Contents

MODULE 1 Higher TIER

6 Probability.

Outcomes of experiments.
Measuring probabilities.
Events that can happen in more than one way.
Probability that an event does not happen.
Mutually exclusive outcomes.
The number of times an event is likely to happen.
Relative frequency.

7 Time series

Reading time-series graphs.
Calculating moving averages.
Plotting moving averages.
Index numbers.

8 Measurements of spread.

Range.
Cumulative frequency tables.
Cumulative frequency curve.
Drawing a cumulative frequency curve.
Using a cumulative frequency curve to find the median.
Quartiles.
Box plots.

9 Compound events.

Compound events.
Independent events.
Conditional events.
Tree diagrams.
Using tree diagrams to find probabilities.

10 Conclusions.

Drawing conclusions.
Comparing data.

Trial papers

Answers

Index

1 Posing the question

This chapter will show you:
- ✓ how to write a hypothesis
- ✓ when statistical methods help to answer a question
- ✓ how to decide what information you need and whether you can get it

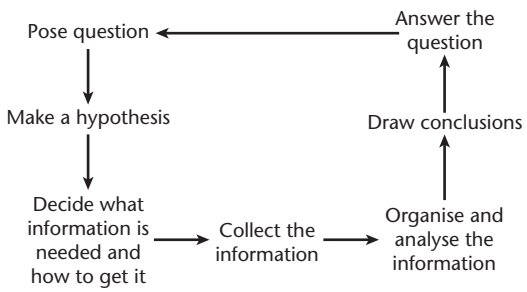
Before you start you need to know:
- ✓ how to use the internet to find information

1.1 Statistical methods

Statistical methods deal with lots of factual information called **data**.

The handling data cycle

Statistical methods use the **handling data cycle**.



Questions that can be answered using statistics need lots of factual information. For example, 'How many people smoke?' can only be answered by knowing the habits of a large number of people.

Some questions can be answered without using statistics.
For example, 'Does that person smoke?' needs one fact only. It does not need statistics!

1.2 Making a hypothesis

A **hypothesis** is a statement that attempts to answer a question.
It may or may not be true.

A hypothesis needs to be tested to find out whether or not it is true.
This means that when you write a hypothesis you must be careful what words you use.
They should include things that can be measured.
They should not include words whose meaning is unclear.

Did you know
that Florence Nightingale (1820–1910) used statistics to convince the British Government to form the first professional nursing service?

She succeeded in reducing deaths in military hospitals in the Crimea from 42% of patients to 12% in six months.

Class discussion
Will statistical methods help answer these questions?
1. How much spending money do sixteen-year-old students have?
2. What time does the next train leave the station?
3. Does this dice give more sixes than any other score?
4. What do people think that Council Tax money should be spent on?
5. What size shoes do I need?
6. How many school places will be needed in ten years' time?
7. Will my mobile phone work in the USA?

Example 1

Write a hypothesis for the question 'Does smoking shorten life?'

Smoking more than 1 cigarette a day reduces life expectancy.

This hypothesis has an exact meaning and can be tested.

Example 2

Give a reason why 'Smokers die young.' is not a good hypothesis.

'Smokers' means different things to different people.

'Young' can also mean different things to different people; 'five' is young to a twenty-year-old and 'fifty' is young to an eighty-year-old.

Exam practice 1A

1 Amy wrote this hypothesis:

Fat people die young

Write down **one** reason why this is not a good hypothesis to test.

'Fat' is not clear enough to measure. You could use 'Clinically obese' which can be measured – it is a body mass index over 30.

2 Write down **one** reason why each of the following is not a good hypothesis to test.
- a Fit people do not become ill very often.
- b Young people have lost interest in politics.
- c More old people vote in a general election than young people.

Did you know
that you can find your body mass index by dividing your weight in kilograms by the square of your height in metres?

3 Write a hypothesis for each question.
- a When this coin is flipped many times, does it show more heads than tails?
- b Are the sentences in broadsheet newspapers longer than sentences in tabloid newspapers?
- c Are people better at estimating the area of a square than the length of a line?
- d Do left-handed people have different reaction times from right-handed people?
- e How old are children when they learn to walk?

A/w 101

4 David asked 'If the price of alcohol is increased, will this stop people getting drunk?'
- a Write down a hypothesis for this question.
- b Give a reason why this might be difficult to test.

1.3 Finding information

The next step is to decide what information is needed and if you can get it. These questions help with your planning.
- What information is needed?
- Does the information exist?
- Can you get this information?

1.4 Primary and secondary data

There are two main sources of information – secondary data and primary data.

Secondary data is information that has already been collected.

You can get secondary data from several places. Newspapers, books and the internet are good places to look.

Primary data is information you collect yourself.

You can get primary data by measuring, asking people questions.

Class discussion

What information is needed to test each of the following hypotheses? Can you use secondary data, in which case where can you look to find it? Will you need primary data, in which case how will you collect it?

1 Runner bean seeds germinate 2 weeks after they are planted.

2 Three-year-old cars are cheaper than one-year-old cars.

3 People who do not eat green vegetables have more days off work than those who do.

4 Most people do not pay all the tax that they should pay.

5 Self-employed people work longer hours and take shorter holidays than employed people.

6 When this spinner is spun, it is more likely to land on 1 than any other number.

(fig A/W: four sided spinner labelled 1, 2, 3, 4.)

Example 3

Max wants to test the hypothesis
'More people over 40 years old use their vote than younger people.'
Answer each of the following questions.

a What information does Max need?

b Does this information exist? Give a reason for your answer.

c How can Max get this information?

a *Peoples' ages and when they have voted.*

b *Yes. People know how old they are and most will know when they have voted.*

c *Max could find some opinion polls that have this information.*
He may also be able find some information on the internet.
Max can collect the information himself by asking people questions.

ICT task

Use a search engine on your computer to find websites giving information on:

a Voting patterns by age.

b Prices of secondhand cars.

1.5 Manageability

The manageability of a task is how easy or difficult it is to carry out.

- How much information do I need?
  You should collect as much data as practical. This can depend on the time available and the cost of collecting the data.
- Is this task manageable?
  This can depend on the amount of time available and the cost of carrying out the task.

Census or survey?

A **census** gives information on all of the things you are investigating.

If you want the heights of oak trees in a small wood, you can probably get all of them. This is a census.

A **sample** is some but not all the things you are investigating.
A **survey** is an investigation that uses a sample.

If you want the heights of oak trees in the UK, you will not be able to get them all. You will have to select some of them. The number that you select depends on the time you have available. Those that you select are the sample.

Did you know?

The government is the only organisation that can demand information from everyone. It does this once every ten years in a National Census.

Exam practice 1B

1 Gerry is investigating the statement
'Girls get better GCSE grades in maths than boys.'
He uses a list of the grades from students in one school.
Is Gerry doing a census or a survey?
Explain your answer.

Explain your answer means write down why you decided it is a census (or a survey).

2 Kay is investigating the statement
'This dice gives more sixes than it should.'
She rolls the dice 100 times and writes down what it shows each time.
Is Kay doing a census or a survey?
Explain your answer.

3 Justin wants to investigate the attitude to further education among the students in his class.
He decides he needs to interview all the students in his class.
He has four weeks to complete the task.
Is this task unmanageable? Explain your answer.

4 Faith has planned a coursework task.
As part of the task she wants to time how long it takes to get to her school from fifty different places.
Write down two reasons why this could be unmanageable.
Photo of high street in town centre

5 Ravi wants to investigate the statement
'People prefer to shop at the supermarket outside the town rather than in the town centre.'
Ravi decides he will get information by questioning people in the supermarket and in the town centre.
Is this a survey or a census?
Explain your answer.

Summary of key points

Before you start a statistical investigation you need to decide:

- Whether the question you are trying to answer can be answered using statistical methods.
- What information you need and whether you can get that information.
- Whether you can get all the possible information or only some of it.
- If the task is manageable.

Most students who get GRADE C or above can:

- decide what primary and/or
- secondary data is needed.

Most students who get GRADE A or above can also:

- pose their own questions and form a hypothesis.

Glossary

census	information from all possible items being investigated
data	factual information
hypothesis	a statement that may or may not be true
primary data	factual information collected by person
sample	same, but not all, of the items being investigated
secondary data	factual information provided by someone else
survey	investigation based on a sample

2 Collecting data

This chapter will show you:
- ✓ the difference between qualitative and quantitative data
- ✓ the difference between discrete and continuous data
- ✓ how to design a data collection sheet
- ✓ how to choose a sample
- ✓ how to design a questionnaire
- ✓ how to deal with missing data from questionnaires

Before you start you need to know:
- ✓ common units for length, mass and time
- ✓ how to simplify a fraction
- ✓ how to find one quantity as a fraction of another
- ✓ how to find a fraction or a percentage of a quantity
- ✓ how to round decimals to the nearest whole number

2.1 Sampling

When you want to collect data you need to decide if you can get all the possible information or only some of it.

The **population** is all the possible items being investigated.
A **sample** is some, but not all, of the possible items.

A **representative** sample reflects the entire population.
An **unrepresentative** sample does not reflect the entire population.

Any results from an unrepresentative sample will be **biased**.

If you want information about oak trees in the UK, the population is all the oak trees in the UK. If you want information about state schools in England, the population is all state schools in England.

Example 1

Julie reads that the council want to make the town centre a traffic-free zone between the hours of 8 a.m. and 10 p.m.
She wants to investigate what support there is for this plan.

a Give reasons why Julie will have to use a sample.

b Julie chooses her sample by asking people in the town centre in the evening.
Give one reason why this is unrepresentative.

a Julie will have to use a sample because she cannot be certain that she can ask everyone living in the town and even if she could it would take far long.

b Some groups of people, such as parents with young children, are unlikely to be included.

Also some people will refuse to give any information and some people will give incorrect information.

When you choose a sample, you must try to avoid bias by not leaving out group of people.

Exam practice 2A

1 A survey on voting intentions in a local election was conducted by telephoning people between the hours of 9 a.m. and 5 p.m.
Give one reason why the sample may not be representative.

Think of one group of people that cannot be reached by telephone at those times.
Write down a short sentence giving this group as your reason.
For example 'People who … will not be part of the sample.'

2 A survey into attitudes to a ban on cars in the city centre was conducted by asking the opinions of people coming out of the car park.
Give one reason why the sample will not be representative.

3 A survey at a railway station found that two-thirds of the people questioned did not own a car.
This survey was used to claim that two-thirds of the people living in the town did not own a car.
Give a reason why this claim is likely to be biased.

4 Kate read that adults spend more than £30 a week on eating out.
This was based on a survey carried out in a restaurant.
Explain why this result is likely to be biased.

5 Sunifa is investigating the connection between obesity and ill health. She decided to carry out a survey of people attending an outpatients clinic at the hospital.
Explain why her sample may not be representative.

6 A local council conducted a survey to find opinions on services that could be reduced to save money. The survey was carried out at an out-of-town supermarket and at the local library.
Write down one group of people whose views would not be likely to be included.

7 A survey needed information from men between the ages of 20 and 60. Three students were each asked to find a sample of 10 men to interview. They each went to a different location.
Give a reason why each of the following locations might introduce bias into the survey.
   a The exit from the local supermarket between 11 a.m. and 12 a.m.
   b The bus stop between 8 a.m. and 9 a.m.
   c The local cinema in the half hour before the film begins.

2.2 Random samples

When you choose a sample, it is important to reduce bias as much as possible.

In a **simple random sample** every possible item in the population has an equal chance of being selected.
This can done by allocating a number to each item and using random numbers to select your sample.

You can produce random numbers using a variety of methods:

The raffle method.

Write the number of each item on a piece of paper then draw the required number of pieces from a box without looking.

This can take a very long time unless the sample and the population are both small.

Random Number Generators.

Random number tables can be used to give as many random numbers as you need.

Most scientific calculators and spreadsheets can also generate random numbers.

(Photo of lottery machine)

The National Lottery uses a variation on the raffle method to select the winning numbers.

Example 2

The 50 members of a dining club are listed in alphabetical order.

a Explain why choosing every 5th member on this list does not give a random sample.

b Explain how you could obtain a random sample of 10 members.

Ph 204 people dining in a club

a Every member does not have an equal chance of being chosen. The members who are not 5th, 10th, 15th, ... on the list have no chance of being chosen.

b Number the list of members from 1 to 50. Generate 10 different random numbers from 1 to 50 and select the corresponding members.

To be a random sample, every member must have an equal chance of being chosen.

You can describe any method that gives random numbers.

Exam practice 2B

1 A school was asked to provide a sample of pupils for a survey. The first child listed on the register in each class was chosen. Explain why this is not a random sample.

2 There are two girls and seven boys in a chess club. One girl and one boy are chosen from the members of the club. Explain why this is not a simple random sample.

Who has a greater chance of being chosen, a boy or a girl?

3 A sample of 5 of these rods is to be chosen.

a David chooses the five rods shown with an arrow because he thinks they are a representative sample. Give a reason why David is wrong.

b Explain how you would choose a random sample.

A/w 201

4 In a competition with 10 prizes, 60 correct entries were received. The correct entries were numbered 01 to 60, and two-digit random numbers were used to select the prize-winning entries. If a number was greater than 60 was generated, the organisers subtracted 60 and used the resulting number. Explain why this method is biased.

ICT task
Find out how to use a spreadsheet to give a set of ten random numbers from 0 to 60.
Use the spreadsheet to select a random sample of 6 students from your class.

2.3 Stratified sampling

A simple random sample may not be representative.
If a population contains 100 women and 50 men, a random sample from this population will not necessarily contain twice as many women as men.

If the data has no relevance to gender, the sample need not be representative of the women and men.

A **stratified sample** produces a representative sample by first dividing the set into groups. Then a simple random sample is selected from each group. The number chosen in each group is the same fraction of the group that the sample is of the population.

Example 3

There are 100 women and 50 men who are members of a gym club. Calculate the number of women and the number of men for a stratified sample of size 12 from the members.

First find the sample size as a fraction of the total membership. Then calculate this fraction of the numbers in each group:
$\frac{2}{25}$ of $100 = \frac{2}{25} \times \frac{100}{1}$
$= 8$, and so on.

The sample size as a fraction of the total membership is $\frac{12}{150} = \frac{2}{25}$

The number of women needed is $\frac{2}{25}$ of the 100 women = 8 women.

The number of men needed is $\frac{2}{25}$ of the 50 men = 4 men.

The advantage of stratified sampling is that it ensures proper representation of each group and still keeps an equal chance of selection for each member of the population.

The disadvantage is that division of a population into groups is not always easy.
For example, the number of people living in Greater London who are under 20 years old can only be an estimate.
The members of the tennis club who are classed as good players or as average players needs human judgment.
Situations like these can introduce bias into the sample.

Class discussion

The owners of a shoe shop want to do a survey. For each of the following data, do they need a simple random sample or a stratified sample? If they need a stratified sample, what categories should they use?

1 The ages of their customers.
2 The amount spent by each customer.
3 Whether the customers buy trainers or ordinary shoes.
4 Whether a customer has to wait more than two minutes before being served.
5 How many customers chose a style that was not in stock in their size.
6 The method the customer uses to pay for their purchase.

Example 4

A consignment of 3560 bottles of water is delivered to a restaurant. The number of bottles of each type of water is given in the table.

Category	Still	Carbonated	Lemon flavoured	Lime flavoured
Number	1280	1980	55	245

How many bottles of each type of water should be selected to give a stratified sample that is 5% of the consignment?

5% of 1280 = 64
5% of 1980 = 99
5% of 55 = 2.75 = 3 to the nearest whole number.
5% of 245 = 12.25 = 12 to the nearest whole number.
We need 64 bottles of still water, 99 bottles of carbonated water, 3 bottles of lemon-flavoured water and 12 bottles of lime-flavoured water.

For a 5% sample, you need 5% of each type of bottle.

Rounding each calculation to the nearest whole number gives a total of 178 bottles for the sample. Check that this is 5% of the total number of bottles delivered.

You need to decide what categories Sam should use and what proportion of each category is needed. You are not asked to find the number needed in each category.

Class discussion

A university education department wants information about pupils in Year 7. They decide to gather information about 5% of the pupils. Here are some ways in which those pupils can be selected.

1 Use a list of all secondary schools in the country. Use a simple random sample of 5% of the schools.
Is this sample likely to be representative of the different types of school, i.e. single sex, selective, independent, comprehensive and so on?
2 Instead of using a simple random sample, every twentieth school in the list can be selected.
Why this is method unlikely to give a sample that is representative of the different types of school.
3 The list of schools can be sorted into categories such as single-sex, religious schools, etc. Select one twentieth of the schools in each category.
This method is likely to give a more representative sample than the first two methods, but what can be done about a school in more than one category?
4 Another method of selection that overcomes the difficulty of categorising schools is to select one twentieth of Year 7 pupils from every school in the country. How can these pupils can be chosen so that any one pupil is as likely as any other to be selected?

Exam practice 2C

1 There are 180 pupils in each of years 7, 8, 9, 10 and 11 in a school.
Sam wants to choose a sample of 90 of these pupils.
Explain how he can choose a stratified sample.

2 The table shows the number of students enrolled in each department of a college.

Department	Catering	Academic	Engineering	Sport	Business
Number of students	72	33	75	66	54

How many students should be chosen from each department to get a 10% stratified sample?

3 There are 180 pupils on the register at a primary school. The table shows the number of pupils in each year.

Year	1	2	3	4	5	6
Number	34	33	29	28	24	21

Mrs Jaffa needs to choose 30 pupils for a survey.

a Explain why choosing 5 pupils from each year is not representative.

b Calculate how many students Mrs Jaffa needs to select from each year to give a stratified sample of 30 pupils.

1 tonne = 1000 kg.

4 A farmer grew the same variety of potatoes in three different fields. The table shows the yield, in tonnes, from each of these fields.

Field	A	B	C	D
Yield (tonnes)	250	220	140	90

a What weight should be selected from each field to give a representative sample of one tonne?

b The purpose of the sample is to estimate the number of damaged potatoes in the whole crop.
What other method of sampling could be used?
Give a reason for your choice.

5 Andy is a pupil at a coeducational school. He is doing a survey into the pupils' attitudes to wearing school uniform in his school.
How should he choose his sample so that his results are representative? Give two reasons for your choice.

6 A food producer has three machines that each fill 100 ml cartons with apple juice.
Machine A fills 200 carton an hour, Machine B fills 150 cartons an hour and Machine C fills 180 cartons an hour.

a Each hour a sample of 5% of all the cartons filled are examined to check the quantity of juice in each carton. How many cartons are examined?

b How many cartons from each machine should be examined to give a stratified sample?

c Explain why a stratified sample is necessary.

7 A company employs 70 manual workers, 25 middle managers and 5 directors.
The finance director needs to choose a sample of 15 of these employees.

a Explain why choosing a random sample of 5 employees from each group is likely to give biased results.

b Calculate the number that should be chosen from each group for a stratified sample.

8 Mrs Smith needs to choose a 10% sample of pupils from Year 10 students in her school.
Mrs Smith divides the pupils into these groups.

Ability Band 1		Ability Band 2		Ability Band 3	
Boys	Girls	Boys	Girls	Boys	Girls
30	40	32	48	12	15

a Explain why these groups may introduce bias into the results.

b Mrs Smith chose a random sample of 4 girls and 4 boys from Ability Band 2. Explain why this does not give a representative sample.

b Calculate the number that should be chosen from each group for a stratified sample.

9 A supermarket sells coffee some of which is fairtrade coffee. The fairtrade coffee usually costs more than the other coffee.
The management wants to conduct a survey into the customers opinions of fairtrade coffee.
Give two factors that should be considered when designing a stratified sample. Explain why you think these factors are important.

Photo of fairtrade coffee grower or someone harvesting coffee beans.

10 A catering company wants to survey the eating habits of pupils in a school. How should they choose their sample to get reliable results?
Give reasons for your decisions.

2.4 Types of data

There are two main types of data.

Qualitative data is data that can only be described in words.

The gender and colour of a group of cats are examples of qualitative data.

Quantitative data is data that can be described using number values.

Height and shoe size are examples of quantitative data.

There are two types of quantitative data.

Discrete data has only certain and exact values.

Shoe sizes are an example of discrete data. They have exact, separate values like 39, 40, $40\frac{1}{2}$. There is no shoe size between 40 and $40\frac{1}{2}$.

Continuous data has values anywhere in a range.

Heights are an example of continuous data. A normal adult's height can be anywhere between 1 m and 2.5 m, so a height can be any value in that range.
A height cannot be measured exactly.

Class discussion

Information is collected about each of the following.
Is the data qualitative, quantitative and discrete or quantitative and continuous?

1 Whether the apples on a tree are eating or cooking apples or both.
2 The weight of apples harvested from a tree.
3 The number of apples harvested from a tree.
4 The time spent watching television each day.
5 The type of programmes watched on television.
6 The lengths of pencils.
7 The lowest temperatures reached each night.

2.5 Designing a data collection sheet

Fig A/W three coins

Before you start to collect data, you need to design a table on which you can record the information.

The form of the table depends on the data you are going to collect.

You can use a simple three column table to record data such as examination grades or the numbers of heads when three coins are tossed.
The first column lists the values you are going to count,
The middle column is used to make a mark each time a value occurs. This is called a **tally**.
The third column is used for the total number of tally marks for each value. This is called the **frequency**.

The completed table is called a **frequency table**.

This table can be used to record the results when three coins are flipped.

Number of Heads	Tally	Frequency
0		
1		
2		
3		
	Total	

Data that has a large number of values, such as examination marks, can be grouped.
When you need to group data, use between three and ten groups.
Make the groups the same size when you can. You may have to make one group slightly larger than the others.

This table can be used to record the marks out of 40 in an examination.

Number of Heads	Tally	Frequency
0		
1		
2		
3		
	Total	

To collect continuous data, you need to decide on a range that will cover all possible values of the data.
The heights of men are likely to be in the range 1 m to 2.5 m.
You can break this range into smaller sections.
These sections are called **class intervals**.

Fig A/W Group of men standing beside a height measure.

The class intervals must cover all possible heights between 1 m and 2.5 m with no gaps.
You can break the range into intervals of width 0.5 m.
You need to be careful how you write these intervals.

If you write them as 1–1.5
1.5–2
2–2.5
you have included 1.5 m and 2 m twice.

The symbol $\leqslant$ means 'less than or equal to' and the symbol $<$ means 'less than'.

You avoid this by starting the first class interval at 1 m then go up to, but do not include, 1.5 m. This class interval is written as $1 \leqslant h < 1.5$, where h metres stands for any height.

The next class interval then starts at 1.5 m and go up to, but does not include 2 m. This is written as $1.5 \leqslant h < 2$.
This makes sure that there are no gaps between the intervals. It also makes sure that there is no overlap.

Example 5

Sam wants to collect the heights of some men.
a Design a table for Sam to record these heights.
b In which class interval would a height of 1.5 m be recorded?

a

Height, h metres	Tally	Frequency
$1 \leqslant h < 1.5$		
$1.5 \leqslant h < 2$		
$2 \leqslant h < 2.5$		

Any height from 1 m up to but NOT including 1.5 m goes here.

This is where a height of 1.5 m goes.

These are the class intervals.

b $1.5 < h < 2$

Exam practice 2D

1 Design a table to record the number of children in a sample of families.

You do not know what the largest number of children is going to be. Add a few blank rows to the table to cope with this.

2 A sample of people were asked to choose a colour from red (R), yellow (Y), green (G), pink (P) or blue (B).

a Design a table to record the information.

b Use this list of choices to complete your table.

Y	R	R	Y	R	P	B	G	R	B	R	B
G	R	B	G	B	G	G	R	B	R	B	R
B	Y	P	B	P	Y	B	Y	G	R	G	B

The first column will list the colours. Don't forget to add a column for frequencies and include the total.

3 A cinema has 50 seats.
Design a table to record the number of seats sold for a sample of performances.

When you are marking tallies, do not go through the list looking for all the Rs, thenYs etc. Work systematically down the columns (or across the rows).

4 A general knowledge quiz is marked out of 30.

a Design a table to record the marks of a sample of adults taking this quiz.

b Use these marks to complete your table.

12	13	18	25	23	17	23	27	17	15
12	19	29	30	24	2	7	6	22	11
5	11	15	26	18	17	9	26		

5 This list gives the weights of some books. The weights are in grams to the nearest gram.

45	52	71	95	51	38	44	62	80	72
81	39	48	60	66	75	36	63	92	108
5	115	95	91	56	72	61	97	110	88

a Copy and complete this frequency table.

Weight, w g	Tally	Frequency
$20 \leqslant w < 50$		
$50 \leqslant w < 80$		
$80 \leqslant w < 110$		
$110 \leqslant w < 150$		

b How many books weighed less than 80 g?

6 This table shows the weights of a sample of women.

Weight, w g	Frequency
$40 \leqslant w \leqslant 60$	36
$60 \leqslant w \leqslant 80$	108
$80 \leqslant w \leqslant 100$	52
$100 \leqslant w < 120$	14

a Write down the number of women whose weight is in the class interval $60 \leqslant w < 80$.
b One of the women in this sample weighs 80 kg.
Write down the class interval in which her weight is recorded.
c How many women were in the sample?
d **i** How many of the sample of women weighed over 100 kg?
ii What proportion of the sample weighed over 100 kg?

You can give the proprotion as a fraction or as a decimal.

7 A data logging machine recorded the weights of bottles of water as they came off filling machines.
The weights were between 930 grams and 1020 grams.
Design a frequency table for these weights.

8 These are the lengths of some planks.
246 cm, 115 cm, 80 cm, 150 cm, 205 cm, 115 cm, 90 cm, 112 cm, 95 cm, 185 cm, 100 cm, 220 cm, 110 cm, 190 cm.
Design and complete a frequency table for these lengths.

Photo of a sawmill.

2.6 Two-way tables

A **two-way table** is used to collect two sets of related information .

A two-way table has the values for one set of information across the top and the values for the other set down the side.

Example 6

Javid wants to collect information about house ownership among men and women.
The two sets of information he wants are:

1 Does a person own a house or not?

2 Is the person male or female?

Design a data collection sheet for Javid.

	Owns a House	Does Not Own a House
Male		
Female		

A man who owns a house would be recorded here.

A two-way table can also be used to record grouped data.
This two-way table is designed to record the salaries and ages of some people.

This means £50 000

This means £0 up to, but not including, £10 000

		Salary					
		£0 –	£10 000 –	£20 000 –	£30 000 –	£40 000 –	£50 000 –
Age, y years	$20 \leq y < 30$						
	$30 \leq y < 40$						
	$40 \leq y < 50$						
	$50 \leq y < 60$						
	$60 \leq y < 70$						

A person between 30 and 40 years old aand earning between £20 000 and £30 000 a year is recorded here.

As you collect the information, you can add the tally marks. Then you can redraw the table and put in the totals.

Exam practice 2E

For questions 1–6 design a two-way table to record the information described.

1 Some men and women are asked whether or not they can swim.

2 Internet connections can be either direct dial or broadband. Some businesses and home users are asked which type of internet connection they have.

3 The number of bathrooms and the number of bedrooms in a sample of houses.

4 The gender (male or female) and the number of credit cards owned by a sample of people.

5 The weights and heights of children starting primary school. Start the heights at 90 cm and the weights at 20 kg.

6 The flowers on geraniums are either white, pink or red. Design a two-way table to record the colour of flowers and the heights of geranium plants three months after planting. Assume the heights start at 21 cm and stop at 60 cm.

7 Diran wants to test the hypothesis 'the number of television sets in a house is related to the number of rooms in the house.' Design a two-way table on which Diran can record the number of rooms and the number of televisions in a house.

8 Tomatoes grown in a greenhouse are either cherry tomatoes, plum tomatoes or regular tomatoes. The weight of tomatoes from each plant is to be recorded. Design a two way table to record this information. Assume the weights are in the range 0 to 3 kg.

You cannot legally be in full-time work until you are 16.

9 Design a two-way table to record the ages of men and of women in full time work.

You cannot legally be in full-time work until you are 16.

2.7 Questionnaires

A **questionnaire** is a form that needs answers written on it.

Questionnaires are useful when you want to get opinions. They make sure that everyone is asked the same questions. But you need to be careful what questions you ask.

Ask questions that will be answered, and answered truthfully.
Add a box for the answer - this is called the **response**.

You may need more than one response. It is usually best to give several answers so that people can tick one of them.

How old are you?

☐ years

This is a personal question;
some people will refuse to answer this,
some people will lie.
Avoid questions like this.

If you want this information, ask people to tick one of a range of ages.

How old are you?

☐ Under 18 years
☐ 18 to 30 years
☐ Over 30 years

- Ask open ended questions that do not expect a certain answer.

Don't you agree that cars should be banned from the town centre.

Yes ☐ No ☐

This is a leading question;
it is worded to expect the answer yes.
It is not suitable.

This is a better question. It does not lead people to any particular answer. It also gives the option of not having to answer Yes or No.

Should cars be banned from the town centre

Yes ☐ Don't know ☐ No ☐

- Ask questions where the answers expected are obvious.

Where are you in the town centre?

This question is not clear. It isn't obvious what information is wanted.

This is clearer. It gives the information that is wanted.

Why are you in the town centre?
You can tick more than one box.

☐ Work ☐ Shopping
☐ Meet friends ☐ Eating out
☐ Entertainment ☐ Live here
☐ Personal busines

- Write the responses so their meaning is exact.

How long do you expect to be in the town centre?

- ☐ A short time
- ☐ An average amount of time
- ☐ A long time

← The responses are not clear. They mean different things to different people.

The responses asked for here are clear and obvious. →

How long do you expect to be in the town centre?

- ☐ Less than 2 hours
- ☐ Between 2 and 6 hours
- ☐ More than 6 hours

When you design a questionnaire, you cannot be sure that the questions are free of problems. Always try them out on a few people so that you can find problems and correct them before you start to collect the information.

Exam practice 2F

1 A questionnaire contained this question.

What is your shoe size?

Why is this question likely to give problems?

2 A questionnaire contained this question.

Don't you agree that children should always walk to school?

Explain why this question is not suitable for a questionnaire.

3 A questionnaire contained this question.

Question: How old are you?

Answer:

Young	Middle aged	Old
☐	☐	☐

a Explain why the question is likely to give problems.
b Give one reason why the headings on the boxes are not suitable.

4 A questionnaire contained this question.

How did you do in your maths homework yesterday?

This question can have more than one meaning.
Write down two possible meanings.

5 Rachael wants to carry out a survey into what people eat at lunch time.
She designed this question to include in her questionnaire.

How much do you eat at lunch time?

a Explain why this question is not clear.
b Design a better question for Rachael to ask.
Include three responses.

6 A survey was carried out on attitudes to making the Town Centre a pedestrian only area.
The questionnaire used asked these questions:
Do cars cause pollution?
Is crossing the roads dangerous because of the traffic?
Should the Town Centre be made pedestrian only?
Explain why you think that the three questions *taken together are unsuitable*.

Mini coursework task

Design a questionnaire to get the information needed for the survey described in the Class Discussion on page 10. Include questions that are needed to place customers in the categories required for any stratified samples that are necessary.

Class discussion

If you are doing a coursework task that needs a questionnaire to be filled in, you need to decide how to carry out your survey.

You can interview each member of your sample and fill in the answers yourself. This ensures that questions are answered sensibly.
However it can take too much time and there may be some people who refuse to answer some of the questions

Alternatively, you can give the questionnaires to each member of your sample, ask them to fill them in and return them to you.
This saves you time, but there are disadvantages.

Discuss ways of dealing with the following problems:
- questionnaires that are not returned
- missing answers
- nonsense answers.

How old are you?
0–10 11–14 15–20
✓ ✓

How tall are you? 2.5 metres

What is your weight ____ kg

Summary of key points

- You cannot always get all the information you need. When you can only use part of the possible information you must try to make sure that it is representative.
- A simple random sample is chosen such that every item that could be chosen has an equal chance of being selected.
- A stratified random sample is where all the items are divided into categories and a random sample is chosen from each category. The size of the sample in each category is the same fraction of that category as the total sample is of the whole set. This ensures a more representative sample.
- Data can be qualitative or quantitative.
- Quantitative data can be discrete or continuous.
- Information can be grouped.
- Data collection sheets can be designed to record information.
- A two-way table can be used to record two related sets of information.
- Questionnaires are used to collect several pieces of information.

Most students who get GRADE C or above can:

- design a question with responses for a questionnaire,
- give reasons why a method of sampling is unrepresentative.

Most students who get GRADE A can also:

- explain why a method for choosing a sample does not give a random sample,
- explain how to find a stratified sample.

Glossary

bias	not chosen by chance
class interval	range for continuous data
continuous	able to have any value within a range
discrete	having exact and separate values
frequency	the number of times that a value (or group of values) occurs
frequency Table	a table listing the possible values, or groups of values, and their frequencies
population	all the possible items being investigated
qualitative	can only be described using words
quantitative	can be described using numbers
questionnaire	a series of questions used to get several pieces of information
random sample	every item that can be selected has an equal chance of selection
representative	reflects the makeup of all items
response	answer to aa question on a questionnaire
sample	a selection of items
stratified sample	a sample drawn from groups of the population
tally	a mark used for counting
two-way table	a table containing two sets of related information
unrepresentative	not reflecting the makeup of all items

3 Representing data

This chapter will show you:
- ✓ how to draw and interpret a bar chart
- ✓ how to draw and interpret a frequency polygon
- ✓ how to draw and interpret a stem-and-leaf diagram
- ✓ how to construct and interpret a histogram

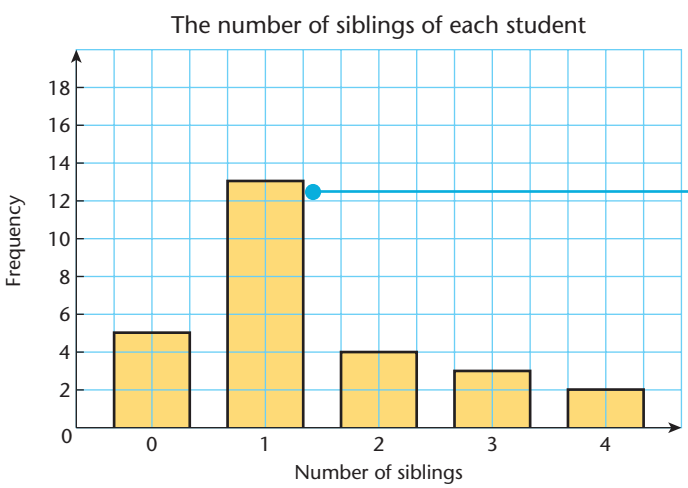
Before you start you need to know:
- ✓ how to scale axes and plot points
- ✓ the meaning of fractions and percentages
- ✓ how to work with decimals
- ✓ how to find the area of a rectangle

3.1 Bar charts and frequency polygons

Bar charts are used to represent data from frequency tables.
A bar chart should have a title to tell you what it represents.

Example 1

This bar chart shows the number of siblings of each student in a group.
How many students have 1 sibling?



Sibling means brother or sister.

This bar comes up to 13 on the frequency scale. It is standding on 1 sibling. This means that there are 13 students who have one brother or sister.

When you want to get information from a bar chart, read the title and the labels on the axes carefully to make sure that you understand what information is shown.

13 students have one sibling.

When you draw a bar chart
- give it a title
- label the axes
- make each bar the same width.

Bar charts can be vertical or horizontal. The bars may be touching or may be lines.

Frequency polygons are also used to represent data in a frequency table. Points are plotted to show the **frequency** of each category. The points are joined with straight lines.

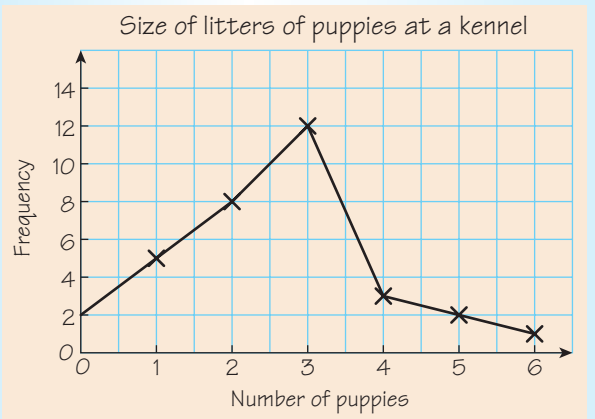
Example 2

Photo: a litter of puppies
The number of puppies per litter born at a kennel was recorded. This frequency table shows the results.

Ph 301 - a litter of puppies

Number of puppies	0	1	2	3	4	5	6
Frequency	2	5	8	12	3	2	1

Draw a frequency polygon to show these results.

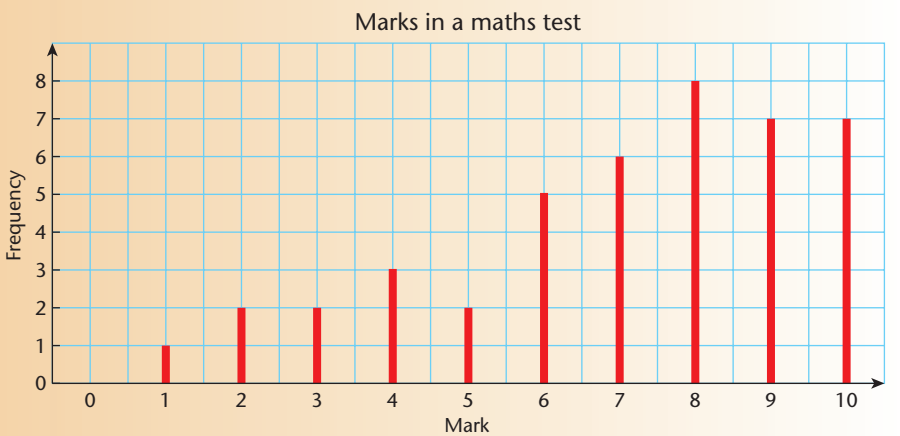


The frequencies are always shown on the vertical axis. The frequencies go from 1 to 12, so scale the axis from 0 to 14.

Treat each number of puppies and its frequency as a pair of coordinates.
Plot these coordinates on a grid and join the points in order with straight lines.

Exam practice 3A

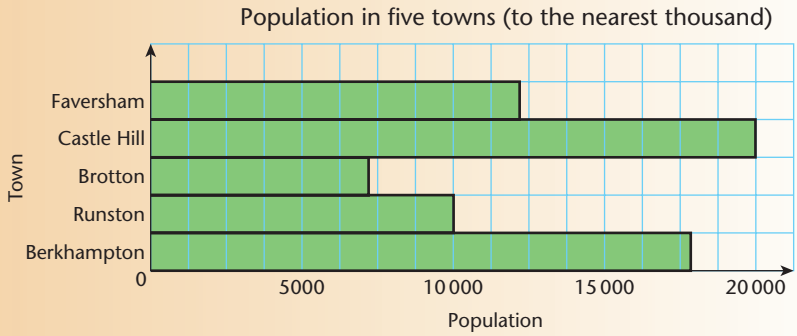
1 This bar chart shows the results of a maths test.



a Write down how many students had a mark of 8.
b What was the lowest mark given and how many students got it?
c Write down the most common mark.
d How many students had a mark of 7 or more?
e How many students sat the test?

'7 or more' means 7, 8, 9 or 10.

2 This bar chart shows the population in five towns.



a Write down the town with the largest population.
b Write down the population of Runston.
c i Which town has the smallest population?
ii Write down the population of this town.

3 Peter did a survey of the types of vehicle passing the school gate one lunch hour. His results are shown in this table.

Type of vehicle	Bicycle	Motorbike	Car	Lorry
Frequency	4	10	25	16

a Calculate the number of vehicles that passed the school gate altogether.
b Which type of vehicle was the most common?
c Draw a bar chart to show the information.

Remember that the bars must have equal widths.

Do not forget to give your chart a title.

4 This table shows the number of goals scored by 32 teams one weekend.

Number of goals scored	Frequency
0	5
1	9
2	10
3	5
4	2
5	1

Photo of football match

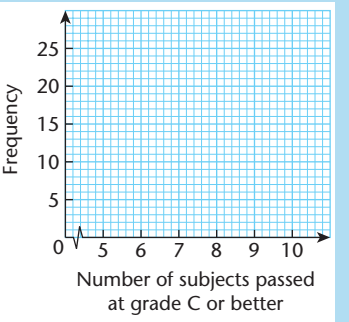
a How many teams scored more than 1 goal?
b Draw a frequency polygon for this data.

5 This table shows the number of GCSE subjects passed at grade C or better by a group of students.

Number of subjects passed	5	6	7	8	9	10
Frequency	14	18	21	12	7	5

a How many students achieved a Grade C or better in 7 or more GCSE subjects?
b Draw a frequency polygon for this data.

Use a grid like this. The zig-zag line shows that the horizontal axis does not start at 0.



6 This table shows the shoe sizes of the girls in a group.

Shoe size	3	$3\frac{1}{2}$	4	$4\frac{1}{2}$	5
Frequency	3	5	10	8	4

a Write down the number of girls in the group.
b Draw a frequency polygon for this data.

Mini coursework task

Kerry said 'Fewer babies were born in January, February and March than in September, October and November'.

a Write down the information you need to test Kerry's hypothesis.
b Collect the information from at least 30 people.
c Show your information on a bar chart.
d Is Kerry right? Give a reason for your answer.

3.2 Frequency polygons for grouped data

A **frequency polygon for grouped data** uses the mid-interval value for each group.

You find the mid-interval value of a group by adding the end values of the group then dividing by 2.

Example 3

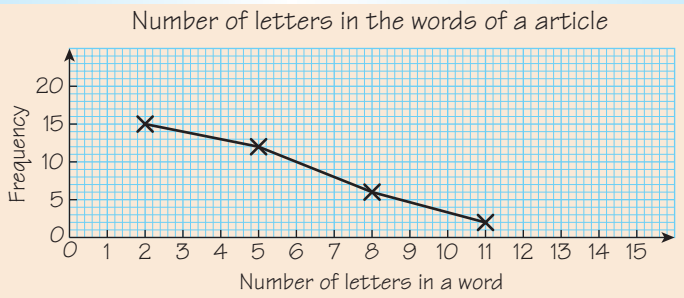
This **grouped frequency table** shows the numbers of letters in the words of a paragraph from a newspaper.

Draw a frequency polygon for this data.

Before you can draw a frequency polygon for this table, you need to find the value in the middle of each group. Add a column to the table for these values.

Number of letters in a word	Frequency	Mid-interval values
1–3	15	2
4–6	12	5
7–9	6	8
10–12	2	11

2 is half-way between 1 and 3



Treat the mid-interval value and the frequency as coordinates to plot the points on the grid. Join the points with straight lines.

Exam practice 3B

1 This table records the number of passengers on the buses that passed a survey point one day.

Number of passengers	Frequency	Mid-interval value
1–12	21	6.5
13–24	26	18.5
25–36	14	
37–48	16	
49–60	5	

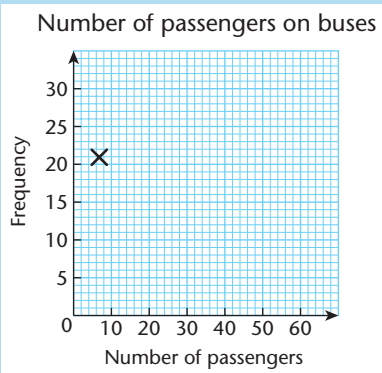
a Copy and complete the table.

b Calculate the number of buses that passed the survey point.

c Draw a frequency polygon to show this data.

The mid-interval value does not have to be a whole number of passengers. 18.5 is halfway between 13 and 24 so it is the mid-interval value.

Use a grid like this. The first point is plotted for you.



2 This table shows the number of competitors in a singing competition.

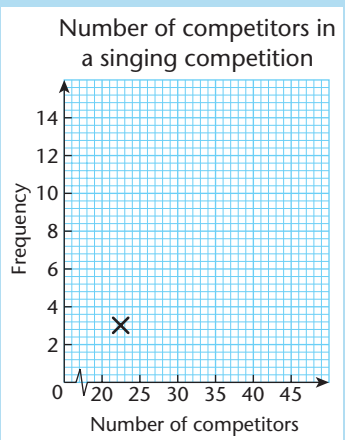
Number of competitors	Frequency
20–25	3
26–30	2
31–35	12
36–40	9
41–45	5

a Copy and complete the table.

b Calculate the number of times the competition has taken place.

c Draw a frequency polygon to show this data.

Use a grid like this. The first point is plotted for you.



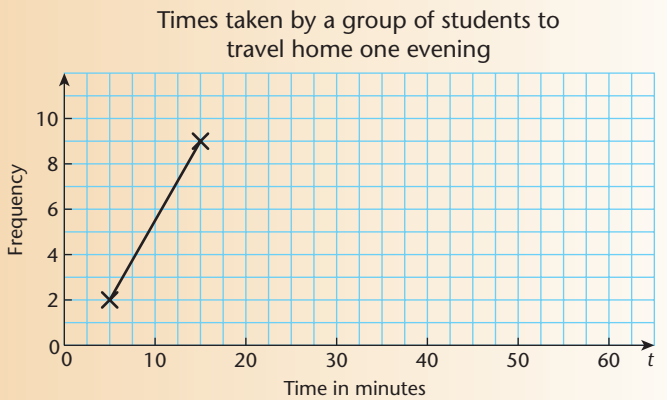
3 This frequency table shows the times taken by a sample of students to travel home one evening.

Time, t minutes	Frequency
$0 \leqslant t < 10$	2
$10 \leqslant t < 20$	9
$20 \leqslant t < 30$	5
$30 \leqslant t < 40$	4
$40 \leqslant t < 50$	2
$50 \leqslant t < 60$	1

This is continuous data. To find the mid-interval value, add the values of the ends of the class, then divide by 2.

a Copy the table and add a column showing the mid-interval values.

b Copy and complete this frequency polygon.



4 Kim recorded the lengths of 50 pea pods in this table.

Length, l cm	Number of pods
$5 \leqslant l < 7$	8
$7 \leqslant l < 9$	14
$9 \leqslant l < 11$	19
$11 \leqslant l < 13$	9

Draw a frequency polygon to illustrate this data.

5 This frequency table shows the heights of a batch of tomato seedlings.

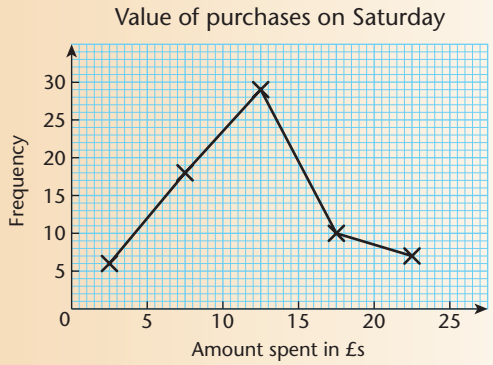
Height, h cm	$0 \leqslant h < 1$	$1 \leqslant h < 2$	$2 \leqslant h < 3$	$3 \leqslant h < 4$	$4 \leqslant h < 5$
Frequency	5	10	25	4	1

a Copy the table and add a row showing the mid-interval values.

b Draw a frequency polygon to show this data.

The mid-interval values are not whole numbers.

6 This frequency polygon shows the amounts of money spent in a butcher's shop by the customers one Saturday.



The table shows the amounts spent by customers on Wednesday.

Amount spent, A pounds	Number of customers
$0 \leqslant A < 5$	8
$5 \leqslant A < 10$	11
$10 \leqslant A < 5$	15
$15 \leqslant A < 20$	8
$20 \leqslant A < 25$	4

a Calculate the number of customers on Saturday.

b Calculate the number of customers on Wednesday.

c Explain why you cannot give the number of customer who spent between £10 and £12 on Wednesday

d Copy the frequency diagram above and draw a frequency polygon for the amounts spent by customers on Wednesday.

7 This table shows the times taken by a group of students to complete a task.

Time (t minutes)	Number of students
$10 \leqslant t < 20$	5
$20 \leqslant t < 30$	25
$30 \leqslant t < 40$	20
$40 \leqslant t < 50$	10
$50 \leqslant t < 60$	7
$60 \leqslant t < 70$	3

Draw a frequency polygon for these times.

3.3 Stem-and-leaf diagrams

A **stem-and-leaf diagram** is a way of grouping data without losing the detail of individual values.

A stem-and-leaf diagram is sometimes called a stem plot.

Getting information from a stem-and-leaf diagram

This stem-and-leaf diagram shows the number of pages in a sample of books.

Number of pages

1 | 20 34 41 92
2 | 9 10 25 36 71 77 86
3 | 10 24 63

1 | 41 means 141

These numbers form the stem. They are like the groups in a frequency table.

This is the key. It tells you that the 1 in the stem is 100 and the leaf, 41, is added to give the value.

These numbers are the leaves.

Example 4

Use the stem-and-leaf diagram above to
a write down the number of books in the sample,
b give the number of books that have more than 300 pages.

a 14
b 3

Count the number of 'leaves' to the right of the vertical line. There are 4 in the first row, 7 in the second and 3 in the third.

The last line shows all books with more than 300 pages. There are 3 entries.

Exam practice 3C

1 The weights of the packages to be delivered by a courier are given in this stem-and-leaf diagram.

Weight

0 | 93 95 98
1 | 8 22 24 25 36 73 83
2 | 14 44 52 88 99

1 | 22 means 122 g

a Find the number of packages the courier has to deliver.
b Write down the weight of the lightest package.
c Write down the number of packages weighing more than 200 g.

2 Cars were tested to see how long they took to reach 100 mph from a standing start.
This stem-and-leaf diagram shows the results.

Time in seconds

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

2 | 3 means 23 seconds

a Work out the number of cars tested.
b Write down the slowest time.
c Write down the fastest time.
d Find the number of cars that took more than 30 seconds.

3 The times taken by the competitors in a 1 km race were recorded. They are shown in this stem-and-leaf diagram.

Time to complete the race

3	21 23 25 37 45 51
4	3 4 4 2 15 34 42 53
5	1 9 14 16
6	0

3 | 21 means 3 minutes 21 seconds

ph 303
Pupils running on a trackj

a Write down the number of competitors who finished.
b Write down the time taken by the person who was last.
c How many times were less than $3\frac{1}{2}$ minutes?
d Find the difference in the time taken by the person who was first and the one who was last.

Drawing a stem-and-leaf diagram

When you draw a stem-and-leaf diagram, you must include
- a key
- a title.

Example 5

This list gives the number of cars in a car park at noon on twenty consecutive Saturdays.

104 99 107 102 115 102 108 98 110 106
108 95 94 118 95 105 114 102 113 97

Draw a stem-and-leaf diagram to represent this data.

Decide on the groups to use for the stem. The numbers are all between 90 and 120, so we will use tens for the stem and units for the leaves.

Number of cars in car park

9	9 8 5 4 5 7
10	4 7 2 2 8 6 8 5 2
11	5 0 8 4 3

Start working across the first row marking each value in the correct place.
Do not try to put them in order at this stage.
Check that you have 20 values.

Number of cars in car park

9	4 5 5 7 8 9
10	2 2 2 4 5 6 7 8 8
11	0 3 4 5 8

11 | 4 means 114

Redraw the diagram with the numbers in order of size. Give a key and a heading.
Check again that you have 20 values.

Exam practice 3D

1 A maths test was marked out of 40.
This stem-and-leaf diagram shows the marks scored by students in the test.

Marks

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

1|4 means 14

a Calculate the number of marks shown in the diagram.
b Write down the lowest mark.
c Two other students got marks of 36 and 29.
Copy the diagram and add these two marks in the correct place.

2 Mina timed her journey to school on seventeen days.
These times are shown in this stem-and-leaf diagram.

Ph 301 - girl cycling to school

Time in minutes

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

2|5 means 25 minutes

The next three days her times were 17 minutes, 19 minutes and 27minutes.

You can use your new diagram for these questions.

a Copy the diagram and add these three values to the stem-and-leaf diagram.
Rewrite your stem-and-leaf diagram so that all the values are in order.
b Write down the number of times the journey took more than 20 minutes.
c Write down the number of times the journey was less than 25 minutes.
d Write down the number of times that the journey was longer than half an hour.

3 A school library bought some new books. This is a list of their prices.

£4.50 £6.20 £5.40 £4.90 £5.70 £4.39
£6.45 £5.95 £4.75 £5.80

Enter the prices one at a time in order. First £4.50, then £6.20, then £5.40 and so on. Do not look for all the £4 books first.

a How many new books were bought?
b Draw a stem-and-leaf diagram to show these prices.

4 These are the recorded playing times, in minutes, of some CDs.

Ph 302 - some CDs

61 37 50 50 59 72 41 47 68 59 59 69 42
59 66 60 61 65 50 55 52 61 45 74 68 39

a How many CDs were recorded?
b Draw a stem-and-leaf diagram to show these times.

5 These are the weights in grams, of some apples.

74 48 102 96 100 85 73 105 59 115 77 82 94 81
79 60 100 89 110 58 62 101 71 73 49 57 76 89

Draw a stem-and-leaf diagram for these weights.

6 The number of mobile telephone calls received by a group of students one week were :

32, 6, 18, 55, 9, 19, 7, 6, 17, 4, 11, 18, 14, 13, 16, 35, 0

a How many students were there?
b Draw a stem-and-leaf diagram for these numbers.

3.4 Histograms

A **histogram** is a diagram that represents continuous data.
A histogram is similar to a bar chart but has two important properties.

The area of each bar represents the frequency of the **class interval**, so frequency = **class width** × height.

Height of the bar = $\frac{\text{frequency}}{\text{class width}}$

There are no gaps between the bars.
The horizontal axis is like a number line. It is scaled evenly along its length.
The vertical axis must start at 0.

The fraction $\frac{\text{frequency}}{\text{class width}}$ is called the **frequency density**.

Example 6

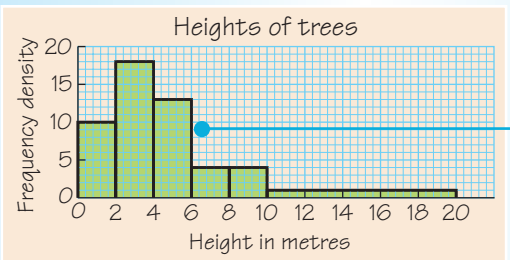
This table gives the heights of trees in a wood.

Height, h metres	$0 \leqslant h < 2$	$0 \leqslant h < 4$	$0 \leqslant h < 6$	$0 \leqslant h < 10$	$0 \leqslant h < 20$
Frequency	20	36	26	16	10

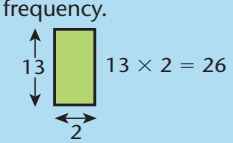
Draw a histogram to show these heights.

Height, h metres	$0 \leqslant h < 2$	$0 \leqslant h < 4$	$0 \leqslant h < 6$	$0 \leqslant h < 10$	$0 \leqslant h < 20$
Frequency	20	36	26	16	10
Class width	2	2	2	4	10
Frequency density	10	18	13	4	1

Add rows to the table for the class widths and the frequency densities.
Frequency density = $\frac{\text{frequency}}{\text{class width}}$



The area of the bars represents the frequency.



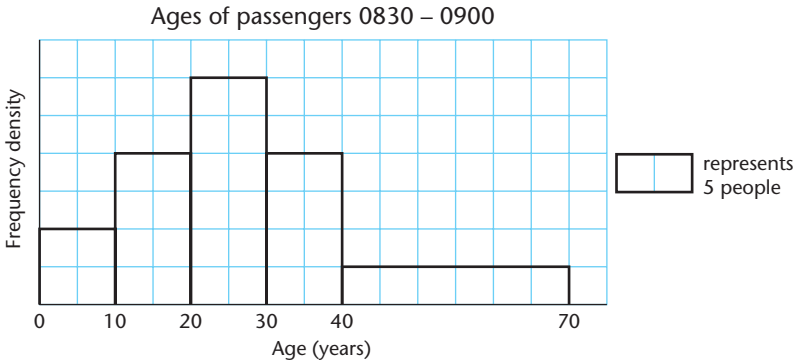
$13 \times 2 = 26$

The frequency for the class $4 \leqslant h < 6$ is 26.

You can construct a frequency table from a histogram provided that it is properly labelled.

Example 7

This histogram shows the ages of people boarding busses between 0830 and 0900 one morning.



The key tells you that 2 squares represents five people.
To work out the frequencies you need to count the squares.

Make a frequency table.

Age, n years	$0 \leqslant n < 10$	$20 \leqslant n < 20$	$20 \leqslant n < 30$	$30 \leqslant n < 40$	$40 \leqslant n < 70$
Frequency	20	20	30	20	15

Exam practice 3E

1 The frequency table shows the ages of 200 people attending a film show.

Age, n metres	$0 \leqslant n < 10$	$20 \leqslant n < 20$	$20 \leqslant n < 30$	$30 \leqslant n < 40$	$40 \leqslant n < 70$
Frequency	20	20	30	20	15

Copy the table and add rows showing the class widths and frequency densities.

Remember to label the vertical axis 'frequency density'

For questions **2** to **4**, copy and complete the table and draw a histogram to illustrate the data

Scale the vertical axis from 0 to 2 in steps of 0.1.

2 The table shows the heights of 65 fourteen-year-old boys.

Age, n metres	$100 \leqslant n < 130$	$130 \leqslant n < 140$	$140 \leqslant n < 145$	$145 \leqslant n < 155$	$155 \leqslant n < 165$	$165 \leqslant n < 185$
Frequency	6	11	8	14	14	12
Class width						
Frequency density						

3 The table shows the times taken by 78 pupils to get t school.

Time, t min	$0 \leqslant t < 5$	$5 \leqslant t < 10$	$10 \leqslant t < 15$	$15 \leqslant t < 18$	$18 \leqslant t < 21$	$21 \leqslant t < 31$
Frequency	6	1	18	9	12	13

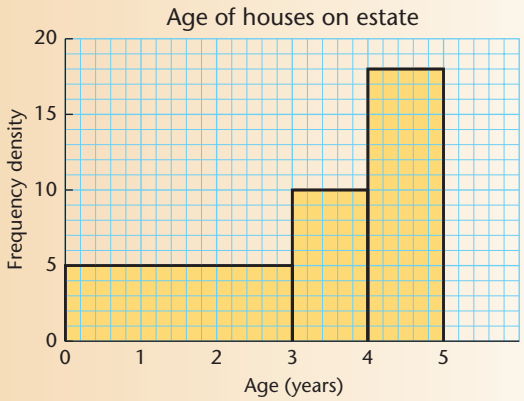
4 The table summarises the results of a survey to find the areas of 185 farms.

Age, A ha	$0 \leqslant A < 4$	$4 \leqslant A < 8$	$8 \leqslant A < 12$	$12 \leqslant A < 17$
Frequency	34	22	38	30

Age, A ha	$17 \leqslant A < 24$	$24 \leqslant A < 28$	$28 \leqslant A < 32$	$32 \leqslant A < 42$
Frequency	28	13	10	10

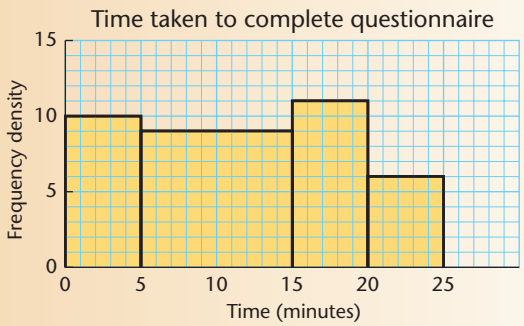
ha means hectares.

5 This histogram shows the ages of houses on a new estate.



a How many houses are between 0 and 3 years old?
b How many houses were more than 3 years old?

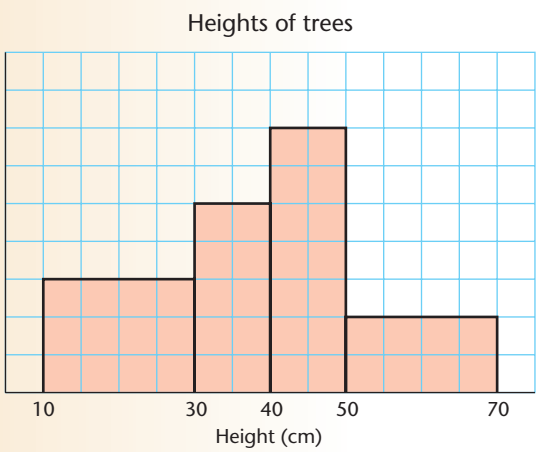
6 This histogram shows the times taken by 225 people to complete a questionnaire.



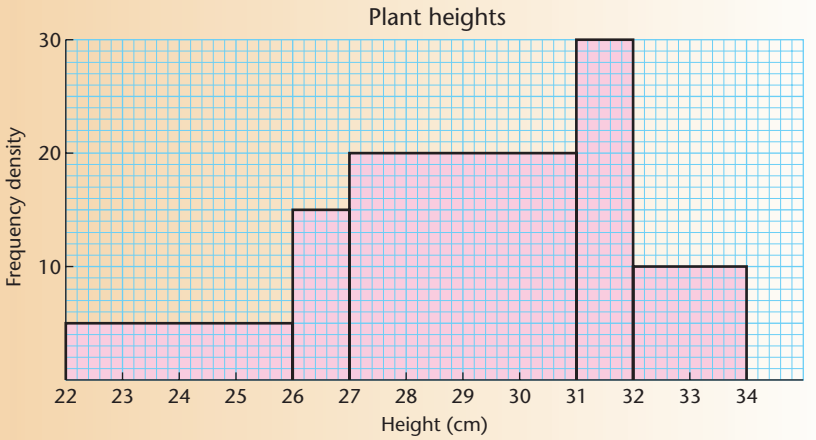
Draw a frequency table to represent this information.

7 This histogram shows the heights of recently planted trees.

There were a hundred trees between 30 and 40 cm tall.
How many trees were:
a between 50 and 70 cm tall
b shorter than 40 cm?



8 The histogram below shows the heights of plants grown by a horticulturalist.



How many plants in the sample are there

a with heights at least 31 cm but less than 32 cm?

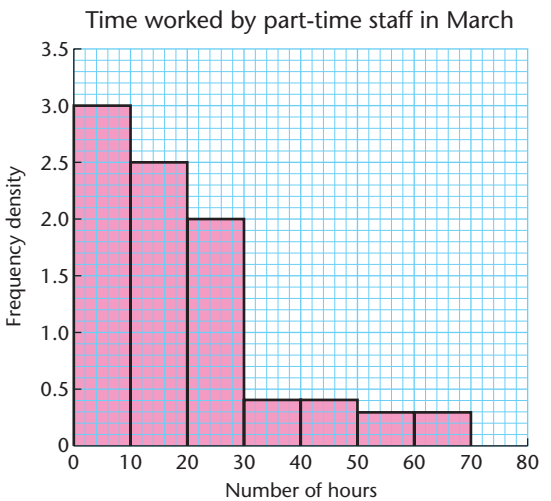
b with heights at least 27 cm but less than 31 cm?

3.5 Using the area under a histogram

The area of a bar in a histogram represents the number of items in the group. This means that the total area of the histogram represents the total number of items in the distribution.

Example 8

This histogram illustrates the number of hours worked in March by each part-time member of staff in one company.



a How many part-time staff are there?

b Estimate the number of part-time staff who worked between 45 and 60 hours in March.
Explain why you cannot give an exact answer.

a 89

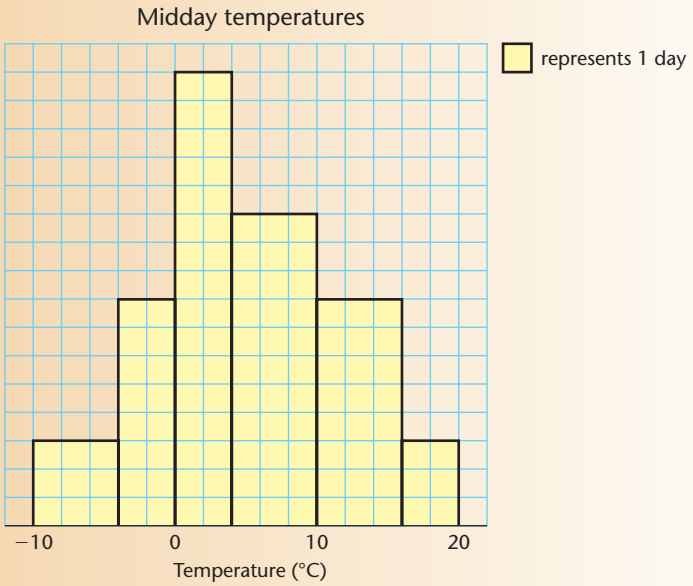
b Area of the histogram between 45 and 60 is
$(5 \times 0.4) + (10 \times 0.3) = 2 + 3 = 5$.
So approximately 5 people worked between 45 and 60 hours.
The number of people who worked 45 hours or any other number of hours is not known.

The first block represents 10×3 people, and so on.

An estimate for the number of people who worked between 45 and 60 hours is represented by the dark shaded area.

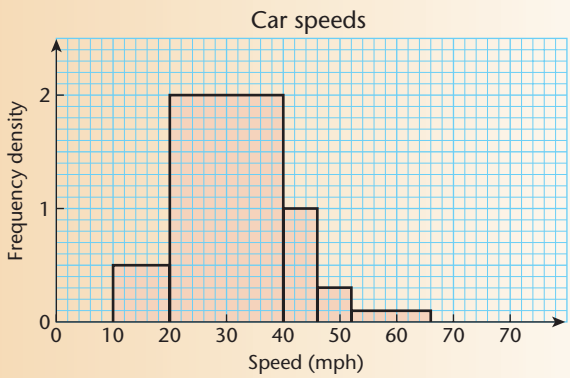
Exam practice 3F

1 This histogram shows the midday temperatures in degrees Celsius for the first 240 days of a year in Northampton.



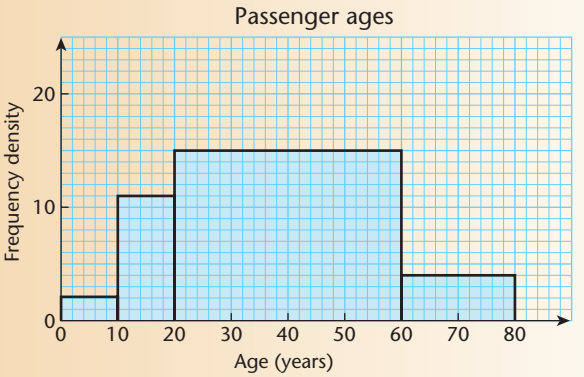
a Why does this chart not need a vertical scale?
b On how many days was the temperature less than 0°C?
c Estimate the number of days on which the temperature was greater than 12°C.

2 A speed camera records the speeds of cars as they enter a 30 mph speed limit zone. The histogram summarises the results for a one hour period.



Estimate the number of cars that were travelling at more than 30 mph.

3 This bar chart summarises the ages of passengers, in completed years, using a bus in the morning rush hour.

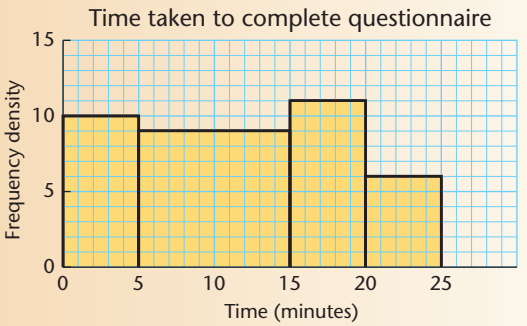


Estimate the number of passengers who are

a under 16 years old

b 65 years old or older.

4 The histogram summarises the times taken by a group of people to complete a questionnaire.



a Estimate the number of people who took less than 10 minutes to fill in the questionnaire.

b Tom said 'More than half the people took under 10 minutes to complete the questionnaire.'
Is Tom correct? Explain your answer.

Summary of key points

- The information in an ungrouped frequency table can be shown in a bar chart or frequency polygon.
- A frequency polygon for grouped data is drawn by plotting the mid-interval values against the frequencies and joining the points with straight lines.
- The mid-interval value is equal to the sum of the end values of the group or class divided by 2.
- Stem-and-leaf diagrams preserve individual values.
- Histograms are used to represent continuous data. The areas of the bars represent the frequencies of the classes.
- Frequency density $= \dfrac{\text{frequency}}{\text{class width}}$.

Most students who get Grade C or above can:

- find mid-interval values for grouped data,
- draw a frequency polygon for grouped data.

Most students who get grade C can also:

- construct a histogram with unequal class intervals.

Glossary

bar chart	a diagram illustrating data from a frequency table
class interval	the range of values of items in a group
class width	the difference between the highest and lowest value in a class interval
data	factual information
frequency	the number of times something happens
frequency table	shows the number of times that each distinct value (or class) occurs
frequency density	frequency divided by class width
frequency polygon	a polygon formed by joining the mid-points of the tops of the bars in a bar chart or histogram
histogram	a chart illustrating continuous data
mid-interval value	the middle value of a group or class
stem-and-leaf diagram	a diagram that groups data without losing individual values

4 Two-way tables and scatter graphs

This chapter will show you:
- ✓ how to get information from a two-way table
- ✓ how to draw a scatter graph
- ✓ how to judge whether there is a relationship between two sets of information

Before you start you need to know:
- ✓ how to scale axes and plot points on a graph
- ✓ how to read values from scaled axes

4.1 Two-way tables

Two-way tables are used to show two groups of related information.

Example 1

This **two-way table** shows the grades obtained by a group of students in two GCSE Modules.

GCSE Grades

	A*	A	B	C	D	E	F	G	U
Module 1	8	18	28	25	20	22	10	2	9
Module 3	4	6	7	15	5	3	2	0	1

See Chapter 2 for more information on two-way tables

a How many students got grade E in Module 1?
b How many students got grade D in Module 3?

Look along the row labelled Module 1 until you get to the column headed C.

Add up the numbers for both rows under the columns headed A*, A, B and C.

a 25
b 111

Exam practice 4A

1 Use the two-way table above to find
- a the number of students who sat Module 3
- b the number of students sitting Module 1, who got grades A* to B.

2 The table shows students in Year 10 with and without jobs.

Students in Year 10

	Job	No Job
Boys	32	25
Girls	28	21

Pic Boy or girll aged 15 doing paper round or working in a shop

- a How many girls have jobs?
- b How many boys do not have jobs?
- c How many students do not have a job?

3 The table shows last year's Key Stage 3 results in a school.

Key Stage 3 Levels

	3	4	5	6	7	8
Boys	8	12	28	25	16	5
Girls	10	12	27	30	21	5

Read questions slowly. Read them several times to check that you understand what you are asked to find.

a How many students were at Level 5?

b How many boys were at Level 6 or higher?

c How many girls were lower than Level 5?

4 The table compares students' grades in GCSE mathematics and English.

Mathematics

English	A*	A	B	C
A*	8	9	7	5
A	9	8	2	3
B	4	6	8	5
C	10	7	6	8

a How many students got a grade C in English?

b How many students got grade a B in both subjects?

c How many students got a grade B?

This means the number of students who got grade B in English or in mathematics. It includes those who got grade B in both subjects.

5 The table shows the number of shops and bars in villages on the South Coast.

Number of bars

Number of shops	0	1	2	3	4
0	1	4	1	0	0
1	1	2	3	1	0
2	0	1	1	0	0
3	3	1	2	0	1
4	1	4	3	2	1
5	0	2	4	2	2
6	0	1	7	4	2

The diagonal of shaded entries in the table below shows the numbers of villages that have the same number of shops and bars.
The numbers above this diagonal give villages with more bars than shops.

Number of bars

Number of shops	0	1	2	3	4
0	1	4	1	0	0
1	1	2	3	1	0
2	0	1	1	0	0
3	3	1	2	0	1
4	1	4	3	2	1

a How many villages have no shops?

b Calculate the total number of villages featured in the table.

c How many villages have more bars than shops?

6 The table shows the height and the numbers of tomatoes on each of a sample of tomato plants.

Height, h cm, of a plant	Number of tomatoes on a plant: 0–19	20–39	40–59	60–79
$0 \leqslant h < 20$	8	7	3	0
$20 \leqslant h < 40$	12	25	8	1
$40 \leqslant h < 60$	6	16	18	17
$60 \leqslant h < 80$	3	10	21	22
$80 \leqslant h < 100$	0	4	12	7

a Calculate the total number of plants featured in this table.

b Calculate the number of plants that had 40 or more tomatoes.

c How many plants less than 60 cm tall had 40 or more tomatoes.

7 The table shows the heights and weights of some adults.

Height, h m,	Weight, w kg: $60 \leqslant w < 70$	$70 \leqslant w < 80$	$80 \leqslant w < 90$	$90 \leqslant w < 100$
$1.4 \leqslant h < 1.6$	5	3	1	0
$1.8 \leqslant h < 1.8$	10	2	1	1
$1.8 \leqslant h < 2.0$	4	6	9	6
$2.0 \leqslant h < 2.2$	2	10	12	10

a How many adults weighed less than 80 kg?

b Calculate the number of adults under 2 m tall who weighed less than 80 kg.

8 The table shows the marks in a maths test and in a science test.

Science Mark	Maths Mark: 1	2	3	4	5
1	2	5	4	1	1
2	6	7	3	2	2
3	7	5	9	9	6
4	3	9	8	10	12
5	2	7	12	11	10

a Find the number of students who got a mark of 4 or more in the maths test.

b How many students got a higher mark in the science test than in the maths test?

Mini coursework task

Adam read that girls do better than boys at GCSE.
He decided to investigate this.

- Write down a hypothesis for him to test.
- Design a two way-table so that Adam can collect the information.

4.2 Scatter graphs

A **scatter-graph** is a way of illustrating information from a two-way table.

Example 2

The table shows the height and shoe size of eight women.

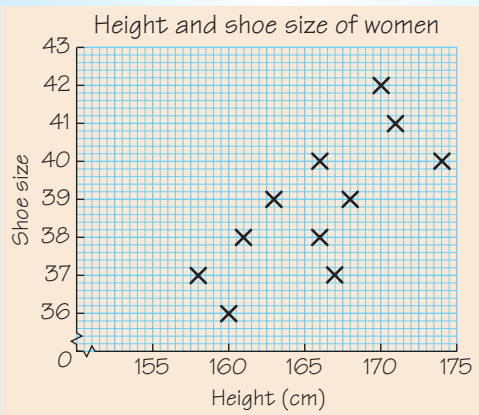
Woman

	A	B	C	D	E	F	G	H	I	J	K	L
Height (cm)	160	166	174	158	166	161	170	171	166	163	167	168
Shoe size, continental	36	38	40	37	38	38	42	41	40	39	37	39

a Draw a scatter graph for this data.

b Jan said that 'Tall people have larger feet than short people.'
Use your scatter graph to test this 'hypothesis'.

a



Draw a grid. Scale the height axis (the horizontal axis) from 155 to 175 and the shoe-size axis (the vertical axis) from 36 to 43.

Plot one point to represent each woman. The coordinates for woman A are 160 across and 36 up.

b The scatter graph shows that taller women tend to have larger feet.

This is only a tendency. It is not always true.

Exam practice 4B

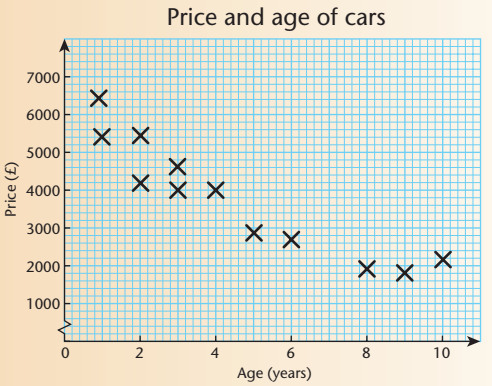
1 **a** In the example 2 we state that information was colleected only from women.
Explain why it is important to specify the gender?

Explain means write down a reason why *you* think it is important.

b How many of the women were taller than 165 cm?

Find the grid line marking a height of 165 cm.
The crosses to the right of this line represent heights greater than 165 cm. You will find it easier to count these if you place a ruler along the grid line.

2 This graph shows the age and price of some second-hand cars.
The cars are all small with similar engine sizes.



a Describe the relationship between the age of a car and its price.

b James says that 'The price of a second-hand car gets lower as the age increases'
Is James correct? Explain your answer.

Write down any two factors that you think will make a car cost more, or cost less.

c Apart from age, give *two* other factors that you think could affect the price of a car.

d Suki says that a ten-year old car is always cheaper than a similar eight-year old car.
Is Suki correct? Explain your answer.

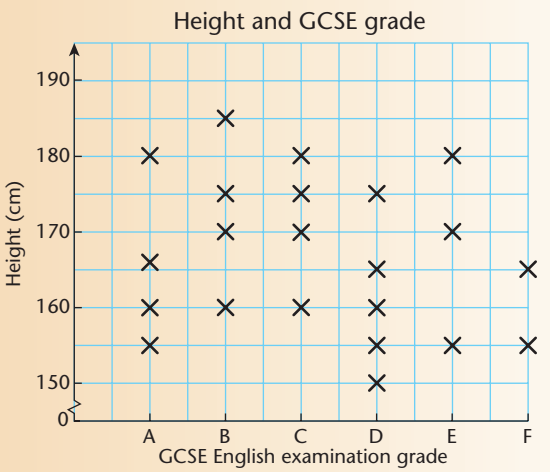
e How many of these cars cost more than £3000?

3 Lim saw this headline in a paper,

'Tall children get better grades at GCSE than short children.'

He did not believe it, so he did some research.

Lim collected data from a group of Year 11 students.
The scatter graph shows the information.

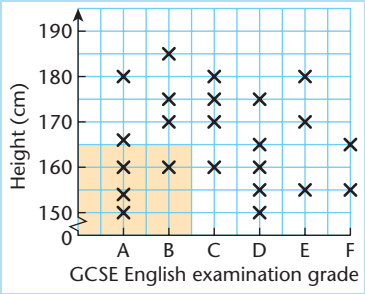


a Does the evidence here support the claim in the newspaper?
Explain your answer.

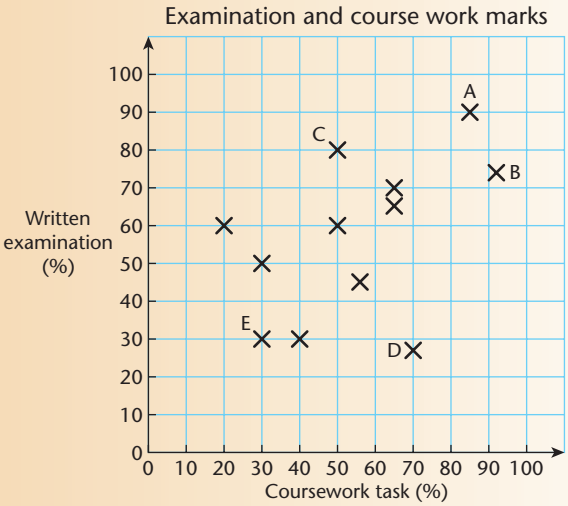
b How many students shorter than 165 cm got grades A or B?

c How can Liam make sure that his results were not biased?

You need the numberr of crosses below the 165 grid that are on grade A aand grade B grid lines.



4 The scatter graph shows the marks of a group of students in a coursework task and in a written examination.



a Which letter refers to the student who got the highest mark on the coursework task?

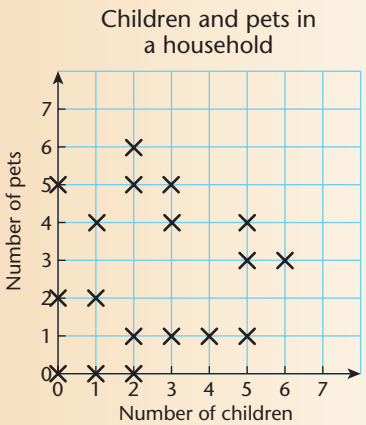
b Which letter refers to a student who got the same mark for the coursework task and for the written examination?

c How many students are represented on the graph?

d How many students got higher marks in their coursework than in their written examination?

Draw a line through the points on the graph that give equal marks. All crosses to the right of this line give higher marks for coursework than for the written exam.

5 The scatter graph shows the number of pets and the number of children in different households.



a How many households are shown?

b Describe the relationship between the number of pets and the number of children in a household.

c Jane said 'Larger families keep more pets than smaller families.'
   i Is Jane correct? Explain your answer.
   ii Sketch a scatter graph that would show that Jane was correct.

6 This table shows the number of rooms and the number of people living in each of twelve houses.

Number of rooms	4	4	5	5	6	6	6	7	7	7	8	8
Number of people	3	5	2	1	6	3	4	3	4	5	2	6

Fig A/W showing two sets of grandparents sitting end-to-end in the same double bed (like in charlie-and-the-chocolate factory

a Show this information on a scatter graph.
Use a scale of 1 cm for 1 unit on each axis.

b Ellie lives in a house with four other people.
Is the house likely to have more than four rooms?

7 A teacher wanted to see if there was a relationship between the mark a pupil got in the mock examination in maths and the mark that the pupil got in the final exam. She recorded the marks of 12 students. These are shown in the table.

Mock mark	72	84	85	41	46	57	44	43	56	80	52	38
Final mark	84	73	77	53	62	54	55	56	48	66	72	50

a Draw a scatter diagram to show this data on a graph.
Use 1 cm to represent 5 marks on both axes.
Mark each axis from 35 to 90.

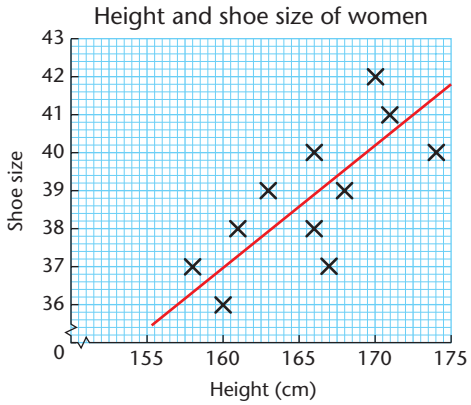
b Does your scatter diagram support the hypothesis that results in the mock exam give a good idea of the mark a student is likely to get in the final exam?

c Tyler got 66 in the mock but failed to take the final exam. Use the scatter graph to estimate what he might have scored in the final exam.

4.3 Line of best fit and correlation

The **line of best fit** is the straight line about which the points on the graph are evenly distributed..

This is the graph from Example 2. It shows the height and shoe size of 8 women. You can draw the line of best fit by eye.



There should be roughly the same number of points above the line and below it. This may mean that none of the points is on the line.

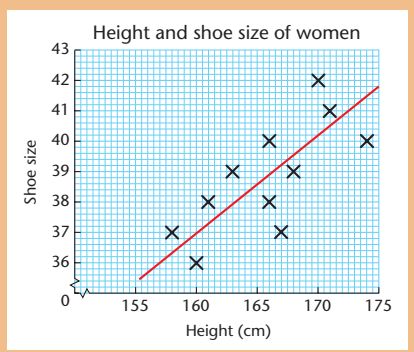
Correlation is the word used to describe the relationship between two sets of data.

Linear correlation means that the points are scattered about a straight line.
The closer the points are to the line, the stronger is the correlation.

There are two types of correlation.

***Positive* linear correlation** is when the line of best fit slopes upwards.

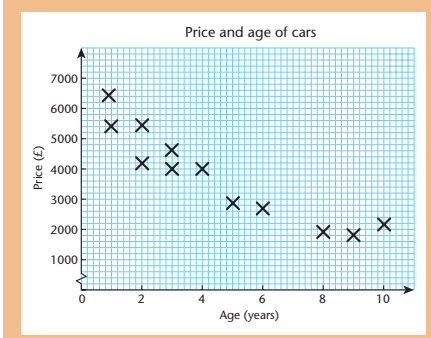
There is positive correlation in this graph.)



Shoe size tends to increase with a person's height.

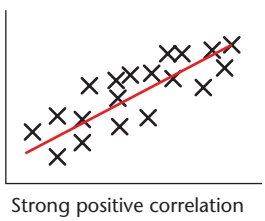
***Negative* linear correlation** is when the line of best fit slopes downwards.

There is negative correlation in this graph.



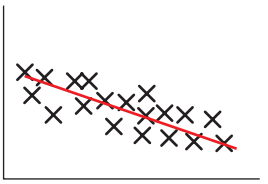
The value of a car goes down as it gets older.

When the points are close to the line, there is a **strong correlation**.



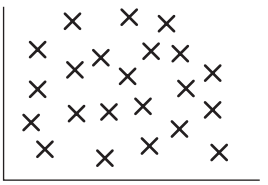
Strong positive correlation

When the points are loosely scattered about the line, there is **weak correlation**.

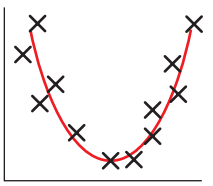


Weak negative correlation

Sometimes the points are so scattered that there is no obvious line. There is **no correlation**.

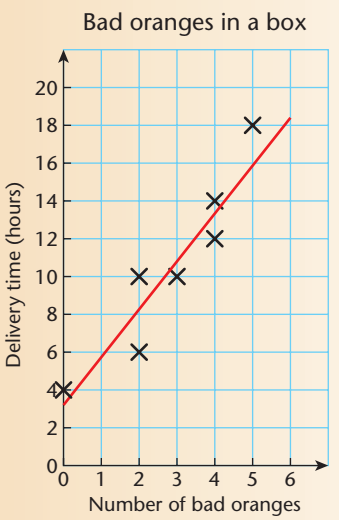


Some graphs have non-linear correlation. There is no obvious line of best fit in this diagram. However the grey curve shows there is some correlation but it is not linear.



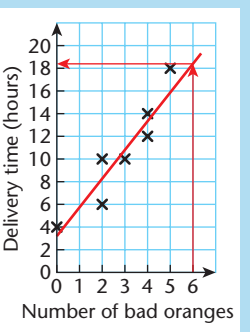
Exam practice 4C

1 This scatter graph shows the number of bad oranges in a box after different delivery times. The line of best fit has been drawn on the graph.

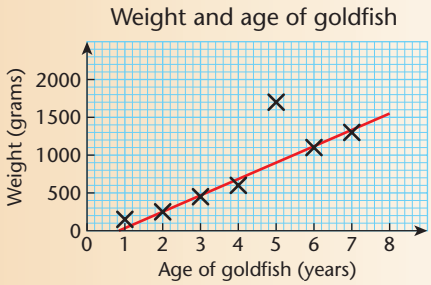


One box contained 6 bad oranges.
Use the diagram to estimate the delivery time for this box.

Find 6 on the the horizontal axis. Go from this point up to the line of best fit and then across to the vertical axis to read off the value.



2 The scatter diagram shows data collected over eight years by a breeder of goldfish.

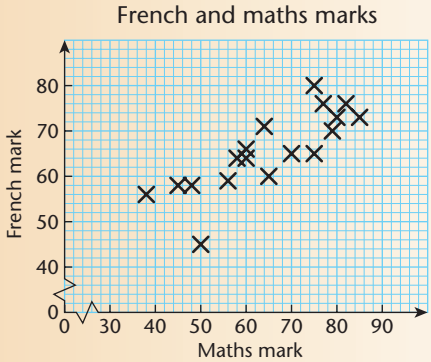


Bowl of goldfish

Look for a point that is much further away from the line than any other point. Use this point and describe its distance from the line as your reason in your explanation.

a Explain why one of the points is likely to have been plotted incorrectly.
b Estimate the average weight of a five-year-old goldfish.

3 The table shows the French and maths marks of 20 pupils in an end of term exam.



Make sure you know what each subdivision on the axes represents.

a Omar got 74 for maths and 68 for French.
On a copy of the diagram add a cross to the scatter diagram to show Omar's marks.
b Draw a line of best fit for all 21 marks on your copy.
c Describe the relationship between the maths marks and the French marks.
d John is good at French.
Is he likely to be good at maths? Explain your answer.
e Ahmed got 65 for his French examination but was absent on the day of the maths examination.
Estimate his likely mark if he had taken the maths examination.

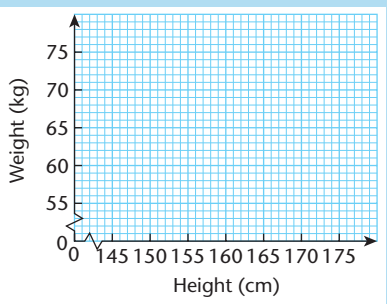
Find 65 on the French Mark axis, then go across to the line of best fit. Go down from this position to the Maths Mark axis and read the value there.

4 This table shows the heights and weights of 10 people.

Height (cm)	150	152	155	158	158	160	163	165	170	175
Weight (kg)	56	62	69	64	57	62	68	66	65	74

a Draw a scatter graph to represent this data.
b Draw a line of best fit.
c Describe the relationship between height and weight.

Use a grid like this.



5 This table shows the number of rooms and the number of people living in each of 15 houses.

Number of rooms	3	4	4	5	5	5	6	6	6	6	7	7	7	8	8
Number of people	2	3	5	4	2	1	6	2	3	4	4	5	3	2	6

a Plot this information as a scatter graph

b Draw a line of best fit on your scatter graph.

c Describe the relationship shown in the scatter graph.

d Use your scatter graph to estimate the number of rooms in a house with 7 people living in it.

e Describe how reliable your answer to part **d** is. Explain your answer.

f A data collection sheet looked like this:

House								
Number of rooms in the house and Number of people in the house								

Give **one** reason why this is not suitable.

6 This table shows the number of pens and pencils and the number of text books that each of 10 pupils have with them in a maths lesson.

Number of pens and pencils	2	3	3	5	6	6	12	15	20	24
Number of text books	4	5	0	3	1	4	6	2	1	6

a Draw a scatter graph to represent this data.

b Give **one** reason why it is not sensible to draw a line of best fit on your diagram.

c A questionnaire contained this question.

HOW MANY BOOKS ARE IN YOUR BAG?

Give **one** reason why this question is not suitable.

7 Match each sentence with the letter of the scatter graph that you think best matches the relationship.

a The number of pages and the number of advertisements in a newspaper.

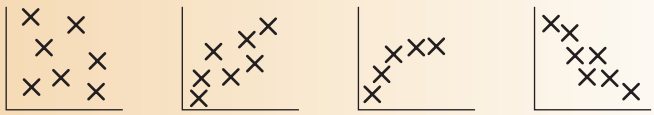
b The length and width of a cucumber.

c The weight of tomatoes produced by a tomato plant and its height.

d The age of an adult and their ability at sport.

e The score on each dice when a red dice and a blue dice are thrown together.

f The number of days a pupil is away from school and the number of days when that pupil is late in handing in coursework.



8 Fahed collected this data about the village he lived in.

Year	1965	1970	1975	1980	1985	1990	1995	2000
Population	100	150	300	550	400	180	400	250
Number of fish in the village pond	20	30	50	80	70	40	60	40

a Plot the population and number of fish as a scatter graph.

b Fahed said 'If we put a lot more fish in the pond, we will get more people living here.'
Fahed is not correct. Explain why he is wrong.

Goldfish in a pond

Mini coursework task

Sarah said that 'A person will live longer if they live in a wealthy country.'

- Look at this data. GNP means Gross National Product. Find the meaning of 'GNP per capita'.

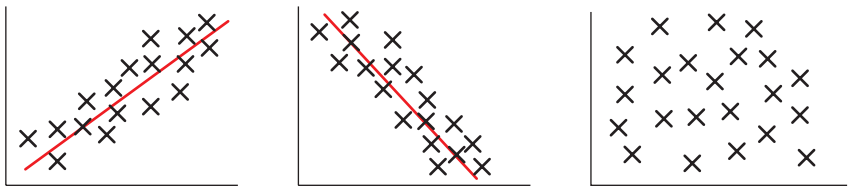
GNP means Gross National Product. Find the meaning of 'GNP per capita'.

Country GNP per capita	*(US$)*	*Life Expectancy (years)*
Luxembourg	41,210	77
Switzerland	40,630	79
Japan	39,640	80
Denmark	29,890	75
France	24,990	78
Canada	19,380	78
UK	18,700	77
New Zealand	14,340	76
Bahamas	11,940	72
Argentina	8,030	72
Oman	4,820	71
Brazil	3,640	67
Jordan	1,510	68
Ecuador	1,390	69
Bolivia	800	60
Pakistan	460	61
Laos	350	52
Nigeria	260	54
Uganda	240	41
Mozambique	80	46

- Write down a hypothesis to test if there is a relationship between the GNP and life expectancy in a country.
- Use the data to test your hypothesis.

Summary of key points

- Two-way tables are used to show different aspects of a set of data.
- scatter graphs are used to show data when you have two pieces of information about a person or an object.
- If the points on a scatter graph look as if they may be scattered about a line, you can judge by eye where this line is and draw it – it is called the line of best fit.
- Correlation describes how close the points are to the line of best fit.



- If there is no obvious the line of best fit, there is no linear correlation.
- No linear correlation does not mean that there is no relationship.

Most students who get GRADE C or above can:

- draw a line of best fit
- describe the correlation between two sets of data.

Glossary

correlation	relationship between two quantities
linear	description of a straight line
line of best fit	straight line drawn on a scatter graph which best fits the scatter of the points
scatter graph	points plotted on a graph to show two sets of information
two-way table	table giving two or more groups of information

5 Averages

This chapter will show you:
- ✓ the meaning of mean, mode and median
- ✓ how to find the mean, mode and median of a set of values.

Before you start you need to know:
- ✓ how to add, subtract, multiply and divide whole numbers
- ✓ how to add, subtract, multiply and divide decimals and fractions
- ✓ how to use simple formulae

5.1 Averages

The mean, median and mode are the three most common averages. **Measures of central tendency** is the collective name for the mean, median and mode.

Did you know
that the mean as defined here is properly called the arithmetic mean. There other means such as the geometric mean and the harmonic mean, but you do not use these at GCSE.

Mean

The most commonly used **average** is the **mean**.
To calculate the mean of a set of values, find the sum of the values and divide that sum by the number of values,

$$\text{Mean} = \frac{\text{Sum of all values}}{\text{Number of values}}$$

Example 1

Kay has 7 pens, Emily has 6 pens and Olivia has 11 pens.
Find the mean number of pens.

Mean number of pens $= \frac{\text{total number of pens}}{\text{number of students}}$

Total number of pens $= 7 + 6 + 11$
$= 24$
Mean $= 24 \div 3$
$= 8$ pens

The mean is not always a whole number.

Example 2

A/w 501

Ian has 2 dogs, Raj has 1 dog and Sara has 1 dog.
Calculate the mean number of dogs per person.

The mean $= \frac{2+1+1}{3} = \frac{4}{3} = 1\frac{1}{3}$ dogs

If they could share the dogs equally, each of the three people would have $1\frac{1}{3}$ dogs.
This is clearly impossible. But $1\frac{1}{3}$ is the mean of the numbers 2, 1 and 1.

Median

The **median** of a set of numbers is the middle value when they have been placed in order of size.

When there are two middle values, the median is the sum of these two values divided by two.

Example 3

These ten students are arranged in order of their height.
Find their median height.

The middle students are the fifth and sixth.

A/w 502

(155 + 157) ÷ 2 = 156

The median height is = 156 cm.

When there are a large number of items, you can find the middle item from this formula:

The middle item of n items is the $\frac{n+1}{2}$th item.

Mode

The **mode** of a set of values is the value that occurs most often.

Example 4

These are some test marks.

6, 4, 6, 8, 10, 6, 3, 8 and 4

Find the mode.

3, 4, 4, 6, 6, 6, 8, 8, 10

The mode is 6 marks.

Put the numbers in order. Then you can see which is the most common.

If the values are all different, there is no mode.
If two or more values occur equally often, there are two or more modes.

Exam practice 5A

1 Find the mean, median and mode of each of the following sets of data.

- **a** Test marks: 6, 8, 5, 6, 7, 9, 6
- **b** Tree heights: 2 m, 3 m, 5 m, 4.5 m, 2.5 m, 6 m, 3m
- **c** Number of pages: 2, 4, 8, 4, 7, 1, 7, 6, 5, 6
- **d** Weights 1.5 g, 2.1 g, 4.4 g, 3.5 g, 3.6 g, 2.1 g, 8.3 g, 5.6 g
- **e** Sums of money £1.20, £4.55, £8.20, £3.15, £2.99

Remember to give the units. The mean, median and mode have the same units as the data.

2 Hussain threw three coins several times and recorded the number of heads that showed at each throw.
His results, put in ascending order, are
0, 0, 0, 0, 0, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3
Find **a** the mode **b** the median **c** the mean.

You can find the middle item by adding 1 to the total number of items, then dividing by 2.
For example, when there are 46 items, the middle one is the (47 ÷ 2)th item.
47 ÷ 2 = 23.5, so the middle items are 23rd and 24th.

3 This stem-and-leaf diagram shows the number of tomatoes harvested from each plant in Joe's greenhouse one week.

Number of tomatoes — Key 1|3 means 13

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

- **a** How many tomato plants does Joe have in his greenhouse?
- **b** Write down the mode.
- **c** Find the median.
- **d** Calculate the mean.

Fig A/W: man tending tomato plants in greenhouse. - A/w 503

4 This stem-and-leaf diagram shows the number of pages in a sample of books.

Number of pages — Key 1|34 means 134

0	01 19 34
1	09 10 25 36
2	10

- **a** Write down the number of pages in the longest book.
- **b** Write down the number of pages in the shortest book.
- **c** Find the mean number of pages per book.

5 Tina got these marks in her last five mathematics tests.

8, 7, 8, 9, 10.

She worked out her mean, median and mode mark for these tests
In her next test Tina scored 6.
Explain how this mark will affect the mean, median and mode.

6 The buses that passed the school gate in four hours were counted.
The mean number of buses was 3 each hour.
How many buses were counted?

The mean = total ÷ number of items, so
total = mean × number of items

7 Twenty-eight students took a maths test.
The mean mark was 15.

a What is the sum of the marks for all these students?

b Carl was away on the day of the test and took it later. His mark was 24.

i Will Carl's mark increase or decrease the average mark for the test? Explain your answer.

ii Find the new mean mark when Carl's mark is added into the total. Give your answer correct to 1 decimal place.

If Carl's mark is more than the old mean, adding it to the total will raise the mean; if it is less than the old mean, adding it to the total will lower the mean.

8 Deena measured the lengths of some leaves. Here are her results:

4.9 cm, 5.2 cm, 5.6 cm, 5.2 cm, 5.7 cm, 5.2 cm.

a Find the mean length of Deena's leaves.
Jonathon measured the lengths of some different leaves. Here are his results:

5.5 cm, 4.7 cm, 5.0 cm and 4.4 cm.

b Find the mean length of Jonathon's leaves.

c Jonathon said 'The mean of all the leaves will be halfway between these two means.'
Is Jonathon correct? Explain your answer.

d Find the mean length of all ten leaves.

9 Eight students took a test. Their mean mark was 68.

a Write down the sum of all the marks.

b Caroline and Greg were absent and took the test later. She scored 59 and he scored 47.
What was the mean mark for the 10 students?

10 The same practical test was given to all the students on an electronics course.
The mean mark for the 18 day students was 55.
The mean mark for the 12 evening students was 50.
Work out the mean mark for the whole group.

You cannot 'average' the two means. You need to work with the total number of marks.

11 Kit has four numbered cards. The mean of the numbers is 4.
He picks another card. The mean of his five numbers is 5.
What number is on the new card?

5.2 Mode, median and mean from ungrouped frequency tables

Frequency tables and bar charts showing ungrouped data give the information in order of size. The mode and the median are easy to find.

Example 5

This table shows the number of words in each sentence on the first page of a book.

Number of words	Frequency
5	8
6	10
7	6
8	4
9	2
10	3

fig A/W an open book

Find **a** the mode **b** the median.

a The mode is 6 words

b $\frac{33+1}{2} = \frac{34}{2}$, so the median is the 17th sentence.

The median is 6 words.

The mode is the number of words with the highest frequency.

Count down the frequencies until you find the 17th sentence. The first 8 sentences have 5 words, adding on the next 10 gives 18 sentences. Therefore the 17th sentence is one of these.

You can find the mean from a frequency table but it needs to be done in an organised way.

Example 6

This table shows the marks out of 5 scored by students in a test. Find the mean mark.

Maths marks in a test

Mark	Frequency
1	2
2	8
3	11
4	6
5	3
Total	30

Maths marks in a test

Mark	Frequency	Frequency x mark
1	2	2
2	8	16
3	11	33
4	6	29
5	3	15
Total	30	90

The sum of all the marks is 90. There are 30 marks so the mean mark is $90 \div 30 = 3$ marks

To find the mean of these marks you need to first add up all the marks. It is easiest to do this in stages.
The table shows that there are 2 lots of 1 mark, so these add up to 2.
There are 8 lots of 2 marks, so these add up to $8 \times 2 = 16$.
Add another column to the table so that you can keep track of your work.

Exam practice 5B

1 Some students gathered this information about themselves.
Find **a** the mode **b** the median **c** the mean.

Number of children in each family	Frequency
1	8
2	12
3	4
4	2

To find the mean, copy the table and add a column like the one in Example 6. Add up the numbers in this column and divide by the total number of families.

2 Jonny watched students leave the music studio in small groups or alone. This table shows what he saw.

Number in group	Frequency
1	3
2	7
3	4
4	4

a How many students did Jonny see leaving?

b Find the mean size of the groups.

3 Joshua tossed three coins several times and recorded the number of heads that showed at each toss. His results are shown in the table.

Number of heads obtained when three coins are tossed	Frequency
0	9
1	7
2	16
3	3

a Write down the mode and the median.
b Find the mean number of heads per toss.

Add a column to the table and remember that $0 \times 9 = 0$.

Class discussion
You can give the mean as a fraction or as a decimal.
Which is better?

4 This table shows the number of defective items found by a technician in each sample of 12 tested.

Number of defective items	Frequency
0	25
1	9
2	4
3	2

a How many defective items were found?
b What is the median number of defective items?
c Find the mean number of defective items.

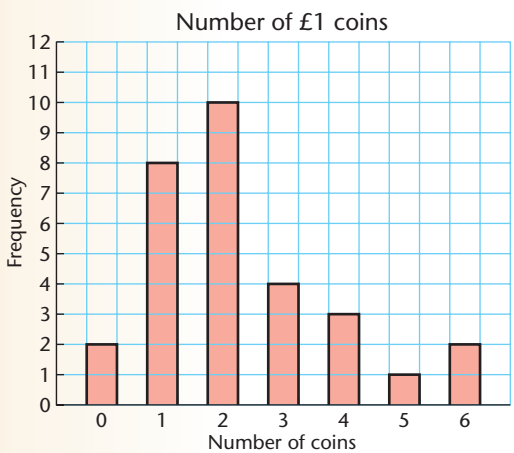
Photo: technician testing electrical circuit.

5 Once every five minutes, Debbie counted the number of people queuing at a checkout. This table summarises her results.

Number of peeople queuing at a supermarket checkout	Frequency
0	4
1	6
2	5
3	2
4	2

a How many times did she count?
b Find the mean number of people queuing.

6 A sample of students were asked to count the number of one pound coins that they had with them.
The distribution of these coins is shown in the bar chart.



a Find the size of the sample.
b Write down the mode and the median.
c Work out the total value of the one pound coins.
d The total sum of money was shared out equally among the students. How much did each student get?

5.3 Finding averages from a grouped frequency distribution

You do not know individual values when data is grouped.
This means that you cannot find exact values for the mode, the median or the mean.
For grouped data, the **modal group** is the group with the highest frequency.

You can also find the group that contains the median.
You can estimate the median value from a histogram.

The area under a histogram represents the frequency. So the median value can be estimated as the value that divides the area in half.

Example 7

Forty-five boxes of oranges were examined and the number of bad oranges in each box was counted.
The results are summarised in this table.

photo: boxes of oranges - one open

Number of bad oranges in a box	0–4	5–9	10–14	15–19
Frequency	29	11	4	1

a Write down the modal group.
b Which group contains the median?

a *0 – 4 oranges*

b *There are 45 boxes.*

$\frac{45+1}{2} = 23$, *so the middle one is the 23rd.*
The median is in the group 0 – 4 oranges.

There are 29 boxes in the first group, so the 23rd is in this group.

The **estimated mean value of a grouped frequency distribution** is given by

$$\frac{\text{sum of all (frequency} \times \text{mid-interval value)}}{\text{sum of frequencies}}$$

Example 8

This grouped frequency table shows the heights of 80 five-year-olds.
Estimate the mean height.

Height, h cm	Frequency
$90 \leqslant h < 95$	3
$95 \leqslant h < 100$	9
$100 \leqslant h < 105$	18
$105 \leqslant h < 110$	29
$110 \leqslant h < 115$	21
	Total: 80

Add two columns to the table: one to list the mid-interval values and one to list the mid-interval values × frequency.

The mid-interval value for the first group is (90 cm + 95 cm) ÷ 2 = 92.5 cm.
Use this as an estimate for the height of every child in this group.
The estimated total height of all the children in this group is 3 × 92.5 cm = 277.5 cm.

Height, h cm	Frequency	Mid-interval value	Frequency x Mid-interval value
$90 \leqslant h < 95$	3	92.5	277.5
$95 \leqslant h < 100$	9	97.5	877.5
$100 \leqslant h < 105$	18	102.5	1845
$105 \leqslant h < 110$	29	107.5	3117.5
$110 \leqslant h < 115$	21	112.5	2362.5
	Total: 80		Total: 8480

The total height of all 80 children is estimated as 8480 cm,
$\frac{8480}{80} = 106$, so the mean height is approximately 106 cm.

Exam practice 5C

1 Jake asked 100 pupils how much they spent in the school tuck shop. His results are shown in the table.

Amount (pence)	Frequency, f
0–24	26
25–49	15
50–74	38
75–99	21

Find an estimate for the mean amount of money spent.

Copy the table and add two columns so that it is like the table above.
Each mid-interval value is the sum of the extreme values divided by 2.
So the mid-interval value of the 25–49 group is (25 + 49) ÷ 2 ÷ 742 = 37.

2 Fifty boxes of peaches were examined and the number of bad peaches in each box was recorded. The results are shown in the table.

Number of bad peaches per box	Frequency
0–4	34
5–9	11
10–14	4
15–19	1

Find an estimate for the mean number of bad peaches per box.

You need to make a table like the one in the example.

3 Ambulance response times to 999 calls are given in this table.

Response time, t minutes	$5 \leqslant t < 10$	$10 \leqslant t < 15$	$15 \leqslant t < 20$	$20 \leqslant t < 25$	$25 \leqslant t < 30$
Frequency	3	6	27	18	6

Find an estimate for the mean response time.

Take care! This table gives the groups going across. Either add two rows for mid-interval value and frequency × mid-interval values under the table or make a new table with the groups going down.

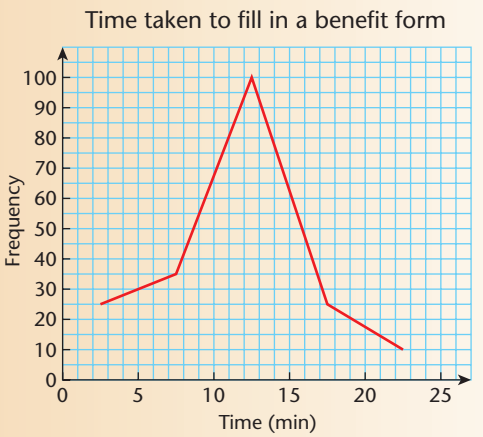
4 Twenty seeds were planted.
Five weeks later, the heights of the resulting plants were measured.
The results are shown in the table.

photo: seed tray with seedlings growing in it

Height, h cm	Frequency
$1 \leqslant h < 4$	2
$4 \leqslant h < 7$	5
$7 \leqslant h < 10$	10
$10 \leqslant h < 13$	3

Find an estimate for the mean height of the seed1ings.

5 A new benefit form was trialed by asking some people to complete it.
The time each person took was recorded and this frequency polygon summarises the results.



a How many people were asked to complete the form ?

b Copy and complete this table.

Mid-interval value, (minutes)	Frequency	Frequency mid-interval value
2.5	25	62.5
7.5	35	
Total:		

c Find an estimate for the mean time taken to complete this form.

6 Twenty-four seeds of the same flower were planted.
Five weeks later, the heights of the new plants were measured and recorded in this table.

Height (h cm)	$1 \leqslant h < 5$	$5 \leqslant h < 10$	$10 \leqslant h < 15$	$15 \leqslant h < 20$
Frequency	3	6	11	4

The modal class is the class or group with the largest number of items in it.

a Write down the modal class.
b In which class interval does the median height of a seedling lie?
c Estimate the mean.

7 Each painter in a group was timed to see how long it took them to complete a job. This table summarises the results.

Time, t minutes	**Number of painters**
$1 \leqslant h < 5$	6
$5 \leqslant h < 10$	9
$10 \leqslant h < 15$	4
$15 \leqslant h < 20$	1

Fig A/W two students each painting a door.

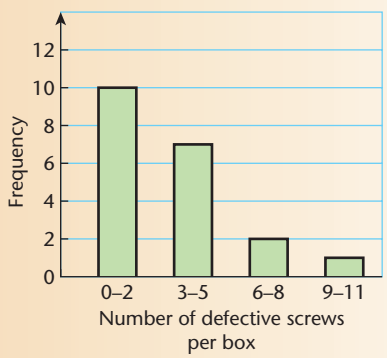
a Write down the modal class.
b Write down the class interval that contains the median time.
c Estimate the mean.

8 This frequency table shows the heights of the 29 students in a class.

Height, t cm	$148 \leqslant h < 150$	$150 \leqslant t < 152$	$152 \leqslant t < 154$	$154 \leqslant t < 156$	$156 \leqslant h < 158$
Frequency	2	5	8	9	5

a Write down the modal class.
b In which class interval does the median height of a student lie?
c Estimate the mean.

9 The bar chart shows the result of an examination of 20 boxes of screws.



a Make a frequency table to represent this information.
b Estimate the mean number of defective screws per box.

10 This histogram summarises the times taken by a group of people to complete a questionnaire.

a Estimate the median time.

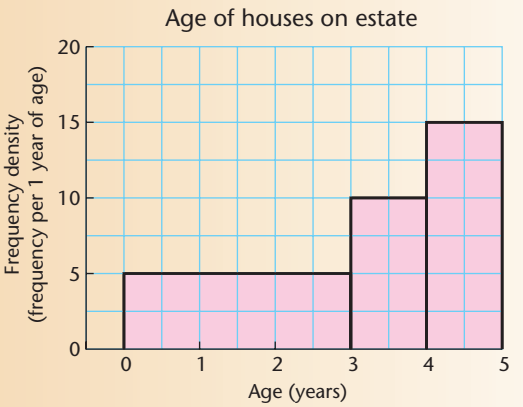
b Copy and complete this frequency table.

Time t minutes	$5 \leqslant t < 15$		
Frequency			

c Write down the modal class

d Estimate the mean time.

11 This histogram summarises the ages of houses on an estate.



Estimate the median age.

5.4 Deciding which average you need

Depending on the situation, you may want to use either the mode, median or mean.

Example 9

Amy wants to buy a printer.
How useful is each of these pieces of information? Give reasons for your answers.

Photo of a photo printer

a The average price of a printer is £95. This is the mean price.

b A middle range printer costs about £80. This is the median price.

c The most common cost is about £70. This is the modal price.

a *The mean price involves adding up all the prices so it takes account of very expensive printers. If Amy does not want an expensive machine, the mean price is not much use.*

b *The median price is quite useful as it shows that half of the models are available for about £80 or less.*

c *The modal price shows that there is more than one printer costing about £70 so there is a choice of printers at this price. It doesn't tell Amy how many printers are priced at about £70; it could be only two or three, so this information is of limited use.*

Class discussion

Which of the mean, median or mode will be the most useful in the following situations?

1 You want to know if you are taller or shorter than most women.

2 The school shop sells several different items.
You need to buy more stock and do not want to run out of the most popular item.

3 You have some information about the temperature in Florida in February.
You need to decide what clothes to take with you for a 10-day holiday to Florida in February.

4 You want an idea of the price you will have to pay for an inkjet photo printer.

Exam practice 5D

For each question, write down whether the mean, mode or median is the most useful average, and whether the range is useful.
Give a reason for your choice.

1 Six pupils got marks of 5, 7, 8, 4, 8 and 4 in a test:
You want one value to show how they did as a group.

2 In different shops, the price of a can of cola is

50 p, 45 p, 65 p, 42 p, 45 p.

You want an idea of the cost of a can of cola.

3 Five people decided to pool their money.
They put in the following amounts: £10, £5, £6, £7 and £12.
You want to know how much each would have if the total amount was divided equally.

4 The ages of the children in a swimming club are

9, 10, 8, 10, 11 ,8, 12, 9, 10, 11, 10, 12.

You want to know the most likely age for a club member.

ICT task

Use an internet search engine to find an online shop selling a large range of photo printers.
List about twenty different prices.
Enter the prices you have found into a spreadsheet.

Use your spreadsheet to find:

- the cheapest printer
- the most expensive printer
- the median price
- the mean price.

Use your spreadsheet program to draw a histogram.

Summary of key points

- There are three different kinds of averages: **mean**, the **median** and the **mode**.
- A set of values always has a mean and a median. It may have no mode, or one or more modes.
- You find the mean by adding all the values then dividing this sum by the number of values.
- The median is the middle value after a set of values have been arranged in order of size. For n values arranged in order of size, the median is the $\frac{n+1}{2}$th value.
  When there are two middle values, the median is halfway between them.
- The mode is the value or values that occur most often.
- The mid-interval value is the sum of the end values of the class divided by 2.
- To find the mean from an ungrouped frequency table, calculate.

$$\frac{\text{sum of all (frequency} \times \text{value)}}{\text{sum of frequencies}}.$$

- The mean of a grouped frequency table can only be an estimate as the individual values are not known.
- To estimate the mean from a grouped frequency table, calculate

$$\frac{\text{sum of all (frequency} \times \text{mid-interval value)}}{\text{sum of frequencies}}.$$

Most students who get GRADE C or above can:

- find the mean of grouped data.

Most students who get GRADE A or above can also:

- decide which average to use and give reasons for the choice.

Glossary

average	any one of the mean, mode or median
mean	the sum of a set of values divided by the number of values
measure of central tendency	the mean, median or mode
median	the middle value after a set of values have been arranged in order of size
mode	the value or values that occur most often