

# 1 COLLECTING AND RECORDING DATA



A local council wants to know whether the facilities for teenagers are adequate in the town.

How could it find out people's views?

How could these views be recorded and presented?

When you have read this chapter you will know how this can be done.

## Objectives

In this chapter you will:

- learn about the statistical problem-solving process and consider different types of data
- discover how to collect, record and interpret data
- look at various sampling methods
- learn how to identify possible sources of bias.

## 1.1 Introduction to statistics

### Objectives

- You can understand the stages of an investigation.
- You can formulate a question in terms of the data needed.
- You can classify data as qualitative (categorical) or quantitative (numerical).
- You can classify quantitative data as discrete or continuous.
- You can choose appropriate units of measurement and convert between metric units.

### Why do this?

To find out how good teachers are at predicting the grades their students will get in an exam, you could carry out a statistical investigation.

### Get Ready

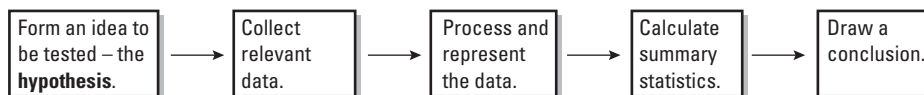
How can you find the following information?

- The average amount of lunch money for your classmates.
- What flights there are from Manchester to Washington D.C.
- How many people voted for the Green Party in the last election.

## Classifying data

### Key Points

- Statistics** is used to provide information. The statistical problem-solving process can be shown as a simple diagram:



- Data that you collect yourself is called **primary data**; data collected by other people is called **secondary data**.
- Qualitative data** can be described in words. For example, the colours of shirts on sale in a shop.
- Quantitative data** are numerical observations. There are two types:
  - Discrete data** can only take certain numerical values. For example the number of carriages on trains.
  - Continuous data** can take any numerical value. For example weights, times, lengths and temperatures are continuous.

## Metric units

- Continuous data is collected in **metric units**.
- The units of measurement used in the UK are metric units.

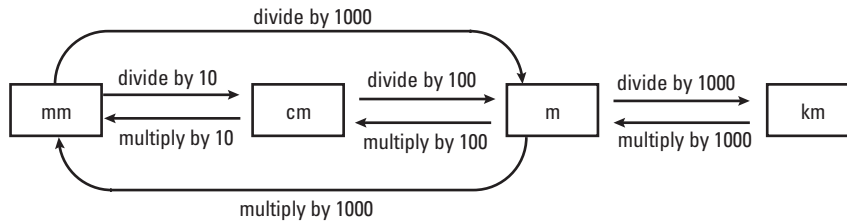
Measurement	Length	Area	Weight	Capacity/Volume
Basic unit	metre (m)	square metre (m <sup>2</sup> )	gram (g)	litre (l)

- The word for each basic unit can be changed into a bigger or smaller unit by adding one of the following words to the front of it.

Word	milli	centi		kilo
Meaning	$\frac{1}{1000}$	$\frac{1}{100}$	1	1000
Example	millimetre = $\frac{1}{1000}$ metre	centimetre = $\frac{1}{100}$ metre	metre	kilometre = 1000 metres

- To change between metric units you only need to multiply or divide by 10, 100 or 1000.
- To change from smaller units to larger ones you divide.

- To change from larger units to smaller units you multiply.  
For example, to convert lengths



- When measuring an object, select an appropriate unit of measure. For example, use metres for the height of a house, kilograms for the weight of a dog and litres for the amount of petrol in a tank.

**Example 1**

An estate agent collects the following information about houses for sale.

Type of house	Number of bedrooms	Garden area	Price
Detached	4	390 m <sup>2</sup>	£321 000
Semi-detached	3	170 m <sup>2</sup>	£184 000
Terraced	3	150 m <sup>2</sup>	£177 000
Flat	2	0	£196 000

Describe the data in each column as qualitative or quantitative. If quantitative, state whether it is discrete or continuous.

Type of house: qualitative

Number of bedrooms: quantitative and discrete

Garden area: quantitative and continuous

Price: quantitative and continuous

See if the data item can be represented as a number. If it cannot it is qualitative data. If it can be given as a number it is quantitative. If quantitative ask yourself 'Can it only take certain values?' If it can it is discrete; otherwise it is continuous.

**Exercise 1A**

Questions in this chapter are targeted at the grades indicated.

- Write down whether each of the following is secondary or primary data.
  - Data collected by you from a government website
  - Data collected by you from a newspaper
  - Data collected by you questioning people in a shopping centre
- Write down whether the following are qualitative or quantitative data.
  - The numbers of students in classes
  - The colour of students' eyes
  - The weight of dogs
  - The floor area of houses
- Write down whether the following are continuous or discrete data.
  - The number of trees in a wood
  - The time taken to run 100 m
  - The length of flower stems
  - The number of animals in a zoo
- Kai knows that he needs 2.6 metres of material to make one curtain.  
Work out how many metres he needs to make 10 curtains.
- A store owner orders a roll of material which is 92 metres long.  
It is to be cut into 100 equal size pieces. How long will each piece be?

6 A garage forecourt fuel tank holds 56 000 litres of petrol.  
The fuel tank of a certain make of lorry holds 200 litres of petrol.  
What is the maximum number of empty lorries that can be filled from the garage forecourt fuel tank?

7 Mr Longton buys 456 metres of flex. At a later date he buys 518 metres of flex.  
He wants as many pieces of flex that are 10 metres long as possible.  
How many 10 metre lengths of flex can he make?

8 Mira is going to make school skirts for her three daughters.  
The table shows the length of material, in centimetres, that she needs for each skirt.

Child	Latika	Nirupa	Saria
Length of material (cm)	130	156	183

- a Work out how many centimetres of material Mira needs altogether.
  - b Work out how many metres of material Mira needs altogether.
- Mira buys 8 metres of material.
- c Work out whether or not she has enough material to make one or more extra skirts for her daughters.

9 A medical researcher wants to find out how effective Drug A is at curing malaria.

- a Write down a hypothesis he could use.
- b What is the next thing that he would need to do?

10 It takes 125 g of flour to make a sponge cake.  
Deirdre wants to make 15 sponge cakes for the village fete.  
Maria has a 1.5 kg bag of flour. Investigate whether she has enough flour to make the 15 cakes.

11 Lemonade is supplied in 2 litre bottles.  
A lemonade glass holds 80 ml.

- a Work out how many glasses can be filled from a 2 litre bottle.

A bottle of lemonade costs £1.65 from a supermarket.

- b How much would it cost to fill 80 glasses?

D  
A03A02  
A03A02  
A03

## 1.2 Sampling methods

### Objectives

- You can collect information about a population by using a sample.
- You can select a simple random sample.
- You know that in a simple random sample each member of the population has an equal chance of being selected.

### Why do this?

A city council wants to know how many people are likely to support the idea of building a swimming pool. They can't ask everybody in the city but they can ask a sample of people.

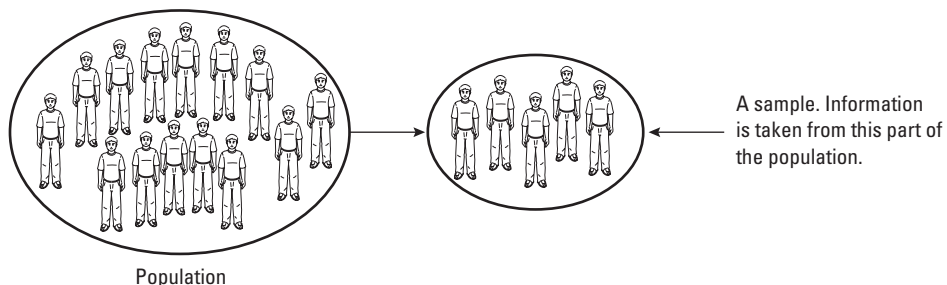
### Get Ready

If it takes 15 seconds for one student to answer a question, how long would it take to get answers from everyone in your class?

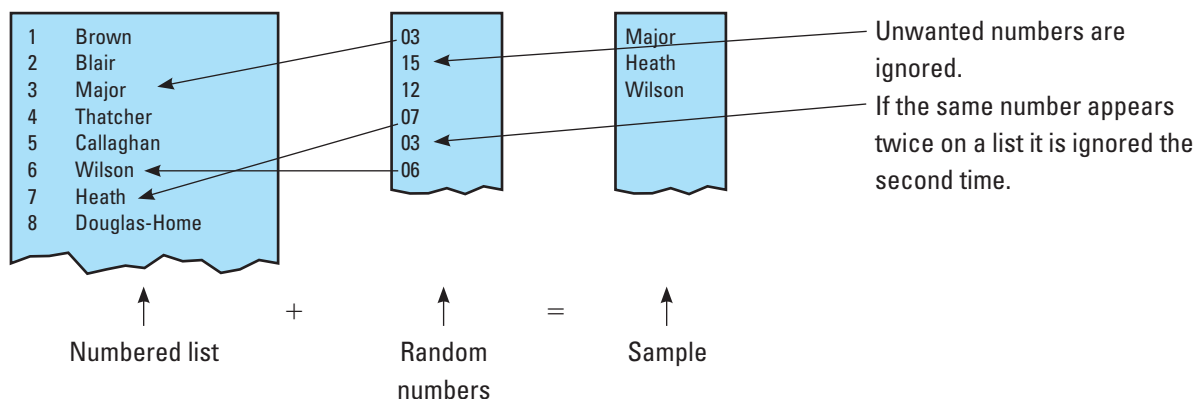


### Key Points

- A small, but carefully chosen, number of people can be used to represent the **population** of a country. These chosen individuals are called a **sample** and the investigation itself is called a sample survey.



- The sample must be **representative** of all the people or items being investigated, with each member of the population having an equal chance of being selected. If it isn't it is **biased**. For example, 'adults only' would be biased. (See Section 1.7 for more information on bias.)
- To make a sample representative, each individual in the sample should be picked at random. This process is known as taking a simple **random sample**.
- To take a simple random sample:
  - Each person or item in the population is given a number.
  - If a sample of 10 is needed, then 10 numbers are selected. This can be either: from a random number table; by a random number generator on a calculator; by using a computer; or by putting the numbers in a hat. The people whose numbers are selected then form the sample.



### Example 2

John is collecting data from each of the 50 students in his year group about the number of brothers and sisters they have.

- a** Give a reason why he might use a simple random sample.

This is an extract from a set of random numbers.

335221170532482146321143059288234171412

- b** Starting at 33 and working across, use the random numbers to give eight numbers less than 50.
- c** Explain how John would use these numbers to take a sample of eight students.

a The number of students is large and it would take a long time to collect that data.

This is one reason for taking a sample. Cost could be another reason.

b 33 21 17 05 32 48 46 11

Start at 33 and take the digits in pairs. 05 counts as the number 5. Numbers like 21 which repeat are ignored when they appear a second time. If the population was 150 you would take the digits in threes.

c He would number the students and select the ones corresponding to these numbers.



### Exercise 1B

C

1 Write down two ways in which you can generate random numbers.

2 Explain what is meant by a simple random sample.

B

3 A call centre has 60 workers. Eight are to be chosen for a new training scheme.

The manager decides to choose a simple random sample of eight.

He uses a calculator to generate random numbers. These are the first few numbers he generates.

21 32 67 54 89 78 90 34 26 45 78 54 35 64 22 ...

Describe how he could use these numbers to get his sample of workers.

## 1.3 Stratified sampling

### Objective

- You can select and use a stratified sample.

### Why do this?

A school has an equal number of boy and girl students. A simple random sample, could contain more boys than girls. A stratified sample would contain an equal number of each.

### Get Ready

- Fifteen of a class of 25 students are girls.
  - What fraction are girls?
  - What fraction are boys?
- There are three classes in Year 11. There are 22 students in class A, 28 in class B and 30 in class C.
  - How many students are in Year 11?
  - What fraction of the students in Year 11 are in class C?

### Key Points

- A population may contain groups in which the observation of interest is likely to differ. For example, if you are looking at the heights of students then the boys' heights are likely to be different to the girls' heights. These groups are called strata (singular stratum).
- A **stratified sample** is one in which the population is split into strata, and a simple random sample is taken from each stratum. The number taken from each stratum should be in **proportion** to the total number in each stratum.

- To find the number to be selected from a stratum:

1. Find what fraction of the population is in the stratum.

$$\text{Fraction in stratum} = \frac{\text{number in stratum}}{\text{number in population}}$$

2. Multiply the fraction in the stratum by the total size of the sample.

$$\text{The number sampled in a stratum} = \frac{\text{number in stratum}}{\text{number in population}} \times \text{total sample size}$$

### Example 3

The table below shows the number in each year group of a school.

A sample of 60 students is to be taken.

How many students from each year group should be in the sample?

Year	7	8	9	10	11
Number of students	150	150	100	100	100



**ResultsPlus**  
Examiner's Tip

Always make sure your individual samples total the required sample size.  
Check:  $15 + 15 + 10 + 10 + 10 = 60$

The sample for Year 7 will be  $\frac{150}{600} \times 60 = 15$

The sample from Year 8 will be the same size as for Year 7.

In each of Years 7 and 8 there are 150 students out of 600 students.

The sample for Year 9 will be  $\frac{100}{600} \times 60 = 10$

Years 10 and 11 will also have a sample size of 10.

In each of Years 9, 10 and 11 there are 100 students out of 600 students.

### Exercise 1C

- 1 A head teacher wants to find out what Year 7 students think about their first term at their new school. He decides to ask a stratified sample of 50 students. The table shows the total number of boys and the total number of girls in Year 7.

Boys	Girls
276	324

Work out the number of boys and the number of girls he should include in the sample.

- \* 2 A call centre allocates work according to the experience of its employees. Those with less than six months' experience do the easier work; those with more than six months' experience do more difficult tasks. There are 150 employees with less than six months' experience and 400 with more than six months' experience. Describe exactly how you would find a stratified sample of 10% of the employees.
- \* 3 A factory owner wants to find out what his employees think about the parking facilities at his factory. He decides to ask a stratified sample of 90 of his workers. The table shows how many people are in each of the six strata he intends to use.

	Office workers	Factory floor workers	Managers
Females	50	250	10
Males	80	490	20

- a Calculate the number of workers he needs to ask in each strata and describe how he should pick the individual members of each strata.

## 1.4 Collecting data by observation and experiment

### Objectives

- You can interpret scales on a range of measuring instruments, including clocks.
- You can design and use data collection sheets.
- You can use tallying methods.
- You can group data into class intervals.
- You can collect data by observation, experiment or data logging.

### Why do this?

If you are collecting data about the number of different types of vehicles passing by, it is easier to keep a record in the form of a data collection sheet rather than trying to remember each total.

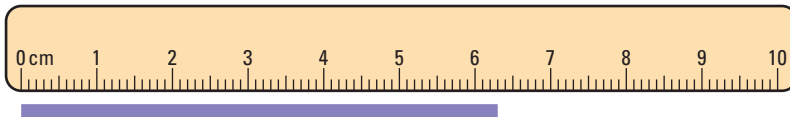
### Get Ready

- What are the numbers given by each set of marks? a |||| b |||||
- Is there a better way of grouping the marks in part b?

## Reading scales

### Key Points

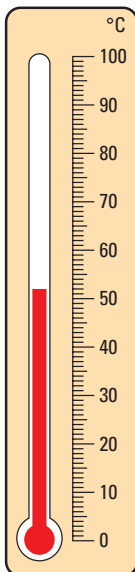
- To collect data you may have to use measuring instruments. You must be able to read different scales on measuring instruments.
- All scales have divisions marked on them and in most cases subdivisions too. The scale below is a ruler.



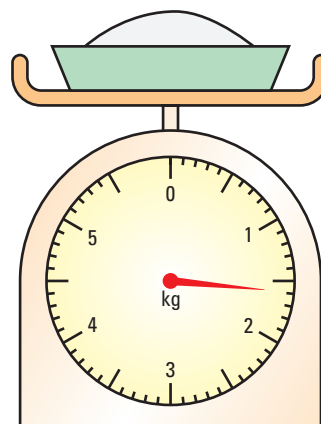
The ruler has cm divisions, each of which is divided into ten 1 mm sections.

If an article being measured is 6 cm and 3 mm long, since  $6\text{ cm} = 60\text{ mm}$ , the article is 63 mm long.

- Some scales are the same as the scale on the ruler. Here are a few examples.



This thermometer shows a temperature of  $52^{\circ}\text{C}$ .



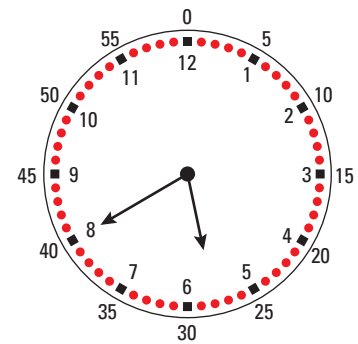
These scales show a weight of 1.6 kg.



## Time

### Key Points

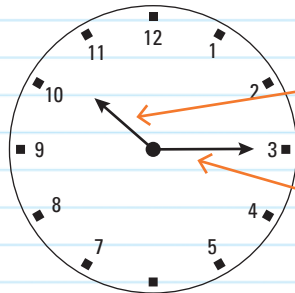
- Clocks have two scales. One shows hours and the other shows minutes.
- A clock shows 12 hours, and 60 minutes, starting from the vertical.
- There are 24 hours in the day. The 12 hour clock does not tell you if it is morning or afternoon. To tell the difference we use am for times before noon (midday) and pm for times after noon.
- Many digital clocks are 24 hour clocks.
- The 24 hour clock times always have four figures.
- Here are some examples of 12 and the corresponding 24 hour clock times.



Time	12 hour clock	24 hour clock
3 hours past midnight	3 am	03:00
Half past 11 in the morning	11.30 am	11:30
20 past three in the afternoon	3.20 pm	15:20
Quarter past 9 in the evening	9.15 pm	21:15

### Example 4

What time is showing on this clock?



The clock's short hour hand shows the time is between 10 and 11.

The clock's long minute hand shows that it is 15 minutes past the vertical position.

The clock shows 15 minutes past 10.

## Collecting data

### Key Points

- There are a number of ways of collecting data.
  - **Collecting data by observation:** if you want to investigate whether a lot of traffic is caused by people taking students to school, you could observe how much traffic there is at school opening time and compare it with how much traffic there is at other times.
  - **Collecting data by experiment:** if you wish to find out how high a tennis ball bounces when dropped from different heights, you could drop a tennis ball from various heights and record how high it bounces. This would enable you to collect data on the bounce of tennis balls.

## Chapter 1 Collecting and recording data

- **Data logging:** when you go to a supermarket till each item is bar coded so that it can be identified at the checkout. The number of each item sold is automatically recorded and this enables the supermarket to know what items are popular and what they need to stock up on.
- When collecting data by observation a **data collection sheet** is used. The following diagram shows a data collection sheet for recording the different types of transport that might pass your door.

Vehicle	Tally	Frequency
Bicycle		8
Bus		3
Car		9
Lorry		5
Motorcycle		6
Van		5

Each time a bicycle passes a **tally** mark is put next to 'Bicycle'.

When the **survey** is complete the tally marks are added together to give the total number of each vehicle. This is known as the **frequency**.

The marks are grouped into fives with the fifth tally mark drawn through the other four.

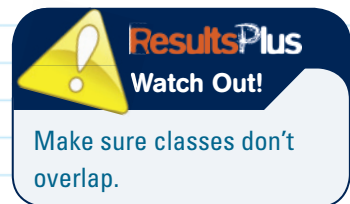
Putting tally marks in fives makes totalling up easier.

- If data is numerical, and widely **spread**, you can group the data into **class intervals**. These class intervals do not have to be the same size.
- When dealing with continuous data you need to make sure the intervals do not overlap. For example, the class intervals for a **variable** such as weight ( $w$ ) will be of the form:  
 $500 \text{ g} \leq w < 550 \text{ g}$  This means that  $w$  is greater than or equal to 500 g but less than 550 g.  
 $500 \text{ g} < w \leq 550 \text{ g}$  This means that  $w$  is greater than 500 g but less than or equal to 550 g.  
 In this case, 500 g and 550 g are the lower and upper class **limits**, while the **class size** is  $550 - 500 = 50 \text{ g}$ .

### Example 5

30 students are asked how many books they read in the Easter holiday. Each student's response is shown below. Draw and fill in a data collection sheet for this information.

1	5	9	9	6	13
6	8	4	5	5	7
9	6	11	3	8	9
14	7	7	2	12	9
3	2	1	4	0	5



Mark	Tally	Frequency
0-4		9
5-9		17
10-14		4
15-19		0

The marks have been grouped together into four equal class groups. The first class includes all the numbers between 0 and 4 inclusive.

In this example, different class intervals could have been chosen. For example:

Mark	Tally	Frequency
0-4		9
5-7		10
8-10		7
11-15		4

**Example 6**

The tally chart below shows the age at marriage of a sample of men.

Age, $a$	Tally	Frequency
$16 < a \leq 20$		
$20 < a \leq 30$		
$30 < a \leq 40$		
$40 < a \leq 50$		
$50 < a \leq 60$		
$60 < a$		

- a Fill in the frequency column.  
 b Write down the most popular age range in which men get married.  
 c Work out how many men in total there were in the sample.

a

Age, $a$	Tally	Frequency
$16 < a \leq 20$		2
$20 < a \leq 30$		13
$30 < a \leq 40$		25
$40 < a \leq 50$		7
$50 < a \leq 60$		3
$60 < a$		1

Add together the tallies:  
 $5 + 5 + 3 = 13$ .

Look for the class with  
 the highest frequency.

b 30 to 40

c 51

Add together all of the frequencies:  $2 + 13 + 25 + 7 + 3 + 1 = 51$ .

**Exercise 1D**

- 1 A road traffic controller keeps a record of the types of traffic using a busy junction during a two-minute rush-hour period. This data is listed below:

Car Car Bus Car Car Car HGV Bike Car Car Car Bus  
 Bus HGV Car Car Car Motorbike Bike Car Car Bus Car HGV  
 Bus Car Car Car Car Bike HGV Car Car Car Car Bike

- a Draw a tally chart to show this data.  
 b Write down the name of the least common type of traffic.  
 c Write down the name of the most common type of traffic.

- 2 Here is some of the data Sally collected on the lengths, in kilometres, of different journeys.  
 5.6 0.86 10.5 8.654 18.49  
 Round all these figures to the nearest whole kilometre.

- 3 Here is some of the data Nassim collected on the lengths, in centimetres, of different screws.  
 12.63 0.96 2.54 1.52 8.45  
 The factory making the screws needs to know the lengths to one decimal place.  
 Round all these figures to one decimal place.

A02  
A03

- 4 Here is part of a railway timetable.

a Which is the fastest train?

Matthew wants to travel from Preston to Birmingham. He wants to arrive before 2 pm.

b Which train should he catch?

Amelia says 'Good, this timetable shows that there is a train that arrives in Wolverhampton at 4.30 pm.'

c Explain why Amelia is wrong.

Station	Train A	Train B	Train C
Carlisle	11:09	12:07	13:00
Preston	12:17	13:17	14:28
Crewe	12:59	13:58	14:59
Wolverhampton	13:31	14:30	15:29
Birmingham	13:56	14:55	15:58

D  
A02

- 5 A shopkeeper asks 30 people entering her shop how many DVDs they have bought in the last three months. The responses are shown below:

3	5	8	9	2	7	4	10	12	3
6	2	4	9	12	13	1	7	7	11
14	3	6	5	8	1	2	7	4	3

Draw and fill in a data collection sheet showing this information. Use equal class intervals starting with the class 0–3.

- 6 A gardener weighs 24 tomatoes produced from plants in his greenhouse. The weights, in grams, are shown below:

60.5	65	64.5	59	67	61.5	67	69
58	59.3	57.2	67	68.5	63	64.2	69
57	57.8	62.4	65.5	67	58	70	75

Weight ( $w$ )	Tally	Frequency
$57 \leq w < 60$		
$60 \leq w < 63$		
$63 \leq w < 66$		
$66 \leq w < 69$		
$69 \leq w$		

- a Copy and complete the data collection sheet for this data.  
 b Write down the most common class.  
 c Write down the least common class.

## 1.5 Questionnaires

### Objectives

- You can collect data by using a questionnaire.
- You can criticise questions for a questionnaire.

### Why do this?

A restaurant may use a questionnaire to get feedback about its service, food and atmosphere if it is looking to make improvements.

### Get Ready

Describe a good method for recording data on a data collection sheet.

### Key Points

- A **questionnaire** is a list of questions designed to collect data. On questionnaires:
  - keep questions short
  - use words that are easily understood
  - do not use **biased questions** that lead the respondent to a particular answer. For example, use 'Do you agree or disagree?', rather than 'You do agree, don't you?'
  - write questions that address a single issue. For example, use 'Do you have a car?' rather than 'Do you have a petrol engine car?'
- There are two types of question to use on questionnaires.
  - An **open question** is one that has no suggested answers.
  - A **closed question** is one that has a set of answers to choose from. It is easier to summarise the data from this type of question. Closed questions will often have an opinion scale to choose from. For example:

Statistics is an important subject.

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Strongly agree	Agree	Disagree	Strongly disagree	Don't know

These are **response boxes**.  
This allows for other answers.

Sometimes a numerical scale is used. For example:

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Under 20	21 to 30	31 to 40	41 to 50	Over 50

The categories do not overlap.

- When designing questionnaires, it is important to ensure that possible answers are clear, do not overlap and cover all possibilities.



### Example 7

Write down what is wrong with each of these questions.

- a** Tick one box to indicate your age group.

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Under 20	20 to 30	31 to 40	40 to 50

- b** How often have you had a medical in the last 4 years? Tick one box.

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Never	Seldom	Sometimes	Often	Very often

- c** Do you agree that people who have regular medicals are less likely to have major illness that goes undetected?

<input type="checkbox"/>	<input type="checkbox"/>
Yes	No

- a** The categories overlap – 40-year-olds could go into two boxes.  
Other answers are not allowed for. Where does a 60-year-old tick?

- b** It is difficult to decide what these words mean.

- c** By asking 'Do you agree ...' you are inviting the answer 'Yes'. This is called a biased question.



## Exercise 1E

D  
A03

- 1 A questionnaire includes the following question.  
'Do you agree that we should build a new road?'

☐

Yes

☐

No

Write down what is wrong with this question.

- 2 A local council wants to know whether or not the residents would like a new swimming pool in the town. It is decided to use a questionnaire. The following questions are suggested.

A: What do you think about the idea of a new pool being built?

B: Do you want a new pool? Yes/No

C: Where should we build a new pool?

D: Is a pool a good idea? Yes/No

Which of the above are open questions and which are closed?

C  
A03

- 3 The management of a theme park have made some changes to the amusements. They want to use a questionnaire to find out what people think about the changes. The following questions are suggested. Write down what is wrong with each of them and design a new question for each that is more suitable.

a What do you think of the new amusements?

Very good ☐

Good ☐

Satisfactory ☐

b How much money would you normally expect to pay for each amusement?

£5–£7 ☐

£7–£8 ☐

More than £8 ☐

c How often do you visit the park each year?

Often ☐

Not very often ☐

- \* 4 A supermarket manager wants to find out if people like the new layout. She decides to use a questionnaire. Write down a suitable question she could use.

A03

## 1.6 Two-way tables

## Objectives

- You can design and use two-way tables.
- You can use information to complete a two-way table.
- You can round numbers to an appropriate degree of accuracy.

## Why do this?

You can use a two-way table to record results such as the drink preferences of boys and girls.

## Get Ready

In a class of 30 students there are:

2 left-handed girls      13 right-handed girls

4 left-handed boys      11 right-handed boys.

- a How many girls are there in the class?      b How many students are left-handed?

### Key Points

- Sometimes we collect two pieces of information, for example gender and eye-colour. To record this we would use a **two-way table**. A two-way table shows the frequency with which data falls into two different categories.

	Blue	Brown	Green	Total
Boys	6	14	5	25
Girls	4	16	5	25
Total	10	30	10	50

This is the number of boys with brown eyes.

This is the number of girls with brown eyes.

This the total number with brown eyes.

This is the total number of boys and girls.

- Sometimes a table is incomplete and has to be filled in before you can answer a question.



### Example 8

Students in Year 11 were asked to choose their favourite drink from a choice of three. Below are the boys' and girls' responses.

A02

#### Girls

Tea Coffee Coffee Tea Soft  
Tea Coffee Tea Coffee Tea  
Soft Soft Tea Tea Soft  
Coffee Coffee Soft Soft Coffee

#### Boys

Coffee Coffee Tea Soft Tea  
Tea Tea Soft Coffee Coffee  
Soft Tea Tea Coffee Coffee  
Soft Tea Coffee Coffee Coffee

- Show this information in a suitable table.
- Write down the girls' top choice of drink.
- Write down the boys' top choice of drink.
- Write down the drink that was chosen by most of the students.

a

	Tea	Coffee	Soft drink	Total
Boys	7	9	4	20
Girls	7	7	6	20
Total	14	16	10	40

The most suitable table is a two-way table. Count up the number of boys that chose tea and enter it here. Do the same for the other drinks and the girls' drinks.

Total the rows and columns.

b Tea and coffee tied.

Look for the highest number in the girls' row.

c Coffee

Look for the highest number in the boys' row.

d Coffee

Look for the drink which has the highest total.

**Example 9**

The following two-way table gives information about people's hair and eye colour.

		Eye colour			Total
		Brown	Green	Blue	
Hair colour	Brown/Black	4	4		16
	Fair	3		4	
	Ginger		1	1	4
	Total	9	8		30

- Complete the table.
- Which eye colour was most frequent?
- Which eye colour was least frequent?


**ResultsPlus**  
**Examiner's Tip**

Look for rows with only one number missing and fill these in first.

The numbers in each row must add up to the row total and the same goes for columns.

**a**

		Eye colour			Total
		Brown	Green	Blue	
Hair colour	Brown/Black	4	4	8	16
	Fair	3	3	4	10
	Ginger	2	1	1	4
	Total	9	8	13	30

The number of blue-eyed black-haired =  $16 - 4 - 4 = 8$

The number of brown-eyed ginger-haired =  $4 - 1 - 1 = 2$

The number of green-eyed fair-haired =  $8 - 4 - 1 = 3$

The total number of fair-haired =  $3 + 3 + 4 = 10$

The total number of blue-eyed =  $30 - 8 - 9 = 13$

- Blue
- Green

**Exercise 1F**

- A number of men and women were asked which type of crisps they liked best. A total of twelve people said Plain, of which seven were men. Six women liked Salt and Vinegar. Fourteen men and twelve women liked Cheese and Onion. There were 28 men in total.

  - Draw and complete a table of the data.
  - How many people liked Salt and Vinegar crisps best?
  - How many people were asked altogether?
- In a supermarket survey 30 men and 30 women were asked whether they preferred orange juice or grapefruit juice. 22 men preferred orange juice. 12 women preferred grapefruit juice.

  - Draw up a two-way table to show this information.
  - How many people liked orange juice best?

- 3** A factory employs 12 supervisors, of which 2 are female; 14 office staff, of which 3 are male; and 120 shop floor workers, of which 38 are female.
- Draw up a two-way table to show this information.
  - Write down the number of female employees.
  - Write down the total number of employees.

## Rounding numbers

### Key Points

- If the number after the place you want to round to is 5 or more, you round up.
- If the number after the place you want to round to is less than 5, you round down.
- You can round to a power of 10, the nearest whole number or to a given number of decimal places.
- When rounding to the nearest whole number the measurement given will be inaccurate by up to half in either direction.  
For example, a reading given as 4 could be between 3.5 and 4.5.
- To round numbers to a given number of significant figures (s.f.), you count that number of digits from the first non-zero digit. If the next digit is 5 or more then you round up. If the next digit is 4 or less you round down.
- Leading zeros in decimals are not counted as significant.

### Example 10

- Write 9736 correct to the nearest 1000.
- Round 9736 correct to the nearest 100.
- Round 9736 correct to the nearest 10.
- Round 5.28 to the nearest whole number.

**a** The answer is 10 000.

The number after the thousands is 7, which is greater than 5, so you round up to 10 000.

**b** The answer is 9700.

The number after the hundreds is 3, which is less than 5, so you round down.

**c** The answer is 9740.

The number after the tens is 6, which is greater than 5, so you round up.

**d** 5.28 rounds to 5 to the nearest whole number.

The number after the decimal point is a 2, which is less than 5, so you round down.

### Example 11

- Round 3.475 to one decimal place.
- Round 5.763 to 2 decimal places.
- Round 2.865 to 2 decimal places.

**a** 3.475 rounds to 3.5 to 1 decimal place.

The number in the second decimal place is a 7, which is greater than 5, so round up.

**b** 5.762 rounds to 5.76 to 2 decimal places.

The number in the third decimal place is a 2, which is less than 5, so round down.

**c** 2.865 rounds to 2.87 to 2 decimal places.

The number in the third decimal place is a 5, so round up.

**Example 12**

Round the following numbers correct to

- a 3 significant figures      b 2 significant figures

- i 4.7084      ii 0.006 375

a i 4.71 (3 s.f.)

The 8 means the 0 will be rounded up to a 1.

ii 0.006 38 (3 s.f.)

The 5 means the 7 will be rounded up to an 8.

b i 4.7 (2 s.f.)

ii 0.0064 (2 s.f.)

The 7 means the 3 will be rounded up to a 4.

**Exercise 1G**

- 1 The table gives some information about the cost of holidays in Greece.

Half Board	3 nights		1 week		10 nights		2 weeks	
Month(s) of Holiday	Adult	Child	Adult	Child	Adult	Child	Adult	Child
October 2009	340	250	469	325	569	450	729	640
Nov/Dec 2009	225	180	315	255	385	300	499	360
January 2010	215	180	315	250	375	290	529	460

Mr and Mrs Caput and their 8-year-old son Aaron decide to go on one of these holidays. They wish to go for 10 nights in November.

- a Work out the total cost of the holiday to the nearest £10.  
b Discuss how they could have reduced the cost of their holiday.

**1.7 Sources of bias****Objectives**

- You can identify possible sources of bias.
- You understand how different sample sizes may affect the reliability of any conclusions drawn.

**Why do this?**

If you want to accurately estimate the average height of students you need to collect reliable data. For example, if you include more boys than girls in your sample then you are likely to get a taller average.

**Key Points**

- When collecting data you should make sure that the data is representative of the population it is taken from. Data that does not do this is said to be biased. There are several types of bias.

**Selection bias**

- If you select only people who shop at a supermarket and ask them what they think about how that supermarket compares with a rival supermarket, you will get a biased opinion, since the people who use the other supermarket are not represented. This is called under-coverage bias.



- If you ask people to fill in a questionnaire and post it back to you, only a certain type of person will bother to respond. The respondents will not be representative of the general public. This is non-response bias.
- If you ask people to text a radio show about a controversial topic, you will get mainly people who have a strong opinion about that topic. This is voluntary response bias.
- Selection bias can be avoided by random selection and random allocation.

### Measurement bias

- If you ask people if they are satisfied, dissatisfied or very dissatisfied, you are likely to get a biased opinion because there are two answers for dissatisfied and only one for satisfied.
- If you ask a question such as, 'You are satisfied, aren't you?' you are more likely to get 'Yes' as an answer. These are called **leading questions**.
- People like to present themselves in the best light, so if you ask them, 'Do you often behave unreasonably?' you are likely to get 'No' for an answer even if it should be 'Yes'.

### Sampling error

- If you take two random samples you are unlikely to get exactly the same result (though they should be close to each other). This is called the sampling error.
- Increasing the sample size will reduce the sampling error. It is difficult to say how big a sample should be as this depends on how varied the population is. However, the larger the sample the more representative it will be of the population and the more accurate the information will be.

#### Example 13

Write down, with reasons, whether or not each of the following is biased.

- You want to find out what people think about a football team. You ask supporters as they enter the ground before a match.
- You wish to find out what proportion of the population has had flu in the last month. You interview people in the doctor's waiting room.
- You ask the first 10 people you meet, 'Do you agree that banning smoking in public places is a good thing?'
- You ask, 'Have you ever been convicted of drink driving?'
- You ask three people what they thought of the Eurovision Song Contest.

- Biased. Non-supporters are not represented.
- Biased. People who do not visit the doctor are not represented.
- Biased. Not everyone has an equal chance of being asked.  
The question is leading the respondent to agree.
- Biased. This is a sensitive question. You are not likely to get a true answer.
- Biased. The sample is too small.



### Exercise 1H

- 1 An examination board wants to get information on schools' views regarding how they respond to queries. They send a questionnaire to a sample of schools in the London area. Is this a biased sample? Give one reason for your answer.

A03

C

- 2 Write down, with reasons, whether or not each of the following are biased.
- A: A hospital wants to know how often people use A & E. They ask all the people attending A & E on one particular Wednesday.
- B: An opinion poll company wants to find out how voters would vote if there were to be an election next week. They conduct a telephone poll of 20 voters in each of 10 towns.
- C: A manufacturer of climbing ropes wants to see if his ropes are of the strength he advertises. He tests a sample. He tests every tenth rope made.
- D: You ask 50 people using a recycling facility what they think about recycling.
- \* 3 One hundred people attend a rally on 'action for climate change'.  
David says, 'That is a lot of people. They must be right.'  
Jody says, 'I disagree.'  
Discuss the views of David and Jody.

## 1.8 Secondary data

### Objective

- You can extract data from lists and tables.

### Why do this?

You want to find out how many accidents there were in your town last year. You can't count these yourself so you have to get the data from published sources.

### Get Ready

If you needed a new mobile phone, where would you look to find one that best suited you at a price you could afford?

### Key Points

- Secondary data can be obtained relatively quickly and cheaply from a number of sources, including reference books, journals, newspapers and the internet. Remember, however, that the data may be inaccurate or out of date. Only use data from a reliable source and check the data against another source if possible.
- It is also possible to obtain secondary data from a database. A **database** is an organised collection of information, usually stored on a computer.
- The spreadsheet below shows part of a database kept on a computer. The entries at the top in red are fields. The entries below in black are the records. They can be easily changed to be arranged in numerical, alphabetical, gender or age order.

ID number	Surname	Forename	Gender	Age
01	Abbot	David	M	32
02	Adair	Jakie	F	27
03	Allison	Paul	M	45
04	Barber	Hassan	M	25
05	Baxter	Jenny	F	38

**Example 14**

Part of a database for second-hand Ford Mondeo cars is shown below.

Vehicle summary	Colour	Engine	Mileage	Price	Year
Ford Mondeo Edge	Black	2000cc petrol	11 549	£9 995	2006
Ford Mondeo Edge	Blue	2000cc petrol	14 100	£10 499	2008
Ford Mondeo Edge	Grey	2000cc petrol	10 400	£11 599	2008
Ford Mondeo Edge	Grey	2000cc petrol	12 654	£11 494	2007
Ford Mondeo Edge	Blue	2000cc petrol	7520	£11 999	2008
Ford Mondeo Zetec	Silver	2000cc petrol	10 078	£11 995	2008
Ford Mondeo Zetec	Silver	2000cc petrol	12 088	£14 995	2008
Ford Mondeo Titanium	Grey	2000cc petrol	11 555	£12 395	2008
Ford Mondeo Zetec	Black	2000cc petrol	5800	£12 895	2008
Ford Mondeo Zetec	Silver	2000cc petrol	12 123	£12 995	2008

- Which four fields could be used to order the data?
- What was the mileage of the car that cost over £13 000?
- What colour was the car that had driven the least number of miles?
- What was the maximum mileage driven by one of these cars?

a Mileage, price, year, colour.

Mileage price and year could be put in numerical order.  
Colour could be put in alphabetical order.

b 12 088

Find the car that had a price greater than £13 000 and look in its mileage column.

c Black

In the mileage column, find the car that had driven the least mileage then look across its row to the colour column.

d 14 100

Look for the largest number in the mileage column.

**Exercise 11**

- The following database gives some information about the CO<sub>2</sub> emissions, in thousand tonnes of carbon dioxide equivalent, in a certain country.

	Year					
	2003	2004	2005	2006	2007	2008
<b>Buses</b>	323	344	355	342	394	421
<b>Cars</b>	6280	6251	6163	6159	6063	6055
<b>HGVs</b>	2147	2154	2235	2295	2162	2221
<b>Motorcycles</b>	39	41	44	40	41	39
<b>Railways</b>	231	209	225	241	245	251

- What were the emissions for motorcycles in 2007?
- Which form of transport produced the most emissions?
- Write down the year when emissions for railways were lowest.
- Write down the method of transport for which the emissions have dropped each year.

- 2 The following database gives information about the weather in a certain town during the first six months of the year.

	Max temp °C	Min temp °C	Air frost days	Sunshine hours	Rainfall mm	Days of rainfall ≥ 1 mm
January	6.4	1.2	10.7	44.3	101.9	15.3
February	6.9	1.3	9.6	72.0	73.4	11.3
March	8.8	2.5	6.3	107.9	78.3	14.1
April	11.4	3.5	3.8	155.1	50.7	10.6
May	15.0	6.1	1.0	214.8	55.0	10.0
June	17.1	9.0	0	197.7	67.9	10.7

- How many days of air frost were there in March?
- Write down the month that had the least number of days of rainfall.
- Which was the sunniest month?
- Which month had the greatest difference between maximum and minimum temperatures?

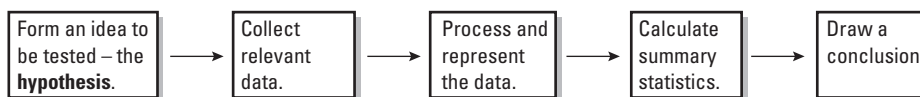
- 3 The database below is part of an agricultural survey of cattle in England in 2006, 2007 and 2008. The numbers of cattle are given in thousands.

	Year		
	2006	2007	2008
<b>Female cattle</b>			
Aged 2 years or more	2550	2531	2475
Total breeding herd (cattle that have calved)	2043	2027	1994
Beef	767	768	758
Dairy	1276	1259	1236
Other female cattle (not calved)	507	504	481
Beef	224	220	216
Dairy	293	284	265
Aged between 1 and 2 years	825	799	778
Beef	512	502	497
Dairy	313	297	282
<b>Male cattle</b>			
Aged 2 years or more	217	217	217
Aged between 1 and 2 years	619	583	578

- Write down the number of female cattle aged between one and two years in 2007.
- What do you notice about the numbers of male cattle aged two years or more throughout the three years?
- Were there more female beef or more female dairy cattle in 2008?
- What conclusions can you draw about the trend in the numbers of cattle over the three years?

## Chapter review

- **Statistics** is used to provide information. The statistical problem-solving process can be shown as a simple diagram:



- **Primary data** is data you collect yourself.
- **Secondary data** is data that has been collected by others.
- **Qualitative data** can be described in words.
- **Quantitative data** are numerical observations.
- **Discrete data** can only take certain numerical values.
- **Continuous data** can take any numerical value.
- A **sample** is part of a population that is used to give information about the population as a whole in a **sample survey**. The sample must be representative of all the people or items being investigated, with each member of the population having an equal chance of being selected.
- A simple **random sample** is one where each person is given the same chance of being included.
- A **stratified sample** is one in which the population is split into groups called strata and a simple random sample is taken from each stratum. The number taken from each stratum is proportional to the size of the stratum.
- Data can be collected by **observation**, **experiment** or **data logging**.
- When collecting data by observation a **data collection sheet** is used.
- When dealing with continuous data you need to make sure the intervals do not overlap.
- If data is numerical, and widely spread, you can group the data into **class intervals**. These class intervals do not have to be the same size.
- A **questionnaire** is a list of questions designed to collect data.
- An **open question** is one that has no suggested answers.
- A **closed question** is one that has a set of answers to choose from.
- When designing questionnaires, it is important to ensure that possible answers are clear, do not overlap and cover all possibilities.
- A **two-way table** shows the frequency with which data falls into two different categories.
- **Biased** data is data that does not represent the population that it is taken from.
- A **database** is an organised collection of information.



### Review exercise

- 1 a In an experiment to look at the growth rate of beans Phoebe measures the height of 5 bean shoots. The heights are 6.7 cm, 7.5 cm, 5.5 cm, 6.3 cm and 7.2 cm.  
Write these measurements to the nearest whole number.
- b James uses a measuring instrument to measure the lengths of 5 bolts. The lengths are 4.56 cm, 3.98 cm, 4.55 cm, 5.67 cm and 3.95 cm.  
Write these measurements to the nearest one decimal place.



- 2 James wants to find out how many text messages people send. He uses this question on a questionnaire.

'How many text messages do you send?'

1 to 10 ☐ 11 to 20 ☐ 21 to 30 ☐ more than 30 ☐

- a Write down **two** things wrong with this question.

James asks 10 students in his class to complete his questionnaire.

- b Give **one** reason why this may not be a suitable sample.

March 2009

- 3 Poppy wants to find out how much time people use their computer for. She uses this questionnaire.

For how much time do you use your computer?

0–1 hours ☐ 3–4 hours ☐

1–2 hours ☐ 4–5 hours ☐

2–3 hours ☐ 5–6 hours ☐

- a Write down **two** things that are wrong with this question.

Poppy gives her questionnaire to all the students in her class. Her sample is biased.

- b Give **one** reason why.

Nov 2008

- 4 Courtney lives in Oxenholme. She has an interview for a job in Glasgow at 10.30 am. The interview will take 45 minutes. She will travel by train. It takes Courtney 20 minutes to walk from her home to the station. It takes 35 minutes for Courtney to walk from the station to the place of her interview. Here is part of the train timetable from Lancaster to Glasgow and from Glasgow to Lancaster.

Lancaster to Glasgow			
Lancaster	06:54	08:08	08:29
Oxenholme	07:08	08:22	08:43
Penrith	07:34		09:19
Carlisle	07:50	09:01	09:21
Glasgow	09:14	10:16	10:46

Glasgow to Lancaster			
Glasgow	14:40	16:00	16:40
Carlisle	15:49	17:09	17:51
Penrith			18:05
Oxenholme	16:24	17:42	18:29
Lancaster	16:38	17:54	18:44

She goes straight to the interview.

She wants to do some shopping in Glasgow but wants to catch the quickest train home after 2 pm.

- a Plan a schedule for Courtney.

The fastest train from Lancaster to Glasgow in the morning is the 08:08.

- b Suggest reasons for this.

	Time
Courtney leaves home	
Train departs Oxenholme	
Train arrives Glasgow	
Courtney arrives for interview	
Interview finished	
Train leaves Glasgow	
Train arrives Oxenholme	
Courtney arrives home	

- 5 The table shows some information about the cost, in £s, of all inclusive holidays to Bahrain. The price is per adult. There is a 20% reduction for children.

Hotel	Economy Class		Business Class	
	3 nights	5 nights	3 nights	5 nights
Metro	469	595	1219	1345
Habtoor	505	655	1350	1505
Hilton	510	659	1259	1410
Atlantis	659	975	1469	1735

Wing, his wife and 12-year-old daughter plan to go to the Hilton Hotel for 5 nights, travelling Business Class.

- How much will Wing have to pay for his family to take this trip?
- Discuss ways in which Wing could reduce the cost of this holiday.

- 6 Naomi wants to find out how often adults go to the cinema. She uses this question on a questionnaire.

'How many times do you go to the cinema?'

☐  
 Not very often

☐  
 Sometimes

☐  
 A lot

- Write down **two** things wrong with this question.
- Design a better question for her questionnaire to find out how often adults go to the cinema. You should include some response boxes.

Nov 2008

- \* 7 Valerie is the manager of a supermarket. She wants to find out how often people shop at her supermarket. She will use a questionnaire.

Design a suitable question for Valerie to use on her questionnaire. You must include some response boxes.

June 2008

- \* 8 Yolande wants to collect information about the number of e-mails the students in her class send. Design a suitable question she could use on a questionnaire. You must include some response boxes.

March 2008

- 9 Melanie wants to find out how often people go to the cinema. She gives a questionnaire to all the women leaving a cinema. Her sample is biased. Give **two** possible reasons why.

March 2008

- \* 10 Amberish is going to carry out a survey about zoo animals. He decides to ask some people whether they prefer lions, tigers, elephants, monkeys or giraffes. Design a data collection sheet that he can use to carry out his survey.

March 2006

D  
A02  
A03

C  
A03

A03

A03

A03

A03

C  
A02  
A03

- \* 11 Angela asked 20 people in which country they spent their last holiday. Here are their answers.

France	Spain	Italy	England
Spain	England	France	Spain
Italy	France	England	Spain
Spain	Italy	Spain	France
England	Spain	France	Italy

Design **and** complete a suitable data collection sheet that Angela could have used to show this information.

March 2004

A03

- 12 The manager of a Country Park asks the following two questions on a questionnaire.

'Do you go to the Country Park?' Sometimes ☐ Often ☐

'How old are you?' 0 to 10 years ☐ 10 to 20 years ☐ Over 20 years ☐

- a What is wrong with each of these questions?  
b For both questions above, write a better version that the manager can use.

A03

- 13 Write down, with reasons, whether or not each of the following is biased.

- a A call centre manager wants to know how easy it is to use the staff reference sheets when answering a call. He asks all the people working on the night shift.  
b A mobile phone company wants to find out what people think about their new pricing contract and randomly select 10% to ask.  
c A town council poses the question 'Do you agree that we are doing a good job in the area of recycling?'

A

- 14 The two-way table shows information about the number of students in a school.

	Year Group					Total
	7	8	9	10	11	
Boys	126	142	140	135	127	670
Girls	134	140	167	125	149	715
Total	260	282	307	260	276	1385

Robert carries out a survey of these students.

He uses a sample of 50 students stratified by gender and by year group.

Calculate the number of girls from Year 9 that are in his sample.

June 2008

- 15 The table shows the number of boys in each of four groups.

Group	A	B	C	D	Total
Number of boys	32	41	38	19	132

Jamie takes a sample of 40 boys stratified by group.

Calculate the number of boys from group B that should be in his sample.

March 2008

- 16** 258 students each study one of three languages.  
The table shows information about these students.

	Language studied		
	German	French	Spanish
Male	45	52	26
Female	25	48	62

A sample, stratified by the language studied and by gender, of 50 of the 258 students is taken.

- a** Work out the number of male students studying Spanish in the sample.  
**b** Work out the number of female students in the sample.

June 2009

**17**

	Male	Female
First year	399	602
Second year	252	198

The table gives information about the numbers of students in the two years of a college course. Anna wants to interview some of these students. She takes a random sample of 70 students stratified by year and by gender.

Work out the number of students in the sample who are male and in the first year.

Nov 2008

- \* **18** 80 children went on a school trip.  
They went to London or to York.  
23 boys and 19 girls went to London.  
14 boys went to York.  
Draw and complete a suitable table of this information.

March 2009



**ResultsPlus**  
Exam Question Report

82% of students answered this sort of question poorly.