statsGCSE

Scaffolded version

Statistics GCSE

Paper 1

Edexcel Foundation - 2025

Notier Tier

Variant 2 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

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 A bookstore monitors the number of novels sold each day of the week. This helps them track which type of books are more popular at different times.

Monday	
Tuesday	
Wednesday	
Thursday	
Friday	

Key:

 represents 8 novels

(a) Find the amount of novels sold on Monday.

(1 mark)

Make sure to look at the key.

(b) Find the amount of novels sold on Tuesday.

(1 mark)

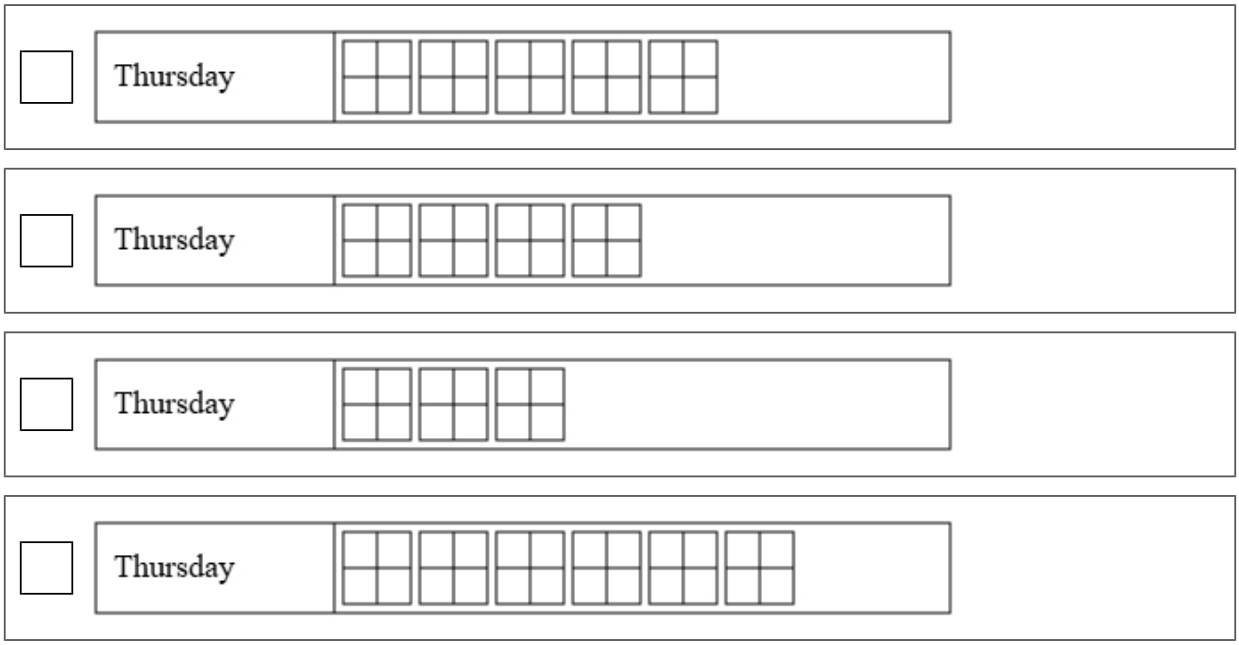
The square is split up into 4 parts. The key shows that four parts represent 8 novels.
Start by finding what one part represents.

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

Show this information on the pictogram.

(1 mark)

Select the correct answer.



(d) Kyler suggests redrawing the pictogram using a key with a whole-square representing 5 novels.

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

(1 mark)

*Select **one** box.*

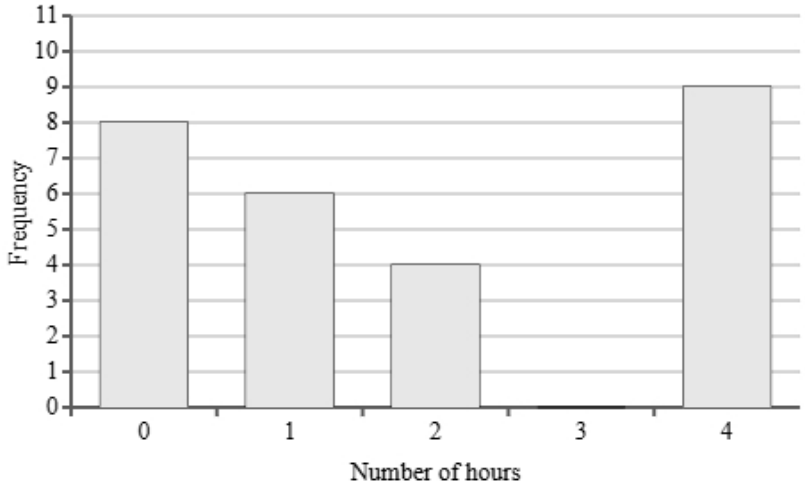
- If Kyler uses 5 novels for the key this will lead to less sales being shown.
- This would be much better because you can fit more squares on.
- The key must always be an even number.
- Monday shows 16 loaves of bread. This would be very difficult to show because 16 has a remainder 1 when divided by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

2 James has collected data on the number of hours people spend watching TV each day.

He asked a total of 30 people.

Each person recorded between 0 and 4 hours of TV time.

The incomplete bar chart displays the number of people who watched 0 hours, 1 hour, 2 hours, and 4 hours of TV.

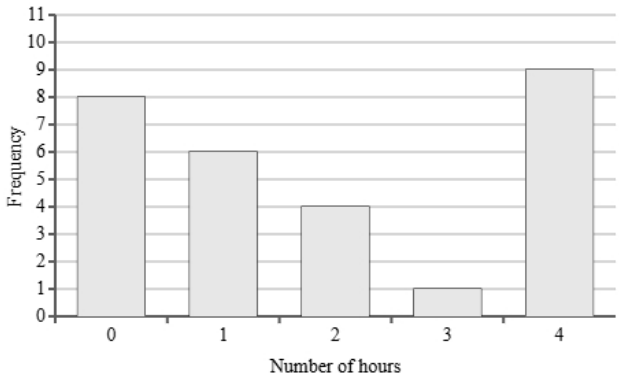


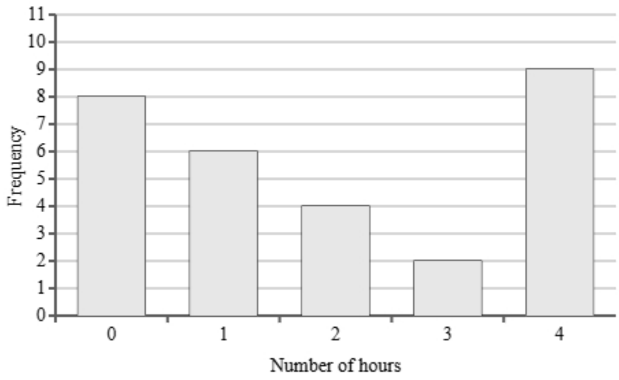
(a) 3 people watched TV for 3 hours.

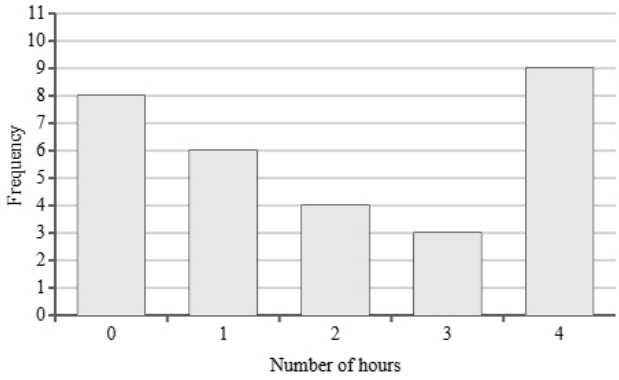
Complete the bar chart.

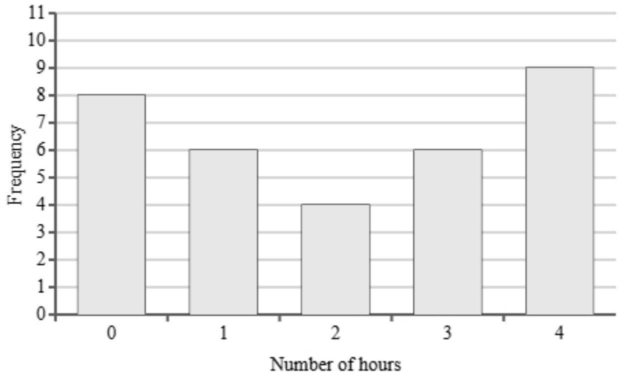
(1 mark)

Select the correct answer.

☐ 

☐ 

☐ 

☐ 

(b) Find how many more people watched no TV than watched 2 hours.

(1 mark)

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

(c) James runs a charity focussed on getting people to watch less TV and be more active.
He thinks that because most people in this area watch 4 hours a day of TV, his charity has been unsuccessful.
Explain why James may not be right.

(1 mark)

*Select **one** box.*

- ☐ More people watched one or less hours of TV a day.
- ☐ People may have watched more than 4 hours.
- ☐ The sample size was small.
- ☐ 8 people watched no TV.

3 A city council is considering adding more public transport routes.

Ethan wants to conduct a survey to learn what all the residents in the city think about the plan.

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

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

(2 marks)

*Select **two** boxes.*

- ☐ Ethan will be able to choose who is in the sample.
- ☐ A sample is cheaper.
- ☐ A sample is quicker.
- ☐ Ethan will be able to explain each question.
- ☐ A sample is more accurate.

(b) Ethan 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.
- ☐ A list of all the members in the population.
- ☐ A list of all the members in the sample.
- ☐ The tally chart or table used in the survey.

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

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

(1 mark)

*Select **one** box.*

- ☐ There will be too many names.
- ☐ The electoral register would also include people's addresses.
- ☐ Only those registered to vote would be included.
- ☐ Unreliable.

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

(2 marks)

*Select **two** boxes.*

- ☐ A pilot survey will include more people.
- ☐ A pilot survey will test questions are understood.
- ☐ A pilot survey will give an idea of response rate.
- ☐ A pilot survey will give more accurate data.
- ☐ A pilot survey will be cheaper than a survey.

(e) Ethan is writing a plan for the investigation into residents' opinions on the new transport routes.

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

You should include:

+ a sampling method

+ a question Ethan 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).*

☐ Ethan should use first-past-the-post sampling.

☐ Ethan should use random sampling.

☐ This will ensure that the students asked are the most knowledgeable.

☐ This will ensure that every resident has an equal chance of being selected.

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

☐ A question could be:
Why do you think the bus routes need to improve?

A question could be:

☐ How satisfied are you with the current bus service?
☐ Very Satisfied ☐ Satisfied ☐ Neutral ☐ Unsatisfied ☐ Very Unsatisfied

☐ 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).*

☐ 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 Andi is studying fin whales and needs to know the average length of a fin whale.

Andi uses the internet to find that the average length of a fin whale is 19.5 metres.

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

(1 mark)

*Select **one** box.*

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

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

(2 marks)

Advantage

*Select **one** box.*

- ☐ The data is numercial.
- ☐ It can be shown in a bar chart.
- ☐ It is easily to access a large amount of data.
- ☐ The data is discrete.

Disadvantage

*Select **one** box.*

- ☐ The data will not be real.
- ☐ Secondary data is always biased.
- ☐ The data may not be in the correct format.
- ☐ The data can not be used for research.

5 David is a local council officer studying the salaries of secondary school teachers in his borough.

He takes a simple random sample of 10 teachers from different schools and asks them to provide their annual salary.

The annual salary of the 10 people are listed:

£38 000	£36 500	£34 500	£35 000	£80 000
£37 000	£36 500	£31 000	£34 000	£37 500

David believes that one of the values is an outlier.

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

(1 mark)

*Select **one** box.*

☐ Individuals are chosen based on specific characteristics to ensure diversity.

☐ Individuals are selected completely at random, with every member of the population has the same probability of being selected

☐ Individuals are chosen in order of their appearance in a list.

☐ Individuals are selected in a way that guarantees equal representation from all subgroups.

(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 within 1 standard deviation of the mean.
- ☐ The value is much higher than the rest of the data.
- ☐ The value has the highest frequency in the dataset.
- ☐ This value is at the exact midpoint between the highest and lowest values.

(e) David 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).*

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

(f) After calculating the mean of the nine salaries without the outlier, David uses this mean in a report to describe all the teachers in the borough.

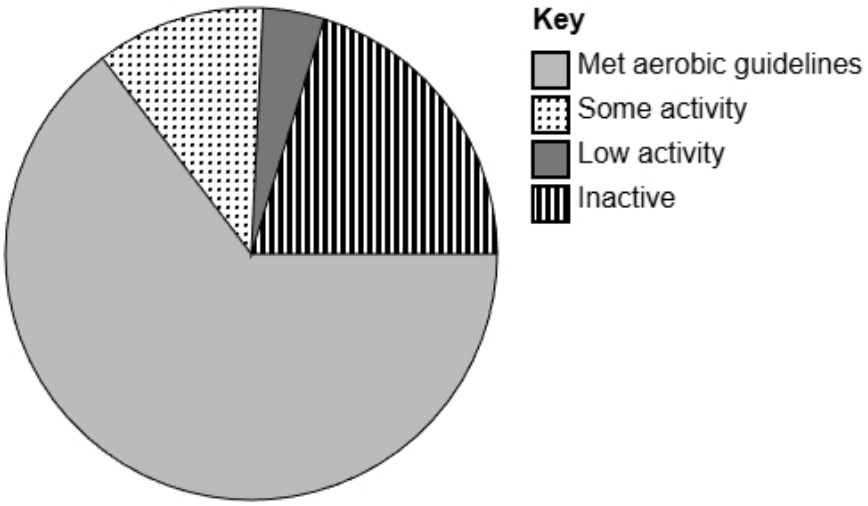
Describe two things that could affect the reliability of her conclusions.

(2 marks)

*Select **two** boxes.*

- ☐ The mean will not measure the spread of the data.
- ☐ The salary of teachers may not truly reflect the hard work they do.
- ☐ One value cannot represent many.
- ☐ Small sample size.
- ☐ Teachers may not want to give their salary.

6 The accurately drawn pie chart shows information about how many people in England met the guidelines for aerobic activity levels in 2021.



(a) Explain how you can tell that most people met the aerobic guidelines in England in 2021 using the pie chart.

(1 mark)

*Select **one** box.*

☐ 'Met aerobic guidelines' is the first value in the key.

☐ 'Met aerobic guidelines' has the largest sector.

☐ 'Met aerobic guidelines' is the most positive response.

☐ 'Met aerobic guidelines' is at the bottom of the pie chart.

(b) The population in England in 2021 was estimated to be 56 million.
Calculate an estimate for the number of people in the UK in 2021 who 'Met aerobic guidelines'.
Round your answer to the nearest million.

(2 marks)

Start by measuring the angle with a protractor.

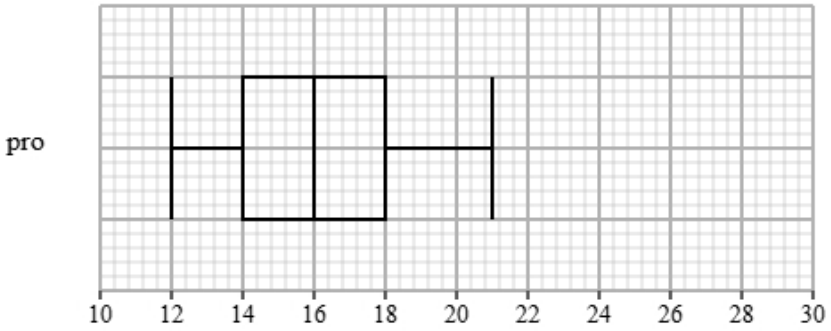
Next find the amount. Remember that there are 360 degrees in a circle.

_____ million

7 Liam recorded the completion times for pro and beginner runners in a 5K race.

Both groups ran the same course.

The box plot presents data on the completion times for the pro runners.



The table gives information about the completion times for the beginner runners.

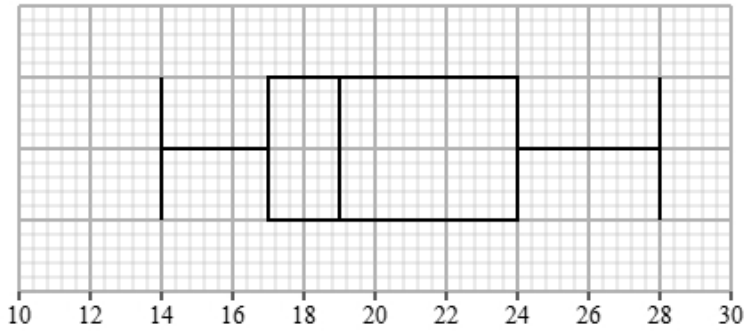
Least tall	Lower quartile	Median	Upper quartile	Most tall
17	19	20	24	30

(a) Draw a box plot for the completion times for the beginner runners.

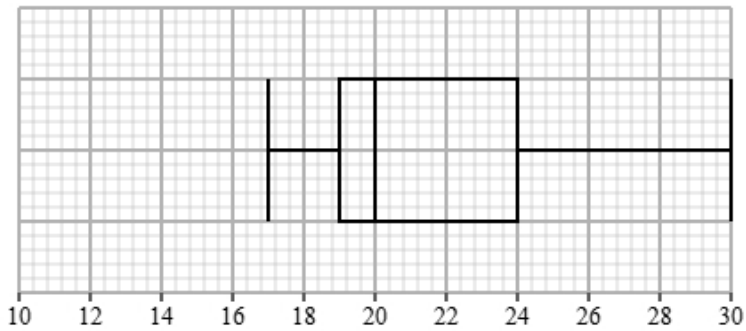
(2 marks)

Select the correct answer.

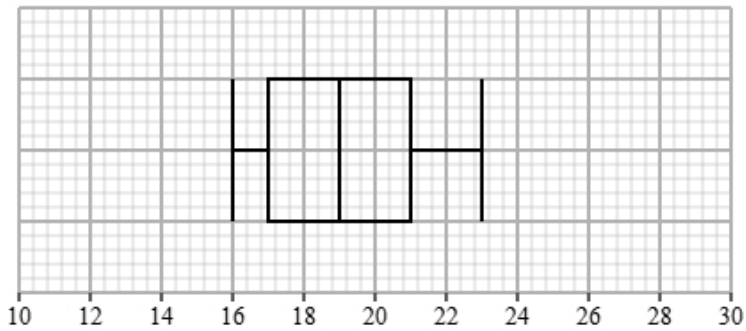
☐ beginner



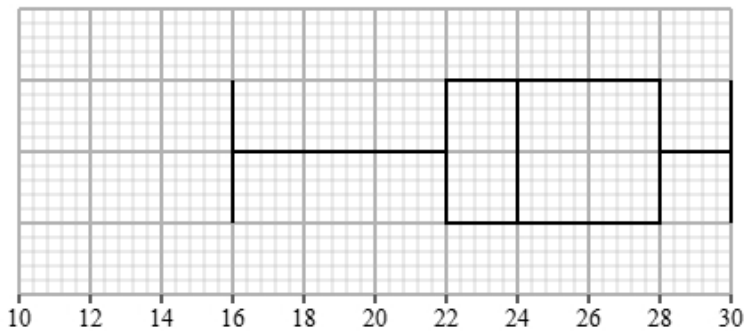
☐ beginner



☐ beginner



☐ beginner



(b) Compare the two distributions of completion times.

Give three comparisons and interpret one of these comparisons.

(4 marks)

*Select **one** box.*

- ☐ The median is bigger.
- ☐ The median completion times for pro runners is greater than beginner runners.
- ☐ The median completion times for pro runners is lower than beginner runners.
- ☐ The median completion times for pro and beginner runners are equal.

*Select **one** box.*

- ☐ The IQR is bigger.
- ☐ The IQR for the completion times of the pro runners is greater than beginner runners.
- ☐ The IQR for the completion times of the pro and beginner runners are equal.
- ☐ The IQR for the completion times of the pro runners is lower than beginner runners.

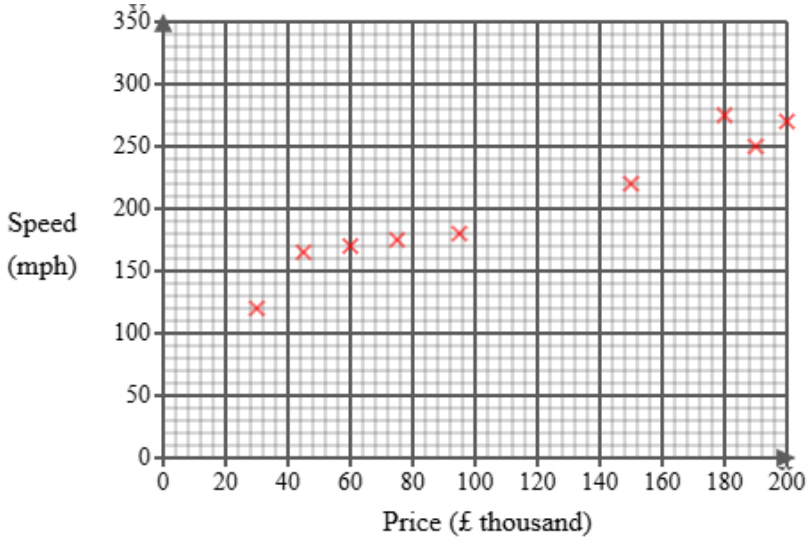
*Select **one** box.*

- ☐ The skews for the completion times of the pro and beginner runners are both positive.
- ☐ The skew for the completion times of the pro runners is symmetrical and the skew for the beginner runners is positive.
- ☐ The skews for the completion times of the pro and beginner runners are both symmetrical.
- ☐ The skew for the completion times of the pro runners is symmetrical and the skew for the beginner runners is negative.

*Select **one** box.*

- ☐ The times for the pro runners are more spread out than the beginner runners.
- ☐ The pro runners are on average slower than the beginner runners.
- ☐ The pro runners are on average faster than the beginner runners.
- ☐ The pro runners are more skewed than beginner runners.

8 Emma recorded data on 11 cars, recording their top speed (in miles per hour) and their price (in thousands of pounds). She represented her findings in the scatter diagram below.



(a) One of the 11 cars has a top speed of 225 mph.

For this car, write down its price.

(1 mark)

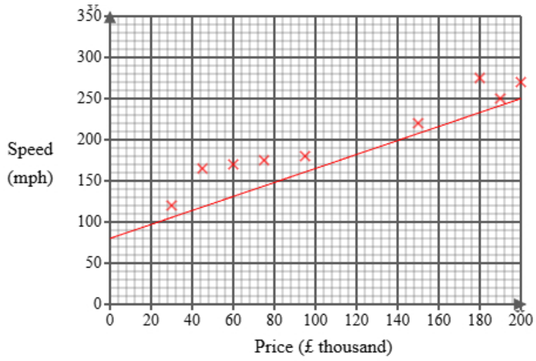
Find the cross on the scatter graph that is at 225 mph on the x-axis (the bottom axis), then read off the value from the y-axis (the side axis).

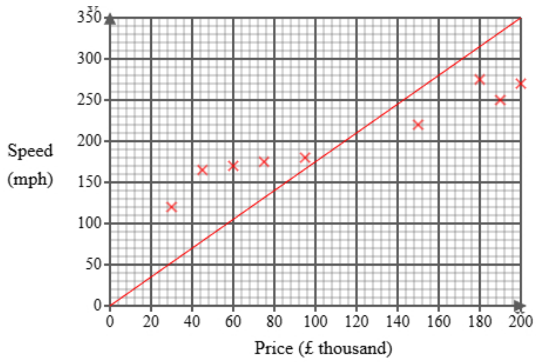
_____ mph

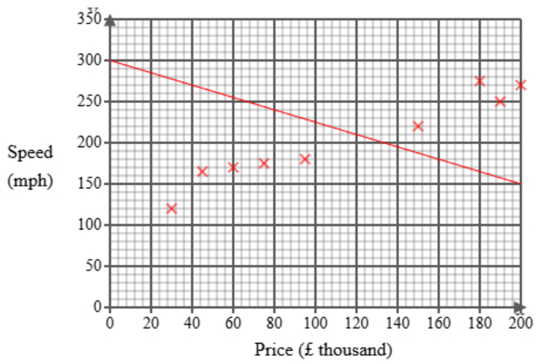
(b) Draw a line of best fit on the scatter diagram.

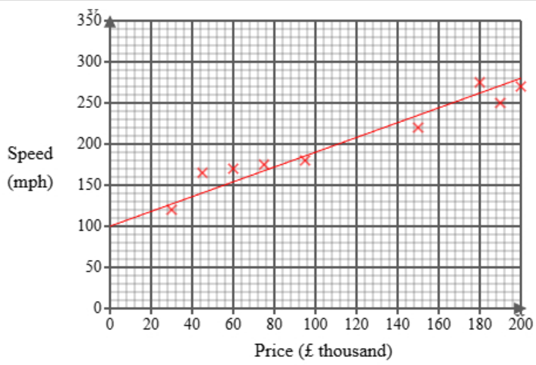
(1 mark)

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

*Select **one** box.*

- ☐ As the price increases the top speed decreases.
- ☐ A high price car will have a low top speed.
- ☐ As the price increases the top speed increases.
- ☐ A high price car will have a high top speed.

(d) A new car will be releasing soon with a price of £250,000.

Emma is planning on using the line of best fit on the scatter diagram to predict the top speed of the new car.

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 the point is inside the range of the data.
- ☐ because the point is outside the range of the data.
- ☐ This is appropriate
- ☐ This is not appropriate

9 Liam investigates the weights of 180 dogs at a pet shelter.

The weights range from 32 kg to 78 kg.

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

Table A

Weight (w kg)	Frequency
$0 < w \leqslant 30$	0
$30 < w \leqslant 60$	87
$60 < w \leqslant 90$	93
$90 < w \leqslant 120$	0
$120 < w \leqslant 150$	0

Table B

Weight (w kg)	Frequency
$30 < w \leqslant 40$	5
$40 < w \leqslant 50$	28
$50 < w \leqslant 60$	54
$60 < w \leqslant 70$	59
$70 < w \leqslant 80$	34

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

(2 marks)

*Select **two** boxes.*

- ☐ Grouped data eliminates the need to use any statistical methods.
- ☐ Grouped data keeps the precision in the data.
- ☐ Grouped data makes it easier to identify outliers.
- ☐ Grouped data is easier to read.
- ☐ Grouped data makes it easier to process large amounts of 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 may introduce too much complexity when analysing trends.
- ☐ Grouped data will lose the accuracy in the data.
- ☐ Grouped data cannot be used to identify patterns.

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

Assess the appropriateness of Liam's claim.

(2 marks)

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

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

(d) Liam wants to work out the average weight of the 180 dogs at the pet shelter.

He decides to use Table B.

Calculate the average weight of the 180 dogs at the pet shelter, giving your answer to 1 decimal place.

(3 marks)

Add midpoint and fw columns onto the table.

Then find the sums of the f and fw.

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

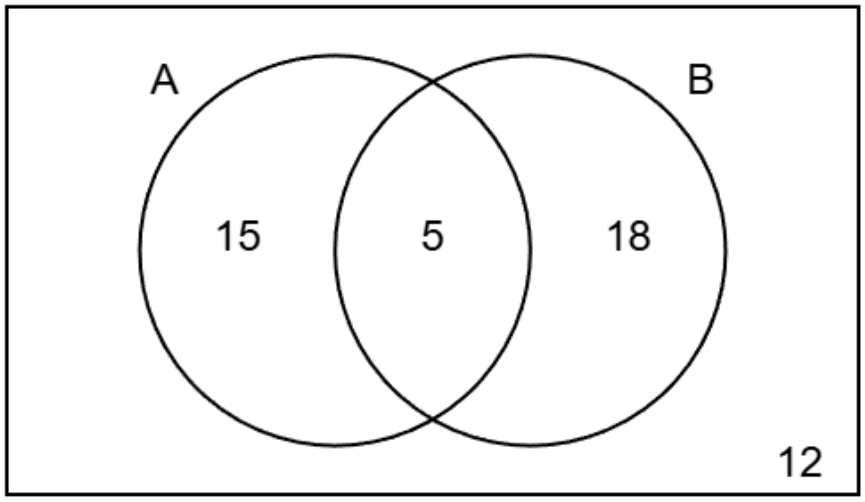
_____ kg

10 The Venn diagram shows information about 50 students in a school.

A is the event that the student plays a musical instrument.

B is the event that the student is part of the school's debate team.

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



(a) In the Venn diagram, explain what the number 5 means.

(1 mark)

*Select **one** box.*

- ☐ The number of students who play a musical instrument **and** are part of the school's debate team
- ☐ The number of students who play a musical instrument but are **not** part of the school's debate team
- ☐ The number of students who play a musical instrument **or** are part of the school's debate team
- ☐ The number of students that do **not** play a musical instrument **or** are part of the school's debate team

(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):

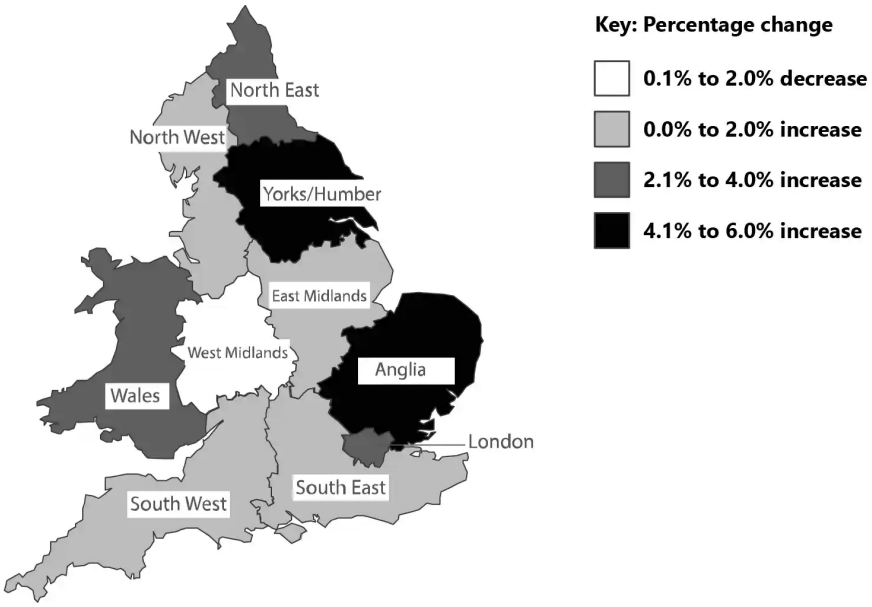
$$P(B) = \frac{\text{amount in B}}{\text{total amount}}$$

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

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

- ☐ P(B) x P(B | A) = 1
- ☐ so they are not independent
- ☐ so they are independent
- ☐ P(B) and P(B | A) are not equal

11 The map below shows the percentage change in domestic tourist visits across different regions of England and Wales between 2011 and 2012.



(a) Write down the percentage change in domestic tourist visits between 2011 and 2012 in South East.

(1 mark)

*Select **one** box.*

☐ 0.1% to 2.0% decrease

☐ 4.1% to 6.0% increase

☐ 2.1% to 4.0% increase

☐ 0.0% to 2.0% 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) Priscilla states that domestic tourism in England and Wales increased overall between 2011 and 2012.

Explain why this may not be the case.

(1 mark)

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

(1 mark)

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

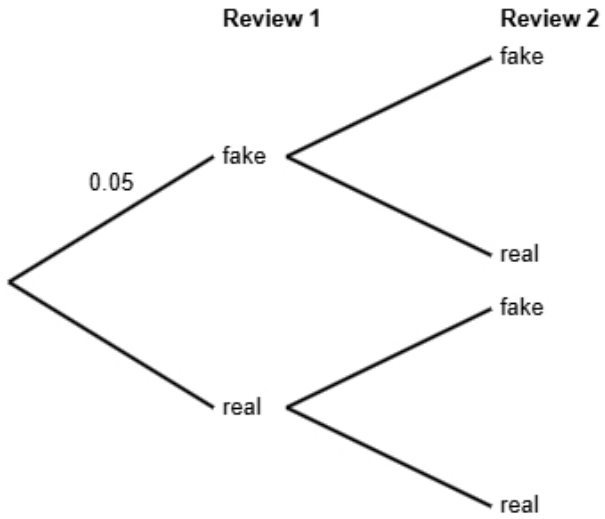
C _ _ _ _ P L E T H

12 Research suggests that 5% of online product reviews are fake.

All other reviews are genuine.

Emma is reading two reviews for a product.

She does not know if each review is fake or real.



(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 reviews are real.

(2 marks)

You will need to find P(real) AND P(real).

Remember, AND means × in probability.

(c) Emma states that the probability that exactly one reviews is fake is less than 10%

Find out whether or not Emma is correct.

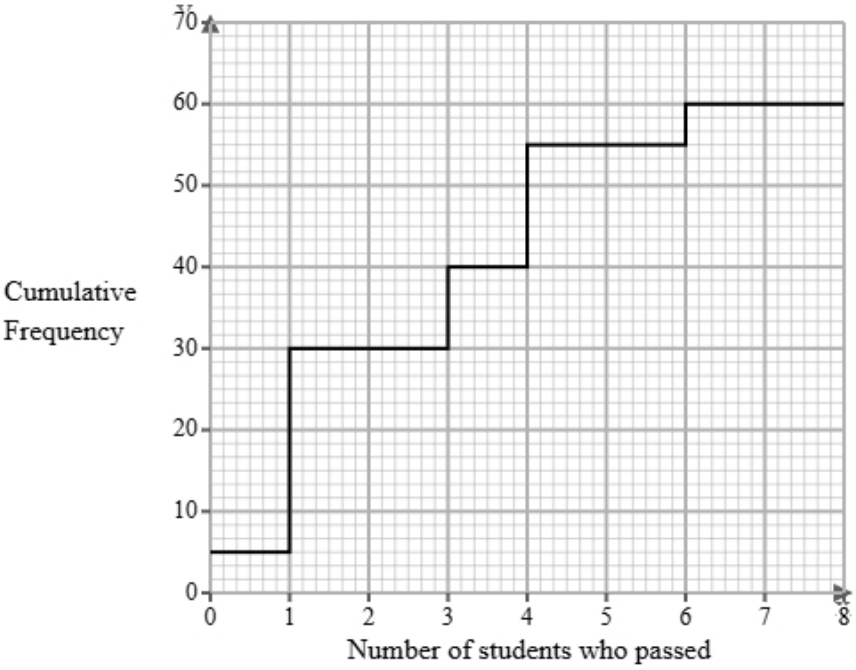
(3 marks)

Find the probability of exactly one reviews is fake (there are two outcomes on the tree diagram).

*Select **one** box.*

- ☐ The probability that exactly one reviews is fake is less than 10%, so Emma is not correct.
- ☐ The probability that exactly one reviews is fake is more than 10%, so Emma is not correct.
- ☐ The probability that exactly one reviews is fake is more than 10%, so Emma is correct.
- ☐ The probability that exactly one reviews is fake is less than 10%, so Emma is correct.

13 The cumulative frequency step polygon shows information about the number of students who passed a daily maths quiz over 60 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 students who passed a daily maths quiz is continuous.

☐ Because the number of students who passed a daily maths quiz is qualitative.

☐ Because the number of students who passed a daily maths quiz is quantitative.

☐ Because the number of students who passed a daily maths quiz is discrete.

(b) Find the mode of the number of students who passed a daily maths quiz.

(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 5 students who passed.

ii) more than 5 students who passed.

(3 marks)

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

Remember, the overall frequency is 60.

i) Exactly 5 students who passed: _____

ii) More than 5 students who passed: _____

(d) In 40 days fewer than x students passed.

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) Peter believes the interquartile range of the number of students who passed 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.