

Collecting and analysing data

In the run up to a general election, opinion polls are taken of which political party people are likely to vote for. The results of just a 1000 peoples' voting intentions are taken very seriously by the media and the politicians.

What's the point?

By selecting a representative sample, surveys allow statisticians to obtain reliable results from manageable amounts of data. The technique is used by polling organisations and throughout industry to monitor quality and productivity etc.

Check in

You should be able to

■ know how to collect data

- 1 Nicola and Maddy wanted to find out how many people catch flu in December. Nicola asked 100 people in the town centre one morning. Maddy looked up flu statistics on the internet.

- a Who collected primary data and who collected secondary data?
b Explain the difference between primary and secondary data.

■ know how to record data

- 2 Lee drew this tally table to collect data on hours of TV people watch:

TV hours	Frequency
3-4	
4-5	
more than 5	

- a Explain why it would be difficult to write a tally mark for
i 4 hours ii 2 hours.
b How could the difficulties discussed in part a be overcome?

■ do simple arithmetic

- 3 Find
a $\frac{1}{2}$ of 45 b $\frac{1}{2}$ of $(41 + 1)$ c $\frac{1}{4}$ of 24 d $\frac{3}{4}$ of $(27 + 1)$

Orientation

What I need to know

KS3 Knowledge of collecting data
Find averages and the range for data sets

N1 Order numbers
Do basic arithmetic

What I will learn

■ Design a survey, select a sample and use a two-way table
■ Calculate measures of average and spread

■ Combine the mean of two data sets

What this leads to

D4 Use statistics to describe data

Opinion polls, Journalism

Rich task

What is the average time taken to travel to school for pupils at your school? Do boys take less time than the girls? Do the pupils who live nearer the school take less time than the pupils who live further away? Investigate and write a report on your results.

This spread will show you how to:

- Design a questionnaire, recognising bias and taking steps to avoid it

Keywords

Biased
Data
Questionnaire
Survey

Organisations carry out statistical surveys to collect data that help them plan for the future.

You can use a **questionnaire** and collect **data** in a **survey**.

You have to choose suitable questions for a questionnaire.

Suitable questions,

- can be answered yes or no
- ask for facts.

Unsuitable questions,

- may be vague
- are leading or **biased**
- could be embarrassing.

Questions must have responses that,

- cover all possible answers
- do not overlap or have gaps.



Do you have an MP3 player?
How many CDs do you own?
none 1-10 11-20 over 20

Do you listen to a lot of music?
Do you agree that Coldplay is the best band in the world?
How cool are you?

Where do you buy your CDs?
internet store other

How much do you spend on CDs each month?
£0 to £14.99 £15 to £29.99
£30 and up

Example

A recently restyled breakfast radio show conducted this survey.

1 What is your opinion of the new breakfast show?

Fantastic Good

2 How long do you listen to the show?

10 min 1 hour 1-2 hours

a Comment on these questions.

b Write two questions for the survey to find out if listeners like the new show and for how long they listen.

a Question 1 does not include all possible responses, for example if you do not like the new show or don't think it's an improvement. In question 2 there are gaps, between 10 min and 1 hour, and an overlap: 1 hour appears twice.

b 1 What is your opinion of the new breakfast show compared to the old one?

Much better Better
Neither better nor worse Not as good

2 How much of the breakfast show do you listen to?

all of it over an hour
half an hour to 1 hour less than half an hour



A02 Functional Maths

1 Katy is doing a survey to find out how often people go to the cinema and how much they spend. She writes this question:

How many times a month do you go to the cinema?

- What is wrong with this question?
- Write an improved question to find out how often people visit the cinema.
- Write a question to find out how much people spend when they go to the cinema.

2 Sally put this question in a questionnaire:

Do you agree that tennis is the best sport?

- What is wrong with this question?
- Write a better question to find out the favourite sport. Include some response boxes.

Sally also wants to find out how often people play sport.

b Design a question for Sally to use. Include some response boxes.

3 James wants to know which flavour crisps he should stock in the school tuck shop. He asks this question:

Do you prefer plain or ketchup flavoured crisps?

- What is wrong with this question?
- Think about what crisps you and your friends like and design a better question for James to use. You should include some response boxes.
- James also put this question in his questionnaire:

How many times have you visited the tuck shop?

Once Lots of times

- Write two things that are wrong with this question.
- Design a better question for James to use. Include some response boxes.

4 Merlin wants to find out how far people would travel to see their favourite band perform. He writes this question:

How far would you travel to see your favourite band?

less than 1 mile 5-10 miles any distance

- What is wrong with this question?
- Design a better question for Merlin to use. Include some response boxes.

Merlin also wants to find out how much people would pay for a ticket to see their favourite band.

b Design a question for Merlin to use. Include some response boxes.

In this exercise you can find out if any questions you write 'work' by testing them out on groups of people in your class. The larger the sample the more reliable the results.

D1.2 Collecting data – choosing a sample

This spread will show you how to:

- Design a questionnaire, recognising bias and taking steps to avoid it
- Understand the concept of random sampling

Keywords

Bias
Population
Random
Sample

It may be time-consuming, too costly, too long or too impractical to collect data from everyone. In these cases you should ask a representative **sample**.

You must choose the sample so that it is not biased. For example, a survey of preferred music using a sample of friends is biased as friends are more likely to have similar opinions.

- A sample should represent a whole **population**.

One way of avoiding **bias** is to use a **random** sample.

- In a random sample each member of the population has the same chance of being included.
- Methods for choosing a random sample include:
 - taking names out of a hat
 - giving everyone a number and using a calculator or random number tables to pick numbers.

The population is the group of people or items being surveyed.

The larger the sample the more reliable the results.

A survey that includes everyone is called a census.

Example

James carries out a survey to find out if people in his town enjoy sport. He stands outside a football ground and surveys people's opinions as they go in to watch a match. Write two reasons why this is not a good sample to use.



People who watch football usually enjoy sport. More men than women go to watch football so the survey could be gender biased.

Example

A train company carried out a survey about a local rail service. They telephoned 100 people from a page of the telephone directory to answer a questionnaire on the rail service.

Write three reasons why this sample could be unrepresentative.

Only people who have a land-line telephone (and are not ex-directory) can be included in the sample.
Only people on one page of the telephone directory are included in the sample.
Some people may not be at home when they are phoned.

Exercise D1.2

Grade D



A02 Functional Maths

- 1** Katy is doing a survey to find out how often people go to the cinema and how much they spend. She stands outside a cinema and asks people as they go in.
Write a reason why this sample could be biased.
- 2** Sally wants to find out how often people play sport. Sally belongs to an athletics club. She asked members in her athletics club.
How could this sample be biased?
- 3** James wants to know which flavour crisps he should stock in the school tuck shop.
 - a** He asks his mum, dad, auntie and uncle.
Explain why this is not a good sample to use.
 - b** He asks only Year 11 at his school.
Explain why this sample could be biased.
 - c** Describe how James could take a sample of 50 people. (There are 1000 people in his school.)
- 4** Merlin wants to find out how far people would travel to see their favourite band perform.
 - a** He asks all his friends.
Write a reason why this sample could be biased.
 - b** He goes into town one Saturday morning and asks anyone listening to music on a MP3 player.
How could this sample be biased?
- 5** Jenny carries out a survey to find out the most popular band. She asks 10 of her friends – all girls.
How could this sample be biased?
- 6** Wayne carries out a survey to find out the most popular car colour. He stands on a street corner and notes the colour of the first 15 cars that pass by.
Write a reason why this sample could be biased.
- 7** Lisa wants to find out how people travel to work.
 - a** She asks people at a bus stop one morning.
How could this sample be biased?
 - b** She opens the telephone directory at a random page and phones everyone on that page.
How could this sample be biased?

D1.3 Designing a data collection sheet – two-way tables

This spread will show you how to:

- Design and use data collection sheets and two-way tables

Keywords

Data
Two-way table

- You can use a data collection sheet to collect **data** from a questionnaire or experiment.
- You can use a **two-way table** to collate the two sets of results.

Example

Two questions on a questionnaire are:

‘Are you male or female?’ and ‘How old are you?’

Design a two-way table to collect this data.

	Under 10	10–19	20–29	30–40	40+	Total
Male						
Female						
Total						

- You can use data in a two-way table to find other results.

Example

The table gives information about Key Stage 4 students at a school.

	Boys	Girls
Year 10	68	117
Year 11	89	126

- Work out the percentage of Key Stage 4 students in Year 10 who are boys.
- Work out the percentage of Key Stage 4 students who are girls.

Find the totals in the table.

	Boys	Girls	Total
Year 10	68	117	185
Year 11	89	126	215
Total	157	243	400

There are 400 students at Key Stage 4 ($68 + 117 + 89 + 126$).

- There are 68 Year 10 boys: $\frac{68}{400} \times 100 = 17\%$
- There are 243 ($117 + 126$) girls altogether: $\frac{243}{400} \times 100 = 60.75\%$

Exercise D1.3

Grade C



A02 Functional Maths

- Katy is doing a survey to find out how often people go to the cinema and how much they spend.

Design a suitable data collection sheet in the form of a two-way table that she could use.

- Sally wants to find out how often people play sport.

She wants to divide the results into those from males and those from females.

Design a suitable data collection sheet in the form of a two-way table that she could use.

- James wants to know which flavour crisps he should stock in the school tuck shop.

He also wants to know which year groups prefer which flavours.

Design a suitable data collection sheet in the form of a two-way table that he could use.

- Merlin wants to find out how far people would travel to see their favourite band perform and how much they would spend on a ticket to watch them.

Design a suitable data collection sheet in the form of a two-way table that he could use.

- Jenny carries out a survey to find out people’s favourite band and how many of that band’s CDs they own.

Design a suitable data collection sheet in the form of a two-way table that she could use.

- Wayne carries out a survey to find the most popular car colour and the most popular make of car.

Design a suitable data collection sheet in the form of a two-way table that he could use.

- Lisa wants to find out how people travel to work and how long it usually takes them.

Design a suitable data collection sheet in the form of a two-way table that she could use.

- The table gives information about the number of students in Years 7–9 that attended a school disco.

	Year 7	Year 8	Year 9
Boys	42	58	96
Girls	78	104	122

- How many students attended the disco?
- Work out the percentage of students that were
 - Year 8 girls
 - in Year 7
 - boys.



D1.4 Averages and spread

This spread will show you how to:

- Find an average and a measure of spread for a data set

Keywords

Average
Interquartile range
Lower quartile
Mean
Measure of spread
Median
Mode
Range
Upper quartile

You can summarise data using an **average** and a **measure of spread**.

p.146

- An average is a single value. There are three types of average:
 - the **mode** is the value that occurs most often
 - the **median** is the middle value when the data are arranged in order
 - the **mean** is calculated by adding all the values and dividing by the number of values.

- Spread is a measure of how widely dispersed the data are. Two measures of spread are:
 - the range
 - the **interquartile range** (IQR).

If there are one or more extreme values the IQR is a better measure of spread than the range.

- Range = highest value – lowest value
- IQR = upper quartile – lower quartile

An extreme value is a value well outside the range of the rest of the data.

Lower quartile
= $\frac{1}{4}(n + 1)$ th value.

Upper quartile
= $\frac{3}{4}(n + 1)$ th value.

Example

Louise collected data on the number of times her friends went swimming in one month.

4 7 22 1 6 2 1 5 6 6 4

Work out the: **a** range **b** mode **c** mean
d median **e** interquartile range.

In order the data are: 1 1 2 4 4 5 6 6 6 7 22

a Range = $22 - 1 = 21$

b Mode = 6

c Mean = $(4 + 7 + 22 + 1 + 6 + 2 + 1 + 5 + 6 + 6 + 4) \div 11$
 $= 64 \div 11 = 5.8$

d There are 11 values.

Median = $\frac{11+1}{2} = 6$ th value = 5

e Interquartile range = upper quartile – lower quartile

Lower quartile = $(\frac{11+1}{4})$ th value Upper quartile = $(\frac{3(11+1)}{4})$ th value
 $= (\frac{12}{4})$ th value $= 9$ th value
 $= 3$ rd value $= 6$
 $= 2$

IQR = $6 - 2 = 4$

The mean does not have to be an integer even if all the data values are integers.

If there are n values
Median value
= $(\frac{n+1}{2})$ th value.

Exercise D1.4

Grade D/C

- For these sets of numbers work out the
 - range
 - mode
 - mean
 - median
 - interquartile range.
 - 5, 9, 7, 8, 2, 3, 6, 6, 7, 6, 5
 - 45, 63, 72, 63, 63, 24, 54, 73, 99, 65, 63, 72, 39, 44, 63
 - 97, 95, 96, 98, 92, 95, 96, 97, 99, 91, 96
 - 13, 76, 22, 54, 37, 22, 21, 19, 59, 37, 84
 - 89, 87, 64, 88, 82, 88, 85, 83, 81, 89, 90
 - 53, 74, 29, 32, 67, 53, 99, 62, 34, 28, 27, 27, 64, 27
 - 101, 106, 108, 102, 108, 105, 106, 109, 103, 105, 107, 104, 104, 105, 105
- For the set of numbers in question **1e**, explain why the interquartile range is a better measure of spread to use than the range.
- For the set of numbers in question **1f**, explain why the mode is not the best average to use.
- Subtract 100 from each of the numbers in question **1g** and write down the set of numbers you get.
 - For your set of numbers in **a**, work out the
 - range
 - mode
 - mean
 - median
 - interquartile range
 - Compare your answers for the measures of spread in part **b i** and **v** and **1g i** and **v**. What do you notice?
 - Add 100 to your answers for the measures of average in part **b ii**, **iii** and **iv**. Compare these answers to the answers you got in question **1g**. What do you notice?
 - Give a reason for what you noticed in parts **c** and **d**.

A02 Functional Maths

- A scientist takes two sets of measurements from her experiment. Her results are:
Set A: 0 99 99 100 100 100 100 100 101 101 200
Set B: 0 0 99 99 100 100 100 101 101 200 200
 - For each set of measurements, work out the
 - range
 - mode
 - mean
 - median
 - interquartile range
 - Discuss what you notice about the measurements and your answers to part **a**.

A03 Problem

- Repeat question **1d** with the addition of a twelfth number 90.

D1.5 Mean of two combined data sets

This spread will show you how to:

- Find the mean of two combined data sets

Keywords

Mean

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

You can combine two data sets to form one larger data set by finding the mean of all the data.

Example

32 students, 12 boys and 20 girls, in class 8Z sat a maths test. The boys' mean mark was 63%. The girls' mean mark is 78%. Work out the mean mark for class 8Z.

Boys: Total sum of marks $63 \times 12 = 756$

Girls: Total sum of marks $78 \times 20 = 1560$

Total sum of marks for boys and girls $756 + 1560 = 2316$

Mean mark of all students

$$2316 \div 32 = 72.375\% = 72\% \text{ to nearest whole mark.}$$

Example

50 students answered a survey question about time spent on the internet one evening. 30 of the students were boys and 20 were girls.

The mean time spent on the internet by all 50 students was 18 minutes. The mean time spent on the internet by the 30 boys was 24 minutes.

Work out the mean time spent on the internet by the 20 girls.

Total time spent on internet by all 50 students:

$$50 \times 18 = 900 \text{ minutes}$$

Total time spent on the internet by the 30 boys:

$$30 \times 24 = 720 \text{ minutes}$$

Total time spent on the internet by the 20 girls:

$$900 - 720 = 180 \text{ minutes}$$

Mean time spent on internet by the 20 girls:

$$180 \div 20 = 9 \text{ minutes}$$



Example

There are 13 boys and 16 girls in a class.

In a test, the mean mark for the boys was p .

In the same test, the mean mark for the girls was q .

Find an expression for the mean mark of all 29 students.

$$\text{Mean} = \frac{13p + 16q}{29}$$

Exercise D1.5

Grade C

- 1 An athletics club has 100 members, 60 boys and 40 girls.

The mean time the boys spent training one day was 86 minutes. The mean time the girls spent training one day was 72 minutes.

Work out the mean time spent training on one day for all 100 members of the athletics club.

- 2 There are 120 students in Year 11 at St Edmunds school. 75 are girls and 45 are boys.

The mean time spent on homework each week for boys is 5.2 hours. The mean time spent on homework each week for girls is 8.6 hours.

Work out the mean time spent on homework for all 120 students in Year 11 at St Edmunds school.

- 3 Of the students in Year 13 at St Edmunds school, 60 boys and 20 girls have passed the driving test.

The mean number of driving lessons that all 80 students had before passing the test was 19.75.

The mean number of driving lessons for the boys was 12.

Work out the mean number of driving lessons for the girls.

A02 Functional Maths

- 4 The mean mark in a statistics test for class 10Z was 84%.

There are 32 students in the class, 12 of whom are girls. The mean mark in the test for these girls was 93%.

Work out the mean mark in the statistics test for the boys.

- 5 Thirty boys and girls were asked how many times they had visited the cinema in the past year.

The average number of times was 5.4.

The average number of times for the 12 boys that were asked was 2.5.

Work out the average number of times the girls in the group visited the cinema in the past year.

- 6 A fitness test was taken by 25 girls and 52 boys.

The average fitness score for the girls was 6.4, and the average fitness score for the boys was 9.2.

Work out the average fitness score for the whole group.

D1.6 Large data sets – averages and range

This spread will show you how to:

- Use frequency tables to find the averages and range of a data set
- Use estimates of averages and range to summarise large data sets

Keywords

Frequency table
Mean
Median
Mode
Range

- You can put large amounts of data in a **frequency table**.
- You use the averages and the **range** to summarise the data.

Example

The table shows the length of the words in the answers to a crossword puzzle.

For these data, work out the

- mode
- median
- mean
- range.

Word length	Frequency
4	3
5	5
6	7
7	8
8	3
9	1

Word length	Frequency	Word length × frequency
4	3	$4 \times 3 = 12$
5	5	$5 \times 5 = 25$
6	7	$6 \times 7 = 42$
7	8	$7 \times 8 = 56$
8	3	$8 \times 3 = 24$
9	1	$9 \times 1 = 9$
Total	27	168

Total number of words.

Total number of letters.

a Mode = 7 Words with 7 letters have the highest frequency.

b $\frac{1}{2}(27 + 1) = 14$ so the 14th value is the median
The 14th value is in the 'Word length 6' group.
Median = 6

c Mean = $\frac{\text{Total number of letters}}{\text{Total number of words}}$
= $168 \div 27 = 6.2$

d Range = $9 - 4 = 5$ Longest – shortest word length.

p.280

Exercise D1.6

Grade D/C

- 1** The tables of data give information about the length of words in four different crosswords.

For each table, copy the table, add an extra working column and find the

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

a

Word length	Frequency
4	2
5	5
6	4
7	2
8	2

b

Word length	Frequency
3	3
4	4
5	9
6	5
7	2

c

Word length	Frequency
4	6
5	3
6	5
7	4
8	2
9	5
10	2

d

Word length	Frequency
3	5
4	4
5	6
6	7
7	7
8	4
9	2

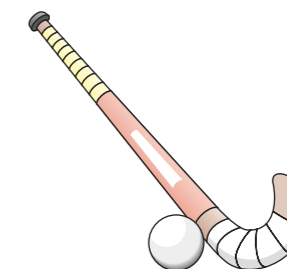
- 2** Jo had 16 boxes of matches. She counted the number of matches in each box. The table gives her results.

Number of matches	Frequency
41	2
42	7
43	4
44	3

Work out the mean number of matches in a box.

- 3** Brian played in 24 hockey matches one season. The table gives information about the number of goals scored in these matches.

Number of goals scored	Frequency
1	6
2	9
3	4
4	2
5	3



Work out the mean number of goals scored.

Summary

Check out

You should now be able to:

- Design an experiment or survey
- Design a questionnaire and use data collection sheets and two-way tables
- Select and justify methods of sampling to investigate a population, including random sampling
- Identify possible sources of bias
- Calculate averages and range of data sets with discrete and continuous data

Worked exam question

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

“How many times do you go to the cinema?”

Not very often

Sometimes

A lot

- a** Write down two things wrong with this question. (2)
- b** Design a better question for her questionnaire to find out how often adults go to the cinema. (2)
You should include some response boxes. (Edexcel Limited 2008)

- a**
- 1 ‘Not very often’, ‘Sometimes’ and ‘A lot’ are vague and will be misunderstood.
 - 2 There is not enough choice of responses. More response boxes are needed.

Write down two reasons.

Another answer for **a** could be: No mention of time in the question.

- b**
- How many times did you go to the cinema last month?
- 0 1–2 3–4 5 or over
-

The question includes a time period.

The response boxes cover all possibilities.

Exam question

A02

- 1** Ali found out the number of rooms in each of 40 houses in a town. He used the information to complete the frequency table.

Number of rooms	Frequency
4	4
5	7
6	10
7	12
8	5
9	2

Ali said the mode is 9
Ali is wrong.

- a** Explain why. (1)
- b** Calculate the mean number of rooms. (3)
- c** Beccy found out the number of rooms in each of 80 houses in the same town. She used the information to complete the frequency table below.

Number of rooms	Frequency
4	10
5	12
6	15
7	18
8	17
9	8

- Find the median number of rooms. (1)
- d** The median number of rooms in Ali’s table is 6

Which of the two medians, Ali’s or Beccy’s, is more likely to give the more reliable estimate for the median number of rooms for a house in this town?
Give a reason for your answer. (1)

(Edexcel Limited 2007)