

**Statistics GCSE****Paper 1**

Edexcel Foundation - 2025

Notier Tier

Variant 1 (same as video)

1ST0/1F

**Instructions**

- Write all answers in the spaces provided.
- Answer all questions.
- You must show all your working.
- There may not be enough space to show all your working out.

**Information**

- This is a practise paper to aid your revision for your exams.
- This site, and all that work on it, have no affiliation or relationship with any exam board.
- This site is not endorsed by any company or charity, unless we state otherwise.




**Copyright**

- Sharing this PDF is strictly forbidden unless you have the author's permission.
- This paper is only authorised to be used by the person who has bought it.
- You may not store this online or any shared area such as an intranet.
- Please contact us if you have any queries.

**Advice**

- You can get support for all these questions at our website: [www.statsgcse.com](http://www.statsgcse.com)
- This paper and more are available on our site with questions that change subtly after each attempt.
- Good luck!

- 1 A bakery tracks the number of loaves of bread sold from Monday to Thursday. This helps them understand which days are busier and plan their stock accordingly.

Monday	
Tuesday	
Wednesday	
Thursday	
Friday	

**Key:**



represents 8 loaves of bread

- (a) Find the amount of loaves of bread sold on Wednesday.

(1 mark)

Make sure to look at the key.

- (b) Find the amount of loaves of bread sold on Monday.

(1 mark)





The square is split up into 4 parts. The key shows that four parts represent 8 loaves of bread. Start by finding what one part represents.

(c) On Thursday, the number of loaves of bread sold was 32.

Show this information on the pictogram.

(1 mark)

Select the correct answer.

<input type="checkbox"/>	Thursday	
<input type="checkbox"/>	Thursday	
<input type="checkbox"/>	Thursday	
<input type="checkbox"/>	Thursday	

(d) Felix suggests redrawing the pictogram using a key with a whole-square representing 5 loaves of bread.

Explain why this key would **not** be suitable.

(1 mark)

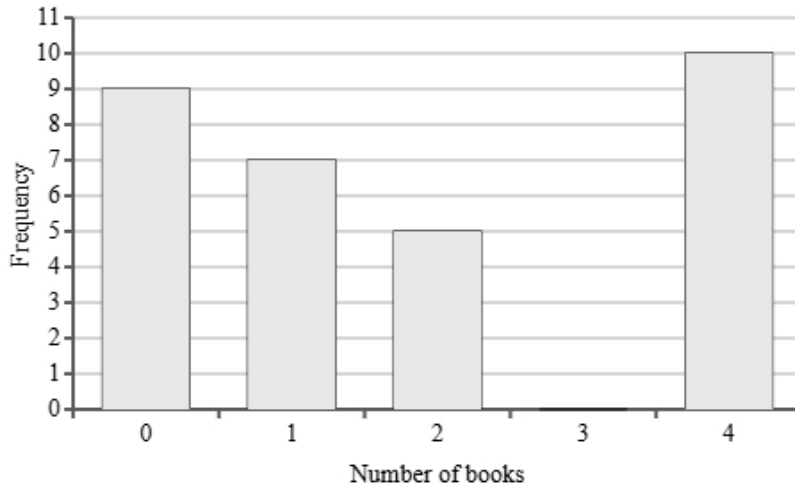
Select **one** box.

- Tuesday shows 6 loaves of bread. This would be very difficult to show because 6 has a remainder 1 when divided by 5.
- If Felix uses 5 loaves of bread for the key this will lead to less sales being shown.
- The key must always be an even number.
- This would be much better because you can fit more squares on.

2 Sarah has surveyed 35 students about the number of books they read in a month.

Each student read between 0 and 4 books.

The incomplete bar chart shows the number of students who read 0 books, 1 book, 2 books, and 4 books.

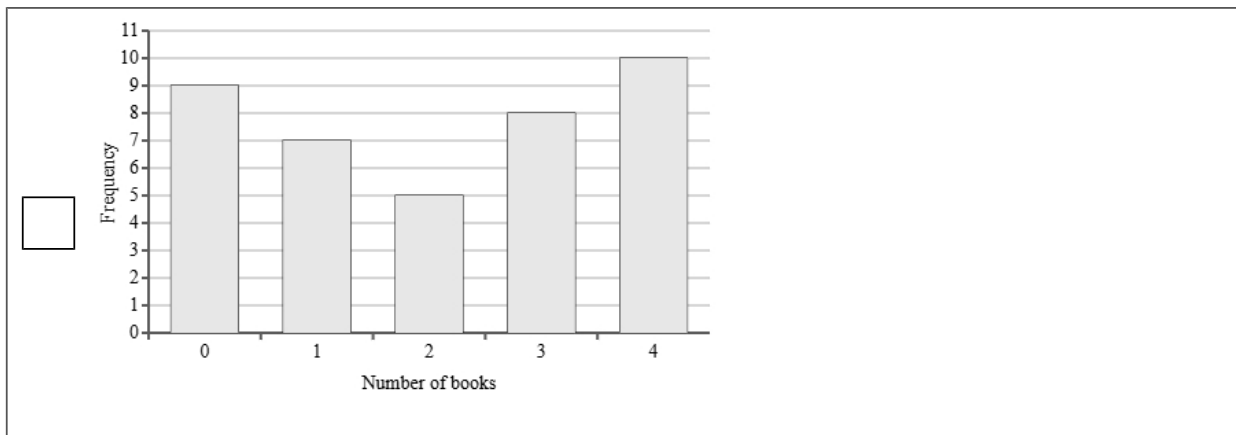
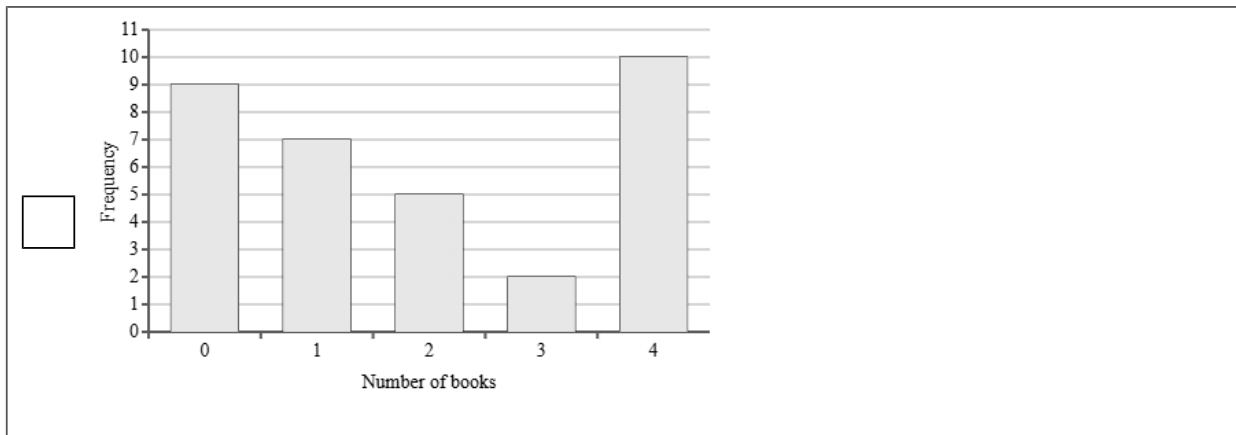
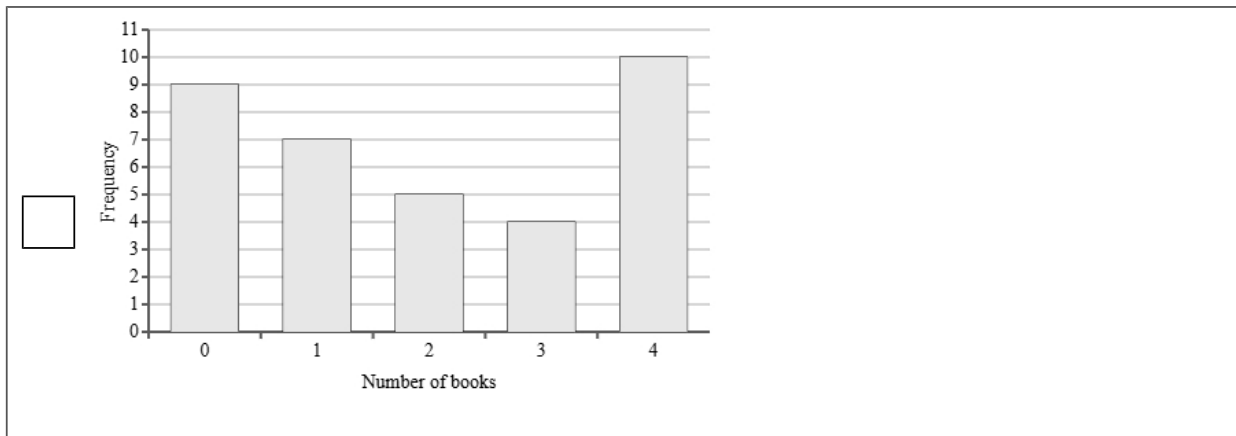
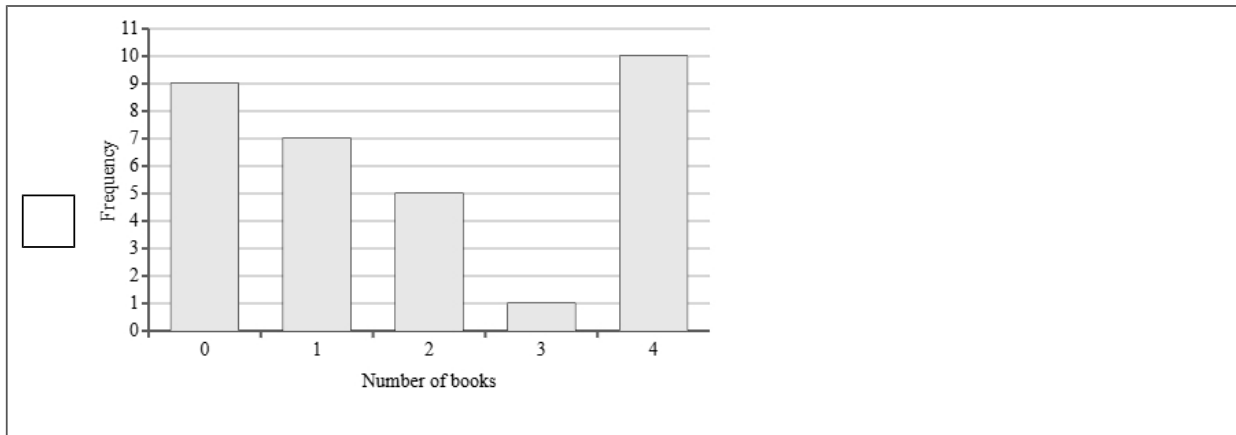


(a) 4 people read 3 books.

Complete the bar chart.

(1 mark)

Select the correct answer.



(b) Find how many more people read no books than 2 books.

(1 mark)

Find the difference between the bar heights (the frequencies) for 0 and 2.

(c) Sarah wants to open a library in the students' school.

She thinks that because most students read 4 books a month then a library will be popular.

Explain why Sarah may not be right.

(1 mark)

Select **one** box.

- The sample size was small.
- Students may have read more than 4 books a month.
- 9 students read no books.
- More students read one or less books a month.

**3** A school is considering introducing a new lunch menu.

Daniel is going to conduct a survey to gather opinions from everyone in the school.

Daniel thinks that he should take a sample rather than a census.

**(a)** Give two reasons why Daniel might think this.

(2 marks)

Select **two** boxes.

A sample is cheaper.

Daniel will be able to explain each question.

A sample is quicker.

Daniel will be able to choose who is in the sample.

A sample is more accurate.

**(b)** Daniel has decided to use the class registers as a sampling frame.

Explain what a sampling frame is.

(1 mark)

Select **one** box.

A list of all the members in the population.

The whole group.

A list of all the members in the sample.

The tally chart or table used in the survey.

(c) Daniel has decided to use the class registers as a sampling frame.

State one problem Daniel may have using the class registers as a sampling frame.

(1 mark)

Select **one** box.

- Unreliable.
- The class registers would also include dates of birth.
- The staff will not be included.
- There will be too many names.

(d) Give **two** reasons why Daniel should conduct a pilot survey.

(2 marks)

Select **two** boxes.

- A pilot survey will help identify problems.
- A pilot survey will give more accurate data.
- A pilot survey will include more people.
- A pilot survey will give an idea of what the results may be.
- A pilot survey will be cheaper than a survey.

(e) Daniel is writing a plan for the investigation into students' opinions on the new lunch menu.

Write down what Daniel should include in the plan and explain why each of the things is appropriate.

You should include:

- + a sampling method
- + a question Daniel should ask in the questionnaire
- + a statistical diagram to show the results of the survey.

(6 marks)

Number the **two** correct statements in the correct order (**two** statements are incorrect).

- Daniel should use stratified sampling.
- This will ensure that the sample is representative of the school.
- This will ensure that more students are asked.
- Daniel should use a case study.

Number the **two** correct statements in the correct order (**two** statements are incorrect).

A question could be:

- How satisfied are you with the current school lunch options?  
 Very Satisfied  Satisfied  Neutral  Unsatisfied  Very Unsatisfied

A question could be:

- Why do you think the school lunch options need to improve?
- The question is open so will be easier and quicker to analyse.
- The question is clear and unbiased, avoiding leading students to a particular answer.

Number the **two** correct statements in the correct order (**two** statements are incorrect).

- This is because it shows frequencies and allows for visual comparisons.
- A tally chart can be used to display the data.
- A bar chart can be used to display the data.
- This is because it can be used also to collect the data.

4 Ryder is studying blue whales and needs to know the average length of a blue whale.  
Ryder uses the internet to find that the average length of a blue whale is 24 metres.

(a) Explain why the statistic collected by Ryder is an example of secondary data.

(1 mark)

Select **one** box.

- Ryder found the data themselves.
- It was collected by another researcher.
- It is a length, not an area.
- It is not as important as Ryder's other data.

(b) State one advantage and one disadvantage of using secondary data.

(2 marks)

**Advantage**

Select **one** box.

- The data will be reliable.
- It is always more accurate.
- It is cheaper to collect.
- The data collection will be more rigorous.

**Disadvantage**

Select **one** box.

- It is time-consuming.
- Ryder cannot control the quality of the data.
- It costs a lot of money to collect data in this way.
- It is never as good.

5 Sophia is an NHS trust manager investigating the working hours of junior doctors. She takes a simple random sample of 12 doctors in a hospital and asks them to report the number of hours they worked last week.

The working hours of the 8 people are listed:

36	41	39	70
40	38	42	38

Sophia believes that one of the values is an outlier.

(a) Describe the meaning of the term 'simple random sample'.

(1 mark)

Select **one** box.

- A sampling method where every individual in the population has an equal chance of being selected.
- A sampling technique that ensures only the most relevant individuals are chosen.
- A method that divides the population into groups before selecting individuals.
- A method where individuals are selected based on convenience and availability.

(b) Work out the mean.

(2 marks)

To find the mean add all the numbers together and then divide by the total amount of numbers.

\_\_\_\_\_ hours

(c) Work out the range.

(1 mark)

To find the range subtract the smallest number from the largest number.

\_\_\_\_\_ hours

(d) Write down the value that is most likely to be an outlier and explain why you think this value is an outlier.

(2 marks)

The outlier is \_\_\_\_\_ hours

Select **one** box.

- This value appears only once in the dataset.
- This value is close to the mean.
- This value is significantly higher than the other numbers.
- This value doesn't match the median.

(e) Sophia removes the outlier.

State whether the mean of the remaining seven values is greater than, is equal to or is less than the mean of all eight salaries.

Give a reason for your answer.

(2 marks)

Number the **two** correct statements in the correct order (**three** statements are incorrect).

- The mean is less.
- This is because the value that Sophia removed is higher than all the other values.
- This is because the sum of values will now be divided by seven.
- The mean is the same.
- The mean is greater.

- (f) After calculating the mean of the seven values without the outlier, Sophia uses this mean in a report to describe all the doctors in the NHS.

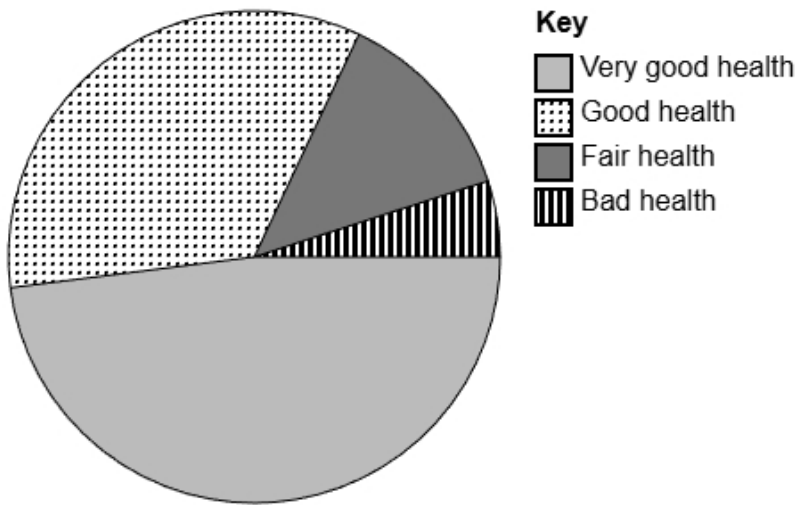
Describe two things that could affect the reliability of her conclusions.

(2 marks)

Select *two* boxes.

- The mean will be the same with and without the outlier.
- Removing the outlier may ignore the true variation in working hours.
- Small sample size.
- The number of hours may not truly reflect the hard work of the doctors.
- One value cannot represent many.

6 The accurately drawn pie chart shows information about how people in the UK rated their health in 2021.



(a) Explain how you can tell that most people viewed themselves as 'Very good health' in the UK in 2021 using the pie chart.

(1 mark)

Select **one** box.

- 'Very good health' is at the bottom of the pie chart.
- 'Very good health' is the first value in the key.
- 'Very good health' is the most positive response.
- 'Very good health' has the largest sector.

(b) The population in the UK in 2021 was estimated to be 67 million.

Calculate an estimate for the number of people in the UK in 2021 who rate themselves as having 'Good health'.

Round your answer to the nearest million.

(2 marks)

Start by measuring the angle with a protractor.

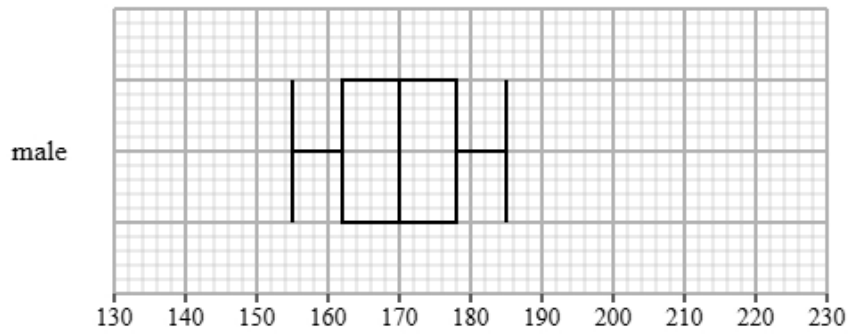
Next find the amount. Remember that there are 360 degrees in a circle.

\_\_\_\_\_ million

7 Sophie recorded the heights of male and female students in a school.

Both groups were measured using the same method.

The box plot shows information about the heights for the male students.



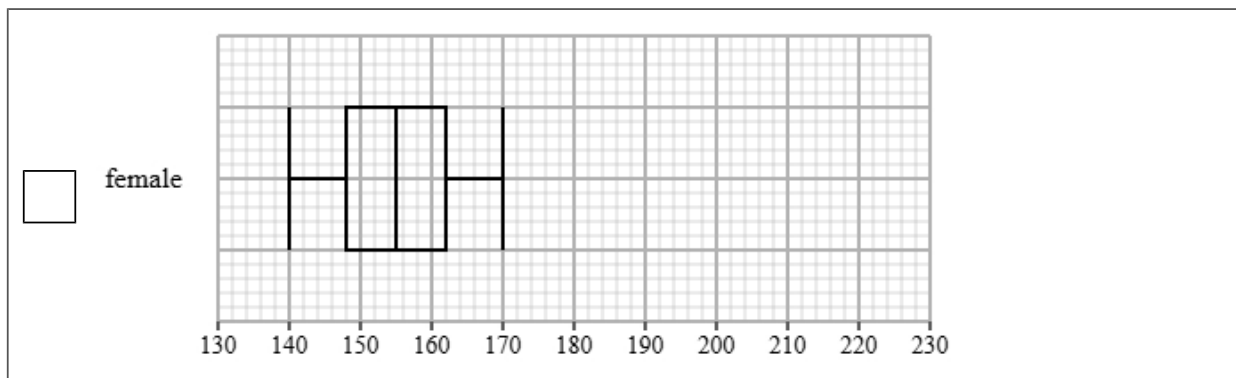
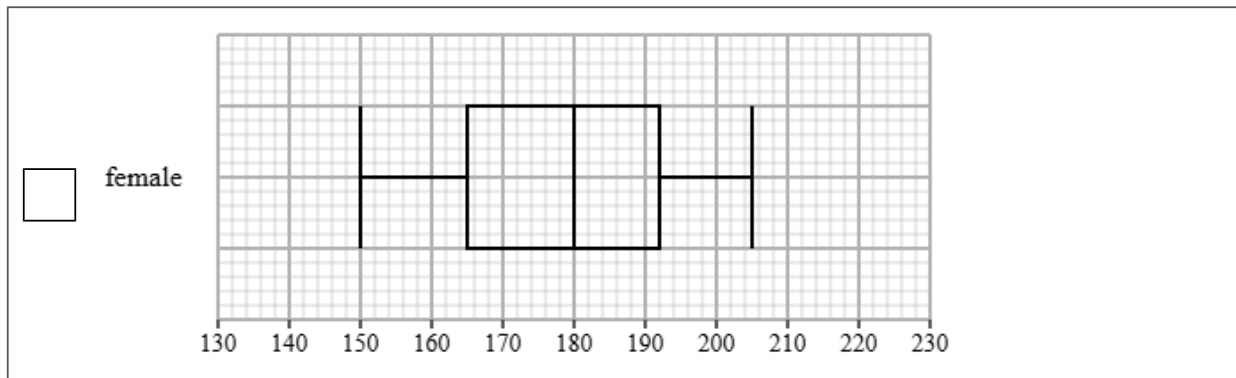
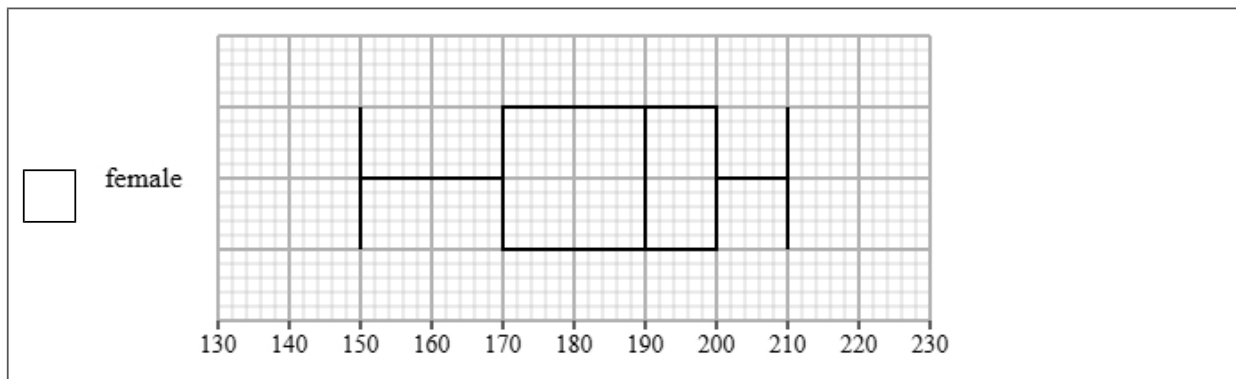
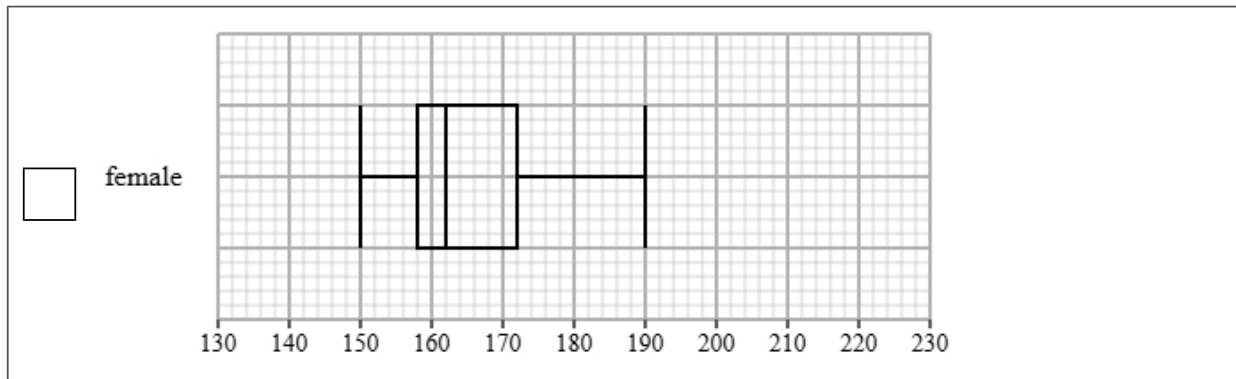
The table gives information about the heights for the female students.

Least tall	Lower quartile	Median	Upper quartile	Most tall
150	158	162	172	190

(a) Draw a box plot for the heights for the female students.

(2 marks)

Select the correct answer.



(b) Compare the two distributions of heights.

Give three comparisons and interpret one of these comparisons.

(4 marks)

Select **one** box.

- The median is bigger.
- The median heights for male and female students are equal.
- The median heights for male students is greater than female students.
- The median heights for male students is lower than female students.

Select **one** box.

- The IQR is bigger.
- The IQR for the heights of the male and female students are equal.
- The IQR for the heights of the male students is greater than female students.
- The IQR for the heights of the male students is lower than female students.

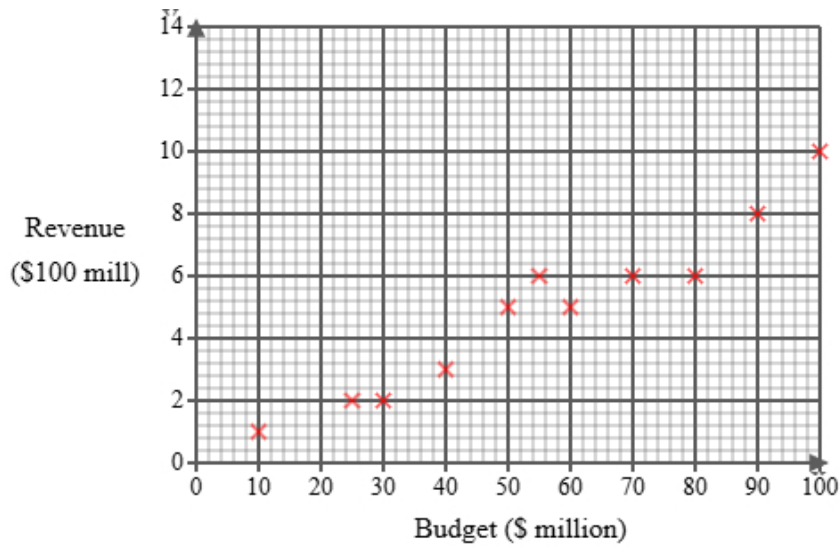
Select **one** box.

- The skews for the heights of the male and female students are both positive.
- The skew for the heights of the male students is symmetrical and the skew for the female students is negative.
- The skew for the heights of the male students is symmetrical and the skew for the female students is positive.
- The skews for the heights of the male and female students are both symmetrical.

Select **one** box.

- The heights for the male students are less spread out than the female students.
- The male students are more skewed than female students.
- The male students are on average taller than the female students.
- The male students are on average shorter than the female students.

- 8 Liam gathered data on 11 movies, recording their production budget (in millions of dollars) and total box office revenue (in hundreds of millions of dollars). He represented his findings in the scatter diagram below.



- (a) One of the 11 movies has a budget of \$55 million.

For this movie, write down its revenue.

(1 mark)

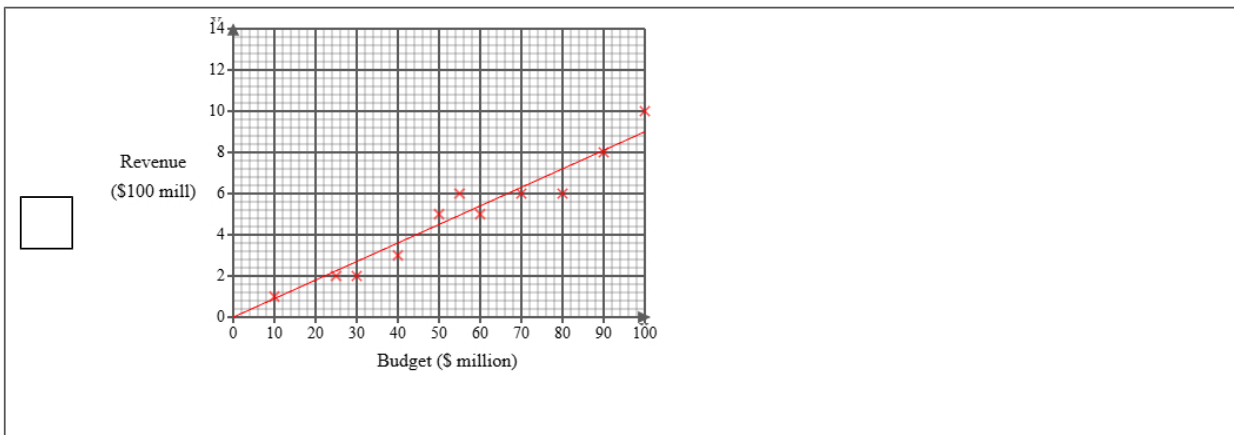
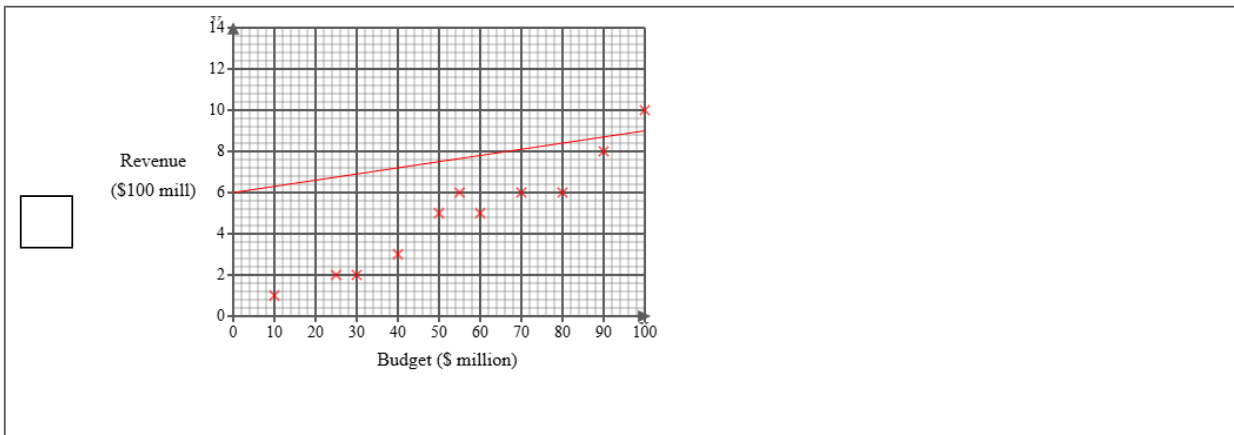
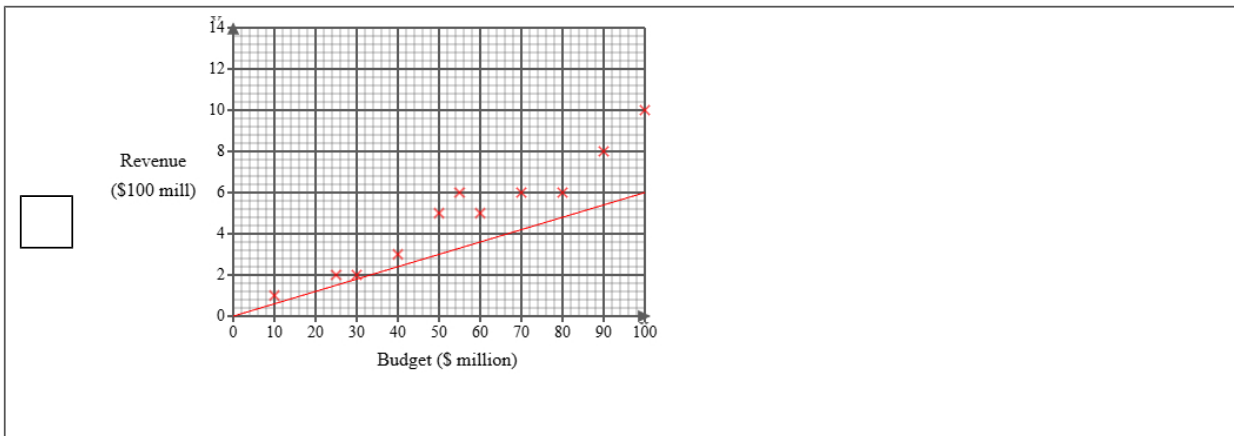
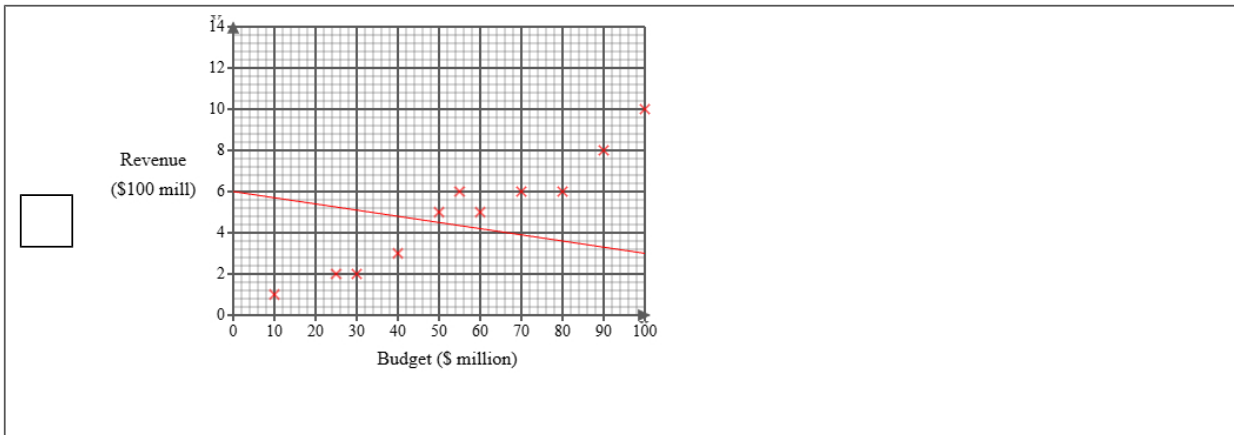
Find the cross on the scatter graph that is at \$55 million on the x-axis (the bottom axis), then read off the value from the y-axis (the side axis).

\$ \_\_\_\_\_

**(b)** Draw a line of best fit on the scatter diagram.

(1 mark)

Select the correct answer.



(c) Describe and interpret the type of correlation shown by the scatter diagram.

(3 marks)

Number the **two** correct statements in the correct order (**three** statements are incorrect).

- weak
- There is no correlation but it is
- strong
- The correlation is negative and
- The correlation is positive and

Select **one** box.

- As the budget increases the revenue increases.
- A big budget movie will have a low revenue.
- A big budget movie will have a high revenue.
- As the budget increases the revenue decreases.

(d) A new movie will be releasing soon with a budget of \$250 million.

Liam is planning on using the line of best fit on the scatter diagram to predict the revenue of the movie.

Explain whether or not it is appropriate to use the line of best fit for this prediction.

(2 marks)

Number the **two** correct statements in the correct order (**two** statements are incorrect).

- because this would be interpolation.
- This is not appropriate
- because this would be extrapolation.
- This is appropriate

9 Sophia investigates the heights of 210 students in a school.

The heights range from 129 cm to 203 cm.

Sophia considers using one of the two possible grouped frequency tables for the results, Table A or Table B, shown below.

**Table A**

Height (h cm)	Frequency
$90 < h \leq 120$	0
$120 < h \leq 150$	37
$150 < h \leq 180$	147
$180 < h \leq 210$	26
$210 < h \leq 240$	0

**Table B**

Height (h cm)	Frequency
$120 < h \leq 140$	13
$140 < h \leq 160$	62
$160 < h \leq 180$	109
$180 < h \leq 200$	24
$200 < h \leq 220$	2

(a) Give **two** advantages of using grouped data rather than raw data.

(2 marks)

Select **two** boxes.

- Grouped data makes it easier to process large amounts of data.
- Grouped data is quicker.
- Grouped data makes it easier to identify outliers.
- Grouped data can make it easier to compare different groups.
- Grouped data keeps the precision in the data.

(b) Give **one** disadvantage of using grouped data rather than raw data.

(1 mark)

Select **one** box.

- Grouped data cannot be used for statistical tests.
- Grouped data will take longer to process large amounts of data.
- Grouped data will lose the detail in the data.
- Grouped data may introduce too much complexity when analysing trends.

(c) Sophia feels that Table B gives more detail than Table A about the results.

Assess the appropriateness of Sophia's claim.

(2 marks)

Select the **three** correct statements (**three** statements are incorrect).

- Sophia's claim is justified.
- In Table A, there are no students in two of the groups, so there is a much better spread of data.
- in Table B, each group has a smaller class width, showing more detail.
- in Table B, each group has a smaller class width, so detail is lost.
- In Table A, there are no students in two of the groups, so these groups are not needed.
- Sophia's claim is not justified.

(d) Sophia wants to work out the average height of the 200 students in the school.

She decides to use Table B.

Calculate the average height of the 200 students in the school, giving your answer to 1 decimal place.

(3 marks)

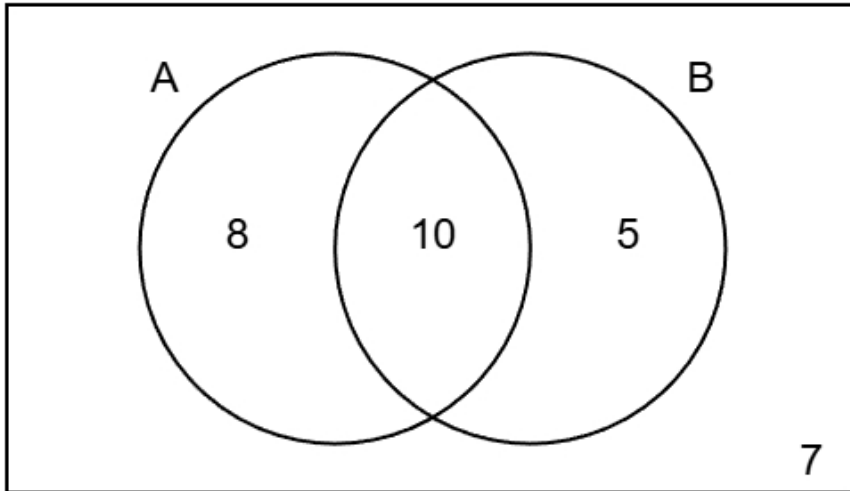
Add midpoint and  $fh$  columns onto the table.

Then find the sums of the  $f$  and  $fh$ .

Find the mean by  $\frac{\sum fh}{\sum f}$

\_\_\_\_\_ cm

- 10 The Venn diagram shows information about 30 books published in 2020.  
A is the event that the book was written by a British author.  
B is the event that the book was a bestseller (sold more than 100,000 copies).  
The numbers in the Venn diagram indicate the number of books.



- (a) In the Venn diagram, explain what the number 10 means.

(1 mark)

Select **one** box.

- The number of books written by a British author **or** were a bestseller
- The number of books written by a British author **and** were a bestseller
- The number of books written by a British author but are **not** a bestseller
- The number of books that were **not** written by a British author **or** were a bestseller

(b) Explain whether or not A and B are independent events by finding  $P(B)$  and  $P(B|A)$ .

(5 marks)

Use these formulae to find  $P(B)$  and  $P(B|A)$ :

$$P(B) = \frac{\text{amount in B}}{\text{total amount}}$$

$$P(B | A) = \frac{P(A \text{ and } B)}{P(A)}$$

---

---

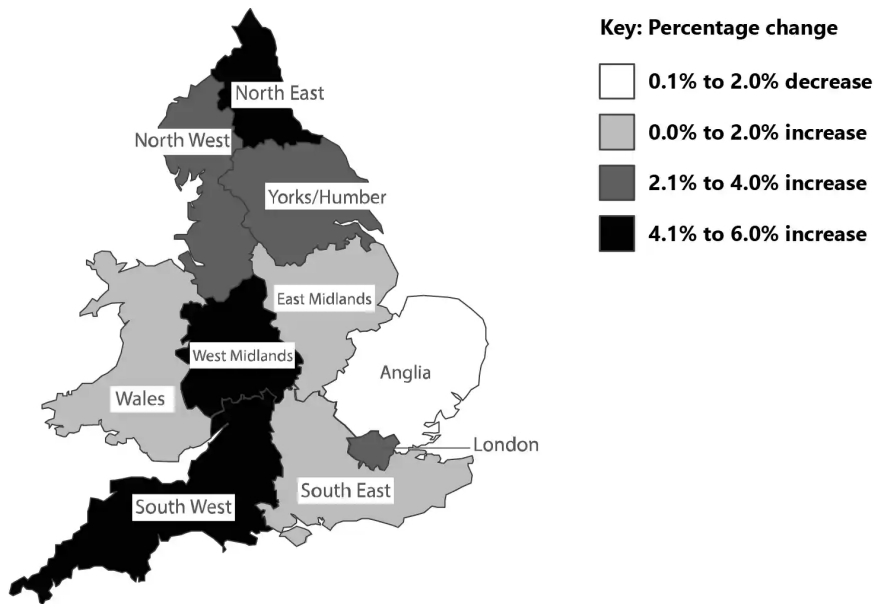
---

---

Number the **two** correct statements in the correct order (**two** statements are incorrect).

- so they are not independent
- $P(B)$  and  $P(B | A)$  are not equal
- so they are independent
- $P(B) \times P(B | A) = 1$

- 11 The map below shows the percentage change in domestic tourist visits across different regions of England and Wales between 2010 and 2011.



- (a) Write down the percentage change in domestic tourist visits between 2010 and 2011 in Wales.

(1 mark)

Select **one** box.

- 0.0% to 2.0% increase
- 4.1% to 6.0% increase
- 0.1% to 2.0% decrease
- 2.1% to 4.0% increase

- (b) There are 10 regions shown.

Find the number of regions that the domestic tourist visits **increased**.

(1 mark)

Make sure to check the key carefully.

Decreases are shown as white, whilst increases are shaded.

- (c) Bryant states that domestic tourism in England and Wales increased overall between 2010 and 2011.

Explain why this may not be the case.

(d) State the statistical name for the type of map shown.

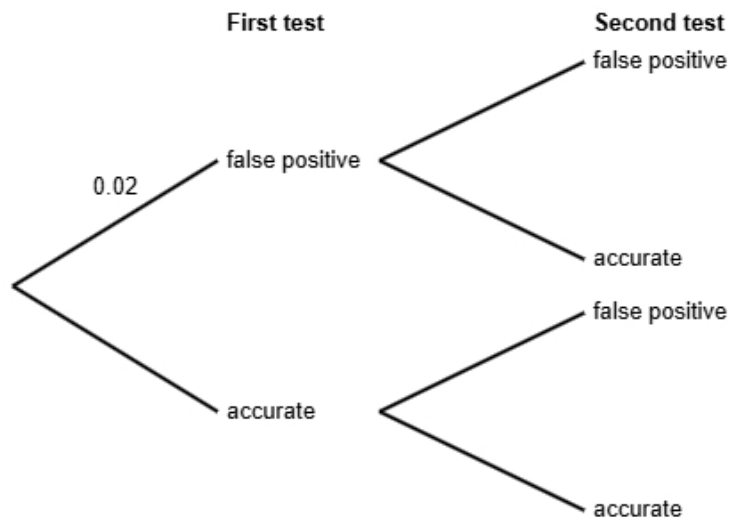
(1 mark)

The correct name starts with a 'C' and ends in 'pleth'.

C \_ \_ \_ \_ P L E T H

---

- 12** It is estimated that 2% of a certain type of medical test gives a false positive result.  
The remaining tests provide accurate results.  
David took two such tests.  
He does not know if each test result is false positive or accurate.



- (a) Complete the probability tree diagram.

(2 marks)

The branches for each stage must add up to 1.  
Each test is independent so will have the same probabilities.

(b) Find the probability that both of David's test results are accurate.

(2 marks)

You will need to find  $P(\text{accurate})$  AND  $P(\text{accurate})$ .

Remember, AND means  $\times$  in probability.

(c) David states that the probability that exactly one of the tests is false positive is less than 4%

Find out whether or not David is correct.

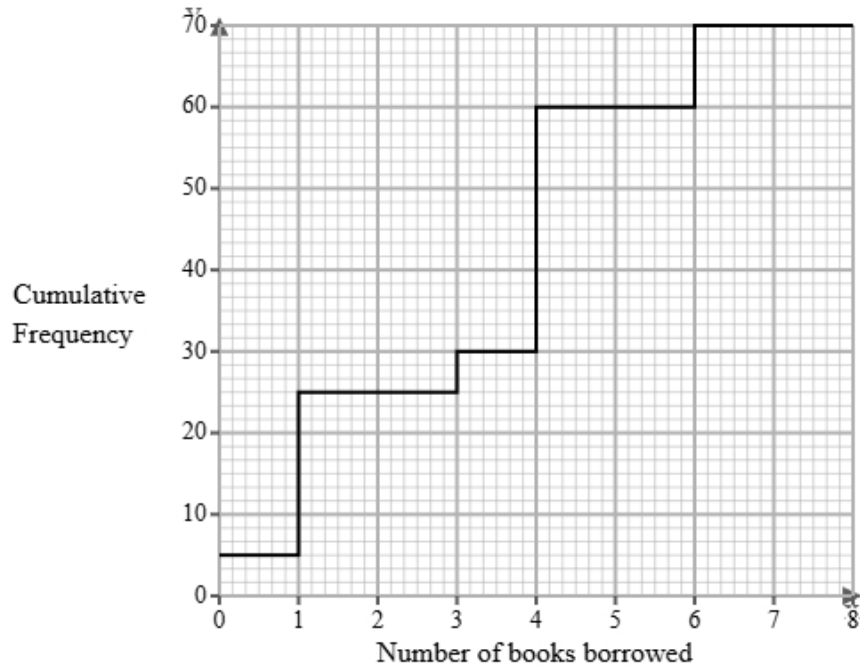
(3 marks)

Find the probability of exactly one of the tests is false positive (there are two outcomes on the tree diagram).

Select **one** box.

- The probability that exactly one of the tests is false positive is more than 4%, so David is not correct.
- The probability that exactly one of the tests is false positive is less than 4%, so David is not correct.
- The probability that exactly one of the tests is false positive is more than 4%, so David is correct.
- The probability that exactly one of the tests is false positive is less than 4%, so David is correct.

- 13 The cumulative frequency step polygon shows information about the number of books borrowed from a local library over 70 days.



- (a) Give a reason why a cumulative frequency step polygon has been used to display this data.

(1 mark)

Select **one** box.

- Because the number of books borrowed from a local library is continuous.
- Because the number of books borrowed from a local library is discrete.
- Because the number of books borrowed from a local library is qualitative.
- Because the number of books borrowed from a local library is quantitative.

- (b) Find the mode of the number of books borrowed.

(1 mark)

The mode is the number that came up the most (the highest frequency).

Look at the cumulative frequency step polygon and see where it 'jumps up' the most.

- (c) Find the number of days where there were:
- i) exactly 2 books borrowed.
  - ii) more than 2 books borrowed.

(3 marks)

The frequency is shown by how much the graph 'goes up' at each point.  
Remember, the overall frequency is 70.

i) Exactly 2 books borrowed: \_\_\_\_\_

ii) More than 2 books borrowed: \_\_\_\_\_

- (d) In 30 days fewer than  $x$  books were borrowed.  
Find the value of  $x$

(1 mark)

Draw a line across from 30 on the graph and see where all the 'jumps up' to this line are under.

\_\_\_\_\_

- (e) Laura believes the interquartile range of the number of books borrowed is 8.  
Explain why the interquartile range for this data cannot be 8.

(1 mark)

Select **one** box.

- The range is 6, so the IQR must be more than 6.
- The range is 6, so the IQR must be less than 6.
- The range is 7, so the IQR must be less than 7.
- The range is 7, so the IQR must be more than 7.