

Online resources



- Auto-marked chapter pre-test
- Video demonstrations of all worked examples
- Interactive widgets
- Interactive walkthroughs
- Downloadable HOTSheets
- Access to all HOTmaths Australian Curriculum courses
- Access to the HOTmaths games library

10 Probability and single variable data analysis

What you will learn

- 10A Probability review **REVISION**
- 10B Venn diagrams and two-way tables
- 10C Using arrays for two-step experiments
- 10D Using tree diagrams
- 10E Using relative frequencies to estimate probabilities
- 10F Mean, median, mode and range **REVISION**
- 10G Interpreting data from tables and graphs
- 10H Stem-and-leaf plots
- 10I Grouping data into classes **FRINGE**

NSW syllabus

**STRAND: STATISTICS
AND PROBABILITY**
**SUBSTRANDS:
PROBABILITY, SINGLE
VARIABLE DATA
ANALYSIS**

Outcomes

A student represents probabilities of simple and compound events.

(MA4–21SP)

A student calculates relative frequencies to estimate probabilities of simple and compound events.

(MA5.1–13SP)

A student describes and calculates probabilities in multi-step chance experiments.

(MA5.2–17SP)

A student analyses single sets of data using measures of location, and range.

(MA4–20SP)

A student uses statistical displays to compare sets of data, and evaluates statistical claims made in the media.

(MA5.1–12SP)

A student uses quartiles and box plots to compare sets of data, and evaluates sources of data.

(MA5.2–15SP)

A student collects, represents and interprets single sets of data, using appropriate statistical displays.

(MA4–19SP)

© NESA, 2012

Once in 100 years

Water from burst pipes, storms or floods can cause extensive damage to a person's home and contents. Different insurers offer different types of cover and differing premiums based on the calculated risk of damage. The higher the risk, the higher the premiums, with some insurers offering cover for flood damage as an optional extra, while others do not offer cover at all.

Since the Brisbane floods of 2011, Australian regulators have introduced a standard definition of flood. All insurers have been made to adopt this definition into their policies since 19 June 2014. Flood damage, it should be noted, is different to storm damage or rainwater damage.

The Brisbane floods of 2011 have resulted in a 45% increase in some people's premiums. In New South Wales it is predicted that 13% of properties will suffer flood damage at some time, with 2% being of high risk.

Many people talk about the 'once in a hundred years' chance of a flood; this is misleading as it simply means that there is a 1% chance of this level of flood in any given year. Although unlikely, it is possible for such a flood to occur more than once in a 100 years. Brisbane has experienced severe flooding in 1893, 1974 and 2011.

1 Write the following as decimals.

a $\frac{1}{10}$

b $\frac{2}{8}$

c 30%

d 85%

e 23.7%

2 Express the following in simplest form.

a $\frac{4}{8}$

b $\frac{7}{21}$

c $\frac{20}{30}$

d $\frac{100}{100}$

e $\frac{0}{4}$

f $\frac{12}{144}$

g $\frac{36}{72}$

h $\frac{36}{58}$

i $\frac{72}{108}$

j $\frac{2}{7}$

3 A six-sided die is rolled.

a List the possible outcomes.

b How many of the outcomes are:

i even?

ii less than 3?

iii less than or equal to 3?

iv at least 2?

v not a 6?

vi not odd?

4 One number is selected from the group of the first 10 positive integers $\{1, 2, \dots, 10\}$. How many of the numbers are:

a odd?

b less than 8?

c greater than or equal to 5?

d no more than 7?

e prime?

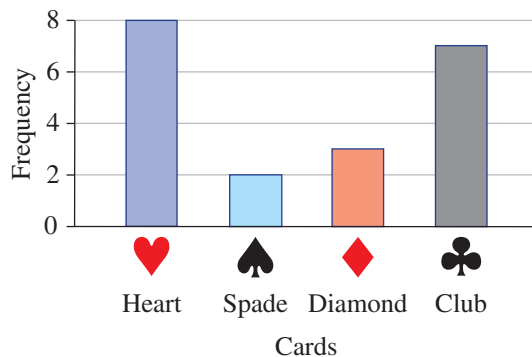
f not prime?

5 Several cards were randomly selected from a pack of playing cards. The suit of each card was noted, the card was replaced and the pack was shuffled. The frequency of each suit is shown in the column graph.

a How many times was a heart selected?

b How many times was a card selected in total?

c In what fraction of the trials was a diamond selected?



6 Consider this simple data set: 1, 2, 3, 5, 5, 7, 8, 10, 13.

a Find the mean (by finding the sum and dividing by 9).

b Find the median (the middle value).

c Find the mode (the most common value).

d Find the range (the difference between the highest and lowest value).

e If one of the numbers is randomly chosen from this data set, find the probability that it is a:

i 5

ii number that is not 5

iii number that is no more than 5

10A Probability review

REVISION



The mathematics used to calculate chance is called probability. We use it to compare the number of favourable outcomes to the total number of outcomes. This shows us how likely it is that the favourable event will occur. The probability of an event occurring is a number between 0 and 1. An impossible event has a probability of 0. An event that is certain to occur has a probability of 1.



We can use probability to find the likelihood of rolling a particular total score with two dice.

Stage

5.3#

5.3

5.3\$

5.2

5.2◇

5.1

4

Let's start: Events and probabilities

As a class group, write down and discuss at least three events that have the following chance of occurring.

- no chance
- very low chance
- medium to low chance
- even (50-50) chance
- medium to high chance
- very high chance
- certainty

For each decimal marked on the number line below, choose the probability description from the list above that best matches it.



Probability terminology

Terminology	Example	Definition
chance experiment	rolling a fair 6-sided die	A chance experiment is an activity which may produce a variety of different results that occur randomly. The example given is a single-step experiment.
trials	rolling a die 50 times	When an experiment is performed one or more times, each occurrence is called a trial. The example given indicates 50 trials of a single-step experiment.
outcome	rolling a 5	An outcome is one of the possible results of a chance experiment.
equally likely outcomes	rolling a 5 rolling a 6	Equally likely outcomes are two or more results that have the same chance of occurring.
sample space	{1, 2, 3, 4, 5, 6}	The sample space is the set of all possible outcomes of an experiment. It is usually written inside braces, as shown in the example.
event	rolling a 2 rolling an even number	An event is either one outcome or a collection of outcomes. It is a subset of the sample space.

Key ideas

Terminology	Example	Definition
compound event	rolling an even number (e.g. 2 or 4 or 6)	A compound event is a collection of two or more outcomes from the sample space of a chance experiment.
mutually exclusive events	rolling a 5 rolling an even number	Two or more events are mutually exclusive if they share no outcomes.
non-mutually exclusive events	rolling a 5 rolling an odd number	Events are non-mutually exclusive if they share one or more outcomes. In the example given, the outcome 5 is shared.
complementary events	rolling a 2 or 3 rolling a 1, 4, 5, or 6	If <i>all</i> the outcomes in the sample space are divided into two events, they are complementary events.
complement	Rolling 2, 3, 4 or 5 is an event. Rolling a 1 or 6 is the complement.	If an experiment was performed and an event did not occur, then the complement definitely occurred.
favourable outcome(s)	In some games you must roll a 6 before you can start moving your pieces.	Outcomes are favourable when they are part of some desired event.
theoretical probability or likelihood or chance	The probability of rolling an odd number is written as: $P(\text{odd}) = \frac{3}{6} = \frac{1}{2} = 0.5 = 50\%$ Probabilities can be expressed as fractions, decimals and percentages.	Theoretical probability is the actual chance or likelihood that an event will occur when an experiment takes place. $P(\text{event}) = \frac{\text{number of favourable outcomes}}{\text{total number of outcomes}}$ Probabilities range from 0 to 1 or 0% to 100%.
experimental probability	A die was rolled 600 times and showed a 5 on 99 occasions. The experimental probability of rolling a 5 on this die is: $P(5) = \frac{99}{600} = 0.165 = 16.5\%$	Sometimes it is difficult or impossible to calculate a theoretical probability, so an estimate can be found using a large number of trials. This is called the experimental probability. If the number of trials is large, then the experimental probability should be very close to the theoretical.
certain likely even chance unlikely impossible	rolling a number below 7 rolling a number below 6 rolling a 1, 2 or 3 rolling a 2 rolling a 7	The probability is 100% or 1. The probability is 50% or 0.5 or $\frac{1}{2}$. The probability is 0% or 0.
the sum of all probabilities in an experiment	$P(1) = \frac{1}{6} \quad P(2) = \frac{1}{6} \quad P(3) = \frac{1}{6}$ $P(4) = \frac{1}{6} \quad P(5) = \frac{1}{6} \quad P(6) = \frac{1}{6}$ $\frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{6}{6} = 1$	The sum of the probabilities of all the outcomes of a chance experiment is 1.
the sum of the probabilities of an event and its complement	$P(\text{rolling 1 or 6}) = \frac{2}{6}$ $P(\text{rolling 2, 3, 4 or 5}) = \frac{4}{6}$ $\frac{2}{6} + \frac{4}{6} = \frac{6}{6} = 1$	The sum of the probabilities of an event and its complement is 1. $P(\text{event}) + P(\text{complementary event}) = 1$

Exercise 10A REVISION

UNDERSTANDING AND FLUENCY

1-9

4-9

1 Use the words in the table below to complete each statement.

even	impossible, certain	probability
complement	sample space	event

- a The _____ is the list of all possible outcomes in an experiment.
- b An _____ is some of the outcomes from an experiment.
- c P stands for _____.
- d An _____ event has a probability of 0; an event that is _____ to occur has a probability of 1.
- e An event with probability of 0.5 has an _____ or 50-50 chance of occurring.
- f The _____ of an event, A, is the event where A does not occur.

2 Jim believes that there is a 1 in 4 chance that the flower on his prized rose will bloom tomorrow.

a Write the chance '1 in 4' as a:

- i fraction
- ii decimal
- iii percentage

'1 in 4' means $\frac{1}{4}$.



b Draw a number line from 0 to 1 and mark the level of chance described by Jim.

3 Copy and complete this table.

Percentage \div 100 = decimal



	Percentage	Decimal	Fraction	Number line
a	50%	0.5	$\frac{1}{2}$	
b	25%			
c			$\frac{3}{4}$	
d				
e		0.6		
f			$\frac{17}{20}$	

- 4 Ten people make the following estimates of the chance that they will get a salary bonus this year.

$$0.7, \frac{2}{5}, 0.9, \frac{1}{3}, 2 \text{ in } 3, \frac{3}{7}, 1 \text{ in } 4, 0.28, \frac{2}{9}, 0.15$$

Order their estimates from lowest to highest.

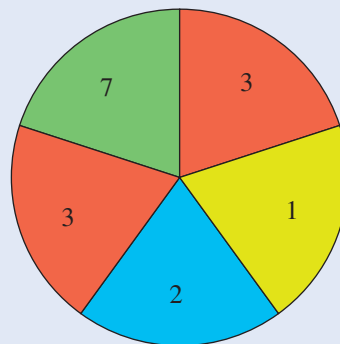
First write each estimate as a decimal.



Example 1 Finding probabilities of events

This spinner has five equally divided sections.

- List the sample space using the given numbers.
- Find $P(3)$.
- Find $P(\text{not a } 3)$.
- Find $P(\text{a } 3 \text{ or a } 7)$.
- Find $P(\text{a number that is at least a } 3)$.



SOLUTION

a $\{1, 2, 3, 3, 7\}$

b $P(3) = \frac{2}{5}$ or 0.4

c $P(\text{not a } 3) = 1 - P(3)$
 $= 1 - \frac{2}{5}$ or $1 - 0.4$
 $= \frac{3}{5}$ or 0.6

d $P(\text{a } 3 \text{ or a } 7) = \frac{2}{5} + \frac{1}{5}$
 $= \frac{3}{5}$

e $P(\text{at least a } 3) = \frac{3}{5}$

EXPLANATION

Use braces, $\{ \}$, and list all the possible outcomes in any order. Include 3 twice to represent its increased chance.

$$P(3) = \frac{\text{number of sections labelled } 3}{\text{number of equal sections}}$$

'Not a 3' is the complementary event of obtaining a 3.

Alternatively, count the number of sectors that are not 3.

There are two 3s and one 7 in the five sections.

Three of the sections have the numbers 3 or 7, which are at least 3; i.e. 3 or more.

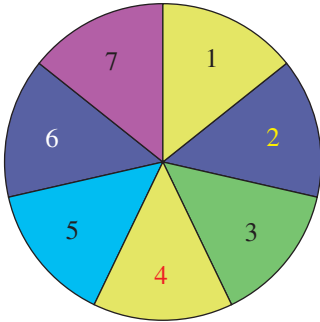
5 The spinners below have equally divided sections. Complete the following for each spinner. Write all probabilities as simplified fractions.

- i List the sample space using the given numbers.
- ii Find $P(2)$.
- iii Find $P(\text{not a } 2)$.
- iv Find $P(\text{a } 2 \text{ or a } 3)$.
- v Find $P(\text{a number which is at least a } 2)$.

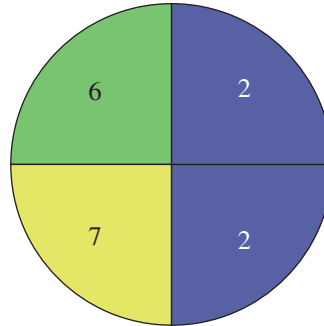
$P(\text{event})$ means 'find the probability of the event'.
'At least 2' means '2 or more'.



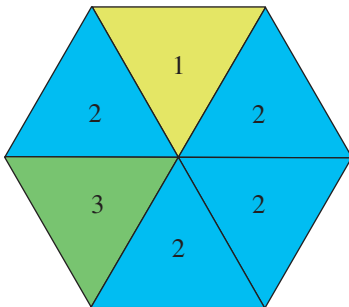
a



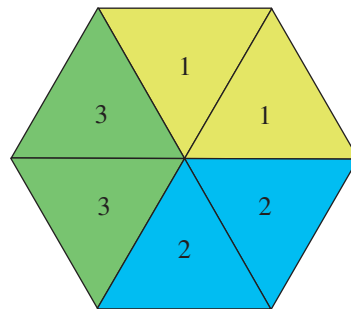
b



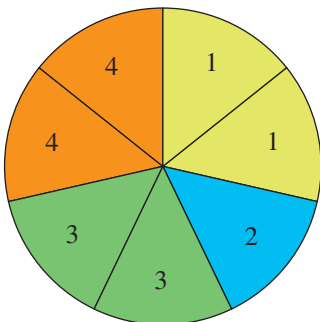
c



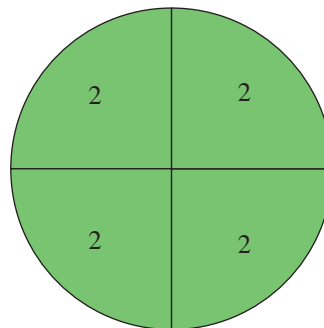
d



e



f



6 Find the probability of obtaining a blue ball if a ball is selected at random from a box that contains:

- a 4 blue balls and 4 red balls
- b 3 blue balls and 5 red balls
- c 1 blue ball, 3 red balls and 2 white balls
- d 8 blue balls, 15 black balls and 9 green balls
- e 15 blue balls only
- f 5 yellow balls and 2 green balls

$$P(\text{blue}) = \frac{\text{number of blue balls}}{\text{total number of balls}}$$



PROBLEM SOLVING AND REASONING

10, 11

10, 12, 13

10 The school captain is to be chosen at random from four candidates. Two are girls (Hayley and Alisa) and two are boys (Rocco and Stuart).

- a List the sample space.
 b Find the probability that the school captain will be:
 i Hayley
 ii male
 iii neither Stuart nor Alisa

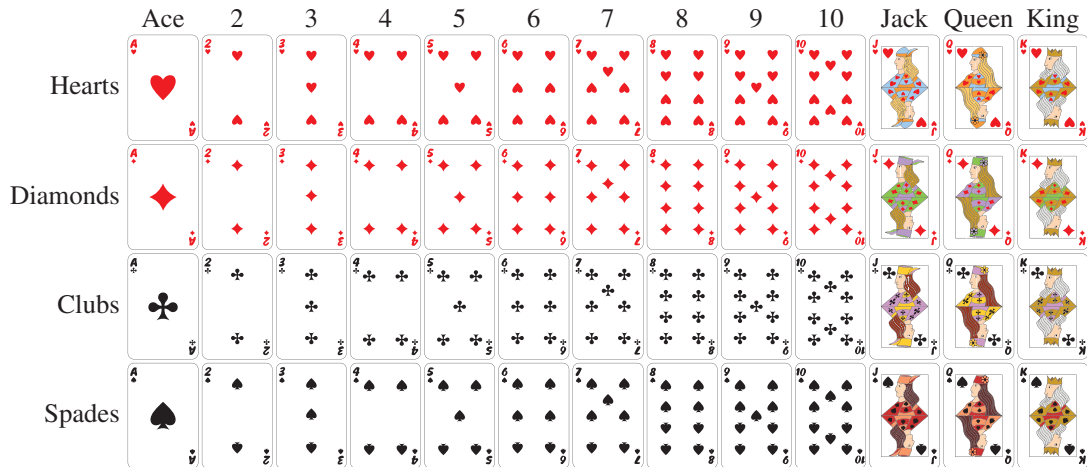
The sample space is a list of the possible students.



11 A card is drawn at random from a pack of 52 playing cards. Find the probability that the selected card will be:

- a the queen of diamonds
 b an ace
 c a red king
 d a red card
 e a jack or a queen
 f any card except a 2
 g any card except a jack or a black queen
 h not a black ace

A deck of 52 cards is shown below.



12 A six-sided die is rolled and the uppermost face is observed and recorded. Find the following probabilities.

- a $P(6)$
 b $P(3)$
 c $P(\text{not } 3)$
 d $P(1 \text{ or } 2)$
 e $P(\text{number less than } 5)$
 f $P(\text{even number or odd number})$
 g $P(\text{square number})$
 h $P(\text{not a prime number})$
 i $P(\text{number greater than } 1)$

'Less than 5' means that 5 is not included.
 1 is not prime.
 Square numbers are $1^2 = 1$, $2^2 = 4$,
 $3^2 = 9$ etc.



- 13** A letter is chosen at random from the word PROBABILITY. Find the probability that the letter will be:
- a** a B
 - b** not a B
 - c** a vowel or a consonant
 - d** not a vowel
 - e** a consonant other than B
 - f** one of the first five letters in the alphabet
 - g** a letter from the word RABBIT
 - h** a letter that is not in the word RABBIT

ENRICHMENT

–

14

Faulty CD player

- 14** A CD contains eight tracks. The time length for each track is shown in the table below.

The CD is placed in a faulty CD player, which begins playing randomly at an unknown place somewhere on the CD, not necessarily at the beginning of a track.

- a** Find the total number of minutes of music available on the CD.
- b** Find the probability that the CD player will begin playing somewhere on track 1.
- c** Find the probability that the CD player will begin somewhere on:
 - i** track 2
 - ii** track 3
 - iii** a track that is 4 minutes long
 - iv** track 4
 - v** track 7 or 8
 - vi** a track that is not 4 minutes long

Track	Time (minutes)
1	3
2	4
3	4
4	5
5	4
6	3
7	4
8	4



10B Venn diagrams and two-way tables



When the results of an experiment involve overlapping categories, it can be very helpful to organise the information into a Venn diagram or two-way table. Probabilities can easily be calculated from these types of diagrams.

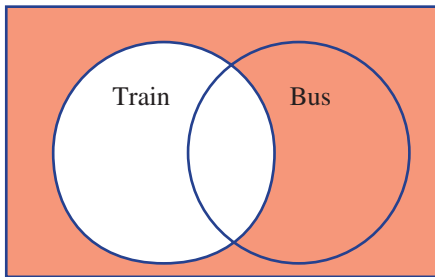
Let's start: Solving puzzles

Work with a partner to find the answer to each puzzle. For puzzles 1–4, draw some Venn diagrams, like the ones shown, to help you. For puzzle 5, use a two-way table.

Stage

5.3#
5.3
5.3\$
5.2
5.2∅
5.1
4

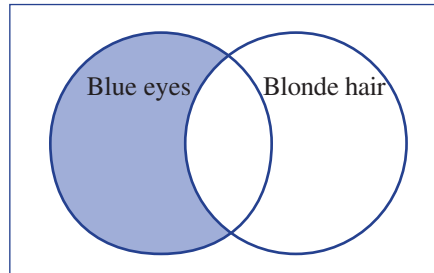
Puzzle 1



12 students travel to school by bus only and 10 students don't travel by either bus or train.

- How many students don't travel by train?

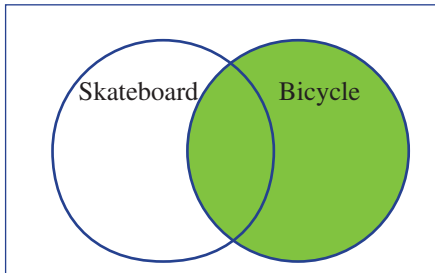
Puzzle 2



14 students in total have blue eyes, and of these students 5 have both blue eyes and blonde hair.

- How many have blue eyes but not blonde hair?

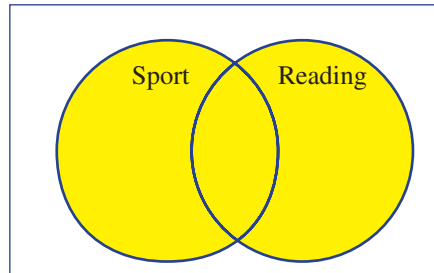
Puzzle 3



11 students own a bicycle only and 4 students own both a skateboard and a bicycle.

- How many students in total own a bicycle?

Puzzle 4



12 students in total like sport, 14 students in total like reading, and 9 students like both sport and reading.

- How many students altogether like either sport or reading or both?

Puzzle 5

A survey of 40 students found that a total of 22 play basketball, 9 play both volleyball and basketball and 6 do not play either basketball or volleyball.

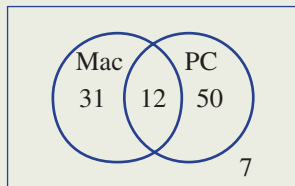
- How many basketball players don't play volleyball?
- How many students in total don't play volleyball?
- How many students in total do play volleyball?

	Basketball	Not basketball	Total
Volleyball			
Not volleyball			
Total			

■ **Venn diagrams** and **two-way tables** help to organise outcomes into different categories.

This example shows the types of computers owned by 100 people.

Venn diagram



Two-way table

	Mac	No Mac	Total
PC	12	50	62
No PC	31	7	38
Total	43	57	100

These diagrams show, for example, that:

- 12 people own both a Mac and a PC
- 62 people own a PC
- 57 people do not own a Mac
- $P(\text{Mac}) = \frac{43}{100}$
- $P(\text{only Mac}) = \frac{31}{100}$
- $P(\text{Mac or PC}) = \frac{93}{100}$
- $P(\text{Mac and PC}) = \frac{12}{100} = \frac{3}{25}$

Venn diagram A diagram using overlapping circles to organise and show two or more sets of outcomes

Two-way table A table used to organise, display and compare two sets of outcomes



Exercise 10B

UNDERSTANDING AND FLUENCY

1–8

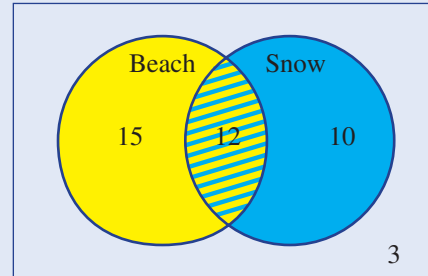
4–9



Example 3 Understanding a Venn diagram

This Venn diagram shows the number of people who enjoy holidays at the beach and at the snow.

- a** How many people are represented in this Venn diagram?
- b** How many people enjoy holidays:
- only at the beach?
 - at the beach (in total)?
 - both at the beach and at the snow?
 - neither at the beach nor at the snow?
- c** How many people do not enjoy holidays at the:
- beach?
 - snow?



SOLUTION

- a** 40 in total
- b i** 15 enjoy the beach only
- ii** 27 enjoy the beach
- iii** 12 enjoy the beach and the snow
- iv** 3 don't enjoy either
- c i** 13 don't enjoy the beach
- ii** 18 don't enjoy the snow

EXPLANATION

Add all the numbers in the Venn diagram.

$$15 + 12 + 10 + 3 = 40$$

There are 15 people in the beach circle who do not appear in the overlapping section.

Add the numbers in the beach circle.

$$15 + 12 = 27$$

12 people are in the overlapping section of the beach and snow circles.

There are 3 people outside the circles.

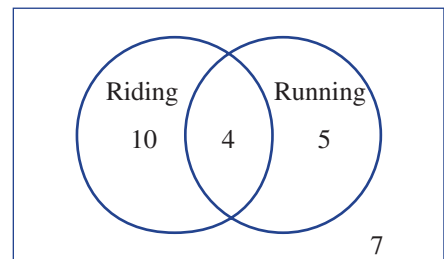
Add all numbers outside the beach circle.

$$10 + 3 = 13$$

Add all numbers outside the snow circle.

$$15 + 3 = 18$$

- 1** This Venn diagram shows the number of people who enjoy riding and running.
- a** How many people in total are represented by this Venn diagram?
- b** How many people enjoy:
- riding only?
 - riding (in total)?
 - running only?
 - running (in total)?
 - both riding and running?
 - neither riding nor running?
 - riding or running?
- c** How many people do not enjoy:
- riding?
 - running?

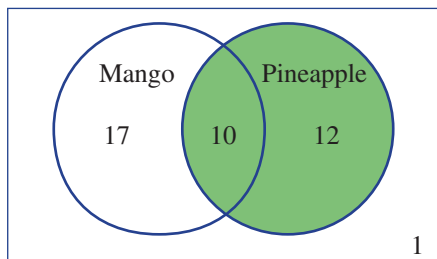


'Riding only' does not include the overlapping section.

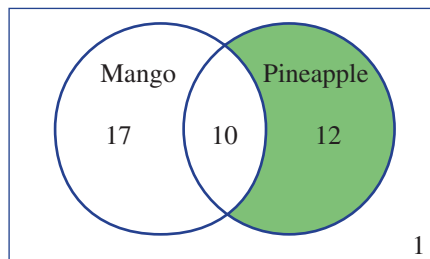


- 2 Each statement is about the shaded area in the Venn diagram. State the missing number in each statement.

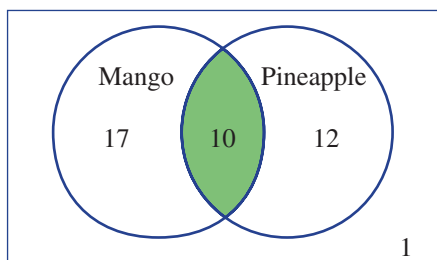
a _____ people in total like pineapple.



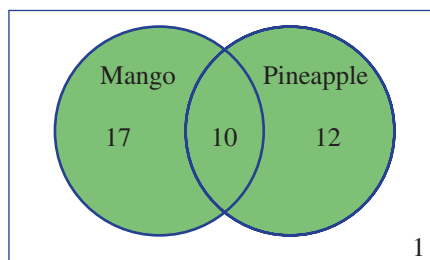
b _____ people like pineapple only.



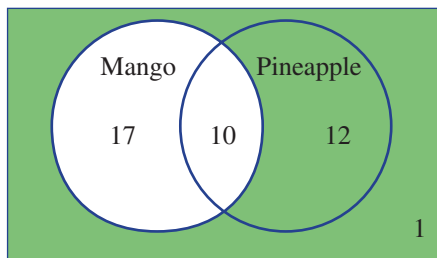
c _____ people like both pineapple and mango.



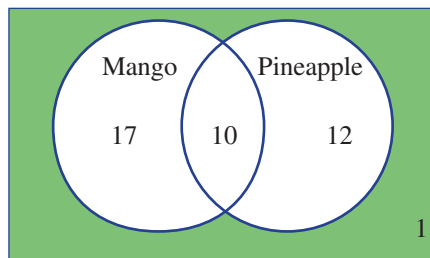
d _____ people like pineapple or mango or both.



e _____ people don't like mango.

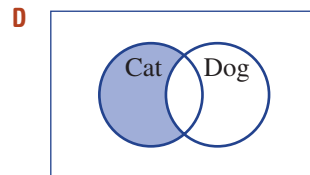
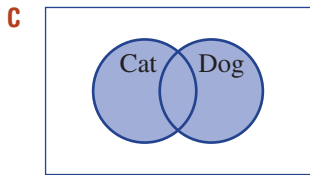
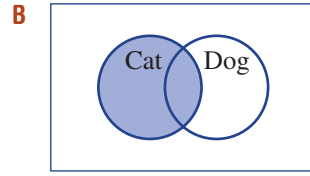
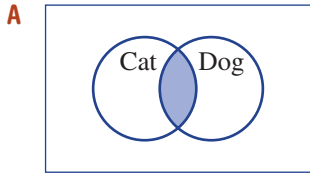


f _____ person likes neither mango nor pineapple.



3 Match the diagrams **A–D** with the given description.

- a Own a cat
- b Own a cat only
- c Own both a cat and a dog
- d Own a cat or a dog or both



'Both a cat and a dog' is where the circles overlap.



Example 4 Completing a two-way table

Copy and complete this two-way table.

	A	Not A	Total
B	7	6	
Not B		2	
Total	17		

SOLUTION

	A	Not A	Total
B	7	6	13
Not B	10	2	12
Total	17	8	25

EXPLANATION

Start with a row or column that has 2 numbers.

'B' row: $7 + 6 = 13$

'A' column: $7 + 10 = 17$

'Not B' row: $10 + 2 = 12$

'Not A' column: $6 + 2 = 8$

Check that the sum of the row totals = the sum of the column totals.

Sum of column totals = $17 + 8 = 25$

Sum of row totals = $13 + 12 = 25$

4 Copy and complete these two-way tables.

a

	A	Not A	Total
B	7	8	
Not B		1	
Total	10		

b

	A	Not A	Total
B	2		7
Not B		4	
Total			20

Start with the column or row that has two numbers in it.



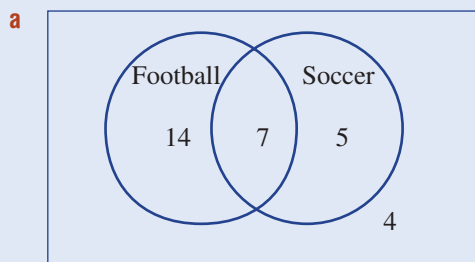


Example 5 Constructing a Venn diagram

A survey of 30 people found that 21 like Australian Rules football and 12 like soccer. Also, 7 people like both football and soccer and 4 like neither football nor soccer.

- a** Construct a Venn diagram for the survey results.
- b** How many people:
- like football or soccer?
 - do not like soccer?
 - like only football?
- c** If one of the 30 people was randomly selected, find:
- $P(\text{like football and soccer})$
 - $P(\text{like neither football nor soccer})$
 - $P(\text{like only soccer})$

SOLUTION



- b**
- 26 like football or soccer or both
 - $30 - 12 = 18$ don't like soccer
 - 14 like football but not soccer
- c**
- $P(\text{like football and soccer})$

$$= \frac{7}{30}$$
 - $P(\text{like neither football nor soccer})$

$$= \frac{4}{30}$$

$$= \frac{2}{15}$$
 - $P(\text{like soccer only})$

$$= \frac{5}{30}$$

$$= \frac{1}{6}$$

EXPLANATION

Place the appropriate number in each category. Place the 7 in the overlap first, then ensure that:

- the total that like football is 21 ($14 + 7 = 21$).
- the total that like soccer is 12 ($5 + 7 = 12$).

The 4 people that like neither are placed outside the circles.

The total number of people who like football, soccer or both is $14 + 7 + 5 = 26$.

12 like soccer, so 18 do not.

21 like football but 7 of these also like soccer.

7 out of 30 people like football and soccer.

The 4 people who like neither football nor soccer sit outside both categories.

5 people like soccer but not football.



Example 6 Constructing a two-way table

At a car yard, 24 cars are tested for fuel economy. Eighteen of the cars run on petrol, 8 cars run on gas and 3 cars can run on both petrol and gas.

- a** Illustrate the situation using a two-way table.
- b** How many of the cars:
- i** do not run on gas? **ii** run on neither petrol nor gas?
- c** Find the probability that a randomly selected car:
- i** runs on gas **ii** runs on only gas **iii** runs on gas or petrol

SOLUTION

a

	Gas	Not gas	Total
Petrol	3	15	18
Not petrol	5	1	6
Total	8	16	24

- b i** 16
ii 1

c i $P(\text{gas}) = \frac{8}{24}$
 $= \frac{1}{3}$

ii $P(\text{only gas}) = \frac{5}{24}$

iii $P(\text{gas or petrol}) = \frac{15 + 5 + 3}{24}$
 $= \frac{23}{24}$

EXPLANATION

Set up a table as shown and enter the numbers (in black) from the given information. Fill in the remaining numbers (in red), ensuring that each column and row adds to the correct total.

The total at the base of the 'Not gas' column is 16. The number at the intersection of the 'Not gas' column and the 'Not petrol' row is 1.

Of the 24 cars, 8 cars run on gas.

Of the 8 cars that run on gas, 5 of them do not also run on petrol.

Of the 24 cars, some run on petrol only (15), some run on gas only (5) and some run on gas and petrol (3).

- 8** Of 50 desserts served at a restaurant one evening, 25 were served with ice-cream, 21 were served with cream and 5 were served with both cream and ice-cream.

- a** Copy and complete this two-way table.

	Cream	Not cream	Total
Ice-cream			
Not ice-cream			
Total			

- b** How many of the desserts:
- i** did not have cream?
ii had neither cream nor ice-cream?
- c** Find the probability that a chosen dessert:
- i** had cream **ii** had only cream **iii** had cream or ice-cream

25 is the total for the ice-cream row.



9 Find the following probabilities using each of the given tables.

First copy and complete each two-way table.

- i $P(A)$ ii $P(\text{not } A)$
- iv $P(A \text{ or } B)$ v $P(B \text{ only})$

a

	A	Not A	Total
B	3	1	
Not B	2		4
Total			

iii $P(A \text{ and } B)$

vi $P(\text{neither } A \text{ nor } B)$

b

	A	Not A	Total
B		4	15
Not B	6		
Total			26

'Neither A nor B' is the same as 'Not A and not B'.

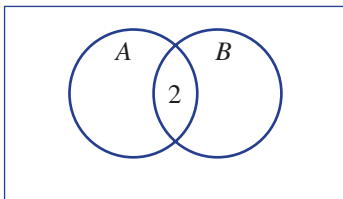


PROBLEM SOLVING AND REASONING 10–12 10, 12–14

10 For each two-way table, fill in the missing numbers then transfer the information to a Venn diagram.

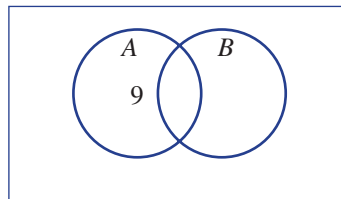
a

	A	Not A	Total
B	2		8
Not B			
Total		7	12



b

	A	Not A	Total
B		4	
Not B	9		13
Total	12		



11 In a group of 17 people, 13 rented their house, 6 rented a car and 3 did not rent either a car or their house.

- a Draw a Venn diagram, showing circles for 'Rents house' and 'Rents car'.
- b How many people rented both a car and their house?
- c Find the probability that one of the people in the group rented only a car.

14 people will be in the circles, but $13 + 6 = 19$. How many must be in the overlap?

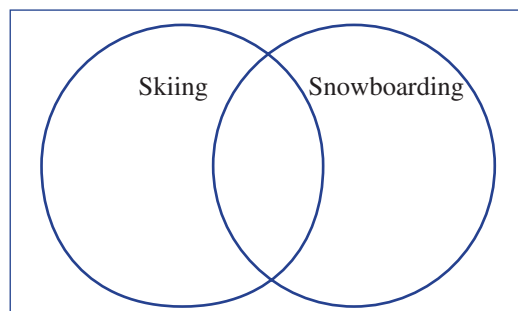


12 One hundred citizens were surveyed regarding their use of water in their garden. Of these, 23 said that they used tank water, 48 said that they used tap water and 41 said that they did not use water on their garden at all.

- a Copy and complete this two-way table.
- b How many people used both tank and tap water?
- c What is the probability that one of the people uses only tap water?
- d What is the probability that one of the people uses tap water or tank water?

	Tank water	Not tank water	Total
Tap water			
Not tap water			
Total			

- 13** All members of a ski club enjoy skiing and/or snowboarding. Seven enjoy only snowboarding, 16 enjoy skiing and 4 enjoy both snowboarding and skiing.



- a** Copy and complete this Venn diagram.
b How many people are in the club?
c If a person from the club is randomly selected, what is the probability of choosing a snowboarder?
d If a person is randomly selected out of the group that likes skiing, what is the probability of choosing a snowboarder?
e What is the probability of choosing a skier out of the group that likes snowboarding?

For part **d**, what fraction of skiers also like snowboarding?



- 14** Of a group of 30 cats, 24 like either tinned or dry food or both, 10 like only dry food and 5 like both tinned and dry food.

- a** Find the probability that a selected cat likes only tinned food.
b Out of the group of cats that eat dry food, what is the probability of selecting a cat that also likes tinned food?

Use a Venn diagram to record the information.



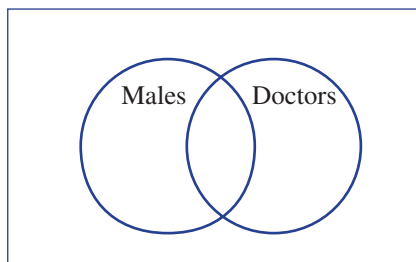
ENRICHMENT

–

15

Numbers challenge

- 15** One hundred people were surveyed and it was found that 55 were males and 30 were doctors. The number of male doctors was 17. Copy and complete this Venn diagram and then determine the number of people in each question below.



- a** The number who are neither male nor a doctor
b The number who are not males
c The number who are not doctors
d The number who are male but not a doctor
e The number who are a doctor but not male
f The number who are female and a doctor
g The number who are female or a doctor

10C Using arrays for two-step experiments



Interactive



Widgets



HOTsheets



Walkthrough

Sometimes an experiment consists of two events, such as tossing a coin and then rolling a die. Or perhaps a card is pulled from a hat and then a spinner is spun. In this type of experiment, we can use arrays to list the sample space.

Let's start: The maths cup

This activity can be run in small groups or as a class.

- 1 Your teacher will draw the table below on the whiteboard. Each student should also have a copy.
- 2 Select one horse as your 'own' (choose a winner!).
- 3 Take turns with other students to roll 2 dice and state the sum of the uppermost faces of the dice.
- 4 The total of each roll refers to the horse number. When its number is rolled, that horse moves another 100 m towards the finish line. A cross is placed in the cell to show the move.
- 5 The winning horse is the first to reach the finish at 1000 m.
- 6 Keep rolling the dice until first, second and third places are decided.



Stage

5.3#

5.3

5.3\$

5.2

5.2◇

5.1

4

	Horse	100 m	200 m	300 m	400 m	500 m	600 m	700 m	800 m	900 m	1000 m
1	SCRATCHED										
2	Greased Lightning										
3	Flying Eagle										
4	Quick Stix										
5	Break a Leg										
6	Slow and Steady										
7	The Donkey										
8	The Wombat										
9	Tooting Tortoise										
10	Ripper Racer										
11	Speedy Gonzo										
12	Pharoah										

Discussion questions

- Do you think that the horse that won your maths cup race would always win?
- In this game, do all the horses have the same chance of winning?
- Why is horse 1 scratched?
- When two dice are thrown, what are all the possible outcomes for the sum of the two uppermost faces?
- What are the horse numbers that are highly likely to win this race?
- What are the horse numbers that are unlikely to win this race?

- A table called an array can be used to list the sample space of a two-step experiment.
- When listing outcomes, it is important to be consistent with the order for each outcome. For example: the outcome (heads, tails) is different from the outcome (tails, heads). So when two coins are tossed, the outcome of coin 1 is written first in each cell of the table.

		Coin 2	
		H	T
Coin 1	H	HH	HT
	T	TH	TT

The sample space is shown in the body of the table; i.e. {HH, HT, TH, TT}.

- The probability is still given by:

$$P(\text{event}) = \frac{\text{number of outcomes where the event occurs}}{\text{total number of possible outcomes}}$$

For example: When two coins are tossed, $P(\text{a tail and a head}) = \frac{2}{4} = \frac{1}{2}$

- Some experiments are conducted without replacement, which means that each individual trial outcome cannot be achieved again. For example: Two letters are chosen from the word CAT. 'Without replacement' means that the first letter chosen is removed and it is not possible to choose it again.

With replacement

		1st		
		C	A	T
2nd	C	(C, C)	(A, C)	(T, C)
	A	(C, A)	(A, A)	(T, A)
	T	(C, T)	(A, T)	(T, T)

9 outcomes

Without replacement

		1st		
		C	A	T
2nd	C	×	(A, C)	(T, C)
	A	(C, A)	×	(T, A)
	T	(C, T)	(A, T)	×

6 outcomes



Exercise 10C

UNDERSTANDING AND FLUENCY

1-6

4-7

- 1 A letter is chosen from the word BYE and a letter is chosen from the word HI. Copy and complete this table to show the sample space.

	B	Y	E
H	HB		
I			

Write 'H' in each cell of the top row.
Write 'B' in each cell of the first column.



- 2 A coin and a four-sided die are tossed.



- a Copy and complete this table to show the sample space.

	1	2	3	4
H				
T				

The sample space is all the outcomes in the central part of the table.



- b How many outcomes are in the sample space?
c What is the probability of tossing an 'H3'?
d What is the probability of tossing a 'T2'?

- 3 A coin is flipped and then a spinner with numbers 1 to 5 is spun. The possible outcomes are listed in the table below.

	1	2	3	4	5
H	H1	H2	H3	H4	H5
T	T1	T2	T3	T4	T5

- a How many outcomes are possible?
b List the four outcomes in which an even number is displayed on the spinner.
c Hence, state the probability that an even number is displayed.
d List the outcomes for which the coin shows tails and the spinner shows an odd number.
e What is $P(\text{T, odd number})$?

- 4 Two coins are tossed and the four possible outcomes are shown below.

		20-cent coin	
		H	T
50-cent coin	H	HH	HT
	T	TH	TT

The same face means either two heads or two tails.



- a What is the probability that the 50-cent coin will be heads and the 20-cent coin will be tails?
b For which outcomes are the two coins displaying the same face?
c What is the probability of the two coins displaying the same face?



Example 7 Using a table for multiple events

A spinner with the numbers 1, 2 and 3 is spun, and then a card is chosen at random from cards containing the letters A, T, H and S.

- Draw a table to list the sample space of this experiment.
- How many outcomes does the experiment have?
- Find the probability of the combination 2S.
- Find the probability of an odd number being spun and the letter H being chosen.

SOLUTION

a

	A	T	H	S
1	1A	1T	1H	1S
2	2A	2T	2H	2S
3	3A	3T	3H	3S

b There are 12 outcomes.

c $P(2, S) = \frac{1}{12}$

d $P(\text{odd}, H) = \frac{2}{12} = \frac{1}{6}$

EXPLANATION

The sample space of the spinner {1, 2, 3} is put into the left column.

The sample space of the cards {A, T, H, S} is put into the top row.

The table has $4 \times 3 = 12$ items in it.

All 12 outcomes are equally likely. Spinning 2 and choosing an S is one of the 12 outcomes.

Possible outcomes are 1H and 3H, so
probability = $2 \div 12$

5 A coin is flipped and then a regular die is rolled.

a Copy and complete this table to list the sample space for this experiment.

	1	2	3	4	5	6
H						
T						

b How many possible outcomes are there?

c Find the probability of the pair H3.

d Find the probability of 'heads' on the coin with an odd number on the die.

6 A letter is chosen from the word LINE and another is chosen from the word RIDE.

a Copy and complete this table to list the sample space.

	R	I	D	E
L				
I				
N				
E				

b How many possible outcomes are there?

c Find $P(N, R)$; i.e. the probability that N is chosen from LINE and R is chosen from RIDE.

d Find $P(L, D)$.

e Find the probability that two vowels are chosen.

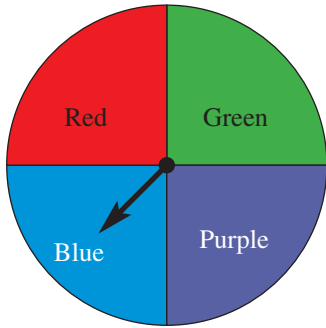
f Find the probability that two consonants are chosen.

g Find the probability that the two letters chosen are the same.

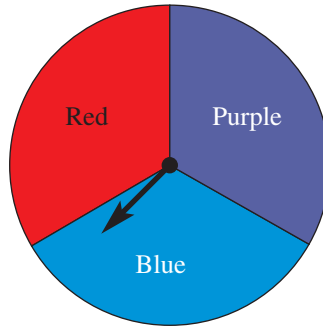
Recall that the vowels are A, E, I, O and U.



7 The spinners shown below are each spun.



Spinner 1



Spinner 2

a Copy and complete this table to list the sample space. Use R for red, P for purple and so on.

		Spinner 2		
		R	P	B
Spinner 1	R			
	P			
	G			
	B			

List the colour from spinner 1 first in each cell.



- b Find the probability that spinner 1 will display red and spinner 2 will display blue.
- c Find the probability that both spinners will display red.
- d What is the probability that one of the spinners displays red and the other displays blue?
- e What is the probability that both spinners display the same colour?

PROBLEM SOLVING AND REASONING

8, 9

8–11

8 Two dice are rolled for a board game. The numbers showing are then added together to get a number between 2 and 12.

Write the sum in each cell. For example, for (1, 1), write 2.



- a Draw a table to show the sample space and state the number of outcomes.
- b Find the probability that the two dice add to 5.
- c Find the probability that the two dice add to an even number.
- d What is the most likely sum to occur?
- e What are the two least likely sums to occur between 2 and 12?

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

- 9 In Rosemary's left pocket she has two orange marbles and one white marble. In her right pocket she has a yellow marble, a white marble and 3 blue marbles. She chooses a marble at random from each pocket.

- Draw a table to describe the sample space.
- Find the probability that she will choose an orange marble and a yellow marble.
- What is the probability that she chooses a white marble and a yellow marble?
- What is the probability that she chooses a white marble and an orange marble?
- Find the probability that a white and a blue marble are selected.
- What is the probability that the two marbles selected are the same colour?

The left-pocket outcomes are W, O, O.



Example 8 Finding the sample space for events without replacement

The letters **T** **R** **E** **K** are written on cards. A letter is chosen and not replaced so that this letter is now unavailable. Then a second letter is chosen.

- In a table, list the outcomes from choosing two letters without replacement.
- Find the probability that the two letters chosen are (E, R), in that order.
- Find the probability of obtaining an outcome with an E in it.

SOLUTION

		1st choice			
		T	R	E	K
2nd choice	T	×	(R, T)	(E, T)	(K, T)
	R	(T, R)	×	(E, R)	(K, R)
	E	(T, E)	(R, E)	×	(K, E)
	K	(T, K)	(R, K)	(E, K)	×

b $P(E, R) = \frac{1}{12}$

c $P(\text{includes E}) = \frac{6}{12} = \frac{1}{2}$

EXPLANATION

List all the outcomes, writing the '1st choice' letter first in each cell. Note that the same letter cannot be chosen twice.

Without replacement, there are 12 outcomes.

6 of the 12 outcomes contain an E.

- 10 Two letters are chosen from the word DOG *without replacement*.

- Copy and complete the given table.

		1st		
		D	O	G
2nd	D	×	(O, D)	(G, D)
	O		×	
	G			×

- Find the probability of obtaining the (G, D) outcome.
- Find the probability of obtaining an outcome with an O in it.

'Without replacement' means the first letter chosen is removed, so it is not possible to choose it again.



- 11** Two digits are selected *without replacement* from the set $\{1, 2, 3, 4\}$.
- a** Draw a table to show the sample space.
- b** Find:
- $P(1, 2)$
 - $P(4, 3)$
- c** Find the probability that:
- both numbers will be at least 3
 - the outcome will contain a 1 or a 4
 - the outcome will contain a 1 and a 4
 - the outcome will not contain a 3

Remember that doubles such as $(1, 1)$, $(2, 2)$ etc. are not allowed.



'At least 3' includes 3.



ENRICHMENT

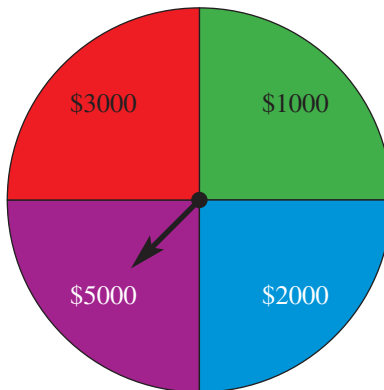
-

12

Game show

- 12** A wheel is spun during a game show to determine the prize money, and then a six-sided die is rolled. The prize money shown on the wheel is multiplied by the number on the die to give the total winnings.
- a** What is the probability that a contestant will win \$6000?
- b** What is the probability that a contestant will win more than \$11000?

Draw a table to show all outcomes.
List the prize in brackets under each outcome.

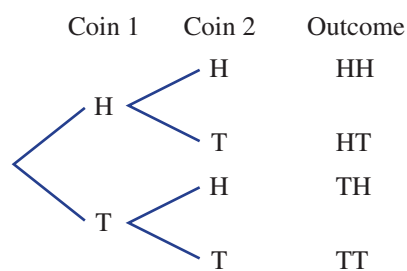


10D Using tree diagrams



When two coins are tossed, we can draw a table to list the sample space. However, if three coins were tossed, we would need a three-dimensional table to list all outcomes. Imagine trying to find probabilities when five coins are tossed!

Another tool that mathematicians use for probability is the tree diagram. The tree diagram shown here describes the four outcomes when two coins are tossed.



Stage

5.3#
5.3
5.3\$
5.2
5.2◊
5.1
4

It is important to be able to read a tree diagram correctly. The first row (HH) represents the outcome where the first coin tossed was heads and the second coin tossed was heads. The third row (TH) represents the outcome where the first coin was tails and the second was heads.

Let's start: Coin puzzle

- If two coins are tossed, rank these outcomes from most likely to least likely.
 - Getting exactly two heads
 - Getting exactly one head and exactly one tail
 - Getting at least one coin showing tails
 - Getting three tails.
- How might the order change if three coins are flipped? Compare your answers with other students.

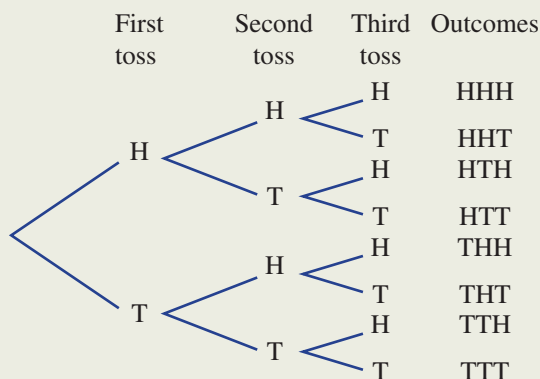
Key ideas

■ **Tree diagrams** are used to list the sample space for multistage experiments with two or more steps.

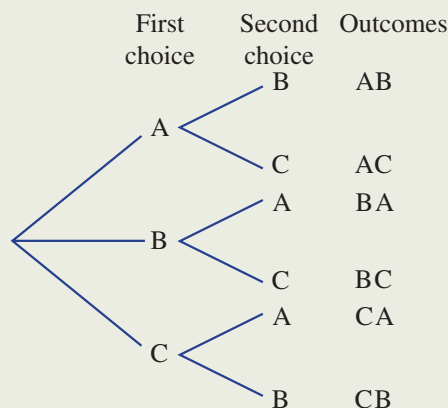
- The outcomes for each stage of the experiment are listed vertically and each stage is connected with branches.

For example:

Tossing a coin 3 times (with replacement)



Selecting 2 letters from {A, B, C} (without replacement)



In these examples, each set of branches produces outcomes that are all equally likely. Tree diagrams can also be used where all outcomes are not equally likely.

Tree diagram A diagram used to show all the possibilities when several different options are available

Exercise 10D

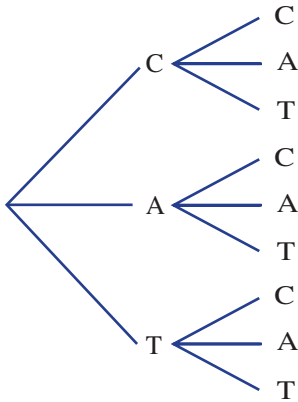
UNDERSTANDING AND FLUENCY

1-6

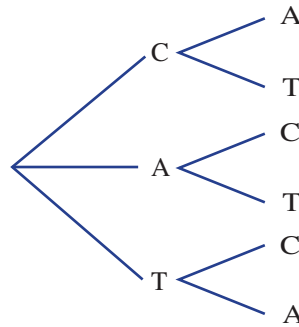
3-7

1 These two tree diagrams show the selection of two letters from the word CAT.

A



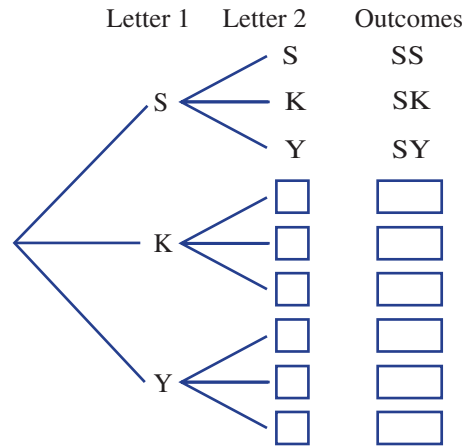
B



- a Which tree diagram shows selection with replacement?
- b Which tree diagram shows selection without replacement?

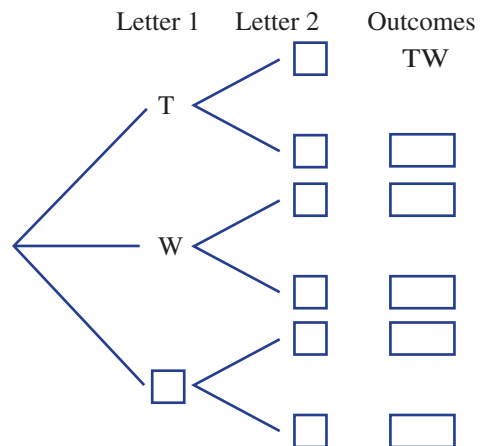
2 The letters S, K and Y are written on cards and placed in a hat. Two letters are randomly selected one at a time.

- a Write the missing word or number from each statement about this event.
 - i 'With _____' means the first letter is returned to the hat before the second letter is selected.
 - ii When the first letter chosen is replaced, there are still _____ letters to choose from for the second letter.
- b Copy and complete this tree diagram for selecting two letters *with replacement*.



3 The letters T, W and O are written on cards and placed in a container. One letter is randomly selected and not returned to the container. A second letter is then selected.

- a Write the missing word or number from each statement about this event.
 - i '_____ _____' means the first letter chosen is placed aside and the second letter is selected from the remaining letters.
 - ii Since the first letter chosen is not replaced, there will be only _____ letters to choose from for the second letter.
- b Copy and complete this tree diagram for selecting two letters *without replacement*.



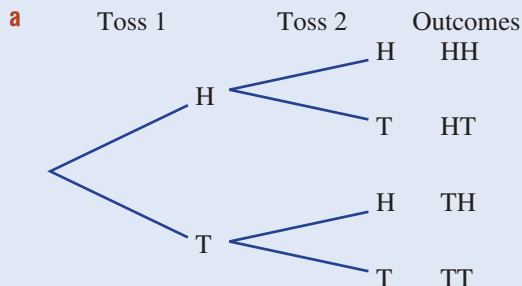


Example 9 Constructing a tree diagram

An experiment involves tossing two coins.

- a** Complete a tree diagram to show all possible outcomes.
b What is the total number of outcomes?
c Find the probability of tossing:
- i** two tails **ii** one tail **iii** at least one head

SOLUTION



EXPLANATION

Tree diagram shows two coin tosses one after the other, resulting in $2 \times 2 = 4$ outcomes.

- b** The total number of outcomes is 4

There are four possibilities in the outcomes column.

c i $P(TT) = \frac{1}{4}$

One out of the four outcomes is TT.

ii $P(1 \text{ tail}) = \frac{2}{4} = \frac{1}{2}$

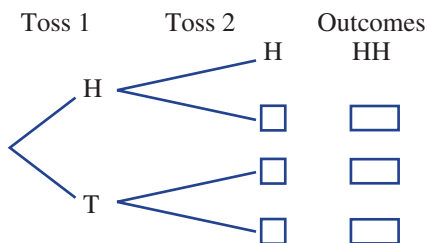
Two outcomes have one tail:
{HT, TH}

iii $P(\text{at least 1 head}) = \frac{3}{4}$

Three outcomes have at least one head:
{HH, HT, TH}

- 4** A coin is tossed twice.

- a** Complete this tree diagram to show all the possible outcomes.
b What is the total number of outcomes?
c Find the probability of obtaining:
- i** two heads
ii exactly one head
iii at least one head
iv at least one tail

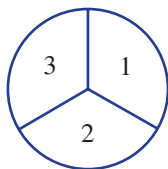


'At least one' means one or two in this case.



- 5** A spinner with numbers 1, 2 and 3 is spun twice.

- a** Draw a tree diagram and list the outcomes.
b Find $P(1 \text{ then } 1)$.
c Find $P(1 \text{ then } 2)$.
d Find $P(1 \text{ and } 2 \text{ spun in either order})$.
e Find $P(\text{both show the same number})$.
f Find $P(\text{numbers add to } 4)$.



Tree diagram headings:

First spin Second spin Outcomes



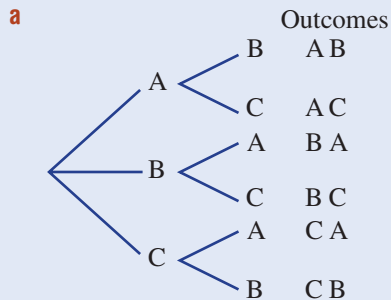


Example 10 Constructing a tree diagram without replacement

Two people are selected without replacement from a group of three: Annabel (A), Brodie (B) and Chris (C).

- a** List all the possible combinations for the selection using a tree diagram.
b Find the probability that the selection will contain:
- Annabel and Brodie
 - Chris
 - Chris or Brodie

SOLUTION



b i $P(\text{Annabel and Brodie}) = \frac{2}{6}$
 $= \frac{1}{3}$

ii $P(\text{Chris}) = \frac{4}{6}$
 $= \frac{2}{3}$

iii $P(\text{Chris or Brodie}) = \frac{6}{6}$
 $= 1$

EXPLANATION

On the first choice there are three options (A, B or C) but on the second choice there are only two remaining.

2 of the 6 outcomes contain Annabel and Brodie (A, B) and (B, A).

4 out of the 6 outcomes contain Chris.

All of the outcomes contain at least one of Chris or Brodie.

- 6** Two people are selected without replacement from a group of three: Donna (D), Elle (E) and Fernando (F).

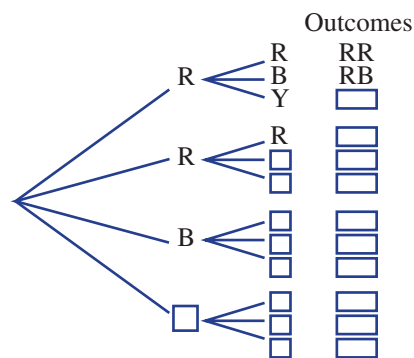
- a** List all the possible combinations for the selection using a tree diagram.
b Find the probability that the selection will contain:
- Donna and Elle
 - Fernando
 - Fernando or Elle

The combinations are the outcomes.



- 7 A drawer contains 2 red socks (R), 1 blue sock (B) and 1 yellow sock (Y) and two socks are selected at random without replacement.

- a Complete this tree diagram.
 b Find the probability of obtaining:
 i a red sock and a blue sock
 ii two red socks
 iii any pair of socks of the same colour
 iv any pair of socks of different colour



PROBLEM SOLVING AND REASONING

8

8, 9

- 8 A coin is tossed three times.
 a Draw a tree diagram and list the outcomes.
 b Find $P(3 \text{ tails})$.
 c Find $P(2 \text{ tails then } 1 \text{ head})$.
 d Find $P(2 \text{ tails and } 1 \text{ head, in any order})$.
 e Which is more likely: getting exactly 3 tails or getting exactly 2 tails?

Use headings: 'Coin 1',
'Coin 2', 'Coin 3', 'Outcomes'



- 9 There are three bottles of white wine and two bottles of red wine on a shelf in a cellar. It is too dark to read the labels so two bottles are randomly selected, one at a time.

Find the probability that:

- a two bottles of different colour wine are selected
 b two bottles of the same colour wine are selected
 c one or more bottles of red wine are selected

Draw a tree diagram
using the 'without
replacement' method.



ENRICHMENT

-

10

Selecting matching clothes

- 10 A man randomly selects a tie from his collection of one green and two red ties. He selects a shirt from a collection of one red and two white. He then selects either a red hat or a black hat. Use a tree diagram to help find the probability that the man selects a tie, shirt and hat according to the following descriptions.
- a A red tie, red shirt and black hat
 b All three items red
 c One item red
 d Two items red
 e At least two items red
 f Green hat
 g Green tie and a black hat
 h Green tie or a black hat
 i Not a red item
 j Red tie or white shirt or black hat

10E Using relative frequencies to estimate probabilities



Interactive



Widgets



HOTsheets



Walkthrough

The sample space is used to calculate theoretical probability. Experimental probability is calculated from the results of a survey, an experiment or a simulation. A larger number of trials make experimental probability calculations more accurate. Experimental probability is used to predict an expected number of results.

Stage

5.3#

5.3

5.3\$

5.2

5.2◊

5.1

4

Let's start: Newspaper theories

A tabloid newspaper reports that of 10 people interviewed in the street, 5 had a dose of the flu. At a similar time a medical student tested 100 people and found that 21 had the flu.

- What is the experimental probability of having the flu, according to the newspaper's survey?
- What is the experimental probability of having the flu, according to the medical student's results?
- Which of the two sets of results would be most reliable and why? Discuss the reasons.
- According to the results from the medical student, how many people would you expect to have the flu in a group of 1000, and why?



- The results of repeated trials of a probability experiment can be used to generate **relative frequencies**, which are sometimes called experimental probabilities.

- **Experimental probability** is calculated using the results of an experiment or survey.

$$\text{Experimental probability} = \frac{\text{number of times the outcomes occurs}}{\text{total number of trials in the experiment}}$$

- The **expected number of occurrences** = probability \times number of trials

$$E = P \times n$$

Relative frequency The proportion (or fraction) of times a particular outcome occurs, written as a fraction, a decimal or a percentage.

Experimental probability Probability based on measuring the outcomes of trials

Expected number of occurrences The expected number of favourable outcomes from an experiment

Key ideas

Exercise 10E

UNDERSTANDING AND FLUENCY

1–6

2–7

- 1 Insert the word *experimental*, *theoretical* or *expected* to complete each statement.
- a _____ probability is calculated from the results of an experiment or survey.
- b _____ probability is calculated from the sample space of an event.
- c When a coin is tossed, there are 2 possible equally likely outcomes. So the _____ probability of obtaining a head is $\frac{1}{2}$.
- d Jayce tossed a coin 20 times and obtained a head 12 times. So the _____ probability of obtaining a head was $\frac{12}{20} = \frac{3}{5}$.
- e A survey found that 3 out of every 5 students have a sister. Out of 1000 students, the _____ number of students with a sister would be $\frac{3}{5} \times 1000 = 600$.
- 2 Write the experimental probability from each set of results listed here.
- a Out of 20 people surveyed, 18 preferred cereal for breakfast.
- b 50 students were surveyed and it was found that 40 had pets.
- c There were 25 boys and 20 girls at the school bus stop one afternoon. If one child is selected from those at the bus stop, state the experimental probability that the child is a boy.

In part c, first find the total number of students at the bus stop.



Example 11 Using experimental probability

Kris plays rugby and has a record of kicking a goal from a penalty kick 4 times out of every 7 attempts.

- a State the experimental probability for Kris to achieve a penalty goal.
- b Calculate the expected number of goals from 28 penalty kicks that Kris takes.

SOLUTION

a $P(\text{goal}) = \frac{4}{7}$

b Expected number = $\frac{4}{7} \times 28 = 16$ goals

EXPLANATION

Experimentally, we know that Kris can kick 4 goals out of every 7 kicks.

Expected number = probability \times number of trials

- 3 Ashleigh has found that she can shoot a basketball through the hoop 4 times out of 10 from the '3-point' area.
- a State the experimental probability for Ashleigh to shoot a basketball through the hoop from the 3-point area.
- b Calculate the expected number of times the ball would go through the hoop from 100 shots Ashleigh makes from the 3-point area.

- 4 The experimental probability of Jess hitting a bullseye on a dartboard is 0.05 (or $\frac{5}{100}$). How many bullseyes would you expect Jess to get if he threw the following number of darts?
- 100 darts
 - 200 darts
 - 1000 darts
 - 80 darts

Expected number = probability
× number of trials



- 5 A bus company surveyed a random selection of 90 people in one suburb. They found that 35 of these people regularly used a bus service from that suburb to the centre of the city.
- State the experimental probability for a person in that suburb to regularly use a bus service from that suburb to the city.
 - If there are 2700 residents in that suburb, find the expected number who would regularly use a bus service from that suburb to the city.



Example 12 Finding the experimental probability

A box contains an unknown number of coloured balls and a ball is drawn from the box and then replaced. The procedure is carried out 100 times and the colour of the ball drawn is recorded each time. Twenty-five red balls were recorded.

- Find the experimental probability for selecting a red ball.
- Find the expected number of red balls if the box contained 500 balls in total.

SOLUTION

- $$P(\text{red ball}) = \frac{25}{100}$$

$$= 0.25$$
- Expected number of red balls in 500

$$= 0.25 \times 500$$

$$= 125$$

EXPLANATION

$$P(\text{red ball}) = \frac{\text{number of red balls drawn}}{\text{total number of balls drawn}}$$

There are 25 red balls and 100 balls in total.

Expected number of occurrences
= probability × number of trials

- 6 A bag contains an unknown number of counters, and a counter is selected from the bag and then replaced. The procedure is carried out 100 times and the colour of the counter is recorded each time. Sixty of the counters drawn were blue.

- Find the experimental probability of selecting a blue counter.
- Find the expected number of blue counters if the bag contained:
 - 100 counters
 - 200 counters
 - 600 counters

Experimental probability
= $\frac{\text{number of blue counters}}{\text{total number of counters selected}}$



- 7 In an experiment involving 200 people chosen at random, 175 people said that they owned a home computer.
- Calculate the experimental probability of choosing a person who owns a home computer.
 - Find the expected number of people who would own a home computer in the following group sizes.
 - 400 people
 - 5000 people
 - 40 people

Expected number =
probability × number of trials



PROBLEM-SOLVING AND REASONING

8–10

8, 10–13

- 8 By calculating the experimental probability, estimate the chance that each of the following events will occur.
- Nat will walk to work today, given that she walked to work five times in the last 75 working days.
 - Mike will win the next game of cards if, in the last 80 games, he has won 32.
 - Brett will hit the bullseye on the dartboard with his next attempt if, in the last 120 attempts, he was successful 22 times.

- 9 This table shows the results of three different surveys of people in Perth about their use of public transport (PT).

Survey	Number who use PT	Survey size	Experimental probability
A	2	10	$\frac{2}{10} = 0.2$
B	5	20	_____
C	30	100	_____

A large survey size makes experimental probability calculations more accurate.



- What are the two missing numbers in the experimental probability list?
 - Which survey should be used to estimate the probability that a person uses public transport and why?
- 10 The results of tossing a drawing pin and observing how many times the pin lands with the spike pointing up are shown in the table. Results are recorded at different stages of the experiment.

Number of throws	Frequency (spike up)	Experimental probability
1	1	1.00
5	2	0.40
10	5	0.50
20	9	0.45
50	18	0.36
100	41	0.41

Which experimental probability would you choose to best represent the probability that the pin will land spike up? Why?

- 11 A six-sided die is rolled 120 times. How many times would you expect the following events to occur?
- Rolling a 6
 - Rolling a 1 or 2
 - Rolling a number less than 4
 - Rolling a number which is at least 5

Use the theoretical probability to calculate the expected numbers.



- 12 The colour of cars driving along a highway was noted over a short period of time and summarised in this frequency table.

Colour	White	Silver	Blue	Green
Frequency	7	4	5	4

- a How many cars had their colour recorded?
 b Find the experimental probability that a car's colour is:
 i blue
 ii white
 c If the colour of 100 cars was recorded, find the expected number of:
 i blue cars
 ii green cars
 iii blue or green cars
- 13 A spinner is divided into three regions not necessarily of equal size. The regions are numbered 1, 2 and 3 and the spinner is spun 50 times. The table shows the results.

Number	1	2	3
Frequency	26	11	13

The angles of the spinner regions are multiples of 90° .



- a Find the experimental probability of obtaining:
 i a 1
 ii at least a 2
 iii a 1 or a 3
 b Based on these results, how many 3s would you expect if the spinner is spun 300 times?
 c In fact, the spinner is divided up using simple and common fractions. Draw and label how you think the spinner regions are divided up.

ENRICHMENT

–

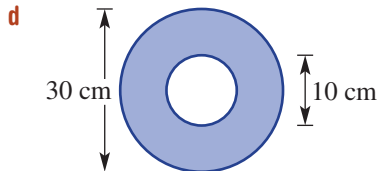
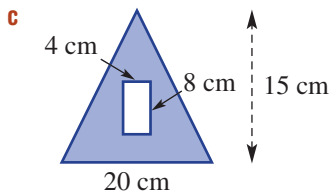
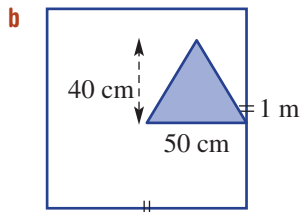
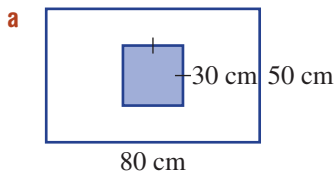
14

Different-shaped dartboards



- 14 One hundred darts are randomly thrown at the given dartboards. No darts miss the dartboard entirely. How many darts do you expect to hit the blue shaded region? Give reasons.

First find the proportion of the entire dartboard taken up by each blue region.



10F Mean, median, mode and range

REVISION



Interactive



Widgets



HOTsheets



Walkthrough

Statistics involves collecting and summarising data. It also involves drawing conclusions and making predictions. For example, the type and amount of product stocked on supermarket shelves is determined by the sales statistics and other measures such as average cost and price range.

Let's start: Game purchase

Arathi purchases 7 computer games at a sale. 3 games cost \$20 each, 2 games cost \$30, 1 game costs \$50 and the last game cost \$200.

- Recall and discuss the meaning of the words *mean*, *median* and *mode*.
- Can you work out the mean, median or mode for the cost of Arathi's games?
- Which of the mean, median or mode gives the best 'average' for the cost of Arathi's games?
- Why is the mean greater than the median in this case?

Stage

5.3#

5.3

5.3\$

5.2

5.20

5.1

4

Key ideas

■ Statisticians use **summary statistics** to highlight important aspects of a data set. These are summarised below.

■ Some summary statistics are called **measures of location**.

- The two most commonly used measures of location are the **mean** and the **median**. These are also called '**measures of centre**' or '**measures of central tendency**'. Mean and median can only be applied to numerical data.
- The **mean** is sometimes called the '**arithmetic mean**' or the '**average**'.

The formula used for calculating the mean, \bar{x} , is:

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

For example, in the following data set

5	7	2	5	1
---	---	---	---	---

the mean is $\frac{5 + 7 + 2 + 5 + 1}{5} = 20 \div 5 = 4$

- The **median** divides an ordered data set into two sets, each of which contain the same number of data values. It is often called the 'middle value'. The median is found by firstly ensuring the data values are in ascending order, then selecting the 'middle' value.

If the number of values is odd, simply choose the one in the middle.

2	2	3	4	6	9	9
---	---	---	---	---	---	---

Median = 4

If the number of values is even find the average of the two in the middle.

2	2	3	4	7	9	9	9
---	---	---	---	---	---	---	---

Median = $(4 + 7) \div 2 = 5.5$

- The **mode** of a data set is the most frequently occurring data value. There can be more than one mode. When there are two modes, the data is said to be **bimodal**. The mode can be very useful for categorical data. It can also be used for numerical data, but it may not be an accurate measure of centre. For example, in the data set below the mode is 10, but it is the largest data value.

1	2	3	10	10
---	---	---	----	----

Mean (\bar{x}) The average value of data values

Median The middle score (or average of two middle scores) when a set of data is arranged in order

- An **outlier** is a data value which is significantly greater than or less than the other data values in the data set. The inclusion of an outlier in a data set can inflate or deflate the mean and give a false impression of the ‘average’.

For example:

In the data set 1, 2, 3, 4, 5, the mean is 3 and median is 3.

If the data value 5 is changed to 20 it is now an outlier.

In the data set 1, 2, 3, 4, 20, the median is still 3, but the mean is now 6, which is greater than four of the five data values. In this case the median is a better measure of centre than the mean.

■ Some summary statistics are called **measures of spread**

- The **range** is a measure of spread.
- The range of a data set is given by the formula:
range = maximum data value – minimum data value

For example, in the data set below the range is $7 - 1 = 6$

5	7	2	5	1
---	---	---	---	---

↑
↑
maximum
minimum

- Later, you will encounter some other measures of spread such as interquartile range and standard deviation.

Mode The most frequently occurring value in a set of data

Outlier Any value that is much larger or much smaller than the rest of the data in a set

Range The difference between the highest data value and the lowest data value in a set of data

Exercise 10F REVISION

UNDERSTANDING AND FLUENCY

1–5, 6(½), 7, 8

5, 6(½), 7–9

1 a To calculate the _____, you add up all the values and divide by the number of values.

b Find the mean of these data sets (round each answer to one decimal place).

i 3 6 8 10 12

ii 6 8 4 12 9 11 16 13

iii 9 12 1 5 7 10 5 7 13 15

iv 32 35 31 45

First find the sum of the scores, then divide by the number of scores.



$$\text{Mean} = \frac{39}{5}$$



2 a The mode is the most _____ value.

b State the mode of these data sets.

i 1 2 2 2 4 5 6

ii 1 4 8 8 9 10 10 10 12

iii 1 5 7 7 8 9 13

iv 21 23 25 23 21 25 27 25

3 a The median is the _____ value when the scores are listed in order.

b Find the median of these data sets.

i 1 3 5 6 8

ii 4 5 7 9 10 11

iii 5 6 1 2 8 9 3

iv 16 12 15 11 10 13 17 20

Order the data sets in iii and iv.



4 a The _____ is the difference between the highest and lowest values.

b Find the range of these data sets.

i 3 6 7 12 15 24

ii 5 12 14 18 23 27

iii 22 24 26 27 30 35

iv 8 14 5 22 27 32 16

Range = highest value – lowest value



5 a An _____ is a score that is much larger or smaller than the rest of the data.

b For each of these data sets, if there is an outlier, state its value.

i 2 4 5 4 6 2 3 36

ii 21 25 3 27 28 24 29 30

iii 13 15 10 12 17 20 19 22

iv 24 21 100 23 25 27 30 25 25



Example 13 Finding measures of centre and the range

For the given data sets, find the:

i mean

ii median

iii mode

iv range

a 5 2 4 10 6 1 2 9 6

b 17 13 26 15 9 10

SOLUTION

$$\begin{aligned} \text{a i Mean} &= \frac{45}{9} \\ &= 5 \end{aligned}$$

ii 1 2 2 4 5 6 6 9 10
Median = 5

iii Mode = 2 and 6

$$\begin{aligned} \text{iv Range} &= 10 - 1 \\ &= 9 \end{aligned}$$

$$\begin{aligned} \text{b i Mean} &= \frac{90}{6} \\ &= 15 \end{aligned}$$

ii 9 10 13 15 17 26
14

$$\begin{aligned} \text{Median} &= \frac{13 + 15}{2} \\ &= 14 \end{aligned}$$

iii No mode

$$\begin{aligned} \text{iv Range} &= 26 - 9 \\ &= 17 \end{aligned}$$

EXPLANATION

$$5 + 2 + 4 + 10 + 6 + 1 + 2 + 9 + 6 = 45$$

Find the sum of all the numbers and divide by the number of values.

First, order the data.

The median is the middle value.

The data set is bimodal since there are two numbers with the highest frequency.

The range is the highest score minus the lowest score.

$$17 + 13 + 26 + 15 + 9 + 10 = 90$$


There are 6 values.

First, order the data.

Since there are two values in the middle, find the average of them.

None of the values is repeated so there is no mode.

The highest score is 26 and the lowest score is 9.


-  **6** For the given data sets, find the:
- | | |
|-----------------|------------------|
| i mean | ii median |
| iii mode | iv range |
- a** 7 2 3 8 5 9 8
b 6 13 5 4 16 10 3 5 10
c 12 9 2 5 8 7 2 3
d 10 17 5 16 4 14
e 3.5 2.1 4.0 8.3 2.1
f 0.7 3 2.9 10.4 6 7.2 1.3 8.5
g 6 0 -3 8 2 -3 9 5
h 3 -7 2 3 -2 -3 4

The data must be in order to locate the middle value (median).



Recall that to subtract a negative, you add a positive. e.g.
 $7 - (-2) = 7 + 2 = 9$



-  **7** These data sets include an outlier. Write down the outlier and then calculate the mean and the median. Include the outlier in your calculations.
- a** 5 7 7 8 12 33
b 2 40 42 45 47
c 1.3 1.1 1.0 1.7 1.5 1.6 -1.1 1.5
d -58 -60 -59 -4 -64


An outlier is much larger or smaller than the other scores.



- 8** Decide whether the following data sets are bimodal.
- a** 2 7 9 5 6 2 8 7 4
b 1 6 2 3 3 1 5 4 1 9
c 10 15 12 11 18 13 9 16 17
d 23 25 26 23 19 24 28 26 27

Bimodal means that there are two modes.





-  **9** This ordered data set shows the number of fish Daniel caught in the 12 weekends that he went fishing during the year.
- 1 2 3 4 4 5 6 7 7 9 11 13
- a** Find the range.
b Find the median.
c Find the mean.
d Find the mode.

PROBLEM-SOLVING AND REASONING

10, 11

10, 12, 13

-  **10** In three running races, Paula recorded the times 25.1 seconds, 24.8 seconds and 24.1 seconds.
- a** What is the mean time of the races? Round to two decimal places.
b Find the median time.

-  **11** This is a data set of six house prices in Darwin.
- \$324 000 \$289 000 \$431 000 \$295 000 \$385 000 \$1 700 000
- a** Which price would be considered the outlier?
b If the outlier was removed from the data set, by how much would the median change? (First work out the median for each case.)
c If the outlier was removed from the data set, by how much would the mean change, to the nearest dollar? (First work out the mean for each case.)

To find the median, list the prices in order, then find the middle value.





Example 14 Finding a data value for a required mean

The hours a shop assistant spends cleaning the store in eight successive weeks are:

8, 9, 12, 10, 10, 8, 5, 10

- Calculate the mean for this set of data.
- Determine the score that needs to be added to this data to make the mean equal to 10.

SOLUTION

$$\begin{aligned} \text{a Mean} &= \frac{72}{8} \\ &= 9 \end{aligned}$$

- Let a be the new score.

$$\text{Require } \frac{72 + a}{8 + 1} = 10$$

$$\frac{72 + a}{9} = 10$$

$$72 + a = 90$$

$$a = 18$$

The new score would need to be 18.

EXPLANATION

$$8 + 9 + 12 + 10 + 10 + 8 + 5 + 10 = 72$$

Sum of the 8 data values is 72.

$72 + a$ is the total of the new data and $8 + 1$ is the new total number of scores. Set this equal to the required mean of 10.

Solve for a .

9 scores have a mean of 10, so the sum of the scores = $9 \times 10 = 90$.

Write the answer.



- A netball player scored the following number of goals in her 10 most recent games:

15 14 16 14 15 12 16 17 16 15

- What is her mean score?
- What number of goals does she need to score in the next game for the mean of her scores to be 16?



- Stevie obtained the following scores on her first five Maths tests: 92 89 94 82 93

- What is her mean test score?
- If there is one more test left to complete, and she wants to achieve an average of at least 85, what is the lowest score Stevie can obtain for her final test?

ENRICHMENT

–

14

Aiming for an A



- A school gives grades in Mathematics each semester according to this table. Raj has scored the following results for four topics this semester, and has one topic to go. 75 68 85 79

- What is Raj's mean score so far?
- What grade will Raj get for the semester if his fifth score is:
 - 50?
 - 68?
 - 94?
- Find the maximum average score Raj can receive for the semester. Is it possible for him to get an A+?
- Find the least score that Raj needs in his fifth topic for him to receive an average of:
 - B+
 - A

Average score	Grade
90–100	A+
80–	A
70–	B+
60–	B
50–	C+
0–	C

10G Interpreting data from tables and graphs



In our everyday lives, it is important to be able to understand many forms of information. Data that is presented in a table or a graph is much easier to interpret than a long list of data. The headings in a table add detail and graphs give a visual comparison between values or categories.

Let's start: An unfair comparison

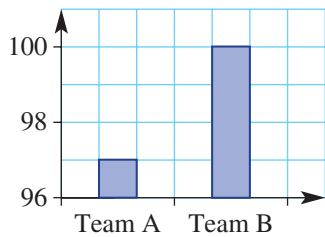
Each pair of graphs below shows the same data. Identify which one of the pair is misleading and discuss the reasons why.

Stage

5.3#
5.3
5.3\$
5.2
5.2∅
5.1
4

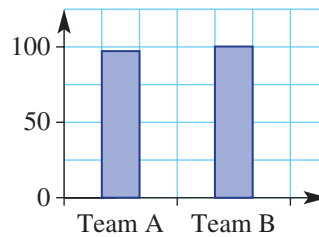
Graph A1

Scores



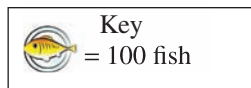
Graph A2

Scores



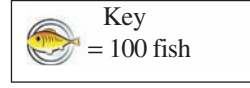
Graph B1

Fish caught



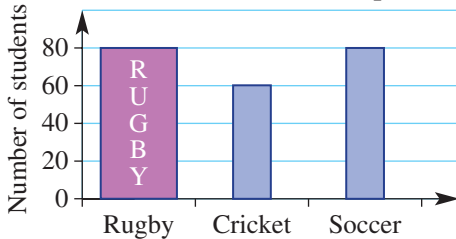
Graph B2

Fish caught



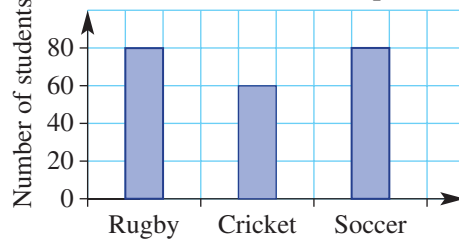
Graph C1

Student numbers for sport

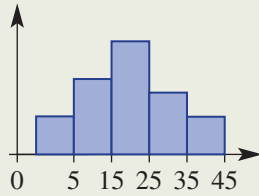
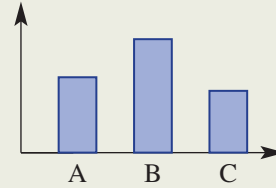
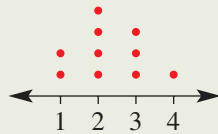
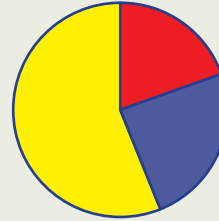


Graph C2

Student numbers for sport

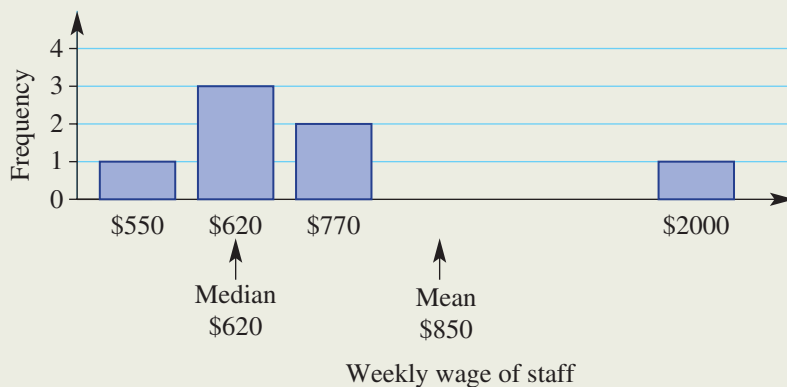


- Graphs can be presented in different forms. Some common examples are:

Histogram**Column graph****Dot plot****Pie chart**

- The titles, scales and column heights or sector angles tell us something about the data that is used to make the graphs.
- Misleading graphs give a false impression about data.
 - When only part of the scale is shown, the difference between results may be exaggerated.
 - If one column has a different size or colour, it may appear to have greater value than the other columns.
 - When pictograph symbols are not the same size, the larger symbol appears to have a higher value.
 - A measure of centre can be misleading if it is not a fair representation of the centre of the data values.

For example, for the data in the histogram below, the mean wage of \$850 is actually larger than all other wages except the outlier of \$2000. So the median (middle) wage of \$620 is a better representation of the 'average' wage in this case.

Weekly wages for a small business

Exercise 10G

UNDERSTANDING AND FLUENCY

1-6

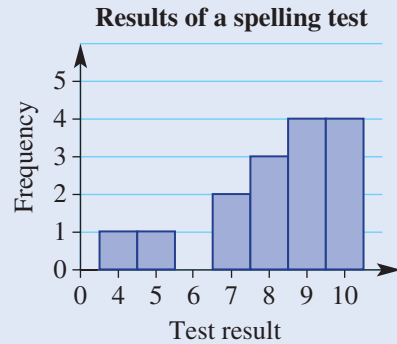
4-7



Example 15 Interpreting histograms

This histogram shows the results of a spelling test out of 10.

- How many results are shown in this histogram?
- List the results in ascending order.
- Calculate the mean.
- Calculate the median.
- What is the range of results from this test?
- Is this data skewed or symmetrical?
- What proportion of results is greater than or equal to 7?



SOLUTION

- 15
- 4, 5, 7, 7, 8, 8, 8, 9, 9, 9, 9, 10, 10, 10, 10
- Mean = $\frac{123}{15}$
= 8.2
- Median = 9
- Range = $10 - 4$
= 6
- Skewed
- $\frac{13}{15}$

EXPLANATION

The frequency shows how many times each score occurred.

Add the frequency values to find the total number of scores. $1 + 1 + 2 + 3 + 4 + 4 = 15$

1 lot of 4, 1 lot of 5, 2 lots of 7, 3 lots of 8, 4 lots of 9, 4 lots of 10.

$4 + 5 + 7 + 7 + 8 + 8 + 8 + 9 + 9 + 9 + 9 + 10 + 10 + 10 + 10 = 123$

The 8th score is the middle score.

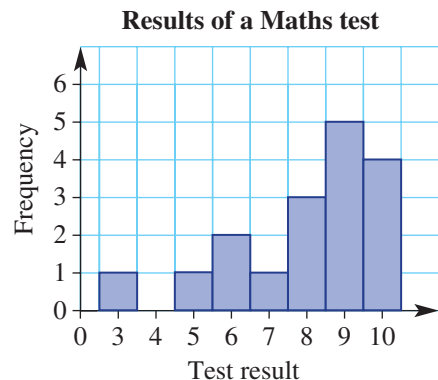
The range is the highest score minus the lowest score.

Results are bunched to the higher scores. The graph is not symmetrical.

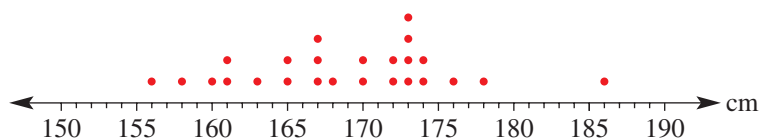
13 scores are greater than or equal to 7 out of a total of 15 scores.



- This histogram shows the results of a Maths test out of 10.
 - How many results are shown in this histogram?
 - List the results in ascending order.
 - Calculate the mean.
 - Calculate the median.
 - What is the range of results from this test?
 - Is this data skewed or symmetrical?
 - What proportion (fraction) of results is greater than or equal to 8?



- 2 This dot plot shows the heights, in cm, of the students in a Year 9 class.



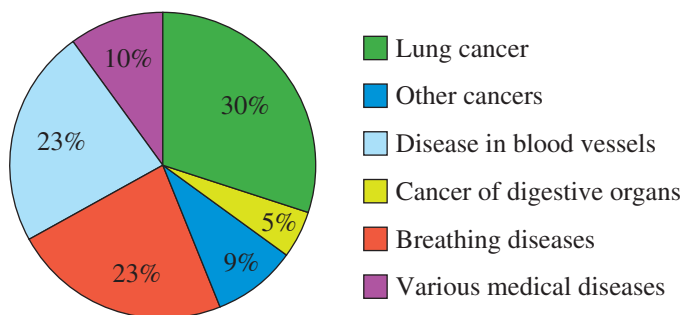
Each dot represents a student.



- How many students have their heights recorded on this dot plot?
- What is the range of heights?
- What is the mode of these heights?
- What is the median height for this class?
- What is the mean height for this class?
- What value is the outlier?

- 3 This pie chart shows the proportions of deaths in Australia from diseases caused by smoking.

Deaths from smoking-related diseases



Around 15000 Australians in total die from smoking-related illness per year.

- What is the total percentage of smoking deaths caused by cancer?
- How many Australian smokers die of smoking-related cancer in a year?
- How many Australian smokers die of smoking-related cancer each day? Round to the whole number.
- How many Australian smokers die from various breathing diseases each day? Round to one decimal place.
- Calculate the sector angle for lung cancer.

$$\text{Number of deaths} = \frac{\text{percentage}}{100} \times \text{total deaths}$$

$$\text{Sector angle} = \text{percentage} \times 360^\circ$$



- 4 The following table shows tide times and heights in December for Yamba, on the NSW north coast. The tide heights (in metres) shown are **red for low tide** and **blue for high tide**. The times are in 24-hour time.

1520 in 24-hour time is
3:20 p.m. in 12-hour time.



Saturday 17		Sunday 18		Monday 19		Tuesday 20		Wednesday 21	
Time	Height	Time	Height	Time	Height	Time	Height	Time	Height
0038	1.14	0145	1.18	0255	1.26	0401	1.37	0502	1.49
0619	0.45	0729	0.49	0847	0.50	1008	0.46	1123	0.38
1243	1.42	1340	1.34	1445	1.26	1556	1.21	1703	1.18
1925	0.29	2018	0.29	2115	0.28	2211	0.26	2307	0.22

- How high is the second high tide on Saturday 17 December?
- What time is the first low tide on Monday 19 December? Write your answer in both 24-hour time and 12-hour time.
- How much later in the morning is the low tide on Tuesday 20 December than the low tide on Monday 19 December?
- What is the difference in height between the two high tides on Wednesday 21 December?
- How long is it between the two high tides on Sunday 18 December?



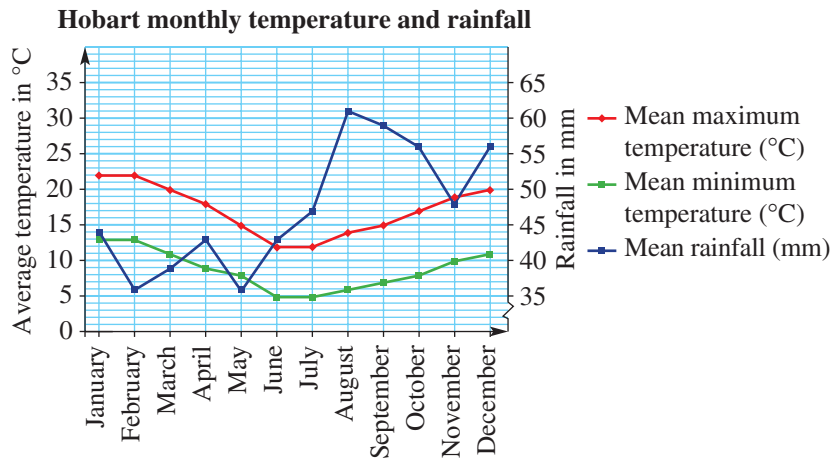
- 5 The following is a table of life expectancy estimates for people in various countries.

Country	Life expectancy in years		
	Overall average	Male	Female
Japan	83	79	86
Switzerland	82	80	84
Australia	81	79	84
Malaysia	74	72	77
Vietnam	74	72	77
Indonesia	71	69	73
Papua New Guinea	57	55	60
South Africa	49	49	50
Mozambique	39	38	39

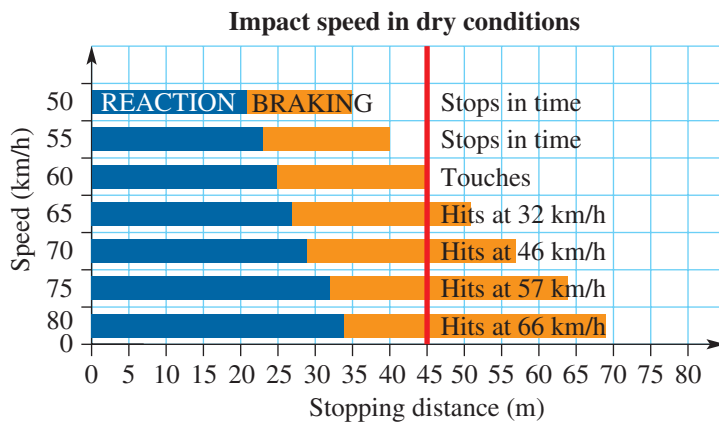
- How many years more do Australian males expect to live than Mozambiquan males?
- How many years more do Australian females expect to live than Papua New Guinean females?
- If the overall world average of life expectancy is 67 years, how far above the average is Japan's overall average life expectancy?
- Find the increase in life expectancy from Indonesian males to Australian males. Now calculate this increase as a percentage of Indonesian male life expectancy. Round to one decimal place.

- 6 The line graphs below show the mean monthly minimum and maximum temperatures and the monthly rainfall in mm for Hobart.

The temperature values are read from the scale on the left and the rainfall values are read from the scale on the right. For example, in January the mean maximum temperature is 22°C and the rainfall is 44 mm.



- What is the mean maximum temperature in April?
 - What is the rainfall in April?
 - What is the mean minimum temperature in July?
 - What is the rainfall in July?
 - During what months is the mean maximum temperature greater than 19°C ?
 - During what months is the rainfall less than 40 mm per month?
 - Which are the wettest two months in Hobart?
 - What is the temperature range in December?
- 7 The following graph shows stopping distances (reaction time and distance travelled while braking) for cars when driving at various speeds on a dry road.



Reaction time is shown by the blue bar, braking time is shown by the orange bar.



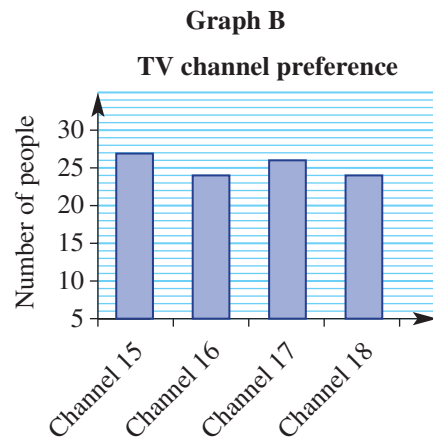
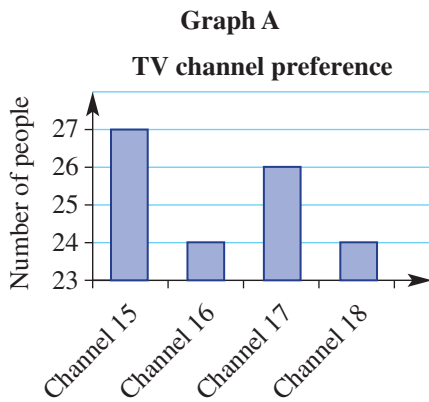
- The red line on this graph represents an object or person in front of the braking car. How far in front of the car is the object or person at the start of the driver's reaction time?
- How many metres are travelled during *reaction time* when driving at 60 km/h?
- How many metres are travelled during *braking time* when driving at 60 km/h?
- How much distance is needed, overall, to stop when driving at 80 km/h?
- By how much does the braking distance increase when driving at 80 km/h compared to 50 km/h?

PROBLEM SOLVING AND REASONING

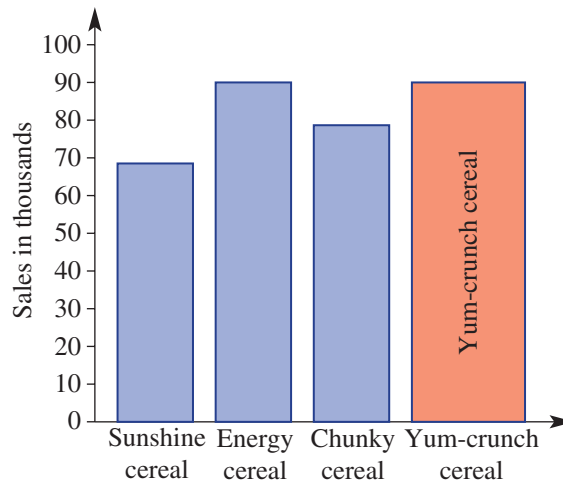
8, 9

8–11

- 8 Here are two column graphs, each showing the same results of a survey that asked people which TV channel they preferred.

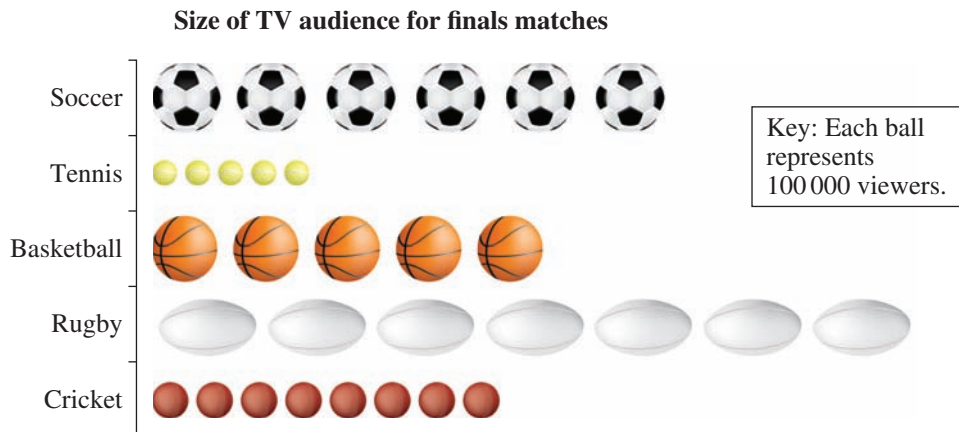


- From graph A, write down how many viewers preferred each channel.
 - Which graph could be titled 'Channel 15 is clearly most popular'?
 - Which graph could be titled 'All TV channels have similar popularity'?
 - What is the difference between the two graphs?
 - Which graph is misleading, and why?
- 9 This column graph shows the number of sales for one month of four popular breakfast cereals. The height of each column represents the sales.



- List the breakfast cereals in order of sales.
- Did Yum-crunch have more sales than the other cereals?
- List three changes that have been made to the Yum-crunch column to make it look as though it has better sales than any of the other cereals.
- How should the columns be drawn so the graph is not misleading?

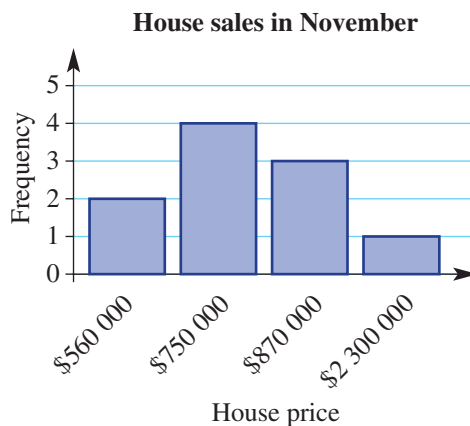
- 10 This pictograph shows the number of TV viewers for the final matches of various sports.



- Which sport *appears* to have had the greatest TV audience?
- List the sports in order, according to their *length* on the pictograph.
- Using the key, determine the audience sizes for the rugby and soccer finals.
- Which sport actually had the largest TV audience? What was the size of its audience?
- List the sports in order according to the audience size calculated by using the key.
- In what way is this graph misleading?
- How should a pictograph be drawn so it is not misleading?



- 11 This histogram shows the price of some houses that were sold during one November in a Sydney suburb.



- How many house sales are shown in this histogram?
- List the price of each house in ascending order.
- Find the mean house price.
- Find the median house price.
- A newspaper headline read 'House prices now average over \$900 000'. Is this 'average' referring to the mean or median price?
- What proportion of house prices were less than the mean value?
- Do you think the mean or the median would better represent the 'average' house price? Give a reason for your answer.

ENRICHMENT

–

12

Viewing distance for TV

- 12 a** Draw a line graph showing the minimum and maximum viewing distances for each TV screen size. Include a key.

TV screen size (length of diagonal in inches)	Maximum viewing distance (cm)	Minimum viewing distance (cm)
26	200	100
30	220	120
34	260	130
42	320	160
47	360	180
50	380	200
55	420	220
60	470	230
65	490	250

- b** If a lounge chair is 1.5 m in front of the TV, which sizes of TV would be suitable?
c If a lounge chair is 3 m in front of the TV, which sizes of TV would be suitable?



10H Stem-and-leaf plots



Interactive



Widgets



HOTsheets



Walkthrough

Stem-and-leaf plots (or stem plots) are commonly used to display a single data set or two related data sets. They help to show how the data is distributed. They retain all the individual data elements so no detail is lost. The median and mode can be easily read from a stem-and-leaf plot because all the data sits in order.



Stage

5.3#

5.3

5.3\$

5.2

5.2∅

5.1

4

Let's start: Ships vs Chops

At a school, Ms Ships' class and Mr Chops' class sit the same exam. The scores are displayed using this back-to-back stem-and-leaf plot. Discuss the following.

- Which class had the most students?
- What were the lowest and highest scores from each class?
- What were the median scores from each class?
- Which class could be described as symmetrical and which as skewed?
- Which class had the better results?

For the data below, digits representing the stem give the tens, and digits representing the leaves give the units.

Leaf Ms Ships' class	Stem	Leaf Mr Chops' class
3 1	5	0 1 1 3 5 7
8 8 7 5	6	2 3 5 5 7 9 9
6 4 4 2 1	7	8 9 9
7 4 3	8	0 3
6	9	1

7|8 means 78

Key ideas

- A **stem-and-leaf plot** is a way to display numerical data.
- Each number is split into a stem (the first digit or digits) and a leaf (the last digit).

For example:

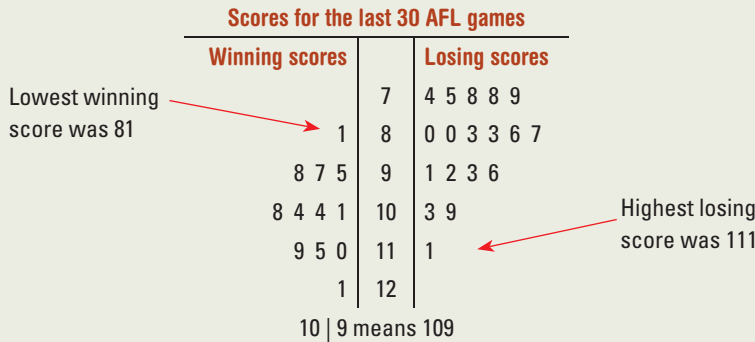
	Stem	Leaf
The number 7 is	0	7
The number 31 is	3	1
The number 152 is	15	2

- Leaves are aligned vertically, getting bigger as you move away from the stem.

Stem-and-leaf plot

A graph that lists numbers in order, grouped in rows

- **Back-to-back stem-and-leaf plots** can be used to compare two sets of data. The stem is drawn in the middle, with the leaves on either side.

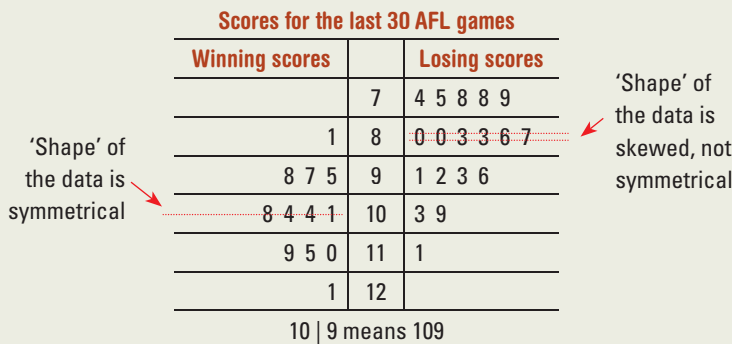


Back-to-back stem-and-leaf plot A visual representation of two data sets

Symmetrical data A distribution of data which is symmetrical either side of the mean and the median

Skewed data A distribution of data which is sloped more to one side than the other. The mean and the median have different values

- **Symmetrical data** will produce a graph that is symmetrical about the centre. The ‘winning scores’ are symmetrical.
- **Skewed data** will produce a graph that lacks symmetry. The ‘losing scores’ are skewed.



Exercise 10H

UNDERSTANDING AND FLUENCY

1–7

4–8

- 1 List the numbers in these stem-and-leaf plots, following the key.

a

Stem	Leaf
3	5 7
4	1 3 8

3 | 1 means 3.1

The same stem goes with each leaf along each row.



b

Stem	Leaf
5	2
6	0 1 7
7	3 5

6 | 3 means 63

- 2 a Rewrite this stem-and-leaf plot with the leaves in order from smallest to largest.

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

- b If the stems are units and the leaves are tenths, write a possible key.

In part b, the key must include a decimal point.



- 3 This stem-and-leaf plot shows the number of minutes Alexis spoke on her phone for a number of calls.

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

2|1 means 21 minutes

Count the leaves to find the number of scores.



- a How many calls are represented by the stem and leaf plot?
 b What is the length of the:
 i shortest phone call?
 ii longest phone call?
 c What is the mode (the most common call time)?
 d What is the median call time (middle value)?
- 4 This back-to-back stem-and-leaf plot shows the thickness of tyre tread on a selection of cars from the city and country.

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

1|3 means 13 mm

06 is just written as 6.



- a How many car tyres were tested altogether?
 b What was the smallest tyre tread thickness in the:
 i city?
 ii country?
 c What was the largest tyre tread thickness in the:
 i city?
 ii country?
 d Find the median tyre tread thickness for tyres in the:
 i city
 ii country
 e Is the distribution of tread thickness for city cars more symmetrical or skewed?
 f Is the distribution of tread thickness for country cars more symmetrical or skewed?

Skewed distributions have only a few scores at one end.





Example 16 Constructing and using a stem-and-leaf plot

Consider the following set of data.

0.3 2.5 4.1 3.7 2.0 3.3 4.8 3.3 4.6 0.1 4.1 7.5 1.4 2.4
5.7 2.3 3.4 3.0 2.3 4.1 6.3 1.0 5.8 4.4 0.1 6.8 5.2 1.0

- Organise the data into an ordered stem-and-leaf plot.
- Find the median.
- Find the mode.
- Describe the symmetry/skewness of the data.

SOLUTION

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

3|4 means 3.4

- Median = $\frac{3.3 + 3.4}{2}$
= 3.35
- Mode is 4.1.
- Data is approximately symmetrical.

EXPLANATION

The minimum is 0.1 and the maximum is 7.5, so stems range from 0 to 7.

Place leaves in order from smallest to largest. Some numbers appear more than once; e.g. two instances of 0.1 means that the leaf 1 appears twice: 0.1, 0.1, ...

There are 28 data values. The median is the average of the two middle values (the 14th and 15th values).

The most common value is 4.1.

The distribution of numbers is approximately symmetrical about the stem containing the median.

- For each of the following data sets:
 - organise the data into an ordered stem-and-leaf plot
 - find the median
 - find the mode
 - describe the symmetry/skewness of the data.

a 41 33 28 24 19 32 54 35 26 28 19 23 32 26 28

b 31 33 23 35 15 23 48 50 35 42 45 15 21 45
51 31 34 23 42 50 26 30 45 37 39

c 34.5 34.9 33.7 34.5 35.8 33.8 34.3 35.2 37.0 34.7
35.2 34.4 35.5 36.5 36.1 33.3 35.4 32.0 36.3 34.8

d 167 159 159 193 161 164 167 157 158 175 177 185
177 202 185 187 159 189 167 159 173 198 200

Draft version: Enter the leaves in the same order as in the question.

Final version: Re-write the stem-and-leaf plot with leaves in order. Remember to include the key.



In part **d**, use 16 as the stem for 167.



- 6 The number of vacant rooms in a motel each week over a 20-week period is shown below.
12 8 11 10 21 12 6 11 12 16 14 22 5 15 20 6 17 8 14 9
- Draw a stem-and-leaf plot of this data.
 - In how many weeks were there fewer than 12 vacant rooms?
 - Find the median number of vacant rooms.



Example 17 Constructing back-to-back stem-and-leaf plots

A shop owner has two jeans shops. The daily sales in each shop over a 16-day period are monitored and recorded as follows.

Shop A

3 12 12 13 14 14 15 15 21 22 24 24 24 26 27 28

Shop B

4 6 6 7 7 8 9 9 10 12 13 14 14 16 17 27

- Draw a back-to-back stem-and-leaf plot.
- Compare and comment on differences between the sales made by the two shops.

SOLUTION

a

Shop A		Shop B
3	0	4 6 6 7 7 8 9 9
5 5 4 4 3 2 2	1	0 2 3 4 4 6 7
8 7 6 4 4 4 2 1	2	7

1|3 means 13

- b Shop A sales are generally between 12 and 28, with one low value of 3. Shop B sales are generally between 4 and 17, with one high value of 27. Shop A has a lot more high values than shop B. Shop B has more low values than shop A.

EXPLANATION

The data for each shop is already ordered. Stems are in intervals of 10. Record leaf digits for shop A on the left and shop B on the right, ordered from the middle to the outside.

Look at both sides of the plot for the highest and lowest values and whether there are a few or many of the small and large numbers.

- 7 For each of the following sets of data:
- draw a back-to-back stem-and-leaf plot
 - state the smallest and largest value in each set and compare the numbers of small and large values in each set
- Set A: 46 32 40 43 45 47 53 54 40 54 33 48 39 43
Set B: 48 49 31 40 43 47 48 41 49 51 44 46 53 44
 - Set A: 1 43 24 26 48 50 2 2 36 11 16 37 41 3 36
6 8 9 10 17 22 10 11 17 29 30 35 4 23 23
Set B: 9 18 19 19 20 21 23 24 27 28 31 37 37 38 39 39 39
40 41 41 43 44 44 45 47 50 50 51 53 53 54 54 55 56
 - Set A: 0.7 0.8 1.4 8.8 9.1 2.6 3.2 0.3 1.7 1.9 2.5 4.1 4.3 3.3 3.4
3.6 3.9 3.9 4.7 1.6 0.4 5.3 5.7 2.1 2.3 1.9 5.2 6.1 6.2 8.3
Set B: 0.1 0.9 0.6 1.3 0.9 0.1 0.3 2.5 0.6 3.4 4.8 5.2 8.8 4.7 5.3
2.6 1.5 1.8 3.9 1.9 0.1 0.2 1.2 3.3 2.1 4.3 5.7 6.1 6.2 8.3

Order the leaves with the smallest on the inside and largest on the outside. State whether each set has a few or many of the small numbers and large numbers.



- 8 Draw a back-to-back stem-and-leaf plot for the final scores of St Kilda and Collingwood in the 24 AFL games given here.

St Kilda: 126 68 78 90 87 118
 88 125 111 117 82 82
 80 66 84 138 109 113
 122 80 94 83 106 68
 Collingwood: 104 80 127 88 103 95
 78 118 89 82 103 115
 98 77 119 91 71 70
 63 89 103 97 72 68

- a In what percentage of their games did each team score more than 100 points?
 b Comment on the symmetry of the distribution of the scores for each team.
 c Which team has scores that are more consistent?
 Which team has more higher scores?

$$\text{Percentage} = \text{fraction} \times 100$$



PROBLEM SOLVING AND REASONING

9, 10

9, 11

- 9 This stem-and-leaf plot shows the time taken, in seconds, by Helena to run 100 m in her last 25 races.
- a Find Helena's median time.
 b What is the difference between the slowest and fastest time?
 c If in her 26th race, her time was 14.8 seconds and this was added to the stem-and-leaf plot, would her median time change? If so, by how much?

Stem	Leaf
14	9
15	4 5 6 6 7 7 7 8 9
16	0 0 1 1 2 2 3 4 4 5 5 7 7
17	2

14|9 means 14, 9 seconds

- 10 Two brands of batteries were tested to determine their lifetime in hours. The data below shows the lifetime of 20 batteries of each brand.

Brand A: 7.3 8.2 8.4 8.5 8.7 8.8 8.9 9.0 9.1 9.2
 9.3 9.4 9.4 9.5 9.5 9.6 9.7 9.8 9.9 9.9

Brand B: 7.2 7.3 7.4 7.5 7.6 7.8 7.9 7.9 8.0 8.1
 8.3 9.0 9.1 9.2 9.3 9.4 9.5 9.6 9.8 9.8

The stems will be the 'units' value and each leaf the 'tenths' value.



- a Draw a back-to-back stem-and-leaf plot for this data.
 b How many batteries from each brand lasted more than 9 hours?
 c Which brand shows the best performance?

- 11 Find the median if all the data in each back-to-back stem-and-leaf plot was combined.

a

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

4|2 means 42

First combine each back-to-back plot into just one stem-and-leaf plot.



b

3	16	0 3 3 6 7 9
9 6 6 1	17	0 1 1 4 8 8
8 7 5 5 4 0	18	2 2 6 7
2	19	0 1

16|3 means 16.3

ENRICHMENT

–

12

Birth weights and smoking

- 12 This back-to-back stem-and-leaf plot below shows the birth weight in kilograms of babies born to mothers who do or don't smoke.

Birth weight of babies

Smoking mothers		Non-smoking mothers
4 3 2 2	2	4
9 9 8 7 6 6 5 5	2*	8 9
4 3 2 1 1 1 0 0 0	3	0 0 1 2 2 3
6 5 5	3*	5 5 5 6 6 7 7 8
1	4	
	4*	5 5 6

2|4 means 2.4 kg
2*|5 means 2.5 kg

- a** What percentage of babies born to smoking mothers have a birth weight of less than 3 kg?
- b** What percentage of babies born to non-smoking mothers have a birth weight of less than 3 kg?
- c** Compare and comment on the differences between the birth weights of babies born to mothers who smoke and those born to mothers who don't smoke.

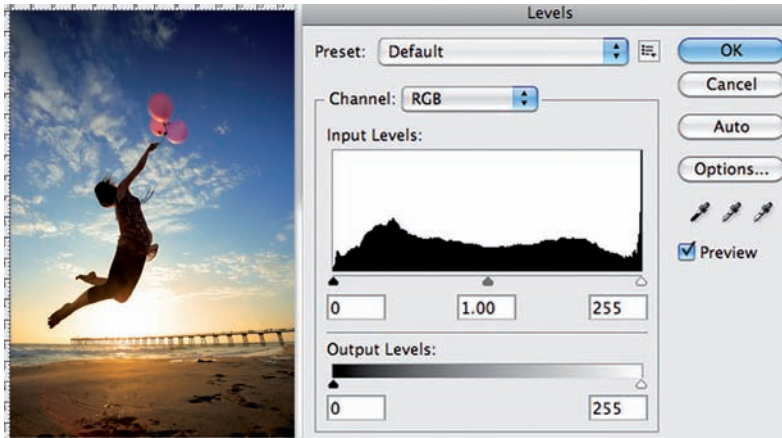


Do babies born to mothers who smoke weigh more or less than those born to mothers who don't smoke?

101 Grouping data into classes FRINGE



For some data, especially large sets, it makes sense to group the data and then record the frequency for each group to produce a frequency table. For numerical data, a graph generated from a frequency table gives a histogram. Like a stem-and-leaf plot, a histogram shows how the data is distributed across the full range of values. A histogram, for example, is used to display the level of exposure of the pixels in an image in digital photography. It uses many narrow columns to show how the light values are distributed across the scale from black to white.

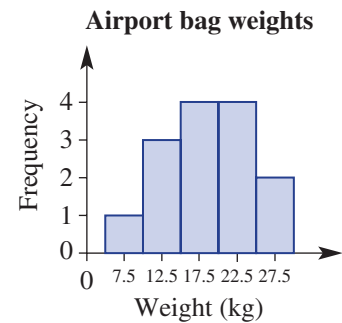


A luminance value histogram used in digital photography software

Let's start: Baggage check

This histogram shows the distribution of the weight of a number of bags checked at an airport.

- How many bags had a weight in the range 10–15 kg?
- How many bags were checked in total?
- Is it possible to determine the exact mean, median or mode of the weight of the bags by looking at the histogram? Discuss.
- Describe the distribution of checked bag weights for the given graph.



- A **frequency table** shows the number of values within a set of categories or **class intervals**.
- Grouped numerical data can be illustrated using a **histogram**.
 - The height of a column corresponds to the frequency of values in that class interval.
 - There are usually no gaps between columns.
 - The scales are evenly spread with each bar spreading across the boundaries of the class interval.
 - A **percentage frequency histogram** shows the frequencies as percentages of the total.
- Like a stem-and-leaf plot, a histogram can show if the data is **skewed** or **symmetrical**.

Frequency table A table summarising data by showing all possible scores from lowest to highest in one column, and the frequency of each score in another column

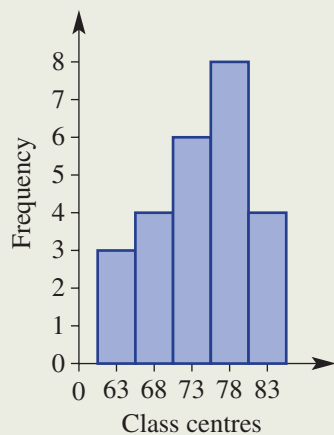
Class interval The name for a range of values that belong in a group of data

Histogram A special type of column graph with no gaps between the columns

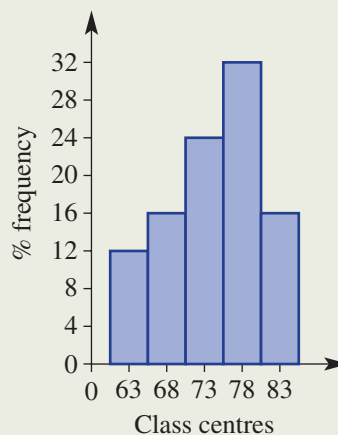
Frequency table

Class	Class centre	Frequency	Percentage frequency
61–65	63	3	12
66–70	68	4	16
71–75	73	6	24
76–80	78	8	32
81–85	83	4	16
	Total	25	100

Frequency histogram



Percentage frequency histogram



Exercise 101 FRINGE

UNDERSTANDING AND FLUENCY

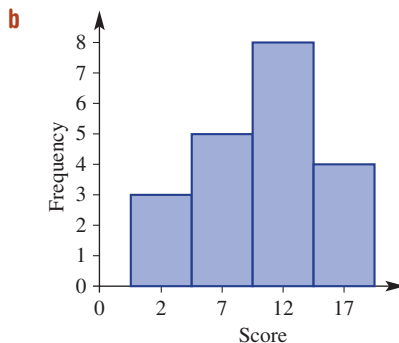
1–7

4–8

- Write the missing word for each of these statements.
 - A group of scores such as 10–15 kg is called a _____.
 - A _____ shows the number of values in each class interval.
 - A column graph that shows frequencies and scores is called a _____.
 - There are _____ between the columns in a histogram.
 - A _____ histogram shows each frequency as a percentage of the total number of scores.
- Write the total number of scores in each of these displays.

a

Class interval	Class centre	Frequency
0–4	2	3
5–9	7	5
10–14	12	8
15–19	17	4

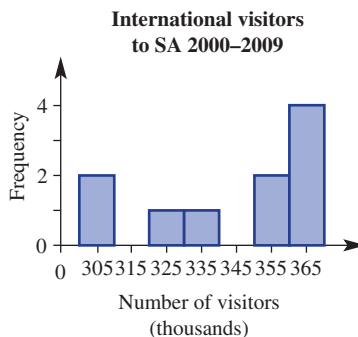


The sum of the frequencies equals the number of scores.



- This frequency histogram shows the number of years for which the number of international visitors to South Australia was within a given range for the decade from 2000 to 2009.

- How many years in the decade were there fewer than 330 000 international visitors?
- Which range of visitor numbers had the highest frequency?

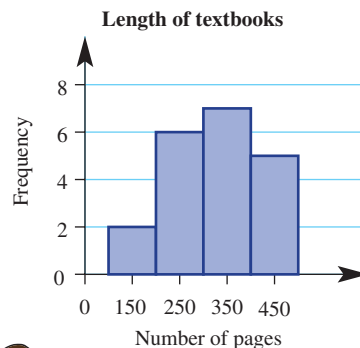


The frequency shows the number of years.



- Some Year 9 students selected a sample of textbooks from the library and recorded the number of pages in each book. The frequency histogram on the right shows their results.

- How many textbooks had between 100 and 200 pages?
- How many textbooks were selected from the library?
- What percentage of textbooks had between:
 - 200 and 300 pages?
 - 200 and 400 pages?



The frequency shows the number of textbooks.

$$\text{Percentage} = \frac{\text{frequency}}{\text{total}} \times 100$$





Example 18 Constructing frequency tables and histograms

The data below shows the number of hamburgers sold each hour by a 24-hour fast-food store during a 50-hour period.

1 10 18 14 20 11 19 10 17 21
 5 16 7 15 21 15 10 22 11 18
 12 12 3 12 8 12 6 5 14 14
 14 4 9 15 17 19 6 24 16 17
 14 11 17 18 19 19 19 18 18 20

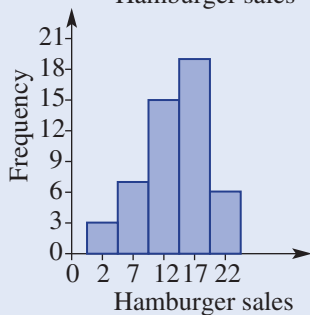
- a** Set up and complete a grouped frequency table, using class intervals 0–4, 5–9 etc. Include a percentage frequency column.
- b** Construct a frequency histogram.
- c** For how many hours did the fast-food store sell:
- fewer than 10 hamburgers?
 - at least 15 hamburgers?

SOLUTION

a

Class interval	Class centre	Frequency	Percentage frequency
0–4	2	3	6
5–9	7	7	14
10–14	12	15	30
15–19	17	19	38
20–24	22	6	12
Total		50	100

- b** Hamburger sales



- c i** $3 + 7 = 10$ hours
- ii** $19 + 6 = 25$ hours

EXPLANATION

Create class intervals of 5 from 0 up to 25, since 24 is the maximum number. Record the number of data values in each interval in the frequency column. Convert to a percentage by dividing by the total (50) and multiplying by 100.

Create a frequency histogram with frequency on the vertical axis and the class intervals on the horizontal axis. The height of the column shows the frequency of that interval.

Fewer than 10 hamburgers covers the 0–4 and 5–9 intervals.

At least 15 hamburgers covers the 15–19 and 20–24 intervals.

- 5 The data below shows the number of ice-creams sold from an ice-cream van over a 50-day period.

0 5 0 35 14 15 18 21 21 36 45 2 8
 2 2 3 17 3 7 28 35 7 21 3 46 47
 1 1 3 9 35 22 7 18 36 3 9 2
 11 37 37 45 11 12 14 17 22 1 2 2

- a** Set up and complete a grouped frequency table using class intervals 0–9, 10–19, 20–29 etc. Include a percentage frequency column.
b Construct a frequency histogram.
c How many days did the ice-cream van sell:
 i fewer than 20 ice-creams?
 ii at least 30 ice-creams?
d What percentage of days were 20 or more ice-creams sold?

'Fewer than 20'
doesn't include
20. 'At least 30'
includes 30.



- 6 The data below shows the mark out of 100 on the Science exam for 60 Year 9 students.

50 67 68 89 82 81 50 50 89 52 60 82 52 60 87
 89 71 73 75 83 86 50 52 71 80 95 87 87 87 74
 60 60 61 63 63 65 82 86 96 88 50 94 87 64 64
 72 71 72 88 86 89 69 71 80 89 92 89 89 60 83

- a** Set up and complete a grouped frequency table, using class intervals 50–59, 60–69, 70–79 etc. Include a percentage frequency column.
b Construct a frequency histogram.
c i How many marks were less than 70 out of 100?
 ii What percentage of marks were at least 80 out of 100?

The frequency for
'50–59' will be the
number of scores from
50–59.



- 7 The number of goals kicked by a country footballer in each of his last 30 football matches is given below.

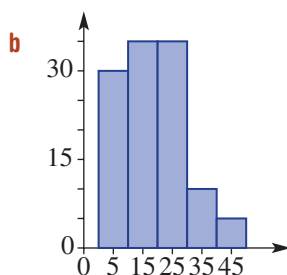
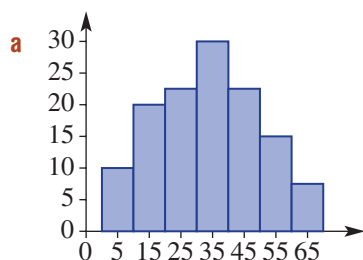
8 9 3 6 12 14 8 3 4 5 2 5 6 4 13
 8 9 12 11 7 12 14 10 9 8 12 10 11 4 5

- a** Organise the data into a grouped frequency table using class intervals of 0–2, 3–5 etc.
b Draw a frequency histogram for the data.
c In how many games did the player kick fewer than six goals?
d In how many games did he kick more than 11 goals?

The frequency for '0–2' will be
the number of scores that are
0, 1 or 2. Label the horizontal
scale 1, 4, 7, 10 etc.



- 8 Which one of these histograms illustrates a symmetrical data set and which one shows a skewed data set?



Skewed data is bunched
either below or above the
middle.



PROBLEM-SOLVING AND REASONING

9, 10

9, 11, 12

9 Write down the missing numbers in these frequency tables; i.e. find the values of the pronumerals.

a

Class interval	Class centre	Frequency	Percentage frequency
0–4	2	1	10
5–9	7	3	c
10–14	12	4	d
a –19	17	b	e
Total		10	f

b

Class interval	Class centre	Frequency	Percentage frequency
40–49	44.5	20	20
a –59	54.5	28	b
60–69	64.5	12	c
70–79	74.5	d	40
Total		100	e



10 The data below shows the length of overseas phone calls (in minutes) made by a particular household over a six-week period.

1.5 1 1.5 1 4.8 4 4 10.1 9.5 1 3
 8 5.9 6 6.4 7 3.5 3.1 3.6 3 4.2 4.3
 4 12.5 10.2 10.3 4.5 4.5 3.4 3.5 3.5 5 3.5
 3.6 4.5 4.5 12 11 12 14 14 12 13 10.8
 12.1 2.4 3.8 4.2 5.6 10.8 11.2 9.3 9.2 8.7 8.5

What percentage of phone calls were more than 3 minutes in length? Answer to one decimal place.

How many calls were 3 minutes or less?



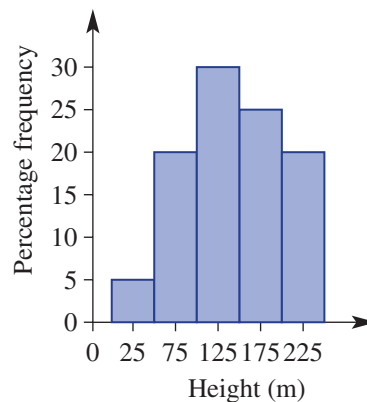
$$\text{Percentage} = \frac{\text{frequency}}{\text{total}} \times 100$$



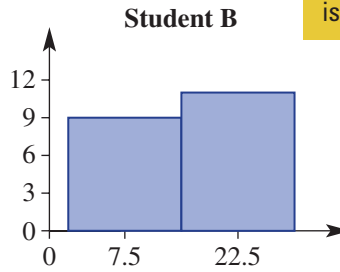
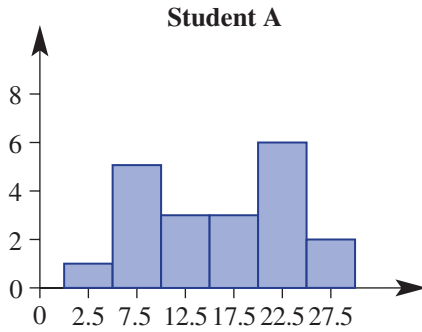
11 This percentage frequency histogram shows the heights of office towers in a city.

- a What percentage of office towers have the following heights?
- Between 50 m and 100 m
 - Less than 150 m
 - No more than 200 m
 - At least 100 m
 - Between 100 m and 150 m or greater than 200 m
- b If the city had 100 office towers, how many would have a height of:
- between 100 m and 150 m?
 - at least 150 m?
- c If the city had 40 office towers, how many would have a height of:
- between 0 m and 50 m?
 - no more than 150 m?

In parts **b** and **c**,
 number of office towers
 = percentage \times total



- 12 Two students show different histograms for the same set of data.



A histogram that shows a lot of detail is more useful.



- a Which histogram is more useful in helping to analyse the data?
 b What would you advise student B to do differently when constructing the histogram?

ENRICHMENT

-

13

The distribution of weekly wages

- 13 The data below shows the weekly wages (in dollars) of 50 people.

400 500 552 455 420 424 325 204 860 894 464 379 563 230
 940 384 370 356 345 380 720 540 654 678 628 656 670 725
 740 750 730 766 760 700 700 768 608 576 890 920 874 860
 450 674 725 612 605 600 548 670

- a What is the minimum weekly wage and the maximum weekly wage?
 b i Organise the data into about 10 class intervals: 200–299, 300–399 etc.
 ii Draw a frequency histogram for the data.
 c i Organise the data into about four class intervals.
 ii Draw a frequency histogram for the data.
 d Discuss the shapes of the two graphs. Which graph represents the data better and why?



- 1 One coin says to another coin: *How does my face appear?*

Answer these questions for each scenario, then match the letter in bold to the answers listed below to find the answer to the riddle.

2 coins are tossed 12 times

$$P(HH) = \mathbf{D}$$

$$P(HT \text{ or } TH) = \mathbf{I}$$

$$\text{Expected number of TT} = \mathbf{R}$$

$$\text{Expected number of one or more tails} = \mathbf{T}$$

1 die is tossed 12 times

$$P(5) = \mathbf{A}$$

$$P(\text{greater than } 2) = \mathbf{N}$$

$$\text{Expected number of '6's} = \mathbf{S}$$

Choosing one letter from the word PUZZLES

$$P(Z) = \mathbf{M}$$

$$P(\text{consonant}) = \mathbf{O}$$

		,							
$\frac{1}{2}$	9	2	3	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{4}$	$\frac{5}{7}$	$\frac{2}{7}$	

- 2 Match each description (i–v) with the most suitable diagram (A–E). Note: There is only one description for each diagram.

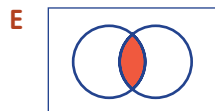
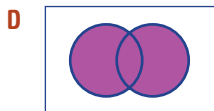
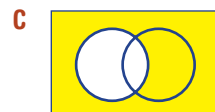
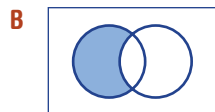
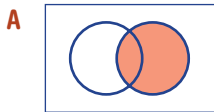
i Don't own an iPhone

ii Own an iPhone only

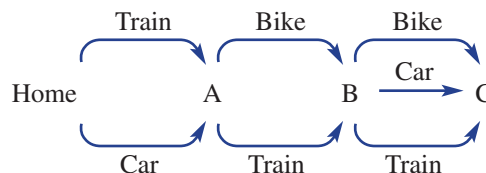
iii Own both an iPad and an iPhone

iv Own an iPad or an iPhone or both

v Own an iPad



- 3 In a class of 28, each student owns a cat or a dog or both. If 18 students own cats and 16 students own dogs, how many students own both a cat and a dog?
- 4 Write down the set of five positive integers that has a mean of 5, a mode of 8 and a range of 6.
- 5 A 6-sided die and a 10-sided die are rolled simultaneously. What total sums have the highest chance of occurring? (A total sum means the sum of the two uppermost faces.)
- 6 Michael needs to deliver parcels to three places (A, B and C, in order) in the city. This diagram shows the different ways that he can travel.



- a Draw a tree diagram showing all the possible combinations of transportation.
- b What is the total number of possible outcomes?
- c Find the probability that Michael will use a different method of transport each time.

Two-step experiments

These can be represented by tables or tree diagrams. They involve more than one component and can occur with or without replacement.

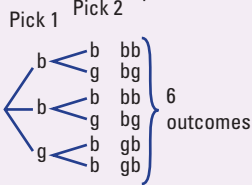
e.g. a bag contains 2 blue counters and 1 green, 2 are selected at random. (I) Table with replacement.

	Pick 1			
	b	b	g	
Pick 2	b	bb	bb	gb
	b	bb	bb	gb
	g	bg	bg	gg

Sample space of 9 outcomes

$P(2 \text{ of same colour}) = \frac{5}{9}$

(II) Tree without replacement



$P(2 \text{ of same colour}) = \frac{2}{6} = \frac{1}{3}$

Probability review

The sample space is the list of all possible outcomes of an experiment.

For all possible outcomes:

$$P(\text{event}) = \frac{\text{number of outcomes where event occurs}}{\text{total number of outcomes}}$$

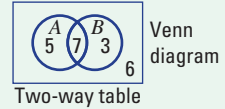
e.g. roll a normal six-sided die

$P(>4) = \frac{2}{6} = \frac{1}{3}$

$P(\text{not } A) = 1 - P(A)$

Venn diagrams and two-way tables

These organise data from two or more categories



Two-way table

	A	Not A	Totals
B	7	3	10
Not B	5	6	11
Totals	12	9	21

i.e. 6 in neither category, 7 in both categories

$P(A) = \frac{12}{21} = \frac{4}{7}$ $P(A \text{ and } B) = \frac{7}{21} = \frac{1}{3}$

$P(B \text{ only}) = \frac{3}{21} = \frac{1}{7}$ $P(A \text{ or } B \text{ or both}) = \frac{15}{21} = \frac{5}{7}$

Probability and single variable data analysis

Experimental probability

This is calculated from results of experiment or survey.

Experimental probability = $\frac{\text{number of times event occurs}}{\text{total number of trials}}$

Expected number = probability \times number of occurrences of trials

Measure of spread

Range = maximum value – minimum value

Stem-and-leaf plots

These display all the data values using a stem and a leaf. An ordered back-to-back stem-and-leaf plot.

	Leaf	Stem	Leaf	
skewed	9 8 7 2	1	0 3	symmetrical
	7 4 3 3	2	2 2 4	
	5 2 1	3	3 6 7 8	
	7	4	4 5 9	
	0	5	0	
	3	5	5 means 35	
			key	

Summarising data: Measures of centre

Mode is the most common value (there can be more than one). Two modes means data is bimodal.

Mean is the average of the data values.

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

Median is the middle value of data that is ordered.

Odd data set	Even data set
2 4 7 10 12	2 4 6 10 15 18
Median	8
	Median = $\frac{6+10}{2}$

An outlier is a value that is not near the rest of the data.

Grouped data

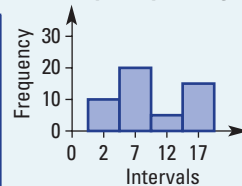
Data values can be grouped into class intervals, e.g. 0–4, 5–9, etc. and recorded in a frequency table.

The frequency or percentage frequency of each interval can be recorded in a histogram.

Frequency table

Class interval	Class centre	Frequency	Percentage frequency
0–4	2	10	20
5–9	7	20	40
10–14	12	5	10
15–19	17	15	30
Total		50	100

Frequency histogram



Multiple-choice questions

1 A letter is randomly chosen from the word XYLOPHONE. The probability that it is an O is:

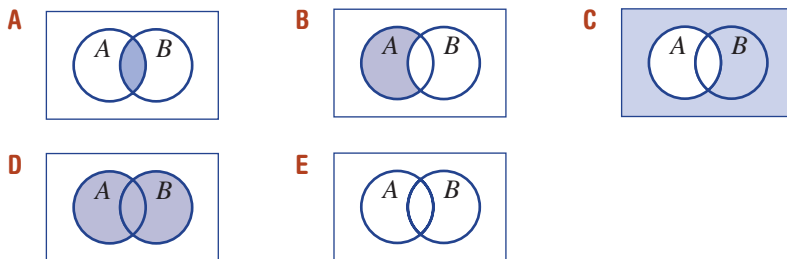
- A $\frac{1}{8}$ B $\frac{2}{9}$ C $\frac{1}{4}$ D $\frac{1}{9}$ E $\frac{1}{3}$

2 The values of x and y in the two-way table are:

- A $x = 12, y = 8$
 B $x = 12, y = 11$
 C $x = 16, y = 4$
 D $x = 10, y = 1$
 E $x = 14, y = 6$

	A	Not A	Total
B		5	9
Not B	8	y	
Total	x		25

- 3 i Which shaded region represents both A and B ?
 ii Which shaded region represents A only?
 iii Which shaded region represents A or B or both?
 iv Which shaded region represents the complement of A ?



4 A bag contains 2 green balls and 1 red ball. Two balls are randomly selected *without replacement*. Using a tree diagram, the probability of selecting one of each colour is:

- A $\frac{1}{2}$ B $\frac{2}{3}$ C $\frac{5}{6}$ D $\frac{1}{3}$ E $\frac{3}{4}$

5 From rolling a biased die, a class finds an experimental probability of 0.3 of rolling a 5. From 500 rolls of the die, the expected number of 5s would be:

- A 300 B 167 C 180 D 150 E 210

6 a The median of the data in this stem-and-leaf plot is:

- A 74 B 71 C 86
 D 65 E 70

b The range of the data in the stem-and-leaf plot is:

- A 3 B 8 C 33
 D 86 E 14

Stem	Leaf
5	3 5 8
6	1 4 7
7	0 2 4 7 9
8	2 6 6
7 4	means 74

7 If Jacob achieved scores of 12, 9, 7 and 12 on his last four language vocabulary tests, what score must he get on the fifth test to have a mean of 11?

- A 16 B 14 C 11 D 13 E 15

Short-answer questions

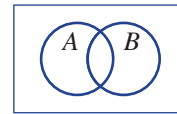
- 1 Determine the probability of each of the following.
- Rolling more than 2 on a normal six-sided die
 - Selecting a vowel from the word EDUCATION
 - Selecting a pink or white jelly bean from a packet containing 4 pink, 2 white and 4 black jelly beans
- 2 From a survey of 50 people, 30 have the newspaper delivered, 25 read it online, 10 do both and 5 do neither.
- Construct a Venn diagram for the survey results.
 - How many people only read the newspaper online?
 - If one of the 50 people were randomly selected, find:
 - $P(\text{have paper delivered and read it online})$
 - $P(\text{don't have it delivered})$
 - $P(\text{only read it online})$



- 3 a Copy and complete this two-way table.

	A	Not A	Total
B		16	
Not B	8		20
Total	17		

- b Convert the information into a Venn diagram, as shown.



- c Find the following:

i $P(\text{not } B)$

ii $P(\text{both } A \text{ and } B)$

iii number of 'A only'

iv number in either A or B or both A and B

- 4 A spinner with equal areas of red, green and blue is spun and a four-sided die numbered 1 to 4 is rolled.

- a Complete a table like the one shown and state the number of outcomes in the sample space.

		Die			
		1	2	3	4
Spinner	red	(red, 1)	(red, 2)		
	green				
	blue				

- b Find the probability that the outcome:

i is red and an even number

ii is blue or green and a 4

iii does not involve blue

- 5 Libby randomly selects two coins from her pocket *without replacement*. Her pocket contains a \$1 coin and two 10-cent coins.

- a List all the possible combinations using a tree diagram.

- b If a chocolate bar costs \$1.10, find the probability that she can hand over the two coins to pay for it.

- 6 A quality controller records the frequency of the types of chocolates from a sample of 120 off its production line.

Centre	Soft	Hard	Nut
Frequency	50	22	48

- a What is the experimental probability of randomly selecting a nut centre?
 b In a box of 24 chocolates, how many would be expected to have a non-soft centre?



- 7 Claudia records the number of emails she receives each weekday for two weeks as follows.
 30 31 33 23 29 31 21 15 24 23

Find the:

- a mean
 b median
 c mode
 d range



- 8 Two mobile phone salespeople are both aiming to be promoted to assistant store manager. The best salesperson over a 15-week period will achieve the promotion. The number of mobile phones they sold each week is recorded below.

Employee 1: 21 34 40 38 46 36 23 51 35 25 39 19 35 53 45

Employee 2: 37 32 29 41 24 17 28 20 37 48 42 38 17 40 45

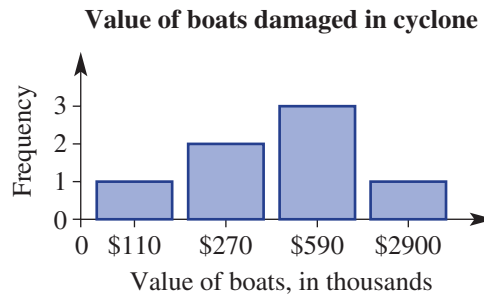
- a Draw an ordered back-to-back stem-and-leaf plot for the data.
 b For each employee, find the:
 i median number of sales
 ii mean number of sales
 c By comparing the two sets of data, state, with reasons, who you think should get the promotion.
 d Describe each employee's data as approximately symmetrical or skewed.
- 9 The data below represents the finish times, in minutes, of 25 competitors in a local car rally race.

134 147 162 164 145 159 151 143 136 155 163 157 168

171 152 128 144 161 158 136 178 152 167 154 161

- a Record the above data in a frequency table in class intervals of 120–129, 130–139 etc. Include a percentage frequency column.
 b Construct a frequency histogram.
 c Determine the:
 i number of competitors that finished in less than 140 minutes
 ii percentage of competitors that finished between 130 and 160 minutes

- 10 This column graph shows the value of some of the boats that were badly damaged in cyclone Yasi in North Queensland in 2011.



- a Find the mean and median boat value.
 b Do you think the mean or the median would better represent the ‘average’ of these boat values? Give a reason for your answer.

Extended-response questions

- 1 The local Sunday market has a number of fundraising activities.

		First spin				
		1	2	3	4	5
Second spin	1	(1, 1)	(2, 1)			
	2					
	3					
	4					
	5					

- a For \$1 you can spin a spinner numbered 1–5 twice. If you spin two even numbers, you receive \$2 (your dollar back plus an extra dollar). If you spin two odd numbers, you receive your dollar back. For any other result you lose your dollar.
- Complete the table shown to list the sample space.
 - What is the probability of losing your dollar (i.e. spinning one odd and one even number)?
 - What is the probability of making a dollar profit (i.e. spinning two even numbers)?
 - In 50 attempts, how many times would you expect to lose your dollar?
 - If you start with \$100 and have 100 attempts, how much money would you expect to end up with?
- b Forty-five people were surveyed as they walked through a market as to whether they bought a sausage and/or a drink from the sausage sizzle. Twenty-five people bought a sausage and 30 people bought a drink, with 15 buying both.
- Construct a Venn diagram to represent this information.
 - How many people bought neither a drink nor a sausage?
 - How many people bought a sausage only?
 - If a person was randomly selected from the 45, what is the probability they bought a drink but not a sausage?
 - Find $P(\text{didn't buy a sausage})$.

- 2 The delay time (in minutes) of the flight departure of the same evening flight of two rival airlines was recorded over 30 consecutive days. The data is shown below.

Airline A


2 11 6 14 18 1 7 4 12 14 9 2 13 4 19
13 17 3 52 24 19 12 14 0 7 13 18 1 23 8

Airline B

6 12 9 22 2 15 10 5 10 19 5 12 7 11 18
21 15 10 4 10 7 18 1 18 8 25 4 22 19 26

- a Copy and complete this ordered back-to-back stem-and-leaf plot for the data. Use 2 lines for each stem, starting with '0' for leaves 0–4, and then '0*' for leaves 5–9 etc.

Airline A		Airline B
	0	
	0*	
	1	
	1*	
	2	
	2*	
	3	
	3*	
	4	
	4*	
	5	
1 2 means 12		1* 5 means 15

- b Does the data for airline A appear to have any outliers (numbers not near the majority of data elements)?
- c  By removing any outliers listed in part b, find the following for each airline, rounding to one decimal place where necessary.
- i The median
 - ii The mean
- d Airline A reports that half its flights for that month had a delay time of less than 10 minutes. Is this claim correct? Explain.

