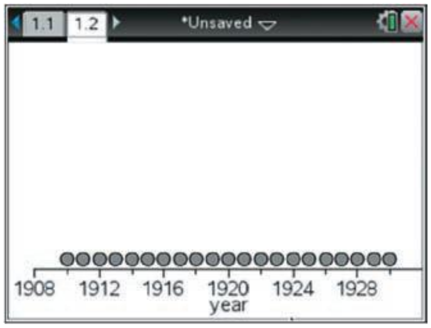
	A year	B births	C	D
1	1910	41104		
2	1911	40623		
3	1912	39200		
4	1913	29239		
5	1914	41300		
6	1915	45601		

STEP 2

Add a Data & statistics page.

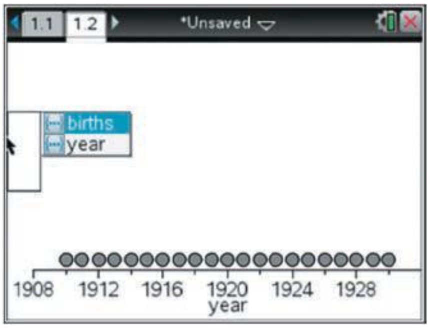
STEP 3

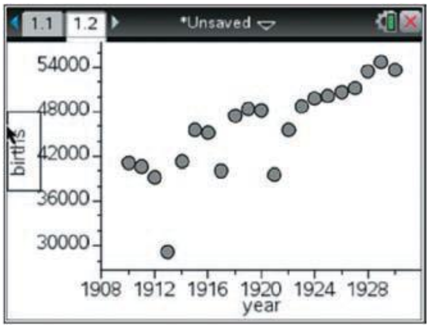
Move the cursor to the Click to add variable box at the bottom of the screen and select year.



STEP 4

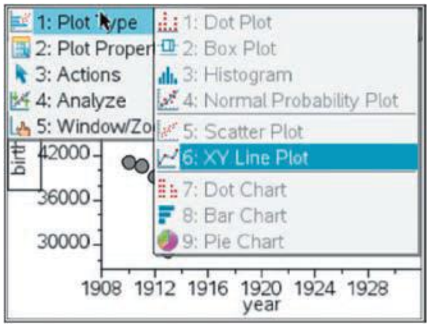
Move the cursor to the Click to add variable box at the left of the screen and select births.

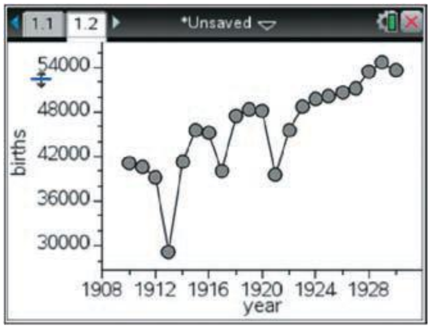




STEP 5

Press menu then 1: Plot Type then 6: XY Line Plot.



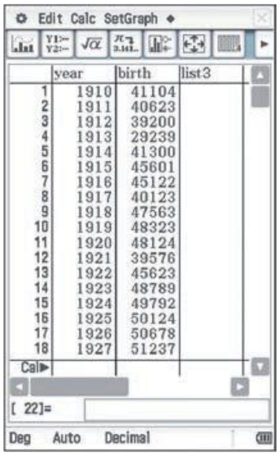


CLASSPAD

STEP 1

Use the **Statistics** application.

Enter the data as shown.



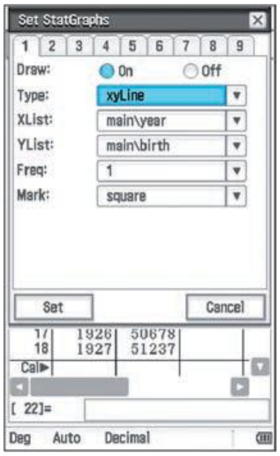
	year	birth	list3
1	1910	41104	
2	1911	40623	
3	1912	39200	
4	1913	29239	
5	1914	41300	
6	1915	45601	
7	1916	45122	
8	1917	40123	
9	1918	47563	
10	1919	48323	
11	1920	48124	
12	1921	39576	
13	1922	45623	
14	1923	48789	
15	1924	49792	
16	1925	50124	
17	1926	50678	
18	1927	51237	

[22]=

STEP 2

Tap **SetGraph**. Make sure that **StatGraph1** is checked, then tap **Setting…** and complete the screen as shown.

Tap Set.



Draw: ● On ○ Off

Type: xyLine

XList: main\year

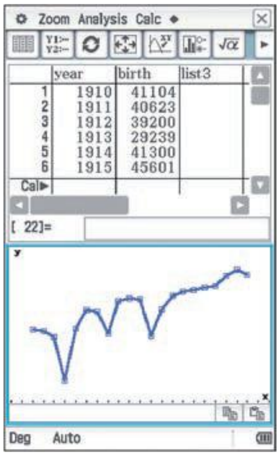
YList: main\birth

Freq: 1

Mark: square

STEP 3

Tap to draw the graph.



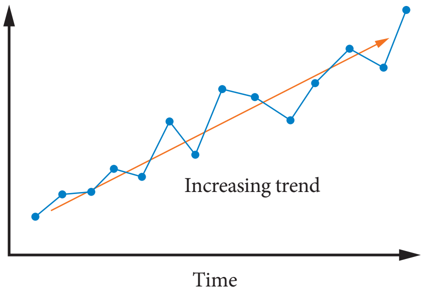
Tap Resize to view just the graph screen.

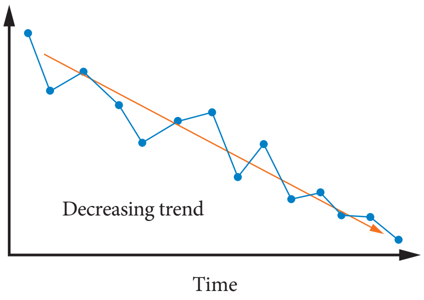
Features of time series plots

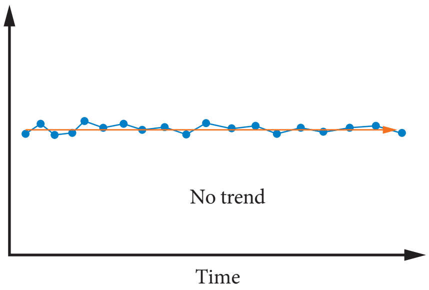
Time series plots can be described by identifying the following features.

Trend

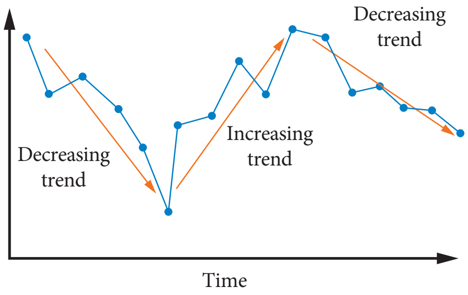
The **trend** is the long-term direction of the data. There will always be some fluctuation, but the data may be increasing or decreasing over the long term.





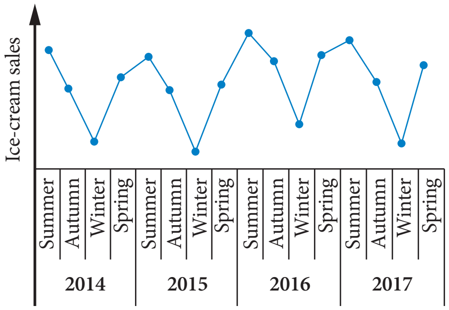


Sometimes certain subsections of the data can be identified as having a long-term trend as shown here:

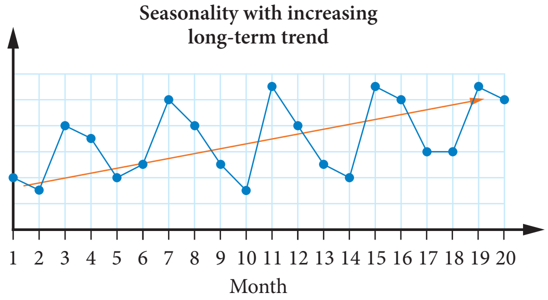


Seasonality

Seasonality describes data that has regular and predictable changes which repeat across a year or less, often matching the actual seasons. For example, ice-cream sales each year would regularly and predictably peak during the summer months and dip during the winter months:

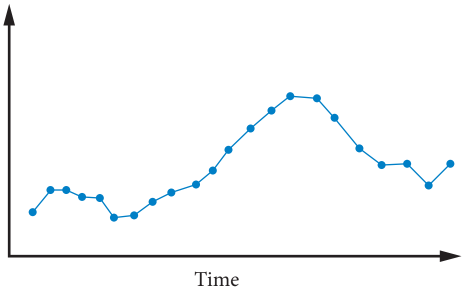


It's possible for data to show both seasonality and a trend at the same time. The following time series has regular peaks every four months as well as having an increasing long-term trend:



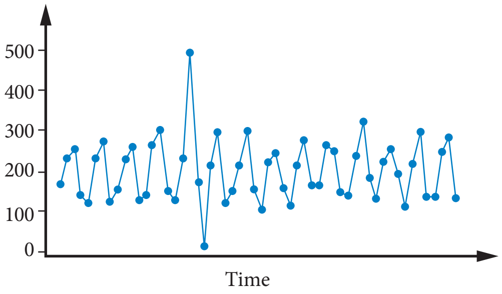
Irregular fluctuations

Sometimes data can appear to occur at random with **irregular fluctuations** that are short-term and have no pattern.



Outliers in time series

Sometimes potential outliers in a time series aren't difficult to identify, as shown in this time series plot.



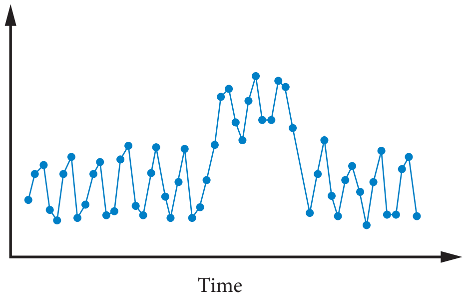
The difficulty lies in working out whether the extreme values are a valid data points and should be included in an analysis or whether they are contaminating the data and should be removed before an analysis. We shouldn't remove outliers without good reason. It's important to try to understand why the outliers have appeared and whether it is likely similar values will continue to occur.

A source of outliers is a one-off unanticipated event. For example, a freak storm over a weather station could distort weather data being monitored over time. Similarly, a large single-day drop in the stock market such as the 'flash crash' of 24 August 2015 could contaminate economic data.

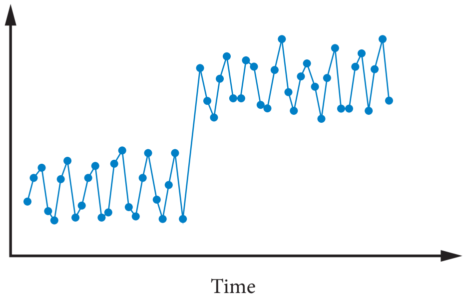
There are other sources of outliers, such as measuring instrument glitches and human error in recording data, which would justify removing them from the data. On the other hand, an outlier on an ECG heart monitor could be an important piece of information in a medical diagnosis.

Structural changes in time series

A **structural change** in a time series occurs when there is an unexpected shift in the data pattern. It is more significant than the appearance of a few outliers, as shown in this example:



Discontinuities are structural changes that are clear breaks in the time series. This plot shows structural change involving an obvious discontinuity:



Structural changes, and discontinuities in particular, are common in the real world and can lead to large forecasting errors and incorrect conclusions. As with outliers, it's important to work out the reasons.

Possible sources of structural change include significant extended events. For example, technological change such as the introduction of geostationary satellite imagery in the 1970s affected weather forecasting and climate data. Similarly a war or a long strike can cause a disruption to economic data over time. Discontinuities can be caused by a change of measuring equipment, new measurement techniques, or a poor experimental set-up.

EXAM PREP 5.1

Time series plots

Prep 1

USING CAS: CONSTRUCTING TIME SERIES PLOTS

An example of time series data in a table is displayed using births in the second column and time (in years) in the first column. Create a time series plot using the data listed.

Births by year in Minnesota from 1910 to 1920

Year	Total live births
1910	11 124
1911	12 000
1912	12 200
1913	13 329
1914	13 500
1915	14 005

Year	Total live births
1916	14 100
1917	14 127
1918	15 163
1919	15 323
1920	16 224

Prep 2

Are the following time series likely to show seasonality?

a the sale of umbrellas

b the sale of shorts

c the sale of underpants

d the sale of milk

e the occupancy rates at a holiday resort

f the money earned by teachers

g the money earned by fruit pickers

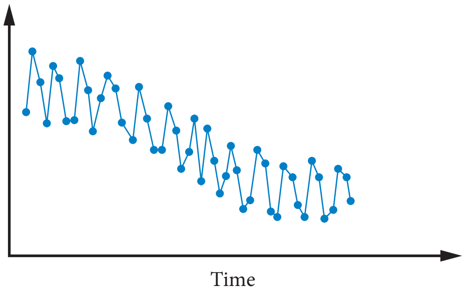
h the number of people attending AFL matches

Prep 3

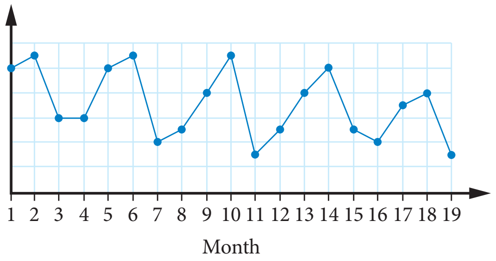
State which of the following features each of these time series plots are showing:

- increasing trend
- decreasing trend
- seasonality
- outliers
- structural change

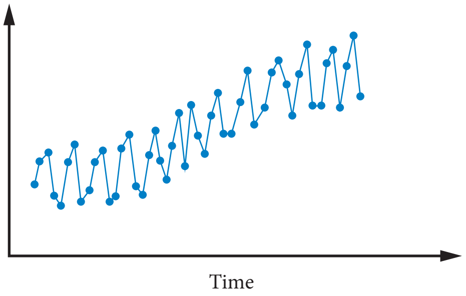
a



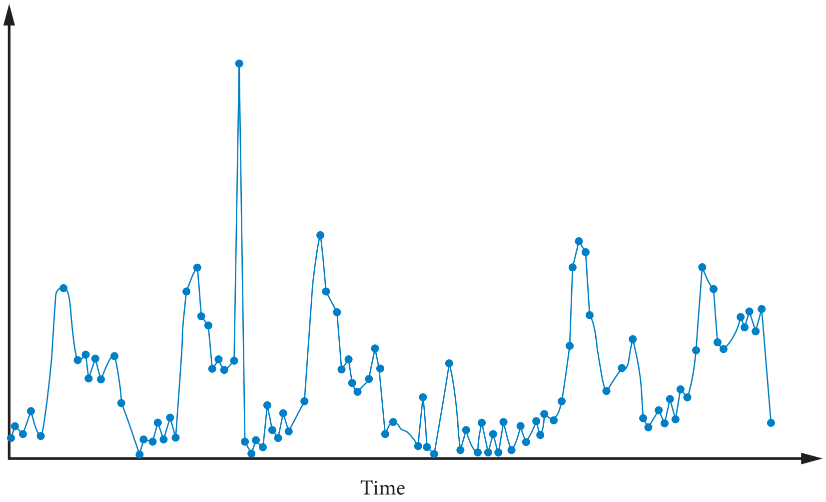
b



c



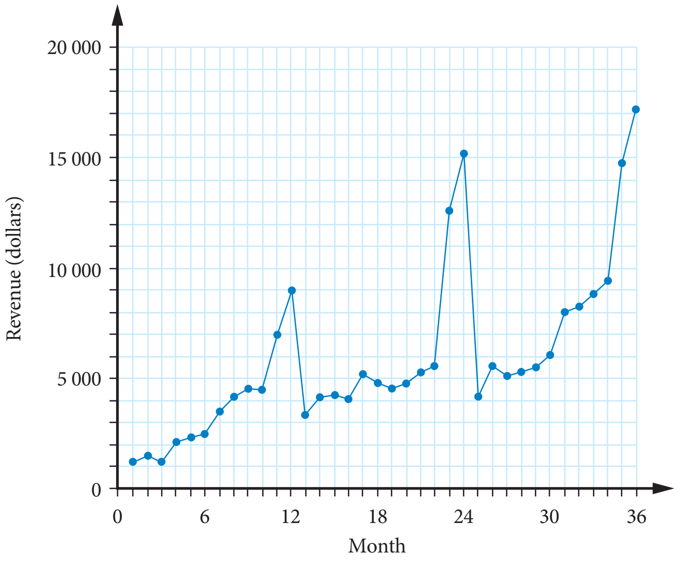
d



EXAM PRACTICE 5.1

Time series plots

Question 1



The time series plot shows the revenue from sales (in dollars) each month made by a Queensland souvenir shop over a three-year period.

This time series plot indicates that, over the three-year period, revenue from sales each month showed

A no overall trend.

B no correlation.

C positive skew.

D an increasing trend only.

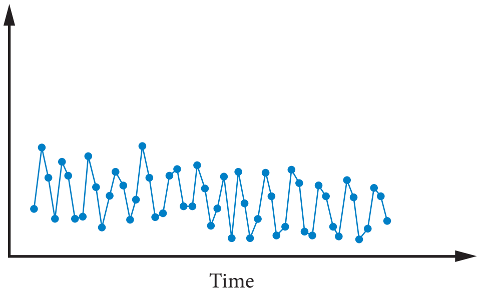
E an increasing trend with seasonal variation.

[VCAA 2007 1CQ11]

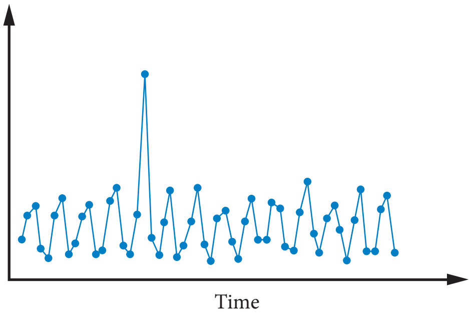
Question 2

Which one of the following time series plots shows a discontinuity?

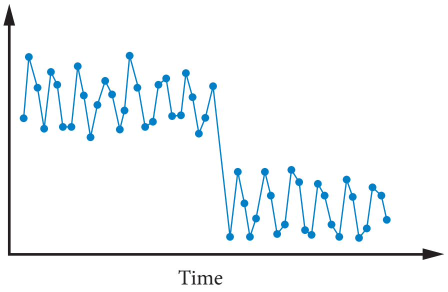
A



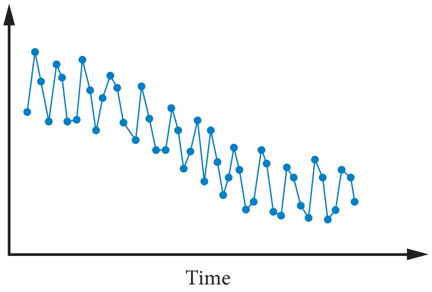
B



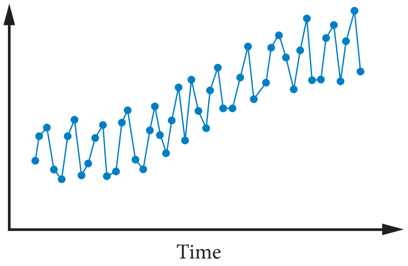
C



D



E



Question 3

A rain gauge collecting data of the amount of rain falling in a particular region was found to be affected by the wind. A wind shield was installed to ensure none of the rain was blown away and the readings were more accurate. As a result, the time series of this data could be affected by which of the following?

A outliers
B seasonality
C decreasing trend
D irregular fluctuations
E discontinuity

Question 4

The temperature of a room is measured at hourly intervals throughout the day. The most appropriate graph to show how the temperature changes from one hour to the next is a

A boxplot.
B stem plot.
C histogram.
D time series plot.
E two-way frequency table.

[VCAA 2012 1CQ5]

Question 5

The time series plot shows the share price of two companies over a period of time.

From the plot, it can be concluded that over the interval 1990–2000, the **difference** in share price between the two companies has shown

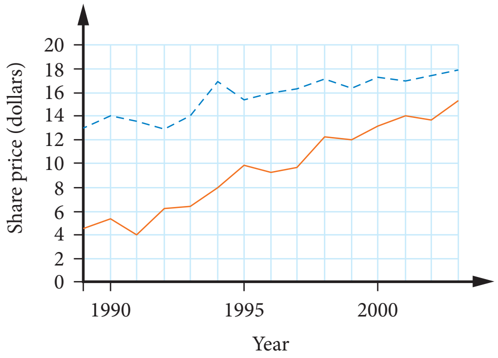
A a decreasing trend.

B an increasing trend.

C seasonal variation.

D a five-year cycle.

E no trend.



[VCAA 2004 1CQ12]

Question 6

The association between **resting pulse rate** (in beats per minute) and **fitness level** (below average, average, above average) is best displayed using

A a histogram.

B a scatterplot.

C a time series plot.

D parallel boxplots.

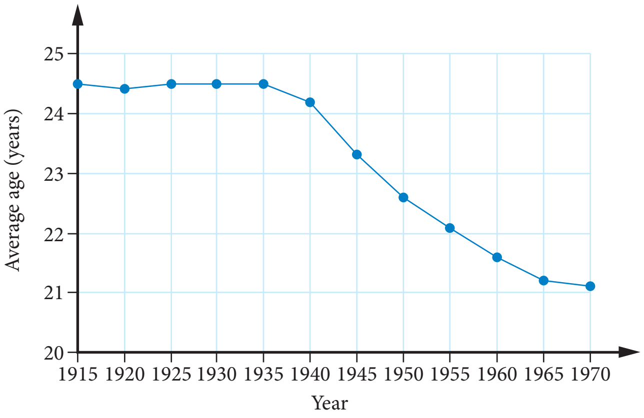
E back-to-back stem plots.

[VCAA 2003 1RCQ11]

Question 7

The following time series plot shows the average age of women at first marriage in a particular country during the period 1915 to 1970.

Use this plot to describe, in general terms, the way in which the average age of women at first marriage in this country has changed during the period 1915 to 1970.



1 mark

[VCAA 2011 2CQ3a]

Question 8

Over recent years, the salaries of Patagonian cricketers have increased rapidly. The following data gives the average salaries in dollars of a large group of these cricketers over the period 1991–2000.

Year	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Average salary	45 000	47 000	50 000	58 000	70 000	78 000	93 000	105 000	126 000	142 000

a This data will be used to predict future average salaries of Patagonian cricketers.

In this analysis, the **explanatory** variable is ☐. 1 mark

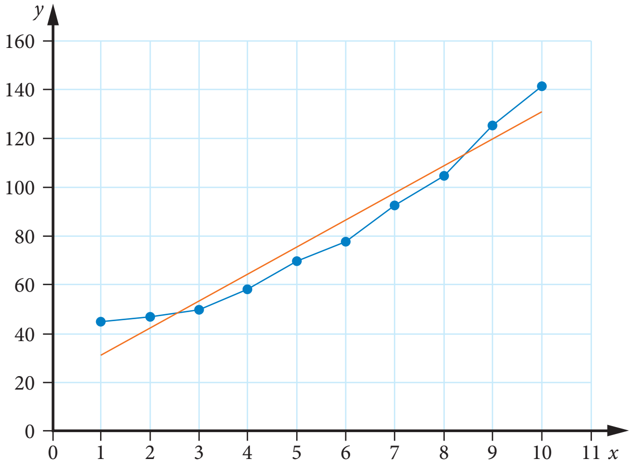
To begin the analysis, the years have been rescaled as $x = 1$ to $x = 10$ (1991 = 1, 1992 = 2, and so on), and average salary rescaled in thousands of dollars as the variable y.

Year	1	2	3	4	5	6	7	8	9	10
Average salary ($000s)	45	47	50	58	70	78	93	105	126	142

This rescaled data is displayed as a time series plot. Also displayed is the least squares line of best fit which has been determined for this rescaled data.

The equation of the least squares line of best fit is

$$y = 20.9 + 10.99x$$

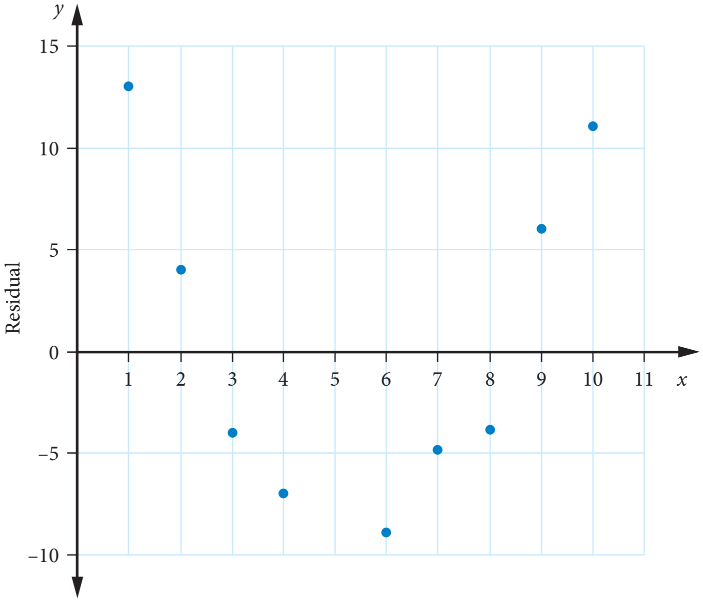


b Using the line of best fit equation $y = 20.9 + 10.99x$ to model the increase in these cricketers' average salaries, we would

i say that the average salary increase for the cricketers was $ ☐ per year.

ii predict that their average salary in the year 2005 will be about $ ☐. 4 marks

From the time series plot, the increase in Patagonian cricketers' salaries over time appears non-linear. This can be confirmed by constructing the corresponding residual plot as shown.



c Copy and complete the plot by

i calculating the value of the residual for 1995.

ii plotting this residual as a point on the graph. 2 marks

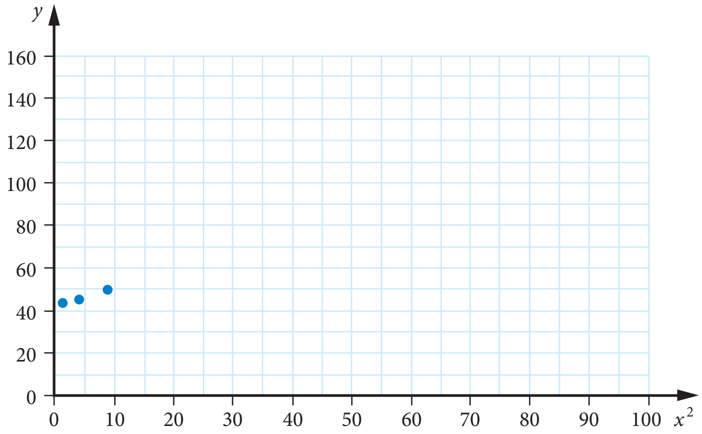
The time series, along with the residual plot, shows that the growth of salaries with time is non-linear. Inspecting the time series, it would appear that it would be appropriate to use an x^2 transformation to transform the data to linearity.

The original data has been reproduced in the table below and an extra row has been added for the transformed variable, x^2.

Year (x)	1	2	3	4	5	6	7	8	9	10
Year^2 (x^2)										
Average salary (y) ($000s)	45	47	50	58	70	78	93	105	126	142

d i Complete the table. 1 mark

ii Copy and complete the time series plot below of the transformed data to show that the x^2 transformation has produced a more nearly linear plot. (Note the first three points have been plotted.) 2 marks

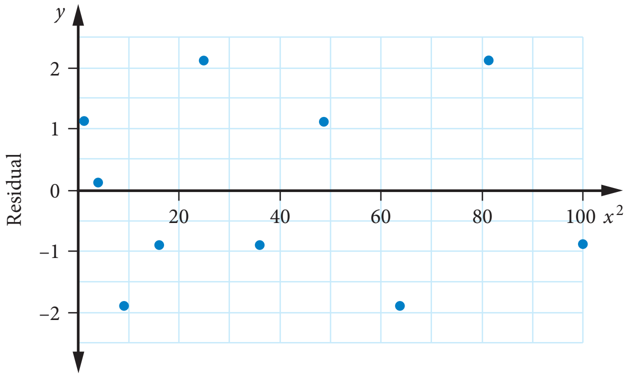


iii Find the equation of the least squares line of best fit for the transformed data. Write the coefficients, correct to 2 decimal places, in the spaces provided.

$y = \boxed{} + \boxed{}\, x^2$ 2 marks

iv Use this line of best fit equation to predict the average salary of this group of Patagonian cricketers in 2005. 2 marks

The residual plot that results from fitting a least squares line of best fit to this transformed data is shown.



v This plot suggests that the x^2 transformation has been successful in linearising the time series plot. What feature of this residual plot supports this conclusion? 1 mark

[VCAA 2002 2C]

Numerical smoothing

5.2

Shutterstock.com/Paul Cowan

Smoothing time series data is a technique for levelling out fluctuations to produce a smoother graph that lets us see the time series trends more clearly. One way of achieving this is through **numerical smoothing** techniques; that is, methods that involve numbers and calculations.

Smoothing using moving means

WS
Moving means

Moving means is the most common numerical smoothing technique. It involves finding a series of means of a fixed number of data points. The simplest method is to smooth using an odd number of data points, such as three, five or seven.

Worked example 1

The following table shows the number of students in a Year 12 Further Mathematics class over the last 10 years.

Year	1	2	3	4	5	6	7	8	9	10
Number of students	25	18	23	21	19	20	18	16	17	15

a Complete a table using the method of three-point moving means.

b Graph the original data and the smoothed data on the same set of axes.

c What does the graph of the smoothed data suggest about the trend in the original data?

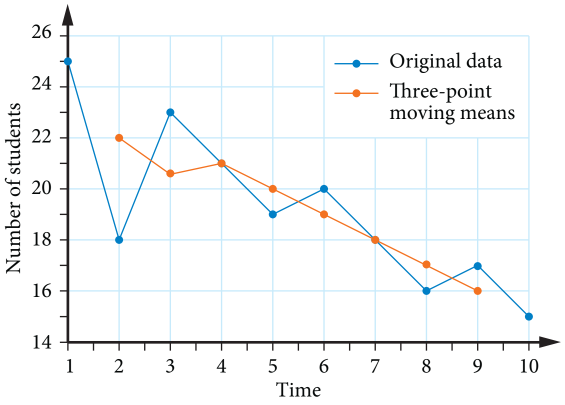
Working

a Set up a table with 3 columns to list Year in list 1, Number of students in list 2, and calculations for three-point moving means in list 3.

Year	Number of students	Three-point moving means
1	25	
2	18	$\frac{25+18+23}{3}=22$
3	23	$\frac{18+23+21}{3}=20.67$
4	21	$\frac{23+21+19}{3}=21$
5	19	$\frac{21+19+20}{3}=20$
6	20	$\frac{19+20+18}{3}=19$
7	18	$\frac{20+18+16}{3}=18$
8	16	$\frac{18+16+17}{3}=17$
9	17	$\frac{16+17+15}{3}=16$
10	15	

b Graph the original data and the smoothed data on the same set of axes.

Note: a CAS/calculator can be used by constructing two time series plots, one for the original data and one for the smoothed data.



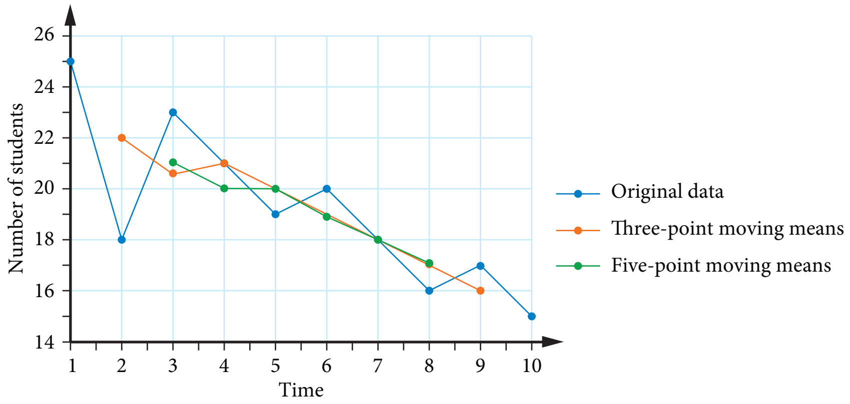
c Write the answer.

The graph of the smoothed data suggests a decreasing trend.

In **three-point moving means smoothing**, the first and last points disappear because the means for these can't be calculated. In **five-point moving means smoothing**, the first two and the last two points disappear. Using the data in the previous worked example five-point moving means smoothing gives us:

Year	Number of students	Five-point moving means
1	25	
2	18	
3	23	$\frac{25+18+23+21+19}{5}=21.2$
4	21	$\frac{18+23+21+19+20}{5}=20.2$
5	19	$\frac{23+21+19+20+18}{5}=20.2$
6	20	$\frac{21+19+20+18+16}{5}=18.8$
7	18	$\frac{19+20+18+16+17}{5}=18$
8	16	$\frac{20+18+16+17+15}{5}=17.2$
9	17	
10	15	

The result of the smoothing using five-point moving means can be seen in the graph. The larger the number of points used for the moving means, the greater the smoothing effect.



Smoothing with an even number of points

Smoothing with an even number of points is more complicated than smoothing with an odd number of points because the centre of an even number of points isn't at one of the original time values. We deal with this using a process called **centring** where we add an extra stage that involves taking two-point moving means of the smoothed values.

Worked example 2

a Use the four-point moving means method with centring to smooth the data given in the table below on the number of births per month over a calendar year in a small country hospital.

b What is the smoothed number of births for April?

c Graph the smoothed data.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
Number of births	10	12	6	5	22	18	13	7	9	10	8	15

Working

a Set up a table with 4 columns and extra rows between time intervals for centring of four-point moving means.

Month	Number of births	Four-point moving means	Four-point moving means with centring
Jan	10		
Feb	12		
		$\frac{10+12+6+5}{4}=8.25$	
Mar	6		$\frac{8.25+11.25}{2}=9.75$
		$\frac{12+6+5+22}{4}=11.25$	
Apr	5		$\frac{11.25+12.75}{2}=12$
		$\frac{6+5+22+18}{4}=12.75$	
May	22		$\frac{12.75+14.5}{2}=13.625$
		$\frac{5+22+18+13}{4}=14.5$	
June	18		$\frac{14.5+15}{2}=14.75$
		$\frac{22+18+13+7}{4}=15$	
July	13		$\frac{15+11.75}{2}=13.375$

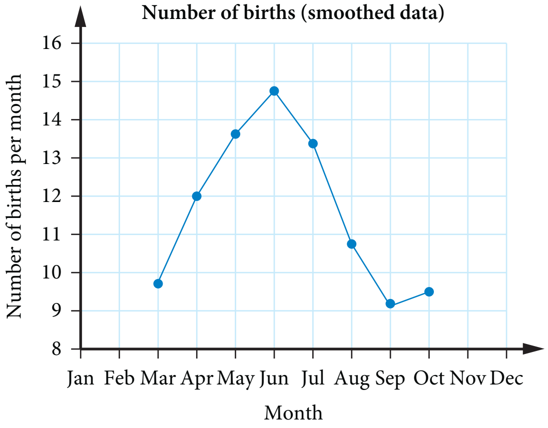
Month	Number of births	Four-point moving means	Four-point moving means with centring
		$\frac{18+13+7+9}{4}=11.75$	
Aug	7		$\frac{11.75+9.75}{2}=10.75$
		$\frac{13+7+9+10}{4}=9.75$	
Sept	9		$\frac{9.75+8.5}{2}=9.125$
		$\frac{7+9+10+8}{4}=8.5$	
Oct	10		$\frac{8.5+10.5}{2}=9.5$
		$\frac{9+10+8+15}{4}=10.5$	
Nov	8		
Dec	15		

b Read the value from the table.

The smoothed number of births for April is 12.

c Graph the smoothed data.

Note this can be graphed as a time series plot using a CAS/calculator.



Choosing the number of points for moving means

How high can you go?

There are a number of factors to take into account when choosing how many points to use for moving means smoothing. The larger the number, the greater the smoothing effect and the clearer the overall trend, but the more data points that are lost. With three-point moving means we lose two data points (one at the start and one at the end); with four-point and five-point moving means we lose four data points, with six-point and seven-point moving means we lose six data points etc. If we go too high, we could end up losing nearly all of our data points.

Matching the moving means to natural cycles

The number of moving means points is usually chosen by looking at the natural cycle of the data being considered.

	No. of points for moving means
Daily sales figures for stores open Monday to Sunday	7
Daily sales figures for stores open only Monday to Friday	5
Monthly figures	12
Quarterly accounts	4

Unless there is an even-numbered natural cycle involved, it's usually better to use an odd number of points.

EXAM PREP 5.2

Numerical smoothing

Prep 1 WORKED EXAMPLE 1 USING CAS: CONSTRUCTING TIME SERIES PLOTS

The introduction of speed cameras in Victoria reduced the number of people killed in road accidents over the period 1990 to 2000. This data is tabulated below.

Year	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Number of road deaths	225	240	201	192	185	160	172	127	132	101	100

a Complete a table using the method of three-point moving means.

b Graph the original data and the smoothed data on the same set of axes.

c What does the graph of the smoothed data suggest about the trend in the original data?

Prep 2 WORKED EXAMPLE 2 USING CAS: CONSTRUCTING TIME SERIES PLOTS

a Use the four-point moving means method with centring to smooth the data given in the table below on the number of sales of textbooks from 2000 to 2012.

b What is the smoothed number of sales for 2009?

c Graph the smoothed data.

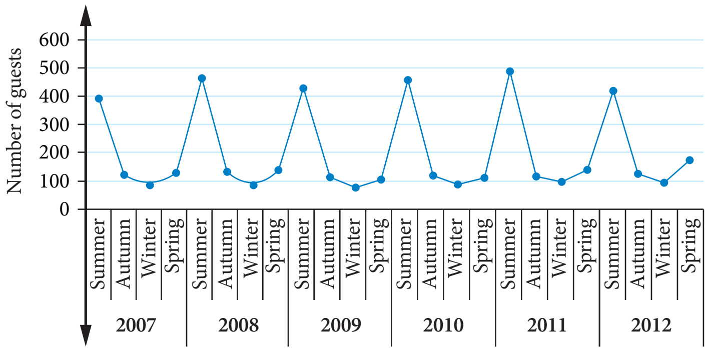
Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Sales of textbooks	2250	2230	2000	2010	1990	3000	2045	2989	3000	1950	2120	2255	2297

EXAM PRACTICE 5.2

Numerical smoothing

Use the following information to answer Questions 1 & 2.

The time series plot displays the number of guests staying at a holiday resort during summer, autumn, winter and spring for the years 2007 to 2012 inclusive.



Question 1

Which one of the following best describes the pattern in the time series?

A random variation only

B decreasing trend with seasonality

C seasonality only

D increasing trend only

E increasing trend with seasonality

[VCAA 2013 1CQ12]

Question 2

The table shows the data from the time series plot for the years 2007 and 2008.

Using four-mean smoothing with centring, the smoothed number of guests for winter 2007 is closest to

A 85

B 107

C 183

D 192

E 200

Year	Season	Number of guests
2007	Summer	390
	Autumn	126
	Winter	85
	Spring	130
2008	Summer	460
	Autumn	136
	Winter	86
	Spring	142

[VCAA 2013 1CQ13]

Question 3

The table below displays the total monthly rainfall (in mm) in a reservoir catchment area over a one-year period.

Month	Jan	Feb	Mar	April	May	June	July	Aug	Sept	Oct	Nov	Dec
Rainfall	9	65	35	99	75	90	133	196	106	56	76	76

Using three-**mean** moving average smoothing, the smoothed value for the total rainfall in April is closest to

A 65 **B** 66 **C** 70 **D** 75 **E** 88

[VCAA 2006 1CQ10]

Question 4

The table below shows the number of broadband users in Australia for each of the years from 2004 to 2008.

Year	2004	2005	2006	2007	2008
Number	1 012 000	2 016 000	3 900 000	4 830 000	5 140 000

Based on data obtained from: www.data.worldbank.org

A two-point moving mean, with centring, is used to smooth the time series. The smoothed value for the number of broadband users in Australia in 2006 is

A 2 958 000 **B** 3 379 600 **C** 3 455 500 **D** 3 661 500 **E** 3 900 000

[VCAA 2011 1CQ13]

Question 5

The data gives the number of accidents recorded at a city intersection each year from 1993 to 2002.

Using a **four**-point moving average (mean) with centring, the smoothed value of the number of accidents in 1995 is

A 7.25 **B** 7.375 **C** 7.5

D 7.625 **E** 8

Year	Number of accidents
1993	13
1994	7
1995	3
1996	9
1997	10
1998	8
1999	7
2000	6
2001	10
2002	11

[VCAA 2003 1CQ12]

Graphical smoothing

5.3

iStock.com/Marccophoto

Smoothing using moving medians

WS

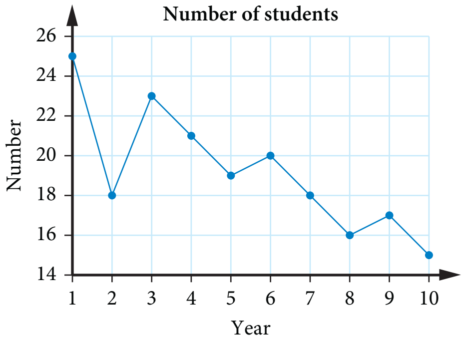
Moving medians

Graphical smoothing is an alternative to numerical smoothing and is based on working directly from the time series plot. The most common graphical smoothing method is **moving medians**, which involves finding a series of medians of a fixed number of data points. Although it's possible to use an even number of data points with this method, we will only be using an odd number.

Since medians, unlike means, are not influenced by outliers, this method has the advantage that extreme data values will be eliminated quickly.

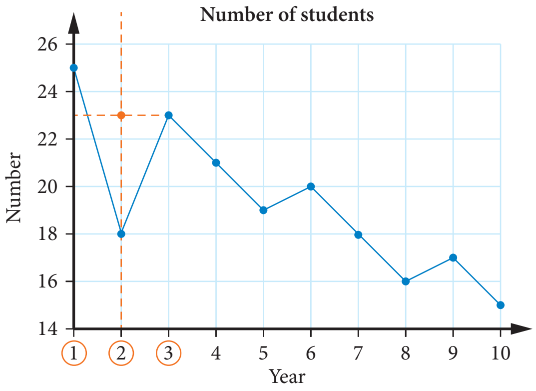
Worked example 3

For this time series plot showing the number of students in a Year 12 Further Mathematics class over the last 10 years, smooth the data using the three-point moving median method.

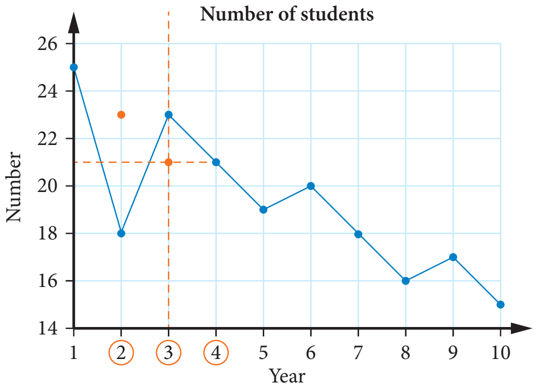


Working

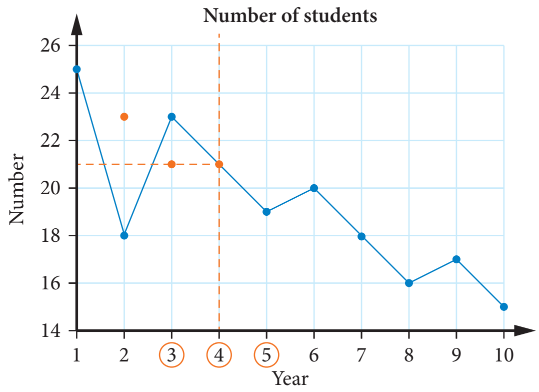
1 Look at the first three time points, find where the median time and the median number intersect, and place that point.



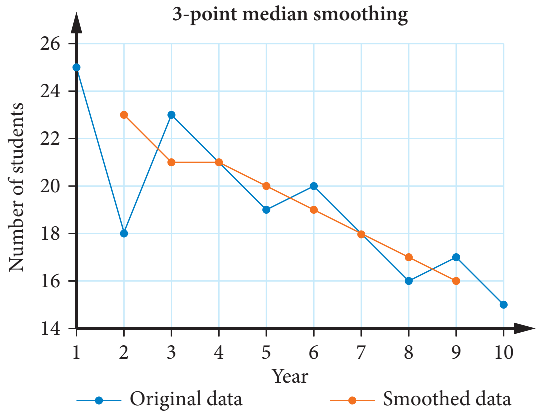
2 Look at the next three time points and repeat.



3 Repeat for the next three points.



4 Repeat the process until you reach the last median and then join the points.



As with three-point moving mean smoothing, two data points are lost with three-point moving median smoothing. Similarly, four data points are lost with five-point moving median smoothing.

Exam hack

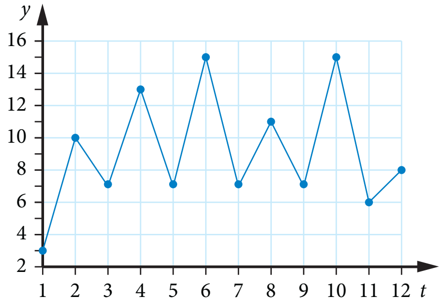
In an exam, watch out when you find a median point in median smoothing which is the same as one of original points. Double check your working. You may have made a mistake. The important step is to work out the middle value viewed from the vertical axis.

EXAM PREP 5.3

Graphical smoothing

Prep 1 WORKED EXAMPLE 3

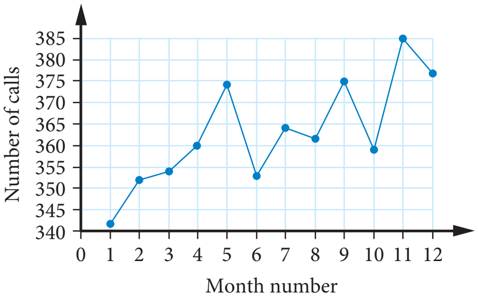
For this time series plot, smooth the data using the three-point moving median method.



EXAM PRACTICE 5.3

Graphical smoothing

Question 1



The time series plot shows the number of calls each month to a call centre over a twelve-month period.

The plot is to be smoothed using five-point **median** smoothing. The smoothed number of calls for month number 10 is closest to

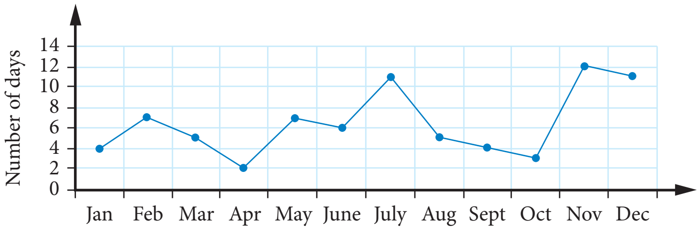
A 358 **B** 364 **C** 371

D 375 **E** 377

[VCAA 2010 1CQ12]

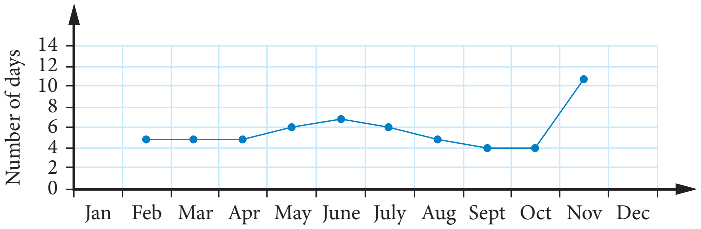
Question 2

The time series plot below shows the number of days that it rained in a town each month during 2011.

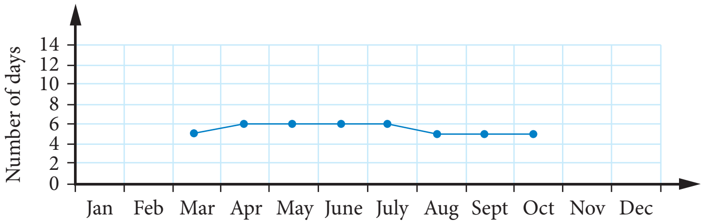


Using five-median smoothing, the smoothed time series plot will look most like

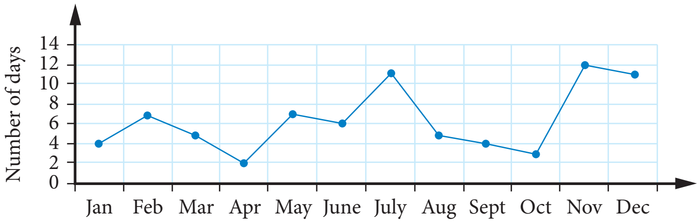
A



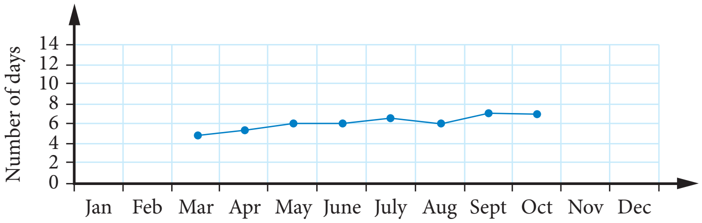
B



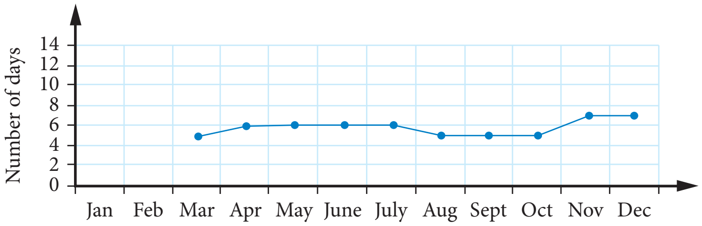
C



D



E



[VCAA 2012 1CQ9]

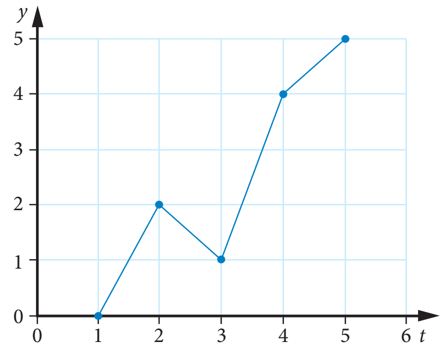
Question 3

The five points shown on the grid have been taken from a time series plot that is to be smoothed using median smoothing.

The coordinates of the median of these five points are

A (3, 1) **B** (3, 2) **C** (3, 2.4)

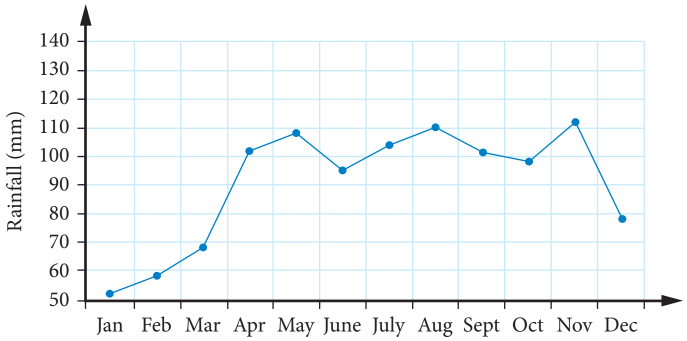
D (3, 2.5) **E** (3, 3)



[VCAA 2005 1CQ10]

Question 4

The time series plot shows the rainfall (in mm) for each month during 2008.



a Which month had the highest rainfall? 1 mark

b Copy the plot and use three-median smoothing to smooth the time series. Plot the smoothed time series on the original plot. Mark each smoothed data point with a cross (×). 2 marks

c Describe the general pattern in rainfall that is revealed by the smoothed time series plot. 1 mark

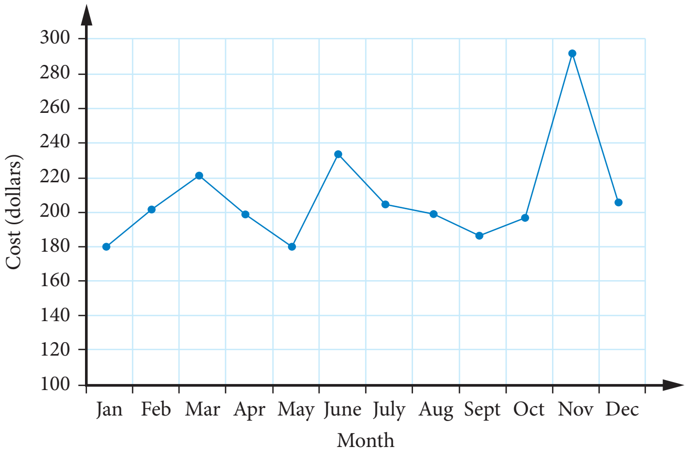
[VCAA 2009 2CQ2]

Question 5

A supervisor believes that there is an increasing trend in the total monthly telephone call costs in the company. She gathers data using the previous year's figures as shown in the table.

Month	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
Cost ($)	180	201	221	198	180	233	205	199	185	197	291	206

To check this, the supervisor constructs the time series plot shown, using the previous year's data.



a Copy the graph and use three-median smoothing to smooth the time series shown. Plot the smoothed time series on the same graph. 2 marks

b Does the smoothed time series support the supervisor's belief that there has been an increasing trend in call costs in the company during the previous year? Briefly explain your answer. 1 mark

c Name a feature (or features) of this particular time series which suggests that it is more appropriate to use **three-median** smoothing rather than **three-mean** moving average smoothing. 1 mark

[VCAA 2003 2CQ2]

Seasonal adjustment

5.4

Interpreting seasonal indices

WS

Seasonal adjustment

As we’ve already seen, some data has regular and predictable changes that repeat across a year or less which we describe as seasonality. When dealing with time series a season is usually a day, a month or a quarter.

It is not effective to smooth seasonal data using the moving means or moving median methods because smoothing two consecutive points might connect two data points with significant seasonal differences. To deal effectively with this sort of data, we need to make **seasonal adjustments**. **Seasonal indices** are used to make seasonal adjustments.

The following example shows the seasonal indices for the number of pizzas sold on each day of the week by a pizza chain:

Mon	Tue	Wed	Thu	Fri	Sat	Sun
0.3	1.05	0.4	0.85	1.4	1.8	1.2

The seasonal indices in the table add up to 7. This is because the season we are looking at is the days of the week. If the seasonal indices were for months, they would add up to 12, and if they were for quarters, they would add up to 4.

The mean of the indices in the table is 1. This means that Tuesdays, Fridays, Saturdays and Sundays have above the daily mean pizza sales, while Mondays, Wednesdays and Thursdays have below the daily mean pizza sales.

Converting the seasonal indices to percentages by multiplying by 100 makes them easier to interpret:

Mon	Tue	Wed	Thu	Fri	Sat	Sun
30%	105%	40%	85%	140%	180%	120%

We can see the following from the percentage table:

- On Tuesdays pizza sales are 5% *above* the daily mean.
- On Wednesdays pizza sales are 60% *below* the daily mean.
- The seasonal index percentages add up to 700%.

The sum of the seasonal indices = the number of seasons

Type of data	No. of seasons	Cycle	Sum of seasonal indices
Daily figures for data from Monday to Sunday	7	A week	7
Daily figures for data from Monday to Friday	5	A week	5
Monthly figures	12	A year	12
Quarterly accounts	4	A year	4

To interpret seasonal indices convert them to percentages.

Calculating and using seasonal indices

The most common seasonal adjustment we make is to **de-seasonalise** data. To do this we use seasonal indices to remove the seasonal component of the time series. De-seasonalisation is a form of smoothing, which takes out the seasonal effects of the data so that a line of best fit can be fitted and long-term trends can be predicted.

To de-seasonalise time series data use the formula:

$$\text{De-seasonalised value} = \frac{\text{Actual value}}{\text{Seasonal index}}$$

To find the actual value given the de-seasonalised value and seasonal index we can rewrite this formula as:

$$\text{Actual value} = \text{De-seasonalised value} \times \text{Seasonal index}$$

Amana Images/Walter Bibikow/JAI

Worked example 4

The quarterly sales figures for the number of cars sold were recorded by a car sales yard for 2015 to 2017.

Year	Sales figures quarter 1	Sales figures quarter 2	Sales figures quarter 3	Sales figures quarter 4
2015	5	7	9	3
2016	4	8	9	4
2017	5	9	10	5

a Calculate the seasonal index for each quarter correct to 4 decimal places.

b Use the seasonal indices to de-seasonalise the data correct to 2 decimal places.

c Plot the original and the de-seasonalised data on the same set of axes.

Working

a **1** Calculate the yearly mean. Calculate the totals for each year, and find the mean for each year by dividing each total by 4.

Year	Q1	Q2	Q3	Q4	Yearly mean
2015	5	7	9	3	$\frac{5+7+9+3}{4}=6$
2016	4	8	9	4	$\frac{4+8+9+4}{4}=6.25$
2017	5	9	10	5	$\frac{5+9+10+5}{4}=7.25$

2 Calculate the quarterly proportions. Divide each quarterly sales figure by the corresponding yearly mean to obtain quarterly proportions. Give answers correct to 4 decimal places.

Year	Q1	Q2	Q3	Q4
2015	$\frac{5}{6}=0.8333$	$\frac{7}{6}=1.1667$	$\frac{9}{6}=1.5000$	$\frac{3}{6}=0.5000$
2016	$\frac{4}{6.25}=0.6400$	$\frac{8}{6.25}=1.2800$	$\frac{9}{6.25}=1.4400$	$\frac{4}{6.25}=0.6400$
2017	$\frac{5}{7.25}=0.6897$	$\frac{9}{7.25}=1.2414$	$\frac{10}{7.25}=1.3793$	$\frac{5}{7.25}=0.6897$

3 Calculate the seasonal indices by finding the mean of the quarterly proportions. Give answers correct to 4 decimal places.

Year	Q1	Q2	Q3	Q4
2015	0.8333	1.1667	1.5000	0.5000
2016	0.6400	1.2800	1.4400	0.6400
2017	0.6897	1.2414	1.3793	0.6897
Total	2.1630	3.6881	4.3193	1.8297
Seasonal index	$\frac{2.1630}{3}$ $=0.7210$	$\frac{3.6881}{3}$ $=1.2294$	$\frac{4.3193}{3}$ $=1.4398$	$\frac{1.8297}{3}$ $=0.6099$

Due to rounding errors the seasonal indices may not add exactly to 4 as expected.

4 Write a summary table for the seasonal indices.

Year	Q1	Q2	Q3	Q4
Seasonal index	0.7210	1.2294	1.4398	0.6099

b **1** De-seasonalise the original time series data using the formula

De-seasonalised value = $\frac{\text{Actual value}}{\text{Seasonal index}}$

Year	Q1	Q2	Q3	Q4
2015	$\frac{5}{0.7210}=6.94$	$\frac{7}{1.2294}=5.69$	$\frac{9}{1.4398}=6.25$	$\frac{3}{0.6099}=4.92$
2016	$\frac{4}{0.7210}=5.55$	$\frac{8}{1.2294}=6.51$	$\frac{9}{1.4398}=6.25$	$\frac{4}{0.6099}=6.56$
2017	$\frac{5}{0.7210}=6.94$	$\frac{9}{1.2294}=7.32$	$\frac{10}{1.4398}=6.95$	$\frac{5}{0.6099}=8.20$

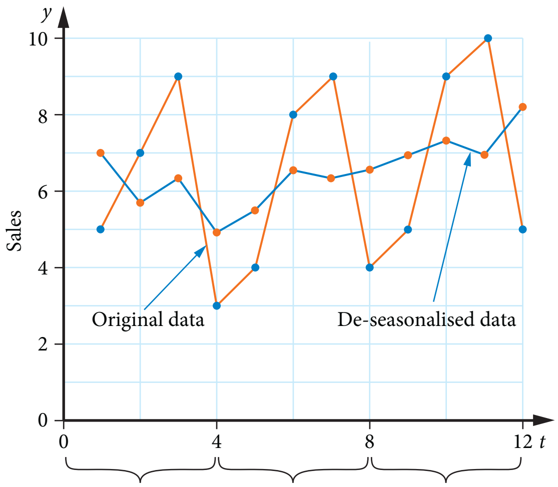
2 Write a summary table for the de-seasonalised values, correct to 2 decimal places.

Year	Q1	Q2	Q3	Q4
2015	6.94	5.69	6.25	4.92
2016	5.55	6.51	6.25	6.56
2017	6.94	7.32	6.95	8.20

9780170371100

c Plot the original and the de-seasonalised time series data on the same set of axes.

Use $t = 1$ to represent the first quarter of 2015, $t = 2$ to represent the second quarter of 2015 etc.



Worked example 5

The following table shows the quarterly seasonal indices for revenue to a publishing company from the sales of maths textbooks.

Quarter	1	2	3	4
Seasonal index	0.7		0.6	1.9

a What is the missing quarter 2 seasonal index?

b To correct for seasonality, by what percentage should the sales for quarter 2 be increased?

c The company predicts that its de-seasonalised quarterly sales will be \$1 000 000 for each quarter.

Based on this, what would you predict the actual sales for quarter 2 to be?

Working

a Use the fact that the seasonal indices need to add to 4 for quarterly data.

$4 - 0.7 - 0.6 - 1.9 = 0.8$

The quarter 2 seasonal index is 0.8.

b Use the formula

$$\text{De-seasonalised value} = \frac{\text{Actual value}}{\text{Seasonal index}}$$

$$\begin{aligned}\text{De-seasonalised value} &= \frac{\text{Actual value}}{\text{Seasonal index}}\\ &= \frac{1}{0.8} \times \text{Actual value}\\ &= 1.25 \times \text{Actual value}\\ &= 125\% \text{ of Actual value}\end{aligned}$$

So to correct for seasonality, the sales for quarter 2 should be increased by 25%.

c Use the formula

$$\text{Actual value} = \text{De-seasonalised value} \times \text{Seasonal index}$$

$$\text{Actual value} = \text{De-seasonalised value} \times \text{Seasonal index}$$

$$\begin{aligned}\text{So Actual sales for quarter 2} &= \$1\,000\,000 \times 0.8\\ &= \$800\,000\end{aligned}$$

EXAM PREP 5.4

Seasonal adjustment

Prep 1

WORKED EXAMPLE 4

The quarterly sales figures for the number of king-sized beds sold were recorded by a furniture shop for 2016 to 2018.

Year	Sales figures quarter 1	Sales figures quarter 2	Sales figures quarter 3	Sales figures quarter 4
2016	11	8	1	3
2017	9	9	3	1
2018	4	9	1	7

a Calculate the seasonal index for each quarter correct to 3 decimal places.

b Use the seasonal indices to de-seasonalise the data.

c Plot the original and the de-seasonalised data on the same set of axes.

Prep 2

WORKED EXAMPLE 5

The following table shows the quarterly seasonal indices for revenue to a company from the sales of a brand of soft drink.

Quarter	1	2	3	4
Seasonal index	1.4	0.8		1.1

a What is the missing quarter 3 seasonal index?

b To correct for seasonality, by what percentage should the sales for quarter 3 be increased?

c The company predicts that its de-seasonalised quarterly sales will be $100 000 for each quarter. Based on this, what would you predict the actual sales for quarter 3 to be?

EXAM PRACTICE 5.4

Seasonal adjustment

Question 1

The quarterly seasonal indices for sales in a shop are shown in the table.

Quarter	1	2	3	4
Seasonal index	1.3	0.9	0.7	1.1

A seasonal index of 1.1 for quarter 4 means that sales in quarter 4 are typically

A 10% above the yearly average.

B 10% below the yearly average.

C 11% below the yearly average.

D 90% above the yearly average.

E 90% below the yearly average.

[VCAA 2005 1CQ11]

Question 2

A garden supplies outlet sells water tanks. The monthly seasonal indices for the revenue from the sale of water tanks are given below. The seasonal index for September is missing.

Seasonal index											
Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
1.26	0.96	0.74	0.48	0.31	0.39	0.75	1.55		1.32	1.25	1.58

The revenue from the sale of water tanks in September 2009 was $104 500.
The de-seasonalised revenue for September 2009 is closest to

A $42 800 **B** $74 100 **C** $104 500 **D** $141 000 **E** $147 300

[VCAA 2010 1CQ13]

Use the following information to answer Questions 3 & 4.

The table below shows the long-term average rainfall (in mm) for summer, autumn, winter and spring. Also shown are the seasonal indices for summer and autumn. The seasonal indices for winter and spring are missing.

	Season			
	Summer	Autumn	Winter	Spring
Long-term average rainfall (mm)	52.0	54.5	48.8	61.3
Seasonal index	0.96	1.01		

Question 3

The seasonal index for spring is closest to

A 0.90 **B** 1.03 **C** 1.13 **D** 1.15 **E** 1.17

[VCAA 2012 1CQ11]

Question 4

In 2011, the rainfall in autumn was 48.9 mm.

The de-seasonalised rainfall (in mm) for autumn is closest to

A 48.4 **B** 48.9 **C** 49.4 **D** 50.9 **E** 54.0

[VCAA 2012 1CQ12]

Question 5

The quarterly seasonal indices for mineral water sales (in litres) of a mineral water supplier are shown in the table below.

	Quarter 1	Quarter 2	Quarter 3	Quarter 4
Seasonal index	1.28	1.02	0.74	0.96

When **de-seasonalised** the amount of mineral water sold in Quarter 1 is 28 098 litres. To the nearest litre, the **actual** amount of mineral water sold in Quarter 1 was

A 7 025 litres. **B** 21 952 litres. **C** 28 098 litres.

D 35 965 litres. **E** 112 392 litres.

[VCAA 2004 1CQ13]

Question 6

The seasonal indices for the first three quarters of a year are shown in the table below.

	Quarter 1	Quarter 2	Quarter 3	Quarter 4
Seasonal index	1.05	0.84	0.92	

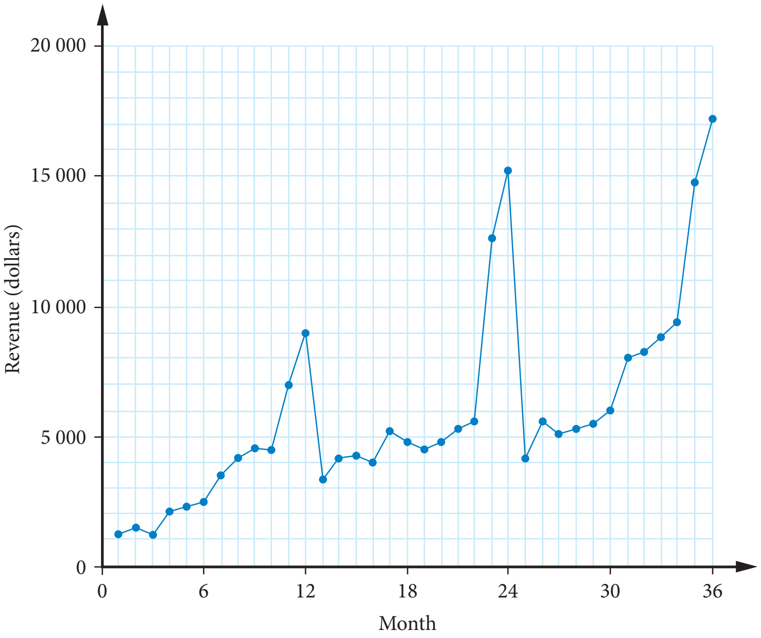
The seasonal index for Quarter 4 is

A 0.88 **B** 0.94 **C** 1.00 **D** 1.08 **E** 1.19

[VCAA 2003 1CQ13]

Question 7

The time series plot shows the revenue from sales (in dollars) each month made by a Queensland souvenir shop over a three-year period. The revenue from sales (in dollars) each month for the first year of the three-year period is shown in the table.



Month	Revenue ($)
January	1236
February	1567
March	1240
April	2178
May	2308
June	2512
July	3510
August	4234
September	4597
October	4478
November	7034
December	8978

If this information is used to determine the seasonal index for each month, the seasonal index for **September** will be closest to

A 0.80 **B** 0.82 **C** 1.16 **D** 1.22 **E** 1.26

[VCAA 2007 1CQ13]

Question 8

The seasonal index for headache tablet sales in summer is 0.80.

To correct for seasonality, the headache tablet sales figures for summer should be

A reduced by 80% **B** reduced by 25% **C** reduced by 20%

D increased by 20% **E** increased by 25%

[VCAA 2011 1CQ12]

Question 9

a This table shows the seasonal indices for rainfall in summer, autumn and winter. Calculate the seasonal index for spring. 1 mark

Seasonal indices			
Summer	Autumn	Winter	Spring
0.78	1.05	1.07	

b In 2008, a total of 188 mm of rain fell during summer.

Using the appropriate seasonal index in the table, determine the de-seasonalised value for the summer rainfall in 2008. Write your answer correct to the nearest millimetre. 1 mark

c What does a seasonal index of 1.05 tell us about the rainfall in autumn? 1 mark

[VCAA 2009 2CQ4]

5.5 Least squares trend lines

WS
Least squares regression line

iStock.com/vkp-australia

As with other data associations we have looked at, we can use a least squares line of best fit (often called a **trend line** for time series) to model time series trends, as long as the data appears to be linear. If it isn't linear, then we would need to use transformation techniques to linearise the data first.

Re-seasonalisation and forecasting

If there is seasonality in the time series, then we need to go through the extra step of de-seasonalising the data first before fitting the least squares line. The least squares line based on the de-seasonalised data can be used to make predictions; however, the result will give a de-seasonalised value. This value then needs to be **re-seasonalised** to give the actual value using:

Actual value = De-seasonalised value × Seasonal index

When we use the least squares line to make predictions outside the data range, the same issues of extrapolation that we've discussed previously apply. In the case of time series, this involves extending into the future, which is called **forecasting**. We can never be certain that the equation of the line will apply in the future, and the further into the future we are trying to predict, the less reliable the equation of the least squares line will be.

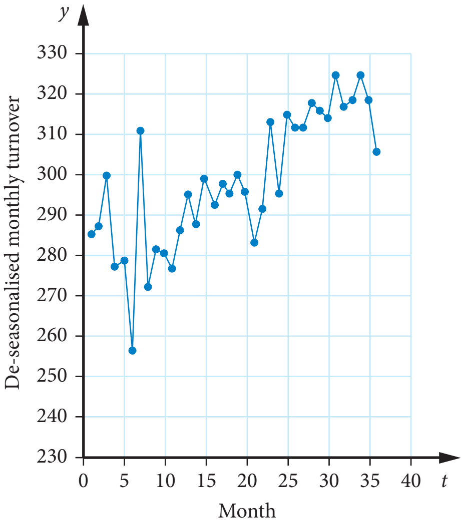
9780170371100

Worked example 6

The following de-seasonalised data represents the monthly turnover, in $millions, in a department store over a period of 3 years.

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2016	285.4	286.9	299.3	277.2	278.6	256.7	311.3	272.4	281.7	280.4	277.1	287.1
2017	294.8	288.3	298.9	292.4	297.3	295.5	300.2	295.5	283.4	291.6	312.9	295.7
2018	314.8	311.8	311.8	317.7	315.6	313.9	324.5	316.7	318.6	324.4	318.4	305.6
Seasonal index	0.841	0.704	0.874	0.935	0.966	0.999	0.869	0.834	0.873	0.978	1.164	1.964

The de-seasonalised data is graphed on the axes. On the horizontal axis, $t = 1$ represents Jan 2016, $t = 2$ represents Feb 2016, and so on.



a Determine the equation of the least squares line of best fit for the de-seasonalised time series data. Round the slope and intercept to 3 significant figures.

b Plot the least squares line of best fit on the time series plot.

c Use this line to estimate the turnover for August of 2018, in millions of dollars. Answer correct to 1 decimal place.

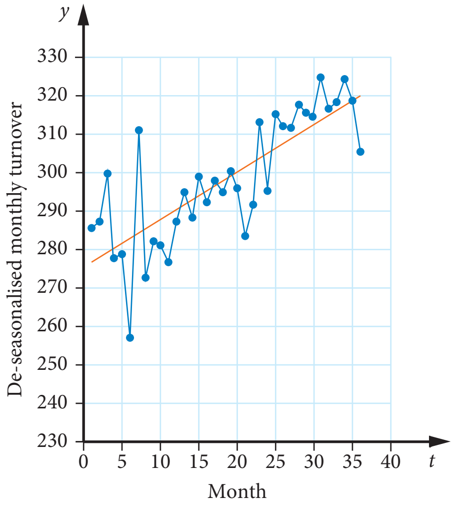
d Use the line to forecast the turnover for May of 2019, in millions of dollars. Answer correct to 1 decimal place.

Working

a Use a CAS/calculator to find the equation of the least squares line of best fit with Month (t) as the explanatory variable and De-seasonalised monthly turnover (y) as the response variable. Write the equation with the slope and intercept rounded to 3 significant figures.

$y = 1.24t + 275$

b Use a CAS/calculator to draw the least squares line of best fit.



c Determine the month number corresponding to August 2018 and substitute the value of t into the least squares line of best fit equation.

Re-seasonalise to convert to the actual value using:

Actual value = De-seasonalised value × Seasonal index

Give answer correct to 1 decimal place.

August of 2018 is $t = 32$

$$\begin{aligned} y &= 1.24t + 275 \\ &= 1.24 \times 32 + 275 \\ &= 314.68 \end{aligned}$$

The estimated de-seasonalised turnover for August 2018 is $314.68 million.

Seasonal index for August is 0.834.

Multiply $314.68 million by 0.834.

$314.68 \times 0.834 = \$262.443\ 12$ million

Estimated turnover for August 2018 is $262.4 million.

d Determine the month number corresponding to May of 2019. Substitute $t = 53$ into the least squares line of best fit equation.

Re-seasonalise to convert to the actual value using:

Actual value = De-seasonalised value × Seasonal index

Give answer correct to 1 decimal place.

May of 2019 is $t = 53$

$$\begin{aligned} y &= 1.24t + 275 \\ &= 1.24 \times 53 + 275 \\ &= 340.72 \end{aligned}$$

The de-seasonalised turnover for May 2019 is forecast to be $340.72 million.

Seasonal index for May is 0.966.

Multiply $340.72 million by 0.966.

$340.72 \times 0.966 = \$329.135\ 52$ million.

Estimated turnover for May 2019 is $329.1 million.

EXAM PREP 5.5

Least squares trend lines

Prep 1 WORKED EXAMPLE 6

The following de-seasonalised data represents the monthly turnover, in dollars, of a small roadside van over a period of 2 years.

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2016	185.4	286.9	199.3	177.2	178.6	256.7	211.3	172.4	181.7	180.4	177.1	287.1
2017	194.8	288.3	198.9	192.4	197.3	295.5	200.2	195.5	183.4	191.6	212.9	195.7
Seasonal index	1.810	0.704	0.774	0.735	0.866	0.899	0.769	1.134	1.073	0.979	1.228	1.064

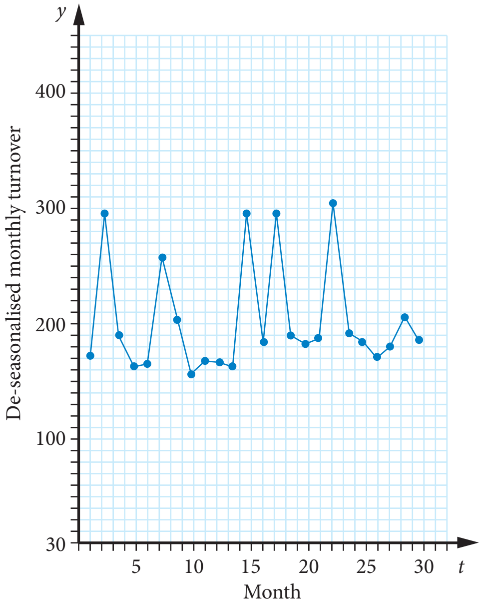
The de-seasonalised data is graphed on the axes. On the horizontal axis, $t = 1$ represents Jan 2016, $t = 2$ represents Feb 2016, and so on.

a Determine the equation of the least squares line of best fit for the de-seasonalised time series data. Round the slope and intercept to 3 significant figures.

b Plot the least squares line of best fit on the time series plot.

c Use this line to estimate the turnover for June of 2017 correct to the nearest ten cents.

d Use the line to forecast the turnover for May of 2018 correct to the nearest ten cents.



EXAM PRACTICE 5.5

Least squares trend lines

Use the following information to answer Questions 1 & 2.

The month-by-month price of a share listed on the Australian Stock Exchange is shown in the time series plot for a 36-month period. Also shown is a least squares line of best fit that has been fitted to the data.

The equation of the least squares line of best fit is

$$\text{share price} = 1.24 + 0.06 \times \text{month}$$

Question 1

The least squares line of best fit predicts that the price of the share after 48 months will be

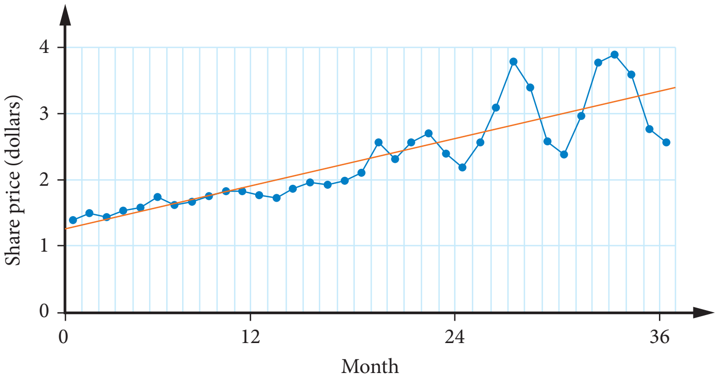
A \$1.96

B \$4.12

C \$5.04

D \$28.80

E \$62.40



[VCAA 2005 1CQ12]

Question 2

Which one of the following statements **best** describes the time series plot for the period shown?

A The share price shows no trend and no change in variability.

B The share price shows no trend and increases in variability.

C The share price shows an increasing linear trend with constant variability.

D The share price shows an increasing linear trend with decreasing variability.

E The share price shows an increasing linear trend with increasing variability.

[VCAA 2005 1CQ13]

Use the following information to answer Questions 3–5.

Month	Jan	Feb	Mar	April	May	June	July	Aug	Sept	Oct	Nov	Dec
Seasonal index	1.30	1.21	1.00	0.95	0.95	0.86	0.86	0.89	0.94		0.99	1.07

The table shows the seasonal indices for the monthly unemployment numbers for workers in a regional town.

Question 3

The seasonal index for October is missing from the table. The value of the missing seasonal index for October is

A 0.93 **B** 0.95 **C** 0.96 **D** 0.98 **E** 1.03

[VCAA 2006 1CQ11]

Question 4

The actual number of unemployed in the regional town in September is 330. The **de-seasonalised** number of unemployed in September is closest to

A 310 **B** 344 **C** 351 **D** 371 **E** 640

[VCAA 2006 1CQ12]

Question 5

A trend line that can be used to forecast the **de-seasonalised** number of unemployed workers in the regional town for the first nine months of the year is given by

de-seasonalised number of unemployed = 373.3 – 3.38 × month number

where month 1 is January, month 2 is February, and so on.

The **actual** number of unemployed for June is predicted to be closest to

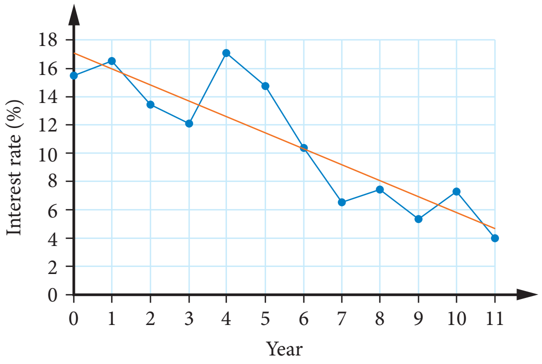
A 304 **B** 353 **C** 376 **D** 393 **E** 410

[VCAA 2006 1CQ13]

Question 6

The time series shown charts the change in interest rates over a period of years.

A trend line has been fitted to the data as shown. The equation of the trend line is closest to



A rate = –1.1 × year.

B rate = 15.6 – 1.5 × year.

C rate = 17 + 1.1 × year.

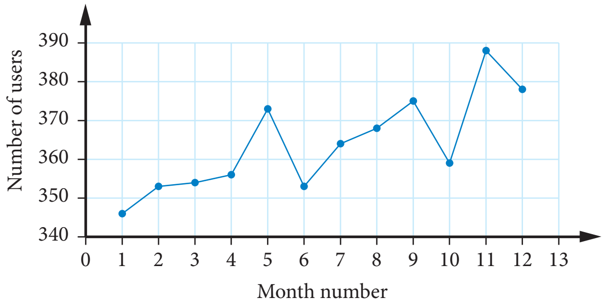
D rate = –1.1 + 17 × year.

E rate = 17 – 1.1 × year.

[VCAA 002 1CQ13]

Use the following information to answer Questions 7–9.

The time series plot shows the number of users each month of an online help service over a twelve-month period.



Question 7

The time series plot has

A no trend.

B no variability.

C seasonality only.

D an increasing trend with seasonality.

E an increasing trend only.

[VCAA 2008 1CQ11]

Question 8

The data values used to construct the time series plot are given below.

Month number	1	2	3	4	5	6	7	8	9	10	11	12
Number of users	346	353	354	356	373	353	364	368	375	359	388	378

A four-point moving mean with centring is used to smooth the time series. The smoothed value of the number of users in month number 5 is closest to

A 357 **B** 359 **C** 360 **D** 365 **E** 373

[VCAA 2008 1CQ12]

Question 9

A least squares line of best fit is fitted to the time series plot.

The equation of this least squares line of best fit is

number of users = 346 + 2.77 × *month number*

Let month number 1 = January 2007, month number 2 = February 2007, and so on.

Using the above information, the line of best fit predicts that the number of users in December 2009 will be closest to

A 379 **B** 412 **C** 443 **D** 446 **E** 448

[VCAA 2008 1CQ13]

Question 10

A trend line was fitted to a de-seasonalised set of quarterly sales data for 2012.

The seasonal indices for quarters 1, 2 and 3 are given in the table. The seasonal index for quarter 4 is not shown.

Quarter number	1	2	3	4
Seasonal index	1.2	0.7	0.8	

The equation of the trend line is

de-seasonalised sales = 256 000 + 15 600 × *quarter number*

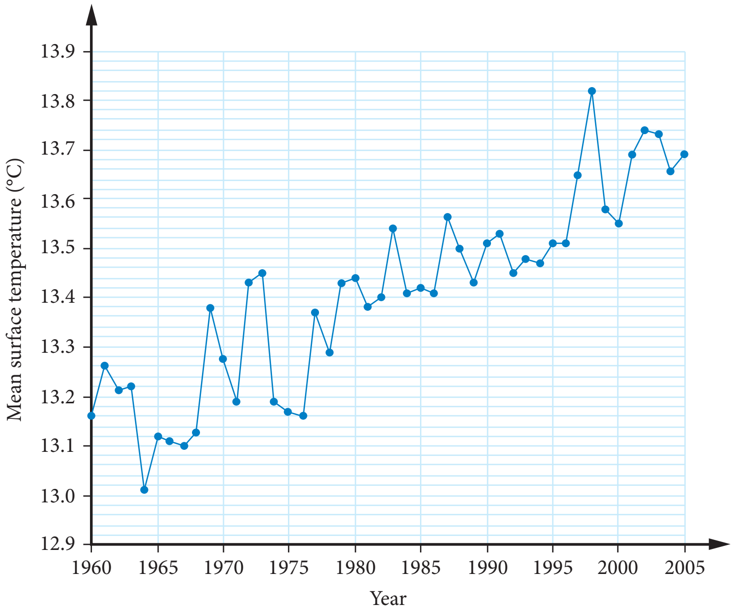
Using this trend line, the **actual** sales for quarter 4 in 2012 are predicted to be closest to

A \$222 880 **B** \$244 923 **C** \$318 400 **D** \$382 080 **E** \$413 920

[VCAA 2012 1CQ13]

Question 11

The mean surface temperature (in °C) of Australia for the period 1960 to 2005 is displayed in the time series plot.



a In what year was the lowest mean surface temperature recorded? 1 mark

The least squares method is used to fit a trend line to the time series plot.

b The equation of this trend line is found to be

mean surface temperature = −12.361 + 0.013 × *year*

i Use the trend line to predict the mean surface temperature (in °C) for 2010. Write your answer correct to 2 decimal places. 1 mark

The actual mean surface temperature in the year 2000 was 13.55°C.

ii Determine the residual value (in °C) when the trend line is used to predict the mean surface temperature for the year 2000. Write your answer correct to 2 decimal places. 1 mark

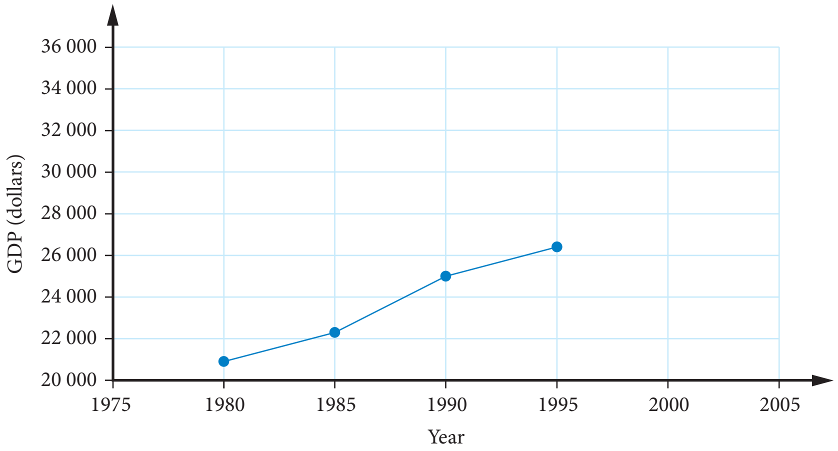
iii By how many degrees does the trend line predict Australia's mean surface temperature will rise each year? Write your answer correct to 3 decimal places. 1 mark

[VCAA 2007 2CQ2]

Question 12

The table shows the Australian gross domestic product (GDP) per person, in dollars, at five-yearly intervals for the period 1980 to 2005.

Year	1980	1985	1990	1995	2000	2005
GDP	20 900	22 300	25 000	26 400	30 900	33 800



a Copy and complete the time series plot above by plotting the GDP for the years 2000 and 2005. 1 mark

b Briefly describe the general trend in the data. 1 mark

In the table below, the variable *year* has been rescaled using 1980 = 0, 1985 = 5 and so on. The new variable is *time*.

Year	1980	1985	1990	1995	2000	2005
Time	0	5	10	15	20	25
GDP	20 900	22 300	25 000	26 400	30 900	33 800

c Use the variables *time* and *GDP* to write down the equation of the least squares line of best fit that can be used to predict *GDP* from *time*. Take *time* as the explanatory variable. 2 marks

d In the year 2007, the *GDP* was $34 900. Find the error in the prediction if the least squares line of best fit calculated in **part c** is used to predict *GDP* in 2007. 2 marks

[VCAA 2010 2CQ3]

Time series

Practice quiz

SUMMARY

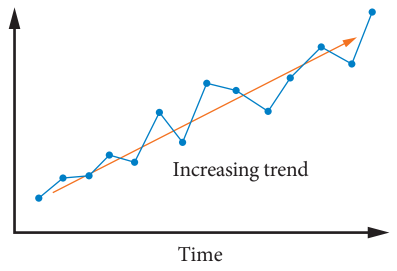
5

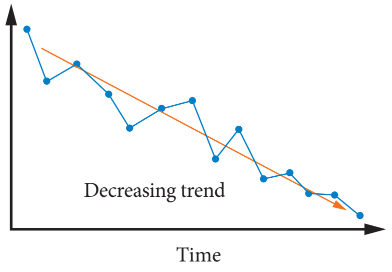
Time series plots

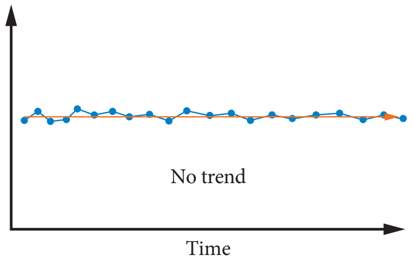
- **Time series data** represents an association between two variables where the explanatory variable is time.
- A **time series plot** is a scatterplot where time is shown on the horizontal axis, the data points are joined and there is only one point for every value of time.

Time series features

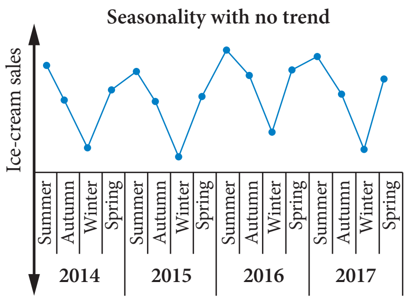
Trend

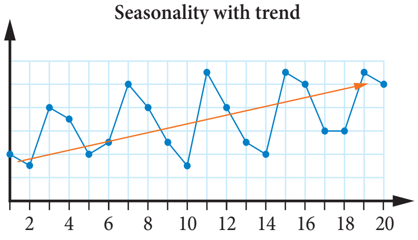




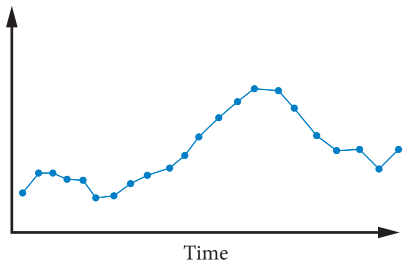


Seasonality

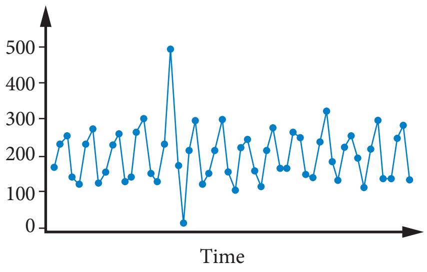

Seasonality with no trend


Seasonality with trend

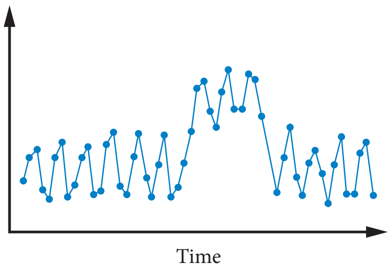
Irregular fluctuations

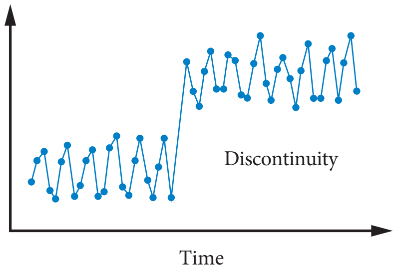


Outliers



Structural change





9780170371100

Smoothing

- **Smoothing** levels out fluctuations in time series to produce a smoother graph so we can see the time series trends more clearly.
- The most common **numerical smoothing** is **moving means**, which involves finding a series of means of a fixed number of data points. An odd number of data points involves one-step smoothing while an even number of data points involves a second smoothing step called **centring**.
- The most common **graphical smoothing** method is **moving medians**, which involves finding a series of medians of a fixed number of data points. We only need to consider using an odd number of data points for this method.
- Since medians, unlike means, are not influenced by outliers, moving medians has the advantage over moving means that extreme data values will be eliminated quickly.
- The number of moving points for both moving means and moving medians is usually chosen by looking at the natural cycle of the data being considered.

Seasonal adjustment

- **Seasonal indices** are used to make seasonal adjustments.
- The sum of the seasonal indices = the number of seasons.

Type of data	No. of seasons	Cycle	Sum of seasonal indices
Daily figures for data from Monday to Sunday	7	A week	7
Daily figures for data from Monday to Friday	5	A week	5
Monthly figures	12	A year	12
Quarterly accounts	4	A year	4

- To interpret seasonal indices convert them to percentages.
- To **de-seasonalise** time series data use the formula:

$$\text{De-seasonalised value} = \frac{\text{Actual value}}{\text{Seasonal index}}$$

- To find the actual value given the de-seasonalised value and the seasonal index, which is called **re-seasonalising**, use the following formula:

$$\text{Actual value} = \text{De-seasonalised value} \times \text{Seasonal index}$$

Fitting a least squares line

- We can use a least squares line of best fit to model time series trend as long as the data appears to be linear, but if there is seasonality then we need to de-seasonalise the data first before fitting the least squares line.
- The least squares line based can be used to **forecast**, but the result will give a de-seasonalised value. This value needs to be **re-seasonalised** to give the actual value.

DATA ANALYSIS

Examination 1

Reading time: (5 minutes)
Writing time: (30 minutes)

- Students are permitted to bring into the examination room: pens, pencils, highlighters, erasers, sharpeners, rulers, one bound reference, one approved graphics calculator or approved CAS calculator or CAS software and, if desired, one scientific calculator. Calculator memory DOES NOT need to be cleared.
- Students are NOT permitted to bring into the examination room: blank sheets of paper and/or white out liquid/tape.
- Students may refer to the sheet of miscellaneous formulas supplied.

Students are NOT permitted to bring mobile phones and/or any other unauthorised electronic devices into the examination room.

Instructions

Choose the response that is correct for the question.

A correct answer scores 1, an incorrect answer scores 0.

Marks will not be deducted for incorrect answers.

No marks will be given if more than one answer is completed for any question.

Unless otherwise indicated, the diagrams are not drawn to scale.

Question 1

The following ordered stem plot shows the areas, in square kilometres, of 27 suburbs of a large city.

Key: 1|6 = 1.6 km^2

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

The median area of these suburbs, in square kilometres, is

A. 3.0 **B.** 3.1 **C.** 3.5 **D.** 30.0 **E.** 30.5

[VCAA 2014 1CQ1]

EXAMINATION 1 – continued

Question 2

The time spent by shoppers at a hardware store on a Saturday is approximately normally distributed with a mean of 31 minutes and a standard deviation of 6 minutes.

If 2850 shoppers are expected to visit the store on a Saturday, the number of shoppers who are expected to spend between 25 and 37 minutes in the store is closest to

A. 16 **B.** 68 **C.** 460 **D.** 1900 **E.** 2400

[VCAA 2014 1CQ2]

Use the following information to answer Questions 3–5.

The following table shows the data collected from a sample of seven drivers who entered a supermarket car park. The variables in the table are:

- *distance* – the distance that each driver travelled to the supermarket from their home
- *sex* – the sex of the driver (female, male)
- *number of children* – the number of children in the car
- *type of car* – the type of car (sedan, wagon, other)
- *postcode* – the postcode of the driver's home

***Distance* (km)**	***Sex* (F = female, M = male)**	***Number of children***	***Type of car* (1 = sedan, 2 = wagon, 3 = other)**	***Postcode***
4.2	F	2	1	8148
0.8	M	3	2	8147
3.9	F	3	2	8146
5.6	F	1	3	8245
0.9	M	1	3	8148
1.7	F	2	2	8147
2.5	M	2	2	8145

Question 3

The mean, $\bar{x}$, and the standard deviation, s_x, of the variable distance are closest to

A. $\bar{x} = 2.5$, $s_x = 3.3$ **B.** $\bar{x} = 2.8$, $s_x = 1.7$

C. $\bar{x} = 2.8$, $s_x = 1.8$ **D.** $\bar{x} = 2.9$, $s_x = 1.7$

E. $\bar{x} = 3.3$, $s_x = 2.5$

[VCAA 2014 1CQ3]

Question 4

The number of categorical variables in this data set is

A. 0 **B.** 1 **C.** 2 **D.** 3 **E.** 4

[VCAA 2014 1CQ4]

Question 5

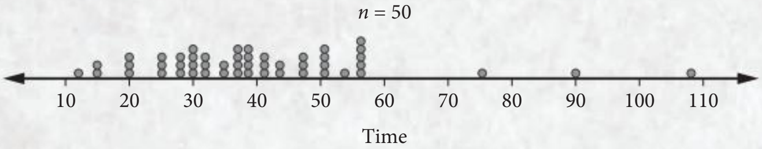
The number of female drivers with three children in the car is

A. 0 **B.** 1 **C.** 2 **D.** 3 **E.** 4

[VCAA 2014 1CQ5]

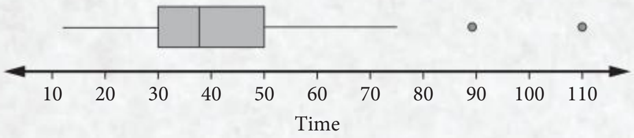
Question 6

The dot plot below shows the distribution of the time, in minutes, that 50 people spent waiting to get help from a call centre.

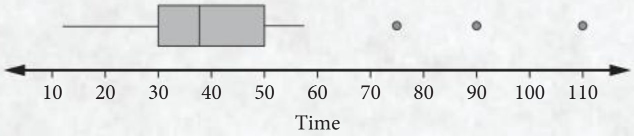


Which of the following boxplots best represents the data?

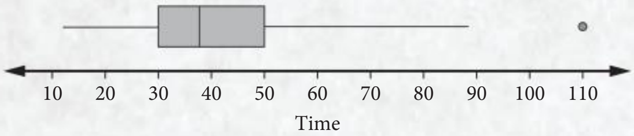
A.



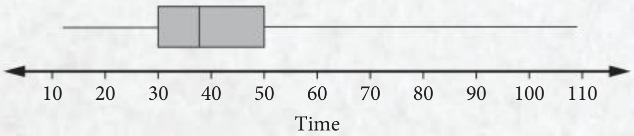
B.



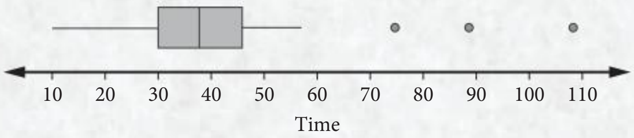
C.



D.



E.

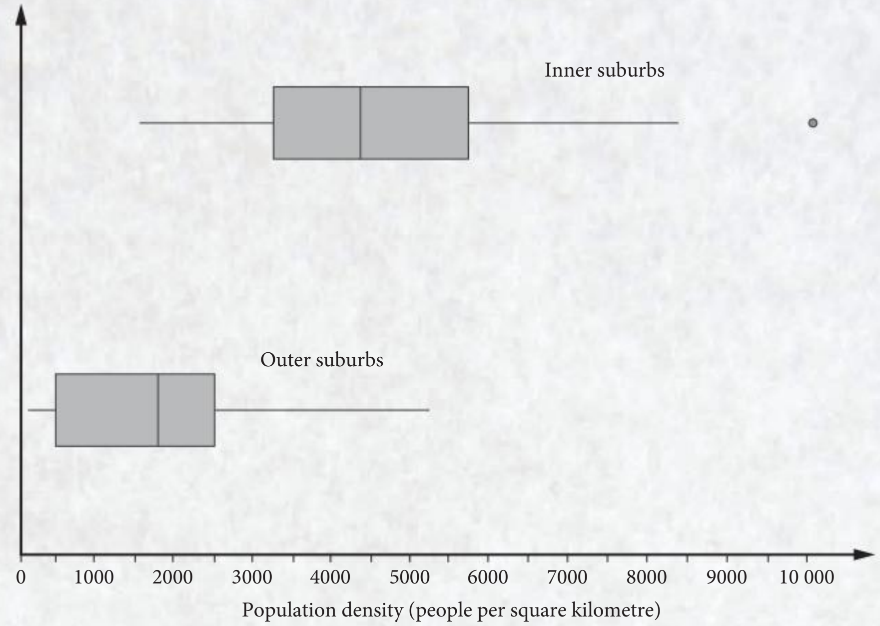


[VCAA 2014 1CQ6]

EXAMINATION 1 – continued

Question 7

The parallel boxplots below summarise the distribution of population density, in people per square kilometre, for 27 inner suburbs and 23 outer suburbs of a large city.



Which one of the following is not true?

A. More than 50% of the outer suburbs have population densities below 2000 people per square kilometre.

B. More than 75% of the inner suburbs have population densities below 6000 people per square kilometre.

C. Population densities are more variable in the outer suburbs than in the inner suburbs.

D. The median population density of the inner suburbs is approximately 4400 people per square kilometre.

E. Population densities are, on average, higher in the inner suburbs than in the outer suburbs.

[VCAA 2014 1CQ7]

Question 8

A single back-to-back stem plot would be an appropriate graphical tool to investigate the association between a car's speed, in kilometres per hour, and the

A. driver's age, in years.

B. car's colour (white, red, grey, other).

C. car's fuel consumption, in kilometres per litre.

D. average distance travelled, in kilometres.

E. driver's sex (female, male).

[VCAA 2014 1CQ8]

EXAMINATION 1 – continued

Question 9

The equation of a least squares line of best fit is used to predict the fuel consumption, in kilometres per litre of fuel, from a car's mass, in kilograms.

The equation predicts that a car weighing 900 kg will travel 10.7 km per litre of fuel, while a car weighing 1700 kg will travel 6.7 km per litre of fuel.

The slope of this least squares line of best fit is closest to

A. –250 **B.** –0.005 **C.** –0.004 **D.** 0.005 **E.** 200

[VCAA2014 1CQ9]

Use the following information to answer Questions 10 & 11.

The seasonal indices for the first 11 months of the year, for sales in a sporting equipment store, are shown in the table below.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Seasonal index	1.23	0.96	1.12	1.08	0.89	0.98	0.86	0.76	0.76	0.95	1.12	▨

Question 10

The seasonal index for December is

A. 0.89 **B.** 0.97 **C.** 1.02 **D.** 1.23 **E.** 1.29

[VCAA 2014 1CQ10]

Question 11

In May, the store sold $213 956 worth of sporting equipment. The de-seasonalised value of these sales was closest to

A. $165 857 **B.** $190 420 **C.** $209 677

D. $218 322 **E.** $240 400

[VCAA 2014 1CQ11]

Question 12

The seasonal index for heaters in winter is 1.25.

To correct for seasonality, the actual heater sales in winter should be

A. reduced by 20% **B.** increased by 20% **C.** reduced by 25%

D. increased by 25% **E.** reduced by 75%

[VCAA 2014 1CQ12]

EXAMINATION 1 – ended

DATA ANALYSIS

Examination 2

Reading time: (5 minutes)
Writing time: (30 minutes)

- Students are permitted to bring into the examination room: pens, pencils, highlighters, erasers, sharpeners, rulers, one bound reference, one approved graphics calculator or approved CAS calculator or CAS software and, if desired, one scientific calculator. Calculator memory DOES NOT need to be cleared.
- Students are NOT permitted to bring into the examination room: blank sheets of paper and/or white out liquid/tape.
- Students may refer to the sheet of miscellaneous formulas supplied.
- All written responses must be in English.

Students are NOT permitted to bring mobile phones and/or any other unauthorised electronic devices into the examination room.

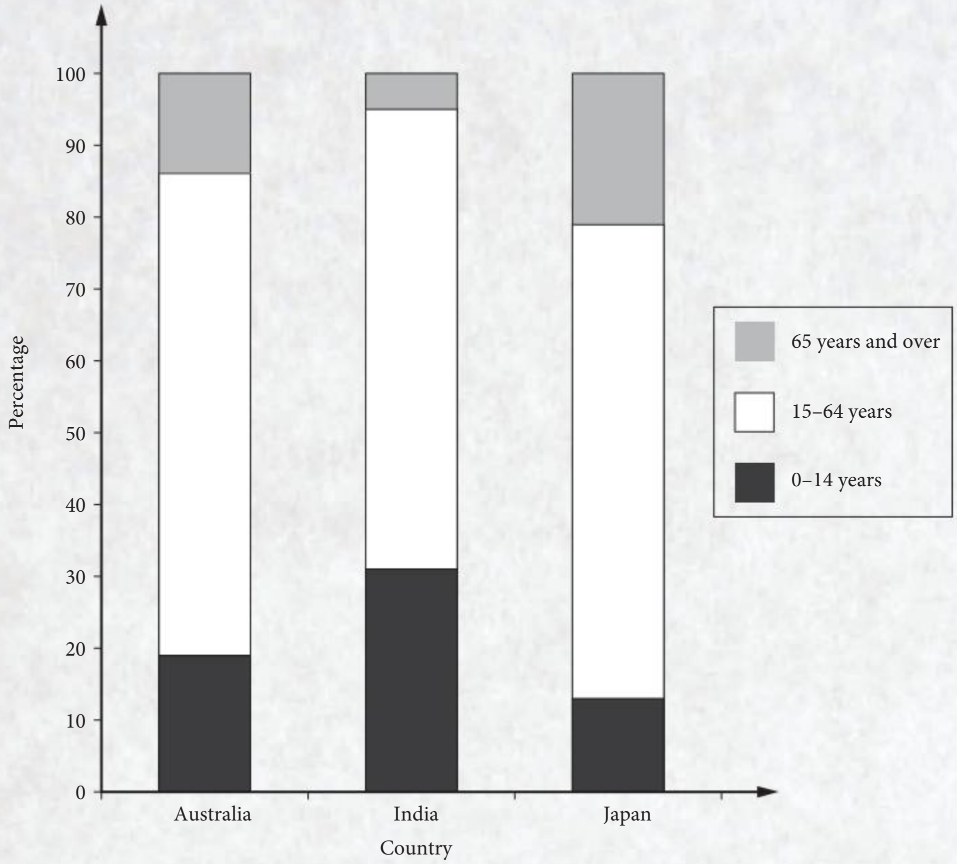
Instructions

You need not give numerical answers as decimals unless instructed to do so. Alternative forms may involve, for example, π, surds or fractions.
Diagrams are not to scale unless specified otherwise.

EXAMINATION 2 – continued

Question 1 (3 marks)

The segmented bar chart below shows the age distribution of people in three countries, Australia, India and Japan, for the year 2010.



a. Write down the percentage of people in Australia who were aged 0–14 years in 2010.
Write your answer correct to the nearest percentage. 1 mark

b. In 2010, the population of Japan was 128 000 000.
How many people in Japan were aged 65 years and over in 2010? 1 mark

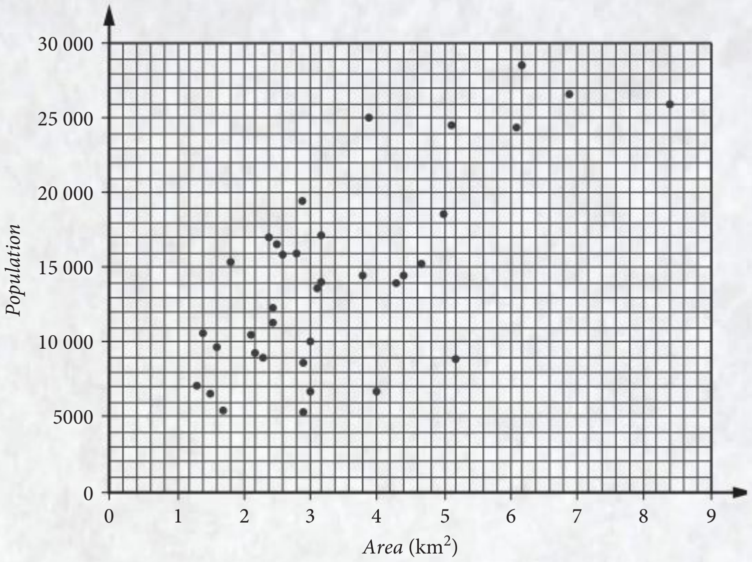
c. From the graph above, it appears that there is no association between the percentage of people in the 15–64 age group and the country in which they live.
Explain why, quoting appropriate percentages to support your explanation. 1 mark

[VCAA 2014 2CQ1]

EXAMINATION 2 – continued

Question 2 (6 marks)

The scatterplot shows the *population* and *area* (in square kilometres) of a sample of inner suburbs of a large city.



The equation of the least squares line of best fit for the data in the scatterplot is

$$population = 5330 + 2680 \times area$$

a. Write down the response variable. 1 mark

b. Draw the least squares line of best fit on the **scatterplot above**. 1 mark

(*Answer on the scatterplot above.*)

c. Interpret the slope of this least squares line of best fit in terms of the variables *area* and *population*. 2 marks

d. Wiston is an inner suburb. It has an area of 4 km^2 and a population of 6690.

The correlation coefficient, r, is equal to 0.668.

i Calculate the residual when the least squares line of best fit is used to predict the population of Wiston from its area. 1 mark

ii What percentage of the variation in the population of the suburbs is explained by the variation in area?

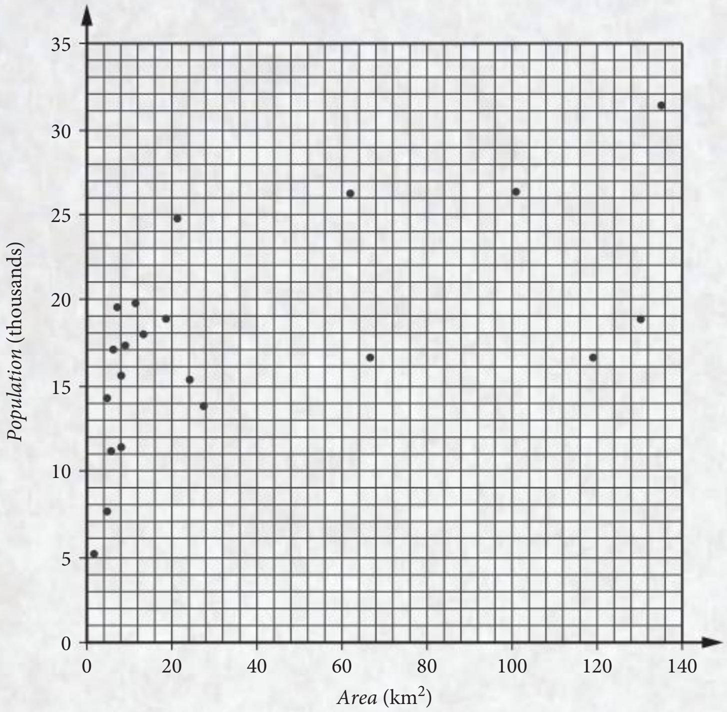
Write your answer correct to 1 decimal place. 1 mark

[VCAA 2014 2CQ2]

EXAMINATION 2 – continued

Question 3 (2 marks)

The scatterplot and table below show the *population*, in thousands, and the *area*, in square kilometres, for a sample of 21 outer suburbs of the same city.



Area (km^2)	*Population* (thousands)
1.6	5.2
4.4	14.3
4.6	7.5
5.6	11.0
6.3	17.1
7.0	19.4
7.3	15.5
8.0	11.3
8.8	17.1
11.1	19.7
13.0	17.9
18.5	18.7
21.3	24.6
24.2	15.2
27.0	13.6
62.1	26.1
66.5	16.4
101.4	26.2
119.2	16.5
130.7	18.9
135.4	31.3

In the outer suburbs, the association between *population* and *area* is non-linear.

A **log** transformation can be applied to the variable *area* to linearise the scatterplot.

a. Apply the **log** transformation to the data and determine the equation of the least squares line of best fit that allows the population of an outer suburb to be predicted from the logarithm of its area.

Write the slope and intercept of this line of best fit in the boxes provided below.

Write your answers correct to 1 decimal place. 1 mark

$$population = \boxed{} + \boxed{} \times \log_{10}(area)$$

b. Use this line of best fit equation to predict the population of an outer suburb with an area of 90 km^2.

Write your answer correct to the nearest one thousand people. 1 mark

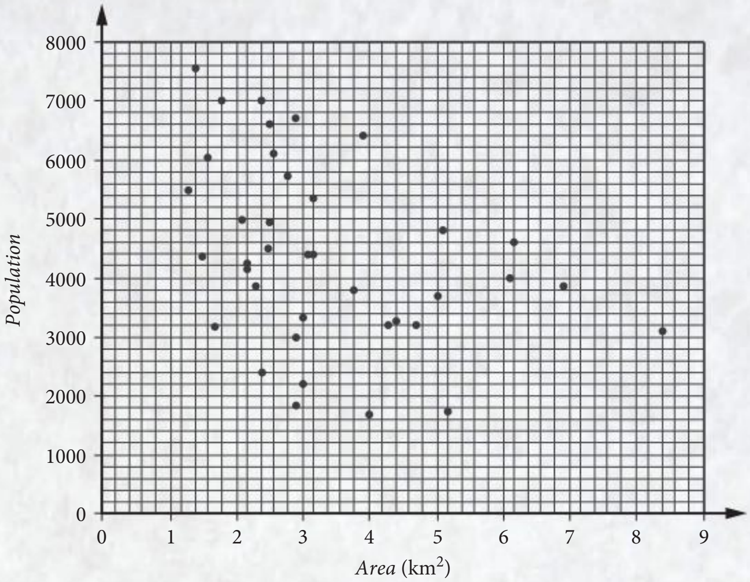
[VCAA 2014 2CQ3]

EXAMINATION 2 – continued

Question 4 (4 marks)

The scatterplot shows the *population density*, in people per square kilometre, and the *area*, in square kilometres, of 38 inner suburbs of the same city.

For this scatterplot, $r^2 = 0.141$.



a. Describe the association between the variables *population density* and *area* for these suburbs in terms of strength, direction and form. 1 mark

b. The mean and standard deviation of the variables *population density* and *area* for these 38 inner suburbs are shown in the table below.

	Population density **(people per km^2)**	***Area*** **(km^2)**
Mean	4370	3.4
Standard deviation	1560	1.6

i One of these suburbs has a population density of 3082 people per square kilometre.

Determine the standard *z*-score of this suburb's population density.

Write your answer correct to 1 decimal place. 1 mark

Assume the areas of these inner suburbs are approximately normally distributed.

ii How many of these 38 suburbs are **expected** to have an area that is two standard deviations or more above the mean?

Write your answer correct to the nearest whole number. 1 mark

iii How many of these 38 inner suburbs **actually** have an area that is two standard deviations or more above the mean? 1 mark

[VCAA 2014 2CQ4]

EXAMINATION 2 – ended

CHAPTER 6

DEPRECIATION AND INTEREST

Prior learning

6.1 First-order recurrence relations

Recurrence relations and sequences

A **recurrence relation** (also known as a **difference equation**) is a rule that specifies a particular term in a sequence using the previous term or terms. Recurrence relations can be used to model real-life situations in areas such as biology and finance that involve growth and decay. We will be looking at their applications to finance.

Consider the sequence 3, 8, 13, 18…

The first term in the sequence is 3 and each subsequent term in the sequence is 5 more than the previous term.

We can write this as $t_0 = 3$, $t_{n+1} = t_n + 5$

which is saying: starting with 3, generate the next term by adding five to the previous term.

It generates the sequence in this way:

$t_0 = 3$

$t_1 = t_0 + 5 = 3 + 5 = 8$

$t_2 = t_1 + 5 = 8 + 5 = 13$

$t_3 = t_2 + 5 = 13 + 5 = 18$

etc.

This is called a **first-order recurrence relation** because it links *two consecutive* terms in the sequence.

A first-order recurrence relation is a rule that specifies how a particular term in a sequence can be found from the previous term in the same sequence.

It consists of

- the starting term (also called initial term), e.g. $t_0 = 3$
- a rule linking two consecutive terms, e.g. $t_{n+1} = t_n + 5$

iStock.com/denphumi

Worked example 1

Find the first four terms of the sequence defined by the recurrence relation $t_0 = 3$, $t_{n+1} = 2t_n + 1$, showing all the steps of the calculations.

	Working
1 The first term is t_0.	$t_0 = 3$
2 The rule $t_{n+1} = 2t_n + 1$ means take the previous term, multiply by 2 and add 1 to get the next term.	$t_1 = 2t_0 + 1$ $= 2 \times 3 + 1$ $= 7$
3 Repeat for the next two terms.	$t_2 = 2t_1 + 1$ $= 2 \times 7 + 1$ $= 15$ $t_3 = 2t_2 + 1$ $= 2 \times 15 + 1$ $= 31$
4 Write down the first four terms of the sequence.	The first four terms of the sequence are 3, 7, 15, 31.

Using CAS Generating sequences through recursive computation

Find the first six terms of the sequences defined by each recurrence relations.

a $t_0 = 12, t_{n+1} = t_n - 4$

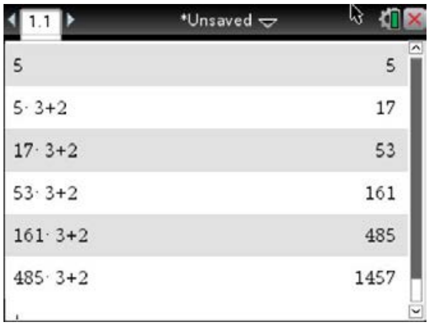
b $t_0 = 5, t_{n+1} = 3t_n + 2$

TI-NSPIRE CAS

a Open a New Document with a Calculator page.
Type 12 and press `enter`.
Type –4. Press `enter`.
Continue to press `enter` until you have the required number of terms.

```
1.1        *Unsaved
12        12
12−4       8
8−4        4
4−4        0
0−4       -4
-4−4      -8
                2/6
```

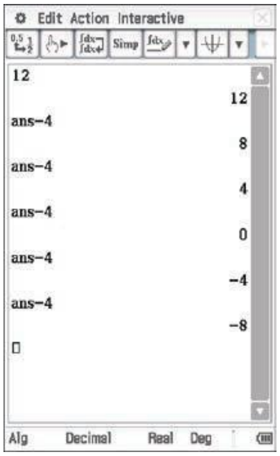
b Open a New Document with a Calculator page.
Type 5 and press `enter`.
Type $\times 3 + 2$ and press `enter`.
Continue to press `enter` until you have the required number of terms.



```
1.1        *Unsaved
5            5
5·3+2       17
17·3+2      53
53·3+2     161
161·3+2    485
485·3+2   1457
```

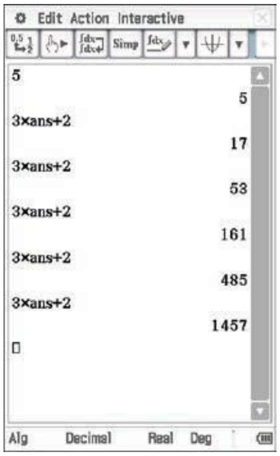
CLASSPAD

a Use the Main $\sqrt{\alpha}$ application.
Type 12 and press `EXE`.
When you press `−`4 after entering 12, it will automatically do `ans` `−`4 and continue until you have the required number of terms.



```
Edit Action Interactive
12
        12
ans−4
         8
ans−4
         4
ans−4
         0
ans−4
        −4
ans−4
        −8
Alg   Decimal   Real   Deg
```

b Use the Main $\sqrt{\alpha}$ application.
Type 5 and press `EXE`.
Press `Keyboard` and tap `Math2` and `ans`.
Press $3 \times$ Ans $+ 2$ then `EXE`.
Continue to press `EXE` until you have the required number of terms.
Note: Ans is in all the Math options from the keyboard, Math1 and Math2 and Math3.



```
Edit Action Interactive
5
           5
3×ans+2
          17
3×ans+2
          53
3×ans+2
         161
3×ans+2
         485
3×ans+2
        1457
Alg   Decimal   Real   Deg
```

Notation for first-order recurrence relations

There are a number of alternative ways of writing first-order recurrence relations. As an example, let's look at the different ways of writing $t_0 = -5, t_{n+1} = 2t_n - 3$.

1 The variable doesn't have to indicated by t. Often the pronumeral used is chosen to fit the context of the question (e.g. V for value or A for amount). For example, all of these are equally correct:

$t_0 = -5, t_{n+1} = 2t_n - 3$

$u_0 = -5, u_{n+1} = 2u_n - 3$

$V_0 = -5, V_{n+1} = 2V_n - 3$

$A_0 = -5, A_{n+1} = 2A_n - 3$

We will call this the standard form. It can be varied in a number of ways.

2 The first term can be written after the rule:

$t_{n+1} = 2t_n - 3$, where $t_0 = -5$

3 The first term can be indicated by t_1 instead of t_0 and the sequence it generates is the same. For example:

$t_1 = -5, t_{n+1} = 2t_n - 3$

(Note the ClassPad gives you a choice of using a_0 or a_1 as the first term and provides other options.)

4 The terms in the rule can be reordered. For example:

$t_0 = -5, t_{n+1} - 2t_n = -3$

5 Since t_{n+1} and t_n represent two consecutive terms in the same way that t_n and t_{n-1} represent two consecutive terms, the recurrence relation can also be written as:

$t_0 = -5, t_n = 2t_{n-1} - 3$

6 Brackets can be used instead of subscripts. The TI-Nspire CAS notation uses:

Initial term is –5, $u(n) = 2u(n-1) - 3$

Worked example 2

Rewrite the following first-order recurrence relations in standard form.

a $u_{n+1} - u_n = 7$, where $u_1 = 10$

b $B_1 = 2, B_n = 4B_{n-1} + 3$

c $a(n) + 5a(n-1) = 40$, where $a(0) = 6$

		Working
a	Reorder and replace subscripts.	$u_0 = 10, u_{n+1} = u_n + 7$
b	Replace subscripts.	$B_0 = 2, B_{n+1} = 4B_n + 3$
c	Replace the bracketed values with standard subscripts and reorder.	$a_0 = 6, a_{n+1} = -5a_n + 40$

Sometimes a term other than the first term is given in the recurrence relation. In these cases, it's still possible to find other terms in the sequence.

Worked example 3

The recurrence relation $w_{n+1} = 3w_n - 4$ can be used to generate a sequence. If $w_3 = -1$, find w_5.

		Working
1	w_3 is known so w_4 can be found using the rule.	$w_4 = 3w_3 - 4$ $= 3 \times (-1) - 4$ $= -7$
2	Find w_5 using w_4.	$w_5 = 3w_4 - 4$ $= 3 \times (-7) - 4$ $= -25$

Graphs of first-order recurrence relations

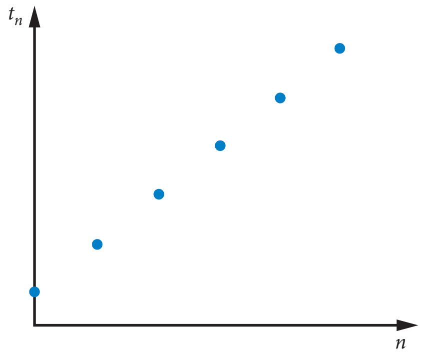
The graphs of first-order recurrence relations that we will be dealing with fall into one of these categories. They should be drawn as a series of as points since they represent discrete values.

Increasing straight line

Linear growth

- t_n has coefficient of 1
- positive number added

e.g. $t_0 = 22$, $t_{n+1} = t_n + 4$

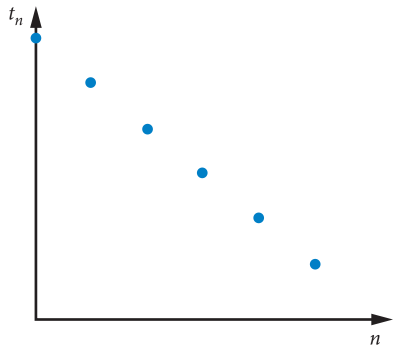


Decreasing straight line

Linear decay

- t_n has coefficient of 1
- positive number subtracted

e.g. $t_0 = 22$, $t_{n+1} = t_n - 4$

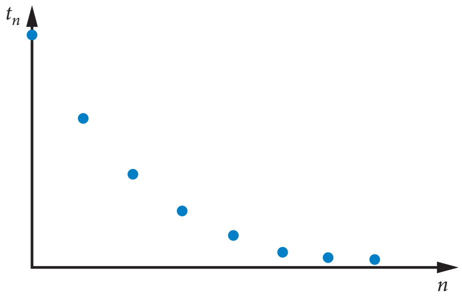


Decreasing curve which never reaches zero

Geometric decay

- t_n has coefficient between 0 and 1
- no numbers added or subtracted

e.g. $t_0 = 10$, $t_{n+1} = 0.75t_n$

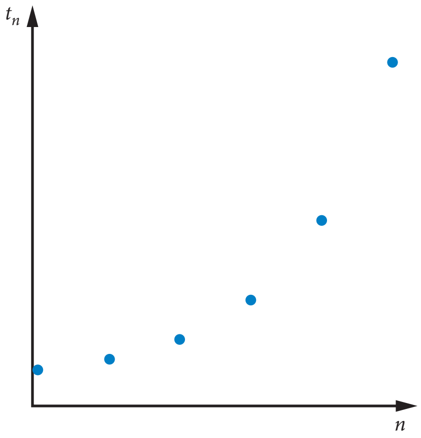


Upward curve

Geometric growth

- t_n has coefficient greater than 1
- no numbers added or subtracted

e.g. $t_0 = 35\ 000$, $t_{n+1} = 1.62t_n$

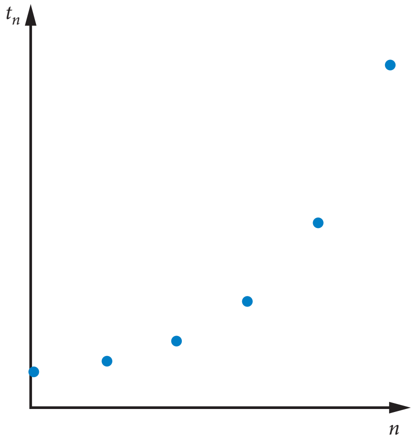


Upward curve

Combined linear growth and geometric growth

- t_n has coefficient greater than 1
- positive number added

e.g. $t_0 = 2$, $t_{n+1} = 2t_n + 5$

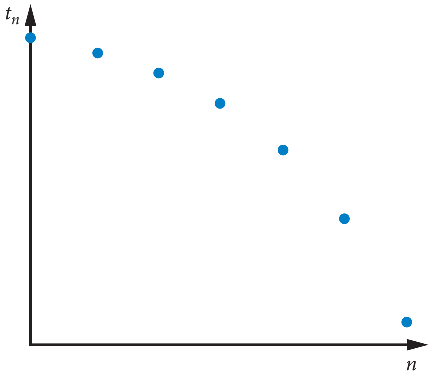


Downward curve

Combined linear decay and geometric growth where the linear decay is greater than the geometric growth

- t_n has coefficient greater than 1
- positive number subtracted

e.g. $t_0 = 15\ 000$, $t_{n+1} = 1.03t_n - 2000$



Using CAS Plotting recurrence relations

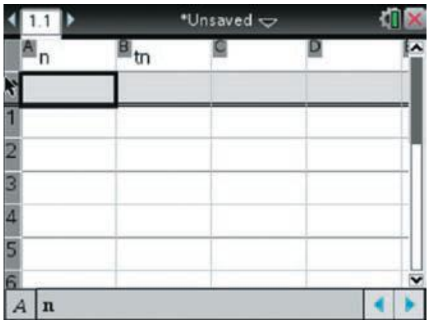
Plot the first five terms of the sequence generated by the recurrence relation $t_0 = -5$, $t_{n+1} = 2t_n - 3$.

TI-NSPIRE CAS

STEP 1

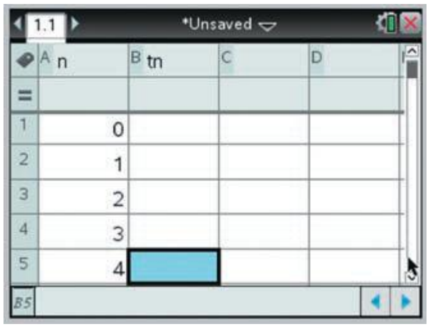
Open a New Document with a Lists & Spreadsheet page.

Name Column A, n and Column B, tn.



STEP 2

To graph the first five terms of a sequence enter the term numbers 0 to 4 in Column A.



	A n	B tn
=		
1	0	
2	1	
3	2	
4	3	
5	4	

STEP 3

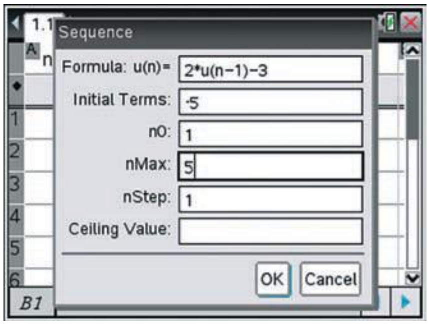
Move the cursor to the formula entry line for Column B.

Press menu

3: Data

1: Generate Sequence

Define the sequence as shown.



Formula: u(n)= 2*u(n−1)−3

Initial Terms: -5

n0: 1

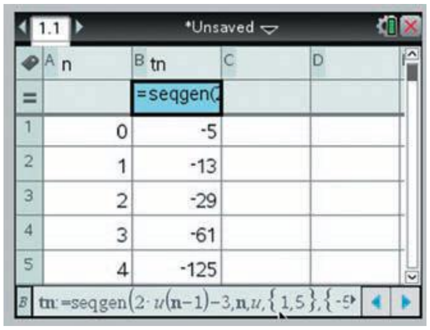
nMax: 5

nStep: 1

Ceiling Value:

STEP 4

Select OK.

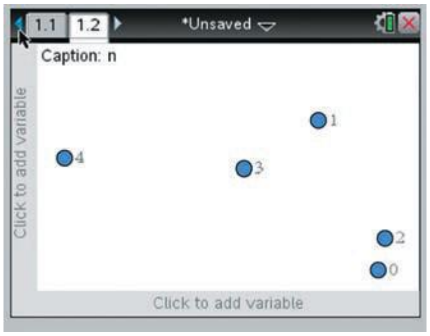


	A n	B tn
=		=seqgen(
1	0	-5
2	1	-13
3	2	-29
4	3	-61
5	4	-125

B tn:=seqgen(2·u(n−1)−3,n,u,{1,5},{-5

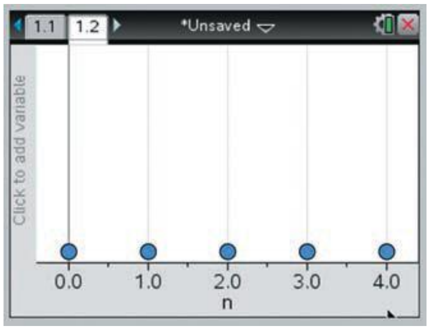
STEP 5

Add a Data & Statistics page.



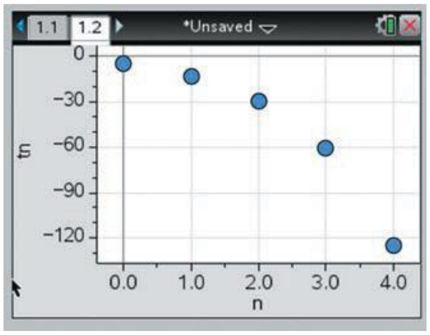
STEP 6

Click to add variable at the bottom of the screen and select n.



STEP 7

Click to add variable at the left of the screen and select tn.



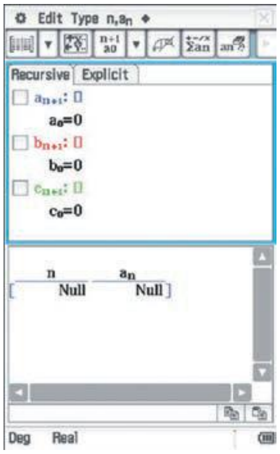
CLASSPAD

STEP 1

Use the Sequence application.

Tap Recursive.

Tap the arrow in the middle of the top tool bar and a drop-down menu will appear with different templates for recurrence relations. Tap n+1 30.



☐ a_{n+1}:

$a_0=0$

☐ b_{n+1}:

$b_0=0$

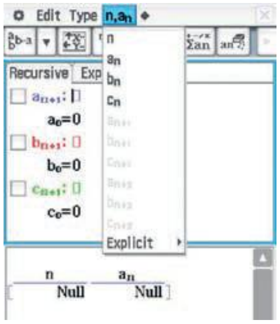
☐ c_{n+1}:

$c_0=0$

n	a_n
Null	Null

STEP 2

Type the recurrence relation in the entry line after a_{n+1}: Tap n,a_n at the top to get a_n.



☐ a_{n+1}:

$a_0=0$

☐ b_{n+1}:

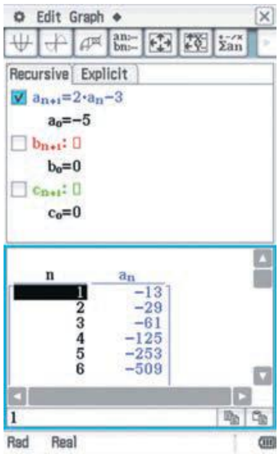
$b_0=0$

☐ c_{n+1}:

$c_0=0$

n	a_n
Null	Null

Make sure that the recurrence relation is checked, then tap . Do this after selecting the terms using scale icon .



☑ $a_{n+1}=2\cdot a_n-3$

$a_0=-5$

☐ b_{n+1}:

$b_0=0$

☐ c_{n+1}:

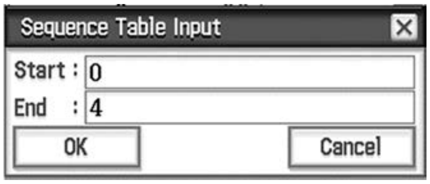
$c_0=0$

n	a_n
1	−13
2	−29
3	−61
4	−125
5	−253
6	−509

1

STEP 3

It is necessary to graph the first five terms of the sequence, so tap and set Start to 0 and End to 4. Tap OK.



Sequence Table Input

Start : 0

End : 4

STEP 4

Tap to set a suitable viewing window.

Tap OK.

View Window

☐ x-log ☐ y-log

xmin : −0.044

max : 4.5

scale : 1

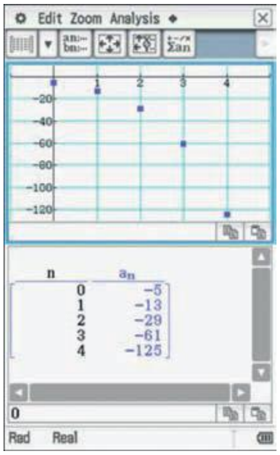
dot : 0.0147532467532468

ymin : −130

max : 5

STEP 5

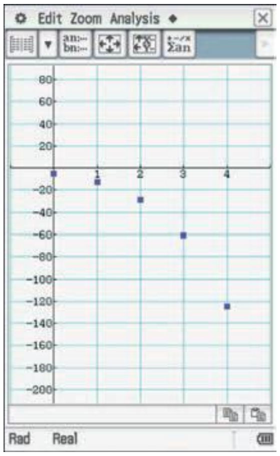
Tap to graph.



n	a_n
0	−5
1	−13
2	−29
3	−61
4	−125

STEP 6

To view only the graph, tap the graph window, then tap Resize .



EXAM PREP 6.1

First-order recurrence relations

Prep 1 WORKED EXAMPLE 1

Find the first four terms of the sequence defined by the recurrence relation $t_0 = 1$, $t_{n+1} = 3t_n + 2$, showing all the steps of the calculations.

Prep 2 USING CAS: GENERATING SEQUENCES THROUGH RECURSIVE COMPUTATION

Using repeated steps, find the first six terms of the sequences defined by each of the following recurrence relations.

a $t_0 = 15, t_{n+1} = t_n - 2$

b $t_0 = 2, t_{n+1} = 4t_n$

c $t_0 = -8, t_{n+1} = t_n + 7$

d $t_0 = 64, t_{n+1} = -0.5t_n$

e $u_0 = 50, u_{n+1} = u_n - 5$

f $w_0 = 3, w_{n+1} = 2w_n + 1$

g $K_0 = 1, K_{n+1} = 3 - 5K_n$

h $P_0 = 40, P_{n+1} = 0.5P_n + 2$

Prep 3 WORKED EXAMPLE 2

Rewrite the following first-order recurrence relations in standard form.

a $A_{n+1} - A_n = -5$, where $A_1 = 12$

b $w_1 = 2, w_n = 2w_{n-1} - 6$

c $u(n) + 10u(n-1) = 13$, where $u(0) = 7$

d $t_{n+1} = 6t_n + 5$, where $t_1 = 1$

e $P_1 = 12, P_n + 3P_{n-1} = -5$

Prep 4 WORKED EXAMPLE 3

The recurrence relation $u_{n+1} = 2u_n + 5$ can be used to generate a sequence. If $u_4 = 9$, find u_6.

Prep 5 USING CAS: PLOTTING A RECURRENCE RELATION

Plot the first six terms of the sequence generated by the recurrence relation $t_0 = 1, t_{n+1} = 1.3t_n - 5$.

EXAMPLE PRACTICE 6.1

First-order recurrence relations

Question 1

The first six terms of a sequence are

8, 2, 8, 2, 8, 2

A recurrence relation that can be used to generate this sequence is

A $t_{n+1} = t_n - 10$

B $t_{n+1} = 10 - t_n$

C $t_{n+1} = 0.25t_n$

D $t_{n+1} = t_n - 6$

E $t_{n+1} = 8 - t_n$

Question 2

The sequence 2, 3, 7, 23, … is generated by the first-order recurrence relation $t_{n+1} = 4t_n + d$, with $t_0 = 2$. The value of d is

A −10 **B** −5 **C** −2 **D** 1 **E** 5

Question 3

The sequence 4, 9, 24, 69, … is generated from the recurrence relation $t_{n+1} = kt_n - 3$, where $t_0 = 4$. The value of k is

A 2 **B** 3 **C** 5 **D** 7 **E** 15

Question 4

The first term in a sequence is 3.

The second term is obtained by multiplying the value of the first term by 5, then subtracting 2. The third term is obtained by multiplying the value of the second term by 5, then subtracting 2. This pattern continues.

The recurrence relation which generates this sequence is

A $t_n = 5t_{n+1} - 2$, where $t_1 = 3$

B $t_{n+1} = 5t_n - 2$, where $t_1 = 3$

C $-5t_{n+1} = t_n - 2$, where $t_1 = 3$

D $5t_{n+1} = t_n - 2$, where $t_1 = 3$

E $t_{n+1} = -5t_n - 2$, where $t_1 = 3$

[VCAA 2005 1NPQ6]

Question 5

The recurrence relation $t_{n+1} = at_n + 6$, where $t_1 = 5$

generates the sequence

5, 21, 69, 213

The value of a is

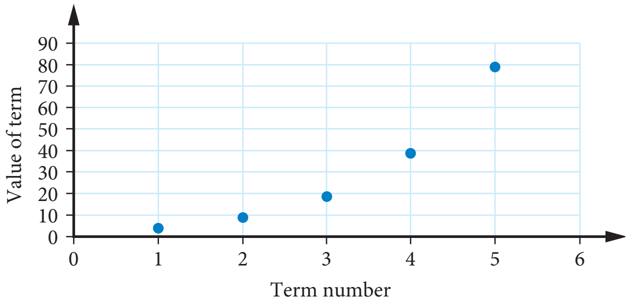
A −1 **B** 3 **C** 4 **D** 15 **E** 16

[VCAA 2007 1NPQ3]

Question 6

The values of the first five terms of a sequence are plotted on the graph shown.

The first-order recurrence relation that could describe the sequence is



A $t_{n+1} = t_n + 5, t_1 = 4$

B $t_{n+1} = 2t_n + 1, t_1 = 4$

C $t_{n+1} = t_n - 3, t_1 = 4$

D $t_{n+1} = t_n + 3, t_1 = 4$

E $t_{n+1} = 3t_n, t_1 = 4$

[VCAA 2006 1NPQ7]

Question 7

The recurrence relation $u_{n+1} = 4u_n - 2$ generates a sequence.

If $u_2 = 2$, then u_4 will be equal to

A 4 **B** 8 **C** 22 **D** 40 **E** 42

[VCAA 2009 1NPQ7]

Question 8

A recurrence relation is defined by

$$f_{n+1} - f_n = 5, \text{ where } f_1 = -1$$

The sequence $f_1, f_2, f_3, \ldots$ is

A 5, 4, 3, … **B** 4, 9, 14, … **C** −1, −6, −11, …

D −1, 4, 9, … **E** −1, 6, 11, …

[VCAA 2006 1NPQ5]

Question 9

The first five terms of a sequence of numbers are 20, 10, 20, 10, 20, …

A recurrence relation that generates this sequence is

A $t_{n+1} = 20 - t_n$ **B** $t_{n+1} = t_n - 20$ **C** $t_{n+1} = 0.5t_n$

D $t_{n+1} = t_n - 10$ **E** $t_{n+1} = 30 - t_n$

[VCAA 2003 1NPQ9]

6.2 Flat rate and unit cost depreciation

While some things tend to increase in value over time (e.g. property, gold, antiques and collectibles), most assets used by businesses (e.g. computers, tablets, equipment and machines) decrease in value over time. We use the term **depreciation** to describe this decrease in value. Depreciation occurs due to age, amount of use, or lack of demand. The true value of a business at any time can't be known unless the depreciated value of assets is taken into account. The new reduced value of an asset at any point in time is called the **future value** (also known as **book value**).

Future value = Purchase price – Depreciation

There are three main ways of depreciating assets:

- flat rate
- unit cost
- reducing balance depreciation

Flat-rate depreciation recurrence relations

Flat rate depreciation (also known as **straight-line depreciation**) calculates the future value of an asset by reducing the value every year by a fixed amount. The amount can be given in dollars or as a fixed percentage of the purchase price.

iStock.com/pete corr

Worked example 4

A ride-on mower is purchased by a gardening business for \$9500. Its value depreciates at a flat rate of 5% per annum.

a Copy and complete the following table to find the future value of the ride-on mower after 5 years.

n	Future value after n years (\$)
0	9500
1	9500 – =
2	– =
3	– =
4	– =
5	– =

b Write a recurrence relation for the future value of the ride-on mower. What shape will the graph be?

c Use the recurrence relation to find a rule for the future value of the ride-on mower after n years.

d Use the rule to find the future value of the ride-on mower after 15 years.

Working

a **1** Calculate the annual depreciation.

$$\text{Depreciation} = 5\% \text{ of } \$9500 = \frac{5}{100} \times \$9500 = \$475$$

2 Use a table to determine the future value of the ride-on mower after 5 years.

Use a CAS/calculator's recursive computation to check your answers.

n	Future value after n years (\$)
0	9500
1	9500 – 475 = 9025
2	9025 – 475 = 8550
3	8550 – 475 = 8075
4	8075 – 475 = 7600
5	7600 – 475 = 7125

3 Write the answer.

The future value of the ride-on mower after 5 years is \$7125.

b Each value is calculated by deducting \$475 from the previous value.

Let V_n = the future value of the ride-on mower after n years.

The recurrence relation is

$V_0 = 9500,\ V_{n+1} = V_n - 475$

Since V_n has a coefficient of 1 and a number is being subtracted, the graph will form a decreasing straight line.

c Use the recurrence relation to find a pattern for a rule.

$V_0 = 9500$

$V_1 = 9500 - 475 = 9500 - 1 \times 475$

$V_2 = (9500 - 475) - 475 = 9500 - 2 \times 475$

$V_3 = (9500 - 475 - 475) - 475 = 9500 - 3 \times 475$

…

$V_n = 9500 - n \times 475$ which can be written as

$V_\text{n} = V_0 - n \times 475$

d Substitute the number of years into the rule and write the answer in words.

$V_n = 9500 - n \times 475$

$n = 15$

$V_{15} = 9500 - 15 \times 475$

$= 2375$

The future value of the ride-on mower after 15 years is $2375.

Flat rate depreciation general rule

The future value after n time periods, V_n, of an asset depreciated using flat rate depreciation is given by the recurrence relation

$V_0 =$ starting value, $V_{n+1} = V_n - d$

where $d =$ the amount depreciated each time period (usually a year)

$$d = \frac{\text{flat rate of depreciation}}{100} \times V_0$$

The general rule for finding the future value of flat rate depreciation after n time periods is

$V_n = V_0 - nd$

Exam hack

The general rule is *not* a recurrence relation. It *doesn't* tell you how a particular term in a sequence can be found from the previous term in the same sequence. It is a formula.

Worked example 5

Steve Gates paid $50 000 for new computer equipment at the start of the 2012 financial year. He used flat rate depreciation to revalue the equipment, and at the start of the 2016 financial year he revalued it at $28 000.

a What is the dollar amount of depreciation from 2012 to 2016?

b What was the annual flat rate of depreciation he used, as a percentage of the purchase price?

Working

a Calculate how much dollar value has been lost during this time.

If the computer equipment has a value of \$50 000 in 2012 and \$28 000 in 2016, it has depreciated by 50 000 – 28 000 = \$22 000.

b **1** List the known values from the question.

The start of 2012 financial year to the start of 2016 financial year is four depreciating periods.

$n = 4$

$V_0 = 50\ 000$

$V_4 = 28\ 000$

2 Use the general rule for flat rate depreciation to find the amount depreciated each time period (i.e. find *d*). Solve using the CAS/calculator solve function if necessary.

$$\begin{aligned} V_n &= V_0 - nd \\ V_4 &= V_0 - 4d \\ 28\ 000 &= 50\ 000 - 4d \\ 4d &= 22\ 000 \\ d &= 5500 \end{aligned}$$

3 Use this dollar amount to find the flat rate of depreciation. Solve using the CAS/calculator solve function if necessary.

$$d = \frac{\text{flat rate of depreciation}}{100} \times V_0$$

$$5500 = \frac{\text{flat rate of depreciation}}{100} \times 50\ 000$$

flat rate of depreciation = 11%

Unit cost depreciation recurrence relations

Unit cost depreciation occurs when an asset is depreciated according to the amount of use it has had, not according to its age. When using the unit cost method of depreciation, the amount of depreciation is determined by applying a rate per unit of use.

Worked example 6

A photocopier is purchased by a business for \$25 000. It depreciates at a rate of 25 cents for every 100 copies made. Each year it makes 150 000 copies.

a Copy and complete the following table to find the future value of the photocopier after 3 years.

n	Future value after *n* years (\$)
0	25 000
1	25000 – =
2	
3	

b Write a recurrence relation for the future value of the photocopier. What shape will the graph be?

c Use the recurrence relation to find a rule for the future value of the photocopier after *n* years.

d Use the rule to find the future value of the photocopier after 15 years.

e The business anticipates that after the third year of use the photocopier will make only 120 000 copies a year. Add two more rows to the table to find the future value after 5 years.

Working

a **1** Determine the rate of depreciation per unit of use.

25 cents per 100 copies $= \frac{0.25}{100} =$ \$0.0025 per copy

2 Calculate the annual depreciation.

Unit cost depreciation = Rate of depreciation per unit of use × Number of units of use

Each year it makes 150 000 copies, so

Annual depreciation = 0.0025 × 150 000
= \$375

3 Use a table to determine the future value of the photocopier after 3 years.

Use a CAS/calculator's recursive computation to check your answers.

n	Future value after *n* years (\$)
0	25 000
1	25 000 – 375 = 24 625
2	24 625 – 375 = 24 250
3	24 250 – 375 = 23 875

4 Write the answer.

The future value of the photocopier after 3 years is \$23 875.

b Each value is calculated by deducting \$375 from the previous value.

Let V_n be the future value of the photocopier after *n* years.

$V_0 = 25\,000,\ V_{n+1} = V_n - 375$

Since V_n has a coefficient of 1, the graph will form a decreasing straight line (as long as the use is the same each year).

c Use the recurrence relation to find a pattern for a rule.

$V_0 = 25\,000$
$V_1 = 25\,000 - 375 = 25\,000 - 1 \times 375$
$V_2 = (25\,000 - 375) - 375 = 25\,000 - 2 \times 375$
$V_3 = (25\,000 - 375 - 375) - 375 = 25\,000 - 3 \times 375$
…
$V_n = 25\,000 - n \times 375$ which can be written as
$V_n = V_0 - n \times 375$

d Substitute the number of years into the rule and write the answer in words.

$V_n = 25\,000 - n \times 375$

$n = 15$

$V_{15} = 25\,000 - 15 \times 375$

$= 19\,375$

The future value of the photocopier after 15 years is $19 375.

e **1** Calculate the new annual depreciation.

Unit cost depreciation = Rate of depreciation per unit of use × Number of units of use

Annual depreciation = 0.0025 × 120 000

= $300

2 Add two more years to the table to show the calculations.

n	Future value after *n* years ($)
0	25 000
1	25 000 – 375 = 24 625
2	24 625 – 375 = 24 250
3	24 250 – 375 = 23 875
4	23 875 – 300 = 23 575
5	23 575 – 300 = 23 275

3 Write the answer.

The future value of the photocopier after 5 years is $23 275.

Unit cost depreciation general rule

For unit cost depreciation:

Unit cost depreciation = Rate of depreciation per unit of use × Number of units of use

The future value after n time periods, V_n, of an asset depreciated using unit cost depreciation is given by the following recurrence relation *if the usage is the same each time period*:

V_0 = starting value, $V_{n+1} = V_n - d$

where d is the amount depreciated each time period (usually a year)

d = Rate of depreciation per unit of use × Number of units of use

The general rule for finding the future value of unit cost depreciation after n time periods *if the usage is the same each time period* is

$V_n = V_0 - nd$

Exam hack

Once you have calculated *d*, the recurrence relation and general rule for unit cost depreciation becomes the same as for flat rate depreciation. This is only true, however, as long as the usage is the same each year. In practice, for unit cost depreciation, the amount deducted each year will vary. For example, a car being depreciated according to the number of kilometres travelled each year will have a different kilometre count each year. This means that *d* will change every year. As a result, graphs for unit cost depreciation take the form of decreasing straight lines only if the level of use is the same each year.

Worked example 7

A delivery van was purchased for \$80 000. The van's value depreciates at a rate of 25 cents per kilometre.

a Find the value of the van after it has travelled a total distance of 150 000 kilometres.

b What is the number of kilometres the van has travelled before it depreciates to \$20 000?

			Working
a	**1**	Find the rate of depreciation per unit of use.	25 cents per kilometre means the rate of depreciation per unit of use = \$0.25 per kilometre.
	2	Calculate the depreciation amount.	Depreciation = Rate of depreciation per unit of use × Number of units of use Depreciation = 150 000 × \$0.25 = \$37 500
	3	Find the future value after it has been depreciated by this amount.	Future value = Purchase price – Depreciation = \$80 000 – \$37 500 = \$42 500
	4	Write the answer in words.	The depreciated value of the van is \$42 500.
b	**1**	Find the depreciation amount from the future value and purchase price. Solve using the CAS/calculator solve function if necessary.	Future value = Purchase price – Depreciation \$20 000 = \$80 000 – Depreciation Depreciation = \$80 000 – \$20 000 Depreciation = \$60 000
	2	Find the number of units of use. Solve using the CAS/calculator solve function if necessary.	Depreciation = Rate of depreciation per unit of use × Number of units of use 60 000 = 0.25 × Number of units of use $\text{Number of units of use} = \frac{60\,000}{0.25}$ = 240 000
	3	Write the answer in words.	The van has travelled 240 000 km before it depreciates to \$20 000.

Flat rate and unit cost depreciation

Prep 1 WORKED EXAMPLE 4

An industrial vacuum cleaner is purchased by a cleaning business for $10 500. Its value depreciates at a flat rate of 7% per annum.

a Copy and complete the table to find the future value of the industrial vacuum cleaner after 5 years.

n	Future value after *n* years ($)
0	10 500
1	10 500 – =
2	– =
3	– =
4	– =
5	– =

b Write a recurrence relation for the future value of the industrial vacuum cleaner. What shape will the graph be?

c Use the recurrence relation to find a rule for the future value of the industrial vacuum cleaner after *n* years.

d Use the rule to find the future value of the industrial vacuum cleaner after 12 years.

Prep 2 WORKED EXAMPLE 4

Use the general rule for flat rate depreciation after *n* time periods to find the future value of each of these.

a A coffee machine is purchased for $6000 and its value depreciates at a flat rate of 5% per annum. Using the general rule for flat rate depreciation after *n* time periods, what is its future value after

i 8 years? **ii** 12 years? **iii** 15 years?

b A video conference system is purchased for $35 000 and its value depreciates at a flat rate of 10% per annum. Using the general rule for flat rate depreciation after *n* time periods, what is its future value after

i 7 years? **ii** 8 years? **iii** 9 years?

Prep 3 WORKED EXAMPLE 5

Melinda Jobs paid $60 000 for new office furniture at the start of the 2011 financial year. She used flat rate depreciation to revalue the equipment, and at the start of the 2016 financial year she revalued it at $24 000.

a What is the dollar amount of depreciation from 2011 to 2016?

b What was the annual flat rate of depreciation she used, as a percentage of the purchase price?

Prep 4 WORKED EXAMPLE 6

A colour laser printer is purchased by a business for \$20 000. It depreciates at a rate of 35 cents for every 50 pages printed. Each year it prints 80 000 copies.

n	Future value after n years (\$)
0	20 000
1	
2	
3	

a Copy and complete the table to find the future value of the colour laser printer after 3 years.

b Write a recurrence relation for the future value of the colour laser printer. What shape will the graph be?

c The business anticipates that after the third year of use the colour laser printer will only print 60 000 pages a year. Add two more rows to the table to find the future value after 5 years.

Prep 5 WORKED EXAMPLE 7

A vehicle was purchased for \$75 000. The van's value depreciates at a rate of 30 cents per kilometre.

a Find the value of the vehicle after it has travelled a total distance of 140 000 kilometres.

b What is the number of kilometres the vehicle has travelled before it depreciates to \$15 000?

EXAM PRACTICE 6.2

Flat rate and unit cost depreciation

Question 1

Which one of the following situations would be best modelled by the recurrence relation

$$V_0 = 25\,000,\ V_{n+1} = V_n - 1200$$

A A car originally purchased for \$25 000 is depreciated at a rate of \$1200 for each kilometre it travels.

B A car originally purchased for \$23 800 is depreciated at a rate of \$1200 for each kilometre it travels.

C A car originally purchased for \$23 800 is depreciated at a rate of \$1200 each year.

D A car originally purchased for \$25 000 has been depreciated to \$1200.

E A car originally purchased for \$25 000 is depreciated at a rate of \$1200 each year.

Question 2

A photocopier bought for \$20 000 is depreciated at a rate of \$3000 every year. Which of the following is *not* true?

A The depreciation method used is flat rate depreciation.

B The recurrence relation which can be used to model this is $V_0 = 20\ 000$, $V_{n+1} = V_n - 3000$, where V_n is the future value of the photocopier after n years.

C The rule for finding the future value after the nth year is $V_n = V_0 - 3000n$, where V_n is the future value of the photocopier after n years.

D The value of the photocopier after 2 years is \$17 000.

E The depreciation method used is straight-line depreciation.

Question 3

An asset purchased for \$10 000 is depreciated by a flat rate of 15% in the first year and a flat rate of 30% each year thereafter. Which calculation gives the value of the asset at the end of the third year?

A 10 000 – 1500 – 3000 – 3000

B 10 000 – 1500 – 3000

C 10 000 – 3000 – 3000 – 3000

D 10 000 – 1500 – 1500 – 3000

E 10 000 – 1500 – 1500 – 1500

Question 4

A van is purchased for \$56 000.

Its value depreciates at a rate of 42 cents for each kilometre that it travels. The value of the van after it has travelled 32 000 km is

A \$13 400 **B** \$26 880 **C** \$29 120 **D** \$32 480 **E** \$42 560

[VCAA 2011 1BRMQ3]

Question 5

A delivery truck when new was valued at \$65 000.

The truck's value depreciates at a rate of 22 cents per kilometre travelled.

After it has travelled a total distance of 132600 km, the value of the truck will be

A \$14 300 **B** \$22 100 **C** \$22 516 **D** \$29 172 **E** \$35 828

[VCAA 2009 1BRMQ4]

Question 6

A machine that makes boxes costs \$45 000.

Its value depreciates by 5 cents for every box it makes.

Each year it makes 120 000 boxes.

The depreciated value of this machine at the end of **two** years is

A \$33 000 **B** \$38 000 **C** \$39 000 **D** \$45 000 **E** \$115 000

[VCAA 2005 1BRMQ4]

Question 7

A photocopier is depreciated by \$0.04 for each copy it makes.

Three years ago the photocopier was purchased for \$48 000.

Its depreciated value now is \$21 000.

The total number of copies made by the photocopier in the three years is

A 108 000 **B** 192 000 **C** 276 000 **D** 525 000 **E** 675 000

[VCAA 2006 1BRMQ5]

Question 8

A car is valued at \$30 000 when new.

Its value is depreciated by 25 cents for each kilometre it travels.

The number of kilometres the car travels before its value depreciates to \$8000 is

A 32 000 **B** 55 000 **C** 88 000 **D** 120 000 **E** 550 000

[VCAA 2007 1BQ2]

Question 9

Zoltan is running a convenience store. He purchases equipment for \$6500. It is anticipated that the equipment will last five years and have a depreciated value of \$2000.

Assuming the straight-line method of depreciation, the equipment is depreciated annually by

A \$400 **B** \$900 **C** \$1027 **D** \$1300 **E** \$4500

[VCAA 2003 1BQ5]

Question 10

A machine is purchased for \$36 000. Its value depreciates at a rate of \$0.16 for each unit it produces. On average, the machine produces 24 000 units a year.

Using the unit cost method of depreciation, the value of the machine after six years of use is closest to

A \$3840 **B** \$7200 **C** \$12 960 **D** \$23 040 **E** \$32 160

[VCAA 2002 1BQ6]

Question 11

Rae paid \$40 000 for new office equipment at the start of the 2007 financial year.

At the start of each following financial year, she used flat rate depreciation to revalue her equipment. At the start of the 2010 financial year she revalued her equipment at \$22 000.

The annual flat rate of depreciation she used, as a percentage of the purchase price, was

A 11.25% **B** 15% **C** 17.5% **D** 35% **E** 45%

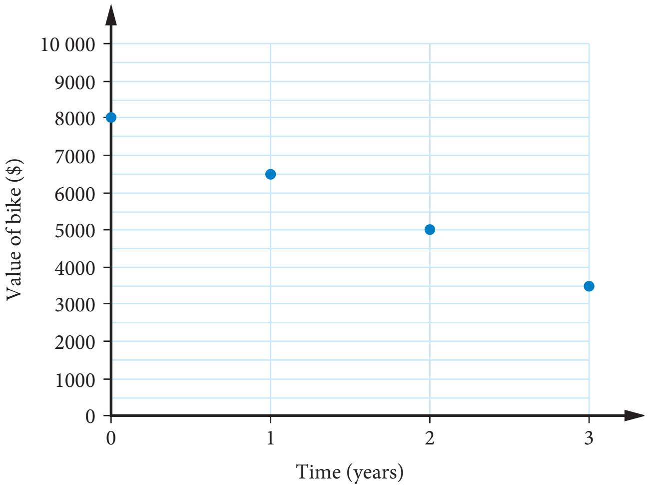
[VCAA 2010 1BRMQ8]

Question 12

Hugo is a professional bike rider.

The value of his bike will be depreciated over time using the flat rate method of depreciation.

The graph shows his bike's initial purchase price and its value at the end of each year for a period of three years.



a What was the initial purchase price of the bike? 1 mark

b **i** Show that the bike depreciates in value by $1500 each year. 1 mark

ii Assume that the bike's value continues to depreciate by $1500 each year.

Determine its value five years after it was purchased. 1 mark

The unit cost method of depreciation can also be used to depreciate the value of the bike.

In a two-year period, the total depreciation calculated at $0.25 per kilometre travelled will equal the depreciation calculated using the flat rate method of depreciation as described above.

c Determine the number of kilometres the bike travels in the two-year period. 1 mark

[VCAA 2013 2BRMQ1]

Question 13

Khan will depreciate his $900 fax machine for taxation purposes. He considers two methods of depreciation.

Flat rate depreciation

Under flat rate depreciation the fax machine will be valued at $300 after five years.

i Calculate the annual depreciation in dollars. 1 mark

Unit cost depreciation

Suppose Khan sends 250 faxes a year. The $900 fax machine is depreciated by 46 cents for each fax it sends.

ii Determine the value of the fax machine after five years. 1 mark

[VCAA 2007 2BRMQ3b]

Question 14

Michelle intends to own a $17 000 car for 15 years. At the end of this time its value will be $3500.

a By what amount, in dollars, would the car's value depreciate annually if Michelle used the flat rate method of depreciation? 1 mark

b Determine the annual flat rate of depreciation correct to 1 decimal place. 1 mark

[VCAA 2008 2BRMQ4]

6.3 Reducing balance depreciation

Reducing balance depreciation recurrence relations

Reducing balance depreciation calculates the future value of an asset by reducing it every year by a fixed percentage *of its value in the preceding year*. With flat rate depreciation, the percentage each year is of the purchase price, so it's the same dollar amount every year. With reducing balance depreciation, the dollar amount of depreciation changes every year.

iStock.com/nicolamargaret

Worked example 8

A business purchased a photocopier for $10 000 in 2014. It is depreciated using reducing balance depreciation at a rate of 18% per annum. Give all answers to the nearest dollar.

a Copy and complete the table below to find

i the future value of the photocopier after 5 years

ii the amount of depreciation in the fourth year.

n	Depreciation after n years ($)	Future value after n years ($)
0	–	10 000
1	$\frac{18}{100} \times 10000 = 1800$	10 000 – 1800 = 8200
2		
3		
4		
5		

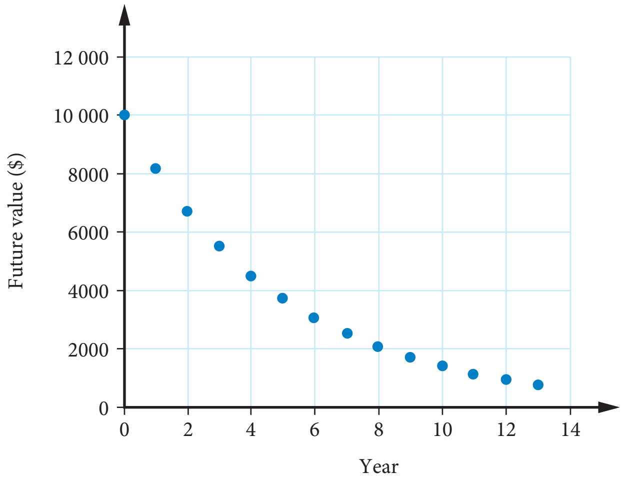
b Write down a recurrence relation that gives the value of the photocopier after *n* years.

c Use the recurrence relation to find a rule for the future value of the photocopier after *n* years.

d The photocopier is replaced when its value first falls below \$3000. At the start of which year will the copier be replaced?

e How much is the photocopier depreciated by in the sixth year?

f Use the graph to find the photocopier's approximate value after 8 years.



Working

a **1** Calculate the percentage of successive future values and subtract from the previous future value.

Use a CAS/calculator's recursive computation where possible.

Give all values to the nearest dollar, but don't round until after all the calculations have been done.

(Note you may get close to but not exactly equal to the answers given if you round too early.)

n	Depreciation after *n* years (\$)	Future value after *n* years (\$)
0	–	10 000
1	$\frac{18}{100} \times 10000 = 1800$	10 000 – 1800 = 8200
2	$\frac{18}{100} \times 8200 = 1476$	8200 – 1476 = 6724
3		
4		
5		

n	Depreciation after *n* years (\$)	Future value after *n* years (\$)
0	–	10 000
1	1800	8200
2	1476	6724
3	1210	5514
4	992	4521
5	814	3707

2 Read from the table.

i The future value of the photocopier after 5 years is \$3707.

ii The amount of depreciation in the fourth year is \$992.

b If the photocopier depreciates by 18% per annum, the value of the photocopier in any one year is 82% of its value the previous year.

Let V_n = the future value of the photocopier after n years.

$V_0 = 10\ 000,\ V_{n+1} = 0.82V_n$

c Use the recurrence relation to find a pattern for a rule.

$V_0 = 10\ 000$
$V_1 = 0.82 \times 10\ 000 = 0.82^1 \times 10\ 000$
$V_2 = 0.82 \times (0.82 \times 10\ 000) = 0.82^2 \times 10\ 000$
$V_3 = 0.82 \times (0.82 \times 0.82 \times 10\ 000) = 0.82^3 \times 10\ 000$
…
$V_n = 0.82^n \times 10\ 000$ which can be written as
$V_n = 0.82^n \times V_0$ or $V_n = 0.82^n V_0$

d Substitute values into the rule to find when the future value first falls below \$3000.

$V_6 = 0.82^6 \times 10\ 000 = 3040.07$
$V_7 = 0.82^7 \times 10\ 000 = 2492.85$
The photocopier will be replaced after $n = 7$ years.

Establish which year this term corresponds to and write the answer.

n	Year
0	Start 2014
1	Start 2015
2	Start 2016
3	Start 2017
4	Start 2018
5	Start 2019
6	Start 2020
7	Start 2021

$n = 7$ corresponds to the start of 2021.

The photocopier is replaced at the start of 2021.

e The depreciation after six years occurs at V_6. So the depreciation *in* the sixth year is $V_5 - V_6$.

Give your answer to the nearest dollar.

$V_5 = \$3707.40$

$V_6 = \$3040.07$

Depreciation in the sixth year $= V_5 - V_6$
$= \$3707.40 - \$3040.07$

\$667 rounded to the nearest dollar.

f Read from the graph.

The value of the photocopier after 8 years is around \$2000.

Reducing balance depreciation general rule

The future value after n time periods, V_n, of an asset depreciated using reducing balance depreciation is given by the recurrence relation

V_0 = starting value, $V_{n+1} = (1 - \frac{r}{100}) \times V_n$

where r is the depreciation rate per time period

The general rule for finding the future value of reducing balance depreciation after n time periods is

$$V_n = (1 - \frac{r}{100})^n \times V_0$$

where r is the depreciation rate per time period

Depreciation *in* the nth time period = $V_{n-1} - V_n$.

Worked example 9

A truck was bought for $250 000 and is being depreciated on a reducing balance basis rate of 8% per annum. Use the general rule for reducing balance depreciation after n time periods to find the future value of the truck after 9 years to the nearest dollar.

		Working
1	State the rule and find the known values.	$V_n = (1 - \frac{r}{100})^n \times V_0$ $V_0 = 250\ 000$ $n = 9$ $r = 8$
2	Substitute in the known values, and solve using the CAS/calculator solve function.	$V_9 = (1 - \frac{8}{100})^9 \times 250\ 000$ $= 118\ 040$
3	Write the answer in words.	The future value of the truck after 9 years is $118 040.

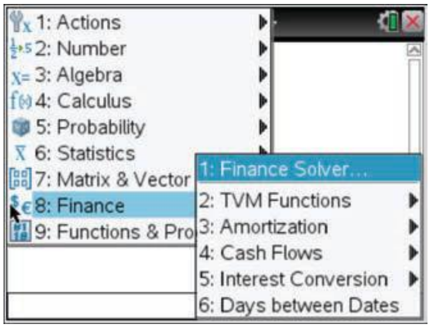
Reducing balance depreciation and finance solvers

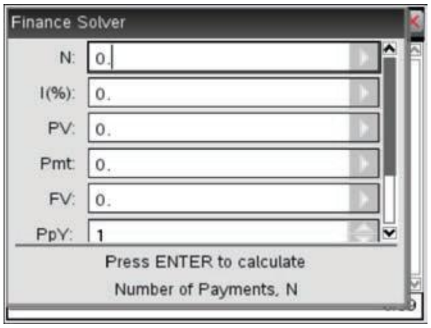
Finance solvers (also known as **TVMs**, which stands for Time Value of Money), such as 'Finance Solver' on the TI-Nspire CAS or 'Financial Application' on the ClassPad, can be used to answer questions on reducing balance depreciation.

TI-NSPIRE CAS

Open a New Document with a Calculator Page.

Press menu, 8(Finance) and 1(Finance Solver…).




Finance Solver

N: 0.

I(%): 0.

PV: 0.

Pmt: 0.

FV: 0.

PpY: 1

Press ENTER to calculate
Number of Payments, N

The fields for the Finance Solver are defined as follows.

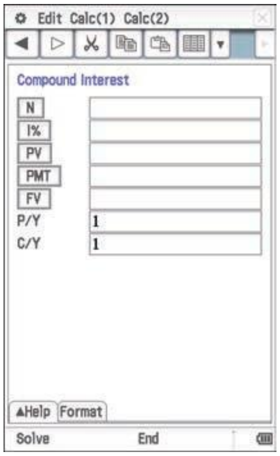
- **N** is the number of time periods.
- **I%** is the interest rate as a percentage per annum.
- **PV** is the **present value** (for depreciation calculations, this is negative since you have spent money to purchase the asset).
- **Pmt** is the value of any payments being made (for depreciation calculations, this is zero).
- **FV** is the future value (this will be positive since the future value of an asset for depreciation calculations is the value of the asset to you at that time).
- **PpY** is the number of payments per year (for depreciation calculations this is always 1 since depreciations are done once a year).
- **CpY** is the number of times in a year interest is compounded (for the problems we work on, this will always be the same as PpY, so for depreciation calculations it will always be 1).
- **PmtAt** is when the payment is made, at the 'beginning' or 'end' of the time period. Depreciation occurs at the end of the time period so leave PmtAt set to END.
- At the bottom of the screen, enter when the payment is made, at the 'beginning' or 'end' of the time period. Depreciation occurs at the end of the time period so set this to END.

When using the Finance solver, press tab to move between fields when entering data. Once the data is entered, press tab to move the cursor to the field representing the unknown quantity and press enter.

CLASSPAD

Tap Menu and the 2nd page, followed by the Financial application.

Tap **Calc(1)** on the top toolbar followed by **Compound Interest**.



P/Y 1

C/Y 1

We use the Compound Interest option on the ClassPad to best do depreciation calculations. The fields for the Compound Interest application are defined as follows.

- **N** is the number of time periods.
- **I%** is the interest rate as a percentage per annum.
- **PV** is the present value (for depreciation calculations, this is negative since you have spent money to purchase the asset).
- **PMT** is the value of any payments being made (for depreciation calculations, this is zero).
- **FV** is the future value (this will be positive since the future value of an asset for depreciation calculations is the value of the asset to you at that time).
- **P/Y** is the number of payments per year (for depreciation calculations this is always 1 since depreciations are done once a year).
- **C/Y** is the number of times in a year interest is compounded (for the problems we work on, this will always be the same as PpY, so for depreciation calculations it will always be 1).
- Compound interest is paid at the end of the time period so make sure that ‘End’ is showing at the bottom of the screen in the middle.

Tapping Help at the bottom left of the screen gives the definition for a selected field. Once the data is entered, tap the unknown quantity to find its value.

Using CAS Reducing balance depreciation with finance solvers

Vince bought a car six years ago for $26 500. After being depreciated on a reducing balance basis, the car is now valued at $9000. Calculate, correct to 1 decimal place, the percentage annual rate of depreciation of the value of Vince's car over six years.

STEP 1

Write down values of the known quantities in each field and write a question mark next to the unknown quantity.

N = 6
I% = ?
PV = −26500
Pmt or PMT = 0
FV = 9000
PpY or P/Y = 1
CpY or C/Y = 1
PmtAt = END

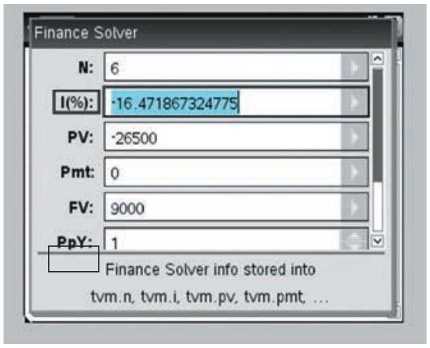
The cell I% is initially left blank.

STEP 2

Enter the data in the finance solver or financial application.

TI-NSPIRE CAS

Move the cursor to the data entry line for I% and press enter.



Finance Solver
N: 6
I(%): -16.471867324775
PV: -26500
Pmt: 0
FV: 9000
PpY: 1
Finance Solver info stored into tvm.n, tvm.i, tvm.pv, tvm.pmt, ...

CLASSPAD

Tap Menu and the 2nd page, followed by the Financial application.

Tap Calc(1) on the top toolbar followed by Compound Interest.

Fill in the fields on the screen using information from the question, including when the payment is made (Begin or End).

Tap I%.

Compound Interest
N 6
I% −16.47186732
PV −26500
PMT 0
FV 9000
P/Y 1
C/Y 1
Solve End

STEP 3

Write the answer, correct to the number of decimal places asked for.

The percentage annual rate of depreciation is 16.5%.

9780170371100

Exam hack

The crucial part of using finance solvers is to get the negatives right. If money is moving *away* from you, the value should be negative, and if the money is moving *to* you it's positive.

EXAM PREP 6.3

Reducing balance depreciation

Prep 1 WORKED EXAMPLE 8

A business purchased a workstation for $12 000 in 2016. It is depreciated using reducing balance depreciation at a rate of 16% per annum.

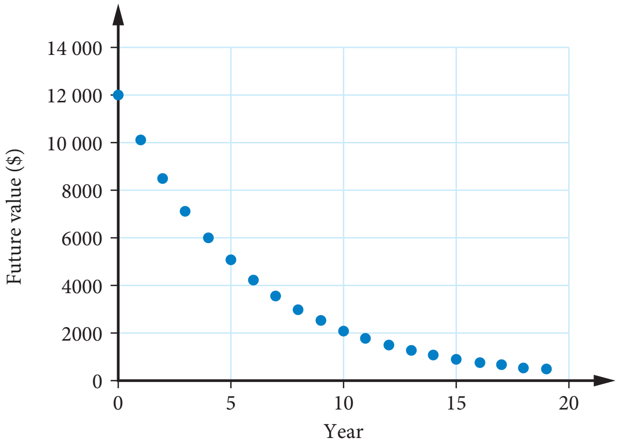
a Copy and complete the table to find to the nearest dollar

i the future value of the workstation after 5 years

ii the amount of depreciation in the fourth year.

n	Depreciation after n years ($)	Future value after n years ($)
0	–	12 000
1	$\frac{16}{100} \times 12\,000 = 1920$	$12\,000 - 1920 = 10\,080$
2		
3		
4		
5		

b Use the graph to find the workstation's approximate value after 14 years.



Prep 2 WORKED EXAMPLE 8

A company car is purchased for $80 000 at the start of 2016. It is depreciated on a reducing balance basis at a rate of 14% per annum. Give all your answers to the nearest dollar.

a Write down a recurrence relation that gives the value of the car after *n* years.

b Use the recurrence relation to find a rule for the future value of the car after *n* years.

c The car is replaced when its value first falls below $50 000. At the start of which year will the car be replaced?

d How much is the car depreciated by in the sixth year?

Prep 3 WORKED EXAMPLE 9

A bus was bought for $350 000 and is being depreciated on a reducing balance basis rate of 9% per annum. Use the general rule for reducing balance depreciation after *n* time periods to find the future value of the bus after 10 years to the nearest dollar.

Prep 4 USING CAS: REDUCING BALANCE DEPRECIATION WITH FINANCIAL SOLVERS

Saskia bought a car seven years ago for $28 000. After being depreciated on a reducing balance basis, the car is now valued at $8500. Calculate, correct to 1 decimal place, the percentage annual rate of depreciation of the value of Saskia's car over seven years.

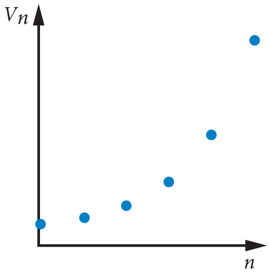
EXAM PRACTICE 6.3

Reducing balance depreciation

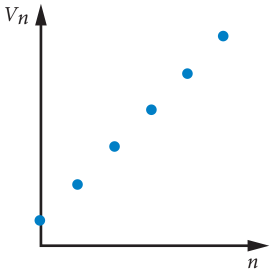
Question 1

A machine is purchased for $50 000 and is depreciated on a reducing balance basis at a rate of 12% per annum. Which of the following would best match the shape of the graph of n against V_n, the value of the machine after n years?

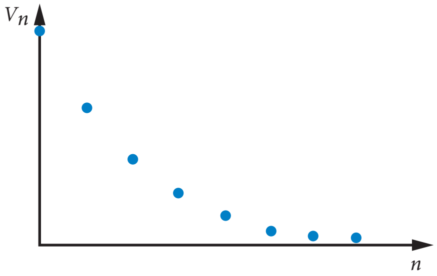
A



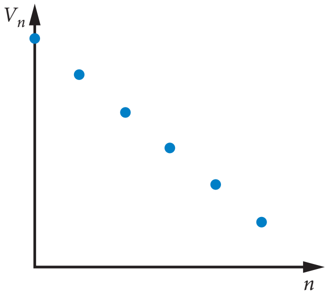
B



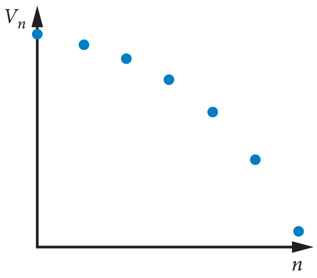
C



D



E



Question 2

A business purchased a computer for \$4000. It is depreciated using reducing balance depreciation at a rate of 15% per annum. What are the two values for $n = 2$ in the depreciation table?

n	Depreciation after *n* years (\$)	Future value after *n* years (\$)
0	–	4000
1		
2		

A 600 and 3400
B 510 and 2890
C 600 and 510
D 15 and 4000
E $\frac{15}{100}$ and 3400

Question 3

Which of the following recurrence relations best models a car purchased at \$32 000 being depreciated using reducing balance depreciation at a rate of 11% per annum?

A $V_n = 0.11^n \times 32\,000$
B $V_n = 0.89^n \times 32\,000$
C $V_0 = 32\,000,\ V_{n+1} = 0.11V_n$
D $V_0 = 32\,000,\ V_{n+1} = 0.89V_n$
E $V_0 = 32\,000,\ V_n = 0.89V_{n+1}$

Question 4

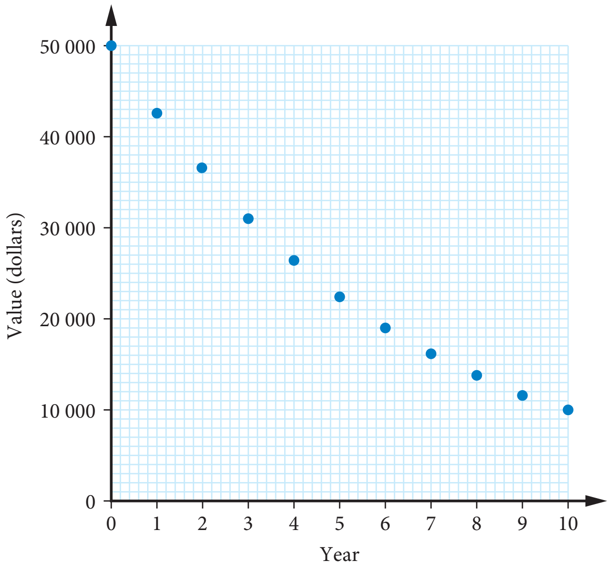
If V_n is the value after n years of an asset depreciated using reducing balance depreciation, which of the following represents the amount of depreciation in the 7th year?

A $V_7 - V_6$
B $V_6 - V_7$
C V_7
D V_6
E V_8

Question 5

A new kitchen in a restaurant cost \$50 000. Its value is depreciated over time using the reducing balance method.

The value of the kitchen in dollars at the end of each year for ten years is shown in the graph.



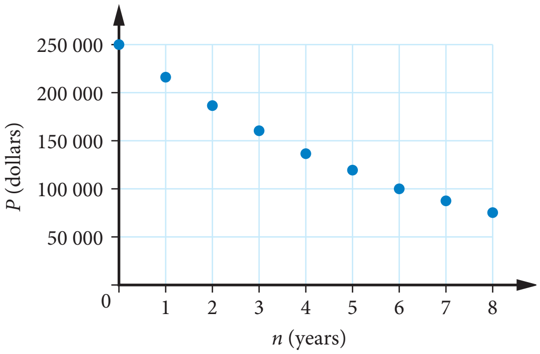
Which one of the following statements is **true**?

A The kitchen depreciates by \$4000 annually.
B At the end of five years, the kitchen's value is less than \$20 000.
C The reducing balance depreciation rate is less than 5% per annum.
D The annual depreciation rate increases over time.
E The amount of depreciation each year decreases over time.

[VCAA 2007 1BRMQ5]

Question 6

The following graph shows the decreasing value of an asset over eight years.



Let P be the value of the asset, in dollars, after n years.

A rule for evaluating P could be

A $P = 250\,000 \times (1 + 0.14)^n$

B $P = 250\,000 \times 1.14 \times n$

C $P = 250\,000 \times (0.14)^n$

D $P = 250\,000 \times (1 - 0.14)^n$

E $P = 250\,000 \times (1 - 0.14) \times n$

[VCAA 2012 1BRMQ7]

Question 7

A file server costs $30 000.

The file server depreciates by 20% of its value each year. After three years its value is

A $6000 **B** $12 000 **C** $15 360 **D** $19 200 **E** $24 000

[VCAA 2010 1BRMQ5]

Question 8

A new air-conditioning unit was purchased for $5000 on 1 January 2009.

On 1 January of each year after 2009 its value is depreciated by 20% using the reducing balance method. The value of the air conditioner will be below $1500 for the first time on 1 January

A 2012 **B** 2013 **C** 2014 **D** 2015 **E** 2016

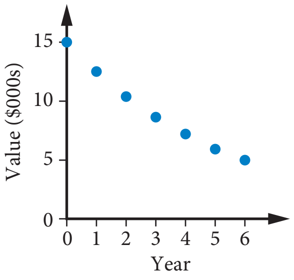
[VCAA 2009 1BRMQ5]

Question 9

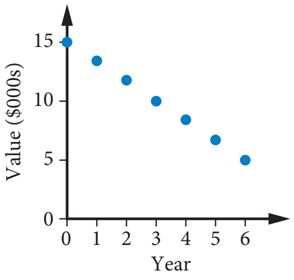
A machine is purchased for $15 000. Using the reducing balance method of depreciation, its future value after six years will be $5000.

The graph that best represents the value of the machine at the end of each year over the six-year period is

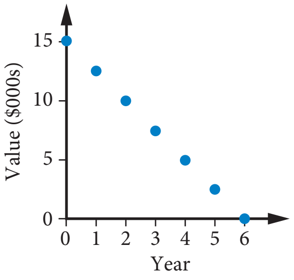
A



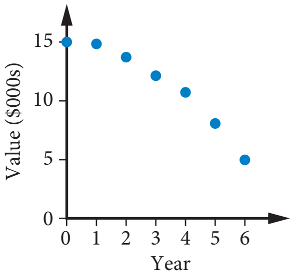
B



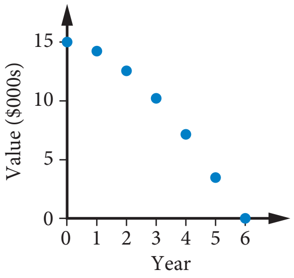
C



D



E



[VCAA 2004 1BRMQ8]

Question 10

A computer originally purchased for $6000 is depreciated each year using the reducing balance method. If the computer is valued at $2000 after four years, then the annual rate of depreciation is closest to

A 17%
B 24%
C 25%
D 28%
E 33%

[VCAA 2008 1BRMQ3]

Question 11

A car is purchased for $25 000. The value of the car is to be depreciated each year by 20% using the reducing balance method.

In the fourth year, the car will depreciate in value by

A $2048
B $2560
C $5000
D $10 240
E $14 760

[VCAA 2013 1BRMQ8]

Question 12

A company purchased a machine for \$60 000. For taxation purposes the machine is depreciated over time. Two methods of depreciation are considered.

a **Flat rate depreciation**

The machine is depreciated at a flat rate of 10% of the purchase price each year.

i By how many dollars will the machine depreciate annually? 1 mark

ii Calculate the value of the machine after three years. 1 mark

iii After how many years will the machine be \$12 000 in value? 1 mark

b **Reducing balance depreciation**

The value, V, of the machine after n years is given by the formula $V = 60\,000 \times (0.85)^n$.

i By what percentage will the machine depreciate annually? 1 mark

ii Calculate the value of the machine after three years. 1 mark

iii At the end of which year will the machine's value first fall below \$12 000? 1 mark

c At the end of which year will the value of the machine **first** be less using flat rate depreciation than it will be using reducing balance depreciation? 2 marks

[VCAA 2006 2BRMQ1]

Question 13

The value of some equipment will be depreciated using the unit cost method.

The initial value of the equipment is \$8360. It will depreciate by 22 cents per hour of use. On average, the equipment will be used for 3800 hours each year.

a Calculate the depreciated value of the equipment after three years. 1 mark

b Show that, in any one year, the flat rate method of depreciation with a depreciation rate of 10% per annum will give the same annual depreciation as the unit cost method. 1 mark

c After how many years will equipment be written off with a depreciated value of \$0? 1 mark

d Suppose the reducing balance method is used to depreciate the equipment instead of the unit cost method.

The initial value of the equipment is \$8360. It will depreciate at a rate of 14% per annum of the reducing balance. Find, correct to the nearest dollar, the depreciated value of the equipment after ten years. 1 mark

[VCAA 2012 2BRMQ2]

Question 14

Brad buys a coffee machine with an initial value of \$3100. He considers two methods of depreciating the value of the coffee machine.

a Suppose the value of the machine is depreciated using the reducing balance method over three years and reducing at a rate of 15% per annum.

What is the depreciated value of the machine after three years? Write your answer correct to the nearest dollar. 2 marks

b Alternatively, suppose that the machine is depreciated using the unit cost depreciation method. Brad sells 15 000 cups of coffee per year and the unit cost per cup is 3.0 cents.

Determine the depreciated value of the machine after three years. Write your answer correct to the nearest dollar. 2 marks

c Brad wants the depreciated value of the machine after three years to be the same when calculated by both methods of depreciation. What would the unit cost per cup have to be for this to occur? Write your answer in cents, correct to 1 decimal place. 2 marks

[VCAA 2003 2BRMQ2]

Question 15

Remy borrows \$650 to buy a digital camera.

Remy uses his camera for work and he wants to depreciate its value over five years. He can either use a flat rate method of depreciation at the rate of 12% per annum or a reducing balance method of depreciation at the rate of 15% per annum.

Which method gives the greater total depreciation over five years? Explain your answer. 1 mark

[VCAA 2004 2BRMQ 1d]

Question 16

For taxation purposes, Stan will depreciate his \$4000 computer over five years. At the end of five years the future value of his computer will be \$1000.

a If Stan uses **flat rate depreciation**, determine the annual depreciation rate. 2 marks

b If Stan uses **reducing balance depreciation**, determine the annual depreciation rate. Write your answer correct to 1 decimal place. 2 marks

[VCAA 2005 2BRMQ2]

Question 17

The books in Khan's office are valued at \$10 000.

a Calculate the value of these books after five years if they are depreciated by 12% per annum using the reducing balance method. Write your answer correct to the nearest dollar. 1 mark

Khan believes his books should be valued at \$4000 after five years.

b Determine the annual reducing balance depreciation rate that will produce this value. Write your answer as a percentage correct to 1 decimal place. 2 marks

[VCAA 2007 2BRMQ4]

Question 18

Michelle purchased a $17 000 car. The car's value depreciates at the rate of 10% per annum using the reducing balance method.

a By what amount, in dollars, does the car's value depreciate during Michelle's third year of ownership? 2 marks

b After how many years of ownership will the car's value first be below $7000? 1 mark

[VCAA 2008 2BRMQ3]

Question 19

The golf club management purchased new lawn mowers for $22 000.

a Use the flat rate depreciation method with a depreciation rate of 12% per annum to find the depreciated value of the lawn mowers after four years. 2 marks

b Use the reducing balance depreciation method with a depreciation rate of 16% per annum to calculate the depreciated value of the lawn mowers after four years. Write your answer in dollars correct to the nearest cent. 1 mark

c After four years, which method, flat rate depreciation or reducing balance depreciation, will give the greater depreciation? Write down the greater depreciation amount in dollars correct to the nearest cent. 1 mark

[VCAA 2009 2BRMQ4]

Question 20

The value of Hugo's bike, purchased for $7500, will be depreciated each year using the reducing balance method of depreciation. One year after it was purchased, this bike was valued at $6375.

Determine the value of the bike five years after it was purchased. Write your answer correct to the nearest dollar. 2 marks

[VCAA 2013 2BRMQ3d]

Question 21

Sally bought her car five years ago for $23 600 and it is now worth $7000. Calculate, correct to 1 decimal place, the percentage annual rate of depreciation of the value of Sally's car over five years

a on a flat rate basis 2 marks

b on a reducing balance basis. 2 marks

[VCAA 2002 2BRMQ3]

Simple interest

6.4

Interest is the cost of using money that isn't yours. When you lend or invest money you earn interest, and when you borrow money you pay interest. **Simple interest** is interest calculated as a percentage of the **principal** (the amount of money invested or borrowed). When borrowing money it can also be referred to as **flat rate interest**.

iStock.com/AleksVF

Simple interest recurrence relations and graphs

The recurrence relation for a simple interest investment is similar to the recurrence relation for flat rate depreciation. The difference is we are adding rather than deducting a fixed amount in each time period.

For example, an amount of $2000 invested in an account earning 6% p.a. simple interest will have 6% of $2000 = $120 added every year.

Let V_n = the value of the investment at the end of nth year.

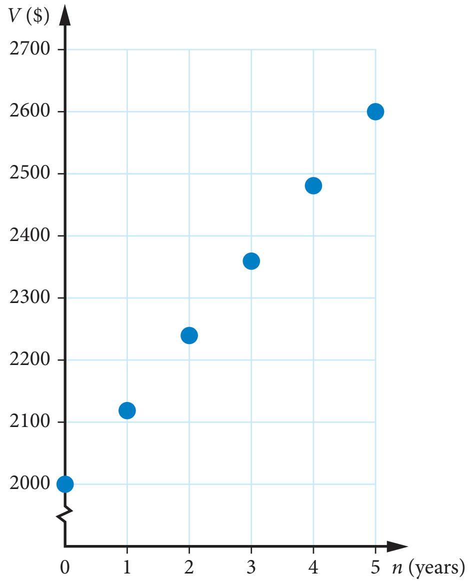
This means the recurrence relation is $V_0 = 2000,\ V_{n+1} = V_n + 120$

A graph is an easy way to show the amount of interest that is earned over a certain period for an investment. Graphs provide a valuable tool for comparing different investments.

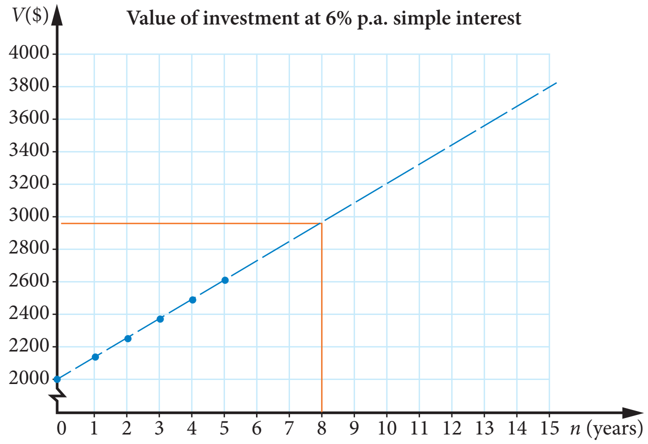
A graph of simple interest against time will always resemble an increasing straight-line graph. As with all the financial graphs we're dealing with, the dots are never to be joined because interest is calculated in discrete periods (e.g. yearly or monthly), not calculated continuously.

For example, $2000 invested in an account earning 6% p.a. simple interest will give the following graph for the amount in the account over time.

Value of investment at 6% p.a. simple interest



The graph can be used to estimate amounts for larger values of n. For example, this extended graph shows the approximate value of the investment after 8 years.



Simple interest general rule

The future value after n time periods, V_n, of a simple interest investment is given by the recurrence relation

V_0 = starting value, $V_{n+1} = V_n + d$

where d = the amount of interest earned per period

$$d = \frac{\text{interest rate}}{100} \times V_0$$

This is the same as flat rate depreciation except we are adding a value each time period, not subtracting it.

The general rule for finding the future value of a simple interest investment after n time periods is

$V_n = V_0 + nd$

nd = the amount of interest earned after n time periods.

Worked example 10

Sasha invests \$4000 in an account earning 5% p.a. simple interest.

a Write a recurrence relation for this investment after n years.

b Write the general rule for finding the future value of the simple interest investment after n years.

c How much interest will she earn in 6 years?

d What amount will her investment grow to in 7 years?

	Working
a **1** Calculate the annual amount of interest earned.	5% of \$4000 $= \frac{5}{100} \times \$4000$ $= \$200$
2 Write the recurrence relation.	Let V_n = the amount in the account at the end of nth year. $V_0 = 4000$ $d = 200$ The recurrence relation is $V_0 = 4000,\ V_{n+1} = V_n + 200$
b Write the general rule.	$V_n = V_0 + nd$ $V_n = 4000 + 200n$
c The amount of interest earned after n time periods is nd.	$n = 6$ $d = 200$ Sasha will earn 6 × 200 = \$1200 interest in 6 years.
d Use the general rule.	$V_n = V_0 + nd$ $n = 7$ $V_7 = 4000 + 200 \times 7$ $= 5400$ Sasha's investment will grow to \$5400 in 7 years.

EXAM PREP 6.4

Simple interest

Prep 1

Write down a recurrence relation for each simple interest investment after n years.

a \$8000 invested at 14% p.a. **b** \$6000 invested at 12% p.a. **c** \$7500 invested at 5% p.a.

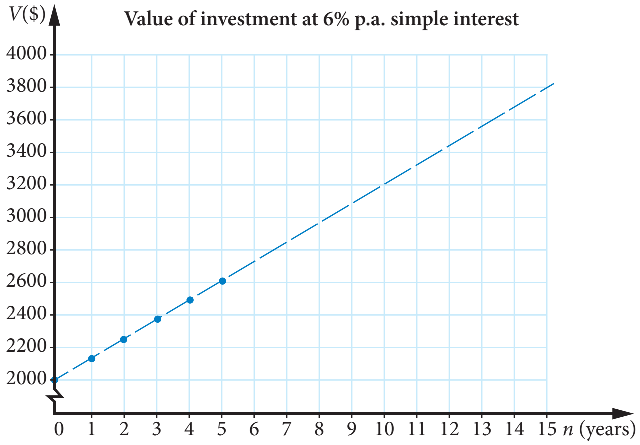
Prep 2

Use the simple interest graph to estimate the value of the investment when \$2000 is invested at 6% p.a. for

a 9 years

b 12 years

c 15 years.



Prep 3

WORKED EXAMPLE 10

Stephen invests \$10 000 at 4% p.a. simple interest.

a Write a recurrence relation for this investment after n years.

b Write the general rule for finding the future value of a simple interest investment after n years.

c How much interest will he earn in 7 years?

d What amount will his investment grow to in 8 years?

EXAM PRACTICE 6.4

Simple interest

Question 1

\$48 000 is invested at a simple interest rate of 4% per annum. Which of the following recurrence relations models the amount after the nth year?

A $A_1 = 48\,000, A_{n+1} = A_n + 192$

B $A_0 = 48\,000, A_{n+1} = A_n - 1920$

C $A_0 = 48\,000, A_{n+1} = A_n + 1920$

D $A_0 = 48\,000, A_{n+1} = A_\mathrm{n} + 19\,200$

E $A_0 = 48\,000, A_{n-1} = A_n + 1920$

Question 2

Which of the following rules could be used to find the future value of $20 000 invested at a simple interest rate of 4% per annum?

A $V_n = 20\,000 + 8000 \times n$

B $V_n = 20\,000 + 8 \times n$

C $V_n = 20\,000 - 800 \times n$

D $V_n = 20\,000 + 800 \times n$

E $V_0 = 20\,000,\ V_{n+1} = V_n + 800$

Question 3

$4000 is invested at a simple interest rate of 5% per annum.

The amount of interest earned in the first year is

A $20 **B** $200 **C** $220 **D** $420 **E** $2000

[VCAA 2006 1BRMQ1]

Question 4

An amount of $800 is invested for two years at a simple interest rate of 4% per annum.

The total amount of interest earned by the investment is

A $32 **B** $64 **C** $160 **D** $320 E $640

[VCAA 2009 1BRMQ1]

Question 5

Sarah invests $37 000 at a simple interest rate of 4% per annum.

The total amount of interest earned in two years is

A $1480 **B** $2960 **C** $5920 **D** $38 480 **E** $39 960

[VCAA 2004 1BRMQ1]

Question 6

$3000 is invested at a simple interest rate of 6.5% per annum.

The total interest earned in three years is

A $195.00 **B** $580.50 **C** $585.00 **D** $3623.85 **E** $3585.00

[VCAA 2012 1BRMQ2]

Question 7

$6000 is invested in an account that earns simple interest at the rate of 3.5% per annum.

The total interest earned in the first four years is

A $70 **B** $84 **C** $210 **D** $840 **E** $885

[VCAA 2010 1BRMQ2]

6.5 Compound interest

Compound interest recurrence relations

Most investments are **compound interest** investments. The interest is calculated at the end of a certain time period. The interest is added onto the principal and then the interest for the next time period is calculated using this new amount. This continues for the full term of the investment.

Consider the table below, which compares $3000 invested at 5% p.a. compound interest compounding yearly to 5% p.a. simple interest after the *n*th year.

n	Compound interest ($)	Value of investment ($)	Simple interest ($)	Value of investment ($)
0	–	3000	–	3000
1	$\frac{5}{100}\times 3000=150$	$3000+150=3150$	$\frac{5}{100}\times 3000=150$	$3000+150=3150$
2	$\frac{5}{100}\times 3150=157.50$	$3150+157.50=3307.50$	$\frac{5}{100}\times 3000=150$	$3150+150=3300$
3	$\frac{5}{100}\times 3307.50=165.38$	$3307.50+165.38=3472.88$	$\frac{5}{100}\times 3000=150$	$3300+150=3450$
4	$\frac{5}{100}\times 3472.88=173.64$	$3472.88+173.64=3646.52$	$\frac{5}{100}\times 3000=150$	$3450+150=3600$

The values in the compound interest part of the table are generated in this way:

New value = old value + 0.05 × old value

Using recurrence relation notation where V_n is the value of the investment at the end of the *n*th year:

$$V_{n+1}=V_n+0.05V_n$$

$$V_{n+1}=(1+0.05)V_n$$

$$V_{n+1}=1.05V_n$$

So the compound interest investment can be generated by the recurrence relation:

$$V_0=3000,\ V_{n+1}=1.05V_n$$

Note that this is similar to the recurrence relation for reducing balance depreciation except the coefficient of V_n is greater than 1. While assets are nearly always depreciated annually, compound interest is often paid in other time periods such as monthly and quarterly.

Amana Images/Klaus Ohlenschläger/dpa

Compound interest general rule

The general rule for compound interest can be found in the following way:

$V_0 = 3000$

$V_1 = 1.05 \times 3000 = 1.05^1 \times 3000$

$V_2 = 1.05 \times (1.05 \times 3000) = 1.05^2 \times 3000$

$V_3 = 1.05 \times (1.05 \times 1.05 \times 3000) = 1.05^3 \times 3000$

…

$V_n = 1.05^n \times 3000$ which can be written as

$V_n = 1.05^n \times V_0$ or

$V_n = 1.05^n V_0$

The future value at the end of the *n*th compounding period, V_n, of a compound interest investment is given by the recurrence relation:

$V_0 =$ starting value, $V_{n+1} = (1+\frac{r}{100}) \times V_n$

where $r =$ the interest rate per compounding period.

The general rule for finding the future value of a compound interest investment after *n* compounding periods is

$V_n = (1+\frac{r}{100})^n \times V_0$

$r =$ the interest rate per compounding period.

Interest earned *in* the *n*th time period $= V_n - V_{n-1}$.

Worked example 11

Alan inherits \$35 000 and decides to invest it in an account where he earns interest of 6% p.a. compounded *monthly*.

a Find r, the percentage interest rate per compounding period.

b Write down a recurrence relation that can be used to describe the value of his investment at the end of each month.

c What is the rule for the future value of the investment?

d Use the rule to find the value of the investment after 7 years to the nearest cent.

		Working
a	Divide the annual interest rate by the number of compounding periods.	6% per annum compounded monthly $= \frac{6}{12}$ % per month $= 0.5\%$ So $r = 0.5$
b	**1** Define the variable.	Let V_n be the value of the investment at the end of the nth month.
	2 Write the recurrence relation.	$V_0 =$ starting value, $V_{n+1} = (1+\frac{r}{100}) \times V_n$ $V_0 = 35\ 000,\ V_{n+1} = (1+\frac{0.5}{100}) \times V_n$ $V_0 = 35\ 000,\ V_{n+1} = 1.005V_n$
c	Write the rule.	$V_n = (1+\frac{r}{100})^n \times V_0$ $V_n = (1+\frac{0.5}{100})^n \times 35\ 000$ $V_n = 1.005^n \times 35\ 000$
d	**1** Calculate how many compounding periods are involved.	7 years = 84 months $n = 84$
	2 Use the rule to find value of the investment. Write the answer in words, rounding to the nearest cent.	$V_{84} = 1.005^{84} \times 35\ 000$ $V_{84} = 53\ 212.937\ 26$ The value of the investment after 7 years is \$53 212.94

Exam hack

Always check how often compounding occurs. When interest is not compounded yearly, the values of n and r need to be adjusted accordingly. For example, if interest is compounding quarterly, n will be the number of quarters (e.g. 6 years = 24 quarters) and r will be the interest rate per quarter (e.g. 5% p.a. compounded quarterly means $r = \frac{5}{4}\%$).

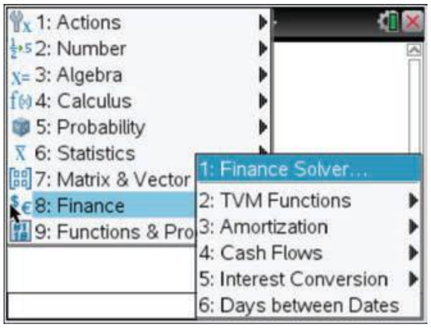
Using CAS Compound interest and finance solvers

Finance solvers can be used to answer questions on compound interest.

TI-NSPIRE CAS

Open a New Document with a Calculator Page.

Press menu, 8: Finance and 1: Finance Solver …



The fields for the Finance Solver are defined as follows.

N is the number of time periods.

I% is the interest rate as a percentage per annum.

PV is the **present value** (when investing this is negative, as you are giving the bank this amount).

Pmt is the value of any payments being made (for compound interest investments where there is only the initial deposit, this is zero).

FV is the future value (this will be positive as this is the money you get back).

PpY is the number of payments per year.

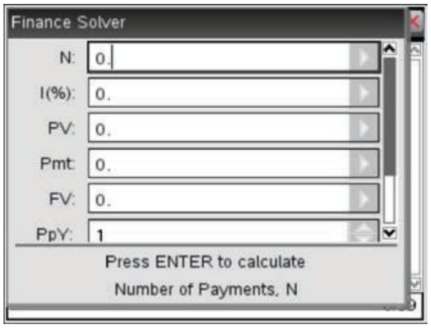
CpY is the number of times in a year interest is compounded.

PpY and **CpY** take the same value for all compound interest calculations.

PmtAt is when the payment is made, at the 'beginning' or 'end' of the time period. Compound interest is paid at the end of the time period so leave PmtAt set to END.

When using the Finance solver, press tab to move between fields when entering data.

Once the data is entered, press tab to move the cursor to the field representing the unknown quantity and press enter.



Finance Solver

N: 0.
I(%): 0.
PV: 0.
Pmt: 0.
FV: 0.
PpY: 1

Press ENTER to calculate
Number of Payments, N

CLASSPAD

Tap Menu and the 2nd page, followed by the Financial application.

Tap **Calc(1)** on the top toolbar followed by **Compound Interest**.

The fields for the Compound Interest application are defined as follows.

N is the number of time periods.

I% is the interest rate as a percentage per annum.

PV is the present value (when investing this is negative, as you are giving the bank this amount).

PMT is the value of any payments being made (for compound interest investments where there is only the initial deposit, this is zero).

FV is the future value (this will be positive as this is the money you get back).

P/Y is the number of payments per year.

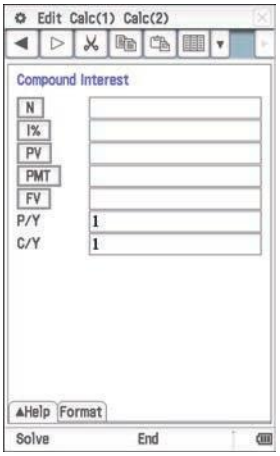
C/Y is the number of times in a year interest is compounded.

P/Y and **C/Y** take the same value for all compound interest calculations.

Compound interest is paid at the end of the time period so make sure that 'End' is showing at the bottom of the screen in the middle.

Tapping Help at the bottom left of the screen gives the definition for a selected field.

Once the data is entered, tap the unknown quantity to find its value.



P/Y 1

C/Y 1

End

Using CAS Calculating future value and compound interest with finance solvers

Carina invested $36 000 in a term deposit earning 8.1% p.a. compounding yearly. Give answers to the nearest cent.

a What is the value of her investment at the end of five years?

b How much interest did she earn?

a **STEP 1**

Write down values of the known quantities.

Interest is compounded yearly, so N is 5 and both PpY (P/Y) and CpY (C/Y) are 1.

The interest rate per annum is 8.1% so I% is 8.1.

$36 000 is being invested so PV is –36 000.

N = 5

I% = 8.1

PV = –36000

PMT = 0

FV = ?

P/Y = 1

C/Y = 1

PmtAt = END

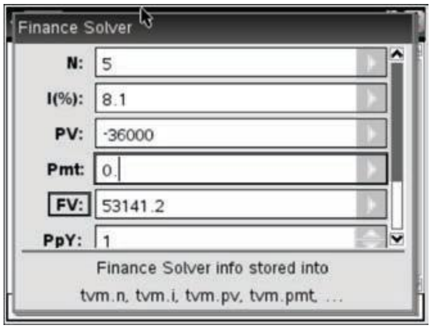
STEP 2

Enter the data in the finance solver or financial application.

STEP 3

TI-NSPIRE CAS

Move the cursor to the data entry line for FV and press enter.



Finance Solver

N: 5

I(%): 8.1

PV: -36000

Pmt: 0

FV: 53141.2

PpY: 1

Finance Solver info stored into tvm.n, tvm.i, tvm.pv, tvm.pmt, ...

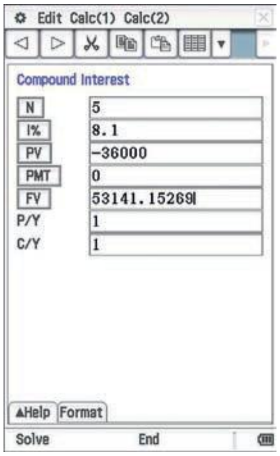
CLASSPAD

Tap Menu and the 2nd page, followed by the Financial application.

Tap **Calc(1)** on the top toolbar followed by **Compound Interest**.

Fill in the fields on the screen using information from the question.

Tap FV.



Compound Interest

N 5

I% 8.1

PV –36000

PMT 0

FV 53141.15269

P/Y 1

C/Y 1

STEP 4

Write the answer.

Carina will have $53 141.15 at the end of five years.

b **STEP 1**

Subtract the principal from the value of the final investment to find the interest earned.

Interest earned
= value of investment after 5 years – the principal
= $53 141.15 – $36 000
= $17 141.15

STEP 2

Write the answer.

Carina earned $17 141.15 interest.

Using CAS Calculating compound interest time with finance solvers

How long will it take for an investment of $12 000 to grow to $24 000 if it is invested at 8% p.a. compounded monthly?

STEP 1

Write down the values of the known quantities.

FV will be positive as that is the amount paid back so FV is 24 000.

The interest rate per annum is 8% so I% is 8.

Interest is compounded monthly so it is paid twelve times in a year. PpY (P/Y) and CpY (C/Y) are both 12.

$12 000 is being invested so PV is –12 000.

N = ?

I% = 8

PV = –12000

PMT = 0

FV = 24000

PpY or P/Y = 12

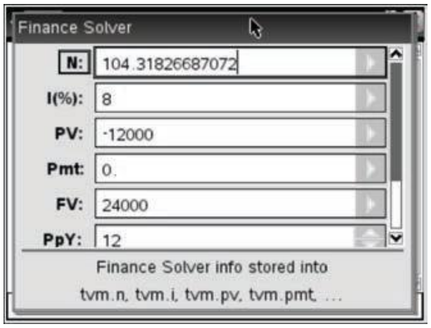
CpY or C/Y = 12

STEP 2

Enter the known values.

TI-NSPIRE CAS

Move the cursor to the data entry line for N and press enter.

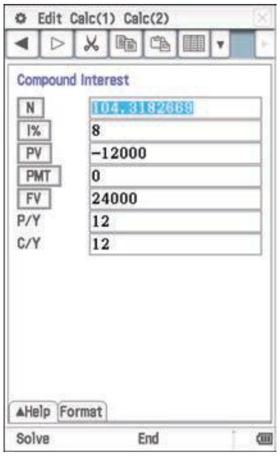


N: 104.31826687072
I(%): 8
PV: -12000
Pmt: 0.
FV: 24000
PpY: 12

Finance Solver info stored into
tvm.n, tvm.i, tvm.pv, tvm.pmt, ...

CLASSPAD

In the **Financial** application, select the **Compound Interest** function as before. Enter the information of the question into the relevant fields. Tap N.



N 104.3182669
I% 8
PV −12000
PMT 0
FV 24000
P/Y 12
C/Y 12

STEP 3

Interest is compounded monthly so N is the number of months.

N = 104.318… months

STEP 4

The value of the investment needs to reach $24 000, so investing for 104 months will not be quite long enough. Therefore, round the answer up to the nearest month.

It will take 105 months or 8 years and 9 months for the value of the investment to reach $24 000.

Compound interest graphs

A CAS/calculator can be used to create compound interest graphs and allow us to compare simple and compound interest investments.

Using CAS Creating interest graphs

Emily has $6000 to invest. She is comparing two different accounts. They offer:

Account 1: simple interest at 6.5% p.a.

Account 2: compound interest at 6% p.a. compounded annually.

a Create a graph that shows the value of each investment over a 10-year period on the same set of axes.

b What would your advice to Emily be?

a **STEP 1**

Define V_n and find the rule.	Let V_n = the value of the investment after n years.
Account 1	6.5% of 6000 = 390
Write the rule.	$d = 390$ $V_n = V_0 + nd$ $V_n = 6000 + 390n$
Account 2	$V_n = (1+\frac{r}{100})^n \times V_0$
Write the rule.	$V_n = 6000 \times 1.06^n$

Construct a graph of each investment over 10 years using a CAS calculator.

TI-NSPIRE CAS

STEP 2

Open a New Document with a Lists & Spreadsheet page.

Give column A the label n.

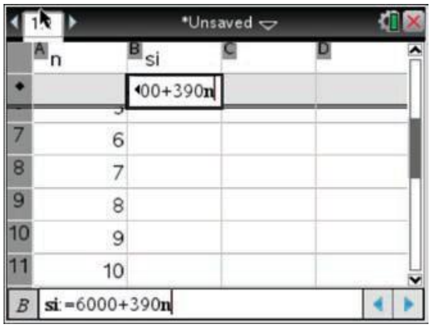
Type in the values 0 to 10 inclusive.

Give column B the label si.

In the shaded formula area, type the rule

= 6000 + 390n

Then press enter.

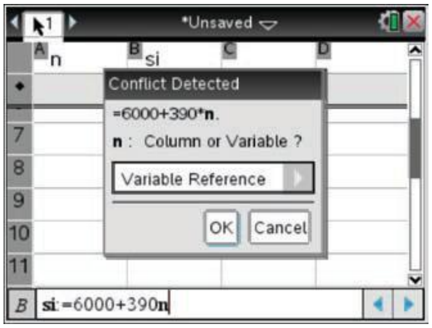


	A n	B si	C	D
◆		◂00+390n		
7	6			
8	7			
9	8			
10	9			
11	10			

B si:=6000+390n

STEP 3

In the pop-up window that appears, select Variable Reference then click OK.



Conflict Detected

=6000+390*n.

n : Column or Variable ?

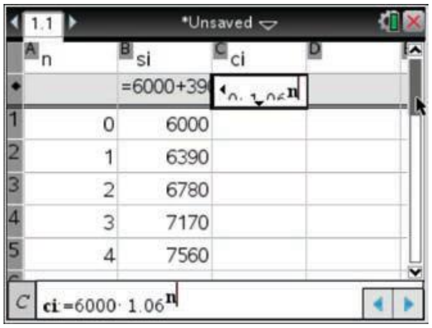
B si:=6000+390n

STEP 4

Give column C the label ci.

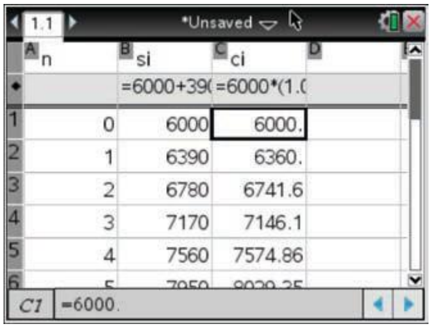
In the shaded formula area type the rule $= 6000 \times 1.06^{n}$

Then press enter.
In the pop-up window that appears select variable reference then click OK.



A n	B si	C ci	D
	=6000+39	1.06^{n}	
0	6000		
1	6390		
2	6780		
3	7170		
4	7560		

C $ci:=6000\cdot 1.06^{n}$

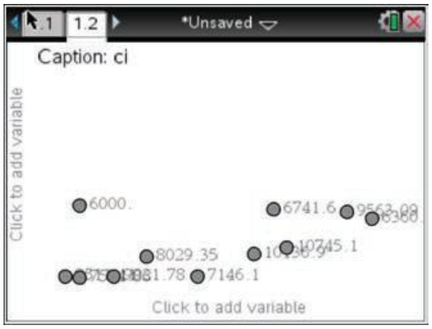


A n	B si	C ci	D
	=6000+39(	=6000*(1.(	
0	6000	6000.	
1	6390	6360.	
2	6780	6741.6	
3	7170	7146.1	
4	7560	7574.86	

C1 =6000.

STEP 5

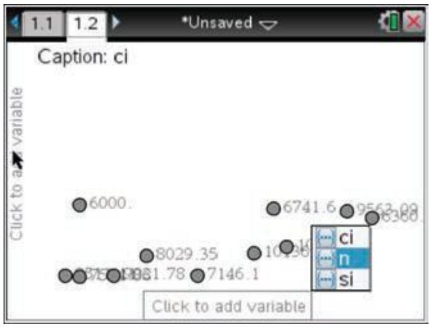
Add a Data & statistics page.



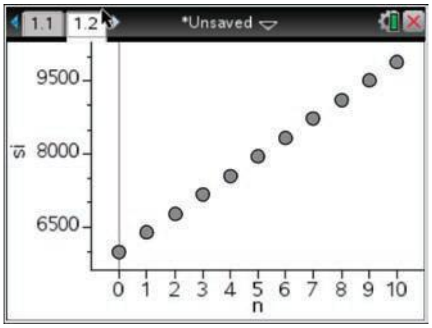
STEP 6

Move the cursor to the ‘Click to add variable’ box at the bottom of the screen and select n.

Move the cursor to the ‘Click to add variable’ box at the left of the screen and select si.



This gives a graph showing the value of the simple interest investment over 10 years.

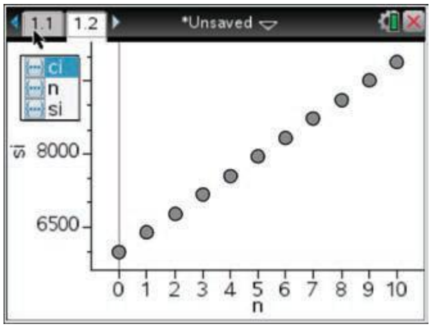


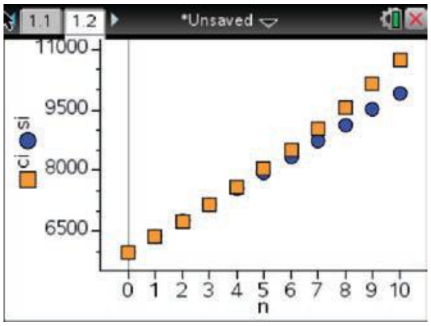
STEP 7

Press menu then 2: Plot properties then 8: Add Y variable.

Select ci.

This graph now displays the value of each investment over 10 years.





CLASSPAD

STEP 2

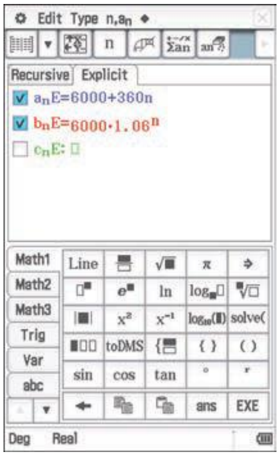
Use the Sequence application.

Tap on the Explicit tab.

Type 6000 + 360n next to anE:

Type 6000 × 1.06^n next to bnE:

Make sure that both anE and bnE are checked.



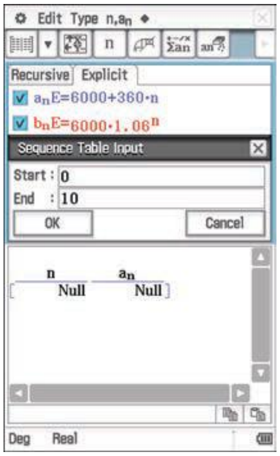
☑ $a_nE=6000+360n$

☑ $b_nE=6000\cdot1.06^n$

☐ c_nE:

STEP 3

Tap and fill in set Start to 0 and end to 10. Tap OK.



☑ $a_nE=6000+360\cdot n$

☑ $b_nE=6000\cdot1.06^n$

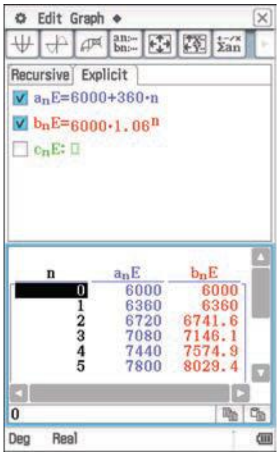
Start : 0

End : 10

n	a_n
Null	Null

STEP 4

Tap ▦ to create a table showing the value of each investment.



☑ $a_nE=6000+360\cdot n$

☑ $b_nE=6000\cdot 1.06^n$

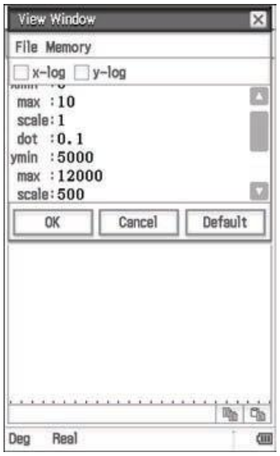
☐ c_nE: □

n	a_nE	b_nE
0	6000	6000
1	6360	6360
2	6720	6741.6
3	7080	7146.1
4	7440	7574.9
5	7800	8029.4

0

STEP 5

Tap ⊞ set xmin: 0 and continue to complete the screen as shown.



☐ x-log ☐ y-log

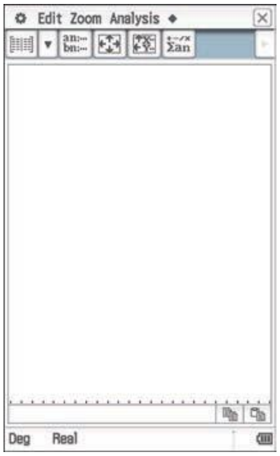
max :10
scale : 1
dot :0.1
ymin :5000
max :12000
scale : 500

This enables the y-axis to be numbered.

STEP 6

Tap ⊞, then Resize ⊟.

You may not see a graph yet.

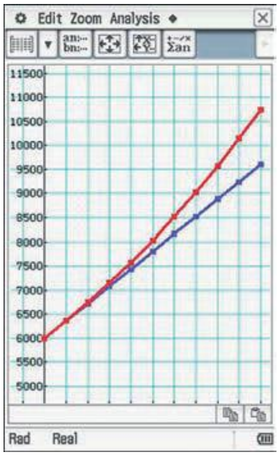


STEP 7

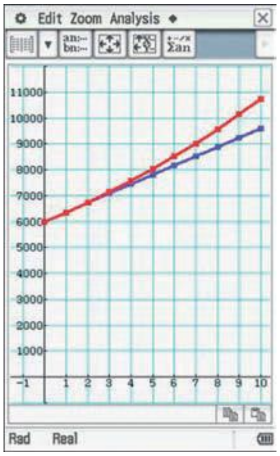
Tap OK.

This graph displays the value of each investment over 10 years. By setting xmin to −2 you get this one, with y-axis labels. No information or accuracy is lost.

It is also possible to drag graphs round and expand and contract them by pinching two fingers, the same as for a smartphone.



The *x*-axis can be labelled by making ymin 1000 and keep the grid by using a scale of 1000 for *y*.



b STEP 8

Looking at the graphs, the simple interest graph lies above the compound interest graph until approximately $n = 4$ years. Then the compound interest graph lies above the simple interest graph.

If Emily is going to invest for less than 4 years she would be better off with Account 1, but if she is going to invest for longer than that Account 2 is the better option.

Effective interest rates

The interest rate quoted for a loan or investment is also known as the **nominal interest rate**. The amount of interest paid for a compound interest investment depends mainly on the nominal interest rate and the term. However, it is also affected by the frequency of the compounding periods. The **effective interest rate**, *e*, is used to compare the interest paid on loans (or investments) with the same nominal annual interest rate, *i*, but with different compounding periods, *c* (daily, monthly, quarterly, annually, other).

As a decimal, the effective interest rate is the interest that would be paid on \$1 invested for 1 year. As percentage, it is the interest that would be paid on \$100 invested for 1 year.

The effective interest rate can be found using:

$$e = \left(1 + \frac{i}{c}\right)^c - 1$$

where

e = effective yearly interest rate as a decimal

i = nominal yearly interest rate as a decimal

c = number of compounding periods in a year

Worked example 12

Liana is looking to invest her money. She has done some research and found the best offers from three different banks.

Bank 1: 8.45 % p.a. compounded daily

Bank 2: 8.6 % p.a. compounded monthly

Bank 3: 8.7 % p.a. compounded six-monthly

Which bank should Liana choose if she wants to earn the most interest?

Working

1 Find the effective rate of interest for each bank.

Bank 1

Write the values of the known variables.

Substitute into the formula.

$e = ?,\ i = 8.45\% = 0.084,\ c = 365$

$$e = \left(1 + \frac{0.0845}{365}\right)^{365} - 1$$

$$= 0.08816\ldots$$

$$\approx 8.82\%\text{ p.a.}$$

Bank 2

Write the values of the known variables.

Substitute into the formula.

$e = ?,\ i = 8.6\% = 0.086,\ c = 12$

$$e = \left(1 + \frac{0.086}{12}\right)^{12} - 1$$

$$= 0.08947\ldots$$

$$\approx 8.95\%\text{ p.a.}$$

Bank 3

Write the values of the known variables.

Substitute into the formula.

$e = ?,\ i = 8.7\% = 0.087,\ c = 2$

$$e = \left(1 + \frac{0.087}{2}\right)^{2} - 1$$

$$= 0.08889\ldots$$

$$\approx 8.89\%\text{ p.a.}$$

2 Compare and write the answer.

Liana should choose Bank 2 as it pays the higher effective rate of interest and will therefore pay more interest.

EXAM PREP 6.5

Compound interest

Prep 1 WORKED EXAMPLE 11

Michelle inherits $55 000 and decides to invest it in an account where she earns interest of 9% p.a. compounded monthly.

a Find *r*, the percentage interest rate per compounding period.

b Write down a recurrence relation that can be used to describe the value of her investment at the end of each month.

c What is the rule for the future value of the investment?

d Use the rule to find the value of the investment after five years, to the nearest cent.

Prep 2 USING CAS: CALCULATING FUTURE VALUE AND COMPOUND INTEREST WITH FINANCIAL SOLVERS

Evan invested $80 000 in an account earning 6.5% p.a. compounded yearly. Give answers to the nearest cent.

a What is the value of his investment at the end of four years?

b How much interest did he earn?

Prep 3 USING CAS: CALCULATING COMPOUND INTEREST TIME WITH FINANCIAL SOLVERS

How long will it take for an investment of $8000 to grow to $14 000 if it is invested at 6.5% p.a. compounded monthly?

Prep 4 USING CAS: CREATING INTEREST GRAPHS

For these two scenarios:

Scenario 1: Simple interest at 5.8% p.a for 10 years

Scenario 2: Compound interest at 7.2% p.a. compounded annually for 10 years

a on the same set of axes graph the values to which an investment of $1000 will grow over a 10-year period

b state which is the better option.

Prep 5 WORKED EXAMPLE 12

Determine the effective interest rate for each of the following investments. Answer as a percentage, correct to 2 decimal places.

a 9% p.a. compounding monthly

b 11% p.a. compounding weekly

c 12% p.a. compounding half-yearly

d 6% p.a. compounding daily

Prep 6 WORKED EXAMPLE 12

Georgio has the choice of three investments at his bank. He can invest at 7.6% p.a. compounded quarterly, 7.5% p.a. compounded monthly or 7.4% p.a. compounded daily.

Calculate the effective rate for each option and decide which is best.

EXAM PRACTICE 6.5

Compound interest

Question 1

The recurrence relation $t_0 = 15\,000,\ t_{n+1} = t_n \times 1.06$ can represent the yearly value of an investment of

A $6000 at 15% p.a. compounded yearly

B $15 000 at 10.6% p.a. compounded yearly

C $10 600 at 15% p.a. compounded yearly

D $15 000 at 6% p.a. compounded yearly

E $15 000 at 1.06% p.a. compounded yearly

Question 2

Which one of the following investments has the highest effective interest rate?

A 5.8 % p.a. compounded monthly

B 5.4 % p.a. compounded daily

C 5.3 % p.a. compounded fortnightly

D 5.6% p.a. compounded six-monthly

E 5.85 % p.a. compounded quarterly

Question 3

The recurrence relation $t_0 = 27\,000,\ t_{n+1} = 1.05t_n$ represents the yearly value of an investment of

A $5000 at 27% p.a. compounded yearly

B $27 000 at 10.5% p.a. compounded yearly

C $10 500 at 27% p.a. compounded yearly

D $27 000 at 5% p.a. compounded yearly

E $27 000 at 1.05% p.a. compounded yearly

Question 4

Which has the lowest effective interest rate?

A 7.9% p.a. compounded quarterly

B 7.7% p.a. compounded fortnightly

C 7.5% p.a. compounded monthly

D 7.45% p.a. compounded daily

E 7.58% p.a. compounded six-monthly

Question 5

$4000 has been invested at 10% p.a. compound interest compounding yearly. What are the two values for $n = 2$ in the compound interest table below where interest is paid at the end of the nth year?

n	Compound interest ($)	Value of investment ($)
0	–	4000
1	400	$4000 + 400 = 4400$
2	?	?

A 440 and 4400

B 400 and 4800

C 400 and 4400

D 400 and 4000

E 440 and 4840

Question 6

$5000 is invested at 6% p.a. compounding yearly. Which of the following is *not* true, given V_n is the future value of the investment after n years?

A The recurrence relation which can be used to model this is $V_0 = 5000$, $V_{n+1} = 1.06V_n$.

B The rule for finding the future value after the nth year is $V_n = 1.06^n V_0$.

C $V_5 = 1.06^5 \times 5000$

D The interest earned in the seventh year is $V_7 - V_6$.

E $V_8 = 1.06 \times V_9$

Question 7

An amount of $22 000 is invested for three years at an interest rate of 3.5% per annum, compounding annually. The value of the investment at the end of three years is closest to

A $2310

B $9433

C $24 040

D $24 392

E $31 433

[VCAA 2011 1BRMQ2]

Question 8

An investment of $16 000 is made at 4% interest per annum, compounding yearly. The value of the investment at the end of the second year is

A $17 280.71

B $17 305.60

C $17 325.71

D $21 120.00

E $21 897.10

[VCAA 2002 1BRMQ2]

Question 9

Ardy invests $150 000 for six years at an interest rate of 3.5% per annum, compounding annually. The value of the investment at the end of the six years is

A $31 500.00

B $34 388 30

C $178 107.00

D $181 500.00

E $184 388.30

[VCAA 2004 1BRMQ2]

Question 10

The following graph represents the growth of an investment over several years.

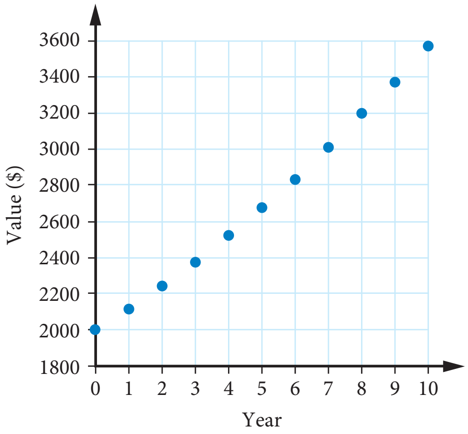
If A dollars is the value of the investment after n years, then a rule for describing the growth of this investment could be

A $A = 2000 \times (1.06)^n$ **B** $A = 2000 \times (0.06)^n$

C $A = 2000 \times 1.06n$ **D** $A = 2000 \times 0.06n$

E $A = 2000 + (1.06)^n$

[VCAA 2002 1BRMQ8]



Question 11

Heather invests \$45 000 at 4% per annum for five years compounding annually.

The total amount of interest earned is

A \$1800 **B** \$2100 **C** \$9000 **D** \$9750 **E** \$54 750

[VCAA 2003 1BRMQ3]

Question 12

Sam and Charlie each invest \$5000 for three years.

Sam's investment earns simple interest at the rate of 7.5% per annum.

Charlie's investment earns interest at the rate of 7.5% per annum compounding annually.

At the conclusion of three years, correct to the nearest cent, Sam will have

A \$86.48 less than Charlie. **B** \$86.48 more than Charlie.

C \$132.23 less than Charlie. **D** \$132.23 more than Charlie.

E the same as Charlie.

[VCAA 2008 1BRMQ6]

Question 13

Tim invests \$3000 in a term deposit account that adds 6.5% interest annually, calculated on the account balance at the end of each year.

The interest paid in the fourth year is

A \$195.00 **B** \$221.16 **C** \$235.55 **D** \$3623.85 **E** \$3859.40

[VCAA 2005 1BRMQ6]

Question 14

\$10 000 is invested at a rate of 10% per annum compounding half-yearly.

The value, in dollars, of this investment after five years is given by

A $10\,000 \times 0.10 \times 5$ **B** $10\,000 \times 0.05 \times 10$ **C** $10\,000 \times 0.05^{10}$

D $10\,000 \times 1.05^{10}$ **E** $10\,000 \times 1.10^5$

[VCAA 2007 1BRMQ6]

Question 15

$15 000 is invested for 12 months.

For the first six months the interest rate is 6.1% per annum compounding monthly.

After six months the interest rate increases to 6.25% per annum compounding monthly.

The total interest is

A $926 **B** $935 **C** $941 **D** $953 **E** $965

[VCAA 2012 1BRMQ8]

Question 16

$10 000 is invested for five years. Interest is earned at a rate of 8% per annum, compounding quarterly. Which one of the following calculations will give the total interest earned, in dollars, by this investment?

A $10\,000 \times 1.02^{5} - 10\,000$

B $10\,000 \times 1.02^{20} - 10\,000$

C $10\,000 \times 1.08^{5} - 10\,000$

D $10\,000 \times 1.08^{20} - 10\,000$

E $10\,000 \times 1.02^{20}$

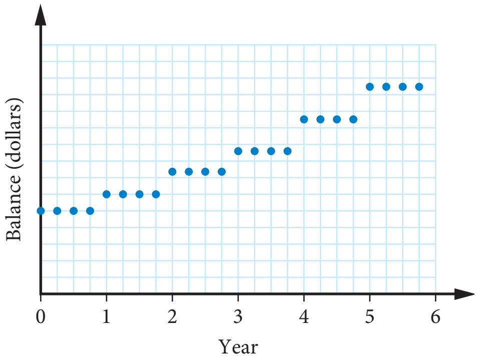
[VCAA 2013 1BRMQ3]

Question 17

The points on the graph show the balance of an investment at the start of each quarter for a period of six years.

The same rate of interest applied for these six years.

In relation to this investment, which one of the following statements is **true**?



A interest is compounding annually and is credited annually

B interest is compounding annually and is credited quarterly

C interest is compounding quarterly and is credited quarterly

D simple interest is paid on the opening balance and is credited annually

E simple interest is paid on the opening balance and is credited quarterly

[VCAA 2006 1BRMQ8]

9780170371100

Question 18

Anthony invested \$15 000 in an account. It earned r% interest per annum, compounding monthly.

The amount of interest that is earned in the third year of the investment is given by

A $15\,000\left(1+\frac{r}{1200}\right)^{3} - 15\,000\left(1+\frac{r}{1200}\right)^{2}$

B $15\,000\left(1+\frac{r}{1200}\right)^{36} - 15\,000\left(1+\frac{r}{1200}\right)^{24}$

C $15\,000\left(1+\frac{r}{100}\right)^{3} - 15\,000\left(1+\frac{r}{100}\right)^{2}$

D $15\,000\left(1+\frac{r}{100}\right)^{36} - 15\,000\left(1+\frac{r}{100}\right)^{24}$

E $15\,000\left(1+\frac{r}{1200}\right)^{4} - 15\,000\left(1+\frac{r}{1200}\right)^{3}$

[VCAA 2011 1BRMQ7]

Question 19

An amount of \$8000 is invested for a period of four years.

The interest rate for this investment is 7.2% per annum compounding quarterly.

The interest earned by the investment in the fourth year (in dollars) is given by

A $4\times\left(\frac{7.2}{100}\times 8000\right)$

B $8000\times 1.018^{4} - 8000\times 1.018^{3}$

C $8000\times 1.018^{16} - 8000\times 1.018^{12}$

D $8000\times 1.072^{4} - 8000\times 1.072^{3}$

E $8000\times 1.072^{16} - 8000\times 1.072^{12}$

[VCAA 2008 1BRMQ9]

Question 20

Lim invested \$8000 in an investment account, earning r% interest per annum, compounding quarterly. The balance in dollars, after five years, is given by

A $8000\left(1+\frac{r}{100}\right)^{5}$

B $8000\left(1+\frac{r}{100}\right)^{20}$

C $8000\left(1+\frac{r}{400}\right)^{5}$

D $8000\left(1+\frac{r}{400}\right)^{20}$

E $8000\left(1+\frac{r}{1200}\right)^{60}$

[VCAA 2003 1BRMQ7]

Question 21

Binnie invests \$12 000 for five years at an interest rate of 3.6% per annum, compounding annually.

The amount of interest she earns during the third year of the investment is closest to

A \$463.66 **B** \$470.41 **C** \$480.36 **D** \$1343.22 **E** \$1823.57

[VCAA 2004 1BRMQ7]

Question 22

Michelle decided to invest some of her money at a higher interest rate. She deposited $3000 in an account paying 8.2% per annum, compounding half-yearly.

a Write down an expression involving the compound interest formula that can be used to find the value of Michelle's $3000 investment at the end of two years. Find this value correct to the nearest cent. 2 marks

b How much interest will the $3000 investment earn over a four-year period? Write your answer correct to the nearest cent. 1 mark

[VCAA 2008 2BRMQ2]

Question 23

The golf club's social committee has $3400 invested in an account which pays interest at the rate of 4.4% per annum compounding quarterly.

a Show that the interest rate per quarter is 1.1%. 2 marks

b Determine the value of the $3400 investment after three years. Write your answer in dollars correct to the nearest cent. 1 mark

c Calculate the interest the $3400 investment will earn over **six** years. Write your answer in dollars correct to the nearest cent. 1 mark

[VCAA 2009 2BRMQ3]

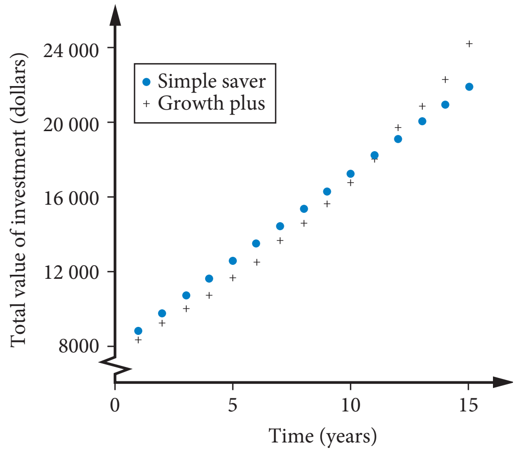
Question 24

Simple Saver is a simple interest investment in which interest is paid annually.

Growth Plus is a compound interest investment in which interest is paid annually.

Initially, $8000 is invested with both Simple Saver and Growth Plus. The graph shows the total value (principal and all interest earned) of each of these investments over a 15-year period.

The increase in the value of each investment over time is due to interest.



a Which investment pays the highest annual interest rate, Growth Plus or Simple Saver? Give a reason to justify your answer. 1 mark

b After 15 years, the total value (principal and all interest earned) of the Simple Saver investment is $21 800. Find the amount of interest paid annually. 1 mark

c After 15 years, the total value (principal and all interest earned) of the Growth Plus investment is $24 000.

- **i** Write down an equation that can be used to find the annual compound interest rate, *r*. 1 mark
- **ii** Determine the annual compound interest rate.

  Write your answer as a percentage correct to 1 decimal place. 1 mark

[VCAA 2010 2BRMQ3]

Depreciation and interest

Practice quiz

First-order recurrence relations

- A **first-order recurrence relation** is a rule that specifies how a particular term in a sequence can be found from the previous term in the same sequence. It consists of
  - a rule linking two consecutive terms (e.g. $t_{n+1} = t_n + 5$)
  - the value of the starting term (e.g. $t_0 = 3$).

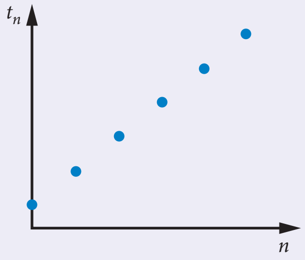
Graphs of first-order recurrence relations

Increasing straight line

Linear growth

- t_n has coefficient of 1
- positive number added

e.g. $t_0 = 22$, $t_{n+1} = t_n + 4$

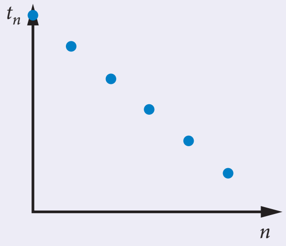


Decreasing straight line

Linear decay

- t_n has coefficient of 1
- positive number subtracted

e.g. $t_0 = 22$, $t_{n+1} = t_n - 4$

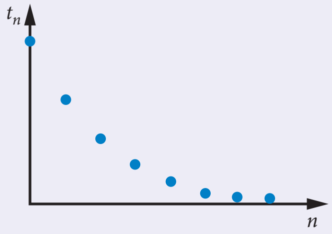


Decreasing curve which never reaches zero

Geometric decay

- t_n has coefficient between 0 and 1
- no numbers added or subtracted

e.g. $t_0 = 10$, $t_{n+1} = 0.75t_n$

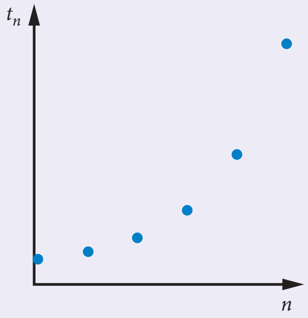


Upward curve

Geometric growth

- t_n has coefficient greater than 1
- no numbers added or subtracted

e.g. $t_0 = 35\,000$, $t_{n+1} = 1.62t_n$

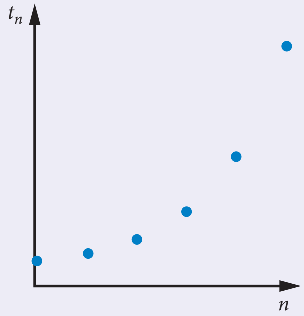


Upward curve

Combined linear growth and geometric growth

- t_n has coefficient greater than 1
- positive number added

e.g. $t_0 = 2$, $t_{n+1} = 2t_n + 5$

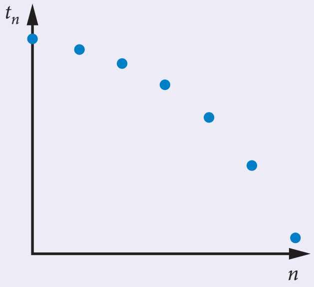


Downward curve

Combined linear decay and geometric growth where the linear decay is greater than the geometric growth

- t_n has coefficient greater than 1
- positive number subtracted

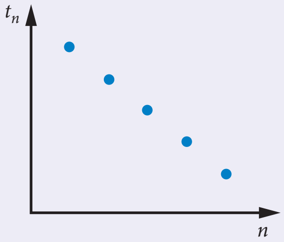
e.g. $t_0 = 15\ 000$, $t_{n+1} = 1.03t_n - 2000$



Depreciation

- **Depreciation** is the decrease in value of assets bought by a business over time.
- **Future value** is the reduced value of an asset at any point in time.
  Future value = Purchase price – Depreciation

Flat rate depreciation



- **Flat rate depreciation** is where the future value of an asset is reduced by a fixed amount every year, expressed either in dollars or as a fixed percentage *of the purchase price*.
- The future value after n time periods, V_n, of an asset depreciated using flat rate depreciation is given by the recurrence relation

  V_0 = starting value, $V_{n+1} = V_n - d$

  where d = the amount depreciated each time period (usually a year)

  $$d = \frac{\text{flat rate of depreciation}}{100} \times V_0$$

  e.g. $V_0 = 9500$, $V_{n+1} = V_n - 1425$.
- The general rule for finding the future value of flat rate depreciation after n time periods is

  $V_n = V_0 - nd$

Unit cost depreciation

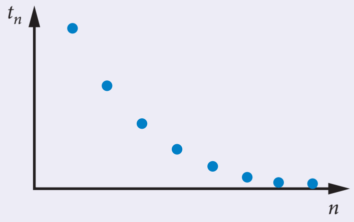
- **Unit cost depreciation** is where the future value of an asset is reduced every year according to the amount of use it has had, not according to its age.
- Depreciation = Rate of depreciation per unit of use × Number of units of use.
- The future value after n time periods, V_n, of an asset depreciated using unit cost depreciation is given by the following recurrence relation *if the usage is the same each time period*

  V_0 = starting value, $V_{n+1} = V_n - d$

  where d is the amount depreciated each time period (usually a year)

  d = Rate of depreciation per unit of use × Number of units of use.
- The general rule for finding the future value of unit cost depreciation after n time periods *if the usage is the same each time period* is $V_n = V_0 - nd$.
- Graphs for unit cost depreciation only look the same as those for flat rate depreciation (i.e. decreasing straight line) if the level of use is the same each year.

Reducing balance depreciation



- **Reducing balance depreciation** is where the future value of an asset is reduced every year by a fixed percentage *of its value in the preceding year*.
- The future value after n time periods, V_n, of an asset depreciated using reducing balance depreciation is given by the recurrence relation:

  V_0 = starting value, $V_{n+1} = (1 - \frac{r}{100}) \times V_n$

  where r is the depreciation rate per time period
  e.g. $V_0 = 10\ 000$, $V_{n+1} = 0.82V_n$.
- The general rule for finding the future value of reducing balance depreciation after n time periods is $V_n = (1 - \frac{r}{100})^n \times V_0$

  where r is the depreciation rate per time period.
- Depreciation *in* the nth time period = $V_{n-1} - V_n$.

Finance solvers for reducing balance depreciation

- **N** is the number of time periods.
- **I%** is the interest rate as a percentage per annum.
- **PV** is the **present value** (for depreciation calculations, this is negative since you have spent money to purchase the asset).
- **Pmt** or **PMT** is the value of any payments being made (for depreciation, this is zero).
- **FV** is the future value (this will be positive since the future value of an asset for depreciation calculations is the value of the asset to you at that time).
- **PpY** or **P/Y** is the number of payments per year (for depreciation calculations this is always 1 since depreciations are done once a year).
- **CpY** or **C/Y** is the number of times in a year interest is compounded (for the problems we work on, this will always be the same as PpY or P/Y, so for depreciation calculations it will always be 1).

Simple interest

- **Simple interest** is interest calculated as a percentage of the **principal** (the amount of money invested or borrowed).

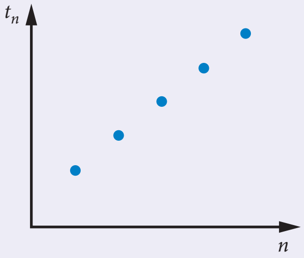
  The future value after n time periods, V_n, of a simple interest investment is given by the recurrence relation

  $V_0 =$ starting value, $V_{n+1} = V_n + d$

  where $d =$ the amount of interest earned per period

  $d = \dfrac{\text{interest rate}}{100} \times V_0$

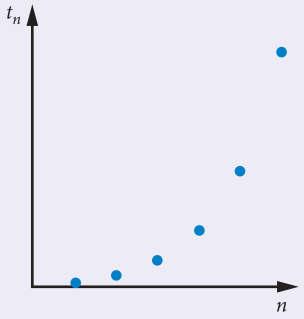
  e.g. $V_0 = 9500$, $V_{n+1} = V_n + 1425$.



- The general rule for finding the future value of a simple interest investment after n time periods is $V_n = V_0 + nd$.

  $nd =$ the amount of interest earned after n time periods.

Compound interest



- **Compound interest** is interest calculated at the end of a certain time period, added onto the principal, and then calculated for the next time period using this new amount.
- The future value at the end of the nth compounding period, V_n, of a compound interest investment is given by the recurrence relation

  $V_0 =$ starting value, $V_{n+1} = (1 + \frac{r}{100}) \times V_n$

  where $r =$ is the interest rate per compounding period.

  e.g. $V_0 = 35\ 000$, $V_{n+1} = 1.062V_n$.
- The general rule for finding the future value of a compound interest investment after n compounding periods is $V_n = (1 + \frac{r}{100})^n \times V_0$

  $r =$ is the interest rate per compounding period.
- Interest earned *in* the nth time period $= V_n - V_{n-1}$.
- When interest is not compounded yearly, the values of n and r need to be adjusted accordingly. For example, if interest is compounding quarterly, n will be the number of quarters

  (e.g. 6 years = 24 quarters) and r will be the interest rate per quarter (e.g. 5% p.a. compounded quarterly means $r = \frac{5}{4}$%).
- The **nominal interest rate** is the interest rate quoted for a loan or investment.
- The **effective interest rate** is the interest rate after the compounding periods have been taken into account.
- The effective interest rate formula is: $e = \left(1 + \frac{i}{c}\right)^c - 1$

  where:

  $e =$ effective yearly interest rate as a decimal

  $i =$ nominal yearly interest rate as a decimal

  $c =$ number of compounding periods in a year.

Finance solvers for compound interest

- **N** is the number of time periods.
- **I%** is the interest rate as a percentage per annum.
- **PV** is the **present value** (when investing, this is negative as you are giving the bank this amount).
- **Pmt** or **PMT** is the value of any payments being made (for compound interest investments where there is only the initial deposit, this is zero).
- **FV** is the future value (this will be positive as this is the money you get back).
- **PpY** or **P/Y** is the number of payments per year.
- **CpY** or **C/Y** is the number of times in a year interest is compounded. (These take the same value for all compound interest calculations as PpY or P/Y.)

CHAPTER 7

REDUCING BALANCE LOANS AND ANNUITIES

Prior learning

7.1 Reducing balance loans

Reducing balance loan recurrence relations

WS Reducing balance loans

A **reducing balance loan** is one where regular payments are made and interest is calculated on the balance still owing after each **repayment** is made. Over the life of the loan:

- the balance still owing reduces with each repayment
- the interest charges reduce
- every repayment includes a combination of interest and principal
- most of the early payments are used to pay the interest
- most of the later payments are used to pay the principal.

Most home loans and car loans are reducing balance loans. The repayment of a debt with a fixed schedule in regular instalments over a period of time is called **amortisation**.

Alamy/Robert Harding Picture Library Ltd

9780170371100

Worked example 1

A housing loan of \$450 000 is taken out for 15 years. Interest is charged at 6% per annum adjusted monthly. The debt is to be repaid in monthly instalments of \$3797.

a Copy and complete the amortisation table below showing the balance remaining after the first 3 months, giving all values to the nearest cent.

Payment number n	Payment made (\$)	Interest paid (\$)	Principal reduction (\$)	Balance remaining (\$)
0	0.00	0.00	0.00	450 000.00
1	3797.00	$\frac{6}{12}\times\frac{1}{100}\times 450\,000.00$ $= 2250.00$	$3797.00 - 2250.00$ $= 1547.00$	$450\,000.00 - 1547.00$ $= 448\,453.00$
2	3797.00	$\frac{6}{12}\times\frac{1}{100}\times 448\,453.00 =$		
3				

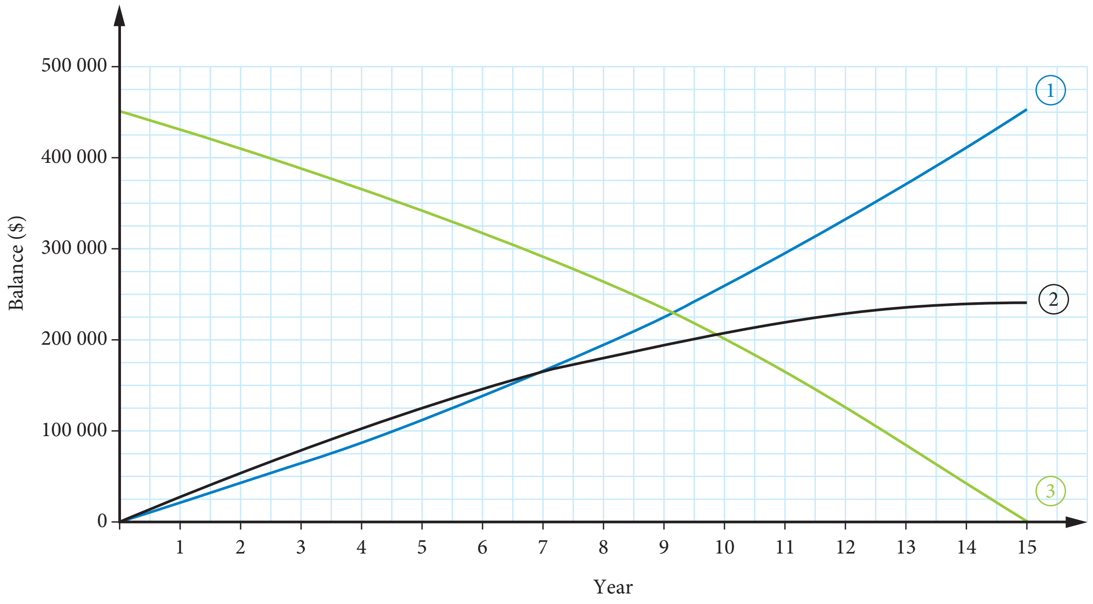
b **i** How many monthly payments need to be made until the loan is paid off?

ii What is the total payment made after the loan has been paid off?

c Write a recurrence relation for the balance remaining on the loan after n payments.

d Use the recurrence relation and a CAS/calculator's recursive computation to find the balance after 5 months.

e The graph for the 15 years of the loan is shown below. Identify which of the three graphs shows the balance of the loan.



f Which of the other two graphs represents the total interest paid on the loan and which represents the total principal paid on the loan?

Working

a Calculate the interest paid using the monthly rate:

$\frac{6}{12}\% = \frac{6}{12} \times \frac{1}{100} = \frac{6}{1200} = 0.005$

on the balance remaining from the previous month.

Calculate the principal reduction by subtracting the interest paid from the payment made each month.

Calculate the balance remaining by subtracting the principal reduction from the previous month's balance.

Give all values to the nearest cent.

Payment number n	Payment made ($)	Interest paid ($)	Principal reduction ($)	Balance remaining ($)
0	0.00	0.00	0.00	450 000.00
1	3797.00	2250.00	1547.00	448 453.00
2	3797.00	2242.27	1554.73	446 898.27
3	3797.00	2234.49	1562.51	445 335.76

b **i** Multiply the number of years by 12.

ii Multiply the number of months by the regular payment.

$15 \times 12 = 180$ monthly payments

total payments $= 180 \times \$3797.00 = \$683\ 460.00$

c Balance = previous balance + interest calculated on previous balance − regular payment

Let B_n = the balance of the loan after n payments.

The recurrence relation is

$B_0 = 450\ 000,\ B_{n+1} = B_n + 0.005B_n - 3797$

which can be written as

$B_0 = 450\ 000,\ B_{n+1} = (1 + 0.005)B_n - 3797$

or

$B_0 = 450\ 000,\ B_{n+1} = 1.005B_n - 3797$

d Use a CAS/calculator's recursive computation to repeat the recurrence relation steps 5 times.

Starting value is 450 000.

Use Ans × 1.005 − 3797 and Enter / EXE five times to get 442 187.26

The balance after 5 months is $442 187.26.

e The balance remaining should be decreasing to 0.

The graph of the balance is number 3.

f Use the fact that in a reducing balance loan most of the early payments are used to pay the interest and most of the later payments are used to pay the principal.

Graph 2 shows the total interest paid and graph 1 shows the total principal paid.

As pointed out in the last chapter, graphs based on recurrence relations should be drawn as points not as a continuous line. However, sometimes, as in the previous worked example, there are so many discrete time periods close together (12 monthly payments) that the graph looks like a continuous line.

The balance on a reducing balance loan, B_n, after the nth repayment is given by the recurrence relation

B_0 = amount borrowed, $B_{n+1} = (1 + \frac{r}{100})\, B_n - D$

where

r is the interest rate per repayment period

D is the regular repayment

$$\begin{aligned}\text{Total interest paid} &= \text{Total repaid} - \text{Amount borrowed}\\ &= \text{Number of payments made} \times \text{Regular payment} - \text{Amount borrowed}\\ &= n \times D - B_0\end{aligned}$$

$$\begin{aligned}\text{Interest after } n \text{ payments} &= \text{Total repaid after } n \text{ time periods} - \text{Reduction in the amount owing on the loan}\\ &= n \times D - (B_0 - B_n)\end{aligned}$$

The graph of a reducing balance loan will always have this shape: 

For the loans and investments in this chapter, the frequency of the regular payments is the same as the compounding period for the interest rate. So, for example, if the interest rate compounds monthly, the regular repayments are monthly. If the interest rate compounds annually, the regular repayments are annual.

Worked example 2

Carlos borrows $40 000 on 15 March 2016 to landscape his garden. The bank lends him this amount with interest calculated monthly at a rate of 9.6% per annum. To repay this loan, Carlos is required to make repayments of $873.69 each month for five years.

a Determine the amount still owing on 15 July 2016.

b How much interest is charged during this period?

		Working
a	**1** Write a recurrence relation for the balance remaining on the loan after n payments.	Let B_n = the balance of the loan after n payments $r = \frac{9.6}{12} = 0.8$ $B_0 = 40\,000$, $B_{n+1} = (1.008)B_n - 873.69$
	2 Determine how many payments are involved.	The first payment is 15 April 2016, so 15 July 2016 will be the fourth payment.

3 Use a CAS/calculator's recursive computation to find the balance after four payments.

Payment number *n*	Balance remaining ($)
0	40 000.00
1	39 446.31
2	38 888.19
3	38 325.61
4	37 758.52

The amount owing on 15 July 2016 is $37 758.52.

b Interest after n payments = Total repaid – Reduction in the amount owing on the loan = $n \times D - (B_0 - B_n)$.

Interest after 4 payments = $4 \times D - (B_0 - B_4)$

$D = \$873.69$
$B_0 = 40\ 000$
$B_4 = 37\ 758.52$

Interest after 4 payments
$= 4 \times \$873.69 - (\$40\ 000 - \$37\ 758.52)$
$= \$3494.76 - \$2241.48$
$= \$1253.28$

The interest charged during this period is $1253.28.

Small differences in answers can occur due to rounding for many questions related to financial modelling with recurrence relations. Because of the compounding effect, these differences tend to increase with each subsequent term of the recurrence relation.

Reducing balance loan amortisation tables

Amortisation tables allow you follow how the payment of the reducing balance loan works at each stage and to answer questions about the loan.

Worked example 3

Karl has taken out a short term personal loan of $20 000 for a gap year trip to Europe. The interest rate is 18% p.a. compounding quarterly and it needs to be repaid by making four quarterly payments of $5570.

Answer these questions using the amortisation table of the loan given below.

a **i** Is the loan paid out exactly after the four payments?

ii What should the last payment be to fully pay out the loan?

b **i** Using the fact that Total interest paid = Total repaid – Amount borrowed, how much total interest does he pay?

ii Explain why your answer to **i** doesn't match the total interest in the table.

c Write a recurrence relation for the balance remaining on the loan after n payments and use it to verify that that Balance remaining column in the amortisation table is correct.

Payment number n	Payment made ($)	Interest paid ($)	Principal reduction ($)	Balance remaining ($)
0	0.00	0.00	0.00	20 000.00
1	5570.00	900.00	4670.00	15 330.00
2	5570.00	689.85	4880.15	10 449.85
3	5570.00	470.24	5099.76	5350.09
4	5570.00	240.75	5329.25	20.84
Total	22 280.00	2300.84	19 979.16	

	Working
a **i** Check the balance remaining after the last payment.	No, the balance remaining after the last payment is $20.84.
ii Add the balance remaining to the last payment.	The last payment should be 5570.00 + 20.84 = $5590.84
b **i** Total interest paid = Total repaid – Amount borrowed = Number of payments made × Regular payment – Amount borrowed	Total interest paid = $n \times D - B_0$ $= 4 \times 5570 - 20\,000$ $= 22\,280 - 20\,000$ = $2280
ii Explain the reason based on the amount added onto the last payment.	The difference between the two figures is the extra amount of $20.84 paid with the last payment. By being paid with the last payment, it is taken out of the interest calculations.
c Find the recurrence relation and use a CAS/calculator's recursive computation to verify the Balance remaining column in the table.	Let B_n = the balance of the loan after n payments. The recurrence relation is $B_0 = 20\,000,\ B_{n+1} = 1.045B_n - 5570$

EXAM PREP 7.1

Reducing balance loans

Prep 1

For each of the following, find, correct to 3 decimal places, the value of k in the recurrence relation A_0 = amount borrowed, $A_{n+1} = kA_n$ − regular repayment that models the loan.

a a loan of $230 000 for 14 years with interest charged at 5% per annum adjusted monthly to be repaid with regular monthly instalments

b a loan of $230 000 for 14 years with interest charged at 5% per annum adjusted annually to be repaid with regular annual instalments

c a loan of $230 000 for 14 years with interest charged at 5% per annum adjusted quarterly to be repaid with regular quarterly instalments

d a loan of $16 000 for 4 years with interest charged at 7.5% per annum adjusted weekly to be repaid with regular weekly payments

e a loan of $14 500 for 3 years with interest charged at 9.5% per annum adjusted fortnightly to be repaid with regular fortnightly payments

Prep 2

WORKED EXAMPLE 1

A housing loan of $380 000 is taken out for 15 years. Interest is charged at 9% per annum adjusted monthly. The debt is to be repaid in monthly instalments of $3854.

a Copy and complete the amortisation table below showing the balance remaining after the first 3 months, giving all values to the nearest cent.

Payment number n	Payment made ($)	Interest paid ($)	Principal reduction ($)	Balance remaining ($)
0	0.00	0.00	0.00	380 000.00
1	3854.00	$\frac{9}{12}\times\frac{1}{100}\times 380\,000.00$ $= 2850.00$	3854.00 − 2850.00 = 1004.00	380 000.00 − 1004.00 = 378 996.00
2	3854.00	$\frac{9}{12}\times\frac{1}{100}\times 378\,996.00 =$		
3				

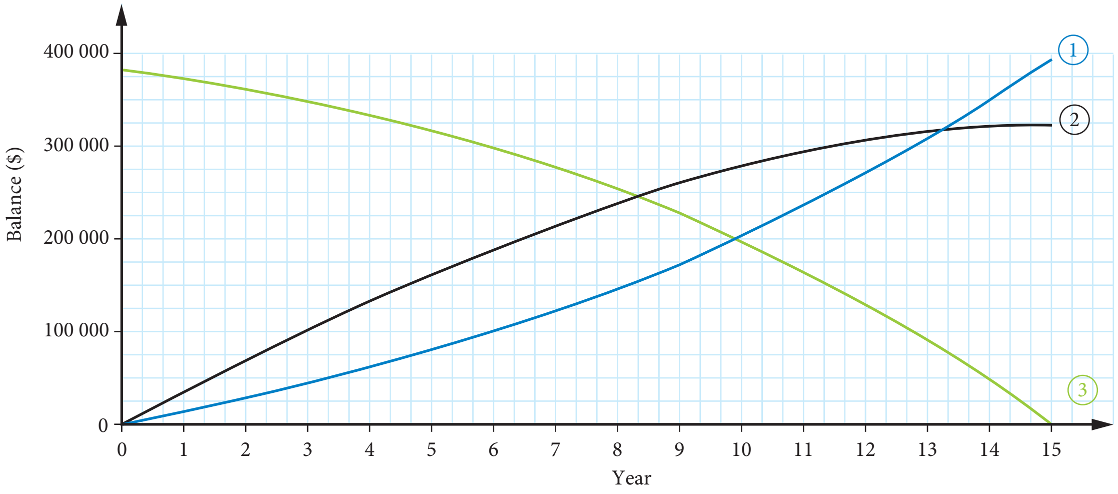
b **i** How many monthly payments need to be made until the loan is paid off?

ii What is the total payment made after the loan has been paid off?

c Write a recurrence relation for the balance remaining on the loan after n payments.

d Use the recurrence relation and a CAS/calculator's recursive computation to find the balance after 5 months.

e The graph for the 15 years of the loan is shown below. Identify which of the three graphs shows the balance of the loan.



f Which of the other two graphs represents the total interest paid on the loan and which represents the total principal paid on the loan?

Prep 3

Paul borrowed $45 000 for a holiday. He makes a monthly repayment on the loan of $850 and interest is calculated at a rate of 11% per annum adjusted monthly.

a Write a recurrence relation that gives the amount still owing on the loan after *n* months.

b How much does Paul still owe after four years?

Prep 4

WORKED EXAMPLE 2

Gaetano borrows $70 000 for a new swimming pool on 1 June 2016. The terms of the loan are that interest is charged at 8% per annum compounded monthly and Gaetano makes regular monthly repayments of $849.30.

a Determine the amount still owing on 1 October 2016.

b How much interest is charged during this period?

Prep 5

WORKED EXAMPLE 3

Chester has taken out a short-term personal loan of $10 000 for a car. The interest rate is 16% p.a. compounding quarterly and it needs to be repaid by making four quarterly payments of $2750.

Answer these questions using the amortisation table of the loan given on page 350.

a **i** Is the loan paid out exactly after the four payments?

ii What should the last payment be to pay out the loan fully?

b **i** Using the fact that Total interest paid = Total repaid – Amount borrowed, how much total interest does he pay?

ii Explain why your answer to **i** doesn't match the total interest in the table.

c Write a recurrence relation for the balance remaining on the loan after n payments and use it to verify that that Balance remaining column in the amortisation table is correct.

Payment number n	Payment made ($)	Interest paid ($)	Principal reduction ($)	Balance remaining ($)
0	0.00	0.00	0.00	10 000.00
1	2 750.00	400.00	2 350.00	7 650.00
2	2 750.00	306.00	2 444.00	5 206.00
3	2 750.00	208.24	2 541.76	2 664.24
4	2 750.00	106.57	2 643.43	20.81
Total	11 000.00	1 020.81	9 979.19	

EXAM PRACTICE 7.1

Reducing balance loans

Question 1

Marina borrows \$200 000 at a rate of 6% per annum calculated monthly on the reducing balance. Each month, Marina repays \$1700.

If A_n is the amount owing after the nth repayment, a recurrence relation that describes the way in which the amount owing on this loan changes is

A $A_0 = 200\ 000,\ A_{n+1} = 0.06A_n - 1700$

B $A_0 = 1700,\ A_{n+1} = 0.06A_n - 200\ 000$

C $A_0 = 200\ 000,\ A_{n+1} = 0.005A_n - 1700$

D $A_0 = 200\ 000,\ A_{n+1} = 1.005A_n - 1700$

E $A_0 = 1700,\ A_{n+1} = 1.005A_n - 200\ 000$

Use the following information to answer Questions 2 & 3.

The recurrence relation $A_0 = 15\ 000,\ A_{n+1} = 1.012A_n - 365$ is used to model the amount owing on a reducing balance loan.

Question 2

What are the first three terms of the recurrence relation?

A 15 000, 15 000, 15 000

B 15 000, 16 344, 17 480

C 15 000, 14 815, 14 627.78

D 15 000, 14 662.42, 13 987.34

E 15 000, 14 635, 14 270

Question 3

Which of the following does the recurrence relation model?

A a reducing balance loan of $15 000 with interest of 1.012% per annum calculated monthly and monthly repayments of $365

B a reducing balance loan of $15 000 with interest of 14% per annum calculated monthly and monthly repayments of $365

C a reducing balance loan of $1012 with interest of 15% per annum calculated monthly and monthly repayments of $365

D a reducing balance loan of $365 with interest of 10% per annum calculated monthly and monthly repayments of $15 000

E a reducing balance loan of $15 000 with interest of 0.012% per annum calculated monthly and monthly repayments of $365

Use the following information to answer Questions 4–6.

A reducing balance loan has the following amortisation table.

Payment number n	Payment made ($)	Interest paid ($)	Principal reduction ($)	Balance remaining ($)
0	0.00	0.00	0.00	12 000.00
1	3 300.00	480.00	2 820.00	9 180.00
2	3 300.00	367.20	2 932.80	6 247.20
3	3 300.00	249.89	3 050.11	3 197.09
4	3 300.00	127.88	3 172.12	24.97
Total	13 200.00	1 224.97	11 975.03	

Question 4

The principal is

A $12 000.00
B $3300.00
C $13 200.00
D $11 975.03
E Not shown in the table

Question 5

Which one of the following is *not* true?

A The regular payments are $3300.00.

B The interest paid reduces with every payment.

C The principal increases with every payment.

D The balance reduces with every payment.

E The balance after the first payment is $9180.00.

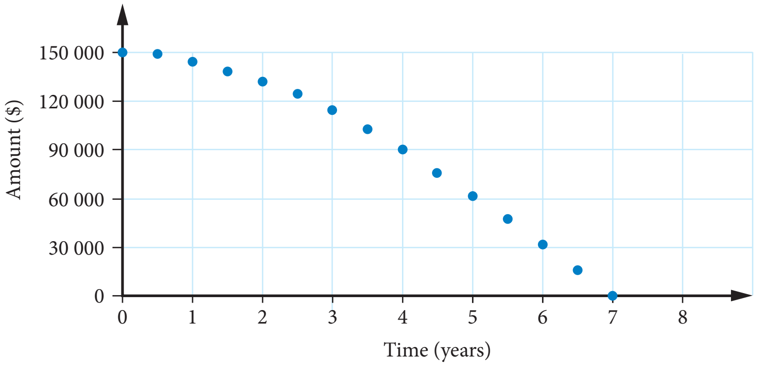
Question 6

After the first three payments have been made, if the loan is to be paid out fully in four payments, what would the last payment need to be?

A $24.97 **B** $3324.97 **C** $13 200.00 **D** $3300.00 **E** $2000.00

Question 7

Which of the following could this graph model?



A the amount of compound interest earned on a 7-year investment

B the amount in a bank account earning compound interest over 7 years

C the value of an asset over 7 years with flat rate depreciation

D the balance owing on a reducing balance loan over 7 years

E the value of an asset over 7 years with reducing balance depreciation

Question 8

Sally planned to repay a loan fully with six equal monthly repayments of $800.

Interest was calculated monthly on the reducing balance.

Sally missed the third payment, but made a **double** payment of $1600 in the fourth month.

Which of the following statements is true?

A The same amount of interest is paid each month.

B The amount owing after three months is the same as the amount owed after two months.

C The amount owing after three months is less than the amount owed after two months.

D To fully repay the loan, Sally will pay less than $4800.

E To fully repay the loan, Sally will pay more than $4800.

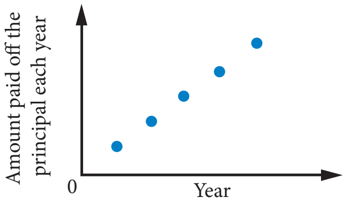
[VCAA 2005 1BRMQ9]

Question 9

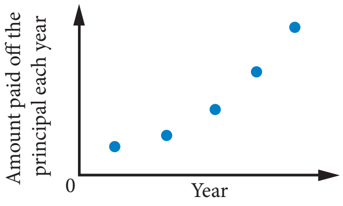
Ernie took out a reducing balance loan to buy a new family home.

He correctly graphed the amount **paid off** the principal of his loan each year for the first five years. The shape of this graph (for the first five years of the loan) is best represented by

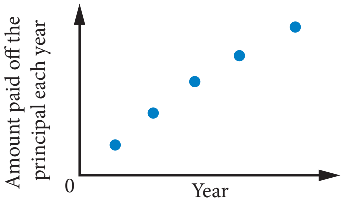
A



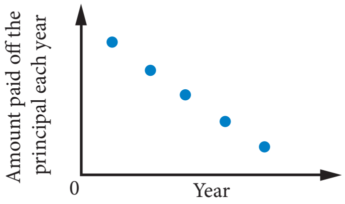
B



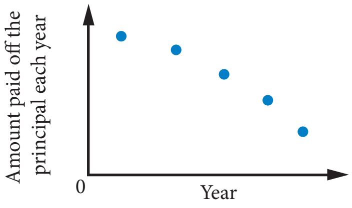
C



D



E



[VCAA 2008 1BRMQ7]

Question 10

Adeline wishes to take out a business loan for $40 000. She has two options available to her.

Option 1: Interest charged at 10% per annum adjusted monthly and an additional management fee of $50 per month.

Option 2: Interest charged at 10.5% per annum adjusted monthly with no additional fees.

Adeline plans to make regular monthly repayments of $800.

a Determine the amount owed using each option after one year.

b Which loan should Adeline choose?

7.2 Reducing balance loans and finance solvers

Loan repayment problems

Finance solvers can be used to answer questions on reducing balance loans.

Calculating reducing balance loans using finance solvers

For a calculation relating to a reducing balance loan, the fields for the Financial Solver or Financial Application (Compound interest) are as follows.

N is the number of time periods.

I% is the interest rate as a percentage per annum.

PV is the present value (for a loan, this is the amount borrowed and is positive since this amount is being given by the bank).

Pmt or **PMT** is the value of the regular payments being made (for a loan, this is negative because this is being paid to the bank).

FV is the future value (this will be 0 if the loan is to be fully repaid, otherwise it is negative, as it is money that needs to be paid to the bank).

PpY or **P/Y** is the number of payments per year.

CpY or **C/Y** is the number of times in a year interest is compounded.

PpY or **P/Y** and **CpY** or **C/Y** usually take the same value for reducing balance loan calculations.

Set **PmtAt** to **End** since reducing balance loan payments are made at the end of the time periods.

Amana Images/Eric Audras/PhotoAlto

Using CAS Calculating repayment amounts, total amounts repaid and total interest

Massimo borrows $140 000 from a building society at 7.5% p.a. compound interest, adjusted monthly. Find:

a the monthly repayment if the loan is to be repaid after 15 years

b the total amount repaid on the loan

c the amount of interest paid on the loan.

---

a **STEP 1**

Use a finance solver to find the monthly repayment.

N is the number of months in 15 years.

I% is the annual interest rate.

PV is positive since the bank has given Massimo the $140 000.

Pmt or PMT is what needs to be found, so leave this field blank.

FV is 0 since the loan will be fully repaid in 15 years.

PpY or P/Y and CpY or C/Y are both 12 since repayments are made monthly and interest compounds monthly.

Set PmtAt to END for TI and End at the bottom of the screen for the ClassPad, as the payments are made and interest calculated at the end of the month.

N = 180
I% = 7.5
PV = 140000
Pmt or PMT =
FV = 0
PpY or P/Y = 12
CpY or C/Y = 12

STEP 2

TI-NSPIRE CAS

Move the cursor into the Pmt field and press enter.

CLASSPAD

Tap FV.

Note that the result is negative since this is money being paid to the bank.

Pmt or PMT = –1297.817304

STEP 3

Write the answer correct to the nearest cent.

Massimo's monthly repayment is $1297.82.

---

b Massimo makes 180 monthly payments of $1297.82.

Amount repaid = 180 × 1297.82
= $233 607.60

---

c The amount of interest paid is the difference between the total amount Massimo repaid and the amount he borrowed.

Interest
= Amount repaid – amount borrowed
= 233 607.60 – 140 000
= $93 607.60

Exam hack

For a reducing balance loan, the amount borrowed (PV) and the payment (PMT) must have different signs when using a finance solver.

Using CAS Calculating the amount owed and the number of repayments

Lily borrows $30 000 at a rate of 8.5% per annum, adjusted monthly. She makes repayments of $650 per month.

a Find how much is still owing on this loan after two years.

b How many repayments are required in order for her to fully repay the loan?

a	**STEP 1** Use a finance solver to find the amount still owing after 2 years. N is the number of months in 2 years. I% is the annual interest rate. PV is positive, since the bank has given Lily the $30 000. Pmt or PMT is –650, since this amount is being given to the bank each month. FV is what needs to be found so leave this blank. PpY or P/Y and CpY or C/Y are both 12 since repayments are made monthly and interest compounds monthly. PmtAt should be END (TI), or make sure End is displayed (ClassPad).	N = 24 I% = 8.5 PV = 30000 Pmt or PMT = –650 FV = PpY or P/Y = 12 CpY or C/Y = 12
	STEP 2 **TI-NSPIRE CAS** Move the cursor into the FV field and press enter. **CLASSPAD** Tap FV. Note that the result is negative, since this is money to be paid back to the bank.	FV = –18 598.558 69
	STEP 3 Write the answer correct to the nearest cent.	After two years, Lily still owes $18 598.56 on this loan.

b **STEP 1** Use a finance solver to find the number of payments required to repay the loan fully. N is what needs to be found so leave this blank. I% is the annual interest rate. PV is positive, since the bank has given Lily the $30 000. Pmt or PMT is −650, since this amount is being given to the bank each month. FV is 0 since the loan will be fully repaid. PpY or P/Y and CpY or C/Y are both 12, since repayments are made monthly and interest compounds monthly. PmtAt should be END (TI), or make sure End is displayed (ClassPad).	N = I% = 8.5 PV = 30000 Pmt or PMT = −650 FV = 0 PpY or P/Y = 12 CpY or C/Y = 12
STEP 2 **TI-NSPIRE CAS** Move the cursor into the N field and press enter. **CLASSPAD** Tap FV.	N = 56.0888…
STEP 3 Round up to the nearest whole number, since she needs to make more than 56 repayments. The last repayment will be less than $650.	It takes 57 payments in order to fully repay the loan.

Using CAS Calculating the annual interest rate

A loan of $200 000 is to be fully repaid over 25 years with quarterly repayments of $3985.14. Find the interest rate per annum correct to 1 decimal place.

STEP 1 N is the number of quarters in 25 years. I% is what needs to be found so leave this field blank. PV is positive since it has been loaned to the borrower. PMT is negative since this is being paid back to the lender. FV is zero since the loan is fully repaid. PpY or P/Y and CpY or C/Y are both 4 since repayments are made quarterly. PmtAt should be END (TI), or make sure End is displayed (ClassPad).	N = 100 I% = PV = 200000 PMT = −3985.14 FV = 0 PpY or P/Y = 4 CpY or C/Y = 4

STEP 2 **TI-NSPIRE CAS** Move the cursor into the I% field and press enter. **CLASSPAD** Tap I%.	I% = 6.2999…
STEP 3 Round to 1 decimal place.	The annual interest rate is 6.3%.

EXAM PREP 7.2

Reducing balance loans and finance solvers

Prep 1

USING CAS: FINDING REPAYMENT AMOUNTS, TOTAL AMOUNTS REPAID AND TOTAL INTEREST

Find the repayment required to repay each of the following loans fully.

- **a** $5000 borrowed at 10% per annum for 6 years with monthly repayments and interest compounded monthly
- **b** $75 000 borrowed at 9.25% per annum for 10 years with fortnightly repayments and interest compounded fortnightly
- **c** $300 000 borrowed at 7.5% per annum for 30 years with monthly repayments and interest compounded monthly
- **d** $35 000 borrowed at 12% per annum for 10 years with quarterly repayments and interest compounded quarterly

Prep 2

USING CAS: FINDING REPAYMENT AMOUNTS, TOTAL AMOUNTS REPAID AND TOTAL INTEREST

Tayla borrows $250 000 from a bank at 8.25% p.a. compound interest, adjusted monthly. Find

- **a** the monthly repayment if the loan is to be repaid after 25 years
- **b** the total amount repaid on the loan
- **c** the amount of interest paid on the loan.

Prep 3 USING CAS: FINDING THE AMOUNT OWED AND THE NUMBER OF REPAYMENTS

Find the number of repayments required to pay off each of the following loans.

a $56 000 borrowed at 9.5% p.a. with monthly repayments of $835 and interest compounded monthly

b $28 000 borrowed at 12.5% p.a. with quarterly repayments of $2250 and interest compounded quarterly

c $260 000 borrowed at 8.25% p.a. with monthly repayments of $2050 and interest compounded monthly

d $45 000 borrowed at 10.25% p.a. with fortnightly repayments of $1031 and interest compounded fortnightly

Prep 4 USING CAS: FINDING THE AMOUNT OWED AND THE NUMBER OF REPAYMENTS

Michael borrows $50 000 at a rate of 8.0% per annum, adjusted monthly. He makes repayments of $620 per month.

a Find how much is still owing on this loan after three years.

b How many repayments are required in order for him to repay the loan fully?

Prep 5 USING CAS: FINDING THE ANNUAL INTEREST RATE

A loan of $250 000 is to be fully repaid over 20 years with quarterly repayments of $5642.56. Find the interest rate per annum correct to 1 decimal place.

EXAM PRACTICE 7.2

Reducing balance loans and finance solvers

Question 1

A loan of $17 500 is to be paid back over four years at an interest rate of 6.25% per annum on a reducing monthly balance.

The monthly repayment, correct to the nearest cent, will be

A $364.58 **B** $413.00 **C** $802.08 **D** $1156.77 **E** $5079.29

[VCAA 2009 1BRMQ7]

Question 2

A loan of $300 000 is to be repaid over a period of 20 years. Interest is charged at the rate of 7.25% per annum compounding quarterly.

The quarterly repayment to the nearest cent is

A $2371.13 **B** $5511.46 **C** $7113.39 **D** $7132.42 **E** $7156.45

[VCAA 2010 1BRMQ7]

Question 3

A loan of $300 000 is taken out to finance a new business venture.

The loan is to be repaid fully over twenty years with quarterly payments of $6727.80.

Interest is calculated quarterly on the reducing balance.

The annual interest rate for this loan is closest to

A 4.1% **B** 6.5% **C** 7.3% **D** 19.5% **E** 26.7%

[VCAA 2008 1BRMQ8]

Question 4

Rho takes a 20-year loan of $172 000 at 6% per annum, compounding monthly and with monthly repayments. To fully repay the loan in 20 years, the amount he must repay each month is

A $716.67 **B** $1216.54 **C** $1232.26 **D** $9058.63 **E** $10 320.00

[VCAA 2002 1BRMQ4]

Question 5

A loan of $250 000 is to be paid back over a period of 20 years at an interest rate of 7.4% per annum, compounding monthly.

To the nearest dollar, the monthly repayment is closest to

A $1963 **B** $1999 **C** $2998 **D** $4343 **E** $13 326

[VCAA 2004 1BRMQ6]

Question 6

Swee borrowed $150 000 at 6.2% per annum compounding monthly.

The repayments are $1100 per month. The balance of the loan at the end of five years is closest to

A $0 **B** $84 000 **C** $127 000 **D** $137 000 **E** $148 000

[VCAA 2003 1BRMQ4]

Question 7

Robin takes out a reducing balance loan of $100 000 with quarterly repayments of $2150.

After seven years of quarterly repayments, Robin still owes $80 000.

Correct to 1 decimal place, the interest rate per annum for this loan is

A 6.3% **B** 8.2% **C** 12.9% **D** 18.9% **E** 24.7%

[VCAA 2009 1BRMQ8]

Question 8

An amount of $130 000 is borrowed at an interest rate of 7.5% per annum, compounding monthly. The loan is fully repaid over ten years with equal monthly repayments.

Which of the following statements is **not** true?

A The monthly interest rate is 0.625%.

B No money will be owed after 10 years.

C The total number of repayments is 120.

D A monthly repayment of $1500 will reduce the length of the loan.

E At the end of five years, the amount of the principal still owing will exceed $65 000.

[VCAA 2004 1BRMQ9]

Question 9

Petra borrowed $250 000 to buy a home. The interest rate is 7% per annum, calculated monthly on the reducing balance over the life of the loan. She will fully repay the loan over 20 years with equal monthly instalments.

The total amount of interest she will pay on the loan is closest to

A $215 000 **B** $266 000 **C** $281 000 **D** $350 000 **E** $465 000

[VCAA 2007 1BRMQ9]

Question 10

An investor borrows $200 000 for five years to buy an apartment.

The interest rate is 8.5% per annum compounding monthly.

It is an interest-only loan; that is, at the end of five years, the investor will still owe $200 000. He is required to make monthly repayments.

Correct to the nearest cent, his monthly repayment will be

A $666.67 **B** $1416.67 **C** $1757.67 **D** $4103.31 **E** $6789.95

[VCAA 2005 1BRMQ5]

Question 11

Sally wants to borrow $20 000 for four years. Interest is calculated quarterly on the reducing balance at an annual rate of 8%.

Sally can afford to repay this loan at $1500 per quarter.

Will this enable her to repay the loan in four years? Explain. 2 marks

[VCAA 2002 2BRM Q4b]

Question 12

Brad sees a coffee machine for sale at Discount King for $3100. The terms of the sale require no deposit and monthly repayments over two years at an interest rate of 9% per annum, calculated monthly on the reducing balance.

The loan is paid out in two years.

a What is the monthly repayment for this loan? Write your answer in dollars, correct to 2 decimal places. 1 mark

b What is the total cost of the machine from Discount King on these terms? Write your answer correct to the nearest dollar. 1 mark

[VCAA 2003 2BRM 1 c ii and iii]

Question 13

A company anticipates that it will need to borrow $20 000 to pay for a machine.

It expects to take out a reducing balance loan with interest calculated monthly at a rate of 10% per annum.

The loan will be fully repaid with 24 equal monthly instalments.

Determine the total amount of interest that will be paid on this loan.

Write your answer to the nearest dollar. 2 marks

[VCAA 2006 2BRM Q4]

Question 14

An area of a club needs to be refurbished.

$40 000 is borrowed at an interest rate of 7.8% per annum.

Interest on the unpaid balance is charged to the loan account monthly.

Suppose the $40 000 loan is to be fully repaid in equal monthly instalments over five years.

a Determine the monthly payment, correct to the nearest cent. 1 mark

b If, instead, the monthly payment was $1000, how many months will it take to repay the $40 000 fully? 1 mark

c Suppose no payments are made on the loan in the first 12 months.

- **i** Write down a calculation that shows that the balance of the loan account after the first 12 months will be $43 234, correct to the nearest dollar. 1 mark
- **ii** After the first 12 months, only the interest on the loan is paid each month. Determine the monthly interest payment, correct to the nearest cent. 1 mark

[VCAA 2012 2BRM Q3]

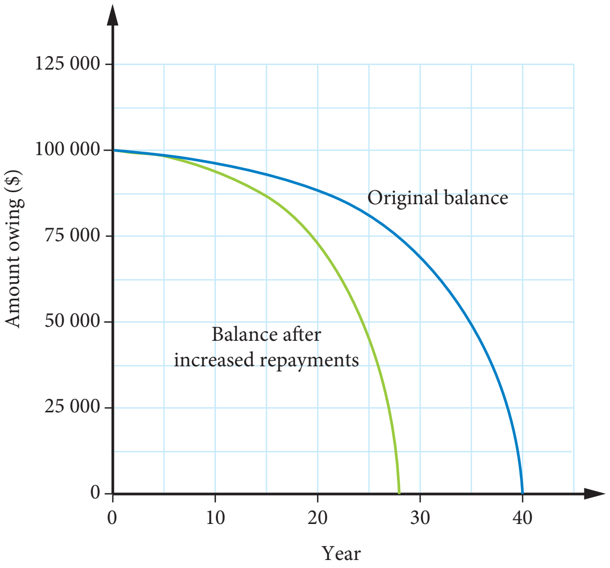
Changing the terms of a loan

7.3

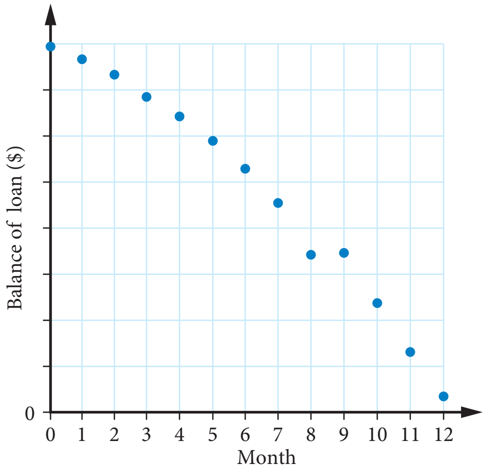
It is common for loans to have flexible terms. Interest rates are usually variable rather than fixed, meaning that they change over the life of the loan. A borrower may choose to change the frequency of their repayments or the size of the repayment. Therefore, the life of a loan can be extended or reduced as required. Any change to a term will also have an effect on the total interest paid and the total cost of the loan.

Graphs of changes in loans

Graphs are often helpful when looking at the effect of changes in terms. 'Original balance' on the graph shows the reducing balance on a 40-year loan of $100 000 at an interest rate of 9% per annum compounding monthly with regular monthly payments of $771. 'Balance after increased repayments' on the graph shows the effect on the same loan when regular payments were increased by $100 each month starting from year 5. The loan ends up being paid off 14 years earlier, and the total interest paid is reduced by $105 156.



The following graph shows the effect of failing to make the regular payment in the ninth month of a 12 month loan and then doubling the next payment to try to make it up. Since interest is still charged during the ninth month, the reducing balance increases. The end result, assuming the regular payments continue, is that the loan isn't fully paid off during the 12 months.



Using CAS Changing the interest rate

Christian borrows $100 000 at 8% per annum compounding monthly, to be repaid over 10 years with repayments of $1213.28 each month. After 5 years, the interest rate is reduced to 7.5% per annum.

a Find the new repayment required for Christian to pay off the loan in the same amount of time.

b How much does Christian save due to the interest rate cut?

---

a **STEP 1**

First find the amount still owing after 5 years.

N is the number of months in 5 years.

I% is the annual interest rate.

PV is positive, since the bank has given Christian $100 000.

Pmt or PMT is negative, since this is money given to the bank.

PpY or P/Y and CpY or C/Y are both 12 since repayments are made monthly and interest compounds monthly.

$N = 5 \times 12 = 60$

I% = 8

PV = 100000

Pmt or PMT = −1213.28

FV =

PpY or P/Y = 12

CpY or C/Y = 12

STEP 2

TI-NSPIRE CAS

Move the cursor into the FV field and press enter.

CLASSPAD

Tap FV.

FV = −59 836.57

After five years, the amount still owing is $59 836.57.

STEP 3

Now find the new repayment.

There are 5 years remaining to repay the loan so N is the number of months in 5 years.

I% is the new annual interest rate.

The amount still owing after the first five years is the present value for the second stage of the loan, so PV = 59 836.57.

The loan needs to be fully repaid, so FV = 0.

PpY or P/Y and CpY or C/Y are both 12 since repayments are made monthly and interest compounds monthly.

$N = 5 \times 12 = 60$

I% = 7.5

PV = 59836.57

Pmt or PMT =

FV = 0

PpY or P/Y = 12

CpY or C/Y = 12

STEP 4 Find the monthly repayment.	Pmt or PMT = −1199.00 So Christian's monthly repayment reduces to \$1199.00 as a result of the interest rate cut.
b The interest rate cut applies to the last five years of the loan, so it is only necessary to find the difference between the total amounts that would have been repaid at the different rates.	Total repaid with old repayment amount $= 60 \times \$1213.28$ $= \$72\,796.80$ Total repaid with new repayment amount $= 60 \times \$1199.00$ $= \$71\,940.00$ Interest saved $= \$72\,796.80 - \$71\,940.00$ $= \$856.80$

Amana Images/Lee White

Using CAS Changing the repayments

Andrew took out a reducing balance loan for $60 000 with interest calculated monthly at a rate of 10.25% per annum. He makes monthly repayments of $918.41.

a How long will it take Andrew to repay the loan?

After 3 years, he increases his monthly repayment to $1200.

b By how many months is the length of the loan reduced?

c Find the amount saved in interest by increasing the size of the monthly repayment.

a	**STEP 1** Find the number of months needed to repay the loan fully. I% is the annual interest rate. PV is positive, since this is money given to Andrew by the bank. Pmt or PMT is negative, since this is money given to the bank by Andrew. FV is 0, since the loan is to be fully repaid. PpY or P/Y and CpY or C/Y are both 12, since repayments are made monthly and interest compounds monthly.	N = I% = 10.25 PV = 60000 Pmt or PMT = −918.41 FV = 0 PpY or P/Y = 12 CpY or C/Y = 12
	STEP 2 Find the number of payments, N. Write the answer.	N = 95.999… It takes Andrew 96 months, or 8 years, to repay the loan.
b	**STEP 1** First find the amount still owing on the loan after 3 years. N is the number of months in 3 years. I% is the annual interest rate. PV is positive, since this is money given to Andrew by the bank. Pmt or PMT is negative, since this is money given to the bank by Andrew. PpY or P/Y and CpY or C/Y are both 12, since repayments are made monthly and interest compounds monthly.	N = 3 × 12 = 36 I% = 10.25 PV = 60000 Pmt or PMT = −918.41 FV = PpY or P/Y = 12 CpY or C/Y = 12

STEP 2 Find the future value, which is the amount still owing after 3 years.	FV = −42 975.73 Andrew still owes \$42 975.73 after 3 years.
STEP 3 Now find the length of time it will take to repay the remainder of the loan with the new repayment amount. I% is the new annual interest rate. The amount still owing after the first three years is the present value for the second stage of the loan, so PV = 42 975.73. Pmt or PMT is −1200, since this is the new repayment amount and it is money paid to the bank. The loan needs to be fully repaid, so FV is 0. PpY or P/Y and CpY or C/Y are both 12, since repayments are made monthly and interest compounds monthly.	N = I% = 10.25 PV = 42975.73 Pmt or PMT = −1200 FV = 0 PpY or P/Y = 12 CpY or C/Y = 12
STEP 4 Find the number of payments, N.	N = 42.93 It takes Andrew 43 months to pay the remainder of the loan.
STEP 5 Find the total number of months required to pay the loan with the new repayment after 3 years.	Total time to repay loan = 3 × 12 + 43 = 79 months
STEP 6 With the original repayment, the loan takes 96 months to repay. Find the difference between the lengths of time taken to pay the loan with the original repayment amounts and the new ones. Write the answer.	Time difference = 96 − 79 = 17 months The length of the loan reduces by 17 months when the repayment amount is increased after 3 years.
c **STEP 1** Find the total amount repaid with the repayments of \$918.41 for 96 months.	Total with original repayments = 96 × \$918.41 = \$88 167.36

Find the total amount repaid with repayments of \$918.41 for 36 months, then \$1200 for 42 months, then the size of the final payment, which will be less than \$1200. Find the amount owing after 42 months. Find the total amount paid.	N = 42. Find FV. FV = −1107.99 The final payment will be \$1107.99. Total with new repayments = 36 × \$918.41 + 42 × \$1200 + \$1107.99 = \$84 570.75
Find the difference in the amounts repaid to determine the amount saved. Write the answer.	Amount saved = \$88 167.36 − \$84 570.75 = \$3596.61 By increasing the size of his regular repayment after 3 years, Andrew saves \$3596.61 in interest.

EXAM PREP 7.3

Changing the terms of a loan

Prep 1 USING CAS: CHANGING THE INTEREST RATE

Tom borrows \$200 000 at 7.5% per annum compounding monthly, to be repaid over 15 years with repayments of \$1854.02 each month. After 5 years, the interest rate is reduced to 7.2% per annum.

a Find the new repayment required in order for Tom to pay off the loan in the same amount of time.

b How much does Tom save due to the interest rate cut?

Prep 2 USING CAS: CHANGING THE REPAYMENTS

Molly took out a reducing balance loan for \$75 000 with interest calculated monthly at a rate of 9.5% per annum. She makes monthly repayments of \$874.78.

a How long will it take for Molly to repay the loan?

After 5 years, she increases her monthly repayment to \$1000.

b By how many months is the length of the loan reduced?

c Find the amount saved in interest by increasing the size of the monthly repayment.

Changing the terms of a loan

Question 1

To purchase a house Sam has borrowed $250 000 at an interest rate of 4.45% per annum, fixed for ten years. Interest is calculated monthly on the reducing balance of the loan. Monthly repayments are set at $1382.50.

After 10 years, Sam renegotiates the conditions for the balance of his loan. The new interest rate will be 4.25% per annum. He will pay $1750 per month.

The total time it will take him to pay out the loan fully is closest to

A 17 years **B** 20 years **C** 21 years **D** 22 years **E** 23 years

[VCAA 2009 1BRMQ9]

Question 2

Ravi has a loan of $135 000 at 7% per annum interest, compounding monthly. The loan is to be repaid monthly over 20 years. The scheduled repayments are $1046.65 per month. However, he finds that he can afford to pay $1200 per month and decides to do so for the duration of the loan.

The amount of time this will save in paying off the loan is closest to

A 6 months **B** 1 year **C** 5 years **D** 10 years **E** 15 years

[VCAA 2002 1BRMQ7]

Question 3

Xavier borrows $45 000 from the bank to buy a car.

He is offered a reducing balance loan for three years with an interest rate of 9.75% per annum, compounding monthly.

He can repay this loan by making 36 equal monthly payments.

Instead, Xavier decides to repay the loan in 18 equal monthly payments.

If there are no penalties for repaying the loan early, the amount he will save is closest to

A $2697 **B** $3530 **C** $3553 **D** $6581 **E** $7083

[VCAA 2011 1BRMQ9]

Question 4

Teresa borrowed \$120 000 at an interest rate of 7.67% per annum, compounding monthly.

The loan is to be repaid with equal monthly payments.

She decides to repay the loan by making monthly payments of \$1430.

Which of the following statements is **true**?

A She will pay out the loan fully in less than ten years.

B The amount of interest that she pays on the loan will increase each year.

C After four years the amount that she owes on the loan will be less than \$80 000.

D Every monthly payment that she makes reduces the amount that she owes on the loan by the same amount.

E Monthly payments of \$1560 (instead of \$1430) will enable her to repay this loan in less than nine years.

[VCAA 2011 1BRMQ8]

Question 5

Peter borrows \$80 000 for ten years at 5.6% per annum, compounding monthly, with monthly repayments of \$555.

Which one of the following statements is **true**?

A The loan will be fully paid out in ten years.

B At the end of five years, the balance of the loan will be \$40 000.

C The amount of interest paid each month during the loan increases.

D Weekly repayments of \$132 compounding weekly would reduce the period of the loan.

E If one extra payment of \$2000 is to be made, it would be better to make it at the end of year eight than at the end of year two.

[VCAA 2003 1BRMQ9]

Question 6

The following information relates to the repayment of a home loan of \$300 000.

- The loan is to be repaid fully with monthly payments of \$2500.
- Interest compounds monthly.
- After the first monthly payment has been made, the amount owing on the loan is \$299 000.

Which one of the following statements is true?

A After two months, \$297 995 is still owing on the loan.

B \$1000 of interest has been paid in the first month.

C The loan will be fully repaid in less than 15 years.

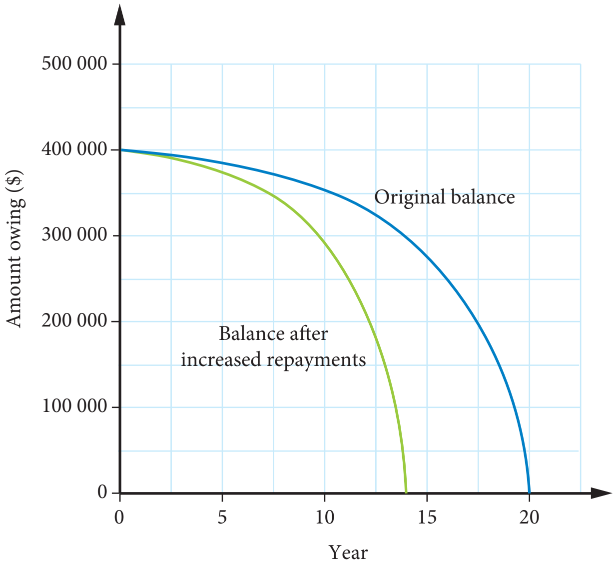
D Halfway through the term of the loan, the amount still owing will be \$150 000.

E Payments of \$2750 rather than \$2500 per month will reduce the time to repay the loan fully by more than three years.

[VCAA 2013 1BQ9]

Question 7

The following graph shows the balance of an original loan and the revised balance of the same loan where the regular monthly payments were increased from the second year. Which of the following is **not** true?



A The amount borrowed was $400 000.

B The original loan would have taken 20 years to pay back.

C The revised loan will take over 12 years to pay back.

D The revised loan will be fully repaid exactly 5 years earlier than the original loan.

E The balance on the revised loan reduces more rapidly than on the original loan.

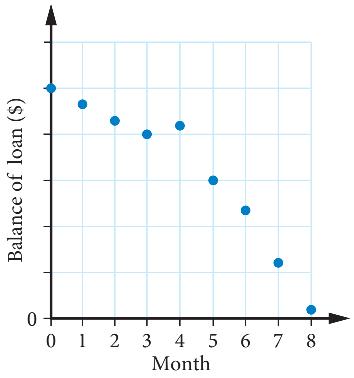
Question 8

Peter took out a reducing balance loan where interest was calculated monthly. He planned to repay this loan fully, with eight equal monthly payments of $260.

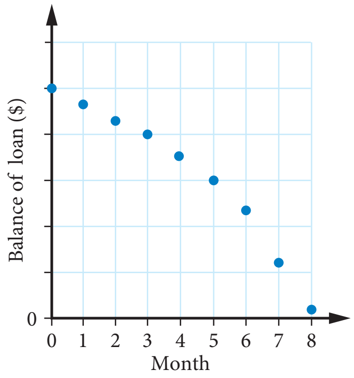
Peter missed the fourth payment, but made a **double** payment of $520 in the fifth month. He then continued to make payments of $260 for the remaining three months.

Which graph could show the balance of the loan each month over the eight-month period?

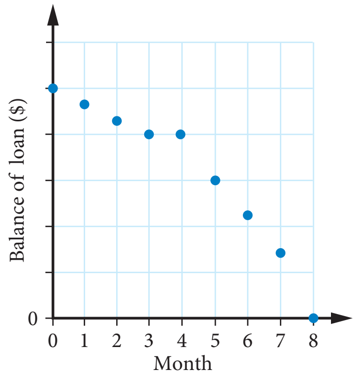
A



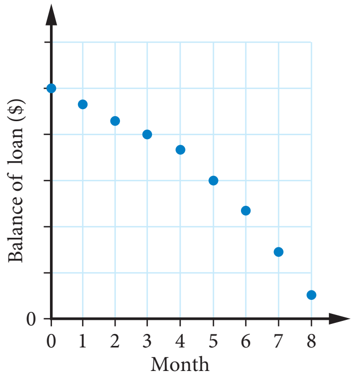
B



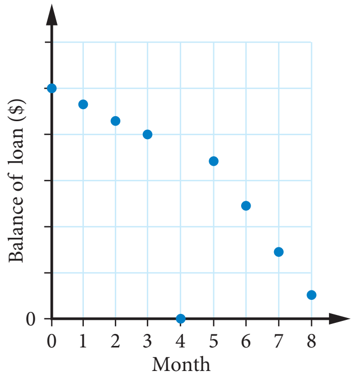
C



D



E



[VCAA 2012 1BRMQ9]

Question 9

Anna borrows \$12 000 at 7.5% interest per annum, compounding monthly. The loan is to be fully repaid over four years by equal monthly repayments.

a Determine the monthly repayment for this loan. Write your answer correct to the nearest cent. 1 mark

b Determine the total amount of interest paid on the loan after four years. 1 mark

c After six equal repayments have been made, how much has Anna paid off the loan? Write your answer correct to the nearest dollar. 1 mark

d At the end of six months the interest rate increases to 8.0% per annum. Anna still has to pay out the balance of the loan completely within the original period of the loan. Determine the new monthly repayments that now apply. Write your answer correct to the nearest dollar. 1 mark

[VCAA 2004 2BRM Q2 b c d eii]

Question 10

Khan decides to extend his home office and borrows \$30 000 for building costs. Interest is charged on the loan at a rate of 9% per annum compounding monthly.

Assume Khan will pay only the interest on the loan at the end of each month.

a Calculate the amount of interest he will pay each month. 1 mark

b Suppose the interest rate remains at 9% per annum compounding monthly and Khan pays \$400 each month for five years.

Determine the amount of the loan that is outstanding at the end of five years.

Write your answer correct to the nearest dollar. 1 mark

c Khan decides to repay the \$30 000 loan fully in equal monthly instalments over five years.

The interest rate is 9% per annum compounding monthly.

Determine the amount of each monthly instalment. Write your answer correct to the nearest cent. 1 mark

[VCAA 2007 2BRM Q2]

Question 11

A home buyer takes out a reducing balance loan of $250 000 to purchase an apartment.

Interest on the loan will be calculated and paid monthly at the rate of 6.25% per annum.

a The loan will be fully repaid in equal monthly instalments over 20 years.

- **i** Find the monthly repayment, in dollars, correct to the nearest cent. 1 mark
- **ii** Calculate the total interest that will be paid over the 20-year term of the loan. Write your answer correct to the nearest dollar. 2 marks

b After 60 monthly repayments have been made, what will be the outstanding principal on the loan?

Write your answer correct to the nearest dollar. 1 mark

By making a lump sum payment after nine years, the home buyer is able to reduce the principal on his loan to $100 000. At this time, his monthly repayment changes to $1250. The interest rate remains at 6.25% per annum, compounding monthly.

c With these changes, how many months, in total, will it take the home buyer to fully repay the $250 000 loan? 1 mark

[VCAA 2010 2BRM Q4]

Question 12

Tania takes out a reducing balance loan of $265 000 to pay for her house.

Her monthly repayments will be $1980.

Interest on the loan will be calculated and paid monthly at the rate of 7.62% per annum.

a **i** How many monthly repayments are required to repay the loan?

Write your answer to the nearest month. 1 mark

ii Determine the amount that is paid off the principal of this loan in the first year.

Write your answer to the nearest cent. 1 mark

Immediately after Tania made her twelfth payment, the interest rate on her loan increased to 8.2% per annum, compounding monthly.

Tania decided to increase her monthly repayment so that the loan would be repaid in a further nineteen years.

b Determine the new monthly repayment. Write your answer to the nearest cent. 1 mark

[VCAA 2011 2BRM Q4]

7.4 Annuities

WS Annuity problems

Annuity recurrence relations

An **annuity** is a type of investment. A sum is invested and interest is compounded at some fixed rate. Withdrawals are made at regular intervals, usually until the value of the investment is \$0. Annuities are often used by people in their retirement who invest a large sum so that they can receive a regular set amount of income.

iStock.com/shironosov

Worked example 4

A retired couple has \$450 000 invested at 6% per annum adjusted monthly and they withdraw \$3797 every month.

a Copy and complete the step-by-step table below showing the value of the investment for the first 3 months, giving all values to the nearest cent.

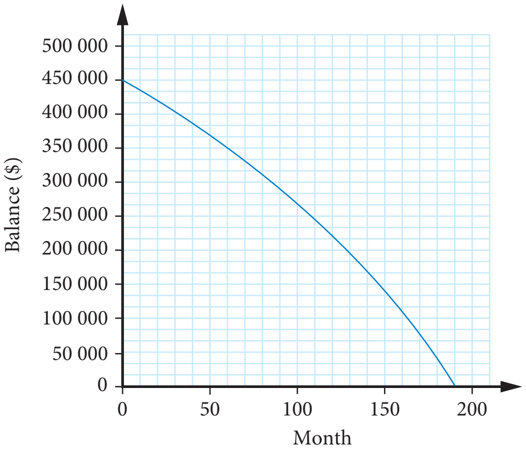
Withdrawal number n	Withdrawal made (\$)	Interest earned (\$)	Principal reduction (\$)	Value of investment (\$)
0	0.00	0.00	0.00	450 000.00
1	3797.00	$\frac{6}{12}\times\frac{1}{100}\times 450\ 000.00$ $= 2250.00$	3797.00 – 2250.00 = 1547.00	450 000.00 – 1547.00 = 448 453.00
2	3797.00	$\frac{6}{12}\times\frac{1}{100}\times 448\ 453.00 =$		
3				

9780170371100

b Write a recurrence relation for the value of the investment after *n* withdrawals.

c Use the recurrence relation and a CAS/calculator's recursive computation to find the value of the investment after 5 months.

d From the following graph of this annuity, how many years will it take for the couple to run out of money?



e Compare this example to the reducing balance loan example in Worked example 1. What do you notice?

Working

a Calculate the interest earned using the monthly rate:

$$\frac{6}{12}\% = \frac{6}{12} \times \frac{1}{100} = \frac{6}{1200} = 0.005$$

on the value of the investment.

Calculate the principal reduction by subtracting the interest from the withdrawal.

Calculate the value of the investment by subtracting the principal reduction from the previous investment.

Withdrawal number *n*	Withdrawal made ($)	Interest earned ($)	Principal reduction ($)	Value of investment ($)
0	0.00	0.00	0.00	450 000.00
1	3797.00	2250.00	1547.00	448 453.00
2	3797.00	2242.27	1554.73	446 898.27
3	3797.00	2234.49	1562.51	445 335.76

Give all values to the nearest cent.

b	Value of the investment = previous investment value + interest calculated on previous investment value – regular withdrawal	Let V_n = the value of the investment after n withdrawals. The recurrence relation is $V_0 = 450\,000,\ V_{n+1} = V_n + 0.005V_n - 3797$ which can be written as $V_0 = 450\,000,\ V_{n+1} = (1 + 0.005)V_n - 3797$ or $V_0 = 450\,000,\ V_{n+1} = 1.005V_n - 3797$
c	Use a CAS/calculator's recursive computation to repeat the recurrence relation steps 5 times.	Starting value is 450 000. Use Ans × 1.005 – 3797 and Enter / EXE five times. The value of the investment after 5 months is \$442 187.26.
d	Read from the graph.	The couple will run out of money after 180 months, which is 15 years.
e	Compare the figures.	The amounts involved are the same. The mathematics behind the two examples is identical, although the two situations in real life are opposites. The first example involves someone paying money into an account, while this example involves someone withdrawing from an account.

The value of an annuity, V_n, after n withdrawals is given by the recurrence relation

$$V_0 = \text{principal},\ V_{n+1} = \left(1 + \frac{r}{100}\right) V_n - D$$

where
r is the interest rate per withdrawal period
D is the regular withdrawal

$$\begin{aligned}\text{Total interest earned} &= \text{Total withdrawals} - \text{Principal}\\ &= \text{Number of withdrawals made} \times \text{Regular withdrawal} - \text{Principal}\\ &= n \times D - B_0\end{aligned}$$

Interest earned after n withdrawals
= Total withdrawals after n time periods – Reduction in the value of the annuity
$= n \times D - (V_0 - V_n)$

Although annuities are investments and reducing balance loans are loans, they are based on the same recurrence relation.

The graph of an annuity will always have the same shape as a reducing balance loan:

Annuities and finance solvers

Annuities can be investigated and analysed using finance solvers in much the same way that reducing balance loans can.

For a calculation relating to an annuity, the fields for the Financial Solver or Financial Application (Compound Interest) are as follows.

N is the number of time periods.

I% is the interest rate as a percentage per annum.

PV is the present value (for an annuity, this is the amount invested and is negative since this amount is being given to the bank).

Pmt or **PMT** is the value of the regular payments being made (for an annuity, this is positive because this is being paid by the bank).

FV is the future value (this is 0 at the end of the annuity. Otherwise, it is positive as it is money that will be given back by the bank at the end of the investment period).

PpY or **P/Y** is the number of payments per year.

CpY or **C/Y** is the number of times in a year interest is compounded.

PpY or **P/Y** and **CpY** or **C/Y** usually take the same value for annuities calculations.

Using CAS Calculating how long an annuity will last

Simeon purchases a $400 000 annuity. Interest is paid at 8% per annum compounded monthly. If he receives monthly payments of $3500, how long will the annuity last? Answer correct to the nearest month.

STEP 1

Use a finance solver to find how long the annuity will last.	N =
I% is the annual interest rate.	I% = 8
PV is negative, since Simeon has effectively given the bank the $400 000.	PV = −400000
	Pmt or PMT = 3500
Pmt or PMT is positive since this amount is being given by the bank each month.	FV = 0
FV is 0, since the annuity will last until it no longer has any value.	PpY or P/Y = 12
	CpY or C/Y = 12
PpY or P/Y and CpY or C/Y are both 12, since payments are made monthly and interest compounds monthly.	

STEP 2 **TI-NSPIRE CAS** Move the cursor into the N field and press enter. **CLASSPAD** Tap FV.	$N = 215.979\ldots$
STEP 3 Write the answer correct to the nearest month.	Simeon's annuity will last for 216 months.

Using CAS Calculating how much to withdraw from an annuity

Christina invests $350 000 with interest paid at 7.5% per annum compounded monthly. She receives monthly payments from this investment. What monthly payment will she receive if she wishes to receive payments for 15 years?

STEP 1 Use a finance solver to find the monthly payment. N is the number of months in 15 years. I% is the annual interest rate. PV is negative, since Christina has given the bank the $350 000. Pmt or PMT is positive, since this amount is being given by the bank each month. FV is 0, since the annuity is to last for 15 years. PpY or P/Y and CpY or C/Y are both 12, since payments are made monthly and interest compounds monthly.	$N = 15 \times 12 = 180$ I% = 7.5 PV = −350000 Pmt or PMT = FV = 0 PpY or P/Y = 12 CpY or C/Y = 12
STEP 2 **TI-NSPIRE CAS** Move the cursor into the Pmt field and press enter. **CLASSPAD** Tap FV.	Pmt or PMT = 3244.54
STEP 3 Write the answer.	Christina will receive monthly payments of $3244.54 for 15 years.

Exam hack

For an annuity where regular payments are made to the investor, PV and Pmt or PMT must have opposite signs.

EXAM PREP 7.4

Annuities

Prep 1

WORKED EXAMPLE 4

A retired couple has \$380 000 invested at 9% per annum adjusted monthly and they withdraw \$3854 every month.

a Copy and complete the step-by-step table below showing the value of the investment for the first 3 months, giving all values to the nearest cent.

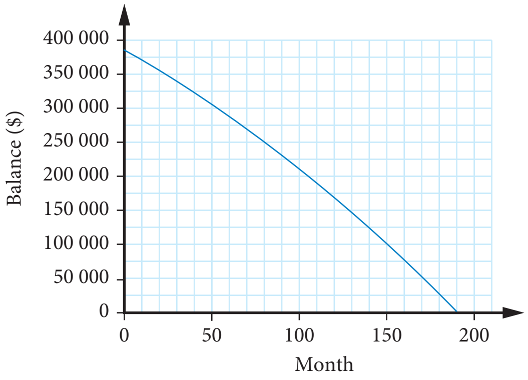
Withdrawal number n	Withdrawal made (\$)	Interest earned (\$)	Principal reduction (\$)	Value of investment (\$)
0	0.00	0.00	0.00	380 000.00
1	3854.00	$\frac{9}{12}\times\frac{1}{100}\times 380\,000.00$ $= 2850.00$	$3854.00 - 2850.00$ $= 1004.00$	$380\,000.00 - 1004.00$ $= 378\,996.00$
2	3854.00	$\frac{9}{12}\times\frac{1}{100}\times 378\,996.00 =$		
3				

b Write a recurrence relation for the value of the investment after n withdrawals.

c Use the recurrence relation and a CAS/calculator's recursive computation to find the value of the investment after 5 months.

d From the following graph of this annuity, approximately how much money would the couple have left after $12\frac{1}{2}$ years?

e Compare this question to Exam prep 7.1 Prep 2. What do you notice?



Prep 2

USING CAS: CALCULATING HOW LONG AN ANNUITY WILL LAST

Ashton purchases a \$500 000 annuity. Interest is paid at 7% per annum compounded monthly. If he receives monthly payments of \$3200, how long will the annuity last? Answer correct to the nearest month.

Prep 3

USING CAS: CALCULATING HOW MUCH TO WITHDRAW FROM AN ANNUITY

Anton invests \$450 000 with interest paid at 6.5% per annum compounded monthly. He receives monthly payments from this investment. What monthly payment will Anton receive if he wishes to receive payments for 20 years?

Annuities

Question 1

Which of the following can be modelled by the recurrence relation

$A_0 = 100\ 000,\ A_{n+1} = 1.0675A_n - 15\ 000$?

A an investment of \$100 000 at 6.75% interest per annum compounding annually where regular annual payments of \$15 000 are added

B an investment of \$100 000 at 1.0675% interest per annum compounding annually where regular annual payments of \$15 000 are added

C an investment of \$100 000 at 1.0675% interest per annum compounding annually where regular annual withdrawals of \$15 000 are made

D an investment of \$100 000 at 6.075% interest per annum compounding monthly where regular monthly withdrawals of \$15 000 are made

E an investment of \$100 000 at 6.75% interest per annum compounding annually where regular annual withdrawals of \$15 000 are made

Question 2

At the start of 2015, Felicity invests \$20 000 at 7.5% per annum. At the end of each year, after interest has been paid, she withdraws \$1600. No other withdrawals or deposits are made. If A_n represents the value of her investment at the end of year n and $A_0 = \$20\ 000$, which one of the following recurrence relation gives the value of her investment at the end of year n?

A $A_{n+1} = 1.075A_n - 1600$

B $A_{n+1} = 0.075A_n + 1600$

C $A_{n+1} = 1.075(A_n - 1600)$

D $A_{n+1} = 0.075(A_n + 1600)$

E $A_{n+1} = 1.075^n A_n - 1600$

Question 3

Branislawa has an annuity of \$120 000 which is earning interest of 6% per annum, compounded monthly. It pays her a fixed amount each month for 18 years. What would this amount be closest to?

A \$1012

B \$909

C \$1013

D \$6988

E \$910

Question 4

A retiree invests \$250 000 in an income-generating account that earns 4.54% per annum on the minimum yearly balance. Interest is credited to the account at the end of each year. Immediately after the interest is credited, a cheque for the amount of \$15 000 is paid out of the account to the retiree. The money remaining in the account is then reinvested. The same procedure is followed each year until the account runs out of money.

If A_n dollars is the amount of money invested at the start of the nth year, then a recurrence relation that describes the way in which the amount of money in the account decreases is

A $A_{n+1} = 1.0454A_n - 15\,000$ where $A_1 = 250\,000$

B $A_{n+1} = 1.0454A_n - 250\,000$ where $A_1 = 15\,000$

C $A_{n+1} = 0.0454A_n - 15\,000$ where $A_1 = 250\,000$

D $A_{n+1} = 4.54A_n - 15\,000$ where $A_1 = 250\,000$

E $A_{n+1} = 0.0454A_n - 250\,000$ where $A_1 = 15\,000$

[VCAA 2002 1NPQ9]

Question 5

Gregor invests \$10 000 and earns interest at a rate of 6% per annum compounding quarterly.

Every quarter, after interest has been added, he withdraws \$500.

At the end of four years, after interest has been added and he has made the \$500 withdrawal, the value of the remaining investment will be closest to

A \$3720 **B** \$4220 **C** \$5440 **D** \$21 660 **E** \$22 160

[VCAA 2005 1BQ7]

Question 6

An annuity of \$14 000 with an interest rate of 7.5% per annum compounded quarterly from which regular quarterly withdrawals of \$2000 are made can be modelled by the recurrence relation

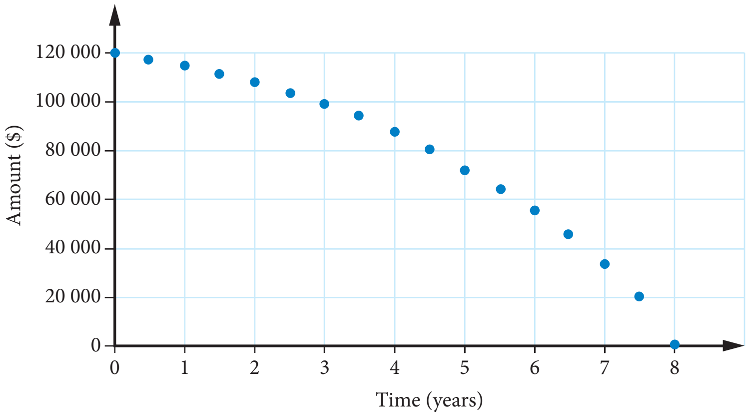
$A_0 = 14\,000,\ A_{n+1} = kA_n - 2000$

What would the value of k be closest to?

A 1.075 **B** 1.018 75 **C** 1.0188 **D** 0.075 **E** 1.0063

Question 7

Which of the following annuities could this graph model?



A an annuity of $120 000 at 6% interest compounding annually with regular annual withdrawals over eight years

B an annuity of $120 000 compounding annually over eight years with annual withdrawals of $23 000

C an annuity of $120 000 at 5.5% interest compounding six-monthly with regular six-monthly withdrawals over eight years

D an annuity of $120 000 compounding monthly over six years with monthly withdrawals of $2451

E an annuity of $120 000 at 8% interest compounding six-monthly with regular six-monthly withdrawals over seven years

Use the following information to answer Questions 8 & 9.

The step-by-step detail for the first four months of an annuity is shown in the following table.

Withdrawal number n	Withdrawal made ($)	Interest earned ($)	Principal reduction ($)	Value of investment ($)
0	0.00	0.00	0.00	160 000.00
1	2000.00	640.00	1360.00	158 640.00
2	2000.00	634.56	?	157 274.56
3	2000.00	629.10	1370.90	155 903.66
4	2000.00	623.61	1376.39	?

Question 8

The missing principal reduction after two months is

A \$1365.44 **B** \$725.44 **C** \$157 274.56 **D** \$1360.00 **E** \$1369.30

Question 9

The missing value of the investment after 4 months is

A \$155 903.66 **B** \$155 903.66 + \$1376.39 **C** \$154 527.27

D \$155 903.66 – \$1370.90 **E** \$157 274.56 – \$155 903.66

Question 10

Ani invests \$200 000 at a rate of 9% p.a. compounded monthly.

Each month, after interest is paid, she withdraws \$2000 and the remaining amount is reinvested.

The value of her investment, A_n, at the end of month n is given by a recurrence relation of the form

$A_0 = a$, $A_{n+1} = kA_n + d$.

a Find the values of k, d and a.

b For how long, correct to the nearest month, will Ani receive a monthly income from this investment?

c How long would it last if she withdrew \$3000 a month?

d How long would it last at an interest rate of 8% with withdrawals of \$2000?

7.5 Perpetuities

A **perpetuity** (also known as a **perpetual annuity**) is a type of annuity where a permanently invested amount of money provides regular payments that continue forever. In a perpetuity the regular payment per time period is equal to the interest earned on the investment over that period. The balance of the amount invested stays the same forever. Often scholarships or grants are set up as perpetuities.

Amana Images/Andrew Lichtenstein

A perpetuity is a special case of an annuity, so let's start off with the recurrence relation for the value of an annuity, V_n, after n withdrawals:

$V_0 = \text{principal},\ V_{n+1} = (1 + \frac{r}{100})\ V_n - D$

In a perpetuity the value never changes so $V_{n+1} = V_n = V_0$.

This means the recurrence relation for a perpetuity gives

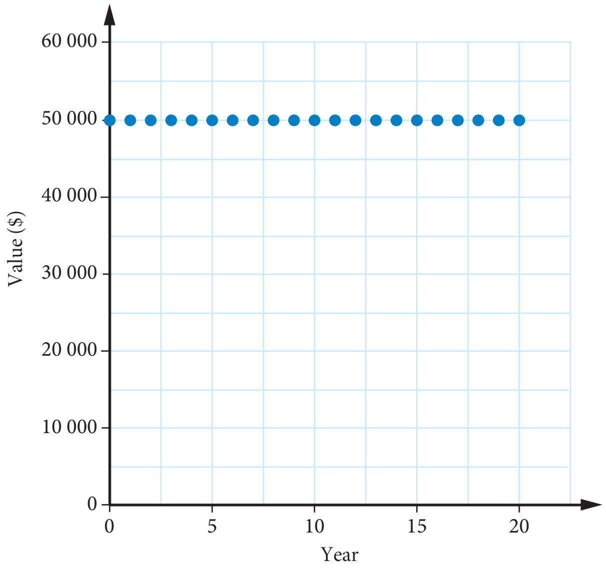
$V_0 = (1 + \frac{r}{100}) V_0 - D$

Rearranging the equation gives us:

$$\begin{aligned} D &= (1 + \frac{r}{100})\ V_0 - V_0 \\ &= V_0 + \frac{r}{100}\ V_0 - V_0 \\ &= \frac{r}{100}\ V_0 \\ &= \frac{V_0 \times r}{100} \end{aligned}$$

9780170371100

The graph of the balance of a perpetuity over time follows a horizontal line. For example, the graph of a perpetuity with a value of \$50 000, for the first 20 years, would look like:



For a perpetuity

$$D = = \frac{V_0 \times r}{100}$$

where

D is the regular withdrawal

V_0 is the principal

r is the interest rate per withdrawal period

A perpetuity is an investment. The equivalent loan, where the regular payments always exactly equal the interest, is called an interest-only loan. In an interest-only loan you continue to owe the original amount that you borrowed.

Worked example 5

Jacob has \$500 000 to set up a perpetuity for his granddaughter Sophie. He invests the money in bonds that return 3.5% per annum compounded quarterly. How much will Sophie receive each quarter from this investment?

Working

1 Write down the formula for a perpetuity investment.

$$D = \frac{V_0 \times r}{100}$$

2 Write down the known values of any variables.

The amount invested is \$500 000.

The interest rate per quarter is $\frac{3.5}{4}$%.

$$V_0 = 500\,000$$

$$r = \frac{3.5}{4}$$

3 Substitute the values of these variables into the formula and evaluate.

$$\begin{aligned} D &= \frac{V_0 \times r}{100} \\ &= \frac{500\,000 \times 3.5}{400} \\ &= 4375 \end{aligned}$$

4 Write the answer.

Sophie will receive \$4375 quarterly from this investment.

Worked example 6

Lucy wishes to set up a scholarship fund in her name so that each year an amount of $5000 is awarded to a promising cross-country runner at the school she attended. If interest on the investment averages to 4% per annum simple interest, compounded yearly, how much should she invest?

	Working
1 Write down the formula for a perpetuity investment.	$D = \frac{V_0 \times r}{100}$
2 Write down the known values of any variables. The regular payment per year is $5000. The interest rate per annum is 4%.	$D = 5000$ $r = 4$
3 Substitute the values of these variables into the formula. Solve for V_0, using a CAS/calculator solve function if necessary.	$D = \frac{V_0 \times r}{100}$ $5000 = \frac{V_0 \times 4}{100}$ $V_0 = \frac{5000 \times 100}{4}$ $= 125\ 000$
4 Write the answer.	Lucy needs to invest $125 000.

You can't use finance solvers for perpetuity problems because they require you to enter a value, *n*, for time periods, and time periods aren't relevant for perpetuities. Finance solvers are for problems where there is compounding involved. There is no compounding in a perpetuity. Perpetuity problems are essentially the same as simple interest problems.

EXAM PREP 7.5

Perpetuities

Prep 1 WORKED EXAMPLE 5

Mila invests $650 000 in a perpetuity that returns 4.2% per annum compounded annually. How much will Mila receive each year from this investment?

Prep 2 WORKED EXAMPLE 6

Emile sets up a scholarship fund in his name so that each year an amount of $3000 is awarded to the top mathematics student in their final year at the school he attended. If interest on the investment averages 4.5% per annum simple interest, compounded yearly, how much should he invest?

Perpetuities

Question 1

Naoto receives a sum of $200 000 and invests the money in a perpetuity that pays him a fixed amount each month. The interest rate is 4% p.a. The amount that he receives each month is

A $153.85 **B** $307.69 **C** $615.40 **D** $666.67 **E** $8000

Question 2

Pia invests $800 000 in an ordinary perpetuity to provide an ongoing fortnightly pension for her retirement. The interest rate for this investment is 5.8% per annum.

Assuming there are 26 fortnights per year, the amount she will receive at the end of each fortnight is closest to

A $464 **B** $892 **C** $1422 **D** $1785 **E** $3867

[VCAA 2008 1BRMQ2]

Question 3

Jane invests in an ordinary perpetuity to provide her with a weekly payment of $500.

The interest rate for the investment is 5.9% per annum.

Assuming there are 52 weeks per year, the amount that Jane needs to invest in the perpetuity is closest to

A $26 000 **B** $102 000 **C** $154 000 **D** $221 000 **E** $441 000

[VCAA 2011 1BRMQ5]

Question 4

$100 000 is invested in a perpetuity at an interest rate of 6% per annum.

After 10 quarterly payments have been made, the amount of money that remains invested in the perpetuity is

A $15 000 **B** $40 000 **C** $85 000 **D** $94 000 **E** $100 000

[VCAA 2013 1BRMQ5]

Question 5

Grandpa invested in an ordinary perpetuity from which he receives a monthly pension of $584.

The interest rate for the investment is 6.2% per annum.

The amount Grandpa has invested in the perpetuity is closest to

A $3600 **B** $9420 **C** $94 200 **D** $43 400 **E** $113 000

[VCAA 2006 1BRMQ3]

Question 6

\$360 000 is invested in a perpetuity at an interest rate of 5.2% per annum.

a Find the monthly payment that the perpetuity provides. 1 mark

b After six years of monthly payments, how much money remains invested in the perpetuity? 1 mark

[VCAA 2010 2BRM Q2]

Question 7

Arthur invested \$80 000 in a perpetuity that returns \$1260 per quarter. Interest is calculated quarterly.

a Calculate the annual interest rate of Arthur's investment. 1 mark

b After Arthur has received 20 quarterly payments, how much money remains invested in the perpetuity? 1 mark

c Arthur's wife, Martha, invested a sum of money at an interest rate of 9.4% per annum, compounding quarterly. She will be paid \$1260 per quarter from her investment.

After ten years, the balance of Martha's investment will have reduced to \$7000. Determine the initial sum of money Martha invested.

Write your answer, correct to the nearest dollar. 1 mark

[VCAA 2012 2BRM Q4]

Annuity investments

7.6

Annuity investment recurrence relations

An **annuity investment** is an investment that involves making an initial deposit followed by additional regular payments into an account earning a fixed rate of compound interest.

People often use this type of investment to save for retirement, the deposit for a house, or large purchases such as a car.

The compounding effect on the initial deposit and the additional regular payments means the balance of the investment has the shape of a steeply increasing curve.

Exam hack

With an *annuity* a fixed sum of money is regularly *withdrawn*. With an *annuity investment* a fixed sum of money is regularly *added*. An annuity can also be thought of as an investment, so the important thing to distinguish between the two is to ask the question, is money being regularly withdrawn or regularly added to?

iStock.com/Neustockimages

Worked example 7

Vinh invests $4000 in an account earning 8% p.a. interest compounded annually. He deposits an extra $1200 into the account each year after the initial deposit. The table below shows the progress of the investment for the first three years.

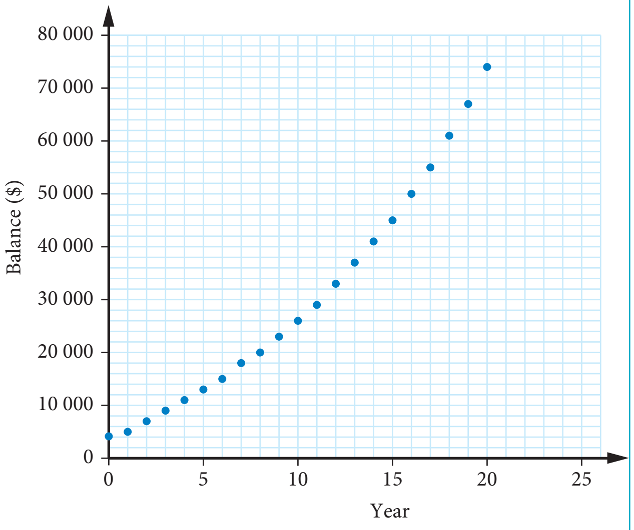
Deposit number *n*	Deposit made ($)	Interest earned ($)	Principal increase ($)	Value of investment ($)
0	0.00	0.00	0.00	4000.00
1	1200.00	$\frac{8}{100} \times 4000 = 320.00$	$1200.00 + 320.00$ $= 1520.00$	$4000.00 + 1520.00$ $= 5520.00$
2	1200.00	$\frac{8}{100} \times 5520 = 441.60$		
3	1200.00			

a Copy and complete the table showing the value of the investment for the first 3 years, giving all values to the nearest cent.

b Write a recurrence relation for the value of the investment after *n* deposits.

c Use the recurrence relation and a CAS/calculator's recursive computation to find the value of the investment after 5 years.

d From the graph of the investment, approximately how much money will Vinh have in the account after 15 years?



Working

a Calculate the principal increase by adding the deposit made and interest earned each year.

Calculate the value of the investment by adding the principal increase to the previous year's investment value.

Give all values to the nearest cent.

Deposit number *n*	Deposit made ($)	Interest earned ($)	Principal increase ($)	Value of investment ($)
0	0.00	0.00	0.00	4000.00
1	1200.00	320.00	1520.00	5520.00
2	1200.00	441.60	1641.60	7161.60
3	1200.00	572.93	1772.93	8934.53

b	Value of the investment = previous investment value + interest calculated on previous investment value + regular deposit	Let V_n = the value of the investment after n deposits. The recurrence relation is $V_0 = 4000,\ V_{n+1} = V_n + 0.08V_n + 1200$ which can be written as $V_0 = 4000,\ V_{n+1} = (1 + 0.08)V_n + 1200$ or $V_0 = 4000,\ V_{n+1} = 1.08V_n + 1200$
c	Use a CAS/calculator's recursive computation to repeat the recurrence relation steps 5 times.	Starting value is 4000. Use Ans × 1.08 + 1200 and Enter / EXE five times. The value of the investment after five years is $12 917.23.
d	Reading from the graph.	After 15 years Vinh will have approximately $45 000 in the account.

The value of an annuity investment, V_n, after n deposits is given by the recurrence relation

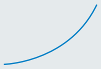
$$V_0 = \text{principal},\ V_{n+1} = (1 + \frac{r}{100})V_n + D$$

where

r is the interest rate per deposit period

D is the regular deposit.

The graph of an annuity investment will always have this shape:



Annuity investments and finance solvers

Finance solvers can be used to solve problems involving annuity investments.

For a calculation relating to an annuity investment, the fields for the Financial Solver or Financial Application (Compound Interest) are as follows.

N is the number of time periods.

I% is the interest rate as a percentage per annum.

PV is the present value (for an annuity investment, this is the amount invested and it is negative since this amount is being given to the bank).

Pmt or **PMT** is the value of the regular payments being made (for an annuity investment, this is negative because this is being paid to the bank).

FV is the future value (this is positive as it is money that will be given back by the bank at the end of the investment period).

PpY or **P/Y** is the number of payments per year.

CpY or **C/Y** is the number of times in a year interest is compounded.

PpY or **P/Y** and **CpY** or **C/Y** usually take the same value for annuities calculations.

Using CAS Calculating annuity investment balances and interest earned

Kim opens a savings account with an initial investment of \$3000 at 6.5% p.a. compounded monthly. At the end of each month, she makes a deposit of \$800. Find

a the amount of her investment after 5 years

b the interest earned.

a	**STEP 1** Use a finance solver to find the amount of the investment after 5 years. N is the number of months in 5 years. I% is the annual interest rate. PV is negative, since Kim has given the bank the \$3000. Pmt or PMT is negative, since this amount is being given to the bank each month. PpY or P/Y and CpY or C/Y are both 12, since payments are made monthly and interest compounds monthly.	$N = 5 \times 12 = 60$ $I\% = 6.5$ $PV = -3000$ Pmt or $PMT = -800$ $FV =$ PpY or $P/Y = 12$ CpY or $C/Y = 12$
	STEP 2 **TI-NSPIRE CAS** Move the cursor into the FV field and press `enter`. **CLASSPAD** Tap `FV`. Note that the result is positive since this is money to be paid to Kim by the bank.	$FV = 60687.63$ The amount of Kim's investment after 5 years is \$60 687.63.
b	**STEP 1** Determine the total amount invested by Kim. She makes an initial deposit of \$3000 and then 60 payments of \$800.	Total invested $= \$3000 + 60 \times \$800$ $= \$51\ 000$
	STEP 2 The interest earned is the difference between the final value of the investment and the total invested by Kim.	Interest = Final value of investment − Total invested $= \$60\ 687.63 - \$51\ 000$ $= \$9687.63$

Using CAS Calculating regular deposits for annuity investments

The owners of a restaurant estimate that $40 000 will be needed in 4 years' time for renovations. They open an investment account with an initial deposit of $2000 and interest of 6% p.a. compounded quarterly. How much must be deposited into the account each quarter in order for them to have enough for the renovations in 4 years' time?

STEP 1

Use a finance solver to find the value of the regular deposit.

N is the number of quarters in 4 years.

I% is the annual interest rate.

PV is negative, since the owners have given the bank the $2000.

FV is positive, since the bank will return $40 000 in 4 years' time.

PpY or P/Y and CpY or C/Y are both 4, since payments are made quarterly and interest compounds quarterly.

$N = 4 \times 4 = 16$

$I\% = 6$

$PV = -2000$

Pmt or PMT =

$FV = 40\ 000$

PpY or P/Y = 4

CpY or C/Y = 4

STEP 2

TI-NSPIRE CAS

Move the cursor into the Pmt field and press enter.

CLASSPAD

Tap FV.

Pmt or PMT = 2809.07

The regular quarterly deposit must be $2089.07.

EXAM PREP 7.6

Annuity investments

Prep 1 WORKED EXAMPLE 7

Megan invests $500 in an account earning 8% p.a. interest compounded annually. She deposits an extra $500 into the account each year after the initial deposit. The table below shows the progress of the investment for the first three years.

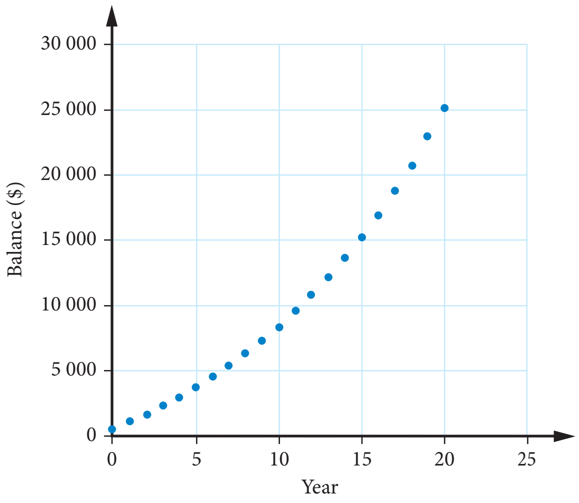
Deposit number n	Deposit made ($)	Interest earned ($)	Principal increase ($)	Value of investment ($)
0	0.00	0.00	0.00	500.00
1	500.00	$\frac{8}{100} \times 500 = 40.00$	$500.00 + 40.00 = 540.00$	$500.00 + 540.00 = 1040.00$
2	500.00	$\frac{8}{100} \times 1040 = 83.20$		
3	500.00			

a Copy and complete the table showing the value of the investment for the first 3 years, giving all values to the nearest cent.

b Write a recurrence relation for the value of the investment after *n* deposits.

c Use the recurrence relation and a CAS/calculator's recursive computation to find the value of the investment after 5 years.

d From the graph of the investment, approximately how much money will Megan have in the account after 15 years?



Prep 2 USING CAS: CALCULATING ANNUITY INVESTMENT BALANCES AND INTEREST EARNED

Hannah opens a savings account with an initial investment of $5000 at 6.2% p.a. compounded monthly. At the end of each month, she makes a deposit of $500. Find

a the amount of her investment after 8 years

b the interest earned.

Prep 3 USING CAS: CALCULATING REGULAR DEPOSITS FOR ANNUITY INVESTMENTS

John and Nancy start a trust fund for their newborn granddaughter, Emma, to finance her university education. They open a savings account earning 7.1% p.a. compounded annually. Their initial deposit is zero. What amount should they deposit each year so that she will have $50 000 after 18 years?

Prep 4

Match the financial application to its graph shape.

i annuity investment

ii annuity

iii perpetuity

iv reducing balance loan

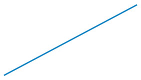
v compound interest investment

vi simple interest investment

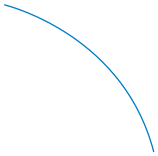
vii reducing balance depreciation

viii flat rate depreciation

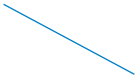
A



B

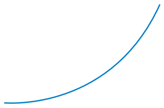


C

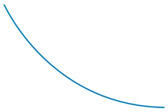


D

E



F



Annuity investments

Question 1

Eric invests \$8000 in an account earning 7.2% p.a. interest compounded annually. He deposits an extra \$8000 into the account each year after the initial deposit. If A_n is the amount in his account after year n, a recurrence relation for determining A_n is

A $A_0 = 8000, A_{n+1} = 1.072A_n + 8000$

B $A_0 = 8576, A_{n+1} = 1.072A_n + 8000$

C $A_0 = 8000, A_{n+1} = 0.072A_n + 8000$

D $A_0 = 8000, A_{n+1} = 1.072(A_n + 8000)$

E $A_0 = 8576, A_{n+1} = 1.072(A_n + 8000)$

Use the following information to answer Questions 2–4.

The step-by-step detail for a four-year annuity investment based on 5% interest p.a. compounded annually is shown in the following table:

Deposit number *n*	Deposit made (\$)	Interest earned (\$)	Principal increase (\$)	Value of investment (\$)
0	0.00	0.00	0.00	2500.00
1	1000.00	?	1125.00	3625.00
2	1000.00	181.25	1181.25	4806.25
3	1000.00	240.31	1240.31	6046.56
4	1000.00	302.33	1302.33	?

Question 2

How much money is in the account after two years?

A \$1181.25 **B** \$1000.00 **C** \$4806.25 **D** \$3625.00 **E** \$1125.00

Question 3

What is the missing interest earned in the first year?

A \$125.00 **B** \$56.25 **C** \$59.06 **D** \$181.25 **E** \$50.00

Question 4

What is the investment value after four years?

A \$6046.56 **B** \$7348.89 **C** \$6286.87 **D** \$3802.33 **E** \$4744.23

Question 5

Jacinta invests \$100 into an account earning 8.4% p.a. compounded weekly.

She deposits a further \$100 each week for 50 weeks.

The value of her investment at the end of the 50 weeks is

A \$5203.10 **B** \$5008.00 **C** \$5311.50 **D** \$5403.85 **E** \$5420

Question 6

A couple estimate that they will need about \$50 000 to cover a wedding and honeymoon in two years' time. They open an investment account with an initial deposit of \$5000 and interest of 4.5% compounded quarterly. How much must be deposited into the account each quarter for them to have enough for the wedding in two years' time?

A \$1587.47 **B** \$1776.65 **C** \$2065.24 **D** \$5350.98 **E** \$6665.09

Question 7

Rebecca invested \$4000 at 5.0% per annum with interest compounding quarterly.

After interest is paid at the end of each quarter, Rebecca adds \$800 to her investment.

The value of her investment at the end of the second quarter, after the \$800 has been added, is closest to

A \$4101 **B** \$4901 **C** \$4911 **D** \$5711 **E** \$6060

[VCAA 2010 1BRMQ9]

Question 8

Lena has some money that she wishes to invest for a period of five years. She is considering three investment options.

a **Investment Option A**

\$10 000 is deposited into an account with an interest rate of 4.8% per annum compounding monthly for five years. Calculate the value of Investment Option A at the end of five years.

Write your answer correct to the nearest cent. 1 mark

b **Investment Option B**

\$4000 is deposited into an account with an interest rate of 4.8% per annum compounding monthly. At the end of each month, for a period of five years, a further \$100 is deposited after interest has been paid.

Determine the value of Investment Option B at the end of five years (immediately after the \$100 has been deposited). Write your answer correct to the nearest cent. 1 mark

c **Investment Option C**

Investment Option B is followed for two years. After this, the amount deposited at the end of each month changes. With the new monthly deposit, Investment Option C is worth \$13 000 at the end of the five years.

i Find the new amount deposited at the end of each month for the remaining three years. Write your answer correct to the nearest cent. 1 mark

ii Determine the total amount of interest earned by Investment Option C over the **five-year** period.

Write your answer correct to the nearest cent. 2 marks

[VCAA 2005 2BRM Q3]

Question 9

\$500 is deposited into an account with an interest rate of 6.5% per annum compounding monthly. Deposits of \$200 are made to this account on the last day of each month after interest has been paid. Determine the total value of this investment at the end of eight years.

Write your answer correct to the nearest dollar. 1 mark

[VCAA 2006 2BRM Q3c]

Question 10

Tom and Patty both decided to invest some money from their savings. Each chose a different investment strategy.

Tom's investment strategy

- Deposit \$5600 into an account with an interest rate of 7.2% per annum, compounding monthly.
- Immediately after interest is paid into his investment account on the last day of each month, deposit a further \$200 into the account.

a Determine the total amount in Tom's investment account at the end of the first month. 1 mark

Patty's investment strategy

- Invest \$8 000 at the start of the year at an interest rate of 7.2% per annum, compounding **annually**.

b The following expression can be used to determine the value of Patty's investment at the end of the first year. Complete the expression by filling in the box.

Value of investment = $8000 \times (1 + \boxed{})$ 1 mark

c At the end of twelve months, Patty has more money in her investment account than Tom.

How much more does she have? Write your answer to the nearest cent. 2 marks

d What annual compounding rate of interest would Patty need in order to earn \$1000 interest in one year on her \$8000 investment?

Write your answer correct to 1 decimal place. 1 mark

[VCAA 2011 2BRM Q2]

SUMMARY

7

Reducing balance loans and annuities

Practice quiz

Reducing balance loans

- A **reducing balance loan** is one where interest is calculated on the amount still owing after each repayment is made.
- The balance on a reducing balance loan, B_n, after the nth repayment is given by the recurrence relation

  $B_0 =$ amount borrowed, $B_{n+1} = (1 + \frac{r}{100})\, B_n - D$

    where
    r is the interest rate per repayment period
    D is the regular repayment

  e.g. $B_0 = 450\ 000$, $B_{n+1} = 1.005B_n - 3797$
- Total interest paid = Total repaid − Amount borrowed
  = Number of payments made × Regular payment − Amount borrowed
  $= n \times D - B_0$
- Interest after n payments = Total repaid after n time periods − Reduction in the amount owing on the loan
  $= n \times D - (B_0 - B_n)$
- The graph of a reducing balance loan will always have this shape: 

Finance solvers for reducing balance loans

- **N** is the number of time periods.
- **I%** is the interest rate as a percentage per annum.
- **PV** is the present value (for a loan, this is the amount borrowed and is positive since this amount is being given by the bank).
- **Pmt** or **PMT** is the value of the regular payments being made (for a loan, this is negative because this is being paid to the bank).
- **FV** is the future value (this will be 0 if the loan is to be fully repaid, otherwise it is negative as it is money that needs to be paid to the bank).
- **PpY** or **P/Y** is the number of payments per year.
- **CpY** or **C/Y** is the number of times in a year interest is compounded.
- **PpY** or **P/Y** and **CpY** or **C/Y** usually take the same value.

9780170371100

Annuities

- An **annuity** is a type of investment. A sum is invested and interest is compounded at some fixed rate. Withdrawals are made at regular intervals, usually until the value of the investment is $0.
- The value of an annuity, V_n, after n withdrawals is given by the recurrence relation

  $V_0 = \text{principal},\ V_{n+1} = (1+\frac{r}{100})V_n - D$ where

  r is the interest rate per withdrawal period

  D is the regular withdrawal

  e.g. $V_0 = 450\ 000,\ V_{n+1} = 1.005V_n - 3797$
- Total interest earned = Total withdrawals − Principal

  = Number of withdrawals made × Regular withdrawal − Principal

  $= n \times D - B_0$
- Interest earned after n withdrawals = Total withdrawals after n time periods − Reduction in the value of the annuity $= n \times D - (V_0 - V_n)$
- The graph of an annuity will always have the same shape as a reducing balance loan: 

Finance solvers for annuities

- **N** is the number of time periods.
- **I%** is the interest rate as a percentage per annum.
- **PV** is the present value (for an annuity, this is the amount invested and is negative since this amount is being given to the bank).
- **Pmt** or **PMT** is the value of the regular payments being made (for an annuity, this is positive because this is being paid by the bank).
- **FV** is the future value (this is 0 at the end of the annuity. At other times, it is positive as it is money that will be given back by the bank at the end of the investment period).
- **PpY** or **P/Y** is the number of payments per year.
- **PpY** or **P/Y** and **CpY** or **C/Y** usually take the same value.

Perpetuities

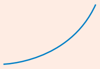
- A **perpetuity** is a type of annuity where a permanently invested amount of money provides regular payments that continue forever and the balance of the amount invested stays the same forever.
- For a perpetuity

  $D = \frac{V_0 \times r}{100}$ where
  D is the regular withdrawal
  V_0 is the principal
  r is the interest rate per withdrawal period
- The graph of the balance of a perpetuity over time follows a horizontal line.

Annuity investments

- An **annuity investment** is an investment that involves making an initial deposit followed by additional regular payments into an account earning a fixed rate of compound interest.
- The value of an annuity investment, V_n, after n deposits is given by the recurrence relation

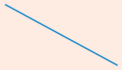
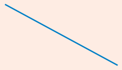
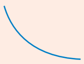
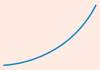
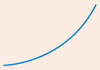
  $V_0 = \text{principal},\ V_{n+1} = (1 + \frac{r}{100})V_n + D$ where
  r is the interest rate per deposit period
  D is the regular deposit

  e.g. $V_0 = 4000,\ V_{n+1} = 1.08V_n + 1200$
- The graph of an annuity investment will always have this shape: 

Finance solvers for annuity investments

- **PV** is the present value (for an annuity investment, this is the amount invested and is negative since it is being given to the bank).
- **Pmt** or **PMT** is the value of the regular payments being made (for an annuity investment, this is negative because this is being paid to the bank).
- **FV** is the future value (this is positive as it is money that will be given back by the bank at the end of the investment period).
- **PpY** or **P/Y** and **CpY** or **C/Y** usually take the same value.

Recursion and financial modelling summary table

Financial application	Growth/Decay	Recursion relation example	Graph shape
Flat rate depreciation	Linear decay	$V_0 = 9500$, $V_{n+1} = V_n - 1425$	
Unit cost depreciation	Linear decay if usage is the same each time period	$V_0 = 25\,000$, $V_{n+1} = V_n - 375$	
Reducing balance depreciation	Geometric decay	$V_0 = 10\,000$, $V_{n+1} = 0.82V_n$	
Simple interest investment	Linear growth	$V_0 = 9500$, $V_{n+1} = V_n + 1425$	
Compound interest investment	Geometric growth	$V_0 = 35\,000$, $V_{n+1} = 1.062V_n$	
Reducing balance loan	Combined geometric growth and linear decay	$V_0 = 450\,000$, $V_{n+1} = 1.005V_n - 3797$	
Annuity	Combined geometric growth and linear decay	$V_0 = 450\,000$, $V_{n+1} = 1.005V_n - 3797$	
Perpetuity	No growth or decay	No recursion relation	
Annuity investment	Combined geometric growth and linear growth	$V_0 = 4000$, $V_{n+1} = 1.08V_n + 1200$	

RECURSION AND FINANCIAL MODELLING

Examination 1

Reading time: (5 minutes)
Writing time: (30 minutes)

- Students are permitted to bring into the examination room: pens, pencils, highlighters, erasers, sharpeners, rulers, one bound reference, one approved graphics calculator or approved CAS calculator or CAS software and, if desired, one scientific calculator. Calculator memory DOES NOT need to be cleared.
- Students are NOT permitted to bring into the examination room: blank sheets of paper and/or white out liquid/tape.
- Students may refer to the sheet of miscellaneous formulas supplied.

Students are NOT permitted to bring mobile phones and/or any other unauthorised electronic devices into the examination room.

Instructions

Choose the response that is correct for the question.

A correct answer scores 1, an incorrect answer scores 0.

Marks will not be deducted for incorrect answers.

No marks will be given if more than one answer is completed for any question.

Unless otherwise indicated, the diagrams are **not** drawn to scale.

Use the following information to answer Questions 1 & 2.

The step-by-step detail for a four year annuity investment based on 10% interest p.a. compounded annually is shown in the following table:

Deposit number n	*Deposit made*	*Interest earned*	*Principal increase*	*Value of investment*
0	0.00	0.00	0.00	$10 000.00
1	$2500.00	$1000.00	$3500.00	$13 500.00
2	$2500.00	?	$3850.00	$17 350.00
3	$2500.00	$1735.00	$4235.00	?
4	$2500.00	$2158.50	$4658.50	$26 243.50

EXAMINATION 1 – continued

Question 1

What is the missing interest earned in the second year?

A. $1350.00 **B.** $3850.00 **C.** $2500.00
D. $1000.00 **E.** $3500.00

Question 2

What is the investment value after three years?

A. $4235.00 **B.** $13 115.00 **C.** $26 243.50
D. $1735.00 **E.** $21 585.00

Question 3

Which one of the following recurrence relations could represent a reducing balance loan?

A. $V_0 = 5000,\ V_{n+1} = V_n - 525$ **B.** $V_0 = 5000,\ V_{n+1} = 0.72V_n$
C. $V_0 = 5000,\ V_{n+1} = 1.06V_n + 1500$ **D.** $V_0 = 5000,\ V_{n+1} = 1.003V_n - 2365$
E. $V_0 = 5000,\ V_{n+1} = 1.062V_n$

Question 4

Amy invests $15 000 for 150 days.

Interest is calculated at the rate of 4.60% per annum, compounding daily.

Assuming that there are 365 days in a year, the value of her investment after 150 days is closest to

A. $15 279 **B.** $15 284 **C.** $15 286 **D.** $15 690 **E.** $16 776

[VCAA 2014 1BRMQ3]

Question 5

A bank approves a $90 000 loan for a customer.

The loan is to be repaid fully over 20 years in equal monthly payments.

Interest is charged at a rate of 6.95% per annum on the reducing monthly balance. To the nearest dollar, the monthly payment will be

A. $478 **B.** $692 **C.** $695 **D.** $1409 **E.** $1579

[VCAA 2014 1BRMQ5]

Question 6

New furniture was purchased for an office at a cost of $18 000.

Using flat rate depreciation, the furniture will be valued at $5000 after four years.

The expression that can be used to determine the value of the furniture, in dollars, after one year is

A. $18\,000 - (4 \times 5000)$ **B.** $18\,000 - \left(\frac{18\,000 - 5000}{4}\right)$ **C.** $18\,000 - \frac{5000}{4}$
D. $\frac{18\,000}{4} - 5000$ **E.** $18\,000 \times 0.726$

[VCAA 2014 1BRMQ7]

EXAMINATION 1 – continued

Question 7

Robert invested $6000 at 4.25% per annum with interest compounding quarterly. Immediately after interest is paid at the end of each quarter, he adds $500 to his investment.

The value of Robert's investment at the end of the third quarter, after his $500 has been added, is closest to

A. $6193 **B.** $7569 **C.** $7574
D. $7709 **E.** $8096

[VCAA 2014 1BRMQ8]

Question 8

Leslie borrowed $35 000 from a bank.

Interest is charged at the rate of 4.75% on the reducing monthly balance.

The loan is to be repaid with 47 monthly payments of $802.00 and a final payment that is to be adjusted so that the loan will be fully repaid after exactly 48 monthly payments.

Correct to the nearest cent, the amount of the final payment will be

A. $0.39 **B.** $3.57 **C.** $802.00
D. $802.39 **E.** $805.57

[VCAA 2014 1BRMQ9]

Question 9

Poh borrowed $25 000 and will fully repay the loan in five years. Interest is charged at a rate of 8.9% per annum, calculated monthly on the reducing balance. The amount that Poh has paid off the principal immediately following her eighth repayment is closest to

A. $1413 **B.** $2379 **C.** $2729
D. $3081 **E.** $4142

Question 10

Paul invests $15 000 at 6% p.a. compounding half-yearly. The value of his investment after 3 years is given by

A. $15\,000 \times 0.06 \times 3$ **B.** $15\,000 \times 0.03 \times 6$ **C.** $15\,000 \times 0.03^6$
D. $15\,000 \times 1.03^6$ **E.** $15\,000 \times 0.06^3$

Question 11

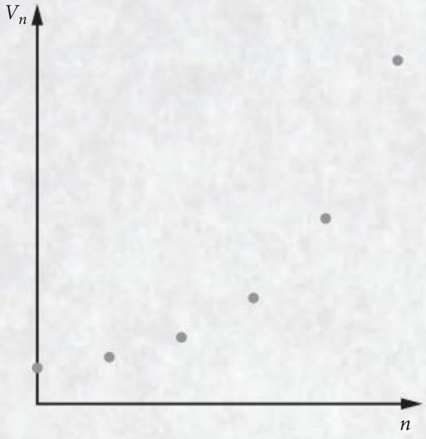
$140 000 is invested for 1 year. The interest rate is 7.2% p.a. compounding monthly for the first four months and then increases to 7.4% p.a. compounding monthly. The total interest earned for the year is closest to

A. $10 080.00 **B.** $10 220.00 **C.** $10 360.00
D. $10 447.95 **E.** $10 618.86

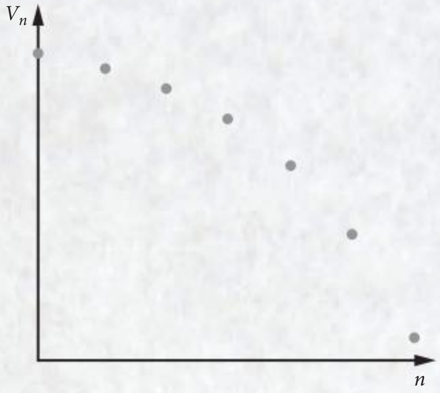
Question 12

Which of the following graphs could represent the value of an asset, V_n, being depreciated by reducing balance depreciation after n time periods?

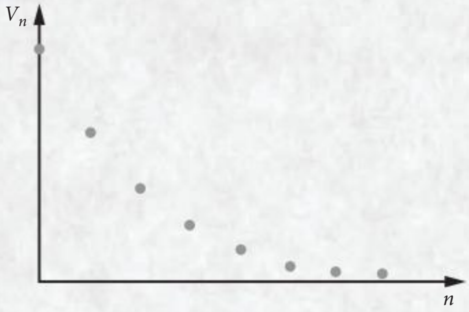
A



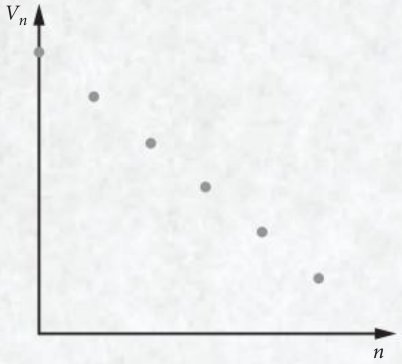
B



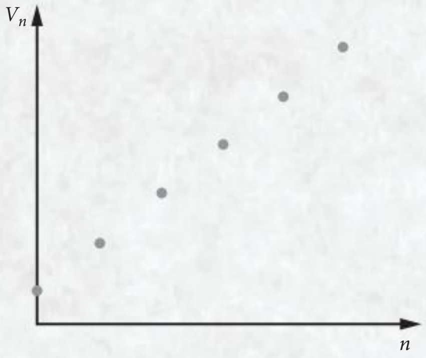
C



D



E



Question 13

A second-hand car is purchased for \$9000. A deposit of \$2500 is paid.

Interest is calculated at the rate of 14.95% per annum on the reducing monthly balance.

The balance and interest will be repaid over two years with equal monthly payments.

The monthly payment is closest to

A. \$315 **B.** \$415 **C.** \$436 **D.** \$575 **E.** \$587

[VCAA 2012 1BRMQ5]

EXAMINATION 1 – continued

Question 14

Jenny borrowed $18 000. She will fully repay the loan in five years with equal monthly payments. Interest is charged at the rate of 9.2% per annum, calculated monthly, on the reducing balance.

The amount Jenny has paid off the principal immediately following the **tenth** repayment is

A. $1876.77 **B.** $2457.60 **C.** $3276.00 **D.** $3600.44 **E.** $3754.00

[VCAA 2006 1BRMQ9]

EXAMINATION 1 – ended

RECURSION AND FINANCIAL MODELLING

Examination 2

Reading time: (5 minutes)
Writing time: (25 minutes)

- Students are permitted to bring into the examination room: pens, pencils, highlighters, erasers, sharpeners, rulers, one bound reference, one approved graphics calculator or approved CAS calculator or CAS software and, if desired, one scientific calculator. Calculator memory DOES NOT need to be cleared.
- Students are NOT permitted to bring into the examination room: blank sheets of paper and/or white out liquid/tape.
- Students may refer to the sheet of miscellaneous formulas supplied.
- All written responses must be in English.

Students are NOT permitted to bring mobile phones and/or any other unauthorised electronic devices into the examination room.

Instructions

You need not give numerical answers as decimals unless instructed to do so. Alternative forms may involve, for example, π, surds or fractions.

Diagrams are not to scale unless specified otherwise.

Question 1 (3 marks)

Hugo won $5000 in a road race and invested this sum at an interest rate of 4.8% per annum compounding monthly.

a. What is the value of Hugo's investment after 12 months?
Write your answer in dollars, correct to the nearest cent. 1 mark

b. **i.** Suppose instead that at the end of each month Hugo added $200 to his initial investment of $5000.
Find the value of this investment immediately after the 12th monthly payment of $200 is made.
Write your answer in dollars, correct to the nearest cent. 1 mark

ii. Assume Hugo follows the investment that is described in part b.i.
Determine the total interest he would earn over the 12-month period.
Write your answer in dollars, correct to the nearest cent. 1 mark

[VCAA 2013 2BRMQ2]

EXAMINATION 2 – continued

Question 2 (2 marks)

Hugo took out a reducing balance loan of \$25 000 to compete in road races overseas. Interest was charged at a rate of 12% per annum compounding quarterly.

His loan is to be repaid fully in four years with equal quarterly payments. After two years, how much of the \$25 000 will Hugo have repaid?

Write your answer, correct to the nearest dollar. 2 marks

[VCAA 2013 2BRMQ4]

Question 3 (2 marks)

A sponsor of the cricket club has invested \$20 000 in a perpetuity. The annual interest from this perpetuity is \$750.

The interest from the perpetuity is given to the best player in the club every year, for a period of 10 years.

a. What is the annual rate of interest for this perpetuity investment? 1 mark

b. After 10 years, how much money is still invested in the perpetuity? 1 mark

[VCAA 2014 2BRMQ2ab]

Question 4 (4 marks)

The cricket club had invested \$45 550 in an account for four years.

After four years of compounding interest, the value of the investment was \$60 000.

a. How much interest was earned during the four years of this investment? 1 mark

Interest on the account had been calculated and paid quarterly.

b. What was the annual rate of interest for this investment?

Write your answer, correct to 1 decimal place. 1 mark

The \$60 000 was re-invested in another account for 12 months.

The new account paid interest at the rate of 7.2% per annum, compounding monthly.

At the end of each month, the cricket club added an additional \$885 to the investment.

c. **i.** The equation below can be used to determine the account balance at the end of the first month, immediately after the \$885 was added.

Complete the equation by filling in the boxes. 1 mark

$$\text{account balance} = 60\,000 \times \left(1 + \boxed{}\right) + \boxed{}$$

ii What was the account balance at the end of 12 months?

Write your answer, correct to the nearest dollar. 1 mark

[VCAA 2014 2BRMQ3]

EXAMINATION 2 – continued

Question 5 (2 marks)

Hugh opens a savings account with an initial investment of $10 000 at 6.5% p.a. compounded monthly. At the end of each month, he makes a deposit of $1000. Find

a. the amount of his investment after 10 years 1 mark

b. the interest earned. 1 mark

Question 6 (4 marks)

Desi receives an inheritance of $70 000. She invests the money at 5% per annum. Interest is paid at the end of each year, and after the interest is paid Desi withdraws $6000. The amount remaining gets invested for another year.

a. Find a recurrence relation of the form $A_{n+1} = rA_n + d$, where $A_1 = a$, that gives the value of the investment, A_n, at the start of the nth year. 2 marks

b. What is the value of the investment at the start of the 10th year? 1 mark

c. What amount could be withdrawn at the end of each year so that the value of the investment remains at $70 000? 1 mark

Question 7 (3 marks)

The initial value of a computer server is $16 720. It will be depreciated using the unit cost method at a rate of 22 cents per hour of use. Each year the equipment will be used for 7600 hours.

a. Calculate the depreciated value of the server after 4 years. 1 mark

b. Determine the annual flat rate of depreciation that will give the same annual depreciation as the unit cost method. 1 mark

c. The reducing balance method of depreciation is used to depreciate the value of the server. Find, correct to the nearest dollar, the value of the equipment after 8 years if the server depreciates at a rate of 12% per year. 1 mark

EXAMINATION 2 – ended

SOLUTIONS

- For multiple choice questions, the tables show the percentage of students who chose each option in the VCAA examination. The correct answer is indicated by highlighting.
- The written response question percentages represent the average percentage score achieved by students in the VCAA examination for that question.
- Highlighting of a question number indicates that working and/or the relevant section from the VCAA examination report has been included.
- Colour coding has been used to show question grading: 🟩 = 80–100%, 🟨 = 60–79%, 🟦 = 0–59%, ⬛ = percentage unknown or not applicable.

CHAPTER 1

Review of data distributions

EXAM PREP 1.1 Data and variables

Prep 1

a	discrete numerical	**b**	continuous numerical
c	nominal categorical	**d**	continuous numerical
e	continuous numerical	**f**	nominal categorical
g	nominal categorical	**h**	ordinal categorical
i	discrete numerical	**j**	ordinal categorical

Prep 2

a range = 24 years, median = 21 years

b range = 23, median = 38.5

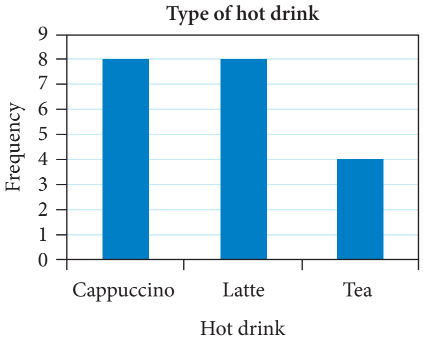
c range = 5.8°C, median = −1.7°C

EXAM PRACTICE 1.1 Data and variables

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1	3	3	91	3	1	2011
Q2	2	1	3	83	11	2003
Q3	7	8	6	6	73	2003
Q4	79	3	14	3	1	2002
Q5	9	62	2	16	11	2003

EXAM PREP 1.2 Tables and charts

Prep 1



The graph shows that cappuccinos and lattes were the most popular hot drinks sold, with tea sales being the lowest.

EXAM PRACTICE 1.2 Tables and charts

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1	1	0	0	0	99	2012
Q2	5	68	6	2	19	2012
Q3	86	8	2	1	3	2010
	Written response %					
Q4	95					2008
Q5	85					2013

Question 4

a 10, 7, 8 **b** 32%

Question 5

a 31 **b** $61\%, \frac{49+45}{153}\times 100 = 61.4379\ldots$

EXAM PREP 1.3 Histograms

Prep 1

a 26 students **b** positively skewed

c

Score	Frequency
5	8
6	6
7	5
8	4
9	2
10	1
Total	26

d 11.5% **e** range = 5

9780170371100

Prep 2

median = 60 because the area of the histogram (i.e. 28 rectangles) is the same either side of 60.

EXAM PRACTICE 1.3 Histograms

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1	2	5	91	2	0	2002
Q2	1	1	4	90	4	2009
Q3	8	6	3	4	79	2009
Q4	1	0	87	12	0	2011
Q5	1	5	3	5	86	2011
Q6	10	12	45	11	21	2011
Q7	2	5	5	85	2	2006
Q8	25	58	11	6	1	2006
Q9	6	2	1	74	17	2005
	Written response					
Q10	70					2007

Question 6

The first four columns on the left of the histogram add to 79, which is close to half of the 160 total number of countries, so the median percentage is close to 20%.

Question 8

The percentage frequency histogram has 100 rectangles. Counting the rectangles upwards from either end shows that the 50th and 51st rectangles lie in the 40–45 interval.

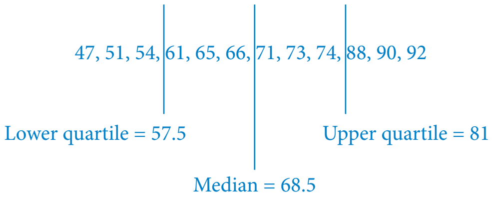
Question 10

a Positively skewed

b **i** 26 **ii** $\frac{20}{103} = 19.4\%$

EXAM PREP 1.4 Boxplots

Prep 1



a

b min = 47, $Q_1 = 57.5$, median = 68.5, $Q_3 = 81$, max = 92

Prep 2

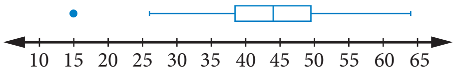
a **i** 1, 3, 5, 8, 11 **ii** 3 to 8

b **i** 7, 10, 12, 13, 16 **ii** 10 to 13

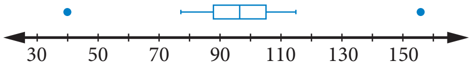
c **i** 1, 2, 6, 8, 10 **ii** 2 to 8

Prep 3

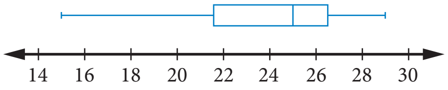
a



b



c



Prep 4

a 25% **b** 24

EXAM PRACTICE 1.4 Boxplots

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1	1	2	14	82	2	2008
Q2	0	1	10	6	83	2008
Q3	13	42	37	6	2	2008
Q4	6	5	3	76	10	2002
Q5	3	5	18	70	4	2010
Q6	13	8	9	55	14	2010
Q7	9	43	14	21	13	2009

Question 3

25% × 79 = 19.75

Examination report

From the boxplot it could be seen that a time of 90 seconds roughly corresponded to the third quartile in the time distribution. Thus, the number of customers who spent more than 90 seconds moving along the aisle was around 25% of 79, or 20 customers (option B) and 42 per cent of students gave this correctly reasoned response. However, 37 per cent of students did not realise that the outliers were already accounted for in determining the top 25% of data values, and incorrectly chose option C.

[VCAA 2008 1RCQ3]

Question 6

From the histogram 21% lie between 179 and 180, and 16% between 180 and 181.

So (21% + 16%) × 300 = 37% × 300 = 111.

Question 7

Reading the % data values from the histogram:

1	2	3	4	5	6	7	8	9	10
0%	17%	12%	6%	12%	19%	28%	5%	1%	0%

Q_1 occurs at 25%. Adding up the percentages from the left, we get to 25% at 3, so $Q_1 = 3$.

Median occurs at 50%. Adding up the percentages from the left, we get to 50% at 6, so median = 6.

Q_3 occurs at 75%. Adding up the percentages from the left, we get to 75% at 7, so $Q_3 = 7$.

These three values match up to B.

Examination report

This question involved matching a boxplot with a given histogram and 43 per cent of students were able to answer correctly. To complete this task successfully, students needed to recognise that a boxplot is a graphical display of the five-number summary of a data set, namely, the minimum value, the first quartile (Q_1), the median (M), the third quartile (Q_3) and the maximum value. As all boxplots had whiskers extending to the same minimum and maximum values, a systematic approach to this question would have been to estimate the values of the median and the first and third quartiles, then look for a match (option B). Students who attempted to answer the question purely by inspection would have found it difficult to obtain the correct answer.

[VCAA 2009 1RCQ6]

EXAM PREP 1.5 Dot plots and stem plots

Prep 1

a

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

Key: 4|5 = 45

b 20 matches **c** 89 points **d** 25%

EXAM PRACTICE 1.5 Dot plots and stem plots

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1	91	2	2	1	3	2007
Q2	0	82	13	4	0	2007
Q3	1	3	93	3	1	2013
Q4	1	86	8	4	1	2013
Q5	2	94	2	0	3	2004
Q6	3	80	5	5	7	2004
Q7						NA
Written response						
Q8	70					2013
Q9	Not reported					2009
Q10	Not reported					2011

Question 8

Examination report

a the mode = 78, the range = 9

b $Q_1 = 75$, $Q_3 = 78$, IQR = 78 − 75 = 3

$Q_1 - 1.5 \times \text{IQR} = 75 - 1.5 \times 3 = 70.5$. Therefore, 70 is an outlier because it is less than 70.5.

This question asked for an explanation of why 70 was an outlier for this group of countries. Many students calculated a value of 70.5 and then wrote that 'it is therefore an outlier'. Further explanation, including a direct comparison between 70 and 70.5, was expected.

Other common, incomplete or unacceptable answers included

70 is outside the bulk of the data – this does not compare 70 with $Q_1 - 1.5 \times \text{IQR}$

it is outside 2SD from the mean – standard deviations do not define an outlier

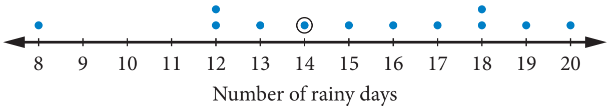
there is no other country close to it – this does not compare 70 with $Q_1 - 1.5 \times \text{IQR}$.

[VCAA 2013 2RCQ2]

Question 9

Examination report

a



A number of students apparently did not see this question and left it blank, even though the verb 'circle' started the sentence and was in bold type.

b **i** 15.5 **ii** 92%

[VCAA 2009 2RCQ1]

Question 10

Examination report

a **i** 25.0 years

Many students did not read this question clearly and found the mean of all the numbers in the stemplot. As a result, a common incorrect answer was 28.04.

a **ii** 28.2 years

b 1.1 years

c $1.5 \times \text{IQR} = 1.5 \times 1.1 = 1.65$ and
$Q_1 - 1.65 = 29.9 - 1.65 = 28.25$

Since 26.0 < 28.25, the age of 26.0 is an outlier.

Many students did not answer this question fully. Most calculated the 28.25 but then did not discuss how 26.0 related to this. Mathematical symbols were sometimes used incorrectly; for example, $28.25 \le 26.0 \le 32.65$.

[VCAA 2011 2RCQ1]

CHAPTER 2

Further data distributions

EXAM PREP 2.1 Log scales

Prep 1

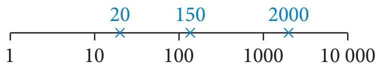
a log base 10 scale **b** log base 10 scale

c linear scale **d** log base 10 scale

e log base 10 scale **f** linear scale

Prep 2

$\log(20) = 1.3$ $\log(150) = 2.2$ $\log(2000) = 3.3$



EXAM PRACTICE 2.1 Log scales

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
D	E	B	D	D	A	B	C

EXAM PREP 2.2 The sample mean and sample standard deviation

Prep 1

a $\bar{x} \approx 23.57, s \approx 2.71$ **b** $\bar{x} = 7.25, s \approx 2.40$

c $\bar{x} \approx 26.29, s \approx 2.42$ **d** $\bar{x} \approx 19.26, s \approx 2.05$

EXAM PRACTICE 2.2 The sample mean and sample standard deviation

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1	3	2	6	1	88	2012
Q2	15	15	60	5	5	2008
Q3	4	30	1	62	3	2004
Q4	11	14	54	8	12	2004
Q5	3	11	59	25	2	2003
Q6	3	14	54	25	4	2003
Q7	53	6	18	15	8	2008
	Written response					
Q8a&b	80					2010
Q8c&d	70					2010

Question 4

Use a CAS/calculator or calculate

$$\frac{0\times6+1\times9+2\times3+3\times1+4\times0+5\times1}{20} = 1.15$$

Question 5

Total mass = 72 × 12 = 864 kg.

Question 6

The extra 2 kg to each person increases the mean mass to 74 kg. The standard deviation is a measure of the spread of a data distribution, so adding 2 kg to each person doesn't affect it (i.e. it stays the same).

Question 7

The mean is greater than the median for positively skewed distributions.

Question 8

Examination report

a 0.5

Common incorrect answers were Norway and 11.

b Median = 28 Range = 56 IQR = 17

c 1|2 4 6

d The distribution is approximately symmetric. This question was not well answered. Many students simply defined median and mean without reference to the given data. Some described the data in vague terms such as 'evenly distributed', while many referred to the absence of outliers. Neither of these terms is sufficient or specific enough to justify the assertion that both mean and median are appropriate measures of centre for this data set. Several explanations suggested that the data was 'symmetrically skewed'; a skewed distribution is not symmetrical.

[VCAA 2010 2RCQ1]

EXAM PREP 2.3 Bell-shaped distributions

Prep 1

a Yes **b** No **c** Yes

d No **e** Yes **f** Yes

g Yes **h** No **i** Yes

j Yes **k** Yes

Prep 2

a 50% **b** 99.7% **c** 2.5%

d 99.7% + 0.15% = 99.85%

e 16% × 426 = 68 customers

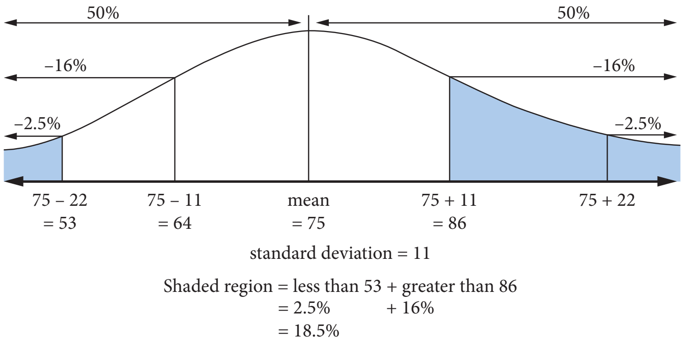
EXAM PRACTICE 2.3 Bell-shaped distributions

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1	2	77	12	8	2	2013
Q2	9	76	10	4	2	2013
Q3	2	10	14	69	4	2009
Q4	10	71	11	5	2	2008
Q5	22	13	15	44	5	2008
Q6	5	8	67	16	4	2006
Q7	2	13	18	61	6	2007
Q8	9	4	58	21	8	2004
Q9	10	6	57	15	11	2005
Q10	25	7	16	47	5	2003
Q11	12	43	27	12	5	2010
Q12	22	41	19	11	7	2002
Q13	2	5	91	2	0	2002
Q14	19	40	16	17	8	2002
Written response						
Q15	70					2012

Question 5

Examination report

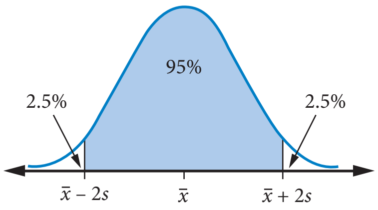
The question required two applications of the 68–95–99.7% rule to obtain the correct answer: one application to determine the percentage of students with pulse rates less than 53 beats/minutes (2.5%) and the second to determine the percentage of students with pulse rates greater than 86 beats/minutes (16%). The percentage of students with pulse rates less than 53 beats/minutes or greater than 86 beats/minutes is then the sum of these two values (18.5% – option D) and 44 per cent of students gave this correct response. In answering any question requiring the use of the 68–95–99.7% rule, a useful strategy is to first draw a normal curve and shade in the required area(s) as defined by the problem statement. This was an essential first step in answering this question.



[VCAA 2008 1RCQ7]

Question 8

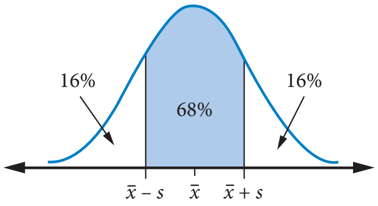
For a bell-shaped distribution, around 95% of the data values lie within two standard deviations of the mean:



$\bar{x} + 2s = 85 + 10 = 95$. So around 2.5% of eggs weigh 95 grams or more. This means 2.5% of eggs are classified as Extra Large.

Question 9

For a bell-shaped distribution, around 68% of the data values lie within one standard deviation of the mean:



$\bar{x} - 1 \times s = 8.8 - 2.2 = 6.6$. So around 16% of this model of car has a fuel consumption less than 6.6 km per litre.

Question 10

For a bell-shaped distribution, around 68% of the data values lie within one standard deviation of the mean.

$\bar{x} - 1 \times s = 64 - 12 = 52$

$\bar{x} + 1 \times s = 64 + 12 = 76$

So 68% of the student scores are between 52 and 78.

There are 2500 students, so 68% of 2500 = 1700 students.

Question 11

Looking at the histogram and adding the percentages from the left gives the 16% point at around 178. Similarly, adding the percentages from the right gives the 16% point at around 182.5. So we know the two data values that are approximately one standard deviation from the mean. Assuming we have an approximately normal distribution:

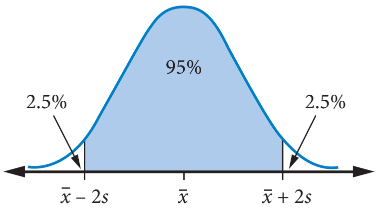
$$\text{standard deviation} \approx \frac{\text{larger data value} - \text{smaller data value}}{2}$$

$$\approx \frac{182.5 - 178}{2}$$

$$= 2.25$$

Question 12

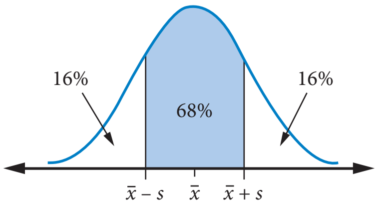
For a bell-shaped distribution, around 95% of the data values lie within two standard deviations of the mean:



$\bar{x} + 2 \times s = 81.1 + 2 \times 17.9 = 116.9$. The heaviest man in the sample is 117 kg. So around 2.5% of men in this population have a mass greater than that of the heaviest man in this sample.

Question 14

We can use the 68–95–99.7% rule because the histogram has the characteristics of a bell-shaped distribution. The symmetry of the histogram tells us the mean is close to 60. The 68–95–99.7% rule tells us that around 68% of the data values lie within one standard deviation of the mean:



This means $60 - 1 \times s = 16\%$

Each rectangle on the histogram is 0.5%, so we need to count 32 rectangles from the left to find where the $60 - 1 \times s$ value is.

Counting 32 rectangles from the left tells us $60 - 1 \times s = 50$

So $s = 10$.

Question 15

Examination report

a **i** 20°C **ii** 23.3%

b 97.5%

1 – 0.025 = 0.975

Some students were unable to apply the 68–95–99.7% rule.

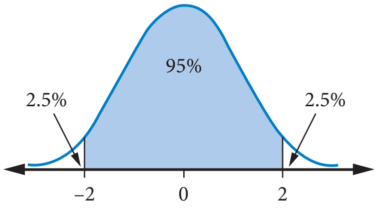
[VCAA 2012 2RCQ1]

EXAM PREP 2.4 Standardised values

Prep 1

a –2

b $z = -2$ means the teacher's height is two standard deviations below the mean.

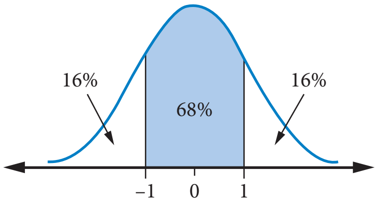


From the diagram, 97.5% of Year 12 teachers are taller people than this particular teacher.

Prep 2

a The z-scores are 1.2, –0.5, 3, –1, –1.2, so the student's best subject was Hospitality.

b Using the following diagram, the student was in the bottom 16% for Psychology and Systems Engineering.



EXAM PRACTICE 2.4 Standardised values

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1	7	3	80	6	4	2010
Q2	6	8	70	6	9	2010
Q3	4	77	12	4	2	2011
Q4	8	10	71	8	2	2011
Q5	72	5	3	6	13	2007
Q6	7	4	16	68	5	2012

EXAM PREP 2.5 Populations and samples

Prep 1

a population **b** population **c** population

d population **e** population

f sample (i.e. it's doubtful that every result on the Drivers Learner Permit Knowledge Test, stretching back for decades, would be included)

g sample **h** population

Prep 2

a $\sigma = 3.9$ **b** $s = 4.2$

Prep 3

All students will have different answers. Show your teacher your calculator screen.

EXAM PRACTICE 2.5 Populations and samples

Q1	Q2	Q3	Q4	Q5
C	B	C	E	A

CHAPTER 3

Associations between two variables

EXAM PREP 3.1 Explanatory and response variables

Prep 1

- **a** the amount of time spent in front of a screen
- **b** the amount of chocolate eaten
- **c** person's age
- **d** person's gender
- **e** the number of television sets in the home
- **f** levels of sleep deprivation

EXAM PRACTICE 3.1 Explanatory and response variables

Q1	Q2	Q3
C	E	C

EXAM PREP 3.2 Associations between two categorical variables

Prep 1

	Gender		
Sunscreen use	**Men**	**Women**	**Total**
Use sunscreen	35	30	65
Do not use sunscreen	5	10	15
Total	40	40	80

Prep 2

	Gender	
Library membership	**Male**	**Female**
Library member	71%	68%
Not a library member	29%	32%
Total	100%	100%

Since the male and female percentages are similar, the table suggests there is no association between gender and library membership.

Prep 3

People living in houses seem to own more refrigerators and freezers than those who live in flats and units. This segmented bar chart suggests that there may be an association between the number of refrigerators and freezers and the type of residence you live in.

EXAM PRACTICE 3.2 Associations between two categorical variables

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1	3	87	6	1	3	2005
Q2	3	1	0	3	93	2005
Q3	3	4	3	2	88	2009
Q4	3	84	7	2	4	2012
Q5	0	0	97	1	0	2004
Q6	13	10	7	4	67	2004
Q7	1	1	3	90	3	2002
Q8	2	2	9	2	85	2002
Q9	18	6	66	4	6	2002
Q10	2	10	80	2	5	2005
Q11	43	13	23	11	9	2005
Q12	1	2	16	79	1	2013
Q13	15	30	45	5	4	2013
Q14	4	11	8	70	7	2012
Q15	22	53	11	10	4	2003
Written response						
Q16	80					2008
Q17	Not reported					2011

Question 11

$$\text{Mean number of mobiles} = \frac{\Sigma xf}{\Sigma f}$$

$$= \frac{0\times34+1\times78+2\times30+3\times12}{154}$$

$$= 1.13$$

Question 13

$$\frac{3}{3+11+8} \approx 0.14 = 14\%$$

Examination report

From the table, the percentage of 'tall' mothers with 'short' daughters is given by

$$\frac{\text{number of tall mothers with short daughters}}{\text{total number of tall mothers}} \times 100\%$$

$$= \frac{3}{3+11+8} = 13.63\ldots\%$$

This answer rounds to 14% (option C).

The other common, but **incorrect**, response was 4% (option B). This response was obtained because the wrong base was used in calculating the percentage, as shown below.

$$\frac{\text{number of tall mothers with short daughters}}{\text{total number of tall mothers}}\times 100\%$$

$$=\frac{3}{82}\times 100\% = 3.65\ldots\%$$

This rounds to 4% (option B).

[VCAA 2013 1RCQ3&4]

Question 15

Number of standard sized houses = 14 + 71 + 47 =132.

47 of these have a high level of water usage.

$$\frac{47}{132}\times 100\% = 35.6\%$$

Question 16

Examination report

The percentaged segmented bar chart does support the opinion that lunch time activity (walked, sat or stood, ran) is associated with Year level. For example, the percentage who ran changed from around 78% to 40% to 10% from Years 6–8 and 8–10.

There are several ways of observing support from the bar chart. In general terms, an association is indicated by the percentage of girls undertaking a particular activity changing with the Year level. Note that this change does not have to be a consistent increase or decrease, but it can be. For example, focusing on the activity 'sat or stood', the percentage of students who sat or stood changed from around 2% to 24% to 68% from Years 6–8 and 8–10. Or, focusing on the activity 'walked', the percentage of students who walked changed from around 20% to 36% to 22% from Years 6–8 and 8–10.

Some students referred to the changes in the percentages for an activity but then did not comment on whether the segmented bar chart **supported the opinion** that the association exists. Others forgot to quote relevant percentages as required to support their argument.

Question 17

Examination report

a 65.5%

b The data for each of the four age groups in the table does the support the opinion that age at first marriage is associated with the year of marriage. For example, of all first marriages, the percentage of women aged 25–29 years increased from 23.4% (1986) to 31.7% (1996) to 34.5% (2006).

Many students tried to work down the table despite the direction to work across a row.

[VCAA 2011 2RCQ2]

EXAM PREP 3.3 Associations between numerical and categorical variables

Prep 1

a

Camera 1	Stem	Camera 2
9 9 9 7 5 3	6	
8 8 6 5 4 4 3 3 2 1 1	7	
9 3 3 3 0	8	
0	9	
2	10	
	11	2 6 6 7 9
	12	0 0 3 3 5 7
	13	0 0 0 1 1 4 5 6 8 9
	14	0 8
	15	
	16	9

Key: 6|3 = 63

b Both cameras are positively skewed.

c **Camera 1:** median = 74, range = 39, IQR = 11.5

Camera 2: median = 130, range = = 57, IQR = 15.5

d Yes, there is a significant difference between the two cameras. Camera 2 has recorded a much higher median speed than Camera 1. Camera 2 also has a greater range and IQR than Camera 1, indicating that the speeds recorded have a greater spread.

Prep 2

a Class A appears positively skewed. Class B appears negatively skewed. Class C is neither skewed nor symmetrical.

b **Class A:** median = 6, range = 5, IQR = 3

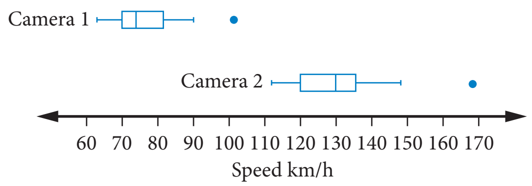
Class B: median = 8, range = 7, IQR = 3

Class C: median = 5, range = 6, IQR = 4.5

c Class B watched more television during the week than Class A. The median number of hours for Class B is 2 hours more than for Class A.

d Of the three classes Class C, not Class B, has the greatest variation amongst the middle 50% of students. The IQR for Class C is greater than for Class B.

Prep 3



Prep 4

a i 12 cm ii 26 cm

b The estimated lengths have a far greater range than the actual lengths.

c The estimated lengths have a far greater IQR than the actual lengths.

d Estimates: Positively skewed; Actual: Approximately symmetrical

e Four of the five-number summary results (min, Q_1, Q_3 and particularly the median) support the view that, in general, people tend to underestimate the length of string. However, the max clearly doesn't support this view. Therefore, the evidence doesn't support Yashneel's view that 'People always underestimate the length of a piece of string' because some people overestimate it.

EXAM PRACTICE 3.3 Associations between numerical and categorical variables

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1	3	1	1	94	1	2009
Q2	4	85	6	3	1	2009
Q3	85	10	2	1	1	2009
Q4	5	13	7	75	1	2006
Q5	1	74	3	1	22	2006
Q6	24	6	10	3	57	2006
Q7	11	15	66	3	5	2007
Q8	1	9	75	6	9	2007
Q9	18	8	2	62	10	2005
Q10	9	9	12	15	54	2011
	Written response					
Q11a	Not reported					2004
Q11b&c	93					2004
Q11d&e	51					2004
Q12a–b	65					2006
Q12c–f	60					2006
Q13	Not reported					2012
Q14	53					2008

Question 6

The temperatures 18 and 19 are outliers because they are less than $Q_1 - 1.5 \times \text{IQR}$.

Question 10

In 2000 over 75% of countries had an average pay rate over \$5.50 per hour. The median average pay rate in 1980 is around \$7.50 per hour. \$5.50 < \$7.50.

Question 11

Examination report

a 23.1

b 10.8

c

6	3
11	18

d Yes. 35.3% of men compared with 14.3% of women are overweight.

Required the comparison of two relevant, comparable and quoted percentages; one for males and the other for females. Students could not just quote raw numbers. No marks were given for 'yes' without any justification. No marks were given for comparing fractions that did not at least have common denominators.

Incorrect answers appeared to result from misinterpretation of the information. Some students ignored the data and expressed their own observations of mass distribution according to gender. Other incorrect answers used fractions such as $\frac{6}{11}$ or $\frac{3}{18}$, or their percentage equivalent.

e i interquartile range and median

Most students gave *median* as one answer. The most common incorrect term was *range*.

ii 23.9

Generally well done, but the most common error came from incorrect rounding techniques. Many students appeared to find the correct number of 23.947 which they rounded off to 23.95, with 2 decimal places. As this number is not **correct to 1 decimal place** students were not awarded the mark for the question. Some did further round 23.95 up to 24.0.

iii The median gives the better indication of the typical BMI as it is not as affected by extreme values as the mean.

This question was not done well as students did not generally appreciate the effect of extreme values on the mean. Nominating the median without some appropriate justification was not sufficient to gain a mark. The most common incorrect response suggested that the *mean* gave the better indication as 'the *mean* takes all data values into account while the *median* is only the middle number'. Another common incorrect interpretation was 'the *mean* gives the average but the *median* is only the midpoint'.

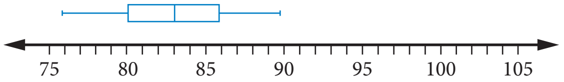
Most students who correctly named the *median* then inappropriately referred to it not being affected so much by *outliers*. This answer scored only one mark as there was no information in the boxplots to suggest the presence of any *outliers*, although *extreme values* were evident.

[VCAA 2004 2RCQ2]

Question 12

a 3.8 b −1.4 c 97.5%

d



e 89.5

f The median height increases with age.

Question 13

Examination report

a The wind direction with the lowest recorded wind speed was south-east.

The wind direction with the largest range of recorded wind speeds was north-east.

b 2, 2, 2, 3, 4, 4, 4, 4

Few students answered this question correctly.

Often, incorrect answers included 3.5 as a value, while others included numbers less than 2 or greater than 4.

The eight wind speeds were recorded correct to the nearest whole number and, since the median = 3.5, the fourth and fifth values either side of this median must be 3 and 4. Further, the lack of whiskers on the relevant boxplot means that the minimum value and the first quartile (Q_1) coincide and thus are both 2. Similarly, the maximum value and Q_3 are both 4.

This means the data set must look like {2, _, _, 3, 4, _, _, 4}.

Since $Q_1 = 2$ and Q_1 represents the midpoint of the second and third data points in this group of eight, both these data points must be 2.

Similarly, since $Q_3 = 4$, the sixth and seventh data points must both be 4.

Hence, the data values are 2, 2, 2, 3, 4, 4, 4, 4.

The available mark was awarded for these eight numbers written in any order.

[VCAA 2012 2RCQ3]

Question 14

Examination report

a 124, 148

b The median *arm span* increases with Year level

The IQR of *arm span* decreases with Year level or the range of *arm span* decreases with Year level

The majority of students erroneously said that the boxplots showed that *arm span* was increasing and explained that this is to be expected as the students are still in a growing stage. To compare the boxplots, one appropriate summary statistic for *arm span* (response variable) had to be compared across the Year levels (explanatory variable). This could have been the median, the range or the IQR.

Some responses seemed to relate to students' own environment and suggested that *arm span* increased with age. The question did not state that an association between Year level and age existed for this data and therefore this could not be assumed without some further information about the sample from which the data had been taken.

c $1.5 \times \text{IQR} = 1.5 \times 10 = 15$

The lower fence is at $Q_1 - 15 = 160 - 15 = 145$

An actual *arm span* of 140 is still lower than this and so is still an outlier.

Many students correctly calculated the lower boundary (fence) for $Q_1 - 1.5 \times \text{IQR}$ and then said that 'it was therefore still an outlier' without any numerical comparison shown or suggested. As this does not explain what 'it' is, nor why 'it' is **still** an outlier, the actual value of 140 cm should be stated as still being lower than the calculated value of the fence. The incorrect value of 147 for the lower 'fence' was given by some students.

[VCAA 2008 2RCQ3]

EXAM PREP 3.4 Associations between two numerical variables

Prep 1

a Positive association, because mass usually increases as height increases.

b No association, because salary level has no effect on a person's heart rate.

c Negative association, because sales usually decrease as car prices increase.

d Positive association, because exam scores usually increase as study effort increases.

e positive
f negative
g no association
h positive
i negative
j no association
k positive
l positive

Prep 2

a negative, linear and strong

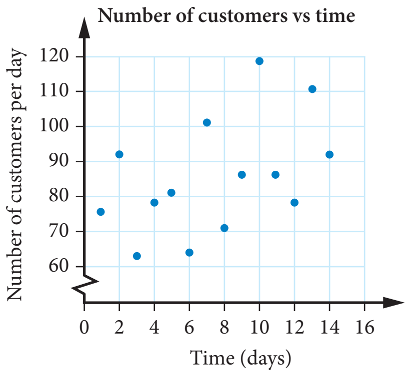
b no association

c negative, linear and weak (Close to no association.)

d positive, linear and moderate

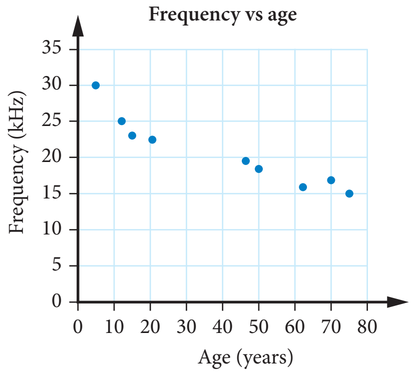
Prep 3

a



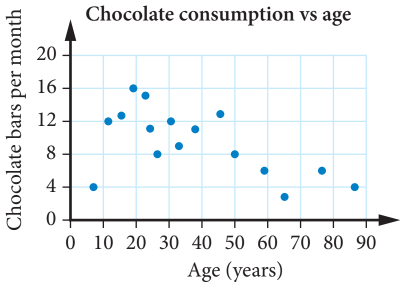
The association between the time and number of customers per day can be described as positive and weak.

b



The association between age and frequency can be described as negative, linear and strong.

c



The association between age and chocolate consumption can be described as negative, linear and moderate.

EXAM PRACTICE 3.4 Associations between two numerical variables

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1					▨	
Q2			▨			
Q3				▨		
Q4	▨					
Q5	23	17	12	10	37	2007

Question 5

The two variables involved are both categorical. The only option that displays two categorical variables is the percentaged segmented bar chart. A scatterplot is used for two numerical variables. Back-to-back stem plots and parallel boxplots are used when one variable is categorical and the other is numerical. A histogram is used when only one numerical variable is involved.

EXAM PREP 3.5 The Pearson correlation coefficient

Prep 1

$r = 0.6484$ This indicates that there is a moderate, positive linear association between yearly income and the number of movies seen in the past year. There is some evidence to suggest that cinema patronage should increase as yearly income increases.

Prep 2

a −0.4
b 0.8
c 0.6
d −0.9
e −0.5
f 0.2

Prep 3

a A weak positive linear association. There is limited evidence that the amount of exercise that a person does increases as their height increases.

b A moderate positive linear association. There is some evidence to support that a person's mass should increase as the number of hours a person spends sitting down increases.

c A strong negative linear association. It can be concluded that the percentage of good peaches should decrease as the time spent in storage increases.

d A moderate positive linear association. There is some evidence to support that sales at a local store should increase as temperature increases.

e No association between the number of pets owned and temperature.

f A weak negative linear association. There is limited evidence to suggest that the number of home cooked meals will decrease as income increases.

EXAM PRACTICE 3.5 The Pearson correlation coefficient

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1		▨				
Q2				▨		
Q3	▨					
Q4	5	4	5	6	79	2013
Q5	8	6	14	66	6	2002
Q6	18	13	6	6	58	2008

Question 6

When the correlation coefficient has a negative value, increase in the explanatory variable is associated with a decrease in the response variable.

EXAM PREP 3.6 Cause and effect

Prep 1

- **a** causation
- **b** causation
- **c** common response (lurking variable would be population)
- **d** coincidence
- **e** confounding (there may be a causation, but there are too many potential confounding variables that could be involved, such as how much the parents value education, the type of books, the type of academic studies undertaken)
- **f** coincidence
- **g** common response (lurking variable would be the size of the fire)
- **h** common response (lurking variable would be the temperature)
- **i** confounding (there are too many other potential variables to be certain, e.g. increasing population, movement to the city, change of government policies, the economy and unemployment, etc.)

Prep 2

It's likely that there is confounding in this association. The secondary school results may have caused the university results, but we can't know because there are too many potential confounding variables that could be involved, such as innate academic ability, the capacity for hard work, economic factors or parental influence.

EXAM PRACTICE 3.6 Cause and effect

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1	10	23	2	62	3	2011
Q2		▨				
Q3	▨					
Q4	▨					
Q5				▨		
Q6			▨			
Q7	▨	▨	▨	▨	▨	

Question 7

- **a** Head lice caused people to be healthy.
- **b** observation
- **c** The sickness caused lice to leave.
- **d** The people of the Hebrides had their explanatory and response variables around the wrong way in the association. They believed that the explanatory variable was the head lice and the response variable was the fever, when in fact the explanatory variable was the fever and the response variable was the head lice.
- **e** Select a group of people without lice who are sick with fever and randomly introduce lice to the hair of half of them. Monitor the fevers of both groups and record how long it takes them to recover. It would be unethical to add to sick people's problems in this way to test such a doubtful proposition. Alternatively, select a group of people with head lice and introduce a sickness that causes fever to half of them. This is also unethical for the same reasons.

CHAPTER 4 Linear associations

EXAM PREP 4.1 The least squares line of best fit

Prep 1

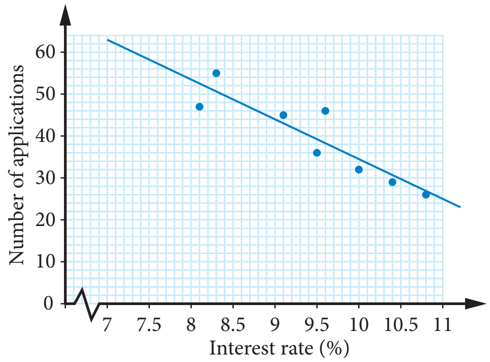
a	**i**	7.42	**ii**	7.4
b	**i**	12.92	**ii**	13
c	**i**	0.36	**ii**	0.36
d	**i**	1800.00	**ii**	1800
e	**i**	20.67	**ii**	21
f	**i**	72.00	**ii**	72
g	**i**	9.80	**ii**	9.8

Prep 2

number of bushfires = 41 − 1.1 × amount of rainfall

Prep 3

- **a** number of applications = 130 − 9.6 × interest rate
- **b**



Prep 4

- **a** The slope is 0.7. This means mass increases by 0.7 kg for every 1 cm increase in height.
- **b** The *y*-intercept is –44. This means that mass is –44 kg when height is 0 cm. Negative mass is clearly impossible. This least squares line of best fit only applies from a certain minimum height.

EXAM PRACTICE 4.1 The least squares line of best fit

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1						
Q2						
Q3	73	6	6	3	11	2012
Q4	11	10	71	4	2	2005
Q5	13	47	9	7	23	2005
Q6	49	22	10	13	6	2009
Q7	70	6	6	9	9	2009
Q8	54	18	15	6	6	2006
Q9	51	11	12	15	11	2010
Q10	8	25	10	39	18	2002
Written response						
Q11	50					2008

Question 5

The slope of the equation = 0.96. This tells us how much the wingspan increases with each one metre increase in length.

Question 6

x is sleeping time, y is life span. Using a CAS/calculator, we get $y = 38.9 - 2.36x$.

Examination report

This question had a success rate of 49 per cent. A further 22 per cent of students chose option B, making the common error of automatically assuming (incorrectly in this case) that the first variable appearing in the table was the explanatory variable.

[VCAA 2009 1RCQ9]

Question 8

$$b = r\frac{s_y}{s_x} = -1.5177,\ a = \bar{y} - b\bar{x} = 30.85$$

Question 9

$$y = a + bx \text{ where } b = r\frac{s_y}{s_x} = -0.673\,464 \text{ and}$$

$$a = \bar{y} - b\bar{x} = 9.169\,36$$

So $y \approx 9.2 - 0.7x$

Question 10

Examination report

This question required fitting a least squares line of best fit to a set of bivariate data. The most commonly chosen alternative (D) indicated that many students were unaware of the importance of identifying the explanatory variable (in this case, waist measurement) and response variable (mass) as part of the process of calculating the equation of a least squares line of best fit. When determining the equation of the least squares line, it **cannot** always be assumed that the first variable listed is the explanatory or x variable. Given that this is the second year in a row that students have made the same error, teachers need to highlight this during review work.

[VCAA 2002 1CQ9]

Question 11

Examination report

a height

The question required students to find 'a linear equation that allows *arm span* to be predicted from *height* and so, for this equation, *arm span* (response variable) depends on *height* (explanatory variable).

b *arm span* = −16 + 1.1 × *height*

The variables *arm span* and *height* were required in the equation rather than y and x.

A common incorrect equation was found by choosing the incorrect explanatory variable.

Some students rounded the coefficients of their equation incorrectly.

There seemed to be much confusion about the term 'coefficient' as many answers consisted of, or included, the values of the Pearson correlation coefficient and the coefficient of determination. The question referred to 'writing an equation' with 'the coefficients rounded to two significant figures.' An equation (as in this question) is made up of variables (such as *height* or *arm span*) and coefficients (the values of a and b in the line of best fit equation $a + bx$ found from the calculator).

c *Arm span* increases by 1.1 cm for each 1 cm increase in *height*.

[VCAA 2008 2RCQ4]

EXAM PREP 4.2 The coefficient of determination

Prep 1

a Height is the explanatory variable. This gives $r^2 \approx 0.687$.

b 68.7% of the variation in shoe size can be explained by the variation in height.

c shoe size = −22 + 0.17 × height

d Yes, this suggests it is an appropriate model due to the high r^2 value.

Prep 2

$r = 0.94$

Prep 3

$r = -0.85$

EXAM PRACTICE 4.2 The coefficient of determination

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1	3	89	5	2	1	2003
Q2	33	16	12	16	23	2003
Q3	72	12	8	6	3	2004
Q4	8	39	38	11	4	2004
Q5	2	3	4	24	67	2011
Q6	57	9	19	8	7	2011
Q7	58	16	10	8	6	2011
Q8	13	6	55	23	3	2007
Q9	6	16	21	51	5	2007
Q10	15	52	28	2	3	2002
Q11	11	7	23	9	49	2013
Extended response						
Q12a	90					2004
Q12b&c	Not reported					2004
Q13	Not reported					2009
Q14	43					2006

Question 2

$r^2 = 0.8198$, so $r = \pm 0.91$. The least squares line of best fit has a negative slope, therefore $r = -0.91$.

Examination report

This question required students to make use of both numerical and graphical information to arrive at the correct answer. First, they needed to know that if $r^2 = 0.8198$ then $r = \pm\sqrt{0.8198}$, and second, that additional information was required to determine which sign to use. The scatterplot clearly showed that the association represented negative correlation so that $r = -0.91$, correct to 2 decimal places. Only 33 per cent of students were able to follow this process through. The fact that 23 per cent of students chose $r = 0.91$ suggested that they may have been unaware that the equation $r^2 = 0.8198$ had two solutions.

[VCAA 2003 1RCQ9]

Question 4

The coefficient of determination = $(-0.9260)^2 = 0.8575$

Examination report

This question required students to find the value of the coefficient of determination given the correlation coefficient of $r = -0.9260$. Only 38% of students obtained the correct response, which was 0.8575 (option C). The majority of students (77%) realised that the coefficient of determination is given by r^2; however, 39% arrived at the **incorrect** solution (−0.8575, option B) because of an apparent inability to correctly square a negative number using a calculator.

This error may have been averted if students realised that the square of a negative number is positive.

The use of graphics calculators is assumed in Further Mathematics examinations. Examiners expect that students will know how to use the technology at a level sufficient to enable them to perform all the basic arithmetic operations required in the course.

[VCAA 2004 1RCQ9]

Question 6

The second column of data is y, the response variable. The third column of data is x, the explanatory variable. Use a CAS/calculator to get $y = 70.3 + 0.790x$.

Question 7

The coefficient of determination $r^2 = 0.258$, so 25.8%.

Question 8

$r^2 = 0.9034^2 = 0.816 = 81.6\%$

Question 9

The slope of the line is 0.87, so there is 0.87 mm increase in *length* for every 1 mm increase in *diameter*.

Question 10

The coefficient of determination measures the predictive power of the association and it is affected by outliers. Removing an outlier increases the predictive power of the least squares line, so the coefficient of determination would increase.

Question 11

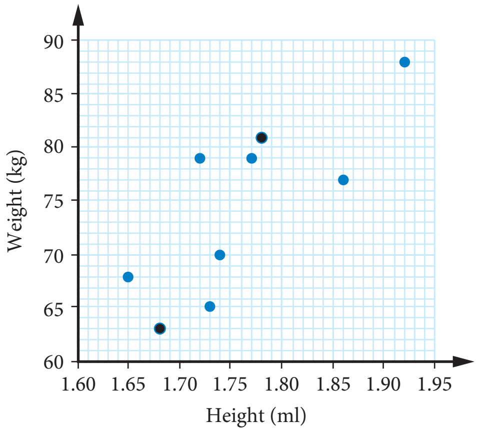
Population density $r^2 = (-0.563)^2 \approx 0.317 = 31.7\%$

House size $r^2 = (0.357)^2 \approx 0.127 = 12.7\%$

Question 12

Examination report

a



As this was the first question and only one mark was allocated for two points, the mark was awarded if either point was correctly plotted.

b –60.8, 76.8

Some students reversed these numbers and indicated some confusion in interpreting their calculator output. One of the two marks was then awarded as long as both numbers were correct and the negative sign preceded 60.8. Some gave integer answers only.

c mass, height

Some reversed these and gained no marks.

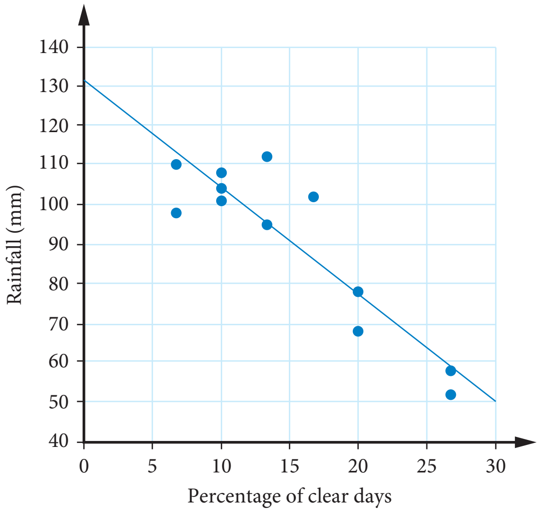
In a few cases, students referred to the *response variable* or *explanatory variable*. This was not accepted unless one of the terms was specifically, and correctly, written as *mass* or *height*.

[VCAA 2004 2RCQ1]

Question 13

Examination report

a



By choosing two points that were close to each other, many students were unable to plot this line within an acceptable accuracy. Some students apparently failed to see this question and left it unanswered. Students are expected to draw straight lines with the aid of a ruler or other straight edge. Freehand drawings are inaccurate and will generally lead to errors that are penalised.

b 37.2 mm, correct to 1 decimal place

c **i** 80.81% of the variation in the *rainfall* can be explained by the variation in the *percentage of clear days*. The coefficient of determination refers to the association in which the variation in the response variable can be explained by the variation in the explanatory variable. It does not imply any causation between the two variables. Unacceptable causation terms for this association include 'is determined by', 'is due to', 'is caused by' and 'is because of'. Other common errors included students reversing the variables, incorrectly suggesting that 80.81% of the variation in percentage of clear days (explanatory value) was explained by the variation in the rainfall (response value). Others incorrectly referred to the absolute values of one or both of the variables rather than to the variation in them. It was apparent that some students transcribed material from their bound reference notes and gave inappropriate answers that made reference to variables such as age, cost or profit.

ii –0.899, correct to 3 decimal places. Many students failed to include the negative sign as indicated by a generally negative slope of the data in the graph.

[VCAA 2009 2RCQ3]

Question 14

Examination report

a 0.53

b **i** 56.9%

Many students did not convert the decimal into a percentage as required and gave an answer of 0.6. Some who did the conversion rounded the decimal first before converting to an incorrect answer of 60%.

ii 56.9% of the variation in height is explained by the variation in age.

Common incorrect interpretations referred to **causation**. A common incorrect answer was '56.9% of the variation in height is **due** to the variation in age.'

[VCAA 2006 2RCQ2]

EXAM PREP 4.3 Making predictions

Prep 1

a 8.95 laps **b** 5.43 laps **c** 2.07 laps

Prep 2

a 2°C **b** 8°C **c** 18°C

Prep 3

a 145 grams, interpolation

b 45 grams, extrapolation

c 370 grams, extrapolation

d The prediction for 25 chocolates is the most reliable as it is inside the data range of 15 to 50; the other two predictions involve extrapolation.

e 21 chocolates

EXAM PRACTICE 4.3 Making predictions

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1						
Q2						
Q3						
Q4	25	10	62	2	1	2008
Q5	2	10	55	11	21	2008
	Written response					
Q6	58					2010

Question 5

Looking at the scatterplot, find the 6th and 7th lowest masses. These are 430 and 450. The median mass will be between them, i.e. 440.

Question 6

Examination report

a Male income

b $350

Common incorrect answers included $0.35, $13 000 and $13 000.35.

c **i** $18 250

ii Making the required prediction involved going beyond the data used (extrapolation) to determine the line of best fit equation.

There was no evidence that the association continued into income levels below those given in the dot plot. Therefore, the mathematical model established by the line of best fit equation cannot be relied upon outside the range of the presented data set.

This question was poorly answered. Many students referred to their own view of the real world rather than the given data. The suggestions that 'Female income is not dependent on male income' or that 'it is a known fact that female income is always lower due to discrimination' were common. Incorrect or unacceptable answers included 'male income must be higher than the female income because that is what the graph shows' or 'any prediction is unreliable, whether or not it is derived from a formula.'

[VCAA 2010 2RCQ2]

EXAM PREP 4.4 Residual analysis

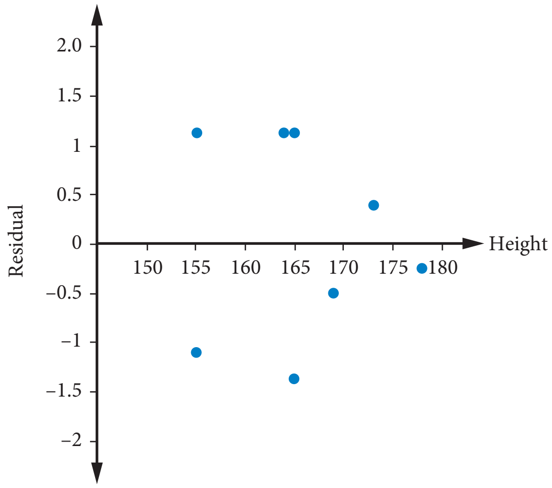
Prep 1

a 0.5 metres

b −0.7 metres

Prep 2

Height (cm)	Femur length (cm)	Residual
178	50.2	−0.3
173	48.4	0.4
165	45.1	1.1
164	44.6	1.1
168	45	−0.5
165	42.6	−1.4
155	39.9	0.8
155	38	−1.1



The data is probably linear because the residual values appear randomly scattered above and below the *x*-axis.

Prep 3

The data value at $x = 11$ is below the least squares line of best fit, which means the residual must be negative. The residual plot has a positive residual value for $x = 11$.

EXAM PRACTICE 4.4 Residual analysis

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1	69	8	9	7	6	2013
Q2	6	73	14	3	3	2010
Q3	8	35	35	17	5	2010
Q4	15	61	9	6	9	2010
Q5	9	10	19	51	11	2012
Q6	4	11	9	28	47	2009
Q7*	6	10	33	10	41	2006
	Written response					
Q8a	90					2007
Q8bi–8e	44					2007
Q9a	60					2013
Q9b&c	50					2013
Q10a-c	60					2003
Q10d-h	41					2003
Q11	40					2012

Question 3

Enter the data values from the scatterplot into a CAS/calculator to calculate the equation of the least squares line of best fit: (20.5, 170.5), (23.5, 165), (23.5, 167.5) etc. Note that the y-intercept is *not* 167 because the horizontal axis doesn't start at zero.

Question 5

Since $Q_3 \geq Q_1$, $\text{IQR} = Q_3 - Q_1 \geq 0$

Question 6

The body mass for a baboon is 10.55. The brain mass is $49.4 + 2.68 \times 10.55 = 77.674$.

Residual = actual – prediction = $179.5 - 77.674 \approx 102$

Examination report

In this question, students were given the body masses and brain masses of nine animals, and the equation to a least squares line of best fit determined from this data. The correct response was chosen by 47 per cent of students. Students were expected to first use the equation of the line of best fit to determine the predicted brain mass for the designated animal (the baboon) and then use this value, along with the animal's actual brain mass, to determine the residual value (the error of prediction). A significant number of students, 28 per cent, correctly completed the first part of the task – correctly calculating the predicted brain mass – but did not go any further and this led them to choose option D.

[VCAA 2009 1RCQ11]

Question 7*

Residual = actual – predicted = $67 - (-20 + 1.11 \times 80) \approx -2$

Examination report

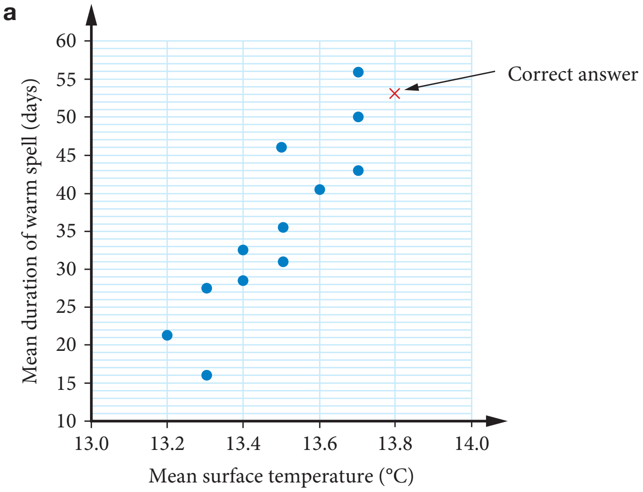
*It should be noted that due to a misprint in the paper for this question, involving the units for the answers, all options were credited as correct.

[VCAA 2006 1RCQ8]

Question 8

Examination report

a

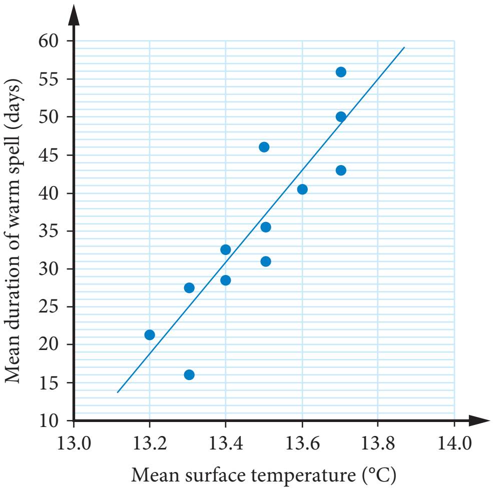


This point had to be on the 13.8 coordinate line, with the value for the duration of the warm spell of 53.1 (fractionally above 53).

b **i** mean duration = –776.9 + 60.3 × temperature

The appropriate variables were required and $y = -776.9 + 60.3x$ received only one mark. Some students got the variables back to front.

ii



A straight line that had to pass through the points (13.2, 19) and (13.8, 55)

A significant number of students missed this question entirely. This highlights the need for students to use their reading time effectively.

To draw a straight line, students are expected to bring a straight edge (for example a ruler) into the examination.

The line was very often badly plotted with two close points apparently chosen to locate the line. It is expected that the points be a reasonable distance apart. Values could have been calculated for mean surface temperatures of 13.2°C and 13.8°C.

Some students incorrectly used the point plotted in part a.

c There is no clear pattern to the random scatter of points.

d 83%

e strong, positive and linear

A common error was to analyse the **residual** plot for strength, direction and form, incorrectly concluding that there was no association between the variables.

[VCAA 2007 2RCQ3]

Question 9

Examination report

a 1.8

$$z = \frac{y - \bar{y}}{s_y} = \frac{91 - 85.6}{2.99} = 1.806\ldots$$

9780170371100

b For the equation $y = a + bx$, the value of

$$b = r\frac{s_y}{s_x} = 0.488 \times \frac{2.99}{5.37} = 0.2717\ldots \approx 0.272 \text{ and}$$

$$a = \overline{y} - b\overline{x} = 85.6 - 0.2717\ldots \times 15.7 = 81.334 \approx 81.3$$

c −2.2

predicted = 81.3 ÷ 0.272 × 14.3 = 85.1896…

∴ residual = 83 − 85.1896… = −2.1896…

Most student correctly found the predicted value but did not continue to find the residual value as required.

[VCAA 2013 2RCQ3]

Question 10

Examination report

a mean = \$6.51, standard deviation = \$3.31 for one mark each.

Most students had no difficulty with finding the correct mean. Some incorrectly used the data for Number of calls instead of Cost. A common incorrect value given for the mean was \$6.50. The standard deviation was very poorly done.

Many students seemed to have little idea of what was required with some quite unlikely answers given despite having the mean correct. Some students incorrectly quoted the population standard deviation rather than that for the sample. A standard deviation of \$3.30 was considered a round-off error.

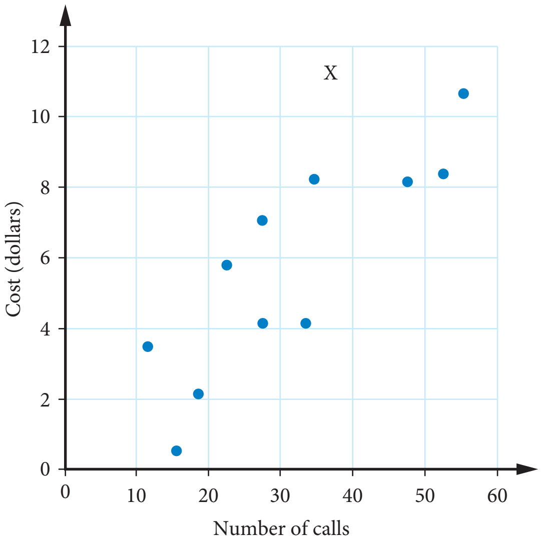
b 75%

c **i** 0.19 **ii** 0.8054

d linear, accept ‘straight’

Unacceptable answers included ‘positive’, ‘strong’ or ‘constant’.

e



The point had to be located in the upper right quarter of the correct square. Many placed the point at the halfway mark, either vertically or horizontally (this was not accepted) or left this question.

f Cost, accept ‘dollars’. A small percentage incorrectly quoted ‘number of calls’.

g **i** 19

This number had to be the coefficient of part c i converted to cents for a consequential mark. An answer of 0.19 was not accepted. This answer often did not match the coefficient stated in part c i.

ii 65%

Many did not seem to relate this to the value of r in part c ii.

h \$1.38 (accepted 1.38)

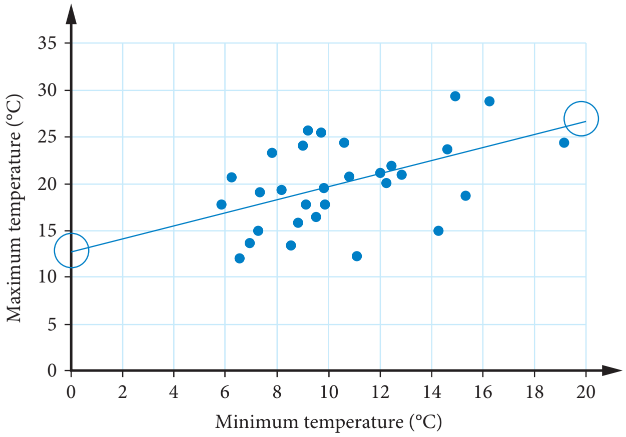
Many simply quoted the predicted cost of \$7.12 without then calculating a residual value. A mark was given here for calculating an appropriate residual value using their equation.

[VCAA 2003 2RCQ1]

Question 11

Examination report

a A line should have been drawn on the graph through (0, 13) and (20, 26.4), circled on the following graph.



Some students did not understand that this question required a line to be drawn on the scatterplot. Many students who attempted to draw the line did so with insufficient accuracy, particularly with regard to the need to pass through or close to the point (20, 26.4).

b On average, it is predicted that, when the minimum temperature is 0°C, the maximum temperature will be 13°C.

c Moderate, positive

d On average, it is predicted that the maximum temperature increases by 0.67°C for each 1°C increase in the minimum temperature.

Some students gave the response and explanatory variable names in reverse order.

e 40%

f −8°C

Predicted value = 13 + 0.67 × 11.1 = 20.437…

Residual value = 12.2 − 20.437 … = −8.237…

The negative sign on the residual was required.

This question was often left unanswered.

Many students incorrectly calculated the residual as 12.2 − 11.1 = 1.1 ≈ 1.

[VCAA 2012 2RCQ2]

EXAM PREP 4.5 Data transformation

Prep 1

a x^2, y^2

b None

c $\frac{1}{x}$, $\log(x)$, $\frac{1}{y}$ or $\log(y)$

d x^2, $\frac{1}{y}$ or $\log(y)$

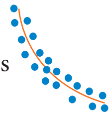
e None

f $\frac{1}{x}$, $\log(x)$, y^2

Prep 2

a 10 seeds **b** 8 seeds **c** 7 seeds

Prep 3

The shape of the data is , so the options to linearise the data are $\frac{1}{x}$, $\log(x)$, $\frac{1}{y}$ or $\log(y)$.

For the $\frac{1}{x}$ transformation

$\frac{1}{x}$	1	0.5	0.333	0.25	0.2
y	28	18	6	3	2

$r^2 = 0.93$

For the $\log(x)$ transformation

$\log(x)$	0	0.301	0.477	0.602	0.699
y	28	18	6	3	2

$r^2 = 0.97$

For the $\frac{1}{y}$ transformation

y	1	2	3	4	5
$\frac{1}{y}$	0.036	0.056	0.167	0.333	0.5

$r^2 = 0.94$

For the $\log(y)$ transformation

x	1	2	3	4	5
$\log(y)$	1.447	1.255	0.778	0.477	0.301

$r^2 = 0.98$

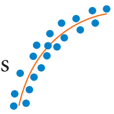
The coefficient of determination of the $\log(y)$ transformation is closest to 1, so $\log(y)$ is the best choice for linearising the data.

The equation of the least squares line of best fit is $\log(y) = 1.77 + -0.31x$

EXAM PRACTICE 4.5 Data transformation

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1	8	15	59	10	7	2009
Q2	2	53	10	3	32	2013
Q3	12	9	52	17	10	2003
Q4	26	9	9	49	7	2007
Q5	42	26	18	7	5	2006
Q6	11	31	11	11	35	2010
Written response						
Q7a	80					2005
Q7b-d	50					2005
Q7e&f	45					2005
Q7g&h	70					2005
Q7i&j	25					2005
Q8	Not reported					2012
Q9	33					2011
Q10	17					2008

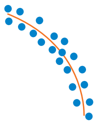
Question 1

The shape of the data is . Using the transformation circle, the options to linearise the data are $\frac{1}{x}$, $\log(x)$ or y^2, where x = *Internet use* and y = *mathematics achievement*. So the least squares line of best fit equation choices must involve $\frac{1}{\text{Internet use}}$ or log (*Internet use*) or (*mathematics achievement*)^2.

Question 2

$(width)^2 = 1.8 + 0.8 \times 120 = 97.8$. So $width = \sqrt{97.8} \approx 9.9$

Question 3

The shape of the data is . Using the transformation circle, the options to linearise the data are x^2 or y^2. So, from the choices given, the correct answer is an x^2 transformation.

Question 4

Use your CAS/calculator to find the least squares line of best fit in terms of x and $\log(y)$.

Question 5

Use your CAS/calculator to find the least squares line of best fit in terms of x^2 and y.

Question 6

Use your CAS/calculator to find the least squares line of best fit in terms of $\frac{1}{x}$ and y. The slope of the line is approximately 249.

Question 7

Examination report

a value

b **i** −0.952

$\sqrt{0.9058} = 0.9517$ was a frequent incorrect answer; however, the negative sign was necessary.

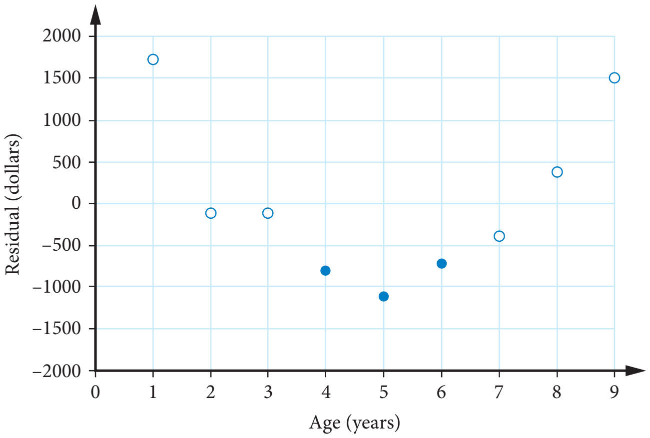
ii 91%, 90.6% and 90.58% were all accepted.

Common incorrect answers quoted the value of r as 95.2%.

c 17 500, −1200

Many students reversed these numbers. These responses scored a maximum of one mark so long as the negative sign was present in −1200.

d



The first mark was given for three points that were on the lines for *Age* = 4, 5 and 6 and were all below the horizontal axis. The second mark was given if the outer two points were within *Residual* = −500 and −1000 and the middle point was within *Residual* = −1000 and −1500.

e Yes: the residual values show a pattern of a hill or mountain shape. This indicates that the association between the original two variables x and y is probably non-linear.

The explanation had to mention that a pattern was evident.

f 0.95

Common incorrect answers were 0.91 and 0.92, which were apparently achieved from looking at some sequence obtained from differences along the log (*age*) row of figures.

g negative, linear and strong

h \$13 100

A common error was for students to disregard the *log* part of the equation and calculate 18 300 − 10 800 × 3 = −14 100.

i the random placement of dots

Several incorrect answers referred only to the points being ‘scattered’. Others mentioned that ‘dots were on both sides of the line’ or ‘the horizontal line is in the middle of the plot’.

j Either $\frac{1}{y}$ or $\log(y)$ (only one was needed). The points closer to the vertical (value) axis need to be compressed more toward the horizontal axis to linearise the data.

Many students answered this poorly. Transformation involving x was common. Vague references to ‘straightening the points’ gave no indication that the students understood which points were affected the most and in which direction. An explanation such as ‘you always do this transformation to a graph shaped this way’ was insufficient.

[VCAA 2005 2RCQ1]

Question 8

Examination report

a 3.4, 6.6

Many students did not answer this question. Some students who did answer this question did not seem to know how to deal with the squared transformation of the response variable.

b 13 km/h

$(ws3pm)^2 = 3.4 + 6.6 \times 24 = 161.8$

So $ws3pm = \sqrt{161.8} = 12.720\ldots$

Many students who performed the correct transformation and got the correct coefficients for the equation in part a gave an incorrect rounded answer of 162. They needed to find the square root to obtain the value of *ws3pm* rather than leave it as $(ws3pm)^2$.

[VCAA 2012 2RCQ4]

Question 9

Examination report

a 2.39, 5.89

Few students found the two answers correctly by correct data entry and log transformation. A common mistake was to ignore the log transformation requirement. For some students it did not register that the response variable was in the first column of the table and hence calculated the line of best fit incorrectly. These students could still access one of the two marks if they then correctly applied a log transformation to the wrong variable.

b 27.7 years

$average\ age = 2.39 + 5.89 \times log(20\ 000) \approx 27.7$

Very few students were able to answer this question correctly as most of those who attempted it did not find log (20 000) in their calculation. Very commonly, answers were inappropriate in the context of the question. For example, many answers were in the thousands, millions or were written as dollars. Answers such as these could not apply to a question about the average age at first marriage.

[VCAA 2011 2RCQ4]

Question 10

Examination report

a $\text{homework hours} = 5.12 + \dfrac{102.90}{\text{TV hours}}$

The majority of students did not seem to understand the term 'reciprocal'.

Of those students who did the correct transformation on their calculator, few were then able to correctly write the equation with their correct coefficients. A common incorrect answer from this group failed to write *TV hours* as a denominator of a fraction. An equation written as 'homework hours = 5.12 + 102.90 (reciprocal) TV hours' was not accepted for full marks.

A common error involved interpreting 'reciprocal' transformation on a 'log' transformation.

The values of the correlation coefficient and coefficient of determination were given as common incorrect responses. Again, the real variable names were required in the equation rather than x and y.

b 13.7 hours

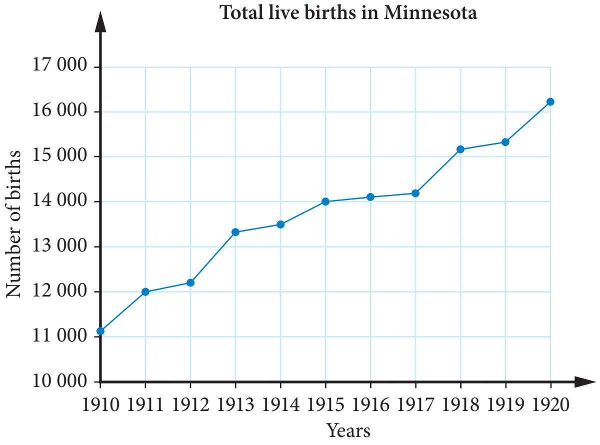
Some unreasonable answers were given for this question without any comment from students that the answer was absurd. For instance, an answer of 1239.9 hours of homework per week would be rather exhausting, and impossible with only 168 hours in a week. Equally, a negative number of hours would also present some difficulties for a student's real study program.

[VCAA 2008 2RCQ5]

CHAPTER 5 Time series

EXAM PREP 5.1 Time series plots

Prep 1



Prep 2

a	Yes	**b**	Yes	**c**	No
d	No	**e**	Yes	**f**	No
g	Yes	**h**	Yes		

Prep 3

a decreasing trend

b decreasing trend and seasonality

c increasing trend

d outlier and structural change

EXAM PRACTICE 5.1 Time series plots

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1	1	1	3	5	91	2007
Q2			▨			
Q3					▨	
Q4	3	5	31	59	2	2012
Q5	31	54	9	3	2	2004
Q6	39	25	13	18	5	2003
Written response						
Q7a	Not reported					2011
Q8	Not reported					2002

Question 4

There are two variables involved, and the explanatory variable is time, so the most appropriate graph is a time series plot.

Question 5

Examination report

This question, which had a 31% success rate, required students to interpret a time series plot displaying the changing price of shares in two companies over a period of time. The price for each share showed an increasing trend over the time period considered. However, the prices of the shares increased at different rates, so that the **difference** in price between the two shares showed a decreasing trend. Students needed to recognise this from the graph to obtain the correct response. The majority of students (54%) gave the incorrect response ‘increasing trend’.

[VCAA 2004 1RCQ12]

Question 6

This involves two variables so it can’t be a histogram. One of the variables (resting pulse rate) is numerical and the other (fitness level) is categorical. The best way of displaying an association between a numerical variable and a categorical variable with two or more levels (in this case 3 levels: below average, average, above average) is through parallel boxplots.

Examination report

The very low success rate of 18 per cent in answering Question 6 indicates that students generally were unfamiliar with the use of parallel boxplots as a method for displaying an association between a numerical variable and a two or more level categorical variable. This knowledge is an explicitly stated requirement of the study design. The fact that 39 per cent chose the ‘histogram’ option, a display used when only one variable is involved, also suggests that many students failed to recognise the bivariate nature of the situation.

Question 7a

Examination report

The average age at first marriage remained relatively constant from 1915 to around 1935, and then decreased during the period 1935 to 1970.

This question was generally answered very poorly. Many students noted a ‘decreasing trend from 1915 to 1970’, without considering the two sections of the plot. Some gave very specific answers rather than answering in general terms. Many students said the data was either ‘positively skewed’ or ‘negatively skewed’, neither of which is an applicable expression for a time-series plot.

[VCAA 2011 2RCQ3]

Question 8

Examination report

a Year

‘Time’ was accepted here but most students obtained the correct answer.

b **i** \$10 990

One mark for 10.99

Many students found an answer of 10.99, not realising this indicated **thousands** of **dollars**.

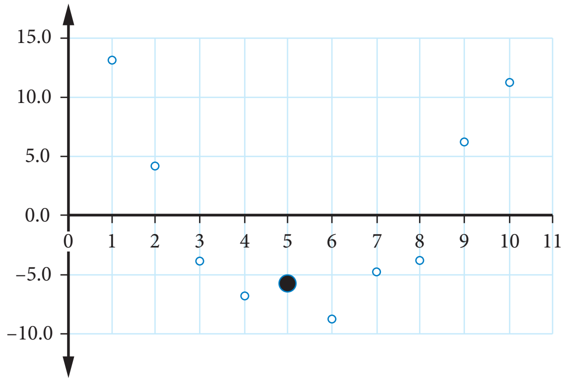
ii \$185 750

One mark for 185.75

c **i** −5.85, +5.85 and 5.9 were also accepted.

Teachers should use the value convention where residual = actual value − predicted value. This convention gives a residual plot that can be visually related to the data and line of best fit plot. For 2002, the sign was ignored in the calculation but a negative value was required in the diagram for 8 c ii.

ii

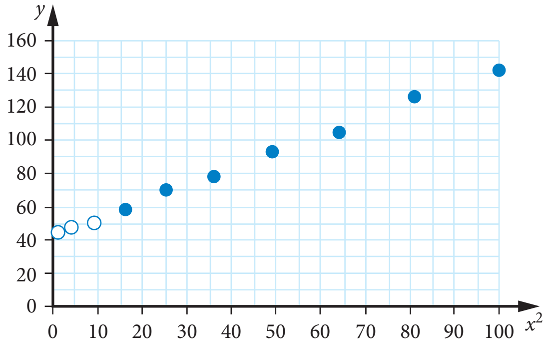


Students who correctly plotted their residual value from 8 c i in the **negative** region gained a consequential mark here.

d **i**

Year^2 (x^2)	1	4	9	16	25	36	49	64	81	100

ii



One mark for a generally correct linear pattern; 2 marks for all points correct.

The most common error was the incorrect plotting of the point for (81, 126), which was often shown around (81, 132).

iii One mark for each correct number and correct to 2 decimal places, although $42.89 + 1x^2$ was accepted. One mark total was awarded if the correct numbers were in the incorrect answer boxes.

Many students still do not make effective use of a graphics calculator to find the equation for the line of best fit.

iv There was one mark for 267.89; and consequential marks were available for correctly substituting 5 or 15 into their equation from 8 d iii but which had to involve the square of *x*.

v There is no pattern in this new residual plot, the points appear to be randomly distributed with respect to the horizontal axis.

This question was not well answered. Many students indicated that there were 'an equal number of points above and below the horizontal axis'. This was not accepted as a pattern could still exist under such a condition.

[VCAA 2002 2RCQ1]

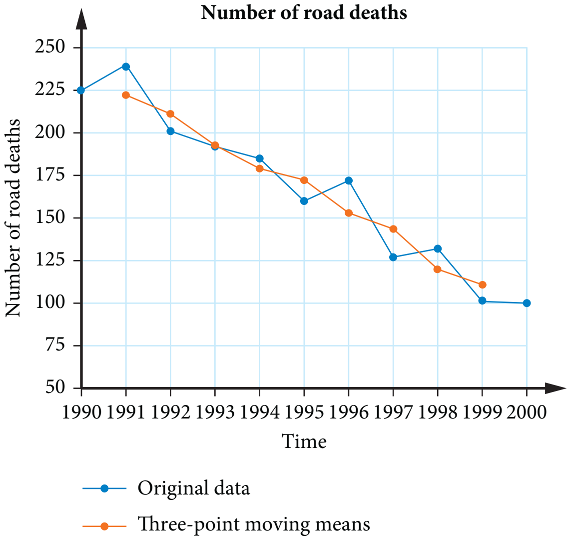
EXAM PREP 5.2 Numerical smoothing

Prep 1

a

Year	Number of road deaths	Three-point moving means
1990	225	
1991	240	$\frac{225+240+201}{3}=222$
1992	201	$\frac{240+201+192}{3}=211$
1993	192	$\frac{201+192+185}{3}=192.67$
1994	185	$\frac{192+185+160}{3}=179$
1995	160	$\frac{185+160+172}{3}=172.33$
1996	172	$\frac{160+172+127}{3}=153$
1997	127	$\frac{172+127+132}{3}=143.67$
1998	132	$\frac{127+132+101}{3}=120$
1999	101	$\frac{132+101+100}{3}=111$
2000	100	

b



c The graph of the smoothed data suggests a decreasing trend.

Prep 2

a

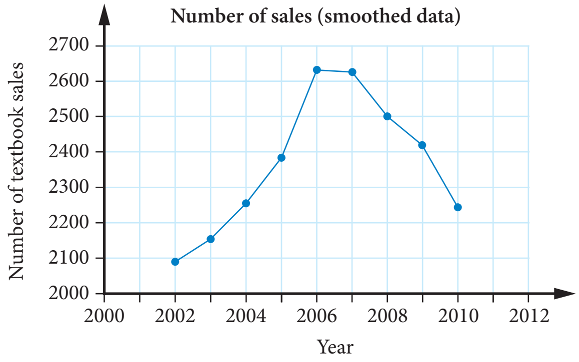
Year	Sales of textbooks	Four-point moving means	Four-point moving means with centring
2000	2250		
2001	2230		
		$\frac{2250+2230+2000+2010}{4}=2122.5$	
2002	2000		2090
		$\frac{2230+2000+2010+1980}{4}=2057.5$	
2003	2010		2153.75
		$\frac{2000+2010+1980+3000}{4}=2250$	
2004	1990		2255.625
		$\frac{2010+1980+3000+2045}{4}=2261.25$	
2005	3000		2383.625
		$\frac{1980+3000+2045+2989}{4}=2506$	
2006	2045		2632.25
		$\frac{3000+2045+2989+3000}{4}=2758.5$	
2007	2989		2627.25
		$\frac{2045+2989+3000+1950}{4}=2496$	

9780170371100

2008	3000		2505.375
		$\frac{2989 + 3000 + 1950 + 2120}{4} = 2514.75$	
2009	1950		2423
		$\frac{3000 + 1950 + 2120 + 2255}{4} = 2331.25$	
2010	2150		2243.375
		$\frac{1950 + 2120 + 2255 + 2297}{4} = 2155.5$	
2011	2255		
2012	2297		

b 2423

c



EXAM PRACTICE 5.2 Numerical smoothing

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1	2	7	80	1	9	2013
Q2	6	11	22	51	11	2013
Q3	3	7	64	13	12	2006
Q4	19	8	10	55	8	2011
Q5	16	9	16	37	22	2003

Question 2

2007	Summer	390		
	Autumn	126		
			182.75	
	Winter	85		191.5
			200.25	
	Spring	130		
2008	Summer	460		

Question 4

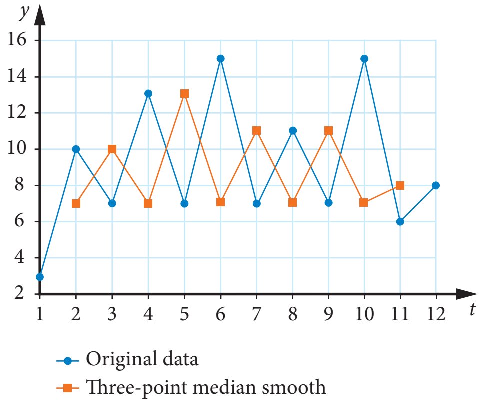
2004	2005		2006		2007	2008
1 012 000	2 016 000		3 900 000		4 830 000	5 140 000
		2 958 000		4 365 000		
			3 661 500			

Question 5

1993	13		
1994	7		
		8	
1995	3		7.625
		7.25	
1996	9		
1997	10		

EXAM PREP 5.3 Graphical smoothing

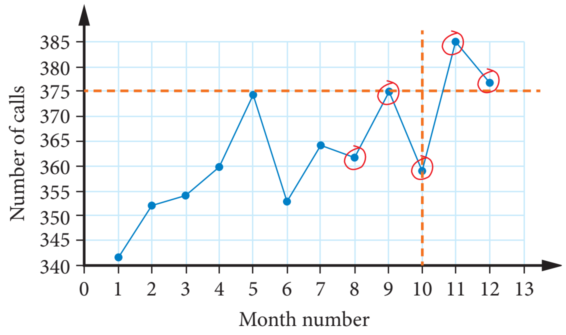
Prep 1



EXAM PRACTICE 5.3 Graphical smoothing

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1	14	12	27	43	4	2010
Q2	12	42	4	34	8	2012
Q3	44	30	8	10	8	2005
Written response						
Q4	Not reported					2009
Q5	27					2003

Question 1



Looking at the last five months, the number of calls for month 9 (i.e. 375) is the middle value (it has two values higher and two values lower). So in the smoothed plot for month 10, the number of calls is 375.

Question 2

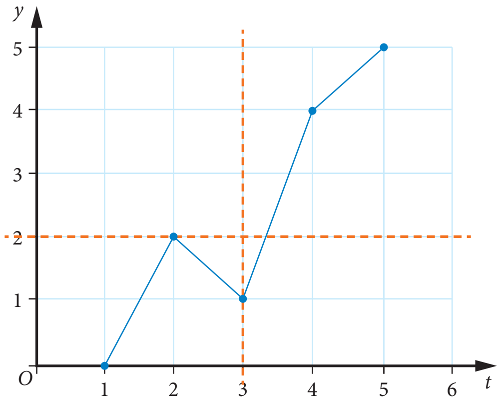
In a smoothed time series, the first two points and the last two points should be missing, so either B or D is correct. B has the correct second point.

Examination report

Students were asked to apply five-median smoothing to a time series plot. Of those students who failed to answer the question correctly, a significant proportion chose the answer that most closely matched five-mean smoothing, suggesting a lack of care when reading the question.

[VCAA 2012 1RCQ9]

Question 3



An alternative way to think of it is:

The x-coordinates in order are 1, 2, 3, 4, 5, so the median is $x = 3$.

The y-coordinates in order are 0, 1, 2, 4, 5 so the median is $y = 2$.

This means the median point is (3, 2).

Examination report

In this question, students were given a set of points on a two-dimensional grid and asked to locate the median point. This is a critical skill in median smoothing. Approximately 44 per cent of students chose option A, which indicated they had determined the median of the set of points with reference to a single number line (the t-axis). However, to obtain the correct answer students needed to then determine the median of the set of points with reference to a second number line (the y-axis) and then combine the results.

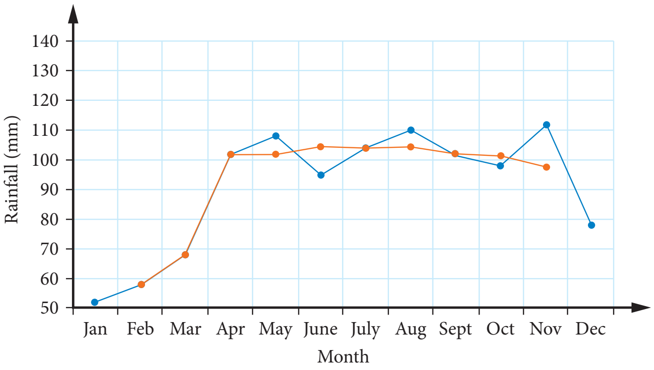
[VCAA 2005 1RCQ10]

Question 4

Examination report

a November

b



[VCAA 2009 2RCQ2]

As stated in the study design, median smoothing is to be treated as a graphical technique. That is, when students are given a time series graph to smooth using median smoothing, they are expected to locate medians directly from the graph by inspection.

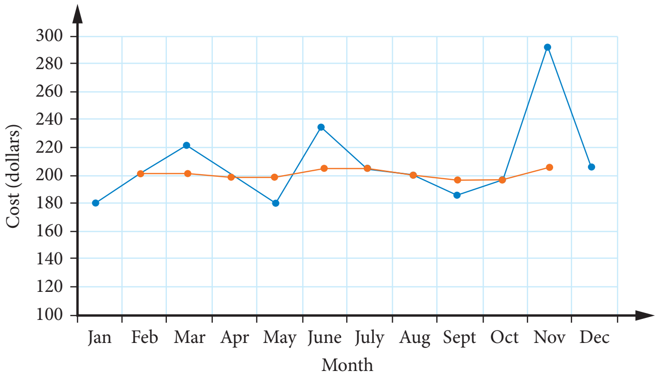
c In the smoothed time series, there are two key trends. Until April, there is an increase in monthly rainfall. It then remains relatively constant for the remainder of the year. In answering such questions, students must always refer to the given data rather than their own personal experiences. Some students referred to the seasons of the year but there was no indication in the data to suggest that these figures belonged to the southern hemisphere. Some students referred to rainfall in January and/or December but these months did not belong to the smoothed data plot.

[VCAA 2009 2RCQ2]

Question 5

Examination report

a



Many seemed to have confused three-median smoothing with three-mean smoothing. The absence of plots for January and for December was essential for full marks. At least one mark was lost if more than two points were in error.

b No. The smoothed time series is basically horizontal and shows no trend. An explanation was needed to support the negation of the supervisor's belief. There was confusion about 'trend' with numerous responses referring to 'an up and down trend' or 'a flat trend that goes up in November' or 'the trend goes up and down'. Students who incorrectly smoothed the data generally missed this mark as they usually indicated an increasing trend, which is not supported by the original data.

c An outlier exists in November. A single outlier can have an undue influence in calculating a mean since its actual *value* is important. A median calculation simply regards the *position* of an outlier as a point 'at the edge', regardless of its value. Students performed poorly on the question. Some asserted that a time series or seasonal data should always be smoothed by the three-median method.

[VCAA 2003 2RCQ2]

EXAM PREP 5.4 Seasonal adjustment

Prep 1

Year	Q1	Q2	Q3	Q4	Yearly mean
2016	11	8	1	3	$\frac{11+8+1+3}{4}=5.75$
	$\frac{11}{5.75}=1.913$	$\frac{8}{5.75}=1.391$	$\frac{1}{5.75}=0.174$	$\frac{3}{5.75}=0.522$	
2017	9	9	3	1	$\frac{9+9+3+1}{4}=5.50$
	$\frac{9}{5.50}=1.636$	$\frac{9}{5.50}=1.636$	$\frac{3}{5.50}=0.545$	$\frac{1}{5.50}=0.182$	
2018	4	9	1	7	$\frac{4+9+1+7}{4}=5.25$
	$\frac{4}{5.25}=0.762$	$\frac{9}{5.25}=1.714$	$\frac{1}{5.25}=0.190$	$\frac{7}{5.25}=1.333$	
Seasonal index	1.437	1.580	0.303	0.679	

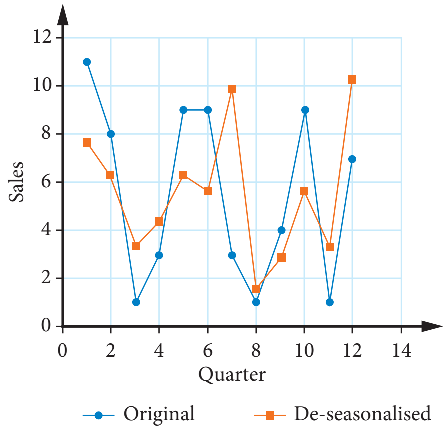
a

Year	Q1	Q2	Q3	Q4
Seasonal index	1.437	1.580	0.303	0.679

b

Year	Q1	Q2	Q3	Q4
2016	7.655	6.263	3.300	4.418
2017	6.263	5.696	9.901	1.473
2018	2.784	5.696	3.300	10.309

c



Prep 2

a 0.7 is the missing seasonal index.

b 40%

c $70 000

EXAM PRACTICE 5.4 Seasonal adjustment

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1	72	7	13	6	2	2005
Q2	5	63	10	11	10	2010
Q3	11	9	60	12	8	2012
Q4	62	7	9	6	16	2012
Q5	4	22	13	58	3	2004
Q6	10	15	11	11	53	2003
Q7	7	11	16	11	53	2007
Q8	7	6	34	42	10	2011
Written response						
Q9	Not reported					2009

Question 3

$$\text{Mean of the four seasons} = \frac{52.0 + 54.5 + 48.8 + 61.3}{4}$$
$$= 54.15$$

$$\text{So the seasonal index for spring} = \frac{61.3}{54.15} \approx 1.13$$

Question 4

$$\text{De-seasonalised value} = \frac{\text{Actual value}}{\text{Seasonal index}} = \frac{48.9}{1.01} \approx 48.4$$

Question 5

Actual value = De-seasonalised value × Seasonal index

$$\text{So Actual value} = 1.28 \times 28\,098$$
$$= 35\,965$$

Question 6

$4.00 - 1.05 - 0.84 - 0.92 = 1.19$

Question 7

The mean of the monthly revenue figures

$$= \frac{\text{Sum}}{12} = \frac{43\,872}{12} = 3656.$$

$$\text{The seasonal index for September} = \frac{4597}{3656} \approx 1.26$$

Question 8

Examination report

In this question, students were asked to determine the percentage by which an actual sales figure must change to obtain the de-seasonalised sales figure, given the seasonal index is 0.80.

A possible solution strategy is as follows.

From the formula sheet:

$$\text{seasonal index} = \frac{\text{actual sales}}{\text{de-seasonalised sales}}$$

Making the de-seasonalised sales the subject of the formula:

$$\text{de-seasonalised sales} = \frac{\text{actual sales}}{\text{seasonal index}}$$

or, for the case in this question,

$$\text{de-seasonalised sales} = \frac{\text{actual sales}}{0.8} = 1.25 \times \text{actual sales}$$

This means that, to obtain the de-seasonalised sales for summer, the actual sales figure must be increased by 25%.

[VCAA 2011 1RCQ12]

Question 9

Examination report

a 1.10

$0.78 + 1.05 + 1.07 + s = 4$

$\Rightarrow s = 1.1$

A common incorrect answer was 1.09.

b 241 mm

c The autumn rainfall is 5% above the average for the four seasons of the year. A common incorrect answer suggested that the autumn rainfall was 5% above the **monthly** average. Students are expected to be clear in their explanations. It was again evident that some students simply copied material from their bound reference notes and gave answers that referred to autumn 'sales'.

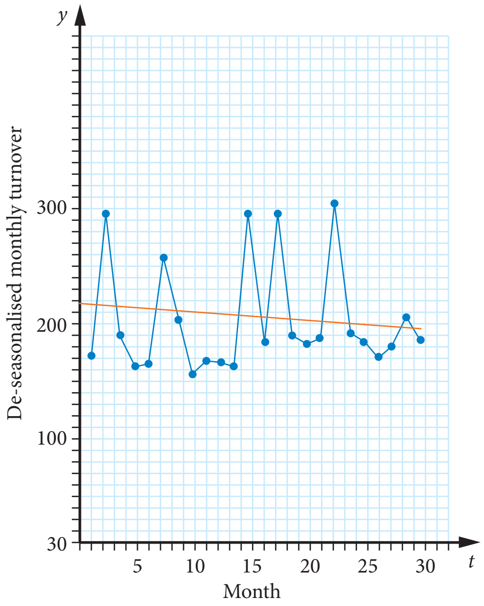
[VCAA 2009 2RCQ4]

EXAM PREP 5.5 Least square trend lines

Prep 1

a $y = 218 - 0.725t$

b



c $184.30

d $170.60

EXAM PRACTICE 5.5 Least squares trend lines

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1	2	90	4	1	3	2005
Q2	1	8	13	8	70	2005
Q3	4	6	25	61	4	2006
Q4	17	10	66	5	1	2006
Q5	29	59	5	3	4	2006
Q6	8	12	11	14	55	2002
Q7	5	2	4	50	39	2008
Q8	6	16	56	14	7	2008
Q9	23	7	5	63	2	2008
Q10	3	10	39	6	41	2012
Written response						
Q11	65					2007
Q12	57					2010

Question 5

De-seasonalised number = 373.3 − 3.38 × 6 = 353.02

Actual number = 353.02 × 0.86 = 303.6

Examination report

In this question, students were given the equation of a trend line for predicting unemployment rates from the month of the year. The trend line had been derived using **de-seasonalised** data. Monthly seasonal indices were also provided. Students were asked to use the trend line to predict the **actual** unemployment rate in June. The majority of students (59 per cent) were able to predict the de-seasonalised unemployment rate from the trend line, but then failed to seasonalise the value they obtained to determine the actual unemployment rate. Only 29 per cent of students answered this question correctly, choosing option A.

[VCAA 2006 1RCQ13]

Question 6

From the graph of the line, the y-intercept is close to 17 and the slope is negative, so the only option that matches this is rate = 17 − 1.1 × year.

Question 7

Examination report

The question required students to choose the correct qualitative description of a time series plot from five alternatives. Only 39 per cent of students chose the correct alternative, 'an increasing trend only' (option E). Fifty per cent of students were apparently distracted by the two or three relatively large random fluctuations in monthly values, and incorrectly concluded that there was evidence of seasonality in addition to the increasing trend (option D).

[VCAA 2008 1RCQ11]

Question 8

Month	No. of users	4-point moving means	4-point moving means with centring
3	354		
4	356		
		359	
5	373		360.25
6	353		
		361.5	
7	364		

So, the smoothed value is closest to 360.

Question 9

December 2009 is the 36th month. So Number of users = 346 + 2.77 × 36 = 445.72

Question 10

Examination report

In this question, students were asked to use a trend line relating de-seasonalised quarterly sales to quarter number in order to predict the actual sales in quarter 4.

A possible solution strategy is as follows.

Use the trend line to predict the de-seasonalised sales in quarter 4.

$$\textit{de-seasonalised sales}\text{ in quarter }4 = 256\,000 + 156\,000 \times 4$$
$$= 318\,400$$

To find the actual sales from the de-seasonalised sales, the seasonal index for quarter 4 is needed.

$$\text{seasonal index}_{\text{quarter }4} = 4 - (1.2 + 0.7 + 0.8) = 1.3$$

Thus

$$\text{actual sales} = \textit{de-seasonalised sales} \times \text{seasonal index}$$
$$= 318\,400 \times 1.3 = 413\,920$$

[VCAA 2012 1RCQ13]

Question 11

Examination report

a 1964

A number of students misread the question and gave the lowest temperature of 13.01°C as their answer.

b **i** 13.77 **ii** −0.22

The negative sign was required, as determined by the correct use of the residual formula on the formula sheet.

iii 0.013

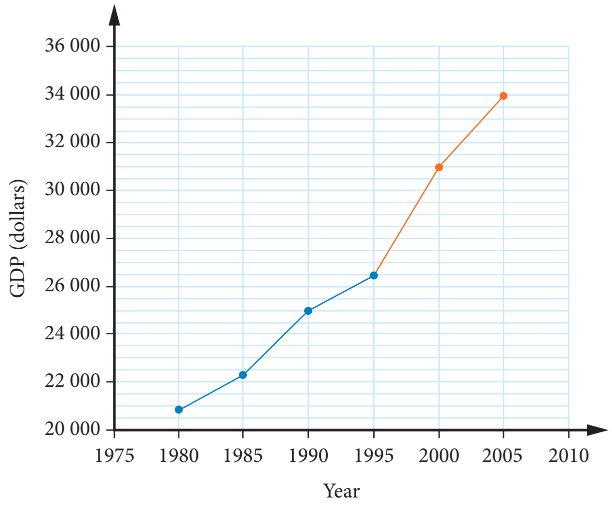
The slope form of the given equation was expected here. Sometimes an incorrect answer was calculated from the data.

[VCAA 2007 2RCQ2]

Question 12

Examination report

a



Many students had difficulty plotting the required values along the vertical scale with intermediate intervals of 500.

b an increasing trend

A common unacceptable answer described the general trend by explaining what occurred in each five-year period.

c $GDP = 20\,000 + 524 \times time$

Students were expected to use the variables given in the question rather than just x and y. While most students found the least squares line of best fit equation using technology, some merely chose two points and found the equation of the straight line that joined them.

d $752

$20\,000 + 524 \times 27 = 34\,148$

Error in prediction $= 34\,900 - 34\,148 = 752$

An answer of −752 was also accepted.

Many students calculated the predicted GDP of $34 148 but then did not continue to find the corresponding error.

A number of students inappropriately substituted GDP = $34 900 into the line of best fit equation, determined that the date could not have been 2007, and stated that 'The error is the time.'

[VCAA 2010 2RCQ3]

EXAMINATION Solutions

DATA ANALYSIS – EXAMINATION 1

1	2	3	4	5	6	7	8	9	10	11	12
B	D	C	D	B	A	C	E	B	E	E	A

Question 1

The median is the 14th entry. Median = 3.1

Question 2

25 and 37 are both 1σ from $\mu = 31$ and 68% of 2850 = 1938

Question 3

Enter data onto your CAS and use Univariate data.

Question 4

The categorical variables are sex, type of car and postcode.

Question 5

Read from table.

Question 6

This comes down to checking which values are outliers. Any waiting time greater than $Q_3 + 1.5 \times \text{IQR}$ is an outlier. $Q_3 + 1.5 \times \text{IQR} = 50 + 1.5 \times 20 = 80$ minutes, so there are two outliers.

Question 7

Both the range and IQR for the inner suburbs are clearly greater than for the outer suburbs, so the population densities are more variable in the inner suburbs than in the outer suburbs.

Question 8

Back-to-back stem plots are used for comparing the distribution of two sets of data values for the same variable.

Question 9

$$\text{Slope} = \frac{6.7 - 10.7}{1700 - 900} = -0.005$$

Questions 10

12 – (sum of January to November values) = 1.29

Questions 11

$$\frac{\$213\,956}{0.89} = \$240\,400$$

Question 12

$$\text{De-seasonalised} = \frac{\text{actual}}{1.25} = 80\% \text{ of actual. So the actual}$$

heater sales should be reduced by 20%.

DATA ANALYSIS – EXAMINATION 2

Question 1

a 19%

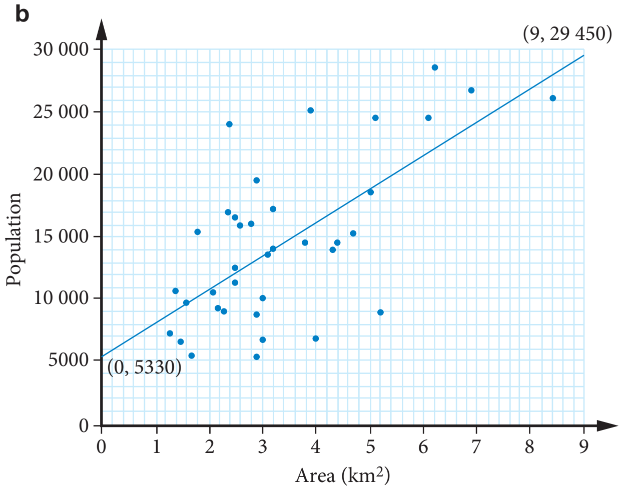
b $128\ 000\ 000 \times \left(\frac{100-77}{100}\right) = 29\ 440\ 000$

c The percentages of the people in the 15–64 age group in all three countries are approximately the same: 67% in Australia, 64% in India and 64% in Japan. This indicates there isn't any association between the percentage of people in the 15–64 age group and the country in which they live.

Question 2

a Population is the dependent (or response) variable.

b



c The slope of the line is 2680, so the rate of increase in population is approximately 2680 people for every increase of 1 km^2 in area.

d **i** Predicted population = 5330 + 2680 × 4 = 16 050

Residual = 6690 − 16 050 = −9360

ii $r^2 = 0.668^2 \approx 0.446$, which is 44.6%

Question 3

a population = $7.7 + 7.7 \times \log_{10}(\text{area})$

b population = $7.7 + 7.7 \times \log_{10} 90 \approx 23$ (thousands)

The population would be 23 000.

Question 4

a $r = -\sqrt{0.141} = -0.3755$, so the association between population density and area is weak, negative and linear.

b **i** $z = \frac{3082-4370}{1560} \approx -0.8$

ii $38 \times \frac{2.5}{100} \approx 1$

iii $\mu + 2\sigma = 3.4 + 2 \times 1.6 = 6.6$. From the graph, two inner suburbs have an area greater than 6.6 km^2.

CHAPTER 6

Depreciation and interest

EXAM PREP 6.1 First-order recurrence relations

Prep 1

1, 5, 17, 53

The first four terms of the sequence are 1, 5, 17, 53.

Prep 2

a 15, 13, 11, 9, 7, 5

b 2, 8, 32, 128, 512, 2048

c −8, −1, 6, 13, 20, 27

d 64, −32, 16, −8, 4, −2

e 50, 45, 40, 35, 30, 25

f 3, 7, 15, 31, 63, 127

g 1, −2, 13, −62, 313, −1562

h 40, 22, 13, 8.5, 6.25, 5.125

Prep 3

a $A_0 = 12,\ A_{n+1} = A_n - 5$

b $w_0 = 2,\ w_{n+1} = 2w_n - 6$

c $u_0 = 7,\ u_{n+1} = -10u_n + 13$

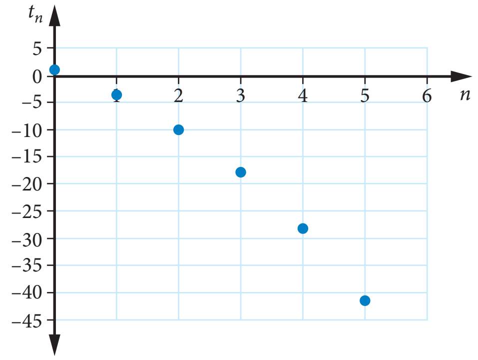
d $t_0 = 1,\ t_{n+1} = 6t_n + 5$

e $P_0 = 12,\ P_{n+1} = -3P_n - 5$

Prep 4

51

Prep 5



EXAM PRACTICE 6.1 First-order recurrence relations

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1		▨				
Q2		▨				
Q3		▨				
Q4	18	72	2	5	3	2005
Q5	13	71	7	4	5	2007
Q6	13	70	4	6	6	2006
Q7	12	19	62	4	3	2009
Q8	9	9	16	55	11	2006
Q9	12	7	19	19	43	2003

Question 8

$f_{n+1} = f_n + 5$. Since $f_1 = -1$, $f_2 = -1 + 5 = 4$.

Question 9

Substitute $t_1 = 20$ into each of the recurrence relations to see which ones give $t_2 = 10$. This means only C, D and E are possibilities. Substitute $t_2 = 10$ to see which ones give $t_3 = 20$. This leaves E as the only option.

EXAM PREP 6.2 Flat rate and unit cost depreciation

Prep 1

a Depreciation = 7% of \$10 500 = \$735

n	Future value after n years (\$)
0	10 500
1	10 500 – 735 = 9765
2	9765 – 735 = 9030
3	9030 – 735 = 8295
4	8295 – 735 = 7560
5	7560 – 735 = 6825

\$6825

b Let V_n = the future value of the industrial vacuum cleaner after n years. The recurrence relation is $V_0 = 10\ 500$, $V_{n+1} = V_n - 735$

Its graph is a decreasing straight line.

c $V_n = V_0 - n \times 735$

d \$1680

Prep 2

a **i** \$3600 **ii** \$2400 **iii** \$1500

b **i** \$10 500 **ii** \$7000 **iii** \$3500

Prep 3

a \$60 000 – \$24 000 = \$36 000 **b** 12%

Prep 4

a

n	Future value after n years (\$)
0	20 000
1	20 000 – 560 = 19 440
2	19 440 – 560 = 18 880
3	18 880 – 560 = 18 320

The future value after 3 years is \$18 320.

b Let V_n be the future value of the colour laser printer after n years. $V_0 = 20\ 000$, $V_{n+1} = V_n - 560$

Since V_n has a coefficient of 1, the graph will form a straight line (as long as the use is the same each year).

c New annual depreciation = 0.007 × 60 000 = \$420

4	18 320 – 420 = 17 900
5	17 900 – 420 = 17 480

The future value after 5 years is \$17 480.

Prep 5

a The depreciated value of the vehicle is \$33 000.

b The vehicle has travelled 200 000 km.

EXAM PRACTICE 6.2 Flat rate and unit cost depreciation

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1					▨	
Q2				▨		
Q3	▨					
Q4	7	4	4	3	82	2011
Q5	2	3	5	13	76	2009
Q6	73	5	13	5	3	2005
Q7	8	5	5	14	67	2006
Q8	19	8	64	6	3	2007
Q9	24	63	3	5	5	2003
Q10	5	6	61	19	9	2002
Q11	11	50	24	6	9	2010
Written response						
Q12	70					2013
Q13	Not reported					2007
Q14	Not reported					2008

Question 11

The start of 2007 financial year to the start of 2010 financial year is three depreciating periods.

$n = 3$

$V_0 = 40\ 000$

$V_3 = 22\ 000$

$V_n = V_0 - nd$

$V_3 = V_0 - 3d$

$22\ 000 = 40\ 000 - 3d$

$d = 6000$

$$d = \frac{\text{flat rate of depreciation}}{100} \times V_0$$

$$6000 = \frac{\text{flat rate of depreciation}}{100} \times 40\ 000$$

flat rate of depreciation = 15%

Question 12

Examination report

a $8000

b **i** 8000 – 6500 = 1500

ii $500

8000 – 5 × 1500 = 500

c 12 000

$0.25 \times k = 3000$

Many students did not attempt this question.

[VCAA 2013 2RBRMQ1]

Question 13

Examination report

i $\frac{(900-300)}{5} = \$120$

ii $900 – 0.46 × 250 × 5 = $325

[VCAA 2007 2RBRMQ3b]

Question 14

Examination report

a $900

$$\frac{17\ 000-3500}{15} = 900$$

b 5.3%

$$\frac{900}{17\ 000} \times 100 = 5.294$$

[VCAA 2008 2RBRMQ4]

EXAM PREP 6.3 Reducing balance depreciation

Prep 1

a

n	Depreciation after *n* years ($)	Future value after *n* years ($)
0	–	12 000
1	1920	10 080
2	1613	8467
3	1355	7112
4	1138	5974
5	956	5019

i $5019

ii $1138

b around $1000.

Prep 2

a Let V_n = the future value of the company car after n years. $V_0 = 80\ 000$, $V_{n+1} = 0.86V_n$

b $V_n = 0.86^n V_0$

c 2020

d $5269 rounded to the nearest dollar.

Prep 3

$136 295

Prep 4

15.7%

EXAM PRACTICE 6.3 Reducing balance depreciation

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1			▨			
Q2		▨				
Q3				▨		
Q4		▨				
Q5	2	8	8	6	75	2007
Q6	12	8	8	63	8	2012
Q7	3	31	57	5	3	2010
Q8	6	14	16	54	9	2009
Q9	49	23	8	16	3	2004
Q10	25	47	13	5	9	2008
Q11	4	29	12	35	19	2013
Written response						
Q12	56					2006
Q13	53					2012
Q14	37					2003
Q15	37					2004
Q16	33					2005
Q17	Not reported					2007
Q18	Not reported					2008
Q19	Not reported					2009
Q20	Not reported					2013
Q21	Not reported					2002

Question 7

Use the recursive computation function on a CAS/calculator.

FV = $15 360

Question 8

The recurrence relation is $V_0 = 5000$, $V_{n+1} = 0.8V_n$. Use the recursive computation function on a CAS/calculator to generate the sequence to see when the value is below $1500 for the first time, using V_0 = 1 January 2009, V_1 = 1 January 2010 etc.

Question 9

The recurrence relation for reducing balance depreciation will always follow the curve:

Question 10

Use a CAS/calculator finance solver.

N = 4

I% = ?

PV = −6000

Pmt or PMT = 0

FV = 2000

PpY or P/Y = 1

CpY or C/Y = 1

PmtAt = END

I% = 24.02%

Question 11

Use the recurrence relation $V_0 = 25\,000$, $V_{n+1} = 0.8V_n$ and a CAS/calculator's recursive computation to generate the sequence. Or alternatively use the general rule for reducing balance depreciation. The depreciation in the fourth year is $V_3 - V_4$ = \$12 800 – \$10 240 = \$2560.

Note this process can be seen in the table below (although there is no need to create the table to answer the question):

n	Depreciation after *n* years ($)	Future value after *n* years ($)	Term
0	–	25 000	V_0
1	$\frac{20}{100} \times 25\,000 = 5000$	25 000 – 5000 = 20 000	V_1
2	$\frac{20}{100} \times 20\,000 = 4000$	20 000 – 4000 = 16 000	V_2
3	$\frac{20}{100} \times 16\,000 = 3200$	16 000 – 3200 = 12 800	V_3
4	$\frac{20}{100} \times 12\,800 = 2560$	12 800 – 2560 = 10 240	V_4

Question 12

Examination report

a **i** $6000 **ii** $42 000 **iii** 8

b **i** 15%

An incorrect answer of 85% was common.

ii $36 847.50

Many students gave this answer despite having part i incorrect.

iii 10th year

c 7th year

[VCAA 2006 2RBRMQ1]

Question 13

Examination report

a $5852

$8360 - 0.22 \times 3800 \times 3 = 5852$

A common error was to give the total depreciation over three years rather than the depreciated value after three years.

b $8360 \times 0.1 = 0.22 \times 3800 = \$836$

This 'show that' question required two calculations, but many students did not follow the instructions and showed only one calculation. Others determined the depreciated *value* for a specific year instead of the annual depreciation *amount* in any one year as required.

c 10 years

$0.22 \times 3800 \times n = 8360$

$\therefore n = 10$

d $1850

$8360 \times 0.86^{10} = 1850.08\ldots$

[VCAA 2012 2RBRMQ2]

Question 14

Examination report

a $1904

A number of students incorrectly treated this as simple interest and reduced $3100 by five lots of (15% of $3100). One mark was available for some demonstration of a single reduction by 15%.

b $1750

The main source of error was converting 3 cents to $0.03 or ignoring the three-year period of depreciation. One mark was available for correct calculation of one year's depreciation.

c 2.7 cents, accept $0.027

This required solution of $1904 = 3100 - (15\,000 \times 3 \times x)$ was generally poorly done. One mark was available for setting up an appropriate equation. Many tried substitution to get a figure close to the result from part a. One mark was available for correctly using incorrect answer from part a.

[VCAA 2003 2RBRMQ2]

9780170371100

Question 15

Examination report

d Flat rate gives a greater depreciation by \$28.41.

Most students calculated the *future value* with the reducing balance depreciation of $650 \times (0.85)^5 = \$288.41$. Many found the flat depreciation amount of $12\% \times 650 \times 5 = \$390$, but then misinterpreted this as being equivalent to the future value.

Several answers ignored the percentages altogether and suggested that 'flat rate will give the greater depreciation as it reduces by the same value while the reducing balance depreciation gets smaller each year'.

[VCAA 2004 2RBRMQ1d]

Question 16

a Depreciation = 4000 −1000 = \$3000

$n = 5$

$V_0 = 4000$

$V_5 = 1000$

$V_n = V_0 - nd$

$V_5 = V_0 - 5d$

$1000 = 4000 - 5d$

$d = 600$ (using CAS/calculator solve function if necessary)

$$d = \frac{\text{flat rate of depreciation}}{100} \times V_0$$

$$600 = \frac{\text{flat rate of depreciation}}{100} \times 4000$$

Annual flat rate of depreciation = 15%

b Use a CAS/calculator finance solver.

Annual reducing balance rate of depreciation = 24.2%

Question 17

Examination report

a $10\,000 \times 0.885 = 5277.319$

$= \$5277$

b 16.7%

N = 5

I = 16.744679

PV = 10 000

PMT = 0

FV = −4000

P/Y = 1

Students were also able to solve $10\,000 \times \left(1 - \frac{y}{100}\right)^5 = 4000$

[VCAA 2007 2RBRMQ4]

Question 18

Examination report

a \$1377

$T_2 - T_3 = 17\,000 \times 0.9^2 - 17\,000 \times 0.9^3 = 1377$

Many students misread this question and found only depreciated value $T_3 = 17\,000 \times 0.9^3 = \$12\,393$.

b 9 years (8.4 years was also accepted)

$17\,000 \times 0.9^n = 7000$

$n = 8.4$

An answer of 8 years was not accepted as the car's value is still not less than \$7000.

[VCAA 2008 2RBRMQ3]

Question 19

Examination report

a \$11 440

Amount of depreciation = $22\,000 \times 0.12 \times 4 = 10\,560$

Depreciated value = $22\,000 - 10\,560 = 11\,440$

Many students did not fully answer this question and found only the amount of depreciation.

b \$10 953.17

$22\,000 \times 0.84^4 = 10\,953.169\,92\ldots$

Many students incorrectly rounded the result to the nearest five cents and gave \$10 953.15.

c Reducing balance method will give the greater depreciation by \$11 046.83.

[VCAA 2009 2RBRMQ4]

Question 20

Examination report

\$3328

Depreciation rate = $\frac{7500 - 6375}{7500} = 15\%$

Value after 5 years = $7500 \times 0.85^5 = \$3327.789\ldots$

[VCAA 2013 2RBRMQ3d]

Question 21

Examination report

a [r = **14.1%**]

One mark for finding depreciation is \$3320 per year. The poor setting out of student work for this question sometimes made method marks difficult to allocate.

b [$r = 21.6\%$]

One mark for $7000 = 23\ 600 \times R^5$ or equivalent.

This question caused difficulties with setting up the relevant equation.

Several used the straight-line method here. Many could not solve $R^5 = \dfrac{7000}{23\ 600}$ which can be done using the equation solver on a graphics calculator.

[VCAA 2002 2RBRMQ3]

EXAM PREP 6.4 Simple interest

Prep 1

a 14% of \$8000 = \$1120

Let V_n = the amount in the account at the end of nth year.

$V_0 = 8000,\ V_{n+1} = V_n + 1120$

b $V_0 = 6000,\ V_{n+1} = V_n + 720$

c $V_0 = 7500,\ V_{n+1} = V_n + 375$

Prep 2

a \$3080 **b** \$3440 **c** \$3800

Prep 3

a $V_0 = 10\ 000,\ V_{n+1} = V_n + 400$

b $V_n = 10\ 000 + 400n$

c Stephen will earn $7 \times 400 = \$2800$ interest in 7 years.

d Stephen's investment will grow to \$13 200 in 8 years.

EXAM PRACTICE 6.4 Simple interest

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1			▨			
Q2				▨		
Q3	6	91	1	1	1	2006
Q4	5	91	1	1	3	2009
Q5	4	90	1	1	3	2004
Q6	3	2	90	2	2	2012
Q7	0	3	4	89	3	2010

EXAM PREP 6.5 Compound interest

Prep 1

a $r = 0.75$

b $V_0 = 55\ 000,\ V_{n+1} = 1.0075V_n$

c $V_n = 1.0075^n \times 55\ 000$

d \$86 112.46

Prep 2

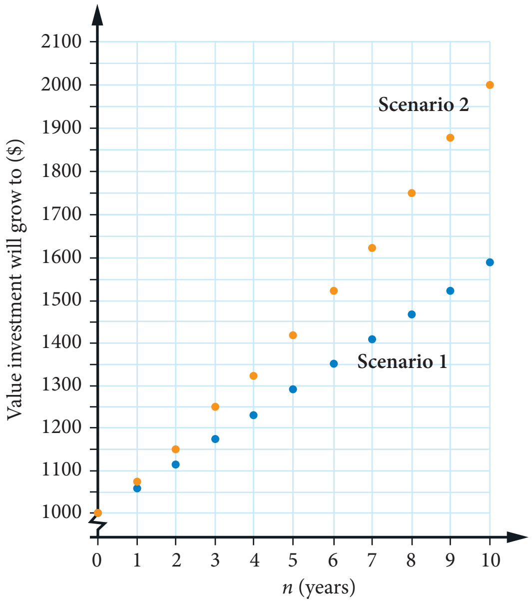
a \$102 917.31 **b** \$22 917.31

Prep 3

104 months = 8 years 8 months

Prep 4

a



b The better option is scenario 2.

Prep 5

a 9.38% **b** 11.61% **c** 12.36% **d** 6.18%

Prep 6

7.82%, 7.76%, 7.68%; so 7.6% compounded quarterly is best.

EXAM PRACTICE 6.5 Compound interest

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1				▨		
Q2					▨	
Q3				▨		
Q4					▨	
Q5					▨	
Q6					▨	
Q7	11	2	4	81	1	2011
Q8	10	79	7	3	1	2002
Q9	10	5	4	9	72	2004
Q10	55	11	12	7	15	2002
Q11	3	4	24	51	18	2003
Q12	51	10	18	8	11	2008
Q13	16	8	47	8	20	2005
Q14	17	10	9	42	21	2007
Q15	16	16	16	42	8	2012

Q16	5	40	16	28	10	2013
Q17	37	20	14	18	10	2006
Q18	7	33	16	40	4	2011
Q19	20	9	33	17	21	2008
Q20	17	39	10	30	4	2003
Q21	29	7	13	46	4	2004
Written response						
Q22	Not reported					2008
Q23	Not reported					2009
Q24ab	35					2010
Q24c	Not reported					2010

Question 10

The graph is showing an investment with compound interest because the amount of interest added each year is increasing. (The increase in the first year is just over \$100, while the increase in the last year was over \$200.) The rule for compound interest growth is in the form

$V_n = \left(1+\frac{r}{100}\right)^n \times V_0$. The only option in this form is A.

Question 11

Find V_5 and subtract it from V_0.

Question 12

Find V_3 for both and compare.

Sam: $V_3 = \$6125$

Charlie: $V_3 = \$6211.48$

Sam has \$6211.48 – \$6125 = \$86.48 less than Charlie.

Question 13

Use $V_n = \left(1+\frac{r}{100}\right)^n \times V_0$.

$$\begin{aligned}\text{The interest paid in the 4th year} &= V_4 - V_3\\ &= \$3859.40 - \$3623.85\\ &= \$235.55\end{aligned}$$

Question 14

Use the rule $V_n = \left(1+\frac{r}{100}\right)^n \times V_0$.

$V_0 = 10\,000$, $r = 5\%$ per compounding period (i.e. half yearly), $n = 10$

So $V_{10} = 10\,000\left(1+\frac{5}{100}\right)^{10} = 10\,000 \times (1.05)^{10}$

Question 15

Use the rule $V_n = \left(1+\frac{r}{100}\right)^n \times V_0$.

Total value

$$= 15\,000 \times \left(1+\frac{1}{12}\times\frac{6.1}{100}\right)^6 \times \left(1+\frac{1}{12}\times\frac{6.25}{100}\right)^6$$

$\approx \$15\,952.92$

Total interest = \$15 952.92 – \$15 000 = \$952.92 ≈ \$953

Question 16

$$V_n = \left(1+\frac{r}{100}\right)^n \times V_0$$

$V_0 = 10\,000$, $r = 2\%$ per compounding period (i.e. quarterly), $n = 20$ (quarterly for 5 years)

$$V_{20} = 10\,000\left(1+\frac{2}{100}\right)^{20} = 10\,000 \times (1.02)^{20}$$

$$\text{Interest} = V_{20} - V_0 = 10\,000 \times (1.02)^{20} - 10\,000$$

Question 17

The balance increases only at the start of each year, not at each quarter, so the interest is credited annually. The interest amount (i.e. the difference between consecutive balances) also increases each year, which means it is compounded annually. (Simple interest would mean the interest amount is the same each year.)

Examination report

This question concerned the growth of an investment, but was formulated graphically rather than numerically. The aim was to test conceptual understanding of the different forms of investment growth. Only 37 per cent of students were successful in answering this question, choosing option A. Teachers and students are reminded that graphical analysis of business and financial situations is a requirement of the study design.

[VCAA 2006 1RBRMQ8]

Question 18

Interest earned in the third year

= value in the third year – value in the second year

$$V_n = \left(1+\frac{r}{100}\right)^n \times V_0$$

$V_0 = 15\,000$, r becomes $\frac{r}{12}$ per compounding period (i.e. monthly), $n = 36$ (monthly for 3 years),

$n = 24$ (monthly for 2 years)

Interest earned in third year

$$= 15\,000\left(1+\frac{r}{100}\right)^{36} - 15\,000\left(1+\frac{r}{100}\right)^{24}$$

Examination report

In this question, many students chose option D:

$15\,000\left(1+\frac{r}{100}\right)^{36} - 15\,000\left(1+\frac{r}{100}\right)^{24}$ and this suggested that these students had used an appropriate method of solution but had used the yearly interest rate r rather than the monthly interest rate $\frac{r}{12}$.

[VCAA 2011 1RBRMQ7]

Question 19

Interest earned in the fourth year

= value in the fourth year – value in the third year.

$$V_n = \left(1+\frac{r}{100}\right)^n \times V_0$$

$V_0 = 8000$, $r = \frac{7.2}{4}$ per compounding period (i.e. quarterly), $n = 16$ (quarterly for 4 years), $n = 12$ (quarterly for 3 years)

Interest earned in fourth year

$$= 8000\left(1+\frac{7.2}{100\times 4}\right)^{16} - 8000\left(1+\frac{7.2}{100\times 4}\right)^{12}$$

$$= 8000(1.018)^{16} - 8000(1.108)^{12}$$

Examination report

In this question, students were asked to determine which of five given expressions, based on the compound interest formula, could be used to determine the interest earned in the fourth year by an $8000 investment. Only 33 per cent of students chose the correct option (C). Common errors included not recognising that interest was calculated quarterly (option B) or not converting the annual interest rate to a quarterly interest rate (option E).

[VCAA 2008 1RBRMQ9]

Question 20

$$V_n = \left(1+\frac{r}{100}\right)^n \times V_0$$

$V_0 = 8000$, r becomes $\frac{r}{4}$ per compounding period (i.e. quarterly), $n = 20$ (quarterly for 5 years)

Balance after 5 years $= V_{20} = 8000\left(1+\frac{r}{400}\right)^{20}$

Examination report

In the question the great number that chose option B indicated that they had failed to recognise that interest was compounding quarterly.

[VCAA 2003 1RBRMQ7]

Question 21

Interest earned *during* the third year

= value in the third year – value in the second year.

$$V_n = \left(1+\frac{r}{100}\right)^n \times V_0$$

$V_0 = 12\,000$, $r = 3.6$, $n = 3$, $n = 2$

Interest during the third year

$$= 12\,000\left(1+\frac{3.6}{100}\right)^3 - 12\,000\left(1+\frac{3.6}{100}\right)^2$$

$$= 12\,000 \times 1.036^3 - 12\,000 \times 1.036^2$$

$$= 463.66$$

Examination report

This question asked students to determine the amount of interest earned **during** the third year of an investment. Only 29% of students gave the correct response (option A). The key word in this question was 'during', which was clearly not recognised by the 46% of students who determined the total amount of interest earned **after** three years.

[VCAA 2004 1RBRMQ7]

Question 22

Examination report

a $A = 3000 \times 1.041^4$

$A = \$3523.09$

Some students did not change the annual rate to a half-yearly rate. Others did not change the number of payments to four. Rounding off the answer to $3523.10 was not accepted.

b $1137.40

$3000 \times 1.041^8 - 3000 = 1137.40$

The finance solver facility of a calculator could readily be used in this question. Often, the principal amount was not subtracted to determine the interest amount.

[VCAA 2008 2RBRMQ2]

Question 23

Examination report

a $\frac{4.4}{4} = 1.1$

b $3876.97

$3400 \times 1.011^{12} = 3876.973\,068\ldots$

Many students incorrectly rounded this result to the nearest five cents and gave $3876.95.

c $1020.86

FV after 24 periods = $4420.858 874…

Interest = 4420.86 – 3400 = 1020.86

Many students incorrectly rounded this result to the nearest five cents and gave $1020.85.

[VCAA 2009 2RBRMQ3]

Question 24

Examination report

a Simple Saver annual interest rate is the highest as the first-year increase is larger.

This question was very poorly answered. The key to understanding the graph was to consider what happens after the first interest payment. Most students seemed to ignore the word 'rate', or they may have confused 'rate' with 'return'. Common incorrect responses were that 'the Growth Plus interest is eventually higher' and 'Growth Plus is compounding and so will eventually be higher'. Others suggested that 'Simple Saver is best for investments under ten years'.

b $920

$$\frac{21\,800-8000}{15}=920$$

c **i** $24\,000=8000\times\left(1+\frac{r}{100}\right)^{15}$

Some students gave their answer as $24\,000=8000\times r^{15}$; however, this was incorrect.

ii 7.6%

[VCAA 2010 2RBRMQ3]

CHAPTER 7

Reducing balance loans and annuities

EXAM PREP 7.1 Reducing balance loans

Prep 1

a $k=\left(1+\frac{r}{100}\right)=\left(1+\frac{5}{12}\times\frac{1}{100}\right)=\left(1+\frac{5}{1200}\right)=1.004$

b $k=\left(1+\frac{r}{100}\right)=\left(1+\frac{5}{100}\right)=1.05$

c $k=\left(1+\frac{r}{100}\right)=\left(1+\frac{5}{4}\times\frac{1}{100}\right)=\left(1+\frac{5}{400}\right)=1.013$

d $k=\left(1+\frac{r}{100}\right)=\left(1+\frac{7.5}{52}\times\frac{1}{100}\right)=\left(1+\frac{7.5}{5200}\right)=1.001$

e $k=\left(1+\frac{r}{100}\right)=\left(1+\frac{9.5}{26}\times\frac{1}{100}\right)=\left(1+\frac{9.5}{2600}\right)=1.004$

Prep 2

a

Payment number *n*	Payment made ($)	Interest paid ($)	Principal reduction ($)	Balance remaining ($)
0	0.00	0.00	0.00	380 000.00
1	3854.00	2850.00	1004.00	378 996.00
2	3854.00	2842.47	1011.53	377 984.47
3	3854.00	2834.88	1019.11	376 965.35

b **i** $15\times12=180$ monthly payments

ii total payments = $180\times\$3854=\$693\,720$

c $B_0=380\,000,\ B_{n+1}=1.0075B_n-3854$

d $374 904.13

e The graph of the balance is number 3.

f Graph 2, graph 1.

Prep 3

a $A_0=45\,000,\ A_{n+1}=\left(1+\frac{11}{1200}\right)A_n-850$

b $18 769.18

Prep 4

a $68 454.10

b $1851.30

Prep 5

a **i** No, the balance remaining after the last payment is $20.81.

ii 2750.00 + 20.81 = $2770.81

b **i** $1000

ii The difference between the two figures is the extra amount of $20.81 paid with the last payment. By being paid with the last payment, it is taken out of the interest calculations.

c $B_0=10\,000,\ B_{n+1}=1.04B_n-2750$

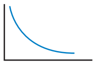
EXAM PRACTICE 7.1 Reducing balance loans

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1				▨		
Q2			▨			
Q3		▨				
Q4	▨					
Q5			▨			
Q6		▨				
Q7				▨		
Q8	9	20	17	8	46	2005
Q9	11	25	13	24	27	2008
	Written response					
Q10	▨					

Question 6

The $24.97 balance left over after the fourth payment needs to be added onto the $3300 to give a final payment of $3324.97.

Question 7

For A and B, the graph would be increasing, not decreasing. Flat rate depreciation has a straight-line graph. Reducing balance depreciation has a graph with the shape:  so E is incorrect. The graph has the shape of a reducing balance loan , so D is the correct option.

Question 8

Because of Sally's missed payment, her balance didn't reduce to the expected level, so she has to pay interest on the missed third payment.

Question 9

Examination report

This question was poorly answered, with only 25 per cent of students choosing the correct option (graph B). The key to answering this question was to realise that, for a reducing balance loan, the amount paid off the principal each year increases with time. It would appear that, by choosing either incorrect option D (24 per cent of students) or E (27 per cent of students), the majority of students confused the amount paid off the principal with the amount still owing on the loan.

[VCAA 2008 1RBQ7]

Question 10

a Option 1: $34 764.35; Option 2: $34 322.39

b Option 2

EXAM PREP 7.2 Reducing balance loans and finance solvers

Prep 1

a $92.63 **b** $442.63

c $2097.64 **d** $1514.18

Prep 2

a $1971.13 **b** $591 339 **c** $341 339

Prep 3

a 96 **b** 16 **c** 300 **d** 48

Prep 4

a $38 379.81 **b** 116

Prep 5

6.6%

EXAM PRACTICE 7.2 Reducing balance loans and finance solvers

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1	11	67	8	7	6	2009
Q2	8	14	10	64	3	2010
Q3	8	53	17	12	9	2008
Q4	18	15	50	11	6	2002
Q5	15	49	15	12	9	2004
Q6	6	28	45	12	9	2003
Q7	45	13	25	11	5	2009
Q8	15	8	8	43	25	2004
Q9	37	11	16	27	8	2007
Q10	5	35	20	23	16	2005

Written response		
Q11	Not reported	2002
Q12	Not reported	2003
Q13	Not reported	2006
Q14	Not reported	2012

Question 3

N = 20 × 4 = 80, I% = ?, PV = 300 000, PMT or Pmt = −6727.80, FV = 0, PpY or P/Y = 4, CpY or C/Y = 4

Solving for I% gives 6.5%.

Question 4

N = 20 × 12 = 240, I% = 6, PV = 172 000, PMT or Pmt = ?, FV = 0, PpY or P/Y = 12, CpY or C/Y = 12

Solving for PMT gives $1232.26.

Question 5

N = 20 × 12 = 240, I% = 7.4, PV = 250 000, PMT or Pmt = ?, FV = 0, PpY or P/Y = 12, CpY or C/Y = 12

Solving for PMT gives $1998.72, which is closest to $1999.

Question 6

N = 5 × 12 = 60, I% = 6.2, PV = 150 000, PMT or Pmt = −1100, FV = ?, PpY or P/Y = 12, CpY or C/Y = 12

Solving for FV gives $127 207.81, which is closest to $127 000.

Question 7

N = 7 × 4 = 28, I% = ?, PV = 100 000, PMT or Pmt = −2150, FV = −80 000, PpY or P/Y = 4, CpY or C/Y = 4

Solving for I% gives 6.3%.

Question 8

Check each option to find which statement is not true:

A: The monthly interest rate is $\frac{7.5}{12} = 0.625\%$. True.

B: The loan is to be fully repaid over 10 years so there will be no money owed after 10 years. True.

C: N = 10 × 12 = 120. True

D: N = 10 × 12 = 120

I% = 7.5, PV = 130 000, PMT or Pmt = ?, FV = 0, PpY or P/Y = 12, CpY or C/Y = 12

Solving for PMT gives $1543.12.

$1500 is less than the monthly repayment required to pay the loan in 10 years, so this repayment would not reduce the life of the loan.

E: N = 5 × 12 = 60, I% = 7.5, PV = 130 000, PMT or Pmt = −1543.12, FV = ?, PpY or P/Y = 12, CpY or C/Y = 12

Solving for FV gives −$77 010.03, so after 5 years the amount of the principal still owing is more than $65 000.

Question 9

N = 20 × 12 = 240, I% = 7, PV = 250 000, PMT or Pmt = ?, FV = 0, PpY or P/Y = 12, CpY or C/Y = 12

Solving for PMT gives $1938.25.

The total amount repaid = 240 × $1938.25 = $465 180.

The amount of interest paid = $465 180 – $250 000 = $215 180.

Question 10

The investor pays the interest only for the first five years since the amount owing at the end of five years is the same as the amount borrowed.

The monthly payment is $1416.67.

Question 11

Examination report

Yes, the loan will be repaid within four years as a minimum required payment of $1473 is less than the proposed $1500.

One mark for correctly finding a balance or for the minimum repayment for a loan over four years. An appropriate justification for their conclusion was required for any mark allocation.

Many students had difficulties with interpreting their calculations. Many achieved a figure of –$503.21 for the Future Value with a finance solver. However, the negative property of this result was often not understood.

A number of students multiplied $1500 by 16 to get a total of $24 000 paid over four years. They then ignored any interest calculation and simply stated that, as this exceeded the $20 000 borrowed, 'the loan must be repaid within the four years'. This argument was not accepted.

To find the quarterly repayment using a finance solver:

N = 4 × 4 = 16, I% = 8, PV = 20 000, PMT or Pmt = ?, FV = 0, PpY or P/Y = 4, CpY or C/Y = 4

Solving for PMT or Pmt gives $1475.14, which is less than her quarterly repayment of $1500 so she will pay the loan in under four years.

[VCAA 2002 2RBQ4b]

Question 12

Examination report

a $141.63

b $ 3399

Most gained the mark in this question. One mark was available for correct use of a wrong answer from part a.

[VCAA 2003 1RBQ1 c ii and 1 c iii]

Question 13

Examination report

$2150

This question was often poorly done, if attempted at all. Students first had to find the total interest paid over 24 months ($922.90 per month) using a finance solver.

Total paid = 24 × 922.90 = $22 149.60

∴ Interest = 22 150 – 20 000 = $2150

Many students rounded off the $922.90 to $923 before progressing with their calculation and therefore arrived at an inaccurate answer. Rounding off should only be done at the last step in a calculation. Others misread the question and simply found the future value (FV) without allowing for any instalments.

To find the monthly repayment using a finance solver,

N = 24, I% = 10, PV = 20 000, PMT or Pmt = ?, FV = 0, PpY or P/Y = 12, CpY or C/Y = 12

Solving for PMT or Pmt gives $ 922.90

[VCAA 2006 1RBQ4]

Question 14

Examination report

a $807.23

N = 60, I% = 7.8, PV = 40 000, **PMT = –807.232 51…**, FV = 0, P/Y = 12, C/Y = 12

b 47 months

N = 46.474 250…, I% = 7.8, PV = 40 000, PMT = –1000, FV = 0, P/Y = 12, C/Y = 12

c **i** $A = 40\ 000 \times \left(1 + \frac{7.8}{1200}\right)^{12}$

c **ii** $281.02

New balance = $43 234

$I = 43\ 234 \times \frac{7.8}{1200} = 281.021\ldots$

Many students did not answer this question.

[VCAA 2012 1RBQ3]

EXAM PREP 7.3 Changing the terms of a loan

Prep 1

a $1829.66 **b** $2923.20

Prep 2

a 12 years (144 months) **b** 14 months

c $3591.64

EXAM PRACTICE 7.3 Changing the terms of a loan

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1	22	11	45	12	9	2009
Q2	9	21	45	14	11	2002
Q3	15	44	20	12	8	2011
Q4	15	12	15	13	43	2011
Q5	15	9	22	32	22	2003
Q6	29	15	21	9	24	2013
Q7						
Q8	20	4	42	6	27	2012
Written response						
Q9	Not reported					2004
Q10	37					2007
Q11	Not reported					2010
Q12	Not reported					2011

Question 1

Find the amount owing at the end of the first 10 years.

N = 120, I% = 4.45, PV = 250 000, PMT or Pmt = −1382.50, FV = ?, PpY or P/Y = 12, CpY or C/Y = 12

Solving for FV gives −181 324.4359

This is now the amount owing.

N = ?, I% = 4.25, PV = 181 324.4359, PMT or Pmt = −1750, FV = 0, PpY or P/Y = 12, CpY or C/Y = 12

Solving for N gives 129.32 ≈129

In total it takes 120 + 129 = 249 months or approximately 21 years.

Question 2

Find the number of repayments using a finance solver:

N = ?, I% = 7, PV = 135 000, PMT or Pmt = −1200, FV = 0, PpY or P/Y = 12, CpY or C/Y = 12

Solving for N gives 183.59 ≈184 which is the number of months required to repay the loan.

Originally it was going to take 20 × 12 = 240 months.

Time saved = 240 – 184 = 56 months or 4 years 8 months, which is closest to 5 years.

Question 3

With 36 equal monthly repayments:

N = 36, I% = 9.75, PV = 45 000, PMT or Pmt = ?, FV = 0, PpY or P/Y = 12, CpY or C/Y = 12

Solving for PMT or Pmt gives $1446.75.

Xavier would repay 36 × $1446.75 = $52 083.

With 18 equal monthly repayments:

N = 18, I% = 9.75, PV = 45 000, PMT or Pmt = ?, FV = 0, PpY or P/Y = 12, CpY or C/Y = 12

Solving for PMT or Pmt gives $2697.39.

Xavier would repay 18 × $2697.39 = $48 553

Amount saved = $52 083 – $48 553 = $3530.

Question 4

N = ?, I% = 7.67, PV = 120 000, PMT or Pmt = –1560, FV = 0, PpY or P/Y = 12, CpY or C/Y = 12

Solving for N gives N = 106.2 so it takes 107 months, or 8 years 11 months, to repay the loan.

Question 5

The equivalent weekly payment to $555 monthly is $\frac{555\times 12}{52} = \$128.08$. Since a weekly payment of $132 exceeds this it will reduce the period of the loan.

Question 6

First determine the annual interest rate.

N = 1, I% = ?, PV = 300 000, PMT or Pmt = –2500, FV = –299 000, PpY or P/Y = 12, CpY or C/Y = 12

Solving for I% gives 6.

Now find the amount owing after 2 months.

N = 2, I% = 6, PV = 300 000, PMT or Pmt = –2500, FV = ?, PpY or P/Y = 12, CpY or C/Y = 12

Solving for FV gives –297 995.

Question 7

The graph shows the revised loan will be fully repaid around 6–7 years earlier than the original loan.

Question 8

Examination report

In this question students were asked to identify the graph that correctly represented the decreasing monthly balance of a loan when the fourth monthly payment is missed but made up with a double payment the following month. The key to answering this question was to recognise that the balance of the loan increases in the fourth month because, while no payment is made, interest is still charged and added to the amount owed.

(Note that in C the balance of the loan in month 4 is the same as month 3, and in D the balance of the loan in month 4 is less than month 3.)

[VCAA 2012 1RBQ9]

Question 9

Examination report

a $290.15

Several examples of $\frac{12\ 000}{48} = \$250$ were found.

b $1927

A consequential mark was available for a correct calculation of
(48 × their reasonable answer part a) − $12 000

c $1311

Very poorly answered. A common incorrect answer was $1740. The required calculation used $n = 6$, $I = 7.5$, $PV = 12\ 000$, $PMT = 290.15$ to find $FV = 10\ 688.78$ as the amount owing after six monthly payments. This was then subtracted off the initial loan of $12 000.

Many students reduced the loan by the value of six repayments without consideration of the interest component in each of those repayments.

d $293

Many students did not use the new interest rate of 8%. A mark was available here if their answer matched $I = 8$ and their numbers from the previous part.

[VCAA 2004 2RB Q2b c d eii]

Question 10

Examination report

a Monthly interest $= \frac{0.09}{12} \times 30\ 000$
$= \$225$

Many students were puzzled by not being given a period of time for this loan. A loan where the interest only is paid each month is not a reducing balance question as the balance does not decrease or increase. It becomes a simple interest calculation for one period (one month in this case). This situation is like a **perpetuity** in reverse, where Khan is paying monthly instalments **in perpetuity**.

b $16 801

N = 60, I = 9, PV = 30 000, PMT = –400,
FV = 16 800.776 04, P/Y = 12

A number of students used N = 5.

c $622.75

N = 60, I = 9, PV = 30 000, **PMT = –622.750 65**, FV = 0,
P/Y = 12

This was a straightforward finance solver question.

[VCAA 2007 2RB Q2]

Question 11

Examination report

a **i** $1827.32

N = 240, I = 6.25, PV = 250 000,
PMT = –1827.3205…, FV = 0, P/Y = 12, C/Y = 12

ii $188 557

1827.32 × 240 = 438 556.80

∴ 438 556.80 – 250 000 = 188 556.80

b $213 18

N = 60, I = 6.25, PV = 250 000, PMT = –1827.32,
FV = –213 117.8071…, P/Y = 12, C/Y = 12

c 212

N = 103.756 59…, I = 6.25, PV = 100 000, PMT = –1250,
FV = 0, P/Y = 12, C/Y = 12

∴ 104 + 108 = 212

Many students did not add the initial 108 weeks at the end.

[VCAA 2010 2RB Q4]

Question 12

Examination report

a **i** 300 months

N = 299.573…, I = 7.62, PV = 265 000, FV = 0,
P/Y = 12, C/Y =12

ii $3694.25

N = 12, I = 7.62, PV = 265 000, PMT = –1980,
FV = –261 305.747…, P/Y = 12, C/Y = 12

and 265 000 – 261 305.75 = $3694.25

b $2265.04

N = 12 × 19, I = 8.2, PV = 261 305.75,
PMT = –2265.04 …, FV = 0, P/Y = 12, C/Y = 12

[VCAA 2011 2RB Q4]

EXAM PREP 7.4 Annuities

Prep 1

a

Withdrawal number *n*	Withdrawal made ($)	Interest earned ($)	Principal reduction ($)	Value of investment ($)
0	0.00	0.00	0.00	380 000.00
1	3854.00	2850.00	1004.00	378 996.00
2	3854.00	2842.47	1011.53	377 984.47
3	3854.00	2834.88	1019.11	376 965.35

b $V_0 = 380\ 000,\ V_{n+1} = 1.0075\ V_n - 3854$

c The value of the investment after 5 months is \$374 904.13.

d approximately \$100 000 left after $12\frac{1}{2}$ years.

e The amounts involved are the same. The mathematics behind the two examples is identical, although the two situations in real life are opposites. Exam prep 7.1 Prep 2 involves someone paying money into an account, while this example involves someone withdrawing from an account.

Prep 2

417 months

Prep 3

\$3355.08

EXAM PRACTICE 7.4 Annuities

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1					▨	
Q2	▨					
Q3					▨	
Q4	46	10	17	23	4	2002
Q5	26	21	25	20	8	2005
Q6		▨				
Q7			▨			
Q8	▨					
Q9			▨			
Written response						
Q10	▨					

Question 4

The regular withdrawal is \$15 000 so A, C and D are the only possibilities.

$\left(1+\frac{4.54}{100}\right)=1.0454$, so A is the correct option.

Question 5

N = 4 × 4 = 16, I% = 6, PV = –10 000, PMT or Pmt = 500, FV = ?, PpY or P/Y = 4, CpY or C/Y = 4

Solving for FV gives –3723.67.

The value of the remaining investment is closest to \$3720.

Question 6

$$k=\left(1+\frac{r}{100}\right)=\left(1+\frac{7.5}{4}\times\frac{1}{100}\right)=\left(1+\frac{7.5}{400}\right)=1.018\ 75$$

Question 7

The balances on the graph are changing every six months and they take 8 years to reduce to zero.

Question 8

2000.00 – 634.56 = \$1365.44

Question 9

155 903.66 – 1376.39 = \$154 527.27

Question 10

a $a = 200\ 000,\ k = 1.0075,\ d = -2000$

b 186 months

c 93 months, with a last payment of \$2299.31

d 166 months, with a last payment of \$682.57

EXAM PREP 7.5 Perpetuities

Prep 1

\$27 300

Prep 2

\$66 666.67

EXAM PRACTICE 7.5 Perpetuities

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1				▨		
Q2	10	7	9	68	5	2008
Q3	20	11	16	7	44	2011
Q4	15	9	22	32	22	2013
Q5	14	41	9	11	23	2006
Q6	▨					2010
Q7	27					2012

Question 3

$$D=\frac{V_0\times r}{100}$$

$$D = 500$$

$$r=\frac{5.9}{52}$$

$$V_0 = ?$$

$$500=\frac{V_0\times 5.9}{5200}$$

$V_0 = 440\ 677.97$ (using a CAS/calculator solve function if necessary)

The amount she should invest is closest to \$441 000.

Question 4

Examination report

To answer this question correctly students needed to realise that, by definition, a simple perpetuity is designed to last indefinitely. This is achieved by paying out no more than the interest it earns. This ensures that a fixed amount can be paid from the perpetuity investment without any reduction in the amount of money originally invested, in this case \$100 000 (option E).

[VCAA 2013 1RBQ5]

Question 5

$$D = \frac{V_0 \times r}{100}$$

$D = 584$

$$r = \frac{6.2}{12}$$

$V_0 = ?$

$$584 = \frac{V_0 \times 6.2}{1200}$$

$V_0 = 113\ 032.26$ (using a CAS/calculator solve function if necessary)

The amount he has invested is closest to \$113 000.

Question 6

Examination report

a \$1560

$(360\ 000 \times 0.052) \times \frac{1}{12} = 1560$

b \$360 000

This question was poorly answered. Most students did not seem to understand that a perpetuity maintains the same principal amount while only regularly paying out 100% of all interest earned.

Many students gave $360\ 000 - 72 \times 1560 = \$247\ 680$ as their answer; however, this was incorrect.

[VCAA 2010 1RBQ2ab]

Question 7

$$D = \frac{V_0 \times r}{100}$$

$D = 1260$

$r = ?$

$V_0 = 80\ 000$

$$1260 = \frac{80000 \times \frac{r}{4}}{100}$$

$$1260 = \frac{80000 \times r}{400}$$

$$r = \frac{1260 \times 400}{80000} = 6.3\%$$ (using a CAS/calculator solve function if necessary)

Examination report

a 6.3%

$$\frac{1260 \times 4}{80\ 000} \times \frac{100}{1} = 6.3$$

This question was poorly answered. A common incorrect answer was $\frac{1260}{80\ 000} \times \frac{100}{1} = 1.575\%$.

b \$80 000

Most students did not show an understanding of a ‘perpetuity’, in which payments are calculated so that its original value (capital) is always preserved.

The most common incorrect answer was $80\ 000 - 20 \times 1260 = \$54\ 800$.

c \$35 208

N = 40, I% = 9.4, **PV = −35 208.002 54…**, PMT = 1260, FV = 7000, P/Y = 4, C/Y = 4

[VCAA 2012 1RBQ4]

EXAM PREP 7.6 Annuity investments

Prep 1

a

Deposit number n	Deposit made (\$)	Interest earned (\$)	Principal increase (\$)	Value of investment (\$)
0	0.00	0.00	0.00	500.00
1	500.00	40.00	540.00	1040.00
2	500.00	83.20	583.20	1623.20
3	500.00	129.86	629.86	2253.06

b The recurrence relation is $V_0 = 500,\ V_{n+1} = 1.08V_n + 500$.

c The value of the investment after 5 years is \$3667.96.

d \$15 000

Prep 2

a \$70 139.96 **b** \$17 139.96

Prep 3

\$1456.56

Prep 4

i E **ii** B **iii** D **iv** B

v E **vi** A **vii** F **viii** C

EXAM PRACTICE 7.6 Annuity investments

	Multiple choice					
	%A	%B	%C	%D	%E	
Q1	▨					
Q2			▨			
Q3	▨					
Q4		▨				
Q5			▨			
Q6				▨		
Q7	6	14	18	47	13	2010
Written response						
Q8	Not reported					2005
Q9	Not reported					2006
Q10	53					2011

Question 7

N = 2, I% = 5, PV = –4000, PMT or Pmt = –800, FV = ?, PpY or P/Y = 4, CpY or C/Y = 4

Solving for FV gives $5710.63, which is closest to $5711.

Question 8

Examination report

a $12 607.41 or $12 706.40

A mark was available here for a reasonable, correct answer using the student's incorrect answer to part a i.

Several students gave an answer of $25 480 396.50 without commenting on its validity.

Money answers should, unless otherwise stated, be written correct to the nearest cent. It was expected that $12 706.40 would not be written to only one decimal place.

b $11 848.58

This question presented a challenge to many students.

The question can be readily calculated with a finance solver. The answer is shown at *FV*.

N = 60, I = 4.8, PV = 4000, PMT = 100, FV = –11 848.58, P/Y = 12

c **i** $129.80

Regardless of prior answers, full, consequential marks were available to all students for parts c i and c ii, but many students did not seem to understand the requirements of these questions.

The value of the investment in part b after two years had to be found first. The correct value for this was $6915.90. A finance solver could then be used with the following data to calculate the new monthly deposits to achieve a $13 000 value after an additional three years.

N = 36, I = 4.8, PV = 6915.90, PMT = 129.80, FV = –13 000, P/Y = 12

c **ii** $1927.20

A mark was available here for students who had calculated the total deposits made in the investment for the three year period using an incorrect answer from part c i.

Total deposits = 4000 + 24 × 100 + 36 × 129.80 = $11 072.80

A mark was then available for finding the final answer by subtracting the total deposits from the $13 000 value at the end of five years.

[VCAA 2005 2RBQ3]

Question 9

$25 935

Question 10

Examination report

a $5833.60 **b** 0.072 **c** $78.42

Tom: Using a finance solver will have $8497.58

Patty: Using formula will have $8576.00

$8576.00 – $8497.58 = 78.42

Many students were able to find the value of Patty's investment correctly but then used their answer from part a (value after first month) instead of calculating Tom's investment after 12 months.

d 12.5%

$$100=\frac{8000\times r}{100}$$

[VCAA 2011 1RBQ2]

EXAMINATION Solutions

RECURSION AND FINANCIAL MODELLING EXAMINATION 1

1	2	3	4	5	6	7	8	9	10	11	12	13	14
A	E	D	C	C	B	D	E	C	D	E	C	A	B

Question 4

$$15\,000\times\left(1+\frac{4.60}{100\times 365}\right)^{150}\approx 15\,286$$

Question 5

Use a CAS/calculator finance solver.

Question 6

$$\text{Depreciation per year}=\frac{18\,000-5000}{4}$$

$$\text{Value after 1 year}=18\,000-\frac{18\,000-5000}{4}$$

Question 7

Use a CAS/calculator finance solver.

Question 8

Using a CAS/calculator finance solver: N = 48, FV = –3.57

The final (i.e. 48th) payment = 802.00 + 3.57 = 805.57.

Question 13

The balance after the deposit = $9000 – $2500 = $6500

I = 14.95% p.a., N = 24

Use a CAS/calculator finance solver.

Question 14

Use a CAS/calculator finance solver.

RECURSION AND FINANCIAL MODELLING EXAMINATION 2

Question 1

Examination report

a $5245.35

$$5000 \times \left(1+\frac{4.8}{100\times 12}\right)^{12} = 5425.3510\ldots$$

Optionally, a calculator TVM solver function could have been used in this question.

An incorrect answer of $5245.40 was often seen.

b i $7698.86

N = 12, **I %** = 4.8, **PV** = –5000, **PMT** = –200,
FV = **7698.8614…**, **P/Y** = 12, **C/Y** = 12

Students attempted to use various formulas that often resulted in calculation errors. The TVM function of technology provides an effective tool for solving annuity problems.

b ii $298.86

7698.86 – (5000 + 12 × 200) = 298.86

Many students did not allow for the $200 that Hugo had added each month for a year.

[VCAA 2013 2BRMQ2]

Question 2

Examination report

$11 029

Students needed to read this question carefully to identify the three steps needed for the solution. Writing and labelling such steps can be very helpful in organising thoughts, but very few students showed any TVM input or other working.

Stage 1 – Find the quarterly payment that would repay the loan in 4 years (16 repayments)

N = 16, **I %** = 12, **PV** = 25 000, **PMT** = **–1990.2712…**,
FV = 0, **P/Y** = 4, **C/Y** = 4

Stage 2 – Use the payment found from Stage 1 to find the principal remaining after 2 years (8 repayments)

N = 8, **I %** = 12, **PV** = 25 000, **PMT** = –1990.27 (from Stage 1 above), **FV** = **–13 971.09…**, **P/Y** = 4, **C/Y** = 4

To complete the answer,

Stage 3

Amount paid off the principal
= (loan value) – (principal remaining after eight months)
= (25 000) – (13 971.09)
= 11 028.91

The convention used above for input values is that any money coming to Hugo (such as a loan to him) is treated as positive. Any money leaving Hugo (such as paying out on a loan or investing money elsewhere) is treated as negative.

In Stage 2 above, the value of PV is positive since it is the loan Hugo had received. The value of PMT is negative because it is money that leaves Hugo when he pays it.

The value of FV is found to be negative. Following the convention used above, this negative value must be interpreted as an amount that will 'leave' Hugo in the future. In other words, the principal remaining is negative and he will still owe $13 971.09 after eight months.

An alternative sign convention for TVM input is merely the opposite of what has been used above. That is, any money coming to Hugo would be regarded as negative, and any money leaving Hugo would be regarded as positive. It does not matter which convention is used as long as interpretations are consistent with the chosen convention.

[VCAA 2013 2BRMQ4]

Question 3

Examination report

a 3.75%

A common, unacceptable answer was 0.0375%.

b $20 000

Many students did not understand what a perpetuity is. A perpetuity balance remains constant since only the interest earned is withdrawn in each compounding period.

[VCAA 2014 2BRMQ2ab]

Question 4

Examination report

a $14 450

b 6.9%

A common error was to treat this as a simple interest question, despite the question stating that there were 'four years of compounding interest'.

c **i** 0.006, 885

Answers equivalent to 0.006 were accepted in the first box, commonly $\frac{\frac{7.2}{100}}{12}$ or $\frac{\frac{7.2}{12}}{100}$. However, many students wrote only $\frac{7.2}{100}$ or $\frac{7.2}{12}$.

A number of students added a power to complete a formula to find the account balance at the end of 12 months, rather than the first month as required.

ii $75 443

N = 12, **I %** = 7.2, **PV** = –60 000, **PMT** = –885,
FV = 75 443.014…, **P/Y** = 12, **C/Y** = 12

[VCAA 2014 2BRMQ3]

Question 5

a $187 524.99 **b** $57 524.99

Question 6

a $A_{n+1} = 1.05A_n - 6000$ where $A_1 = 70\,000$

b $42 433.59 **c** $3500

Question 7

a $10 032 **b** 10% **c** $6013

GLOSSARY AND INDEX

68–95–99.7% rule A rule that says for approximately normal distributions, around 68% of the data values lie within *one* standard deviation of the mean, 95% of the data values lie within *two* standard deviations of the mean, and 99.7% of the data values lie within *three* standard deviations of the mean. (p. 57)

amortisation The repayment of a debt with a fixed schedule in regular instalments over a period of time. (p. 342)

amortisation table A table that shows the step-by-step calculations of how a loan is reduced. (p. 346)

annuity A type of investment where a sum is invested, interest is compounded at a fixed rate and withdrawals are made at regular intervals, usually until the value of the investment is $0. (p. 374)

annuity investment An investment that involves making an initial deposit followed by additional regular payments into an account earning a fixed rate of compound interest. (p. 389)

back-to-back stem plots A statistical graph used when dealing with two sets of data values for the same variable where the original data values are visible. (p. 96)

bar chart A graphical display used for categorical data where the frequency of each different category is shown using a vertical column or a horizontal bar. (p. 7)

bell-shaped distribution *See* normal distribution.

bi-modal A distribution with two modes. (p. 17)

book value *See* future value (in relation to depreciation).

box-and-whisker plot *See* boxplot.

boxplot (**box-and-whisker plot**) A graphical display of numerical data based on the five-number summary, IQR and outliers. (p. 99)

categorical variables Variables that represent qualities that can't be counted or measured. (p. 85)

causation A high level of correlation between two variables where one variable is causing a change in the other. (p. 128)

centring The extra step of taking two-point moving means of the smoothed values when smoothing with an even number of points. (p. 256)

coefficient of determination A measure of how useful a line of best fit is as a linear model for a particular set of data. (p. 171)

coincidence A high level of correlation between two variables occurring by chance without any underlying cause. (p. 143)

common response A high level of correlation between two variables where a third factor is causing the correlation. (p. 129)

compound interest The interest calculated at the end of a certain time period, added on to the principal, and then calculated for the next time period using this new amount. (p. 345, 361)

compounding period The length of the time period that elapses before interest compounds. (p. 345)

confounding A high level of correlation between two variables where it's unclear how the variables are related and there may be an unknown factor involved. (p. 129)

confounding variable An unknown factor that may be causing a high level of correlation between two variables. (p. 129)

contingency table *See* two-way frequency table.

continuous numerical variables Numerical variables that can be measured to an increasing level of accuracy. (p. 3)

data Observations or measurements collected about a variable from which conclusions may be drawn. (p. 2)

dependent variable *See* response variable.

depreciation The decrease in value of assets bought by a business over time. (p. 326)

de-seasonalisation The process of using seasonal indices to remove the seasonal component of time series data, using the formula

$$\text{De-seasonalised value} = \frac{\text{Actual value}}{\text{Seasonal index}}. \text{ (p. 240)}$$

difference equation *See* recurrence relation.

discontinuity (in a time series) A structural change in a time series which is a clear break. (p. 217)

discrete numerical variables Numerical variables that can't be measured to an increasing level of accuracy. (p. 3)

dot plot Graphical display for categorical or numerical discrete data. (p. 32)

effective interest rate The interest rate after the compounding periods have been taken into account. (p. 326)

experimentation A way in which data can be gathered where the researcher actively manipulates a situation to eliminate possible confounding *before* observing it. (p. 144)

explanatory variable (**independent variable**) A variable that we expect to affect another variable. (p. 82)

extrapolation A prediction made outside the original data range. (p. 183)

first-order recurrence relation A rule that specifies how a particular term in a sequence can be found from the previous term in the same sequence consisting of a rule linking two consecutive terms and the value of a term. (p. 270)

five-number summary Five key points in a data distribution consisting of the minimum value, the lower quartile, the median, the upper quartile and the maximum value. (p. 21)

five-point moving means smoothing A numerical smoothing technique which involves finding means of consecutive sets of 5 data points. (p. 256)

five-point moving median smoothing A graphical smoothing technique which involves finding medians of consecutive sets of 5 data points. (p. 235)

flat rate depreciation (**straight-line depreciation**) Depreciation where the future value of an asset is reduced by a fixed amount every year, expressed either in dollars or as a fixed percentage of the purchase price. (p. 284)

forecasting The process of making predictions outside the data range for a time series. (p. 278)

frequency tables Table used to organise large amounts of data with data values in one column and the corresponding frequencies in another. (p. 6)

future value The new reduced value of an asset being depreciated at any point in time or the balance of loans and investments at any point in time. (p. 326)

gradient *See* slope.

graphical smoothing A smoothing technique that involves working directly from a time series plot. (p. 263)

histogram Graphical display for numerical data (discrete or continuous) with vertical joined columns. (p. 119)

independent variable *See* explanatory variable.

interest The cost of using money that isn't yours. (p. 356)

interest-only loan A loan where the regular repayments exactly equal the interest earned and the balance owing on the loan doesn't reduce. (p. 361)

interpolation A prediction made within the original data range. (p. 183)

interquartile range (IQR) The measure of the spread of the middle 50% of the data values. (p. 22)

irregular fluctuation (in time series) Time series data that appears to occur at random with no pattern. (p. 216)

least squares line of best fit (**least squares regression line**) The method of finding a line of best fit that minimises the sum of the squares of the vertical distances between the line and each data point in a scatterplot. (p. 138)

least squares regression line *See* least squares line of best fit.

line of best fit A straight line that is the best approximation for a set of data. (p. 158)

linear scale A scale used on a plot or graph where the same number is *added* to get from one scale mark to the next. (p. 44)

log scale A scale used on a plot or graph where the same number is multiplied to get from one scale mark to the next. (p. 50)

log transformation A transformation involving finding the log of either the x values or the y values. (p. 189)

lower fence The value below which a data point may be considered an outlier. (p. 22)

lower quartile The data point that has 25% of the data below it. (p. 20, 34)

lurking variable A third factor causing a high level of correlation between two variables. (p. 129)

mean The value often referred to in everyday life as the average, calculated using $\bar{x} = \frac{\Sigma x}{n}$ and considered to be a measure of centre. (p. 109)

median The middle value when a data set is ordered from smallest to largest. It is considered to be a measure of centre. (p. 14)

mode The most common data value in a set. It is considered to be a measure of centre. (p. 33)

moving means A numerical smoothing technique which involves finding a series of means of a fixed number of data points. (p. 254)

moving medians The graphical smoothing technique which involves finding a series of medians of a fixed number of data points. (p. 263)

negatively skewed Description of a distribution that has a tail at the lower end. (p. 27)

nominal interest rate The interest rate quoted for a loan or investment. (p. 374)

normal distribution (**bell-shaped distribution**) A distribution with a bell shape which is symmetrical about the mean, peaks in the centre and tails off to zero on both sides. (p. 57)

numerical data Variables represented by quantities or measurements that can be either discrete or continuous. (p. 32)

numerical smoothing A smoothing technique that involves numbers and calculations. (p. 254)

numerical variables Variables that represent quantities that can be counted or measured. (p. 3)

observational study A way in which data can be gathered where the researcher passively observes an existing situation. (p. 130)

ordered stem plots A stem plot where the leaves are ordered from smallest to largest. (p. 34)

outlier An extreme high or low value in the data. (p. 14)

parallel boxplots A graph where two or more boxplots are shown on the same axis. (p. 99)

parallel dot plots A graph where two or more dot plots are shown on the same axes. (p. 110)

Pearson correlation coefficient (**product-moment correlation coefficient**) A number on a scale from −1 to 1 that measures the strength and direction of *linear* associations. (p. 208)

percentage segmented bar chart A segmented bar chart where the bars represent percentages and the height of the bar is 100. (p. 7, 100)

percentage two-way frequency table A two-way frequency table where the data values have been converted into percentages. (p. 99)

percentaging The process of converting values into percentages. (p. 99)

perpetuity A type of annuity where a permanently invested amount of money provides regular payments that continue forever and the balance of the amount invested stays the same forever. (p. 384)

population All items of the group being studied. (p. 129)

population mean The mean of a population. (p. 71)

population parameter A number that describes a characteristic of a population. (p. 71)

population standard deviation The standard deviation of a population. (p. 71)

positively skewed distribution Description of a distribution that has a tail at the upper end. (p. 51)

present value The current value of an asset, loan or investment. (p. 344)

principal The amount of money invested or borrowed. (p. 356)

product-moment correlation coefficient *See* Pearson correlation coefficient.

quartiles The three points that divide a set of data into quarters.

range A measure of the spread of data which is the difference between the largest and smallest data values. (p. 14)

reciprocal transformation A transformation involving taking the reciprocal of either the x values or the y values. (p. 190)

recurrence relation (**difference equation**) A rule that specifies a particular term in a sequence using the previous term or terms. (p. 270)

reducing balance depreciation Depreciation where the future value of an asset is reduced every year by a fixed percentage of its value in the preceding year. (p. 339)

reducing balance loan A loan where interest is calculated on the amount still owing after each repayment is made. (p. 342)

re-seasonalisation The process of using seasonal indices to find the original non-seasonalised value using the formula
Actual value = De-seasonalised value
× Seasonal index. (p. 248)

residual The vertical distance between each data point and the least squares line of best fit.

residual plot A plot of with the explanatory variable on the x-axis and the residual values on the y-axis. (p. 193)

response variable (**dependent variable**) A variable that we expect to be affected or changed by another variable. (p. 82)

Richter magnitude scale A log base 10 scale involving measuring the amplitude of the largest wave of an earthquake. (p. 51)

sample A section of the population that is selected so that it is representative of the population and can be used to draw accurate conclusions about the population. (p. 28)

sample mean The mean of a sample. (p. 55)

sample standard deviation The standard deviation of a sample, which has a slightly different formula from the population standard deviation. (p. 56)

sample statistic A number calculated from sample data that approximates a population parameter. (p. 71)

scatterplot A graph used to compare two numerical variables where the explanatory variable is plotted on the x-axis and the response variable on the y-axis. (p. 126)

seasonal adjustment An adjustment made to time series data to eliminate seasonality. (p. 258)

seasonal indices Values used to make seasonal adjustments to time series data. (p. 269)

seasonality Time series data which has regular and predictable changes repeated across a year or less. (p. 244)

segmented bar chart A statistical graph used to display the information from a two-way frequency table where the bars have several segments, each representing frequency. (p. 7)

shape of data A description of data as symmetrical, positively skewed or negatively skewed. (p. 12)

significant figures (rounding to a number of) A method of rounding involving all the non-zero digits of a number plus the zeros that are included between them or that are final zeros and signify accuracy. (p. 139)

simple interest The interest calculated as a percentage of the amount of money invested or borrowed. (p. 356)

simple random sample The simplest way to get a representative sample from a population in which everything or everyone in the population has an equal chance of being chosen. (p. 83)

slope (**gradient**) (of the least squares line of best fit) The coefficient $b = r\frac{s_y}{s_x}$ in the equation for the least squares line of best fit $y = a + bx$. (p. 127)

smoothing A technique for levelling out fluctuations in time series data to produce a smoother graph which allows us to see trends more clearly. (p. 254)

squared transformation A transformation involving squaring either the x values or the y values. (p. 189)

standard deviation A measurement of the spread of data about the mean. (p. 208)

standardise The process of converting data values from normal distributions to standardised values.

standardised values (**z-scores**) Values calculated using the formula

$$\text{standardised value} = \frac{\text{data value} - \text{mean}}{\text{standard deviation}}$$

that allow us to compare values from different normal distributions. (p. 65)

stem plot (**stem-and-leaf plot**) A graphical display for numerical data that is either discrete or continuous. (p. 103)

stem-and-leaf plot *See* stem plot.

straight-line depreciation *See* flat rate depreciation.

structural change (in statistics) An unexpected shift in the data pattern of a time series. (p. 245)

symmetric distribution A distribution that is symmetric in shape. (p. 12)

three-point moving means smoothing A numerical smoothing technique that involves finding means of consecutive sets of 3 data points. (p. 227)

three-point moving median smoothing A graphical smoothing technique that involves finding medians of consecutive sets of 3 data points. (p. 235)

time series Data representing an association between two variables where the explanatory variable is time. (p. 240)

time series plot A scatterplot where time is shown on the horizontal axis in regular intervals and the data points are joined. (p. 240)

transformation A way of changing a non-linear association so that the association between the two variables becomes closer to a straight line. (p. 207)

trend line A line of best fit for time series. (p. 278)

trend The long-term direction of time series data. (p. 244)

two-point moving means The process of finding means of consecutive pairs of data points used as an extra step when smoothing with an even number of points. (p. 227)

two-way frequency table (**contingency table**) A frequency table used to explore the association between two categorical variables which has at least two categories for each variable. (p. 85)

two-way table A table used to display information when comparing two categorical variables. (p. 87)

unit cost depreciation Depreciation where the future value of an asset is reduced every year according to the amount of use it has had, not according to its age. (p. 282)

upper fence The value above which a data point may be considered an outlier. (p. 22)

upper quartile The data point that that has 75% of the data below it. (p. 20, 34)

variable (in statistics) Something measurable or observable that changes between individual observations or over time. (p. 82)

whiskers Parts of a boxplot that show the minimum and maximum values if there are no outliers. (p. 102)

y-intercept (for least squares line of best fit) The coefficient $a = \bar{y} - b\bar{x}$ in the equation for the least squares line of best fit $y = a + bx$. (p. 139)

z-scores *See* standardised values.

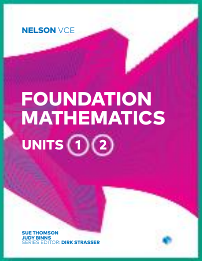
nelsonnet

Register online at:

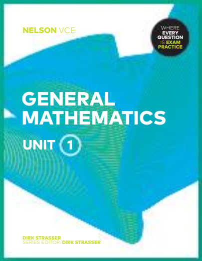
www.nelsonnet.com.au

to access digital resources to support improved learning & teaching*

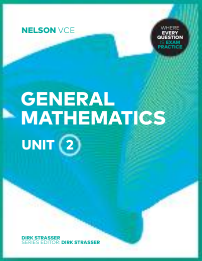
- Introducing Nelson VCE Maths: the only VCE series where every question is exam practice, whose structure and content are based on actual VCE exam questions
- Clear and concise theory and worked examples
- VCE exam questions from 2002 to 2014 integrated into each book
- Contains Exam prep and Exam practice exercises
- Using CAS contains instructions for the TI-Nspire and ClassPad calculators, integrated into all chapters
- Exam hack provides 'insider advice' to help students overcome common errors, achieve learning success and gain valuable marks
- Practice exams in Units 3 and 4 cover a number of related topics in an area of study
- Solutions at the back of the book include worked solutions to selected questions, with examiners' comments and student performance percentages from the VCAA examination reports
- Worked solutions to all questions can be downloaded by teachers*
- All teacher resource material included in the price of the book*



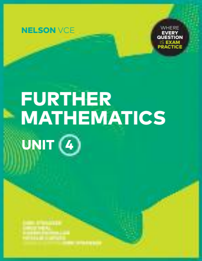
9780170370752
Foundation
Mathematics
Units 1 & 2



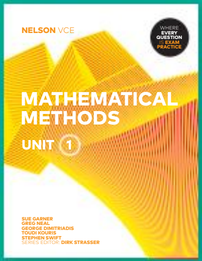
9780170370783
General
Mathematics
Unit 1



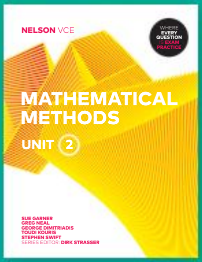
9780170370868
General
Mathematics
Unit 2



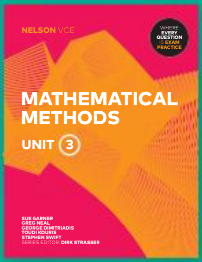
9780170371186
Further
Mathematics
Unit 4



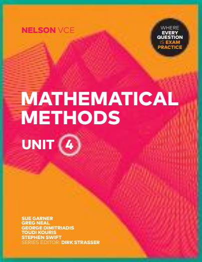
9780170370943
Mathematical
Methods
Unit 1



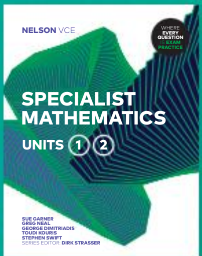
9780170371025
Mathematical
Methods
Unit 2



9780170371261
Mathematical
Methods
Unit 3



9780170371346
Mathematical
Methods
Unit 4



Specialist
Mathematics
Units 1 & 2

Specialist
Mathematics
Units 3 & 4

NELSON
CENGAGE Learning®
For learning solutions, visit **cengage.com.au**

*Please note that complimentary access to NelsonNet is only available to teachers who use the accompanying student textbook as a core educational resource in their classroom. Contact your sales representative for information about access codes and conditions.

ISBN: 978-0170371100
9 780170 371100