AQA

Please write clearly in block capitals.

Centre number | | | | | | Candidate number | | | | |

Surname _____

Forename(s) _____

Candidate signature _____

I declare this is my own work.

# GCSE STATISTICS

**H**

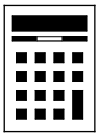
Higher Tier Paper 2

Monday 17 June 2024 Afternoon Time allowed: 1 hour 45 minutes

**Materials**

For this paper you must have:

- a calculator
- mathematical instruments.



**Instructions**

- Use black ink or black ball-point pen. Draw diagrams in pencil.
- Fill in the boxes at the top of this page.
- Answer **all** questions.
- You must answer the questions in the spaces provided. Do not write outside the box around each page or on blank pages.
- If you need extra space for your answer(s), use the lined pages at the end of this book. Write the question number against your answer(s).
- Do all rough work in this book. Cross out any work you do not want to be marked.

**Information**

- The marks for questions are shown in brackets.
- The maximum mark for this paper is 80.
- You may ask for more answer paper and graph paper. These must be tagged securely to this answer booklet.

| For Examiner's Use | |
|---|---|
| Question | Mark |
| 1-4 | |
| 5 | |
| 6 | |
| 7 | |
| 8 | |
| 9 | |
| 10 | |
| 11 | |
| 12 | |
| 13 | |
| 14 | |
| 15 | |
| 16 | |
| 17 | |
| 18 | |
| 19 | |
| **TOTAL** | |

Answer **all** questions in the spaces provided.

*Do not write outside the box*

**1** The index number for the price of a bike is

- 100 in 2020
- 97 in 2022.

What is the percentage change in price from 2020 to 2022?

Circle your answer.

**[1 mark]**

Decrease of 3% Decrease of 97% Increase of 3% Increase of 97%

**2** In a town, 30% of the population are children.

An interviewer decides to collect data in the town centre from a sample of

15 children

and

35 people who are not children.

Which name best describes the interviewer's sampling method?

Circle your answer.

**[1 mark]**

Convenience Quota Random Systematic

*Do not write outside the box*

**3** Matteus is researching the history of his village.

He finds a record of the people who died in 1909 and draws this stem and leaf diagram.

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

**Key:** 1 | 2 = 12 years

Which **one** of these statements about the people who died in 1909 is true?

Tick (✓) a box.

**[1 mark]**

Five people died in their 30s. ☐

More people died younger than 45 years than died older than 45 years. ☐

Nobody died that was older than 70 years. ☐

The youngest person was aged 6 years. ☐

**4** Circle the pair of numbers that do **not** have a geometric mean.

**[1 mark]**

–4, 6 0.5, 6 1, 6 4, 6

4

**Turn over ►**

*Do not write outside the box*

**5** Arne has produced a new fitness app.

He records the number of downloads per week, for 8 consecutive weeks.

| **Week** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| **Downloads** | 500 | 520 | 580 | 600 | 680 | 720 | 600 | 840 |

**5 (a)** Arne says,

"520 people used the app in week 2."

Is he correct?

Tick (✓) a box.

Yes ☐ No ☐ Cannot tell ☐

Give a reason for your answer.

**[1 mark]**

**5 (b)** Here are the data again, with some 4-point moving averages completed.

| **Week** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| **Downloads** | 500 | 520 | 580 | 600 | 680 | 720 | 600 | 840 |
| **Moving Average** | | 550 | 595 | 645 | 650 | | | |

Complete the table by calculating the missing 4-point moving average.

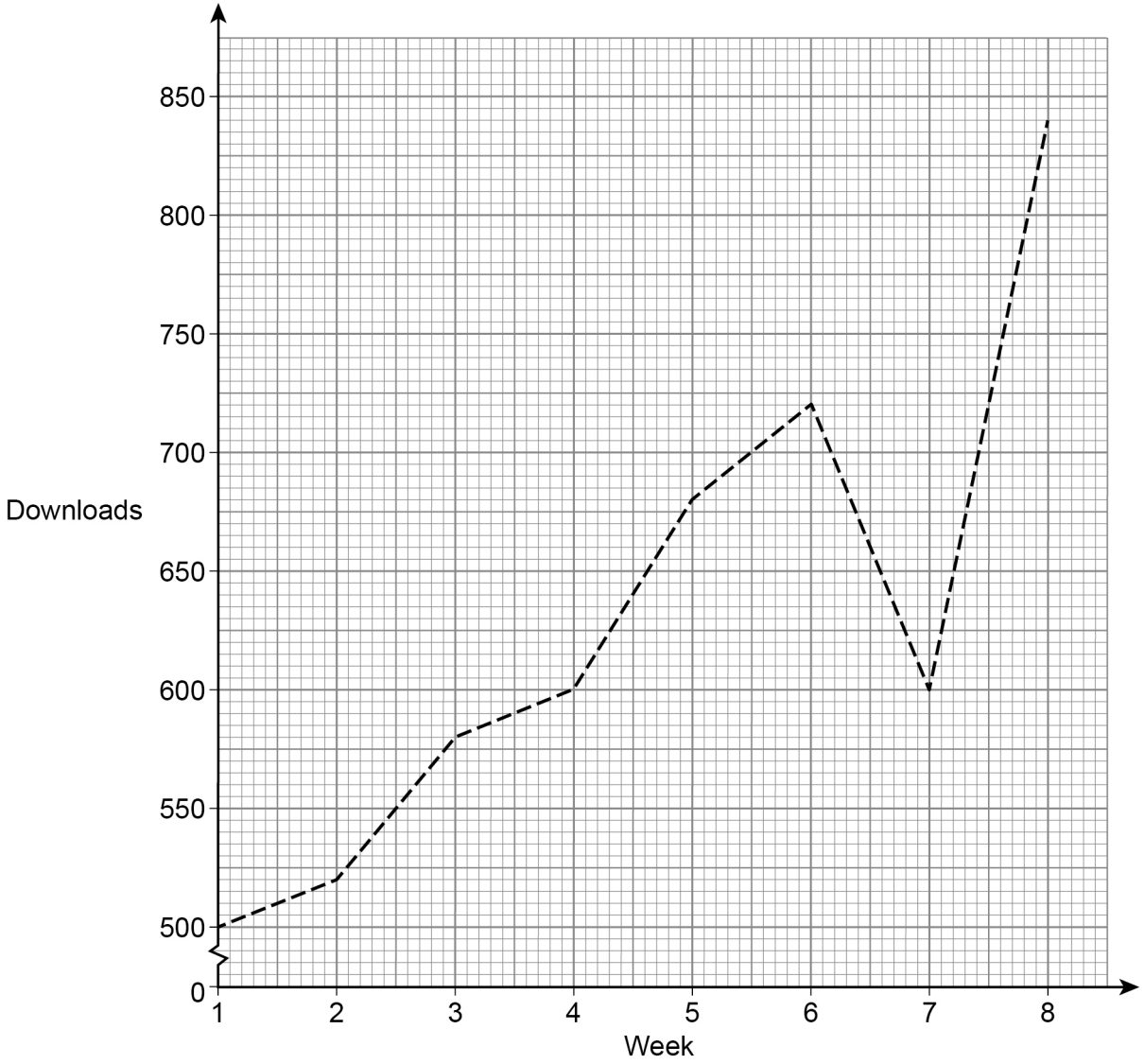
**[2 marks]**

*Do not write outside the box*

**5 (c)** The time series graph shows the data for the 8 weeks.

Plot the 4-point moving averages **and** draw the trend line.

**[3 marks]**




**5 (d)** Describe the trend, in context.

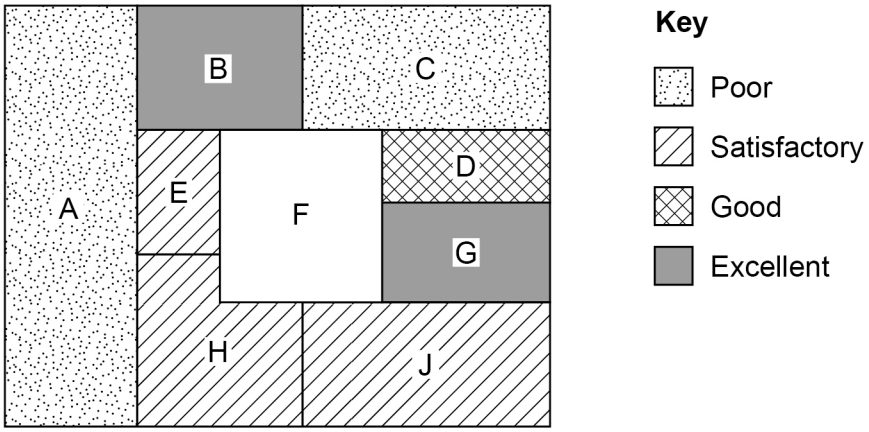
**[1 mark]**

7

*Do not write outside the box*

**6** A farmer rates how well crops grow in her nine fields.

The choropleth map represents the fields on her farm and some of her ratings.




**6 (a)** The farmer has the same number of fields rated good as are rated excellent.

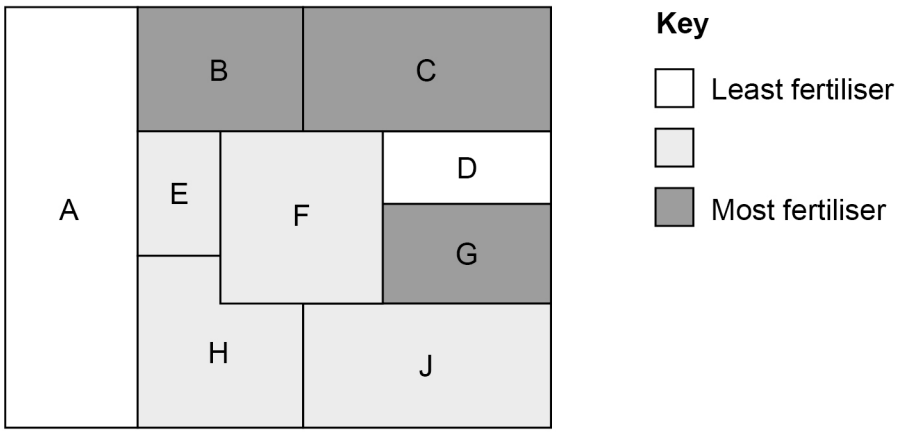
Complete field F on the choropleth map.

**[1 mark]**

*Do not write outside the box*

**6 (b)** Here is a different choropleth map of the same nine fields.

This map shows the amount of fertiliser the farmer used on each field.




The farmer makes the conclusion,

"The fields that had more fertiliser produced better quality crops."

Comment on her conclusion by comparing the **two** choropleth maps.

You should give a reason for and a reason against her conclusion.

**[2 marks]**

Reason for _____

_____

_____

_____

_____

Reason against _____

_____

_____

_____

_____

**Question 6 continues on the next page**

*Do not write outside the box*

**6 (c)** Give **one** way the farmer could gain a greater amount of detail from the choropleth maps.

**[1 mark]**

**6 (d)** Write down **one** possible extraneous variable.

**[1 mark]**

5

*Do not write outside the box*

**7** A type of fraud is when money is taken from you without your permission.

Money can be taken from your bank account, when shopping, or in other ways.

These frauds can take place online or offline.

The table shows data, in thousands, about some of these frauds that took place in 2020–2021.

| | **Bank account** | **Shopping** | **Other** | **Total** |
|---|---|---|---|---|
| **Online** | 809 | 853 | 240 | **1902** |
| **Offline** | 1073 | 116 | 181 | **1370** |
| **Total** | **1882** | **969** | **421** | **3272** |

Adapted from ONS

One of these frauds is selected at random.

**7 (a)** Write down the probability that it happened online.

**[1 mark]**

Answer _____

**7 (b)** Work out the probability that the fraud was **not** a bank account fraud.

**[2 marks]**

_____

_____

Answer _____

**7 (c)** Work out the probability that it was a shopping fraud given that it happened offline.

**[2 marks]**

_____

_____

_____

Answer _____

**5**

*Do not write outside the box*

**8** 342 teenagers were asked if they had ever spent a night alone at home.

58 of these teenagers said that they had.

**8 (a)** Calculate the percentage of these teenagers who said that they had spent a night alone at home.

**[2 marks]**

Answer _____ %

**8 (b)** Estimate the percentage of **all** teenagers who have spent a night alone at home.

Answer _____ %

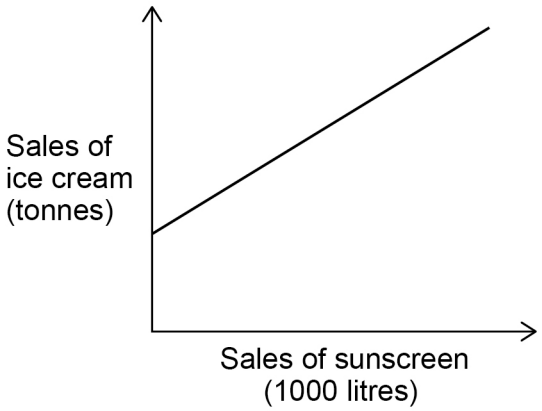
State one assumption you have made.

**[2 marks]**

**4**

*Do not write outside the box*

**9** The diagram shows information about sales of ice cream and sales of sunscreen.




Russ says,

"Increased sunscreen sales are causing increased ice cream sales."

He is not correct.

Write down a factor that may affect **both** sales of sunscreen and sales of ice cream.

**[1 mark]**

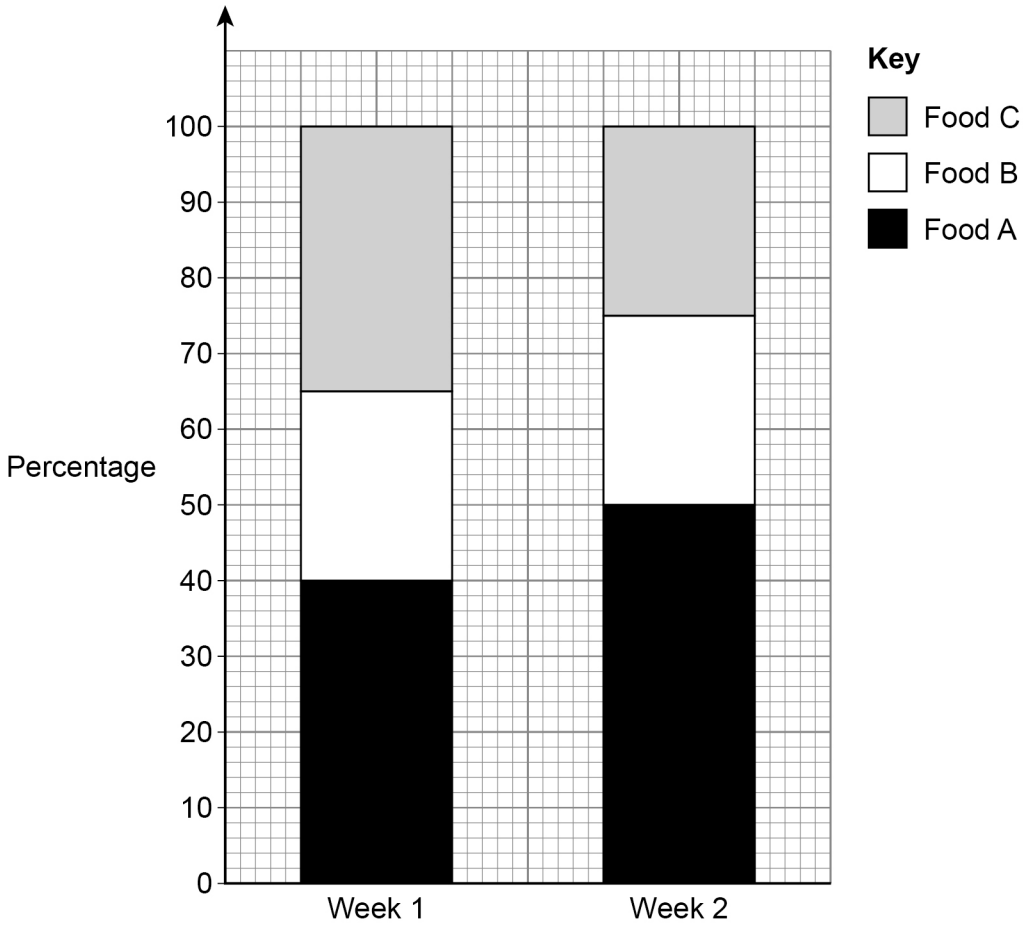
1

**Turn over for the next question**

*Do not write outside the box*

**10** Shona sells bags of three different types of dog food.

The percentage bar chart shows information about the sales of each type, for two weeks.




**10 (a)** Make **two** comparisons of the sales of dog food in Week 1 with the sales in Week 2.

**[2 marks]**

Comparison 1 _____

_____

_____

Comparison 2 _____

_____

_____

*Do not write outside the box*

**10 (b)** Shona says,

"The percentage of food sold in Week 1 that was Food B is 65%."

Comment on this statement.

**[1 mark]**

3

**Turn over for the next question**

*Do not write outside the box*

**11** 300 runners take part in a parkrun.

Franco is investigating how far the runners travel to get to the parkrun.

The table shows information about how they travel to the parkrun.

| | **Car** | **Walk** | **Cycle** | **Other** |
|---|---|---|---|---|
| **Child** | 90 | 28 | 41 | 2 |
| **Adult** | 108 | 12 | 16 | 3 |

Franco uses stratification before sampling 50 runners.

**11 (a)** Give a reason why stratification is appropriate.

**[1 mark]**

**11 (b)** Franco says he needs 33 runners in his sample who travel by car.

Is he correct?

Show working to support your answer.

Tick (✓) a box.

Yes ☐ No ☐

**[2 marks]**

*Do not write outside the box*

**11 (c)** Stef also uses stratification before sampling the same 300 runners.

In her sample, there are exactly 9 runners who are **adult** and travel by **car**.

Work out how many runners are in Stef's sample.

**[2 marks]**

Answer _____

5

**Turn over for the next question**

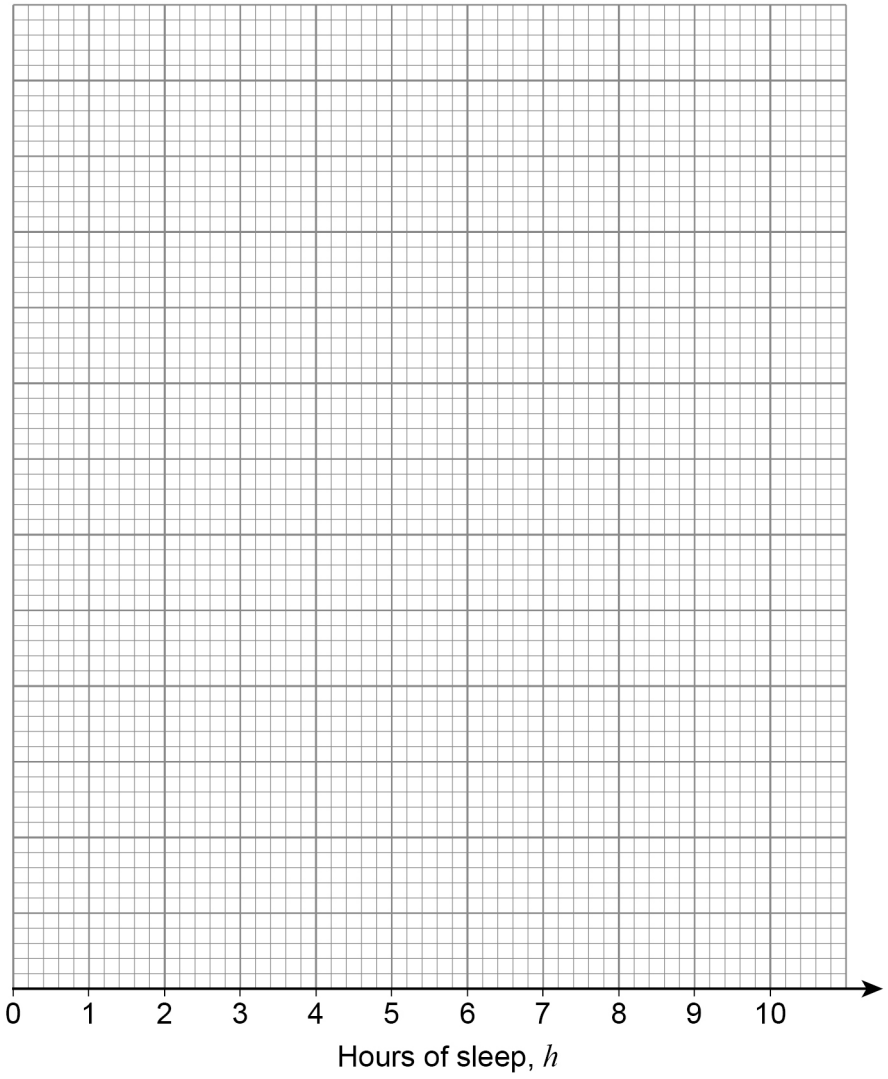
*Do not write outside the box*

**12** Mair is investigating how many hours people sleep per night.

Here is some information about how many hours 120 people slept for one night.

| **Hours of sleep, $h$** | $0 < h \leqslant 2$ | $2 < h \leqslant 4$ | $4 < h \leqslant 6$ | $6 < h \leqslant 8$ | $8 < h \leqslant 10$ |
|---|---|---|---|---|---|
| **Frequency** | 5 | 13 | 42 | 45 | 15 |

**12 (a)** On the grid, draw a cumulative frequency diagram for Mair's data.




**[4 marks]**

*Do not write outside the box*

**12 (b)** Estimate the probability that a person, chosen at random, slept for more than 7 hours that night.

**[2 marks]**

Answer _____

6

**Turn over for the next question**

*Do not write outside the box*

**13** A bag of sweets contains only Aniseed, Butterscotch and Chewies.

There are 40 sweets in the bag.

11 of the sweets are Aniseed.

16 of the sweets are Butterscotch.

**Two** sweets are taken from the bag at random without replacement.

Work out the probability that **at least one** of the sweets is a Chewie.

**[3 marks]**

Answer _____

3

*Do not write outside the box*

**14** A school wants to estimate the number of students who have missed school to avoid a test.

720 students are given this survey with a fair, six-sided dice.

Roll the dice once.

If you roll a six, answer the question below honestly.

If you do **not** roll a six, tick Yes.

Have you ever missed school to avoid a test?

Tick (✓) a box.

Yes ☐ No ☐

**14 (a)** Explain why this is an appropriate method for the school to use.

**[1 mark]**

**14 (b)** 634 students tick Yes.

Work out an estimate of how many students answer the question honestly **and** tick Yes.

**[1 mark]**

Answer _____

**2**

*Do not write outside the box*

**15** Clara is on a course that has three modules.

Each module has a different weighting.

Some information about the weightings and her marks is shown in the table.

| Module | Weighting | Mark |
|---|---|---|
| A | 30 | 26 |
| B | 40 | |
| C | 90 | 74 |

**15 (a)** The weighted mean of Clara's marks is 60

Work out Clara's mark for Module B.

**[5 marks]**

Answer _____

*Do not write outside the box*

**15 (b)** The course leader decides to reduce the weighting of each module by 10

Clara says,

> "The weighting of each module has been reduced by the same amount, so my weighted mean will not change."

Comment on her statement.

**[1 mark]**

6

**Turn over for the next question**

*Do not write outside the box*

**16** Dorota owns a café.

Customers either order to 'eat in' or order to 'take away'.

The probability that a customer will 'eat in' is 0.3

**16 (a)** Dorota says,

"There is a 70% probability that the next customer orders to 'take away'."

Comment on her statement.

**[1 mark]**

**16 (b)** Dorota selects a random sample of 5 orders.

**16 (b) (i)** Use the Binomial distribution to show,

"the probability that **exactly** 4 of these orders are 'eat in' is 0.02835".

**[2 marks]**

*Do not write outside the box*

**16 (b) (ii)** Use the Binomial distribution to calculate the probability that **fewer** than 4 of these orders are 'eat in'.

**[3 marks]**

Answer _____

**16 (b) (iii)** Dorota has assumed that each order is independent.

Comment upon the validity of her assumption.

**[1 mark]**

**7**

**Turn over for the next question**

*Do not write outside the box*

**17** A group of 10 students take a biology test.

They record the time ($x$ minutes) it takes them to complete the test.

Their times are summarised as

mean = 40.4 minutes $\quad\quad \sum x^2 = 17\,960$

**17 (a)** Calculate the standard deviation of the times the students take to complete the biology test.

Use the formula

$$\text{Standard deviation} = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

**[2 marks]**

Answer _____ minutes

**17 (b)** The same 10 students also take a chemistry test.

The times it takes these students to complete the chemistry test are summarised as

mean = 41.5 minutes $\quad\quad$ standard deviation = 18.7 minutes.

Compare statistically the times these students take to complete the two tests.

**[2 marks]**

Comparison 1 _____

Comparison 2 _____

4

*Do not write outside the box*

**18** In a data set,

median $= x$

mean $= 2x$

$x$ is a positive number.

Explain why the set of data **cannot** be normally distributed.

**[1 mark]**

1

**Turn over for the next question**

*Do not write outside the box*

**19** Sid and Bobbi are investigating what music they each like.

Sid has this hypothesis to investigate.

"We like similar music."

They both listen to 8 songs, A to H.

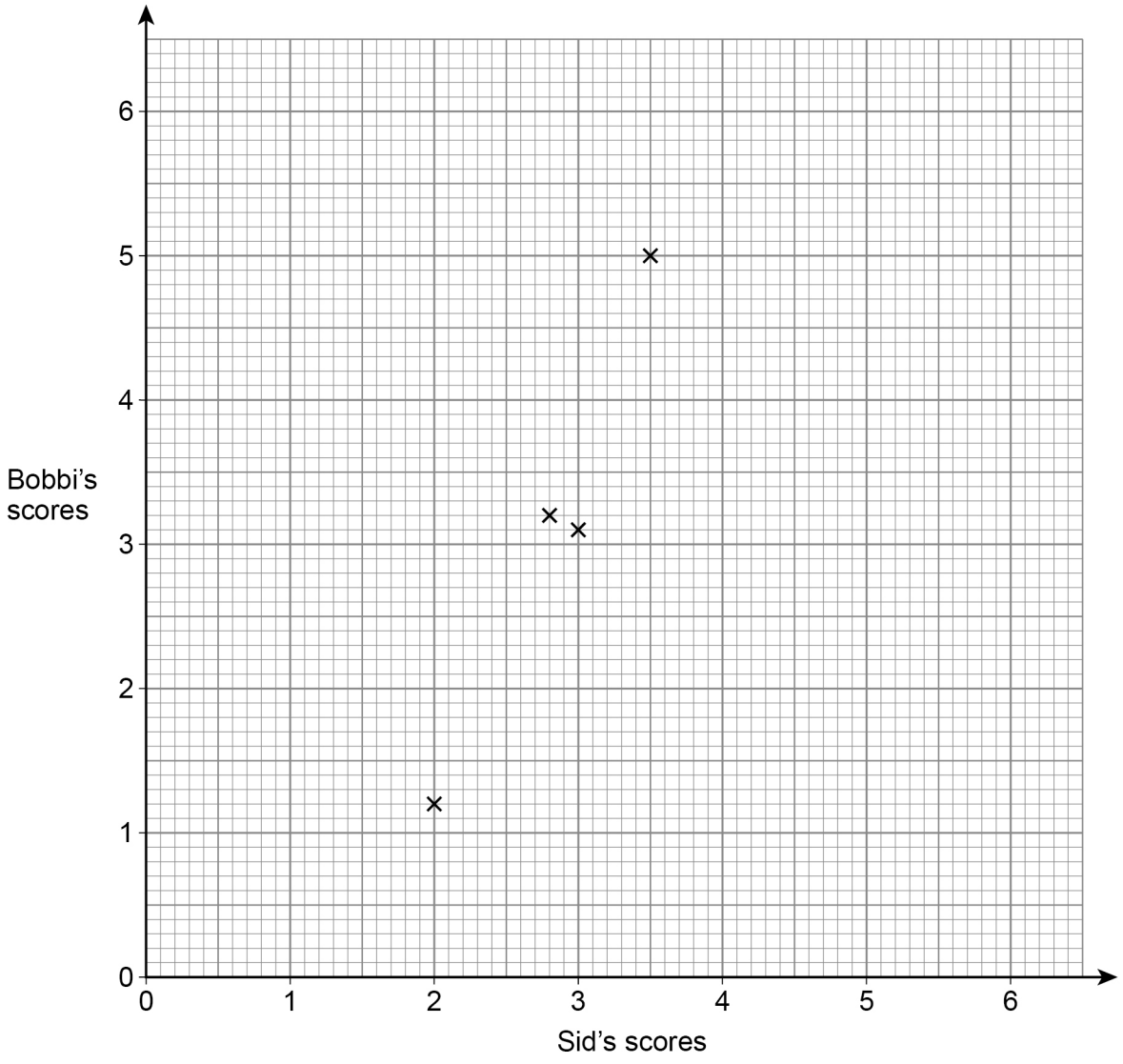
They each score the songs out of 6

| | **A** | **B** | **C** | **D** | **E** | **F** | **G** | **H** |
|---|---|---|---|---|---|---|---|---|
| Sid's scores | 3 | 3.5 | 2 | 2.8 | 1 | 4.7 | 0.7 | 3.9 |
| Bobbi's scores | 3.1 | 5 | 1.2 | 3.2 | 1.7 | 4.6 | 1 | 3.1 |

**19 (a)** Complete the scatter diagram.

The first four points have been plotted for you.

**[2 marks]**



*Do not write outside the box*

**19 (b)** What type of correlation is shown by the scatter diagram?

**[1 mark]**

Answer _____

**19 (c) (i)** By completing appropriate calculations to find the double mean point, draw a line of best fit on the diagram.

**[4 marks]**

**19 (c) (ii)** Sid gives a new song a score of 4.5

Use your line of best fit to estimate the score that Bobbi would give this song.

**[1 mark]**

Answer _____

**19 (c) (iii)** Bobbi gives a different song a score of 6

She extends the line of best fit and uses it to estimate the score that Sid would give for the song.

Explain why her estimate may **not** be reliable.

**[1 mark]**

**Question 19 continues on the next page**

*Do not write outside the box*

**19 (d)** Here are the original scores again.

| | **A** | **B** | **C** | **D** | **E** | **F** | **G** | **H** |
|---|---|---|---|---|---|---|---|---|
| Sid's scores | 3 | 3.5 | 2 | 2.8 | 1 | 4.7 | 0.7 | 3.9 |
| Bobbi's scores | 3.1 | 5 | 1.2 | 3.2 | 1.7 | 4.6 | 1 | 3.1 |

By calculating the value of Spearman's rank correlation coefficient between these two sets of scores, comment on Sid's hypothesis.

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

**[6 marks]**

*Do not write outside the box*

**19 (e)** Suggest **one** possible problem with using the given data to test Sid's hypothesis.

**[1 mark]**

**19 (f)** Bobbi says that they would get a more reliable value for Spearman's rank correlation coefficient if they give each song a score out of 10, instead of out of 6

Is she correct?

Tick (✓) a box.

Yes ☐ No ☐

Give a reason for your answer.

**[1 mark]**

**17**

**END OF QUESTIONS**

*Do not write outside the box*

**There are no questions printed on this page**

**DO NOT WRITE ON THIS PAGE**
**ANSWER IN THE SPACES PROVIDED**

*Do not write outside the box*

| Question number | **Additional page, if required.**<br>**Write the question numbers in the left-hand margin.** |
|---|---|
| | |

*Do not write outside the box*

| Question number | **Additional page, if required.**<br>**Write the question numbers in the left-hand margin.** |
|---|---|
| | |
| |  |