

Types of Data, Sampling and Bias

Solutions

For Pearson Edexcel GCSE Mathematics (1MA1) • Foundation tier

Each question is shown in grey, followed by the working, one step per line. The final answer is in a box. *Marks* shows how the marks are usually given in the exam, and **Tip** points out a common mistake.

1 (3 marks)

Here are some types of data.
For each one, tick the correct box.

	Categorical	Discrete	Continuous
Shoe size		✓	
Temperature at noon			✓
Colour of a car	✓		
Number of goals scored		✓	
Length of a leaf			✓

Categorical data is words, not numbers.

Discrete data can only take certain values (usually counted).

Continuous data is measured, so it can take any value.

Shoe size: only certain values (e.g. 6, $6\frac{1}{2}$, 7), so discrete.

Temperature at noon: measured, so continuous.

Colour of a car: a word, so categorical.

Number of goals scored: counted, so discrete.

Length of a leaf: measured, so continuous.

Answer: Discrete, Continuous, Categorical, Discrete, Continuous

Tip: Shoe size is a number, but you can't have a size of 6.3. It can only take certain values, so it is discrete.

Marks: 3 marks for all five correct. 2 marks for four correct. 1 mark for three correct.

2 (1 mark)

Here are four types of data.

- A* the mass of a parcel
- B* the time taken to walk to school
- C* the number of people in a car
- D* the height of a tree

Which one is discrete data?

Write down the letter.

You count the number of people in a car, so it is discrete.

Mass, time and height are measured, so they are continuous.

Answer: *C*

Tip: Time is measured, so it is continuous, not discrete.

3 (2 marks)

Leo wants to know the favourite subject of students at his school.

He asks 40 students in Year 7.

- (a) Is Leo's sample likely to be representative of the whole school?
Give a reason for your answer.

No.

He only asked Year 7 students.

Students in other years may have different favourite subjects.

Answer: No, only Year 7

Tip: Give a reason that is about this sample: it is only Year 7. A general comment like “it is too small” is not enough here.

- (b) Give one way Leo could improve his sample.

Ask students from every year group.

Choose them at random.

Answer: Random students from all years

4 (2 marks)

The council wants to know how people in a town travel to work.
It asks 30 people at the train station.

- (a) Give one reason why this may not be a good sample.

People at a train station are more likely to travel by train.

So the sample is biased towards train users.

Answer: Biased: people at the station use trains

- (b) A councillor says,

‘If we ask 300 people at the station, the sample will not be biased.’

Is the councillor correct?

You must give a reason for your answer.

Asking more people can make the results more reliable.

But all 300 people would still be at a train station.

They are still more likely to travel by train than other people in the town.

So the sample would still be biased.

Answer: No. Everyone asked would still be at the station, so the sample would still be biased towards train users.

Tip: A bigger sample does not remove bias. Bias comes from **where** and **who** you ask, not from **how many** people you ask.

5 (1 mark)

Tom and Lisa each want to find out how many people in their town own a bike.

Tom asks 10 people chosen at random.

Lisa asks 80 people chosen at random.

Whose results are more likely to represent all the people in the town, Tom's or Lisa's?
You must give a reason for your answer.

Both samples are chosen at random.

Lisa asked more people (80 instead of 10).

A larger sample is more likely to represent the whole town.

Answer: Lisa's, larger sample

6 (2 marks)

A factory made 1600 phone cases one day.

A manager checked a sample of 40 of these phone cases.

3 of the phone cases in the sample were faulty.

Work out an estimate for the number of faulty phone cases made that day.

In the sample, 3 out of 40 phone cases are faulty.

The factory made $1600 \div 40 = 40$ times as many phone cases as the sample.

$$3 \times 40 = 120$$

Answer: 120

Tip: Scale up from the sample to the whole day. Do not just multiply by 3 or divide 1600 by 3.

Marks: 1 mark for $\frac{3}{40} \times 1600$ or $1600 \div 40 = 40$. 1 mark for 120.

7 (3 marks)

Orla runs a café with 600 regular customers.

She is going to sell one new smoothie flavour.

She asks 40 of her regular customers to choose their favourite of four flavours.

Here are the results.

Flavour	Number of customers
Mango	14
Berry	11
Banana	9
Kiwi	6

- (a) Work out an estimate for how many more of the 600 customers would choose mango than kiwi.

In the sample, $14 - 6 = 8$ more customers chose mango than kiwi.

There are $600 \div 40 = 15$ times as many regular customers as there are in the sample.

$$8 \times 15 = 120$$

Answer: 120

Tip: You can scale up each flavour first ($14 \times 15 = 210$ and $6 \times 15 = 90$), then subtract. Either way, remember to scale up to all 600 customers.

Marks: 1 mark for $\frac{8}{40} \times 600$, or for 210 and 90. 1 mark for 120.

- (b) State any assumption you made and explain how this may affect your answer.

Assumption: the 40 customers in the sample are representative of all 600 regular customers.

Effect: if they are not, the real difference could be more or less than 120.

Answer: Sample is representative; if not, answer could be too high or too low

Tip: You need two things: the assumption **and** how it affects your answer (too high or too low).

8 (3 marks)

There are 3000 households in a town.

Hamid and Tess want to find out how many of these households have a pet.

Hamid asks 50 households chosen at random. 21 of them have a pet.

Tess asks 150 households chosen at random. 57 of them have a pet.

- (a) Use all of these results to work out an estimate for the number of households in the town that have a pet.

Altogether, $50 + 150 = 200$ households were asked.

Altogether, $21 + 57 = 78$ of them have a pet.

$$\frac{78}{200} \times 3000 = 1170$$

Answer: 1170

Tip: Add the two samples together first. Do not add the two fractions $\frac{21}{50} + \frac{57}{150}$.

Marks: 1 mark for $\frac{78}{200} \times 3000$. 1 mark for 1170.

(b) Tess says,

‘My results show that exactly 38% of the households in the town have a pet.’

Is Tess correct?

You must give a reason for your answer.

$\frac{57}{150} = 0.38$, so 38% of the households in Tess’s **sample** have a pet.

She only asked 150 of the 3000 households.

A sample only gives an estimate. A different sample can give a different answer:

Hamid’s sample gives $\frac{21}{50} = 42\%$.

So 38% is not exact, and Tess is not correct.

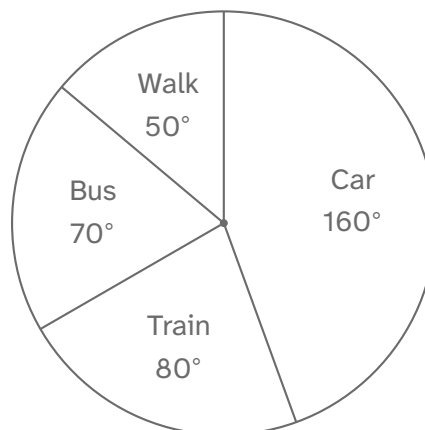
Answer: No. Her 38% comes from a sample, so it is only an estimate for the whole town (Hamid’s sample gives 42%)

Tip: A sample tells you about the sample. You can use it to **estimate** for the whole population, but the true value is very unlikely to be exactly the same.

9 (3 marks)

A sample of the visitors to a museum one month were asked how they travelled to the museum.

The pie chart shows information about their answers.



(a) There were 4500 visitors to the museum that month.

Work out an estimate for the number of these visitors who came by train.

The angle for train is 80° .

The whole circle is 360° .

$$\frac{80}{360} \times 4500 = 1000$$

Answer: 1000

Marks: 1 mark for $\frac{80}{360} \times 4500$. 1 mark for 1000.

- (b) All the visitors in the sample were asked on the same Saturday morning. Explain how this might affect the accuracy of your estimate.

People visiting on a Saturday morning may travel differently from people on weekdays.

So the sample may not be representative of all the visitors.

The estimate may be too high or too low.

Answer: Sample may not be representative, so estimate may be inaccurate

Tip: Link your answer to the day and time: only one Saturday morning was sampled. That is what makes it unrepresentative.
