7

STATISTICS AND PROBABILITY

INVESTIGATING DATA

The Bureau of Meteorology collects weather data across Australia, including rainfall and temperatures. Polling organisations like Roy Morgan and Galaxy Research collect data on people's opinions on a wide range of issues. In NSW, the Bureau of Crime Statistics and Research (BOCSAR) collects data on crimes in different areas of the state. Health departments collect data on medical issues and health care. In all cases the data is analysed to evaluate decisions already made and to inform decision making in the future.

Shutterstock.com/r.classen

Chapter outline

	Working mathematically				
7.01 Organising and displaying data	U	F		R	C
7.02 Types of data	U	F		R	C
7.03 The mean and mode	U	F		R	C
7.04 The median and range	U	F	PS	R	C
7.05 Analysing frequency tables	U	F			
7.06 Dot plots and stem-and-leaf plots	U	F	PS	R	C
7.07 Frequency histograms and polygons	U	F	PS	R	
7.08 Comparing data	U	F	PS	R	C
7.09 Sampling	U	F		R	C
7.10 Designing survey questions	U	F	PS	R	C
7.11 Comparing samples and populations	U	F		R	C

Wordbank

bias Something unwanted that causes a sample to not truly represent the population

categorical data Non-numerical data that can be classified into categories (for example, people's hair colour, or their favourite football team)

frequency How often a value appears in a data set

frequency histogram A special column graph that shows the frequencies of numerical data

frequency polygon A special line graph that shows the frequencies of numerical data

measure of location (or **measure of central tendency**) An average, middle or typical value of a data set; a general name for the mean, median or mode

median The middle value of a data set when the values are arranged in order, or the average of the 2 middle values

numerical data Data that can be measured and counted using values (for example people's height, or the number of devices they own)

U = Understanding | **F** = Fluency | **PS** = Problem solving | **R** = Reasoning | **C** = Communication

Statistics

In this chapter you will:

- interpret and draw divided bar graphs, sector graphs, line graphs, stem-and-leaf plots and dot plots
- organise data into a frequency table
- identify numerical and categorical data, and sort numerical data into discrete and continuous data
- calculate and interpret the mean, median, mode and range of a set of data, including those represented on a graph or plot
- interpret and draw frequency histograms and polygons
- investigate the effect of individual data values, including outliers, on the mean and median
- identify ways of collecting data, including census, surveys, sampling and observation
- design surveys and survey questions that are clear and unbiased
- compare samples and populations, especially their means and proportions

SkillCheck ANSWERS ON P. 564

1 A group of school students visiting Japan have the following ages.

14 13 12 15 13 14 12 14 15 14 14 13 13

Show this information on a dot plot.

2 The masses (in kilograms) of some hospital patients are:

68 59 63 68 54 48 49 64 47 48 59 68 30

a Show this data on an ordered stem-and-leaf plot.

b What is the outlier?

3 A group of people was surveyed about the number of phone calls they made on a Saturday night. The results are shown below.

6 4 0 3 1 2 2 3 5 4 3 3

4 1 0 6 6 3 4 3 1 5 4 3

a Draw a column graph for this data.

b How many people were surveyed?

c What was the most common number of phone calls?

d What was the least common number of phone calls?

e What was the highest number of phone calls?

4 This back-to-back stem-and-leaf plot shows the points scored for and against the Koala Rugby team in each match, at one stage in the season.

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

a What is the highest number of points scored by the Koalas in one game?

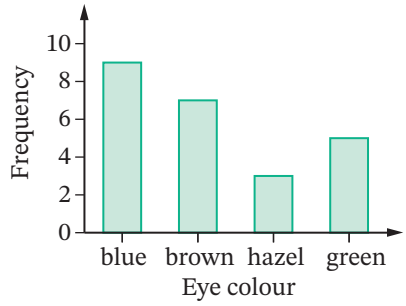
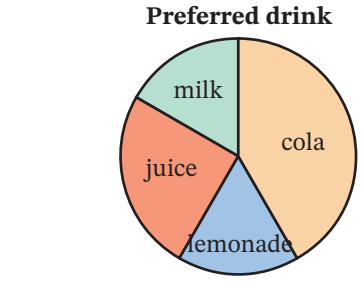
b How many matches have the Koalas played so far?

c How many points have been scored against the Koalas?

Organising and displaying data 7.01

Data is a collection of facts or raw information. Data can be organised and displayed in a variety of ways so that we can **analyse** (study) it in more detail. This table shows 3 examples of statistical graphs.

Displaying data
Every picture tells a story
Where all the cars are red

Graph	Description
Column graph (or bar chart) 	• Mostly used for data that are in categories • Uses columns to show the quantity or frequency of each category • The horizontal axis (going across) shows the categories • The vertical axis (going up) shows the frequency
Sector graph (or pie chart) 	• Shows parts of a whole in circle form • Uses sectors of a circle to show sizes of parts
Divided bar graph 	• Shows parts of a whole in rectangular form • Uses sections of a rectangle to show sizes of parts

Example 1

A sample of 80 people was surveyed on their favourite TV sport, and the results are shown.

TV sport	Australian football	Cricket	Motor racing	Rugby league	Rugby union	Football
Frequency	13	16	8	19	10	14

Display this data on a:

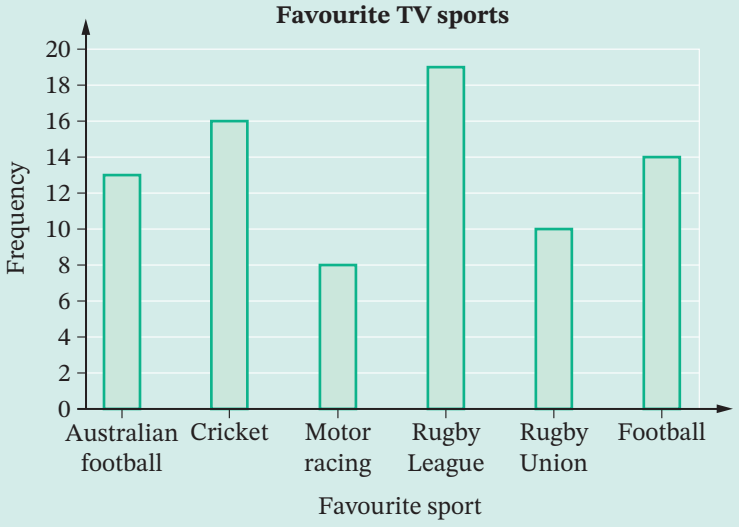
a column graph

b sector graph

c divided bar graph.

Solution

a The highest frequency is 19 (for rugby league), so use a scale of 1 unit = 2 on the 'Frequency' axis. Label both axes, draw a column for each category, then label the graph.



b Draw a large circle, then calculate the angle sizes for each sector using fractions of 360° (because there are 360° in a revolution).

Total number of people in sample = 80.

Sector angles:

Australian football: $\frac{13}{80} \times 360° = 58.5°$

Cricket: $\frac{16}{80} \times 360° = 72°$

Motor racing: $\frac{8}{80} \times 360° = 36°$

Rugby League: $\frac{19}{80} \times 360° = 85.5°$

Rugby Union: $\frac{10}{80} \times 360° = 45°$

Football: $\frac{14}{80} \times 360° = 63°$

Construct the sector graph with the calculated angles, then label the sectors and the graph.

Favourite TV sports

c Draw a rectangle whose length will represent 80, the total frequency (the whole), for example, 12 cm. Then calculate the length of each section using fractions of 12 cm.

Section sizes:

Australian football: $\frac{13}{80} \times 12\text{ cm} = 1.95\text{ cm}$

Cricket: $\frac{16}{80} \times 12\text{ cm} = 2.4\text{ cm}$

Motor racing: $\frac{8}{80} \times 12\text{ cm} = 1.2\text{ cm}$

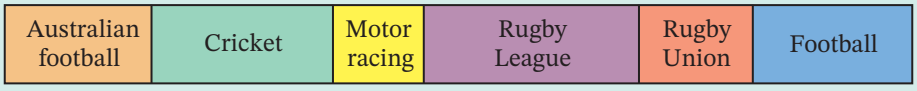
Rugby League: $\frac{19}{80} \times 12\text{ cm} = 2.85\text{ cm}$

Rugby Union: $\frac{10}{80} \times 12\text{ cm} = 1.5\text{ cm}$

Football: $\frac{14}{80} \times 12\text{ cm} = 2.1\text{ cm}$

Construct the divided bar graph with the calculated section lengths, then label the sections and the graph.

Favourite TV sports



Frequency tables

WS Frequency tables

A **frequency table** (or **frequency distribution table**) can be used to efficiently organise and count a large set of numerical data values. The **frequency** of a value is the number of times that value appears in the set of data.

Example 2

20 people were surveyed about the number of cars in their household.

2	2	1	2	0	3	2	1	1	4
1	1	1	2	2	0	3	2	1	2

Complete a frequency table for the data.

Solution

Draw a table with columns for Score, Tally and Frequency, as shown next page.
The lowest score is 0 and the highest is 4, so list the numbers 0 to 4 in the Score column.

For each value in the data set, draw a tally mark in the tally column. Tally means 'to count', and we are counting how often each value occurs.

The first value is a 2, so mark a '|' in the 'Tally' column beside 2. The second value is another 2, so mark another '|' there. Continue until you have counted all of the values.

Score	Tally	Frequency
0	II	
1	𝍸 II	𝍸 means a tally of 5.
2	𝍸 III	
3	II	
4	I	
Total		

For each value, count up the tally marks and write the answer in the 'Frequency' column. Then write the total in the final row and check that it is equal to the total number of values.

Score	Tally	Frequency
0	II	2
1	𝍸 II	7
2	𝍸 III	8
3	II	2
4	I	1
Total		20

EXERCISE 7.01 ANSWERS ON P. 564

Organising and displaying data U F R C

EXAMPLE 1

1 A class of 30 PE students was surveyed about the day of the week that each student trained for their chosen sport.

Day	Monday	Tuesday	Wednesday	Thursday	Friday
Number of students	3	13	4	7	3

Display this data in a sector graph. Label the graph. C

2 The colours of the jelly beans in a jar were counted.

Colour	Red	Black	Purple	Pink	Orange	White
Frequency	11	5	2	4	6	8

Display this data as a column graph, using an appropriate scale. Label the graph and both axes. C

Shutterstock.com/Hayati Kayhan

3 A sample of men were surveyed on the number of hours they slept the previous night.

Number of hours	6	7	8	9	10	11
Frequency	3	9	11	6	5	2

Display this data as a divided bar graph, 9 cm long. Label the graph. C

4 This table shows the number of kilolitres of water used by a household in different areas over a year.

Area	Garden	Shower	Toilet	Kitchen
Water usage (kL)	90	520	140	50

If this information is shown in a divided bar graph of length 10 cm, which of the following is the length of the section representing 'Shower'? Select the correct answer **A**, **B**, **C** or **D**. R

A 1.5 cm **B** 1.9 cm **C** 5.2 cm **D** 6.5 cm

7.01

EXAMPLE 2

5 The results of a spelling quiz (out of 10) for a class of students are as shown. C

7	7	8	9	7	5	4	8	9
9	8	10	8	7	5	5	4	1
8	8	9	7	7	4	8	3	1

a Complete a frequency table for this data.

b What was the highest value?

c How many students scored 4 marks in the quiz?

d What was the most frequent value?

e Draw a column graph of this data.

6 30 customers were asked to name their favourite car colour. B = black, S = silver, W = white, U = blue, R = red. C

B	B	W	R	W	R	B	S	U
W	R	B	S	B	U	W	B	W
S	B	B	R	R	U	W	B	B

a Complete a frequency table for this data.

b Draw a divided bar graph for this data.

c Which is the most popular colour for this group of customers?

7 A sample of women were asked how many siblings (brothers and sisters) they had. C

4	0	2	3	2	4	3	5	4	3
0	0	1	2	3	2	2	1	4	3
1	1	1	5	2	5	0	1	2	1
0	1	3	2	3	4	1	1	3	3

a Complete a frequency table for this data.

b What was the highest number of siblings?

c What was the mode (the most frequent value)?

d Illustrate this data on a sector graph.

Creating a graph

Technology

Graphing life expectancy

Life expectancy describes the age to which a person is expected to live, based on statistical data covering different factors.

1 Enter the data below into a new spreadsheet. It compares the life expectancy at birth for males and females in 10 countries.

	A	B	C	D
1	**Life expectancy at birth**			
2			**Years**	
3	**Country**	**Male**	**Female**	**Total**
4	Australia	80.5	85.0	82.7
5	Bangladesh	72.0	76.5	74.2
6	Brazil	71.2	78.4	74.7
7	Canada	81.1	85.9	83.4
8	Chad	56.5	60.1	79.4
9	China	74.0	78.4	76.1
10	India	68.4	71.2	69.7
11	Spain	79.0	85.2	82.0
12	Sweden	80.4	84.5	82.4
13	USA	78.0	82.5	80.3

Source: www.cia.gov

2 Highlight all 4 columns, including headings. Click **Insert** and **2-D Column Graph**. Give the graph an appropriate title and label the axes.

3 Which country has:

- **a** the lowest total life expectancy?
- **b** the highest total life expectancy?
- **c** the lowest male life expectancy?
- **d** the lowest female life expectancy?
- **e** the highest male life expectancy?
- **f** the highest female life expectancy?

4 Suggest reasons why some countries would have higher life expectancies than others.

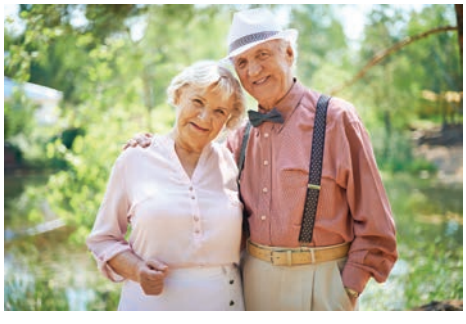
5 In column E, calculate the difference between the male and female life expectancies for each country.

6 Generally, do males or females live longer?

7 Which country has the greatest difference between its male and female life expectancies?

8 Which country has the least difference?

Shutterstock.com/antpkr



Shutterstock.com/Pressmaster

Types of data

7.02

Statistical data may be in the form of categories (such as hair colour or favourite sport), or in the form of numbers (such as soccer score or number of cars owned). There are 2 types of data.

Categorical data can be grouped into categories, such as a type of seafood or the suburb where you live.

Numerical data (or **quantitative data**) can be counted or measured, such as number of children or a person's height.

Example 3

Classify each of the following types of data as categorical or numerical.

a make of car

b weekly pay

c weight of athlete

d favourite radio station

Solution

a Make of car is categorical data. Involves words or categories

b Weekly pay is numerical data. Involves numbers

c Weight of athlete is numerical data. Involves numbers

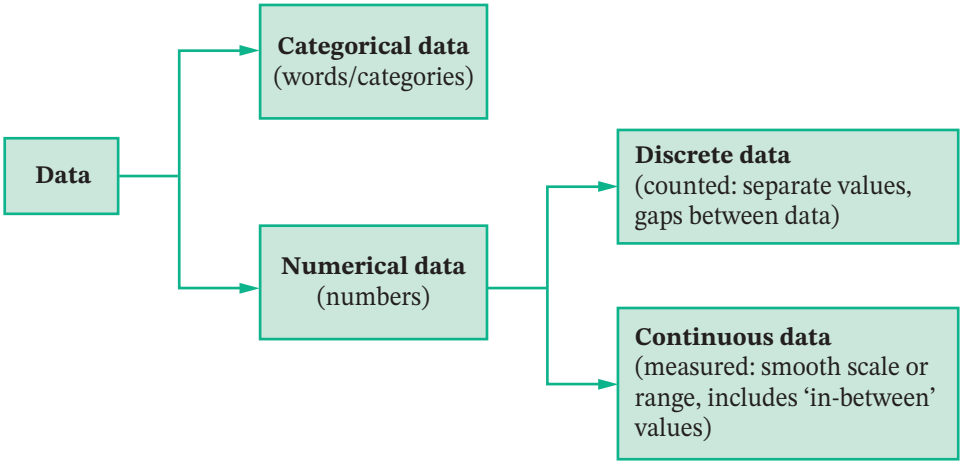
d Favourite radio station is categorical data. Involves words or categories

Types of numerical data

Numerical data can be further divided into **discrete data** and **continuous data**.

Discrete data are counted or measured and can only take on separate, distinct values, with 'gaps' or 'jumps'. 'In-between' values are not allowed. For example, the number of children in a family is discrete because it could be 0, 1, 2, etc., but never 2.3 or $1\frac{2}{5}$.

Continuous data are measured on a 'smooth' scale without any gaps, and can take on a full range of values. 'In-between' values are allowed. For example, a town's temperature is continuous because it could be 28°C, 29°C, 29.5°C, 29.8°C, etc.



Types of data

Example 4

Classify each type of numerical data as discrete or continuous:

a number of rooms in a home

b loudness of aeroplane

c shoe size

d weight of whale

Shutterstock.com/Paul S. Wolf

Solution

a Number of rooms is discrete data. Takes on separate values

b Loudness of aeroplane is continuous data. Takes on a smooth scale of values

c Shoe size is discrete data. Takes on separate values

d Weight of whale is continuous data. Takes on a smooth scale of values

EXERCISE 7.02 ANSWERS ON P. 565

Types of data U F R C

EXAMPLE 3

1 Classify each type of data as categorical (C) or numerical (N). R C

a mass of student

b brand of computers

c price of ice cream

d number of TV sets owned by family

e amount of rainfall

f cause of tooth decay

g sets won in a tennis match

h speed of truck

EXAMPLE 4

2 Classify each type of numerical data as discrete (D) or continuous (C). R C

a number of homes with solar power

b patients' body temperature

c sizes of the crowd at a rock concert

d star ratings of hotel

e times taken for athlete to run 100 metres

f school population

g amount of cough medicine taken by child

h number of road accidents in Grafton

3 Which of the following is **not** discrete data? Select the correct answer **A**, **B**, **C** or **D**. C

A The number of shoes you own

B The number of blue-eyed students at your school

C The number of options in a multiple-choice test

D The maximum daily temperature

Foundation Standard Complex

9780170453141

4 Classify each type of data as categorical (C), numerical discrete (ND) or numerical continuous (NC). R C

a hair colour

b customer rating of a restaurant's level of service (poor, satisfactory, good, very good)

c number of points a player scored in a netball match

d film classification (G, PG, M, MA15+, R)

e the age you turn this year

f height of bridge

g number of staff employed

h country where cars were manufactured

The mean and mode 7.03

When we analyse data, we summarise the information and try to find patterns and trends to draw some conclusions from them. We can use the **mean**, **mode**, **median** and **range** to make comparisons between sets of data.

Statistical measures

Looking for gold

The mean, mode and median are known as **measures of location** (averages) or **measures of central tendency**: they are typical or central values that represent all of the data in the set.

The range is a **measure of spread**: it describes how spread out the values are.

The **mean** is found by adding all the data values and dividing by the number of values.

The mean

The symbol for the mean is $\bar{x}$.

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

- The mean is a single value that represents all of the data.
- *All* of the data values are used in the calculation of the mean.
- The mean is *not* necessarily one of the data values.

The **mode** is the most common or frequent value (or values). A set of data may have more than one mode, or no mode at all.

The mode

The **mode** is the data value (or values) that occurs the most often: the value with highest frequency.

Think: mode = '**m**ost **o**ften'

- The mode can be one or more values.
- The mode does not depend on all of the values in the data set.
- The mode is one (or more) of the data values.

Foundation Standard Complex

Example 5

The number of goals scored by Courtney's soccer team in each match over a season were:

3 2 1 4 0 2 3 1 3 1 1 0

a Find the mean. **b** Find the mean.

Solution

a Mean $\bar{x} = \frac{\text{sum of data values}}{\text{number of values}}$

$= \frac{3+2+1+4+0+2+3+1+3+1+1+0}{12}$

$= \frac{21}{12}$

$= 1.75$

Note that the value of the mean is at the centre of the data.

b The mode is 1.

It is the most common value (occurs 4 times).

The mean from a calculator's statistics mode

Scientific and graphics calculators have a statistics mode (SD or STAT). Follow the instructions in the table below to calculate the mean of the values from Example **5** using your calculator's statistics mode.

Operation	Casio scientific	Sharp scientific
Start statistics mode.	MODE STAT 1-VAR	MODE STAT =
Clear the statistical memory.	SHIFT 1 Edit, Del-A	2nd F DEL
Enter data.	SHIFT 1 Data to get table 3 = 2 =, etc. to enter in column. Press AC to leave table.	3 M+ 2 M+, etc.
Calculate the mean. ($\bar{x} = 1.75$)	SHIFT 1 Var $\bar{x}$ =	RCL $\bar{x}$
Check the number of values. (n = 12)	SHIFT 1 Var n =	RCL n
Return to normal (COMP) mode.	MODE COMP	MODE 0

EXERCISE 7.03 ANSWERS ON P. 565

The mean and mode UFRC

EXAMPLE 5

1 The residents in a street were surveyed about the number of children living in each household. The results were:

2 2 1 2 0 3 2 1 1 4 1 0

a Find the mean, correct to one decimal place.

b Find the mode.

2 For each set of data, find:

i the mean (rounded to 2 decimal places, if needed)

ii the mode(s).

a 6 8 7 6

b 7 9 5 2 2 6 2

c 11 7 7 6 1 2 4 1

d 96 53 96 45 64 76 54 54 96

e 47.31 47.20 47.24 47.21 47.28 47.26

3 Use your calculator's statistics mode to calculate the mean of each data set in question **2**.

4 Find the mode of each set of categorical data.

a Holden, Ford, Nissan, Holden, Hyundai, Ford, Hyundai, Nissan, Hyundai, Toyota, Toyota

b heart, spade, spade, diamond, spade, club, heart, heart, diamond, spade

c Chinese, Korean, Japanese, Korean, Chinese, Japanese, Japanese, Korean, Korean

5 Why is it not possible to find the **mean** of each data set in question **4**? R C

6 In a diving competition, the judges awarded the following scores.

6.2 7.0 8.2 8.8 7.0 9.1 7.0 8.1 8.0

a Find the mean of these scores, correct to 2 decimal places.

b Find the mode of these scores.

c Which measure of location represents this set of scores better? Give a reason for your answer. R C

7 What is the difference between the mean and the mode of the following set of scores?

20 60 20 50 40

Select the correct answer **A**, **B**, **C** or **D**.

A 20 **B** 12 **C** 18 **D** 2

8 The ages of the members of the Carrozza family are:

1 3 11 25 26 30 31 52 58

The ages of the members of the Binns family are: R C

5 8 16 21 25 29 40 63 93

a Calculate the mean age (correct to one decimal place) of each family.

b What is the main difference between these 2 sets of ages?

c What effect does this difference have on the means?

9 2 Year 8 classes scored the following marks for their PE assignments. R C

8Y:	5	5	5	5	6	6	6	7	7	7	8
8Z:	3	3	4	5	5	6	6	7	9	9	10

a Calculate the mean mark (correct to one decimal place) for each class.

b Describe the differences between the way each class has its marks spread out.

c Class 8X scored the following marks, but Hannah was away.

8X:	4	5	5	6	6	6	7	7	7	10

What mark would Hannah have to score so that her class has the same mean as the other 2 classes?

10 The mean points score per game by a basketball team over a season of 30 games was 98. What was the total number of points scored over the season? R

11 The takings at the school canteen were: R

Monday $1600 Tuesday $2400 Wednesday $1500 Thursday $1750

a Find the average takings for these 4 days.

b How much should the canteen earn on Friday to make the average daily takings $2000?

12 **a** Find 4 values that have a mean of 10 and a mode of 11. R

b Find 7 values that have a mean of 13 and a mode of 5.

7.04 The median and range

Statistical measures

Mean, median, mode

Ranges and averages

Data puzzles

Mode, median and mean

The **median** is the middle value of a set of data or the average of the 2 middle values.

The median

When the values of a data set are ordered, the **median** is:

- the middle value if there is an odd number of values
- the average of the 2 middle values if there is an even number of values.

Think: 'Median' sounds like 'medium', which is halfway between small and large.

A 'median' strip is the strip of grass or concrete down the *middle* of a highway.

- The median is a single value that represents all of the data.
- The median does not depend on all of the data values, so it is not affected by outliers (extremely high or low values).
- The median is often one of the data values.

The **range** is a measure of the spread of a data set, and is simply the difference between the highest and lowest values.

Foundation Standard Complex

The range

Range = highest value – lowest value

Example 6

The ages of the members of the Carrozza family are:

3 1 25 58 26 52 30 31 11

Find:

a the median **b** the range

Solution

a First, rewrite the ages in order.

1 3 11 25 (26) 30 31 52 58

4 values — Median — 4 values

Note that the value of the median is at the centre of the data.

There are 9 values, an odd number, so there is one middle value.

$\frac{1}{2} \times 9 = 4\frac{1}{2}$, so round up to 5 and count to the 5th value (circled above) to find the median.

Median = 26

b Range = 58 − 1 Highest value − lowest value

= 57

Example 7

The lap times (in seconds) of 6 cyclists were:

12.5 21.4 10.5 18.7 12.9 10.3

Find:

a the median **b** the range

Solution

a Rewrite the values in order:

10.3 10.5 (12.5) (12.9) 18.7 21.4

3 values — 2 middle values — 3 values

There are 6 values, an even number, so there are 2 middle values.

$\frac{1}{2} \times 6 = 3$, so count to the 3rd and 4th values (circled above) and find their average.

$$\text{Median} = \frac{12.5 + 12.9}{2}$$
$$= 12.7$$

12.7 is halfway between 12.5 and 12.9

b $\text{Range} = 21.4 - 10.3$ Highest value − lowest value
$= 11.1$

Shutterstock.com/Duncan Andison

EXERCISE 7.04 ANSWERS ON P. 565

The median and range U F PS R C

EXAMPLE 6

1 What is the median of this set of data?

10 3 5 9 4 3 10

Select the correct answer **A**, **B**, **C** or **D**.

A 3 **B** 5 **C** 7 **D** 10

2 What is the range of this set of data?

17.1 18.3 14.5 11.8 21.9 14.5

Select **A**, **B**, **C** or **D**.

A 16.4 **B** 15.8 **C** 14.5 **D** 10.1

EXAMPLE 7

3 For each set of data, find:

i the median **ii** the range.

a	11	9	7	4					
b	46	50	48	47	50	40			
c	12	13	11	14	10				
d	84	50	0	28	14	99	51	66	51
e	57	57	57	57	57	57	57		
f	64.2	99	99.2	71.2	93.5	61.8			

Foundation Standard Complex

4 Manish likes to gamble on horse races. His winnings or losses (in dollars) each day are as follows (losses are shown as negative values). R C

8	21	−16	−25	21	−25	16
11	−12	12	−25	17	−25	14

a Calculate the mean, correct to the nearest cent.

b Find the median and the mode.

c As Manish wins more often than he loses, should he continue to gamble on horse races? Give a reason for your answer.

5 11 houses were sold in Jacaranda Avenue. The selling prices are listed here. R C

$580 000	$485 000	$660 000	$758 000	$644 000	$1 670 000
$815 000	$686 000	$666 500	$720 800	$805 000	

a Find the mode.

b Find the median.

c Find the mean.

d Which measure of central tendency best describes this set of house prices? Give a reason for your answer.

6 2 ten-pin bowlers recorded the following scores over 10 games. PS R C

Aldo:	175	175	160	165	180
	145	150	177	177	177
Peter:	173	173	174	175	175
	176	176	177	179	179


Shutterstock.com/Mindscape studio

a Calculate the mean for each bowler.

b Calculate the range for each bowler.

c Who would you choose, and why, if you were asked to present an award for:

i the more talented bowler?

ii the more consistent bowler?

d A third player, Rob, recorded the following scores over 9 games.

Rob:	134	175	175	144	176
	150	187	166	180	

What score would Rob need to achieve in his tenth game to have the same mean score as Aldo?

7 The favourite party food for a group of 3-year-old children was recorded. R

popcorn	fruit	chocolate	fruit	popcorn
chocolate	chocolate	fruit	fruit	fruit

Which is the only statistical measure that can be found for this data?
Select **A**, **B**, **C** or **D**.

A mean **B** median **C** mode **D** range

8 Samir is the coach of the Panthers athletics team. He needs to pick his best runner for the 200 m race. He has recorded the times (in seconds) for the last few races of Daniel and Cristian. R C

Daniel	29.4	31	27.5	32.5	28	
Cristian	28.5	27	30.5	28.5	31.5	28.5

a Calculate the mean for each runner.

b Calculate the range for each runner.

c Who do you think Samir should choose as the runner for the next event? Give a reason for your answer.

9 A group of 8 friends counted the number of letters in their surnames. The results were:

4 6 7 5 4 6 9 ☐

where ☐ represents a missing value.

Find a possible value of ☐ if the median is 6 and the range is 5. PS R

10 **a** Find 8 values that have a median of 8 and a range of 8.

b Find 7 values from 3 to 9 inclusive that have a mode of 4 and a median of 6. PS R

Technology

Mean, mode, median and range

Students' marks

Finding the mean, median, mode and range

Statistical measures calculator

Statistical measures calculator

1 Enter the data below into a new spreadsheet, listing the daily maximum temperatures in Camden for one week in January.

	A	B	C	D	E
1	Day	Temperature (deg C)			
2	Saturday	31.5		Mean	
3	Sunday	32.2		Mode	
4	Monday	33		Median	
5	Tuesday	36		Maximum	
6	Wednesday	36.8		Minimum	
7	Thursday	35.9		Range	
8	Friday	26.7			

2 Enter each formula into the given cell to calculate each statistic.

Cell E2: **=average(B2:B8)**

Cell E3:**=mode(B2:B8)**

Cell E4:**=median(B2:B8)**

Cell E5:**=max(B2:B8)**

Cell E6: **=min(B2:B8)**

Cell E7: **=E5-E6**

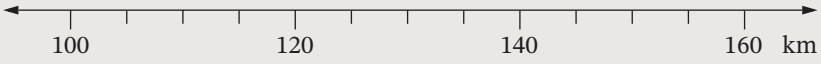
Mental skills 7: Maths without calculators ANSWERS ON P. 565

Reading linear scales

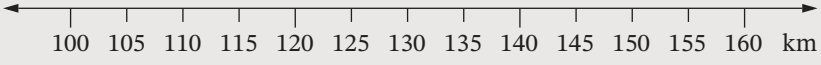
Understanding and reading the scale on a measuring instrument, on a number line or on the axis of a graph is an important mathematical skill.

1 Study each example.

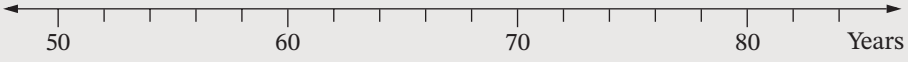
a Complete the missing values on this scale.



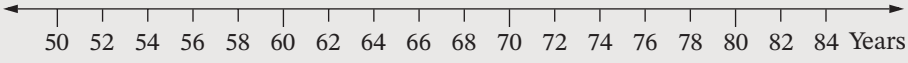
- First, choose 2 values on the scale, say 100 and 120.
- Count the number of intervals ('spaces') between the 2 values. There are 4 intervals between 100 and 120.
- To find the size of each interval, divide the difference between the 2 values by the number of intervals.
- Difference = 120 – 100 = 20 km
- Number of intervals = 4
- Size of an interval = 20 ÷ 4 = 5 km
- Use the calculated size of an interval to complete the missing values.



b Complete the values on this scale.

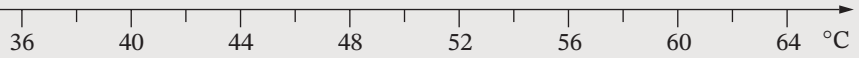


- Choose 50 and 60 on the scale.
- Number of intervals (between 50 and 60) = 5
- Difference (between 50 and 60) = 60 – 50 = 10 years
- Size of an interval = 10 ÷ 5 = 2 years.

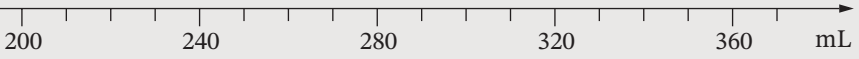


2 Now copy and complete the following scales.

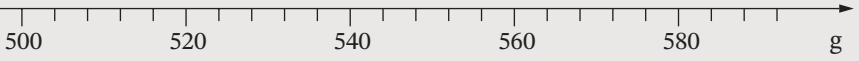
a



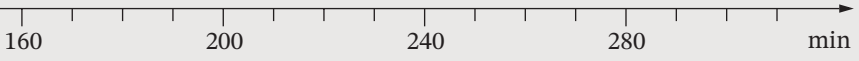
b



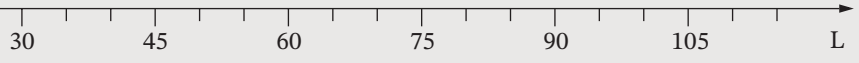
c



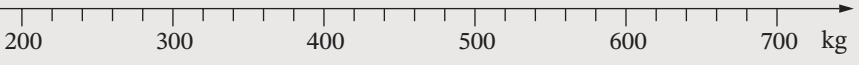
d



e



f



g

h

Investigation

Gym friends

7 friends joined a gym to lose weight.

Name	Sex	Age	Weight loss
Brenda	F	36	4 kg
Peter	M	34	28 kg
Mark	M	28	21 kg
Fiona	F	33	2 kg
Manjeet	F	30	6 kg
Mia	F	34	3 kg
Natisa	F	29	6 kg

Shutterstock.com/Monkey Business Images

In the following questions, round all mean answers to 2 decimal places.

1 What is the mean age?

2 Find the mean age of the males.

3 What is the median age of the group? Who has this age?

4 What is the median age of the males? Who has this age?

5 If the men were not counted, would the mean age or the median age be affected more?

6 Predict what would happen to the mean age if Imraan (aged 21) joined the group. Test your prediction.

7 What was the mean age 2 years ago? How does this compare to the group's mean age now?

8 Compare the mean weight loss with the median weight loss of the whole group. Who has the median weight loss?

9 Compare the mean weight loss and the median weight loss of the females only. Who has the median weight loss now?

10 When Diego joined the group, the mean weight loss of the group became 20 kg. How much weight did Diego lose?

Investigation

Finding the mean of a large set of data

How do we find the mean of a large amount of data? Is there a simpler way than adding all the values, then dividing the total by the number of values?

A group of students was surveyed about the number of hours they watched TV on Saturday:

6	4	0	3	1	2	2	3	5	4	3	3
4	1	0	6	6	3	4	3	1	5	4	3

We can calculate the mean of this set of data using 3 different methods.

Method 1: Adding all the values and dividing by the number of values

1 Find the sum of all the data values.

2 How many values are there in the data set?

3 Hence show that the mean is approximately 3.17 hours.

Method 2: Arranging the values in order first to simplify the addition

1 Sort all of the values in ascending order and group the equal values:

0 0 1 1 1 2 2 3 3 3 3 3 3 3 4 4 4 4 4 5 5 6 6 6

2 To find the sum of all the values, notice that we can multiply each value by its frequency and add those amounts together.

Evaluate $(0 \times 2) + (1 \times 3) + (2 \times 2) + (3 \times 7) + (4 \times 5) + (5 \times 2) + (6 \times 3)$.

3 Hence show that the mean is approximately 3.17 hours.

Method 3: Using a frequency table with an extra column

Score	Frequency	Frequency × Score
0	2	0
1	3	3
2	2	4
3		
4		
5		
6		
Total		

This is a more organised version of Method **2**.

1 Copy and complete the frequency table.

2 What is the total number of values?

3 What is the sum of all the values?

4 Hence show that the mean is approximately 3.17 hours.

7.05 Analysing frequency tables

WS The mean and *fx* tables

Analysing frequency tables

Statistics from a frequency table

Example 8

The number of minutes each train was late at a railway station in one afternoon were:

2 4 1 0 0 3 2 3
3 2 0 3 2 1 3

Copy and complete this frequency table for the data and use it to find:

a the range

b the mode

c the median

d the mean, correct to one decimal place.

Score, x	Frequency (*f*)	*fx*
0		
1		
2		
3		
4		
Total		

Solution

fx means $f \times x$, so in the fx column, multiply each score (x) by its frequency (f):

$3 \times 0 = 0$

$2 \times 1 = 2$

$4 \times 2 = 8$

Score, x	Frequency (f)	fx
0	3	0
1	2	2
2	4	8
3	5	15
4	1	4
Total	15	29

a Range $= 4 - 0$
$= 4$

Highest value − lowest value

b Mode $= 3$

The score with the highest frequency (5)

c There are 15 data values (an odd number), so the median (middle) is the **8th** value. According to the frequency table, the values in order are:

0, 0, 0 1, 1 2,2,(2,)2 3,3,3,3,3 4

Half of 15 is $7\frac{1}{2}$, so round up to 8 and count the 8th value.

The first 3 values are 0, the next 2 values are 1, the next 4 values are 2, and so on.

∴ The 8th value (circled) is 2.

Median $= 2$.

d Mean $\bar{x} = \frac{\text{sum of values}}{\text{number of values}}$.

The fx column multiplies each score by its frequency, so its total, 29, is the sum of all the data values. If we write all the values, we can see that the fx column groups and adds the values.

$\underbrace{0,0,0}_{\text{Sum } 0}\ \underbrace{1,1}_{2}\ \underbrace{2,2,2,2}_{8}\ \underbrace{3,3,3,3,3}_{15}\ \underbrace{4}_{4}$

The f column shows the frequency of each value so its total, 15, is the number of values.

$$\text{Mean} = \frac{\text{sum of } fx}{\text{sum of } f}$$
$$= \frac{29}{15}$$
$$= 1.9333\ldots$$
$$= 1.9$$

Statistics for data in a frequency table

- the **mean** is $\bar{x} = \frac{\text{sum of } fx}{\text{sum of } f}$
- the **mode** is the data value(s) with the highest frequency, f
- the **range** is the difference between the last and first values in the score (x) column

Example 9

The ages of children at a play centre are shown in the frequency table. Copy and complete the table and use it to find:

Score, x	Frequency, *f*	*fx*
5	3	
6	4	
7	5	
8	2	
9	2	
Total		

Shutterstock.com/SeventyFour

a the mode

b the mean

c the median

d the range.

Solution

Score, x	Frequency, *f*	*fx*
5	3	15
6	4	24
7	5	35
8	2	16
9	2	8
Total	16	108

a Mode = 7

Has the highest frequency, 5.

b Mean:

$$\bar{x} = \frac{\text{sum of } fx}{\text{sum of } f} = \frac{108}{16} = 6.75$$

c There are 16 values (an even number), so there are 2 middle values.

5 5 5 6 6 6 6 (7 7) 7 7 7 8 8 9 9

The 2 middle values

$$\text{Median} = \frac{7+7}{2} = 7$$

Half of 16 is 8, so count the 8th and 9th values (circled) and find the average of them.

d Range = 9 − 5 = 4

Note that the mean and median are close to each other and at the centre of the values.

9780170453141

The mean of data in a frequency table using a calculator's statistics mode

Data from a frequency table can be entered into a scientific or graphics calculator using the statistics mode. Follow the instructions in the table below to calculate the mean of the children's ages from the previous page using your calculator's statistics mode.

Operation	Casio scientific	Sharp scientific
Start statistics mode.	MODE STAT 1-VAR SHIFT MODE scroll down to STAT Frequency? ON	MODE STAT =
Clear the statistical memory.	SHIFT 1 Edit, Del-A	2nd F DEL
Enter data.	SHIFT 1 Data to get table 5 = 6 =, etc. to enter in x column 3 = 4 =, etc. to enter in FREQ column AC to leave table	5 2nd F STO 3 M+ 6 2nd F STO 4 M+, etc.
Calculate the mean. ($\bar{x} = 6.75$)	SHIFT 1 Var $\bar{x}$ =	RCL $\bar{x}$
Check the number of values. (n = 16)	SHIFT 1 Var n =	RCL n
Return to normal (COMP) mode.	MODE COMP	MODE 0

EXERCISE 7.05 ANSWERS ON P. 566

Analysing frequency tables U F

In this exercise, round all means to one decimal place where necessary.

1 This table shows how many times a sample of children visited a doctor over 3 months. Copy and complete the table, and use it to find:

a the number of people surveyed

b the median

c the mean

d the mode

e the range.

No. of visits, x	Frequency, f	fx
0	5	
1	6	
2	14	
3	3	
4	5	
8	1	
Total		

EXAMPLE 8, 9

Foundation Standard Complex

2 A sample of hospital patients were surveyed on the number of hours they slept the previous night:

6	7	9	8	10	11	9	10	8	9	7	6
11	9	8	8	8	10	10	9	7	8	9	10

a How many people were surveyed?

b What was the highest number of hours people slept?

c Use a frequency table with an *fx* column to find the mean.

d Find the range.

e Find the median.

3 The judge's scores for students' speeches (out of 10) for an English class are as shown.

7	7	8	9	7	5	4	8	9
9	8	10	8	7	5	5	4	1
8	8	9	7	7	4	8	3	1

a Complete a frequency table for this data.

b What is the mode?

c What is the mean?

d What is the median?

e What is the range?

4 A class of students was asked how many children there were in their families.

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

Complete a frequency table for this data, including an *fx* column, and use it to find:

a the range **b** the mode **c** the median **d** the mean

5 This frequency table shows the heights in metres of a sample of students.

Height (m)	Frequency
1.42	4
1.44	5
1.46	8
1.48	11
1.5	6
1.52	3

Find:

a the number of students in the sample

b the mean correct to 2 decimal places, using your calculator's statistics mode

c the range

d the median

e the mode

6 Carmelina recorded the number of books bought by the people entering her shop in one hour.

No. of books	Frequency
0	7
1	5
2	13
3	4
4	1
5	2
6	1


Shutterstock.com/SpeedKingz

What is the median number of books bought? Select the correct answer **A**, **B**, **C** or **D**.

A 2 **B** 3

C 13 **D** 4

7 30 people were asked how often they went to a cinema in the previous month.

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

Construct a frequency table and use it to find:

a the range **b** the mode

c the median **d** the mean.

8 A pigeon breeders' club has 25 members. This frequency table shows the number of pigeons each member has.

Score, x	Frequency, f
8	3
10	4
14	3
15	6
17	2
18	4
20	3


iStock.com/CaoChunhai

a What is the range?

b What is the median?

c Use your calculator's statistics mode to find the mean.

7.05

Technology

Daily temperatures

1 Enter the following data into a spreadsheet: a frequency table showing the maximum daily temperatures in Coffs Harbour in July.

	A	B	C
1	Temperatures (x)	Frequency (f)	*fx*
2	15	1	
3	16	3	
4	17	3	
5	18	7	
6	19	7	
7	20	2	
8	21	7	
9	22	0	
10	23	1	
11	Total:		
12	Mean:		
13	Range:		

2 How many temperatures are in this data set? In cell B11, enter an appropriate formula for the sum of the Frequency column.

3 To calculate the mean, first complete the *fx* column. In cell C2, enter the formula for multiplying the values in the A and B columns, then **Fill Down** to copy the formula into cells C3 to C10.

4 In cell C11, enter an appropriate formula for the sum of the *fx* column.

5 In cell B12, calculate the mean by entering the formula for dividing the sum of the *fx* column by the sum of the *f* column.

6 In cell B13, calculate the range by entering the formula for subtracting the lowest temperature from the highest temperature.

7.06 Dot plots and stem-and-leaf plots

Dot plots

Stem-and-leaf plots

What data?

Stem-and-leaf plots

A **dot plot** is a graph that uses dots to show the frequency of each data value. A dot plot shows:

- any gaps in the data
- any **clusters**, where values are grouped or bunched together
- any **outliers**: extremely high or low values that are very different from the rest of the data
- how the values are spread out.

An **outlier** 'lies outside' the other values in the data set. Outlier is pronounced 'out-ly-er'

Example 10

An online music store recorded the daily sales figure of a particular album while it was popular, and displayed the results in a dot plot.

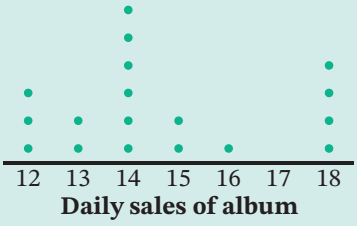
a On how many days were there exactly 13 sales?

b What is the range?

c Find the mode.

d Find the median.

e Calculate the mean, correct to one decimal place.



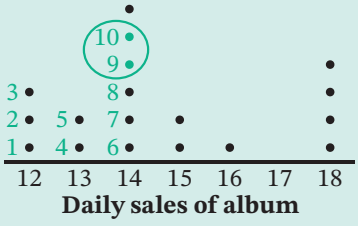
Solution

a 13 sales were made on 2 days — 2 dots above 13

b Range = 18 − 12 = 6 — Highest value − lowest value

c Mode = 14 — The value with the most dots.

d There are 18 values (18 dots). This is an even number, so there are 2 middle values (the 9th and 10th values). By counting the dots, or by crossing out pairs of dots at each end, we can see that the middle values (circled) are 14 and 14.



$$\text{Median} = \frac{14+14}{2} = 14$$

e $\text{Mean: } \bar{x} = \dfrac{\text{sum of values}}{\text{number of values}}$ $\left(\text{or } \dfrac{\text{sum of } fx}{\text{sum of } f}\right)$

Number of values = 18 — 18 dots

$$\bar{x} = \frac{3\times12+2\times13+6\times14+2\times15+16+4\times18}{18}$$
$$= 14.6666\ldots$$
$$= 14.7$$

Note that the mean and the median are close to each other and at the centre of the data.

Note: This answer can also be found using a frequency table and/or your calculator's statistics mode.

Stem-and-leaf plots

A **stem-and-leaf plot** (or **stemplot**) is like a sideways column graph, but one that lists the actual values on the horizontal columns. An **ordered** stem-and-leaf plot shows:

- all the data values, listed from smallest to largest
- any modes (the most common values)
- any clusters or outliers
- how the data are spread out.

Example 11

Back-to-back stem-and-leaf plots

The heights (in centimetres) of a sample group of Year 8 boys and girls were measured.

Girls:	148	155	153	137	155	137	146	165	157	168
Boys:	140	152	143	139	166	144	136	151	130	138

a Show this data on a back-to-back stem-and-leaf plot.

b Find the median for each set of data.

c Find the range for each set of data.

d Comment on the differences between the data for girls and boys.

Solution

a Heights of Year 8 students in centimetres

Girls		Boys
7 7	13	0 6 8 9
8 6	14	(0) (3) 4
7 5 (5) (3)	15	1 2
8 5	16	6

b Median: There are 10 values in each set.

The median is the average of the 5th and 6th values (circled in red on the plot).

Median: Girls $= \frac{153+155}{2} = 154$

Boys $= \frac{140+143}{2} = 141.5$

c Range: Girls = 168 – 137 = 31

Boys = 166 – 130 = 36

d The girls in the sample are generally taller (higher median), but the boys' heights are more spread out (higher range).

EXERCISE 7.06 ANSWERS ON P. 566

Dot plots and stem-and-leaf plots U F PS R C

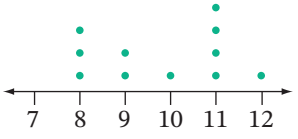
In this exercise, round all mean answers to one decimal place where necessary.

EXAMPLE 10

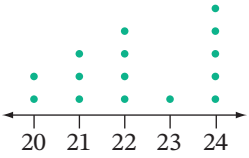
1 For each dot plot, find:

i the range **ii** the mode **iii** the median **iv** the mean.

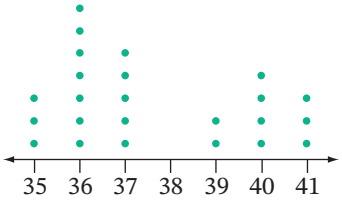
a



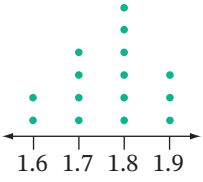
b



c



d



Foundation Standard Complex

9780170453141

2 The number of goals a football player kicked for each match last season were:

2	7	3	2	4	0	3	2	4	3	5
3	1	1	2	4	2	5	3	0	1	0

a Draw a dot plot to represent this data.

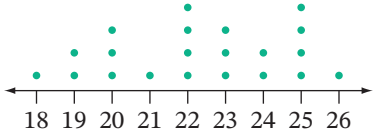
b How many matches were there last season?

c Which value is an outlier?

d Calculate the median.

e Calculate the mean.

3 This dot plot shows the number of students in each class at Northvale Primary School.



Number of students in class

a How many students are in the smallest class?

b Find the range.

c Find the mode.

d Calculate the mean.

e What percentage (to the nearest whole number) of classes have over 20 students?

f Which of the following values is the median? Select the correct answer **A**, **B**, **C**, or **D**.

A 11 **B** 22 **C** 22.5 **D** 23

4 The number of songs Udani downloaded each day was: R

1, 2, 0, 2, 3, 3, 3, 1, 1, 1, 6, 2, 3, 0

a Draw a dot plot for this data.

b How many days are represented by this dot plot?

c What is the outlier?

d Find the median.

e Calculate the mean.

f If this pattern continues in the following weeks, how many songs would she expect to download on average in a week?

5 The quiz marks out of 10 for 2 Year 8 classes are shown. R C

8 Hamper:	3	2	0	1	5	8	6	7	6	3	
	5	4	5	6	7	9	2	5	7		
8 Yen:	7	6	3	7	8	1	9	4	6	7	2
	7	2	8	10	9	9	5	7	8	9	10

a Draw a dot plot of the marks for each class.

b What is the mode of the marks for class 8 Yen?

c What is the median of the marks for class 8 Hamper?

d What is the range of the marks for class 8 Yen?

e Draw a frequency table of the data for class 8 Hamper and use it to calculate the mean.

f Which do you think is the 'better' class? Give a reason for your answer.

EXAMPLE 11

6 For each stem-and-leaf plot below, find:

i the range **ii** the mode **iii** the median **iv** the mean.

a

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

b

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

c

Stem	Leaf
10	0 1 1 2
11	7 8 8 8 9
12	3 6 6 7
13	1 1 1 1 1 8
14	0

d

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

7 A survey shows that the number of students served at the school canteen daily over a 3-week period is:

105 76 97 108 114 106 124 101
112 98 95 105 117 101 92

a Show this information on a stem-and-leaf plot.

b What is the outlier?

c Find the mode.

d What was the highest value?

e On what stems were the values clustered?

f On what percentage (to one decimal place) of days were more than 100 students served?

8 The heights in centimetres of the students in a Science class are:

156 154 158 167 164 163 155 166 159 158 137 163
171 160 164 169 157 161 160 165 145 175 167 167

a Make a stem-and-leaf plot for this data.

b On what stem are the data clustered?

c How many students were measured?

d How many students were shorter than 150 cm?

e Calculate the range.

f Find the median.

9 The times taken, in minutes, to drive between 2 towns along 2 different roads are shown on this back-to-back stem-and-leaf plot. PS R C

Fuller's Ridge Rd		Great Southern Drive
7 5 3 2	2	6 8
9 8 8 5 1 1	3	0 2 4 7 8
3 3 2 0 0	4	0 1 2 6 9 9
2	5	0 1 3 7
1 0	6	2
	7	1

a How many trips were made along the Great Southern Drive?

b How many trips made along Fuller's Ridge Road took over an hour?

c Which road is the faster one to take to drive between the 2 towns? Explain your answer.

d Which statement is true for the data sets above? Select **A**, **B**, **C** or **D**.

A The range is the same for both data sets

B There is a difference of 5 minutes in the median travel times

C The mean travel time for Fuller's Ridge Road is 36 minutes

D A travel time of 71 minutes along the Great Southern Drive is an outlier

10 This back-to-back stem-and-leaf plot shows the number of penalties that 2 football teams received for each match last season. PS R C

Wee Waa Reds		Narrabri Lions
9 8 4 4 2	0	7 9 9
8 6 5 5 2 2 2	1	2 3 3 6 8
3 1 1	2	5 6 8 9
0	3	0 0 1 3

a How many matches did these teams play in one season?

b How many times did the Narrabri Lions receive fewer than 20 penalties in a match?

c Find the highest number of penalties received by the Wee Waa Reds in a game.

d Find the mode for the Narrabri Lions.

e Which value was an outlier for the Wee Waa Reds?

f Which team generally had a higher number of penalties?

g Which team had a greater range in their number of penalties? Give a reason for your answer.

Investigation

Measures of average

1 Create 6 values in a data set where the mean equals the median.

2 Create 5 values in a data set where the median equals the mode.

3 Create 4 values where the mean, the median and mode all differ.

4 Create 5 values where the mean, median and mode are the same.

5 7 judges score a gymnast out of a possible 10. Their values are:

6 7 8 6 8 9 6

The scoring system requires that the lowest and highest scores are rejected and that the mean of the remaining 5 gives the final score. What is the final score?

Shutterstock.com/ITALO

6 In 4 games Dan scores 23, 11, 17 and 26 points. What score must Dan achieve in his next game so that he can average 20 on the 5 games? (Show the working or reasoning that you use.)

7 11 film students rate a movie out of a possible 10. Their ratings are:

7 9 9 4 6 4 8 9 7 8 5

The group cannot agree on a group rating for the movie. In one sentence, comment on each of these suggestions.

a Kim thinks that the rating should be 9 (the mode) because that was the rating given the most often.

b Vlado thinks that the rating should be 7 (the median) because it is not swayed by extremely high or extremely low ratings.

c Thanh thinks that the rating should be 6.9 (the mean) because it considers the ratings of all the students.

What overall rating would you give the film based on these opinions?

8 Find 3 numbers for which the median is 15, the mean is 13 and the range is 12.

9 Find 5 numbers for which the median is 3, the mean is 4, and the mode is 7.

Frequency histograms and polygons 7.07

Numerical data can be graphed on a **frequency histogram**, a special type of column graph, or a **frequency polygon**, a special type of line graph.

WS Statistical match-up
WS Analysing data
Highway elephants

Example 12

20 families were surveyed about how many phones they owned.

Graph this data as:

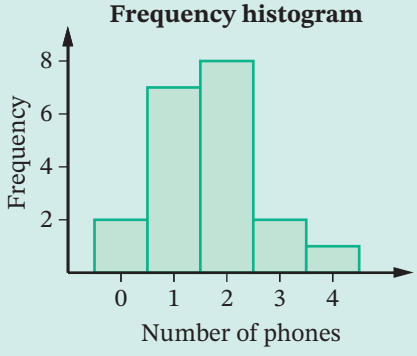
a a frequency histogram

b a frequency polygon.

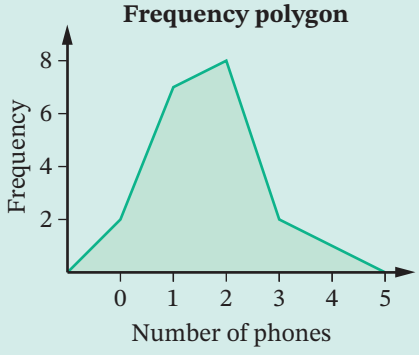
No. of phones	Frequency
0	2
1	7
2	8
3	2
4	1
Total	20

Solution

a



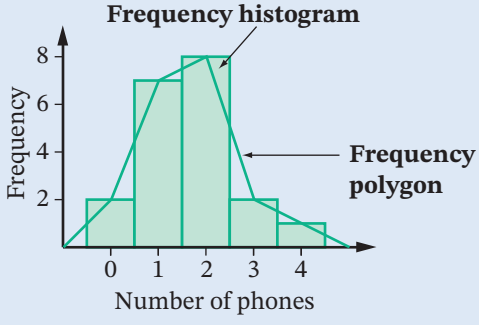
b



Note: A frequency polygon can be constructed from a frequency histogram by joining the midpoints of the tops of the columns.

Frequency histogram

- A column graph that looks like a row of office buildings
- Frequency is shown on the vertical axis
- The values are shown on the horizontal axis
- The columns stand together without gaps between them
- The columns are centred on the values
- There is a gap of half a column width on the left (between the frequency axis and the first column)



Frequency polygon

- A line graph that looks like a mountain
- Frequency is shown on the vertical axis
- The values are shown on the horizontal axis
- Begins and ends on the horizontal axis
- Called a 'polygon' because the graph and the horizontal axis together make a shape with straight sides

EXERCISE 7.07 ANSWERS ON P. 567

Frequency histograms and polygons U F PS R

EXAMPLE 12

1 The students in a class were surveyed about how many pets they owned.

No. of pets	Frequency
0	7
1	11
2	9
3	3

a Graph this data as a frequency polygon.

b How many students were surveyed?

c How many students had no pets?

d What percentage of students had 2 pets?

e Find the median.

f Calculate the mean, correct to one decimal place. Select the correct answer **A**, **B**, **C** and **D**.

A 1.0 **B** 1.3 **C** 1.5 **D** 3

Shutterstock.com/ Eric Isselée

2 Hadieya counted the number of phone calls she made each day for 30 days.

5	3	6	2	4	3	3	5	4	7
5	1	6	3	2	3	5	6	1	6
2	4	3	1	5	4	2	7	3	2

a Complete a frequency table for this data.

b Draw a combined frequency histogram and polygon of the data.

c On how many days did Hadieya make 3 calls?

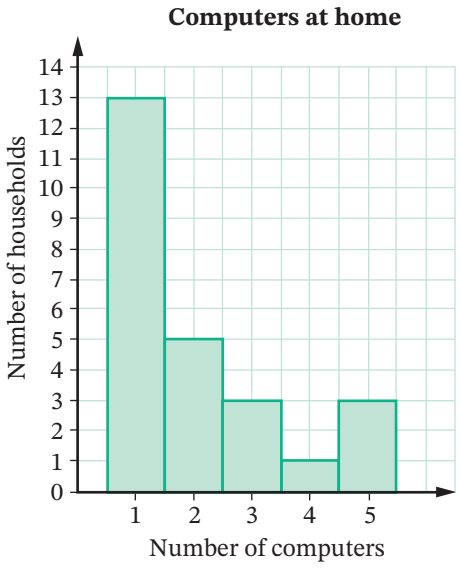
d How many phone calls were made over the 30 days?

e On how many days did Hadieya make fewer than 3 calls?

f What was the mode?

g What was the mean (correct to one decimal place)?

3 This histogram shows the number of computers found in a sample of homes.



a How many households were surveyed?

b Where are most of the values clustered?

c Find the mode.

d Find the range.

e Complete a frequency table, including an *fx* column, for the data and use it to calculate the mean.

f Find the median.

4 The numbers of goals scored by a soccer team in each match last season were:

2	0	1	2	3	4	3	1	2	0
2	1	4	3	2	5	1	2	0	3

a Complete a frequency table for this data.

b Draw a frequency histogram and frequency polygon of the data on the same axes.

c How many matches were played by the team?

d What was the highest number of goals scored?

e What was the mode?

f What was the mean?

g In how many matches were more than 2 goals scored?

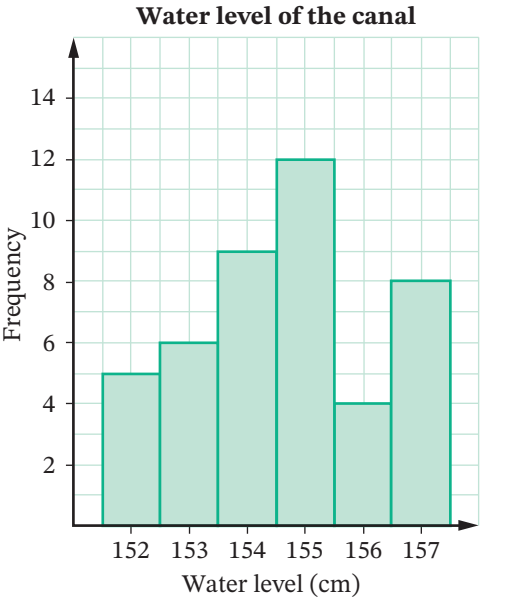
h In what percentage of matches were no goals scored?

Shutterstock.com/Fotokostic

5 The water level of a canal was recorded over 44 days and graphed on this histogram. PS R

a Use the histogram to copy and complete the frequency table.

Water level (cm)	Frequency
152	
153	
154	
155	
156	
157	
Total	



b On how many days was the water level 156 cm high?

c What is the range?

d What is the mode?

e What is the mean, correct to 2 decimal places?

f On how many days was the water level lower than 154 cm?

g What is the median?

h There needs to be at least 155 cm of water in the canal for a boat to travel on it safely. On how many days would the boat be able to use the canal?

7.08 Comparing data

WS Analysing data

Population polygons

Daily rainfall

WS Comparing word lengths

WS Investigating young drivers

This table describes the 4 statistical measures and when it is more appropriate to use each one.

Statistical measure	Features	When is it appropriate to use?
Mean		
$\bar{x} = \frac{\text{sum of values}}{\text{number of values}}$ For a frequency table: $\bar{x} = \frac{\text{sum of } fx}{\text{sum of } f}$	• For numerical data • Depends on all the values in the data set • Affected by outliers • Can be a decimal that is not one of the actual values	When the data set does not have many outliers.
Mode		
Most common value(s)	• For numerical and categorical data • There may be more than one mode, or no mode at all • Not affected by outliers	When the most common value or category is needed.

Foundation Standard Complex

9780170453141

Statistical measure	Features	When is it appropriate to use?
Median		
Middle value, or average of the 2 middle values, when values are arranged in order	• For numerical data • Can be one of the values • Not affected by outliers	When the data set has many outliers.
Range		
Highest value – lowest value	• For numerical data • Depends on the highest and lowest values only	When a measure of spread is needed.

Example 13

A group of 8 children was surveyed about the amount of pocket money (in dollars) they received each week.

20 32 32 40 18 32 18 48

a Find:

i the mean **ii** the mode

iii the median **iv** the range

b If the last student's pocket money increased from \$48 to \$75, describe how this would affect:

i the mean **ii** the mode

iii the median **iv** the range

c Which measure of location best represents the new set of data?

Solution

a i Mean: $\bar{x} = \frac{\text{sum of values}}{\text{number of values}}$

$= \frac{20+32+32+40+18+32+18+48}{8}$

$= \frac{240}{8}$

$= \$30$

ii Mode = \$32

iii Rewriting the 8 values in order:

18 18 20 (32) (32) 32 40 48

There are 2 middle values, 32 and 32 (circled above).

$\text{Median} = \frac{32+32}{2}$

$= \$32$

iv Range $= 48 - 18$

$= \$30$

b If the last student's pocket money increased to \$75, this value would be a high outlier compared to the others.

i The mean should increase.

Mean: $\bar{x} = \frac{267}{8}$

$= \$33.375$

ii The mode will not change because \$32 is still the most common value.

iii The median will not change because the order of the values has not been affected.

18 18 20 (32) (32) 32 40 75

The 2 middle values are still 32 and 32 (circled above).

iv The range will increase because the highest value is now even higher.

Range $= 75 - 18$

$= \$57$

c The median is the best measure of location because it is not affected by the outlier of \$75.

Note:

- The mode (\$32) is a good measure, but it only represents 3 of the values.
- The mean (\$33.375) is the least appropriate because it has been affected by the outlier and is now higher than most of the values.

EXERCISE 7.08 ANSWERS ON P. 567

Comparing data U F PS R C

In this exercise, round all mean answers to one decimal place where necessary.

EXAMPLE 13

1 The distances (in kilometres) that a salesperson travelled each day for 10 days were:

49 77 63 56 49 40 56 40 74 49

a Find:

i the range **ii** the mode **iii** the median **iv** the mean.

b Which measure of location best represents the data?

2 The ages of the mathematics teachers at Westvale Secondary College are:

49 32 37 32 25 41 39 R C

a Find:

i the range **ii** the mode **iii** the median **iv** the mean

b The 39-year-old teacher was replaced by a 22-year-old teacher. Without recalculating, describe how this will affect:

i the range **ii** the mode **iii** the median **iv** the mean

3 In a swimming competition, each swimmer was graded A, B, C, D or E. One squad achieved the results shown. R C

Grade	A	B	C	D	E
No. of swimmers	4	9	8	6	3

a How many swimmers were in the squad?

b What is the mode?

c Why isn't it possible to find the mean or median of these grades?

4 The annual incomes (× $1000) of 12 people are: R C

84 57 121 75 81 134 66 484 210 93 125 90

a Find (in dollars):

i the range **ii** the median **iii** the mean

b Why isn't the mean or mode an appropriate measure of central tendency for the incomes?

c Remove the outlier from the set of data and recalculate the mean and median. Are they closer together now?

5 The heights (in centimetres) of a team of basketball players are: R C

204 196 181 185 190 198 193

a Find:

i the mean **ii** the median **iii** the mode **iv** the range

b The 190 cm player is replaced by a 181 cm player. Without recalculating, describe how this will affect:

i the mean **ii** the median **iii** the mode **iv** the range

6 The cricket batting scores for Sunjeet and Chris over six innings are shown. PS R C

Sunjeet: 29, 35, 40, 26, 28, 39 Chris: 2, 2, 3, 2, 188, 5

a What are the median scores for each player? Select the correct answer **A**, **B**, **C**, or **D**.

A Sunjeet 32, Chris 2 **B** Sunjeet 40, Chris 2.5

C Sunjeet 33, Chris 2.5 **D** Sunjeet 32, Chris 2.5

b Find the mean score for each player.

c Who is the better player if:

i the median is used? **ii** the mean is used?

d Who do you think is the better player? Explain your answer.

7 Hardly Normal Discounts recorded the weekly number of sales of GPS navigators at 2 stores over a 20-week period. PS R C

Forbes: 44, 30, 54, 39, 63, 69, 75, 32, 80, 76, 90, 80, 55, 58, 71, 91, 36, 61, 65, 84

Cobar: 37, 36, 79, 77, 65, 47, 33, 49, 51, 60, 72, 49, 51, 73, 88, 31, 41, 97, 58, 30

a Display this information on a back-to-back stem-and-leaf plot.

b Calculate the mean, mode, median and range for each store.

c Which store performed better? Explain your answer.

d Which sales figure from the Cobar store is an outlier? Which of the measures of centre calculated in part **b** are most affected by this outlier?

8 A netball coach kept a record of the number of goals scored by Christina and Mary during the first half of the season to compare their performances. PS R C

Christina	12	23	25	19	10	35	20	32	19	25	4	27
Mary	13	30	20	24	17	25	17	34	20	29	19	15

a Calculate the mean, mode, median and range for each player.

b Who do you think is the more consistent player? Explain.

c Who is the higher scorer?

d Who do you think is the better player? Why?

e What is the outlier value?

f If the outlier is removed, does it change:

i who the more consistent player is? **ii** who the higher scorer is?

9 2 judges rated 8 contestants in a singing contest. R C

Delta	8	5	6	8	5	9	2	6
Guy	7	8	5	4	9	5	5	6

a Draw a dot plot for each judge's marks.

b Calculate the range and median for each judge.

c Describe differences in the way each judge allocates marks.

7.09 Sampling

Sample vs census

In statistics, the **population** means all of the items being studied, whether they be people, cars, businesses or houses. Statistical data is collected in a variety of ways, such as surveys, Internet polls, interviews and questionnaires. To collect information, statisticians usually survey a large representative group called a **sample**. For example, to investigate opinions about school uniform, we can survey a selection of students, teachers and parents. The bigger the sample, the more representative and accurate it will be.

To collect information about a whole population, **all** people or items must be surveyed. This is called taking a **census**. For example, to plan the locations of childcare centres, the government may need to count the number of babies in different areas.

Sample	Census
Surveys a selected representative group of people or items.	Surveys **all** people or items (in the population).
Only gives **approximate** information about the population. Helps us make predictions about the population.	Gives **exact** information about the population.
Simple and inexpensive to survey.	Complex and expensive to survey.
Can be done quickly.	Takes a lot of time to collect and process the information.

In Australia, a national census is conducted every 5 years to count the Australian population and obtain important information about every person. It takes place in a year ending in 1 or 6, such as 2021 and 2026.

Example 14

Should a sample or a census be used to investigate each fact?

a The most popular car colour.

b Whether energy-saving light globes really use less electricity.

c The number of retired people living on the Gold Coast.

d People's views on which nation will win the Rugby World Cup.

Solution

a Sample, because it is easier and faster than asking all car-owners in Australia.

b Sample, because it is easier and faster than testing every light globe.

c Census, because the exact number is required.

d Sample, because it is very difficult to ask all people

Random sampling and bias

Remember that in probability, a **random experiment** is a situation where every possible outcome has the same chance of occurring. When taking a sample, it is important that each person or item in the population has an equal chance of being chosen to be surveyed. This is called **random sampling**. To find a sample of students to survey about school uniform, we could place students' names in a box, mix them up, and then choose 100 names without looking. Alternatively, we could use a computer or calculator to generate random numbers to select students.

A sample that is not random is called a **biased sample**, and is not truly representative of the population. For example, if you survey the first 100 students to arrive at school today, or the first 100 students on the school roll, then not all students in the school have an equal chance of being selected. Your sample would not represent students who arrive at school later or whose names begin with letters at the end of the alphabet.

Example 15

A random sample of students is to be selected for their views on a new school uniform. Will each of the following sampling methods provide a sample that is representative or biased? If biased, explain why.

a Randomly selecting one student from each class.

b Selecting 20 students from your mathematics class.

c Selecting the students from the school's netball teams.

d Randomly selecting students from the school roll by rolling a die and counting down the roll.

Solution

a Representative. Every student in the school has an equal chance.

b Biased, because your mathematics class does not represent the other Year groups (or even the other Year 8 classes) at your school.

c Biased, because the netball teams do not represent the whole school and will have mainly female students.

d Representative. Every student in the school has an equal chance.

EXERCISE 7.09 ANSWERS ON P. 568

Sampling U F R C

EXAMPLE 14

1 Is a sample or a census more appropriate for each situation? R C

a Testing coffee for taste.

b Finding the number of migrants from New Zealand.

c Investigating people's views on road safety.

d Finding out people's opinions on whether Australia should become a republic.

e Testing a flu vaccine for effectiveness.

f Counting the number of Spanish-speaking children in Australian schools.

g Finding the population of Gundagai.

h Finding the number of people in Australia over 40, to plan for health care.

2 For each situation, state whether a sample or a census would be more appropriate and how the data could be collected. R C

a A car dealer wants to find the level of customer satisfaction with his business.

b To find the most popular TV program on Friday nights.

c To find the busiest airport in Australia.

d A principal wishes to find the least-liked subject in the school.

e The number of people who support whaling.

f The average height of students in your school.

g The amount of recyclable garbage each household produces in a week.

h The number of times the students in your class visit a restaurant in a month.

EXAMPLE 15

7.09

3 The parent association for Nelson Vale High School want to know what people think of a new design for the school uniform. Which of the following samples are NOT biased? Select the correct answer **A**, **B**, **C** or **D**. R

A Students who visit the school office during the day

B Parents as they are picking up their children after school

C Randomly-selected families from an alphabetical list of school families

D Families watching the school sporting teams

4 For each survey, describe the total population, give a suitable way of choosing a representative sample, write any problems that may be encountered and explain how bias could be minimised. R C

a The attitudes of people to smoking in public.

b Whether girls are better than boys at mathematics.

c The choice of music at a school disco.

d The political parties people will vote for at the next state election.

5 In each survey, state whether the sample is random. If not, explain why you think it is biased. R C

a To find participants for a survey of students' study habits, a teacher wrote a message in the daily notices asking for volunteers.

b To ensure that items being made in a factory were of good quality, 2 items coming off the assembly line were tested at a randomly selected time each hour.

c In a poll to find voters' preferences in the next election, every 100th person listed in the White pages phone database was called. If there was no answer, the next person listed was called, and so on.

d In a parent survey about school uniform, the school council posted letters to all parents of students at the school asking for their opinion on the issue.

e In a survey about the best song of the decade, listeners of a rock radio station were asked to ring and nominate a song.

6 The student council wants to survey students on their attitude to mobile phone usage at school. 15 students from each Year group were selected to take part in the survey. R C

a Do you believe the results of this survey would be reliable? Explain your answer.

b What factors should have been considered in selecting the sample?

c How would you have selected the students to be surveyed?

Technology

Random numbers

Random numbers can be used to select people or items for a sample. A calculator or spreadsheet can be set up to generate a random number from 1 to 100, for example, and that number can be matched to the name of a person or item from a list. For a scientific calculator:

Casio	Sharp
Enter the formula RanInt#(1,100) as shown: ALPHA . (1 SHIFT) 100) =	2nd F 7 (RANDOM), select 3 R-INT, + 1 =

This generates a random number from 1 to 100. Press = more times to generate more random numbers.

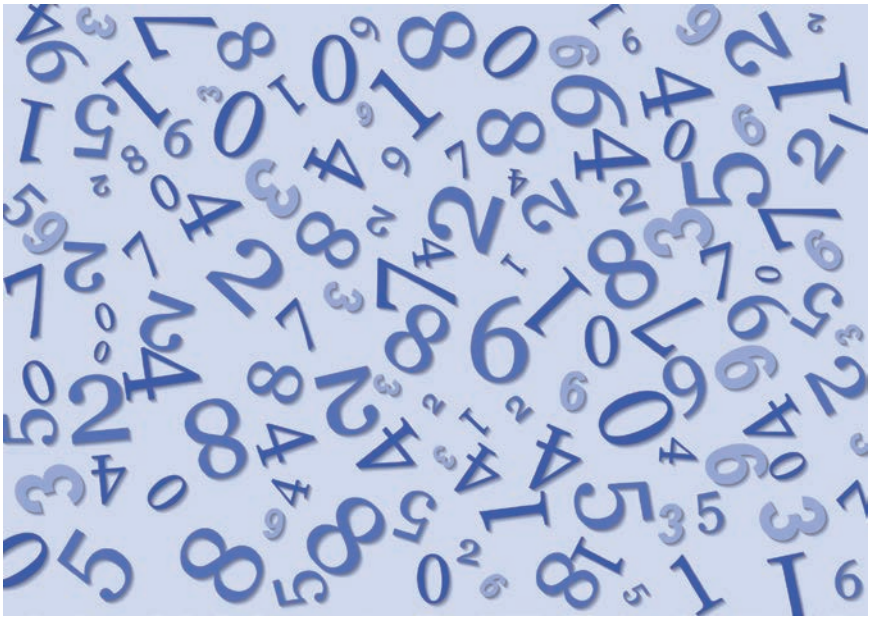
Alternatively, on a **spreadsheet,** type **=INT(RAND()*100+1)** into a cell and press the ENTER key, then the F9 key repeatedly for more random numbers.

1 Use your calculator or spreadsheet to generate 12 random numbers from 1 to 100.

2 Use your calculator or spreadsheet to generate 12 random numbers from 1 to the number of students in your class.

3 Use a class roll to number the students in your class from 1 onwards. Suppose you need to choose a random sample of 9 students from your class for a survey on homework. Use your calculator or spreadsheet to generate 9 random numbers to select the 9 students for your survey.

4 Choose 4 students from each Year 8 maths class to answer some questions. Devise several different methods to ensure that your selection is random. Prepare a short class talk about your methods.



iStock.com/Andril Zakharov

Investigation

Statistics in the media

We live in a world of 24-hour news: from TV, newspapers, websites, Facebook, Twitter, blogs, etc. To detect bias in news, we need to consider where the news item comes from and what types of samples the statistics are based on. Is the information supplied by a reporter, the police, a company executive, a government official or an opinionated blogger? Are the statistics based on a small sample, large sample, unrepresentative sample, or a phone or Internet poll of volunteers? Each may have a particular bias that may influence how the story is reported.

Find examples of news items or surveys reported in a newspaper, magazine or the Internet and investigate the following questions.

a Where did it come from?

b Who wrote the story?

c Does it show any bias?

d Can it be supported by other news providers?

e What type of sampling was used? Was it representative?

f How many people were questioned?

g When was the survey conducted?

Shutterstock.com/ra2 studio

Did you know?

The Australian Census

The first population counts in Australia were known as musters, where people assembled at a location to be counted. The first census when people were recorded at their dwellings was in NSW in 1828. The first simultaneous census across all states was in 1881 and the population was counted as 2.25 million.

Following Federation in 1901, the first national census was developed and undertaken by the Commonwealth in 1911. From 1961, the Census has been held every 5 years in the years ending in 1 and 6.

The current census asks questions relating to the family, where you live, income, religion, level of education and other aspects of modern life.

Newpix.com/Andy Tyndall

What significant change was introduced to the Census in 2016?
Find 5 questions that were asked in the most recent census.

7.10 Designing survey questions

WS Canteen survey

WS Student survey form

When designing a survey, it is important that the questions asked are clear, fair and not biased.

A good survey question:

- uses plain English
- is not vague and does not have a 'double meaning'
- is not too open-ended, and does not allow too many different answers
- asks for only one piece of information
- does not try to influence a particular answer unfairly (this is also called bias)
- protects a person's privacy and does not ask for personal information, such as phone numbers
- provides choices, such as check boxes or a scale of 1 to 5 to select from.

Example 16

Explain what is wrong with each survey question and suggest an alternative.

a 'How many phones are there in your home?'

b 'What is your e-mail address?'

c 'How long did you have to wait before someone answered your call?'

Solution

a This question is not clear. Does it include mobile phones? Do 2 or more phones using the same line count as one phone? A better question would be: 'How many fixed-line phone numbers does your home have?'

b This question is too personal. This should be an optional question, perhaps adding 'May we contact you for further information if required?'

c This question has too many possible answers. It would be better to provide options:

☐ less than 3 minutes ☐ 3 to 5 minutes ☐ over 5 minutes

EXERCISE 7.10 ANSWERS ON P. 568

Designing survey questions U F PS R C

1 Explain what is wrong with each survey question and suggest an alternative. R C

EXAMPLE 16

a 'Are you happy?'

b 'How did you enjoy your meal at our restaurant?'

c 'What is your religion?'

d 'Soldiers have died under our Australian flag, so do you think it should be changed?'

e 'How much do you earn?'

f 'How often do you eat takeaway food?'

g 'What brand of computer do you own?'

A Apple **B** HP **C** Toshiba

Alamy Stock Photo/ Jeffrey Isaac Greenberg 6+

2 Some Year 8 students designed this questionnaire to investigate the breakfast choices of students. They were particularly interested in whether there is a difference between what boys and girls have for breakfast. R C

BREAKFAST QUESTIONNAIRE

Please complete each blank or tick the box or boxes that apply to you.

Age : _____ Male ☐ Female ☐

1 What do you eat for breakfast?

Porridge	☐	Rice bubbles	☐	Eggs	☐
Toast	☐	Corn flakes	☐	Muesli	☐

2 What do you drink?

Orange juice	☐	Milk	☐	Tea	☐
Coffee	☐	Apple juice	☐		

3 Do you think you have a healthy breakfast? Yes ☐ No ☐

Give reasons for your answer:

. . . . . .

. . . . . .

In a small group, discuss these questions.

a Do you think there are enough questions on the questionnaire? If you think more questions need to be added, write them down.

b Do you think the wording of the questions can be improved? If yes, replace the questions.

c Do the questions take into account cultural differences? Is it possible to improve the questions to minimise this problem?

3 As a group activity, design a survey about:

- the types of cars owned by the families of students in the class, or
- the most popular types of food sold in the school canteen.

Determine the suitable wording of each question and provide check box choices where appropriate. PS C

Comparing samples and populations 7.11

Statisticians do not study just one sample. They generally take a number of samples from a population and calculate the statistics of each sample. The results of each sample are meant to estimate the statistics of the whole population. Also, the larger the sample sizes, the better the estimates of the population.

Example 17

The ages of 100 people working at a theme park are:

18	19	28	27	20	20	24	40	24	19
30	35	26	24	22	19	27	27	23	29
28	40	21	17	20	22	23	21	24	23
22	26	36	25	16	24	36	25	19	25
34	39	45	20	35	21	33	27	19	33
18	27	34	30	25	34	37	29	25	27
24	24	16	18	25	21	26	31	26	25
18	19	26	22	25	22	45	38	34	43
57	24	48	18	20	30	28	22	26	35
29	21	21	28	16	22	35	24	18	23

a Randomly select 5 samples of 12 ages from this population of employees and for each sample, calculate the mean to one decimal place, the median, the number of employees aged under 20 and the number aged between 30–39.

b From your sample statistics, estimate the population mean, median, the total number of employees aged under 20 and the total number aged between 30–39.

Solution

a Sample 1: 40, 19, 17, 26, 25, 27, 34, 18, 19, 43, 22, 23

Sample 2: 19, 19, 25, 36, 23, 24, 16, 18, 17, 21, 24, 18

Sample 3: 24, 22, 43, 25, 16, 34, 35, 25, 24, 22, 20, 40

Sample 4: 35, 24, 20, 35, 48, 26, 28, 18, 17, 25, 22, 22

Sample 5: 27, 22, 19, 25, 25, 25, 35, 34, 35, 39, 34, 24

Each sample of 12 ages is selected randomly from the above population

	Mean	Median	No. aged under 20	No. aged 30–39
Sample 1	26.1	24	4	1
Sample 2	21.7	20	6	1
Sample 3	27.5	24.5	1	2
Sample 4	26.7	24.5	2	3
Sample 5	28.7	26	1	5

The values are similar for different samples but not close, because each sample is small and does not represent the whole population accurately.

b We can estimate the population statistics by taking averages of the samples.

$$\text{Estimate of mean} = \frac{26.1 + 21.7 + 27.5 + 26.7 + 28.7}{5}$$
$$= 26.14$$

$$\text{Estimate of median} = \frac{24 + 20 + 24.5 + 24.5 + 26}{5}$$
$$= 23.8$$

$$\text{Average number of employees aged under 20 per sample} = \frac{4 + 6 + 1 + 2 + 1}{5}$$
$$= 2.8$$

$$\therefore \text{Estimated number of employees aged under 20 in population} = \frac{2.8}{12} \times 100$$
$$= 23.333\ldots$$
$$\approx 23$$

$$\text{Average number of employees aged between 30–39 per sample} = \frac{1 + 1 + 2 + 3 + 5}{5}$$
$$= 2.4$$

$$\therefore \text{Estimated number of employees aged between 30–39 in population} = \frac{2.4}{12} \times 100$$
$$= 20$$

iStock.com/andykazie

EXERCISE 7.11 ANSWERS ON P. 568

Comparing samples and populations U F R C

EXAMPLE 17

7.11

1 a Copy this table. R C

	Mean	Median	No. aged under 20	No. aged 30–39
Sample 1				
Sample 2				
Sample 3				
Sample 4				
Sample 5				

b As in Example **17**, randomly select 5 samples of 12 ages from the population of 100 theme park employees and for each sample complete the table above by calculating the mean to one decimal place, the median, the number of employees aged under 20 and the number aged between 30–39.

c Are the values similar or different across the samples? Why?

d From your sample statistics, estimate the population mean, median, the total number of employees aged under 20 and the total number aged between 30–39.

e Add a row of the table for 'Population' and calculate the population mean, median, the total number of employees aged under 20 and the total number aged between 30–39 using all 100 numbers.

f How do the statistics of each sample compare to the population statistics?

g How do your estimated statistics from part **d** compare to the population statistics?

2 a Copy this table. R C

Sample size	Mean	Median	% aged under 20	% aged 30–39
12				
24				
36				
48				
60				
Population				

b Select another random sample of 12 ages from the population given in Example **17** and calculate and record in the table correct to one decimal place (where appropriate) the mean, the median, the *percentage* of employees aged under 20 and the *percentage* aged between 30–39.

c Repeat for random samples of 24 ages, 36 ages, 48 ages and 60 ages, recording your results in the table.

d Are the values similar or different across the samples? Why?

e Copy your results from question **1 e** for the population of 100 ages into the last row of the table.

f How do your sample statistics compare to the population statistics?

g Do larger samples represent the population better? Are their statistics closer to the population statistics?

Power plus ANSWERS ON P. 568

1 A small class achieved these marks in an exam:

66 68 74 76 82 79

a Find the mean of these marks, correct to one decimal place.

b The teacher realised there was an error in the marking and added 3 to each mark. Find the mean of the new marks.

c What effect did the extra 3 marks have on the mean?

d What effect does adding or subtracting the same number for all of the data have on the mean?

2 Yelena sat 5 exams. Her average mark was 74%. What mark should Yelena obtain in her next exam if she wishes her average mark for the 6 exams to be 77%?

3 If the mean of a set of values is 20, must 20 be one of the values? Give an example to illustrate your answer.

4 5 items of data were collected, but the figures were lost. The mean of the data was known to be 8 and the median was 9.

a What was the total of the data?

b What could the data have been?

c If you are told that the range is 7, what could the set of data have been?

5 4 sisters work at the same bank. One earns $500 per week, another earns $800 per week. The mean weekly wage of the 4 sisters is $2000. Is this possible? How?

9780170453141

CHAPTER 7 REVIEW

Language of maths

Qz Statistics

bias	categorical	census	continuous
cluster	discrete	dot plot	frequency histogram
frequency polygon	frequency table	mean	measure of location/ central tendency
measure of spread	median	mode	numerical
outlier	population	random	range
sample	stem-and-leaf plot	survey	

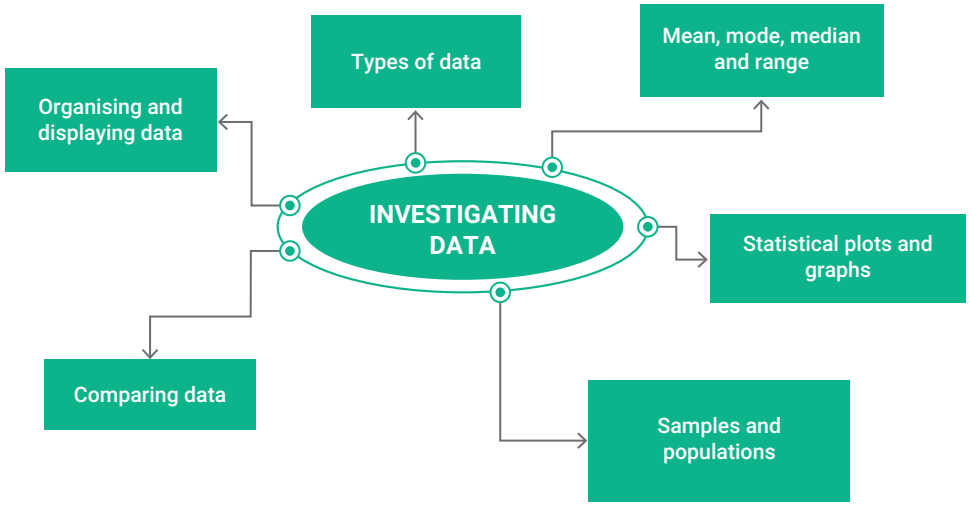
1. Name the **measures of location** or **measures of central tendency**.
2. What word means how often a value is listed in a set of data?
3. What type of data is described by words rather than numbers?
4. Is a **frequency polygon** a column graph or a line graph?
5. What is a survey of the whole population called?
6. What word describes a sample that misrepresents the population or a survey question that tries to lead you to answer in a particular way?

Topic summary

WS Mind map: Investigating data

- What did you find easy in this chapter? What don't you understand? Talk to your teacher.
- Give some examples of situations where statistics and measures of location are used. Explain, where possible, which measure of location is being used.

Print (or copy) and complete this mind map of the topic, adding detail to its branches and using pictures, symbols and colour where needed. Ask your teacher to check your work.



TEST YOURSELF 7

ANSWERS ON P. 568

In this exercise, round all mean values to 2 decimal places where necessary.

7.01

1 20 families were asked how many computers they owned.

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

a Complete a frequency table for this data.
b What was the most frequent value?
c How many people owned 4 computers?
d What percentage of people owned one computer?
e Draw a sector graph of this data.

7.02

2 Classify each type of data below as categorical (C), numerical discrete (ND) or numerical continuous (NC).

a life of a battery
b type of milk sold each day
c number of rooms in home
d volume of water in pools
e person's gender
f weight of bay

7.03

3 For each set of data, find:

i the mean **ii** the mode **iii** the median **iv** the range

7.04

a 6 12 11 12 10 6 6 10 6
b 8 4 1 1 4 1 3 6
c 20 21 23 24 25

7.05

4 The number of goals scored in each soccer match by the Southwest Strikers is shown in this frequency table.

Score, x	Frequency, x	fx
0	3	
1	7	
2	6	
3	4	
4	2	
5	1	
6	1	
Total		

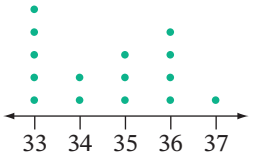
a Copy and complete the frequency table.
b How many matches did the Strikers play?
c Find the range.
d Find the mode.
e Find the median.
f Find the mean.

7.06

5 For each plot below, find:

i the range **ii** the median **iii** the mode **iv** the mean

a



b

Stem	Leaf
12	3 5 6
13	1 1
14	0 8 8 8 8
15	7 8

Foundation Standard Complex

9780170453141

6 The daily number of people visiting a house that was for sale was: 7.07

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

a Arrange this information into a frequency table.

b Draw a frequency histogram and frequency polygon.

c On how many days did the house receive more than 4 visitors?

d Find the mode.

e Find the median.

7 21 Year 8 students were asked to find out the ages of their mothers. The results were: 7.08

30	35	35	35	36	36	36	36	36	37	37
37	37	38	39	39	41	43	43	45	45	

a Find the median.

b Find the mode.

c Find the mean.

d Which measure of location is the most appropriate for representing this data?

e Predict how many 39-year-old mothers there would be in a Year 8 group of 120 students.

f Suppose one of the ages is incorrect: one of the 39s should actually be 40. Without recalculating, describe how this will affect:

i the mean **ii** the median **iii** the mode **iv** the range

8 Determine whether a sample or a census would be more appropriate for each survey below. 7.09

a Electing a Year 8 representative for student council.

b Investigating audience reaction to a new film.

c Testing the safety of a new car.

d Finding how many people live in Australia.

9 A school of 1200 students wants to survey a sample of students to determine how often they use mobile phones. 7.09

a State a suitable way of choosing a representative sample of 50 students.

b Write one problem that may be encountered in conducting the survey.

c Explain how bias could be minimised.

10 Write 4 questions that could be used in the survey described in question **9**. 7.10

11 Explain what is wrong with each survey question and suggest an alternative. 7.10

a 'How big is your house?'

b 'Is it time that Australia grew up and became a republic?'

Foundation Standard Complex