

Statistics GCSE

Paper 1

Edexcel Foundation - 2025

Notier Tier

Variant 4




1ST0/1F

# Answers

Visit our website for tutorials on each question.

[www.statsgcse.com](http://www.statsgcse.com)

- 1 A clothing store tracks how many pairs of jeans are sold each day. This information helps with restocking and running special sales based on demand.

Monday	
Tuesday	
Wednesday	
Thursday	
Friday	

**Key:**



represents 8 pairs of jeans

- (a) Find the amount of pairs of jeans sold on Tuesday.

(1 mark)

32

- (b) Find the amount of pairs of jeans sold on Wednesday.


(1 mark)


46


- (c) On Thursday, the number of pairs of jeans sold was 16.  
Show this information on the pictogram.


(1 mark)

Select the correct answer.

Thursday 

Thursday 

Thursday 

Thursday 

- (d) Halle suggests redrawing the pictogram using a key with a whole-square representing 5 pairs of jeans.  
Explain why this key would **not** be suitable.

(1 mark)

Select **one** box.

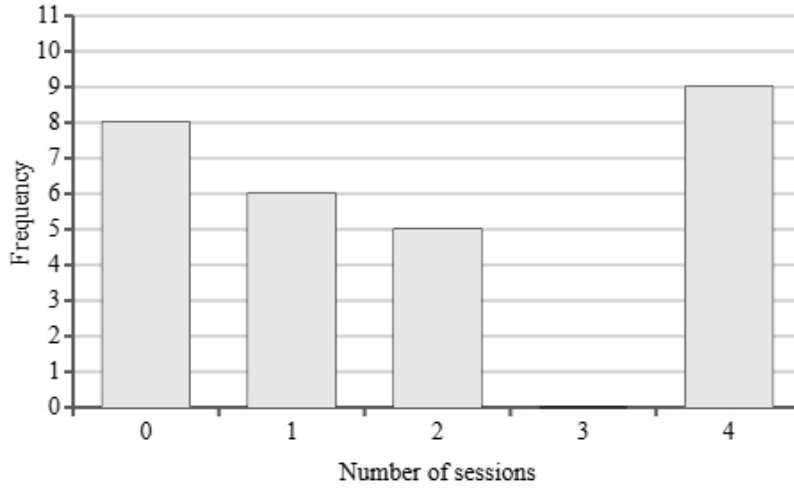
- Tuesday shows 32 loaves of bread. This would be very difficult to show because 32 has a remainder 2 when divided by 5.
- This would be much better because you can fit more squares on.
- The key must always be an even number.
- If Halle uses 5 pairs of jeans for the key this will lead to less sales being shown.

2 Liam has conducted a survey on the number of times people exercise per week.

He asked a total of 30 people.

Each person reported between 0 and 4 exercise sessions per week.

The incomplete bar chart shows the number of people who exercise 0, 1, 2, and 4 times per week.

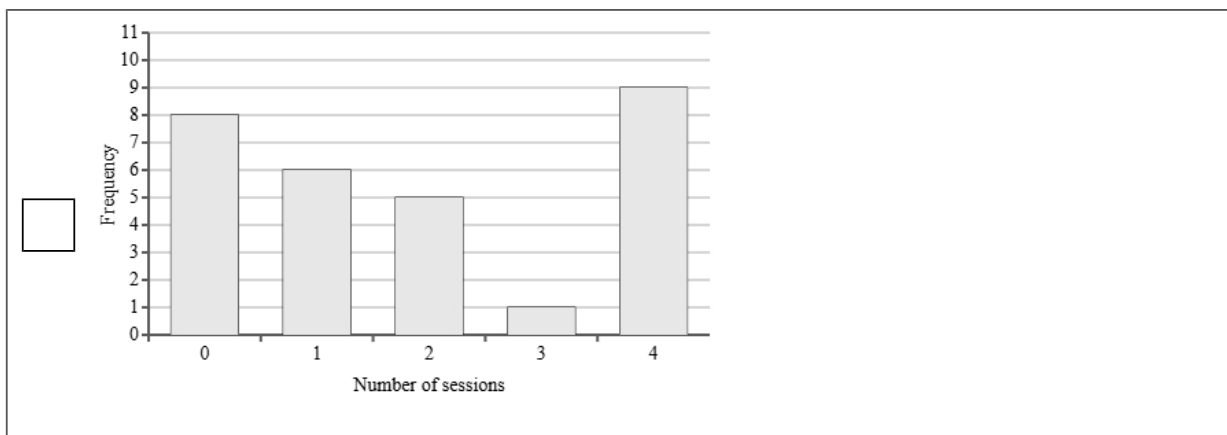
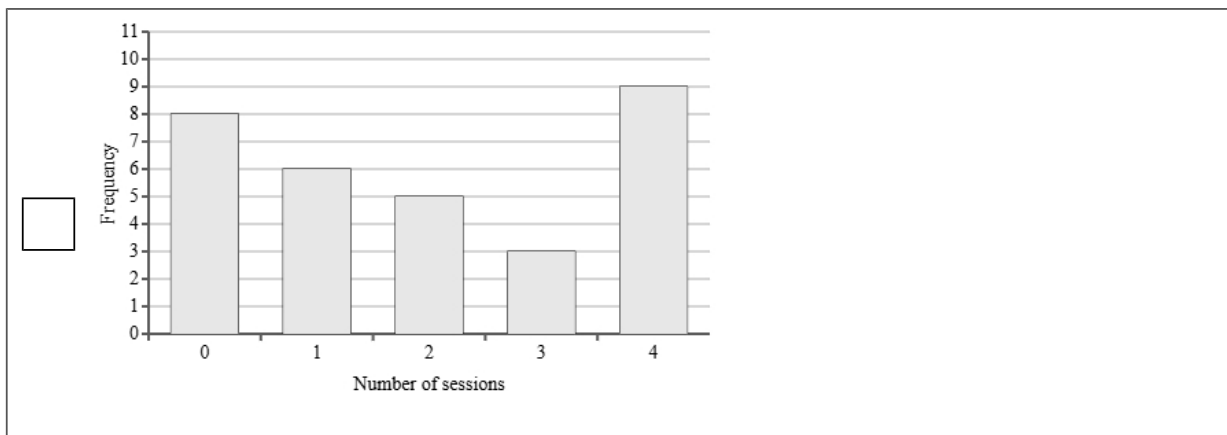
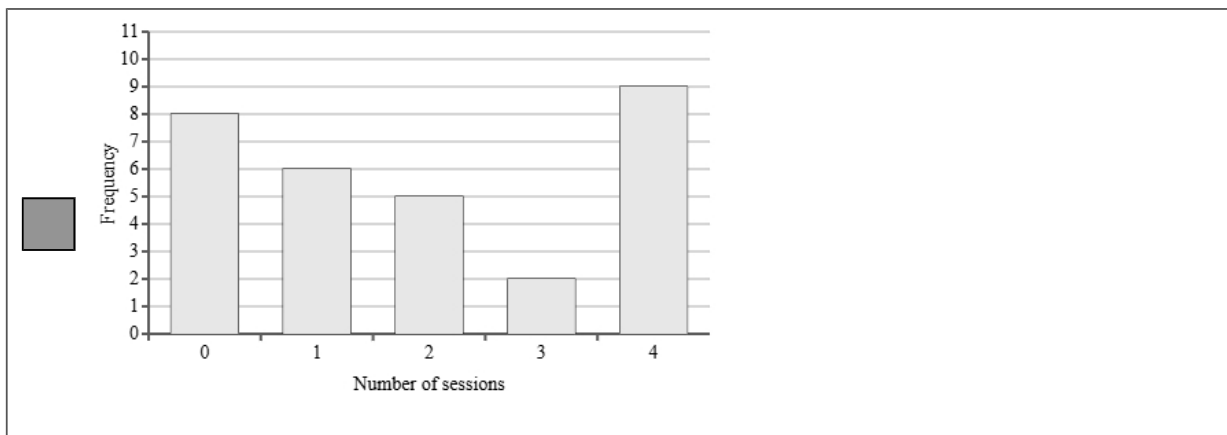
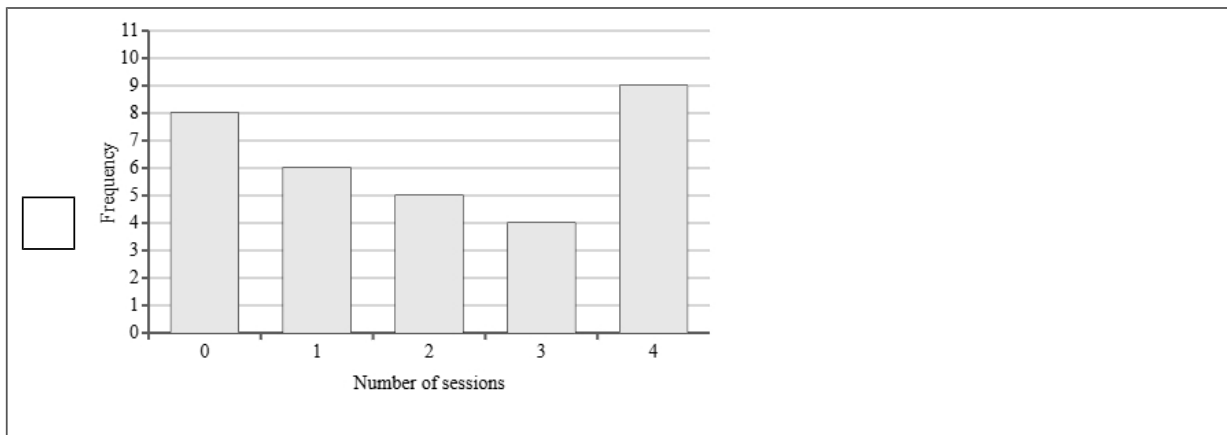


(a) 2 people reported 3 exercise sessions per week.

Complete the bar chart.

(1 mark)

Select the correct answer.



(b) Find how many more people reported no exercise sessions than reported 2 exercise sessions.

(1 mark)

3

(c) All the people in the survey attend Liam's Gym.

He thinks that because most people reported 4 exercise sessions then his gym has been successful.

Explain why Liam may not be right.

(1 mark)

Select **one** box.

- More people reported one or less exercise sessions per week.
- People may have has more than 4 exercise sessions.
- The sample size was small.
- 8 people reported no exercise sessions.

3 A hospital is planning to introduce a new appointment booking system.

Sophia wants to carry out a survey to find out what all patients think about the proposed change.

Sophia thinks that she should take a sample rather than a census.

(a) Give two reasons why Sophia might think this.

(2 marks)

Select *two* boxes.

Sophia will be able to explain each question.

A sample is cheaper.

A sample is quicker.

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

A sample is more accurate.

(b) Sophia has decided to use the electoral register as a sampling frame.

Explain what a sampling frame is.

(1 mark)

Select *one* box.

The whole group.

The tally chart or table used in the survey.

A list of all the members in the sample.

A list of all the members in the population.

(c) Sophia has decided to use the electoral register as a sampling frame.

State one problem Sophia may have using the electoral register as a sampling frame.

(1 mark)

Select **one** box.

The electoral register would also include people's addresses.

Bias.

Sophia may have difficulty gaining access to the electoral register.

There will be too many names.

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

(2 marks)

Select **two** boxes.

A pilot survey will give more accurate data.

A pilot survey will include more people.

A pilot survey will check questions are inoffensive.

A pilot survey will test questions are working as intended.

A pilot survey will be cheaper than a survey.

- (e) Sophia is writing a plan for the investigation into patients' opinions on the new appointment booking system.

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

You should include:

- + a sampling method
- + a question Sophia 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).

- 2 This will ensure that the sample is representative of the population.
- Sophia should use a case study.
- This will ensure that more students are asked.
- 1 Sophia should use stratified sampling.

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

- 2 The question is clear and unbiased, avoiding leading students to a particular answer.
- A question could be:  
Why do you think the appointment booking system needs to improve?
- The question is open so will be easier and quicker to analyse.
- A question could be:
- 1 How satisfied are you with the current appointment booking system?  
 Very Satisfied  Satisfied  Neutral  Unsatisfied  Very Unsatisfied

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

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

4 Otis is studying sperm whales and needs to know the average length of a sperm whale.  
Otis uses the internet to find that the average length of a sperm whale is 13.25 metres.

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

(1 mark)

Select **one** box.

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

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

(2 marks)

**Advantage**

Select **one** box.

- The data is only available in English.
- It is convenient and easy to collect.
- The data is collected specifically for your research.
- It is useful for decision-making.

**Disadvantage**

Select **one** box.

- The data is only available in newspapers.
- The data is never useful for further studies.
- The data may not be in required form.
- Only experts are allowed to use secondary data.

5 Liam is a civil servant working for the Department for Work and Pensions (DWP). He is researching the average weekly working hours of UK employees and takes a simple random sample of 10 workers from various industries and asks them how many hours they worked last week.

The hours worked of the 10 people are listed:

59 54 56 55 90  
56 55 59 58 58

Liam believes that one of the values is an outlier.

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

(1 mark)

Select **one** box.

- Choosing 10 employees from different departments to ensure variety.
- All members of the population have the same likelihood of selection.
- Taking the first 10 employees who volunteer to participate.
- Selecting people who have worked the longest.

(b) Work out the mean.

(2 marks)

\_\_\_\_\_ 60 hours

(c) Work out the range.

(1 mark)

\_\_\_\_\_ 36 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 \_\_\_\_\_ 90 hours

Select **one** box.

This value is part of a regular pattern in the dataset.

The value is significantly higher than the rest.

This value falls between the first and third quartiles.

The value is closer to the mode.

(e) Liam removes the outlier.

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

Give a reason for your answer.

(2 marks)

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

**2** This is because the value that Liam removed is higher than all the other values.

This is because the sum of values will now be divided by nine.

The mean is the same.

**1** The mean is less.

The mean is greater.

- (f) After calculating the mean of the nine values without the outlier, Liam uses this mean in a report to describe all the employees in the UK.

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

(2 marks)

Select *two* boxes.

The mean will not measure the spread of the data.

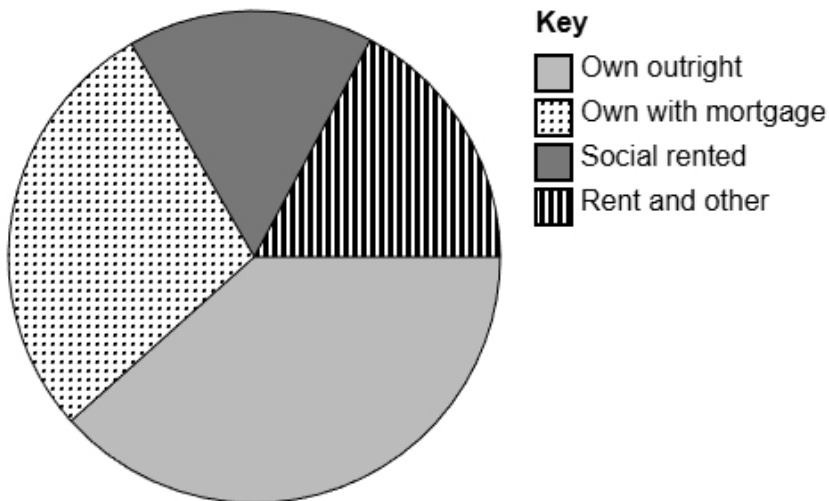
Small sample size.

The working hours may not truly reflect the hard work people do.

One value cannot represent many.

Sample might not be representative.

6 The accurately drawn pie chart shows information about the tenure types for people in Wales in 2021.



(a) Explain how you can tell that most households own their house outright in Wales in 2021 using the pie chart.

(1 mark)

Select **one** box.

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

(b) In 2021 there was an estimated 130 thousand households.

Calculate an estimate for the number of people in the UK in 2021 who's highest level of qualification was 'Own outright'.

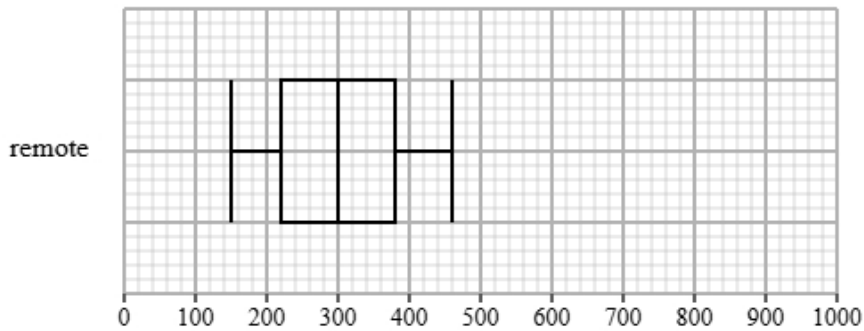
Round your answer to the nearest thousand.

(2 marks)

\_\_\_\_\_ 50 thousand

- 7 Ethan collected the steps for remote and on-site workers in an hour within their day.  
Both groups recorded their steps over the same period.

The box plot presents data on the steps for the remote workers.



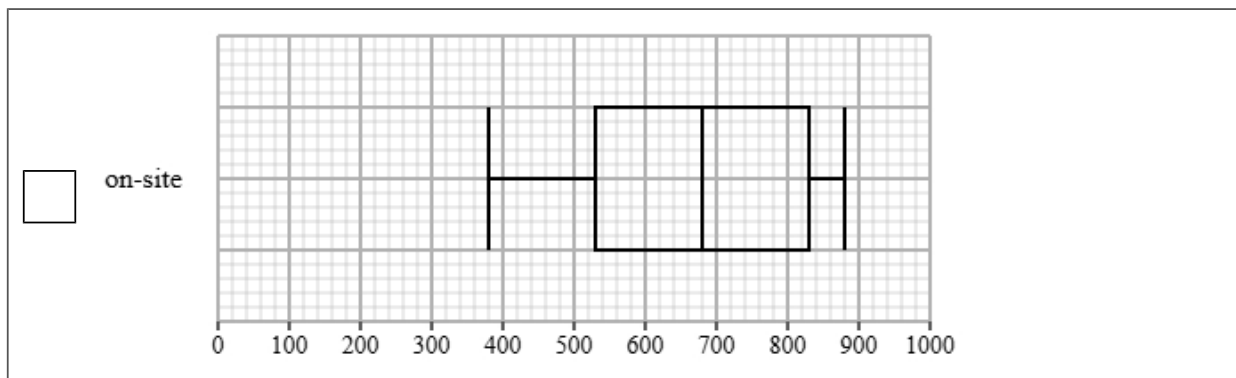
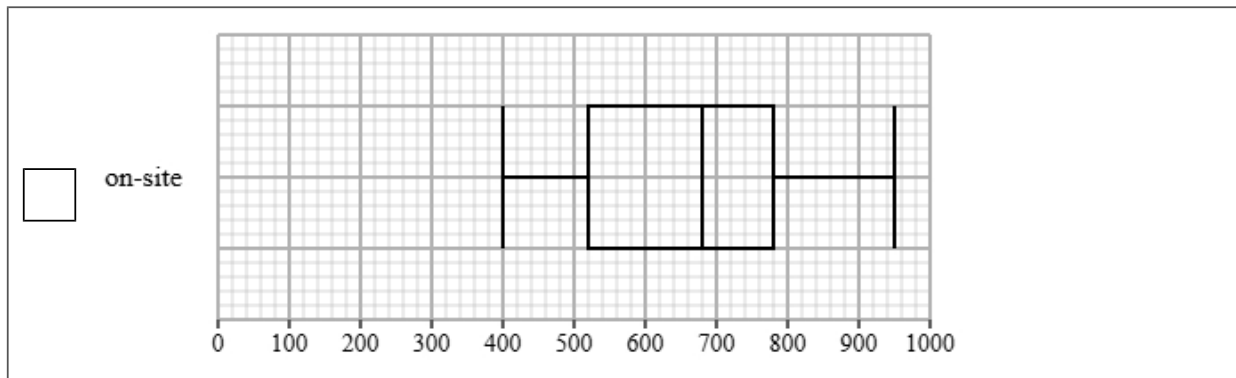
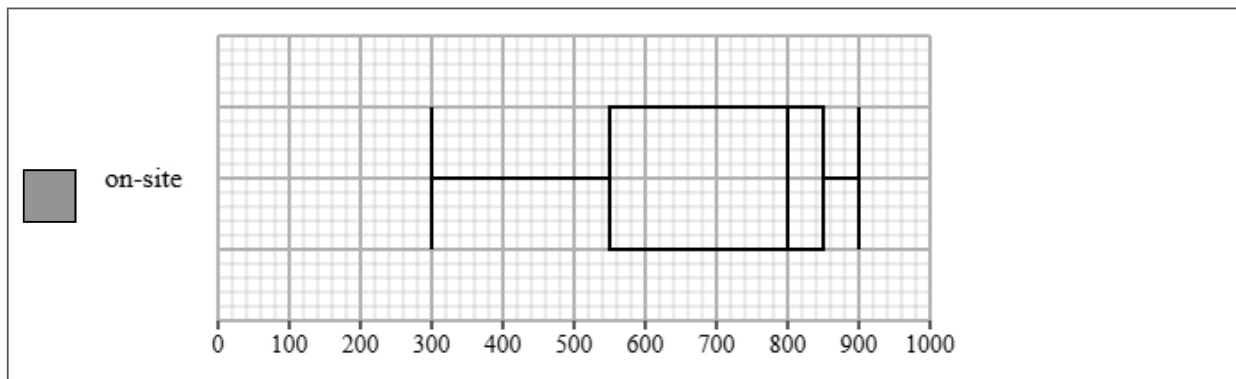
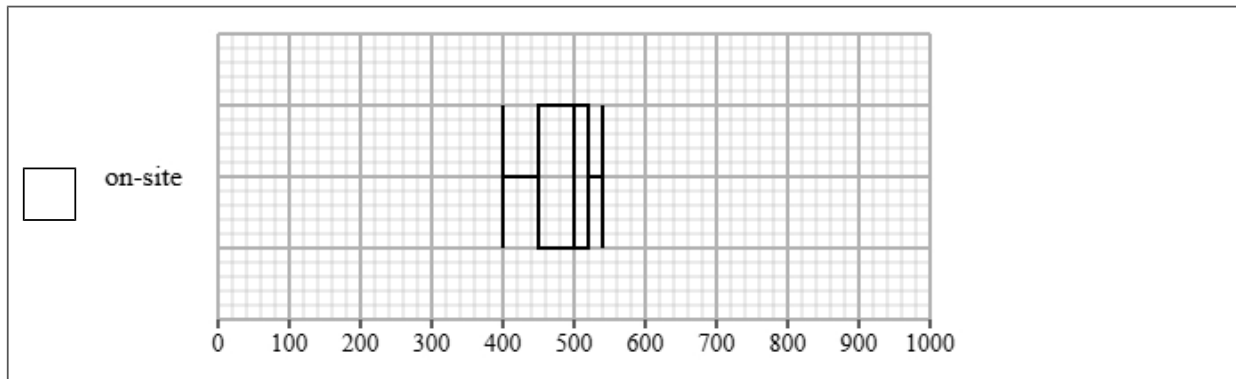
The table gives information about the steps for the on-site workers.

Least tall	Lower quartile	Median	Upper quartile	Most tall
300	550	800	850	900

- (a) Draw a box plot for the steps for the on-site workers.

(2 marks)

Select the correct answer.



(b) Compare the two distributions of steps.

Give three comparisons and interpret one of these comparisons.

(4 marks)

Select **one** box.

- The median is bigger.
- The median steps for remote workers is greater than on-site workers.
- The median steps for remote workers is lower than on-site workers.
- The median steps for remote and on-site workers are equal.

Select **one** box.

- The IQR is bigger.
- The IQR for the steps of the remote workers is greater than on-site workers.
- The IQR for the steps of the remote workers is lower than on-site workers.
- The IQR for the steps of the remote and on-site workers are equal.

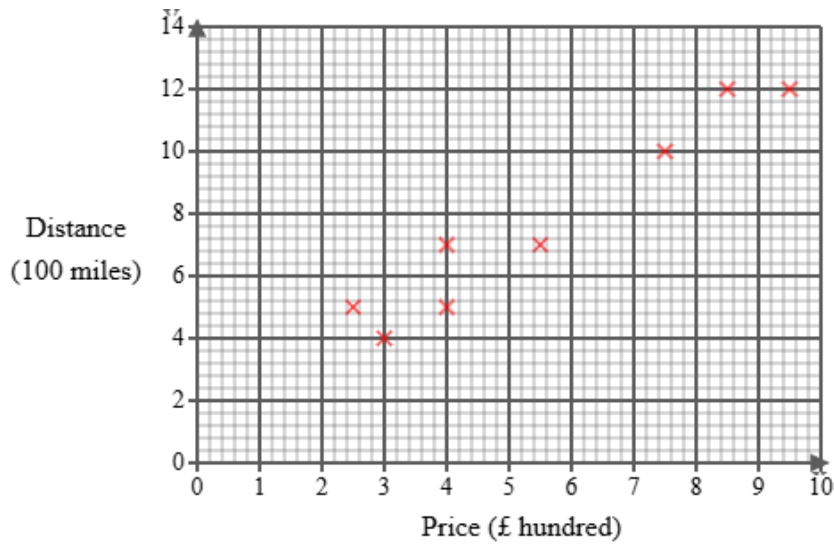
Select **one** box.

- The skews for the steps of the remote and on-site workers are both positive.
- The skew for the steps of the remote workers is symmetrical and the skew for the on-site workers is positive.
- The skew for the steps of the remote workers is symmetrical and the skew for the on-site workers is negative.
- The skews for the steps of the remote and on-site workers are both symmetrical.

Select **one** box.

- The steps for the remote workers are more spread out than the on-site workers.
- The remote workers on average walk less than the on-site workers.
- The remote workers on average walk more than the on-site workers.
- The remote workers are more skewed than on-site workers.

- 8 Jamie collected data on 11 flights, recording the distance (in hundreds of miles) and the price (in hundreds of pounds) of each flight. He represented his findings in the scatter diagram below.



- (a) One of the 11 flights has a price of £550.  
For this flight, write down the price.

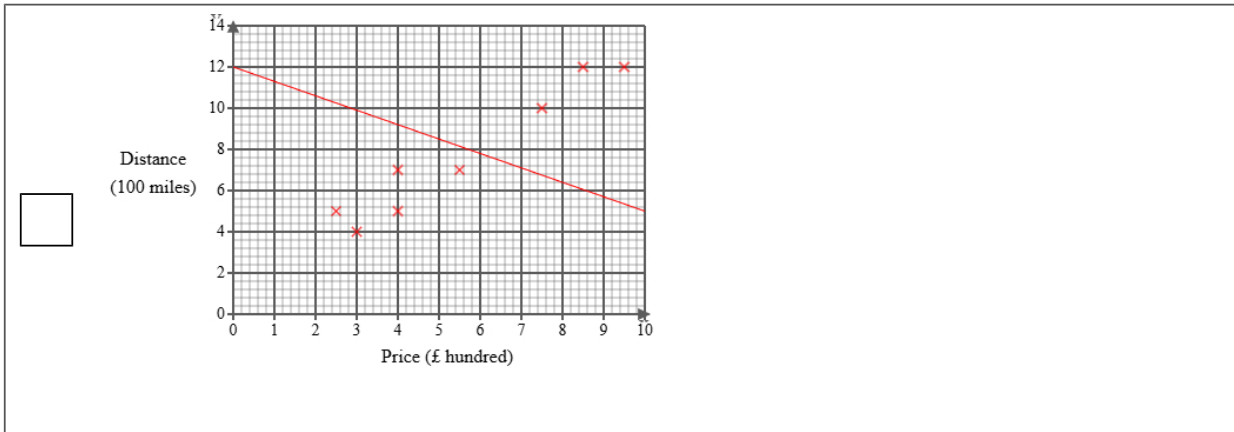
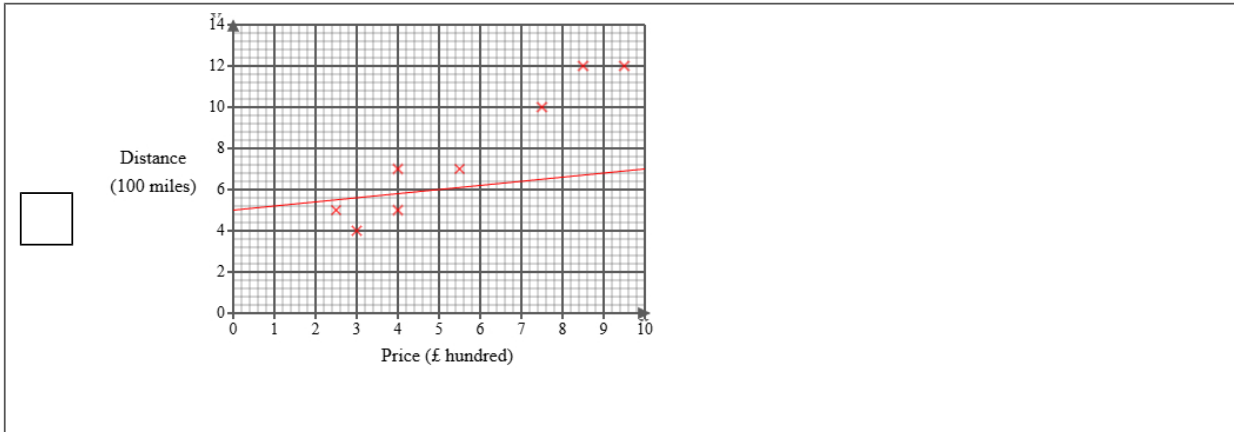
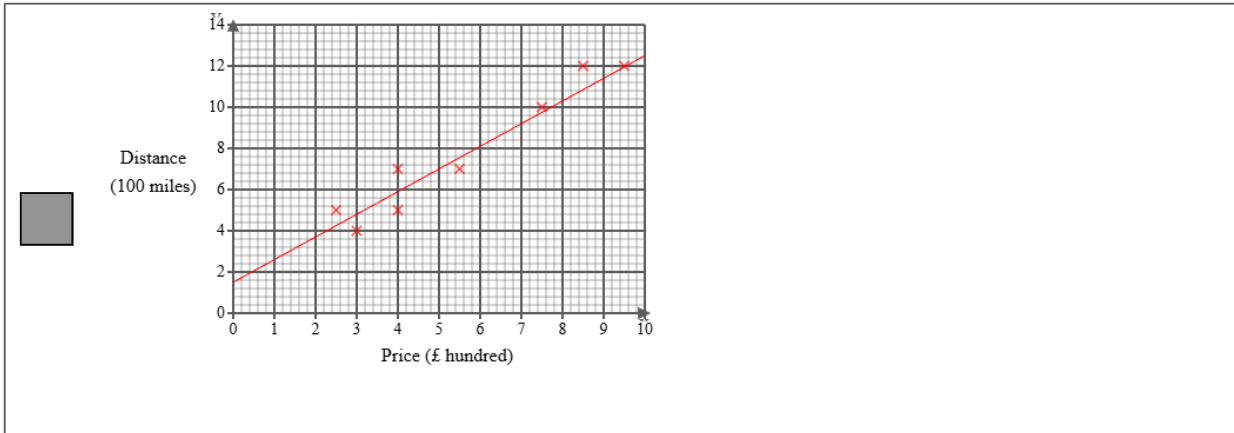
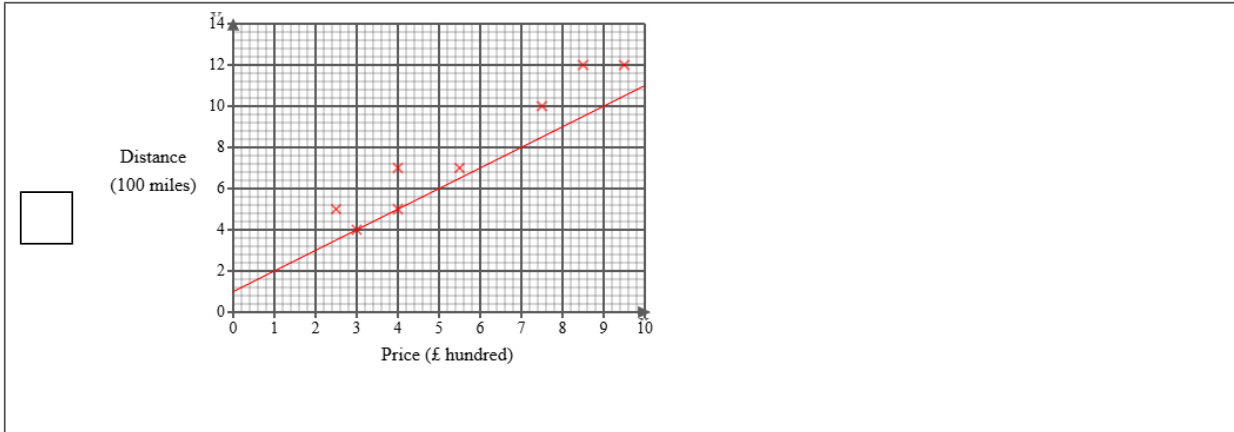
(1 mark)

\_\_\_\_\_ 700 miles

**(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).

1 The correlation is positive and

There is no correlation but it is

The correlation is negative and

weak

2 strong

Select **one** box.

As the price increases the journey distance decreases.

A journey that has a high price will have a low journey distance.

A journey that has a high price will have a high journey distance.

As the price increases the journey distance increases.

(d) An airline has announced a new route at a price of £2000.

Jamie is planning on using the line of best fit on the scatter diagram to predict the distance of the flight.

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).

This is appropriate

2 because the point is after the data and the trend may not continue.

1 This is not appropriate

because the trend will continue.

9 Noah investigates the reaction times (in milliseconds) of 120 people taking a driving test.

The times range from 203 ms to 281 ms.

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

**Table A**

<b>Time (<math>t</math> ms)</b>	<b>Frequency</b>
$170 < t \leq 200$	0
$200 < t \leq 230$	31
$230 < t \leq 260$	67
$260 < t \leq 290$	22
$290 < t \leq 310$	0

**Table B**

<b>Time (<math>t</math> ms)</b>	<b>Frequency</b>
$200 < t \leq 220$	14
$220 < t \leq 240$	40
$240 < t \leq 260$	44
$260 < t \leq 280$	20
$280 < t \leq 300$	2

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

(2 marks)

Select **two** boxes.

- Grouped data doesn't require any calculation, as it displays frequencies directly.
- Grouped data provides a more accurate reflection of trends.
- Grouped data can help to spot patterns in the data.
- Grouped data is easier to represent on graphs.
- Grouped data helps avoid using averages.

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

(1 mark)

Select **one** box.

- Grouped data is much more difficult to read.
- Grouped data cannot be drawn on a graph.
- Grouped data can only calculate estimates of statistical values.
- Grouped data cannot be compared.

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

Assess the appropriateness of Noah's claim.

(2 marks)

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

- In Table A, the data goes from 170 to 210, showing a much wider range of data.
- In Table B, some data could have been less than 200 or more than 300, but would not be shown.
- Noah's claim is justified.
- In Table B, the table starts at 200 and the lowest value is 203 ms and ends at 300 with the highest value at 281 ms.
- Noah's claim is not justified.
- In Table A, all the data is concentrated into three groups.

(d) Noah wants to work out the average reaction times of the 120 people taking a driving test.

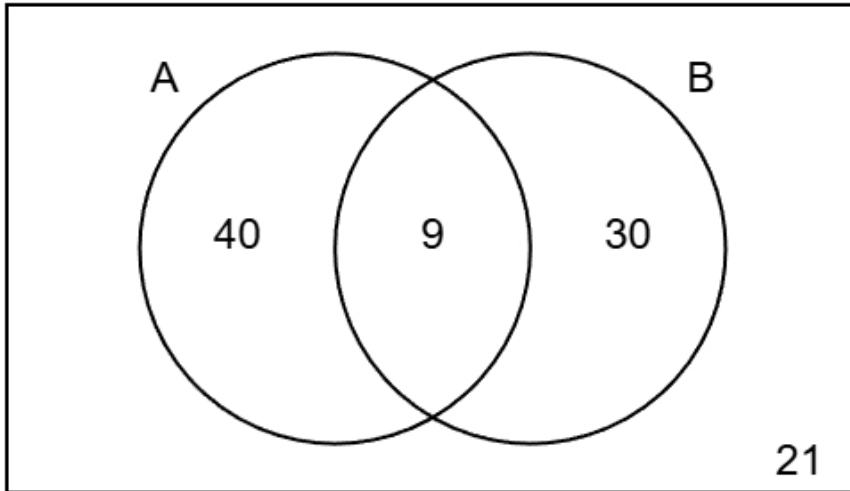
He decides to use Table B.

Calculate the average reaction times of the 120 people taking a driving test, giving your answer to 1 decimal place.

(3 marks)

242.7 ms

- 10 The Venn diagram shows information about 100 customers in shopping centre.  
A is the event that the customer bought a product in the electronics shop.  
B is the event that the customer bought a product in the clothing shop.  
The numbers in the Venn diagram indicate the number of customers.



- (a) In the Venn diagram, explain what the number 9 means.

(1 mark)

Select **one** box.

The number of customers who bought a product the electronics shop **or** a clothing shop

The number of customers who bought a product the electronics shop **and** a clothing shop

The number of customers who bought a product the electronics shop but did **not** buy a product in the clothing shop

The number of customers who did **not** buy a product in the electronics shop **or** the clothing shop

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

(5 marks)

$$P(B) = \frac{39}{100}$$

$$P(B|A) = \frac{9}{49}$$

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

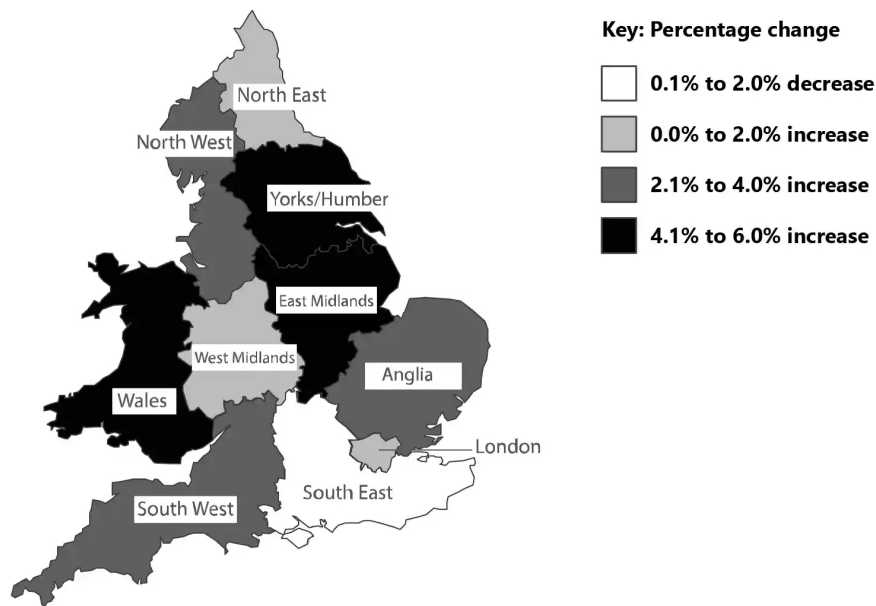
1  $P(B)$  and  $P(B | A)$  are not equal

$P(B) \times P(B | A) = 1$

so they are independent

2 so they are not independent

- 11 The map below shows the percentage change in domestic tourist visits across different regions of England and Wales between 2013 and 2014.



- (a) Write down the percentage change in domestic tourist visits between 2013 and 2014 in Wales.

(1 mark)

Select **one** box.

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

- (b) There are 10 regions shown.

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

(1 mark)

\_\_\_\_\_ 1

- (c) Polly states that domestic tourism in England and Wales increased overall between 2013 and 2014.

Explain why this may not be the case.

(1 mark)

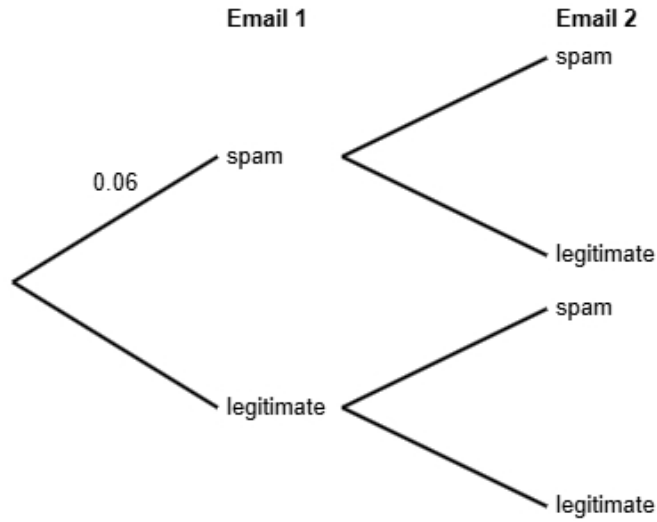
Select **one** box.

26  The map is not detailed enough. [www.statsgcse.com](http://www.statsgcse.com)

We do not know any data after 2014.

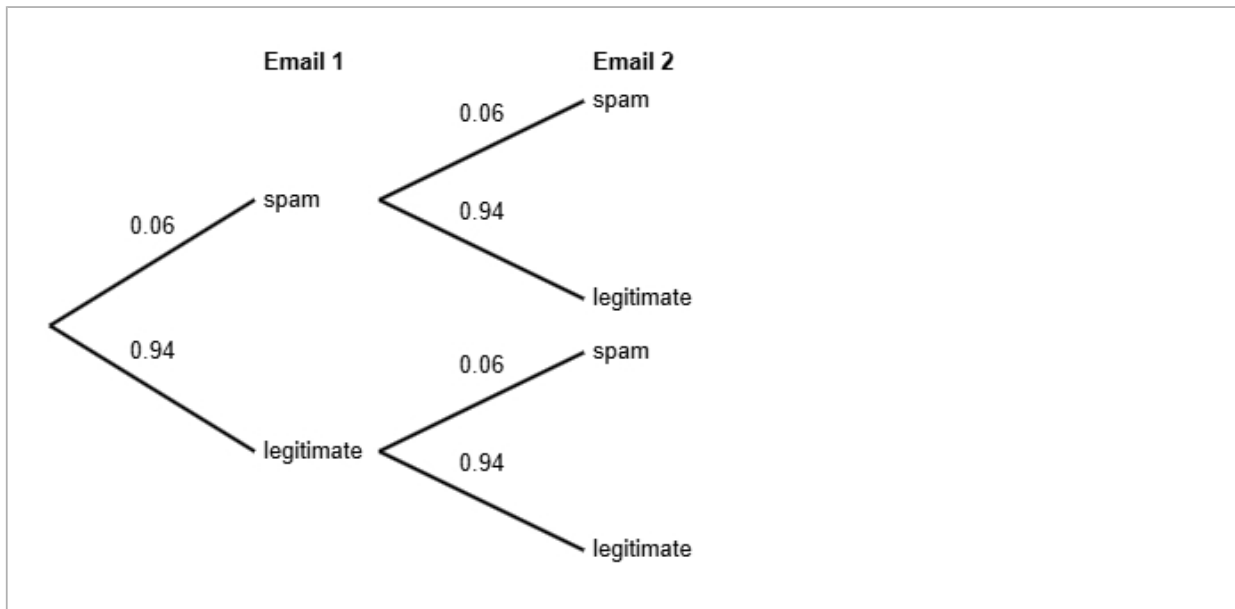
Turn Over ►

- 12** A study shows that 6% of emails received by a certain email provider are spam.  
 All other emails are legitimate.  
 Maria receives two emails in her inbox.  
 She does not know if each email is spam or legitimate.



- (a) Complete the probability tree diagram.

(2 marks)



(b) Find the probability that both of Maria's emails are legitimate.

(2 marks)

0.8836

(c) Maria states that the probability that exactly one email is spam is less than 12%

Find out whether or not Maria is correct.

(3 marks)

$$\begin{aligned} P(\text{spam AND legitimate}) &= 0.06 \times 0.94 \\ &= 0.0564 \end{aligned}$$

$$\begin{aligned} P(\text{exactly one email is spam}) &= 0.0564 \times 2 \\ &= 0.1128 \end{aligned}$$

$$0.1128 = 11.28\%$$

Select **one** box.

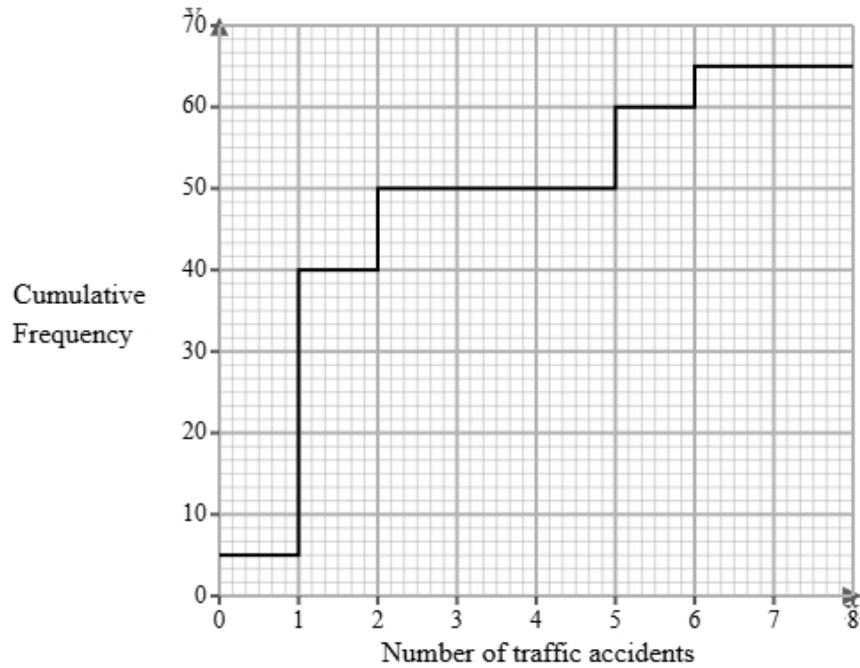
The probability that exactly one email is spam is less than 12%, so Maria is not correct.

The probability that exactly one email is spam is less than 12%, so Maria is correct.

The probability that exactly one email is spam is more than 12%, so Maria is correct.

The probability that exactly one email is spam is more than 12%, so Maria is not correct.

- 13 The cumulative frequency step polygon shows information about number of traffic accidents reported in a neighbourhood over 65 days.



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

(1 mark)

Select **one** box.

- Because number of traffic accidents reported in a neighbourhood is continuous.
- Because number of traffic accidents reported in a neighbourhood is qualitative.
- Because number of traffic accidents reported in a neighbourhood is quantitative.
- Because number of traffic accidents reported in a neighbourhood is discrete.

- (b) Find the mode of the number of traffic accidents.

(1 mark)

\_\_\_\_\_ 1

(c) Find the number of days where there were:

i) exactly 4 traffic accidents.

ii) more than 4 traffic accidents.

(3 marks)

i) Exactly 4 traffic accidents: \_\_\_\_\_ 0

ii) More than 4 traffic accidents: \_\_\_\_\_ 15

(d) In 60 days fewer than  $x$  traffic accidents were reported.

Find the value of  $x$

(1 mark)

\_\_\_\_\_ 6

(e) Rosemary believes the interquartile range of number of traffic accidents reported 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 less than 6.

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

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

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