AQA - Data Collection – GCSE Statistics – 2021

1. June/2021/Paper_1F/No.7

(a) As part of a school project Hakeeb asks 10 of his friends to write down how many **hours** they slept last Sunday night.

These are the 10 values given by his friends.

6	8	6	480	7	9	7	8.5	8	6

(a) (i) Identify the value which appears to be incorrect.

[1 mark]

Answer _____

(a) (ii) Suggest, in context, what might have happened and write down the correct value.

[2 marks]

What might have happened _____

_____

_____

Correct value _____

(b) Here is part of a statement seen in a text book.

'Raw data sometimes need to be 'cleaned' so that...'

(b) (i) What are raw data?

[1 mark]

_____

_____

_____

(b) (ii) What does 'cleaned' mean in this statement?

[1 mark]

_____

_____

_____

(b) (iii) Complete the sentence from the book to give a reason **why** cleaning may take place.

[1 mark]

'Raw data sometimes need to be 'cleaned' so that…' _____

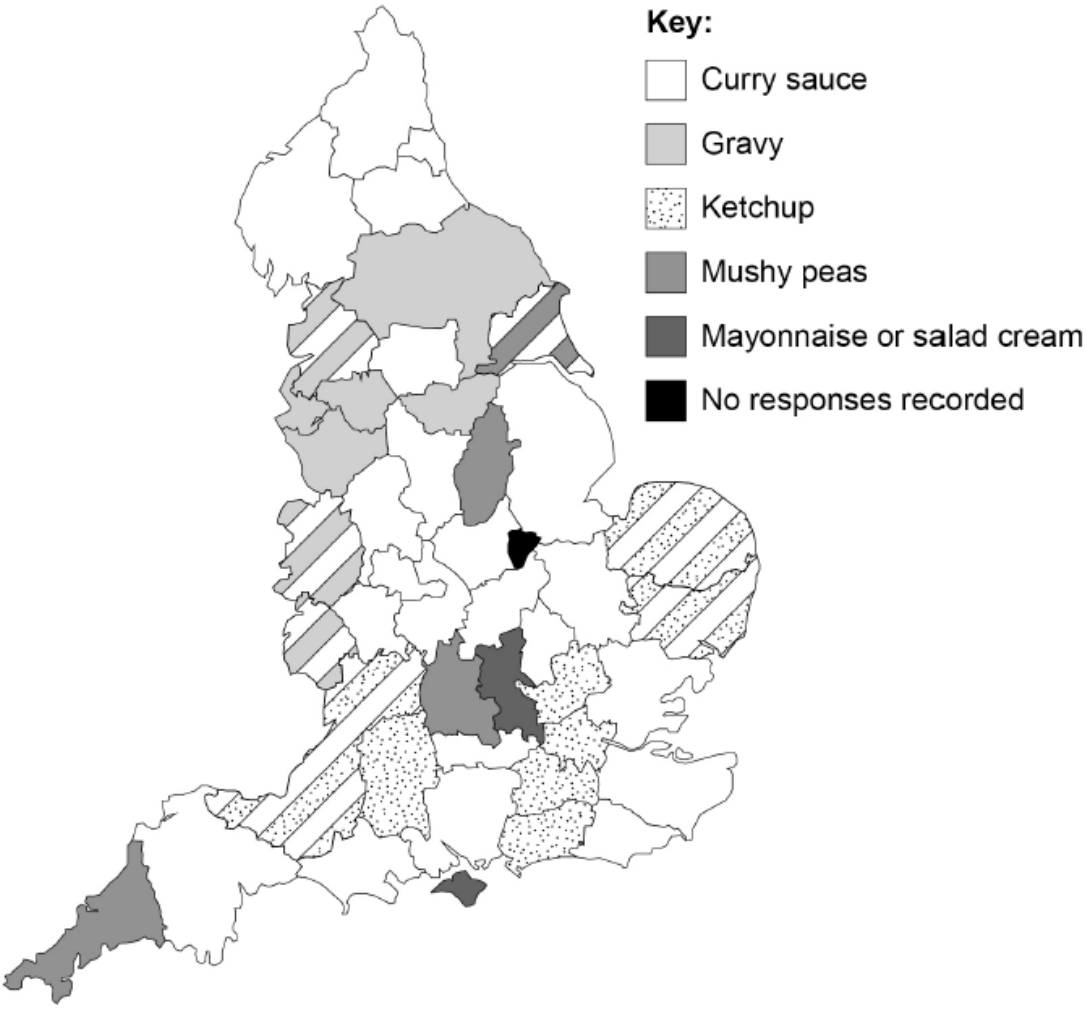
_____

_____

2. June/2021/Paper_1F/No.13

A sample of 670 adults in England were asked what side order they preferred at fish and chip shops.

A striped area indicates two equally popular side orders.



(a) Based on area of land, which is the most popular side order?

[1 mark]

Answer _____

(b) Give **two** reasons why your answer to **part (a)** might not be the side order that **most** people eating fish and chips in England prefer.

[2 marks]

1 _____

_____

2 _____

_____

3. June/2021/Paper_1H/No.2

Here is the definition of a term used in sampling.

'Those who are actually available to be part of a survey or investigation.'

Circle the term being defined.

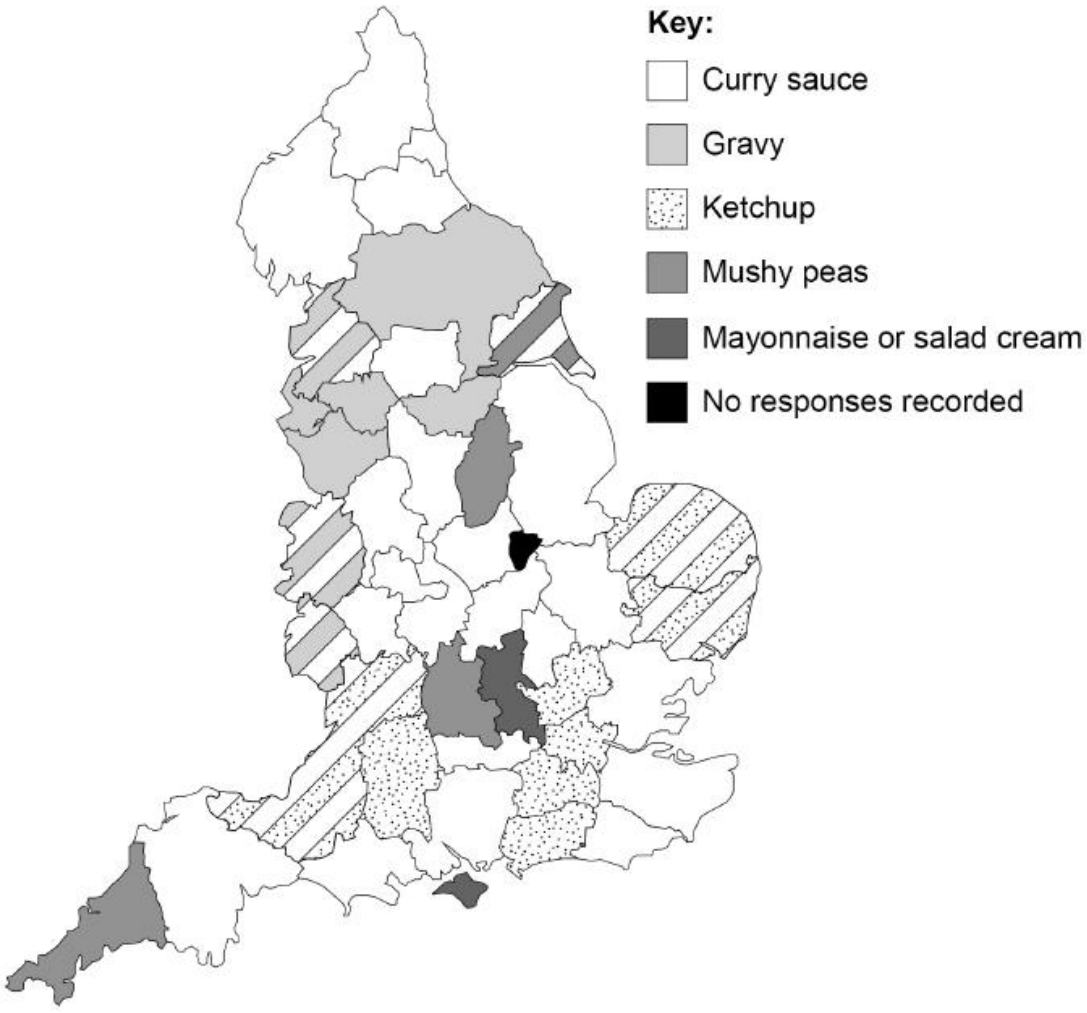
[1 mark]

sample frame census group population

4. June/2021/Paper_1H/No.5

A sample of 670 adults in England were asked what side order they preferred at fish and chip shops.

A striped area indicates two equally popular side orders.



(a) Based on area of land, which is the most popular side order?

[1 mark]

Answer _____

(b) Give **two** reasons why your answer to **(a)** might not be the side order that **most** people eating fish and chips in England prefer.

[2 marks]

1 _____

_____

2 _____

_____

5. June/2021/Paper_1H/No.6(c)

(c) Tom wants a sample of Year 7 students and a sample of Year 11 students to complete a questionnaire for him.

He considers these three sampling methods for Year 7 students.

Method A

Number all the students in Year 7.
Obtain 30 random numbers.
Ask the students whose random numbers come up to complete the questionnaire.

Method B

Wait outside the dinner hall.
Ask the first 30 Year 7 students he sees to complete the questionnaire.

Method C

Choose three Year 7 students from each of the 10 maths sets.
Ask these students to complete his questionnaire.

Name and compare the merits of each sampling method.

Make a reasoned choice of which method Tom should use.

[7 marks]

6. June/2021/Paper_2F/No.1

Which of these sets of data has a different range to the others?

Circle your answer.

[1 mark]

1, 6, 6, 6

2, 3, 5, 7

3, 5, 6, 8

4, 4, 4, 8

7. June/2021/Paper_2F/No.3

Which of these diagrams is suitable for bivariate data?

Circle your answer.

[1 mark]

scatter diagram choropleth map bar chart box plot

8. June/2021/Paper_2F/No.10

A Sixth Form college has 1000 students.

Students on different courses have different numbers of lessons.

Ben and Matt are investigating the hypothesis,

'Students with better GCSE grades have more lessons per week at the college.'

Ben is collecting the information about GCSE grades.

Matt is collecting the information about the number of lessons the students have.

(a) What type of data is 'numbers of lessons'?

Circle your answer.

[1 mark]

ordinal bivariate discrete continuous

(b) Ben gets a list of all the 1000 students who go to the college.

He decides to choose a systematic sample of 50 students using these steps.

Step 1 Start at the 25th student

Step 2 Pick every 50th student on the list

Ben has made an error in each step.

Write down a corrected version of each step.

[2 marks]

Step 1 _____

_____

Step 2 _____

_____

(c) Matt says,

> "I'm going to choose a different sample using random sampling to get data about the number of lessons students have."

Why is this not a good idea?

[1 mark]

_____

_____

(d) Ben and Matt correct their errors and collect appropriate data.

They input all the data to a spreadsheet.

Give **one** reason why it might be helpful to have the data in a spreadsheet.

[1 mark]

(e) Here are the top few rows of the spreadsheet.

Student	Average GCSE grade	Number of lessons per week
1	6	16
2	5	14
3	8	21
4	5	17
5	4	16
6	4	14
7	7	199
8	5	17
9	8	22
10	4	15
11	7	20

(e) (i) Identify the wrongly recorded value.

[1 mark]

Answer _____

(e) (ii) Suggest what they should do with this value.

[1 mark]

(e) (iii) Based on the data you can see, comment on the original hypothesis.

[1 mark]

9. June/2021/Paper_2H/No.2

Which of the following is a measure of the **change** in the cost of goods and services?

Circle your answer.

[1 mark]

Standardised score	Gross domestic product
Average seasonal effect	Consumer price index

10. June/2021/Paper_2H/No.15

You will need the **Data Sheet** to answer this question.

Mark and Fiona carry out roadside safety checks on trucks.

For each truck, they record,

- the number of wheels
- its length
- its mass.

(a) Circle the name given to the data that they collect.

[1 mark]

bivariate multivariate secondary qualitative

(b) Mark records data from trucks travelling on one part of the A1 road.

He records data from,

- the first 20 trucks with 6 or fewer wheels
- the first 20 trucks with more than 6 wheels.

His data are given on the **Data Sheet**.

(b) (i) State **one** problem with the data Mark has collected.

Suggest a solution to deal with this problem.

[2 marks]

Problem _____

_____

Solution _____

_____

(b) (ii) Mark wants to use his data to estimate the mean mass of trucks using this part of the A1 road.

Explain why the data Mark has collected is **not** likely to be suitable for this purpose.

[1 mark]

_____

_____

_____

(c) Fiona carries out her checks on two roads, the A2 and the A229.

(c) (i) Some summary statistics for the lengths of trucks she checks on the **A2** are shown.

mean	10.20 metres
median	9.18 metres
standard deviation (s.d.)	2.90 metres

Calculate the skew of the data.

Use $\text{skew} = \dfrac{3(\text{mean} - \text{median})}{\text{s.d.}}$

[2 marks]

Answer _____

(c) (ii) Fiona says,

> "The data show positive skew, so the trucks below median length have more variable lengths than the trucks above median length."

Has Fiona interpreted the skew correctly?

Tick (✓) a box.

Yes ☐ No ☐

Explain your answer.

[1 mark]

(d)

$$\text{s.d.} = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

Fiona checks the lengths of **30** trucks on the **A229**.

The lengths, x (metres), of these 30 trucks are summarised by

$$\sum x = 267.12 \qquad \sum x^2 = 2538.52 \qquad \text{skew} = 0.43$$

Compare statistically the length of trucks checked on the A229 with those checked on the A2.

You **must** show your working.

[6 marks]

(e) Give a reason why the standard deviation may not be the best measure of spread to summarise Fiona's data.

[1 mark]