

Statistics GCSE**Paper 1**

Edexcel Foundation - 2025

Notier Tier

Variant 5

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
- This paper and more are available on our site with questions that change subtly after each attempt.
- Good luck!

- 1 An electronics store tracks how many headphones are sold each day. This helps them forecast demand and adjust their inventory levels.

Monday	
Tuesday	
Wednesday	
Thursday	
Friday	

Key:

 represents 8 headphones

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

(1 mark)

Make sure to look at the key.

- (b) Find the amount of headphones sold on Tuesday.

(1 mark)





The circle is split up into 4 parts. The key shows that four parts represent 8 headphones.
Start by finding what one part represents.

(c) On Thursday, the number of headphones sold was 24.

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) Trent suggests redrawing the pictogram using a key with a whole-circle representing 5 headphones.

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

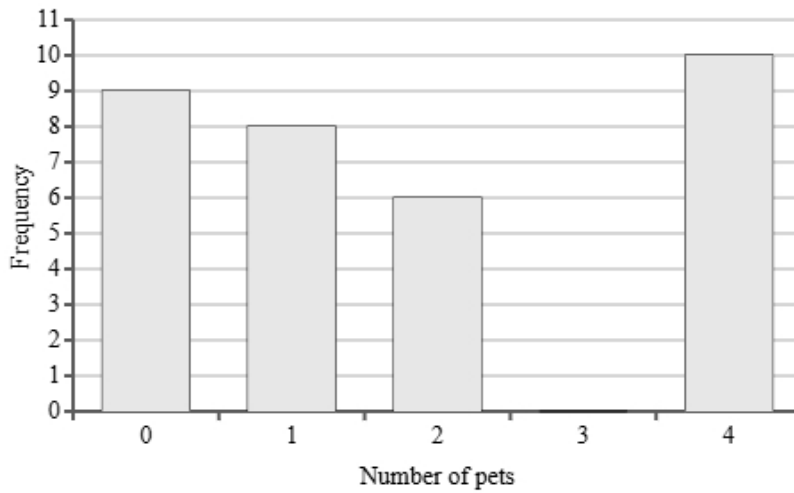
(1 mark)

Select **one** box.

- This would be much better because you can fit more circles on.
- Monday shows 16 loaves of bread. This would be very difficult to show because 16 has a remainder 1 when divided by 5.
- The key must always be an even number.
- If Trent uses 5 headphones for the key this will lead to less sales being shown.

- 2 Sophia has collected data on the number of pets owned by 35 families.
Each family owns between 0 and 4 pets.

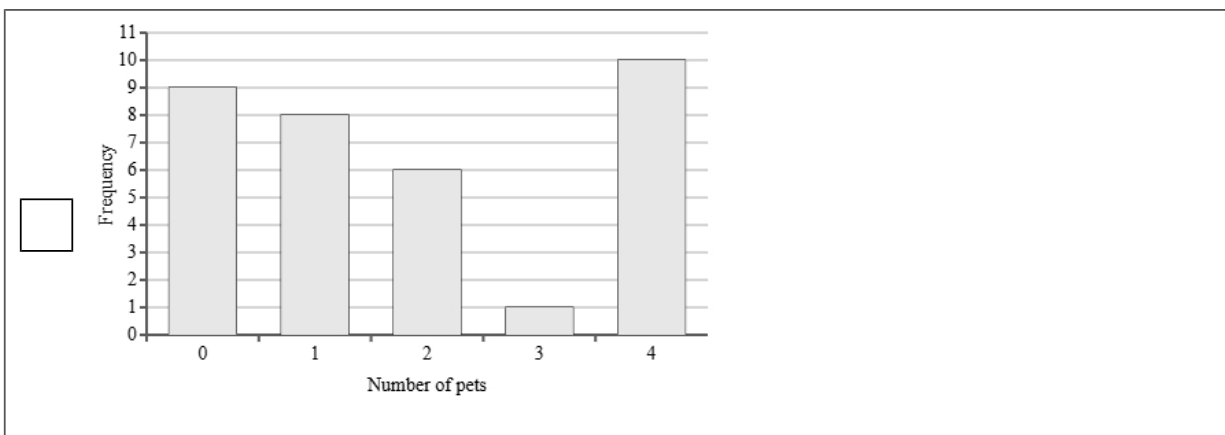
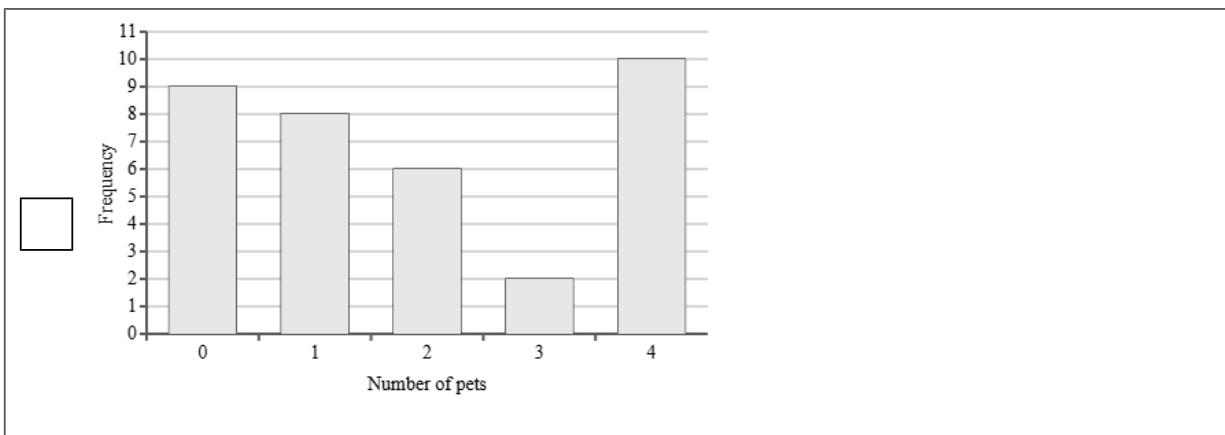
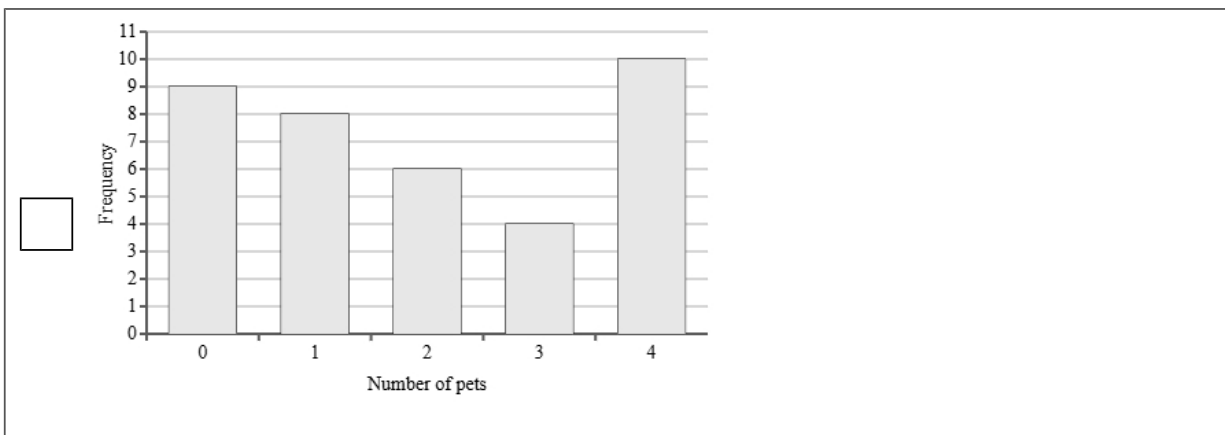
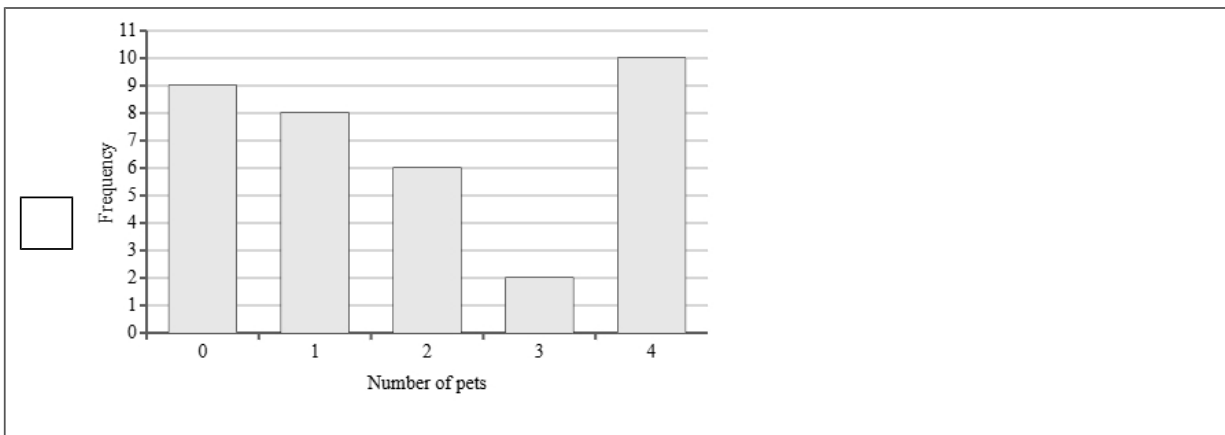
The incomplete bar chart displays the number of families with 0, 1, 2, and 4 pets.



- (a) 2 families owned 3 pets.
Complete the bar chart.

(1 mark)

Select the correct answer.



(b) Find how many more families owned no pets than owned 2 pets.

(1 mark)

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

(c) Sophia wants to open a pet shop in the area where she collected the data.

She thinks that because most people have 4 pets then a pet shop will be popular and sell lots of pet food.

Explain why Sophia may not be right.

(1 mark)

Select **one** box.

9 families have no pets.

Families may have more than 4 pets.

More families have one or less pets.

The sample size was small.

3 A theme park is considering building a new roller coaster.

Mia is carrying out a survey to see what all visitors think about the new attraction.

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

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

(2 marks)

Select *two* boxes.

A sample is cheaper.

A sample is quicker.

Mia will be able to explain each question.

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

A sample is more accurate.

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

Explain what a sampling frame is.

(1 mark)

Select *one* box.

The tally chart or table used in the survey.

A list of all the members in the population.

A list of all the members in the sample.

The whole group.

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

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

(1 mark)

Select **one** box.

- The electoral register would also include people's addresses.
- The electoral register may be out of date.
- There will be too many names.
- Unreliable.

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

(2 marks)

Select **two** boxes.

- A pilot survey will check questions are clear.
- A pilot survey will check questions are inoffensive.
- A pilot survey will give more accurate data.
- A pilot survey will include more people.
- A pilot survey will be cheaper than a survey.

(e) Mia is writing a plan for the investigation into visitors' opinions on the new roller coaster.

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

You should include:

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

- Mia should use tick-tock sampling.
- This will ensure that every visitor has an equal chance of being selected.
- This will ensure that the students asked are the most knowledgeable.
- Mia should use random sampling.

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

A question could be:

- How would you describe the plans for the new roller coaster?
 Very Poor Poor Neutral Good Very Good
- The question is clear and unbiased, avoiding leading students to a particular answer.
- The question is open so will be easier and quicker to analyse.
- A question could be:
What do you think makes the new roller coaster so amazing?

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

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

4 Conor is studying humpback whales and needs to know the average length of a humpback whale.
Conor uses the internet to find that the average length of a humpback whale is 13.5 metres.

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

(1 mark)

Select **one** box.

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

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

(2 marks)

Advantage

Select **one** box.

- It is convenient and quick to access.
- It is always in the form of numbers.
- It is never outdated.
- It is more reliable than government reports.

Disadvantage

Select **one** box.

- It is only useful for short-term studies.
- It disappears after a certain period.
- It is very difficult to control the quality of the data.
- It cannot be compared with primary data.

5 Charlotte is a housing officer for a London council. She is studying the affordability of rental housing in her borough.

Charlotte takes a simple random sample of 8 tenants from different council and private rental properties and asks them how much rent they pay each month.

The rent of the 8 people are listed:

£910	£910	£880	£2500
£900	£875	£825	£880

Charlotte believes that one of the values is an outlier.

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

(1 mark)

Select **one** box.

- To ensure only the most qualified individuals are included in the sample.
- Every person in the population has an identical probability of being included in the sample.
- To ensure that every subgroup in the population is proportionally represented.
- To make data collection easier by selecting the most accessible individuals.

(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.

£ _____

(c) Work out the range.

(1 mark)

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

£ _____

(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 £ _____

Select **one** box.

- This value is significantly higher than the rest of the data points.
- The value is exactly equal to the median.
- This value repeats several times in the data.
- The value is far below the mean.

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

(f) After calculating the mean of the seven rents without the outlier, Charlotte uses this mean in a report to describe all the tenants in London.

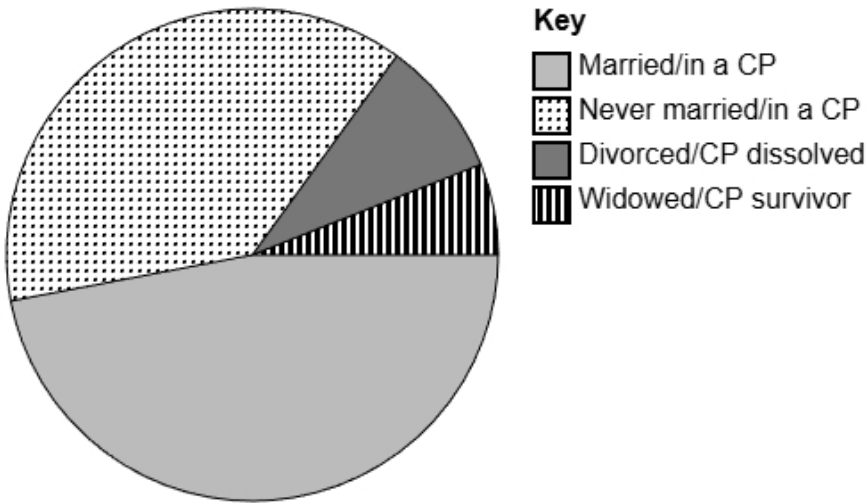
Describe two things that could affect the reliability of her conclusions.

(2 marks)

Select *two* boxes.

- The mean will not measure the range of the data.
- One value cannot represent many.
- Sensitive topic.
- Small sample size.
- The rent people pay may not truly reflect the quality of their dwellings.

- 6 The accurately drawn pie chart shows information about the proportion of adults who are married or in a civil partnership (CP) in England and Wales in 2021.



- (a) Explain how you can tell that most adults in 2021 were married or in a civil partnership using the pie chart.

(1 mark)

Select **one** box.

- 'Married/in a CP' is the first value in the key.
- 'Married/in a CP' is the most positive response.
- 'Married/in a CP' has the largest sector.
- 'Married/in a CP' is at the bottom of the pie chart.

- (b) The population in the England and Wales in 2021 was estimated to be 60 million.

Calculate an estimate for the number of adults in England and Wales in 2021 who were 'Never married/in a CP'.

Round your answer to the nearest million.

(2 marks)

Start by measuring the angle with a protractor.

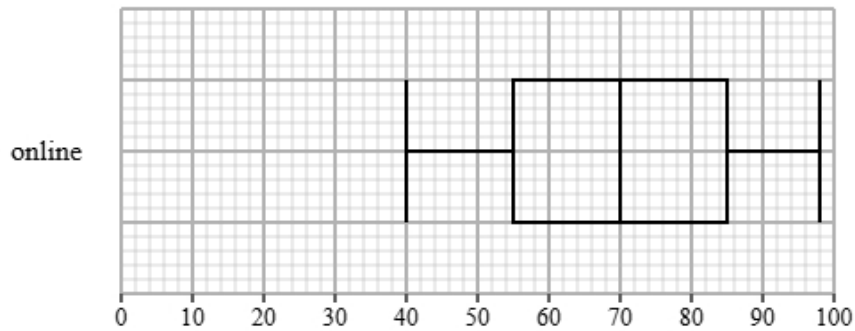
Next find the amount. Remember that there are 360 degrees in a circle.

_____ million

7 Isla collected the satisfaction scores for online and in-store customers.

Both types of shopping experiences were rated by the same number of customers.

The box plot presents data on the satisfaction scores for the online customers.



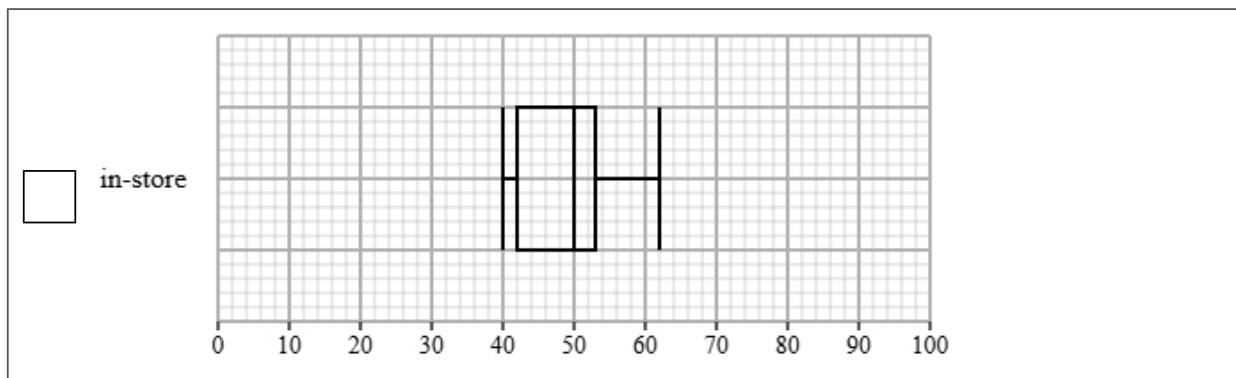
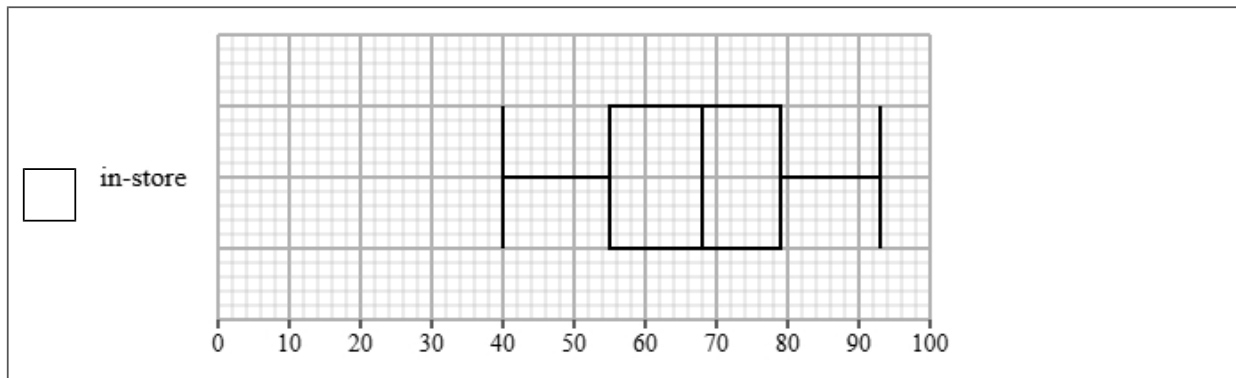
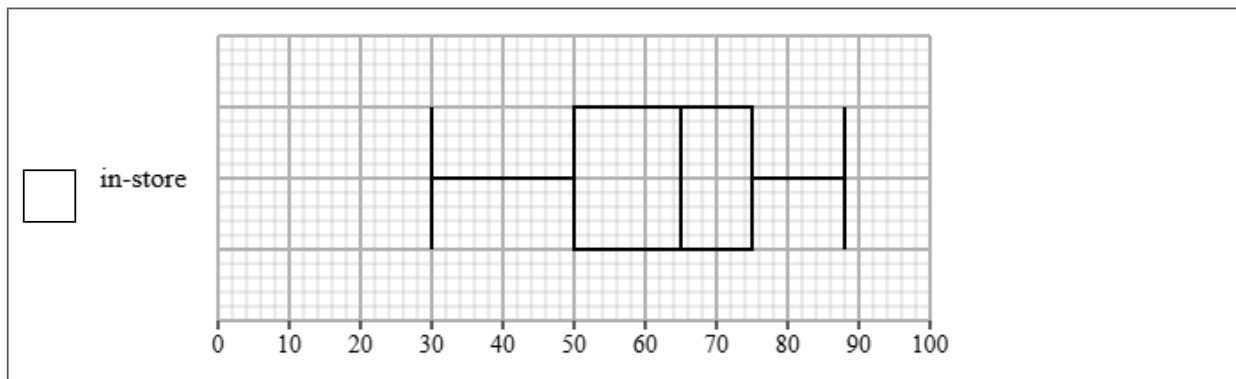
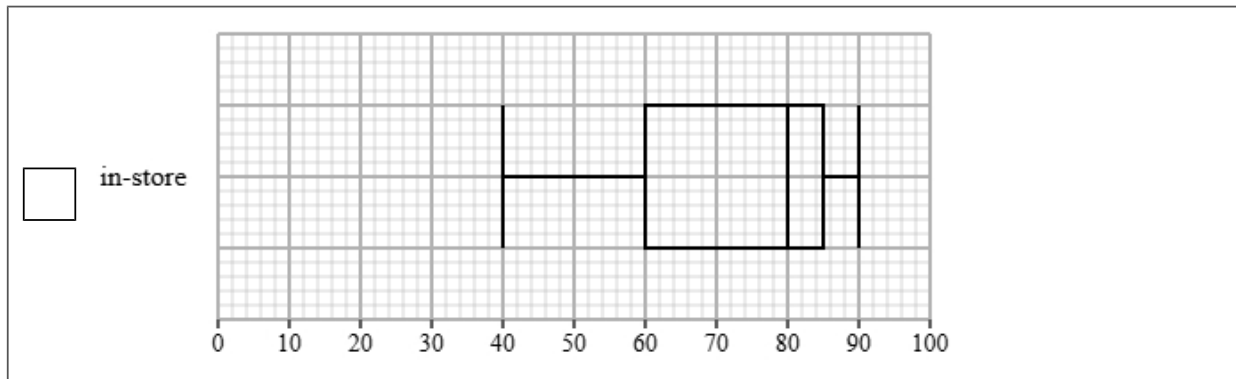
The table gives information about the satisfaction scores for the in-store customers.

Least tall	Lower quartile	Median	Upper quartile	Most tall
40	60	80	85	90

(a) Draw a box plot for the satisfaction scores for the in-store customers.

(2 marks)

Select the correct answer.



(b) Compare the two distributions of satisfaction scores.

Give three comparisons and interpret one of these comparisons.

(4 marks)

Select **one** box.

- The median is bigger.
- The median satisfaction scores for online customers is greater than in-store customers.
- The median satisfaction scores for online and in-store customers are equal.
- The median satisfaction scores for online customers is lower than in-store customers.

Select **one** box.

- The IQR is bigger.
- The IQR for the satisfaction scores of the online customers is greater than in-store customers.
- The IQR for the satisfaction scores of the online and in-store customers are equal.
- The IQR for the satisfaction scores of the online customers is lower than in-store customers.

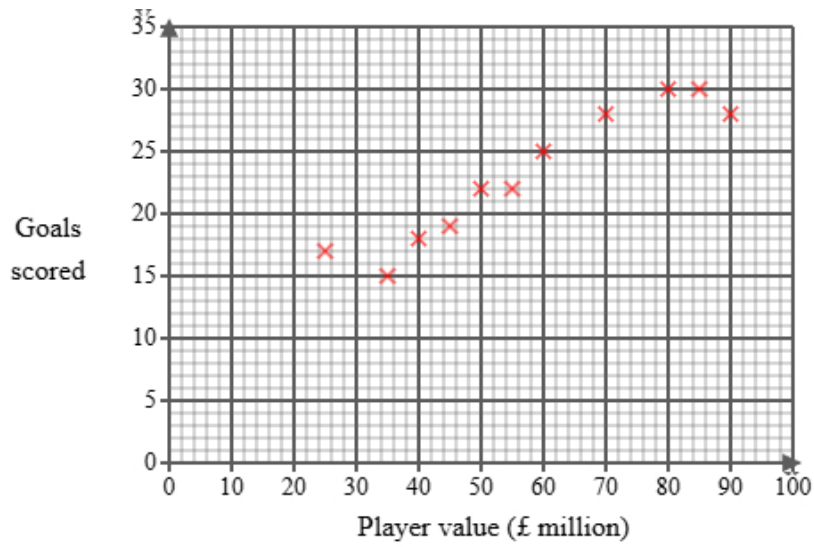
Select **one** box.

- The skews for the satisfaction scores of the online and in-store customers are both positive.
- The skews for the satisfaction scores of the online and in-store customers are both symmetrical.
- The skew for the satisfaction scores of the online customers is symmetrical and the skew for the in-store customers is negative.
- The skew for the satisfaction scores of the online customers is symmetrical and the skew for the in-store customers is positive.

Select **one** box.

- The scores for the online customers are less spread out than the in-store customers.
- The online customers on average have higher satisfaction scores than the in-store customers.
- The online customers are more skewed than in-store customers.
- The online customers on average have lower satisfaction scores than the in-store customers.

- 8 Alice collected data on 11 football strikers, recording the value (in millions of pounds) and the goals scored last season of each player. She represented her findings in the scatter diagram below.



- (a) One of the 11 strikers scored 15 goals last season.

Write down the value of the striker.

(1 mark)

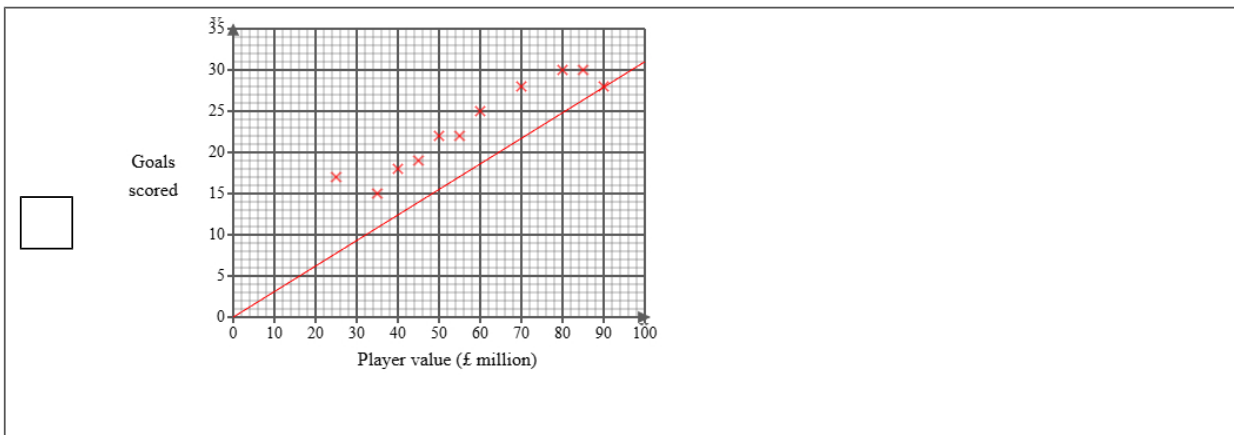
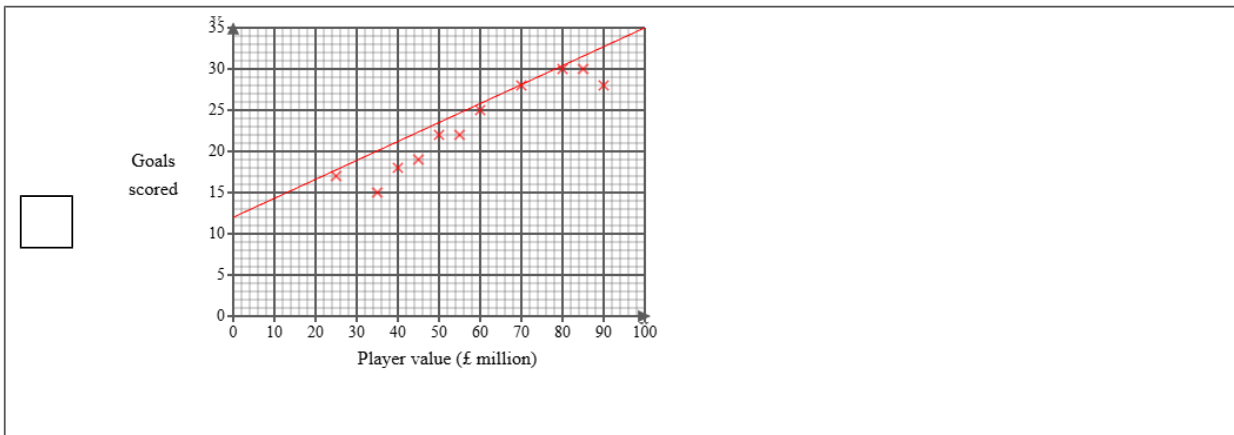
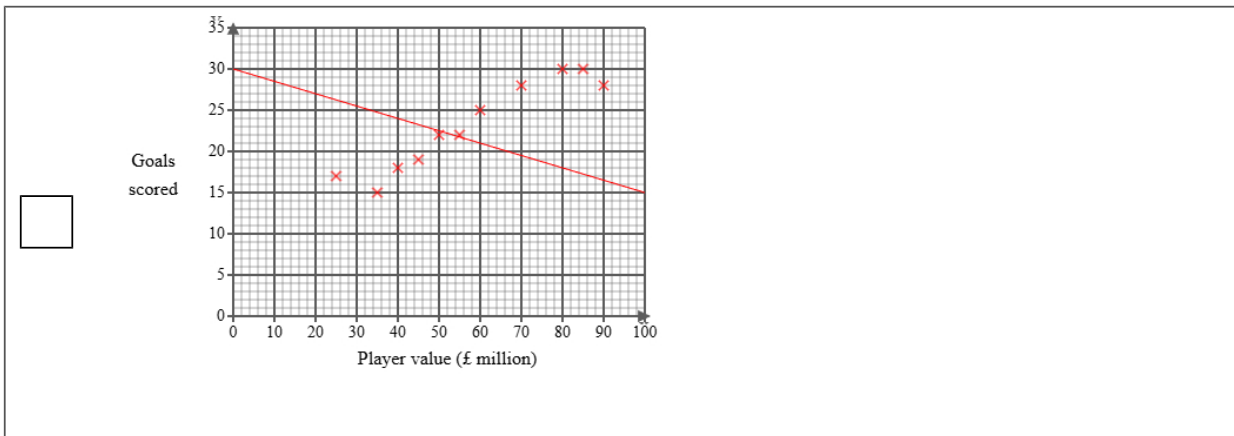
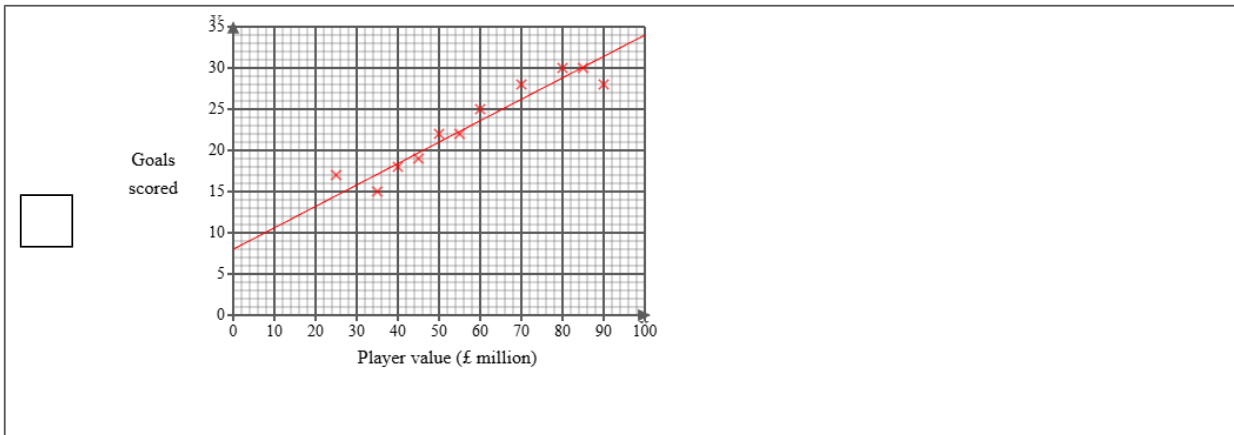
Find the cross on the scatter graph that is at 15 million on the x-axis (the bottom axis), then read off the value from the y-axis (the side axis).

£ _____

(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).

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

Select **one** box.

- As the value increases the goals scored decreases.
- A striker with a high value will have a low goals scored.
- As the value increases the goals scored increases.
- A striker with a high value will have a high goals scored.

(d) A new strike has just been signed for value of £150 million.

Alice is planning on using the line of best fit on the scatter diagram to predict the goals the strike will score this season.

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 not appropriate
- This is appropriate
- because the point is after the data and the trend may not continue.
- because the trend will continue.

9 Olivia investigates the number of pages in 170 books from a library.

The pages range from 302 pages to 389 pages.

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

Table A

Pages	Frequency
$270 < p \leq 300$	0
$300 < p \leq 330$	42
$330 < p \leq 360$	94
$360 < p \leq 390$	34
$390 < p \leq 420$	0

Table B

Pages	Frequency
$300 < p \leq 320$	19
$320 < p \leq 340$	58
$340 < p \leq 360$	59
$360 < p \leq 380$	26
$380 < p \leq 400$	8

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

(2 marks)

Select **two** boxes.

- Grouped data gives more precise results.
- Grouped data ensures that there are no gaps in the data.
- Grouped data is easier to process large amounts of data.
- Grouped data doesn't require any calculation, as it displays frequencies directly.
- Grouped data is easier to represent on graphs.

(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 will lose the detail in the data.
- Grouped data cannot be used with decimals.
- Grouped data cannot show patterns in the data.

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

Assess the appropriateness of Olivia's claim.

(2 marks)

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

- In Table A, all the data is concentrated into three groups.
- Olivia's claim is justified.
- In Table A, the data goes from 270 to 420, showing a much wider range of data.
- In Table B, the table starts at 300 and the lowest value is 302 pages and ends at 400 with the highest value at 389 pages.
- Olivia's claim is not justified.
- In Table B, some data could have been less than 300 or more than 400, but would not be shown.

(d) Olivia wants to work out the average number of pages in the 170 books from a library.

She decides to use Table B.

Calculate the average number of pages in the 170 books from a library, giving your answer to 1 decimal place.

(3 marks)

Add midpoint and fp columns onto the table.

Then find the sums of the f and fp .

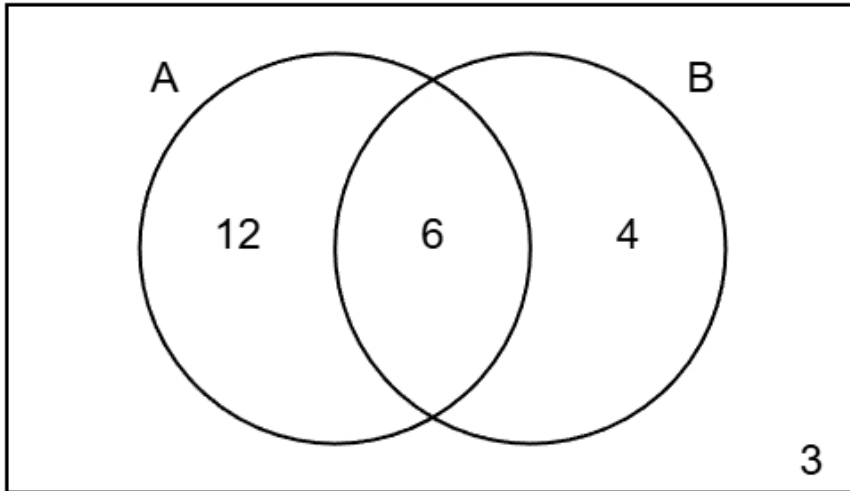
Find the mean by $\frac{\sum fp}{\sum f}$

10 The Venn diagram shows information about 25 research papers published in 2021.

A is the event that the paper was published in a peer-reviewed journal.

B is the event that the paper received more than 500 citations.

The numbers in the Venn diagram indicate the number of papers.



(a) In the Venn diagram, explain what the number 6 means.

(1 mark)

Select **one** box.

The number of research papers published in a peer-reviewed journal but did **not** receive more than 500 citations

The number of research papers published in a peer-reviewed journal **or** received more than 500 citations

The number of research papers published in a peer-reviewed journal **and** received more than 500 citations

The number of research papers that were **not** published in a peer-reviewed journal **or** receive more than 500 citations

(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 $P(B|A)$:

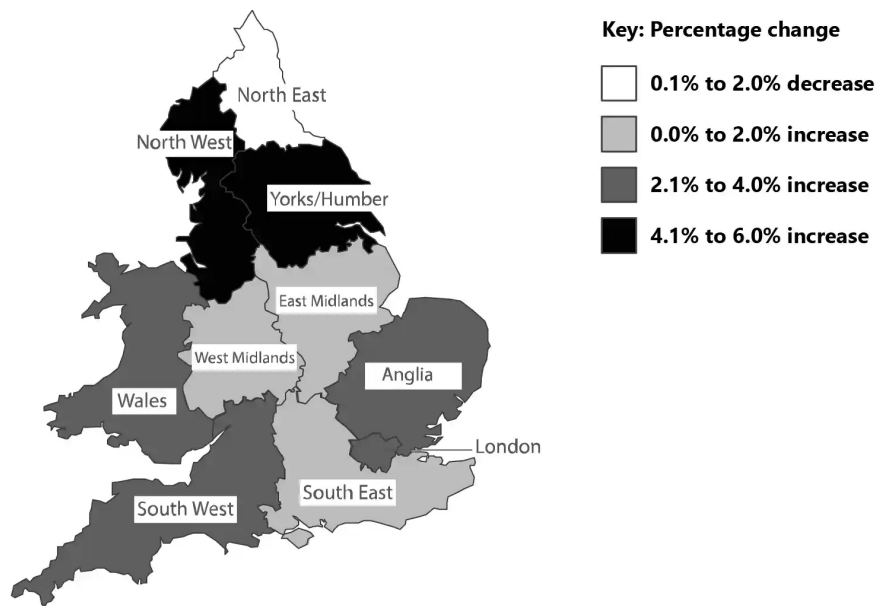
$$P(B) = \frac{\text{amount in B}}{\text{total amount}}$$

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

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

- $P(B) \times P(B | A) = 1$
- $P(B)$ and $P(B | A)$ are not equal
- so they are independent
- 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 2014 and 2015.



- (a) Write down the percentage change in domestic tourist visits between 2014 and 2015 in North West.

(1 mark)

Select **one** box.

- 0.1% to 2.0% decrease
- 0.0% to 2.0% increase
- 2.1% to 4.0% increase
- 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)

Make sure to check the key carefully.

Decreases are shown as white, whilst increases are shaded.

- (c) Hadley states that domestic tourism in England and Wales increased overall between 2014 and 2015.

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'.

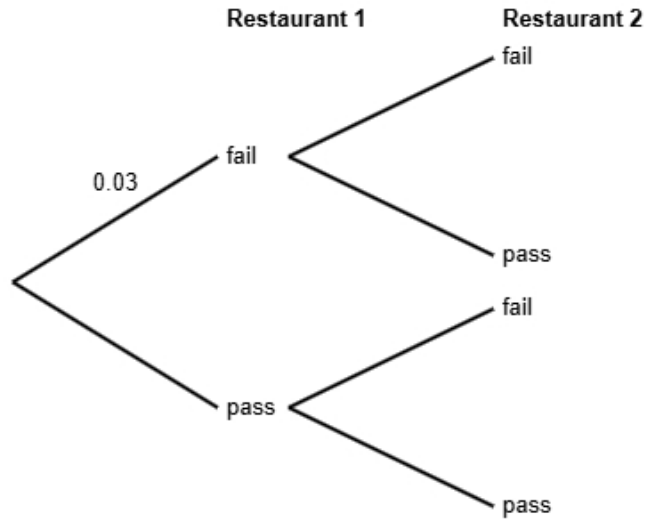
C _ _ _ _ P L E T H

12 Health inspections show that 3% of restaurants in a city fail basic hygiene standards.

The rest pass the required health standards.

Emma is a food safety officer and will be inspecting two different restaurants.

She does not know if each restaurant will fail or pass.



(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 Emma's restaurants have an outcome of pass.

(2 marks)

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

Remember, AND means \times in probability.

(c) Emma states that the probability that exactly one restaurant outcome is fail is less than 6%

Find out whether or not Emma is correct.

(3 marks)

Find the probability of exactly one restaurant outcome is fail (there are two outcomes on the tree diagram).

Select **one** box.

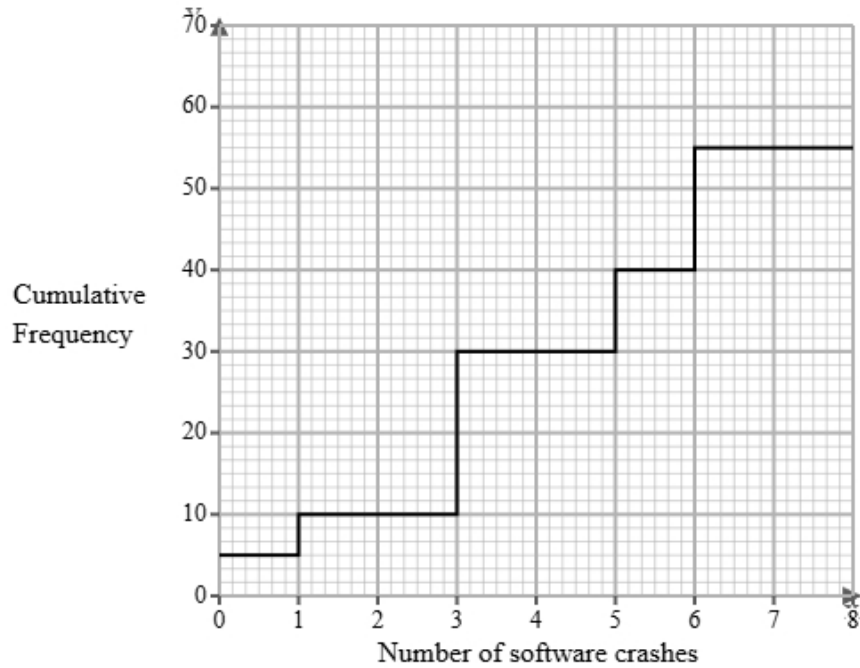
The probability that exactly one restaurant outcome is fail is less than 6%, so Emma is correct.

The probability that exactly one restaurant outcome is fail is more than 6%, so Emma is not correct.

The probability that exactly one restaurant outcome is fail is more than 6%, so Emma is correct.

The probability that exactly one restaurant outcome is fail is less than 6%, so Emma is not correct.

- 13 The cumulative frequency step polygon shows information about number of software crashes experienced by a company's servers over 55 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 software crashes experienced by a company's servers is continuous.
- Because number of software crashes experienced by a company's servers is discrete.
- Because number of software crashes experienced by a company's servers is quantitative.
- Because number of software crashes experienced by a company's servers is qualitative.

- (b) Find the mode of the number of software crashes.

(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 4 software crashes.
 - ii) more than 4 software crashes.

(3 marks)

The frequency is shown by how much the graph 'goes up' at each point.
Remember, the overall frequency is 55.

i) Exactly 4 software crashes: _____

ii) More than 4 software crashes: _____

- (d) In 40 days fewer than x software crashes were experienced.
Find the value of x

(1 mark)

Draw a line across from 40 on the graph and see where all the 'jumps up' to this line are under.

- (e) Delia believes the interquartile range of number of software crashes 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 6, so the IQR must be more 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.