Scaffolded version

Statistics GCSE

Paper 2

2025

Edexcel Foundation

Variant 2 | 1ST0/2F

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.

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 Lisa asked her colleagues in her office which coffee they prefer from the choices Espresso, Latte, Cappuccino and Mocha.
Their answers are shown below:

Latte Cappuccino Espresso Latte Mocha Espresso
Espresso Latte Cappuccino Mocha Espresso Latte Espresso
Espresso Espresso Cappuccino Espresso Mocha Latte Latte

(a) Complete the tally and frequency columns on the frequency table for the data.

(2 marks)

Select the correct answer.

☐

Coffee	Tally	Frequency
Espresso	𝍸 𝍸 I	11
Latte	𝍸 III	8
Cappuccino	IIII	4
Mocha	IIII	4

☐

Coffee	Tally	Frequency
Espresso	𝍸 IIII	9
Latte	𝍸 II	7
Cappuccino	IIII	4
Mocha	IIII	4

☐

Coffee	Tally	Frequency
Espresso	𝍸 III	8
Latte	𝍸 I	6
Cappuccino	III	3
Mocha	III	3

☐

Coffee	Tally	Frequency
Espresso	𝍸 IIII	9
Latte	𝍸 II	7
Cappuccino	III	3
Mocha	III	3

(b) Find the number of colleagues in Lisa's office.

(1 mark)

Each person gave one answer, so count the number of answers.

____

(c) Compare the amount of colleagues who chose Espresso with the amount who chose Mocha.

(1 mark)

*Select **one** box.*

☐ The colleagues selected Mocha and Espresso in equal amounts.

☐ More colleagues selected Mocha than Espresso.

☐ More colleagues selected Espresso than Mocha.

(d) Lisa finds the mode from the data.
Give a reason why the mode is an appropriate average to use in this situation.

(1 mark)

*Select **one** box.*

☐ The data is numeric.

☐ The data is qualitative.

☐ The data is continuous.

☐ The mode will be quicker to find.

(e) Give **one** advantage of displaying this data in a tally chart rather than leaving it as raw data.

(1 mark)

*Select **one** box.*

☐ It will reduce mistakes with the data.

☐ The data is more organised.

☐ A tally chart shows more data.

☐ It is more reliable.

2 The table shows the results of a survey into Weekly Physical Activity.
For example, 30% of all respondents said "Went to the gym" but only 25% of 16-24 year olds said this.
Some respondents participated in more than one option.

Physical Activity	**All People %**	**Gender**		**Age**			
		Male	**Female**	**18-26**	**27-36**	**37-56**	**57+**
Went to the gym	30%	25%	35%	25%	27%	33%	35%
Played outdoor sports	70%	72%	68%	78%	72%	69%	61%
Did home workouts	42%	37%	47%	38%	41%	43%	46%
Did yoga or pilates	15%	11%	19%	12%	16%	17%	15%
Went running or jogging	28%	34%	22%	36%	22%	26%	28%
Attended fitness classes	28%	21%	35%	11%	19%	36%	46%
Went cycling	39%	40%	38%	48%	46%	35%	27%
Walked for exercise	42%	45%	39%	34%	42%	34%	58%
None of the above	5%	6%	4%	7%	5%	5%	3%

(a) A person who selected **Attended fitness classes** is chosen at random.
Describe the most likely gender and age of the person.

(2 marks)

Gender

*Select **one** box.*

- [ ] Female
- [ ] Male

Age group

Select the correct boxes.

- [ ] 57+
- [ ] 37-56
- [ ] 27-36
- [ ] 18-26

(b) Describe the relationship in the data between attending fitness classes and age.

(1 mark)

*Select **one** box.*

- [ ] The proportion of people attending fitness classes decreases with age.
- [ ] The proportion of people attending fitness classes is the same at all ages.
- [ ] The proportion of people attending fitness classes is highest at 57+.
- [ ] The proportion of people attending fitness classes increases with age.

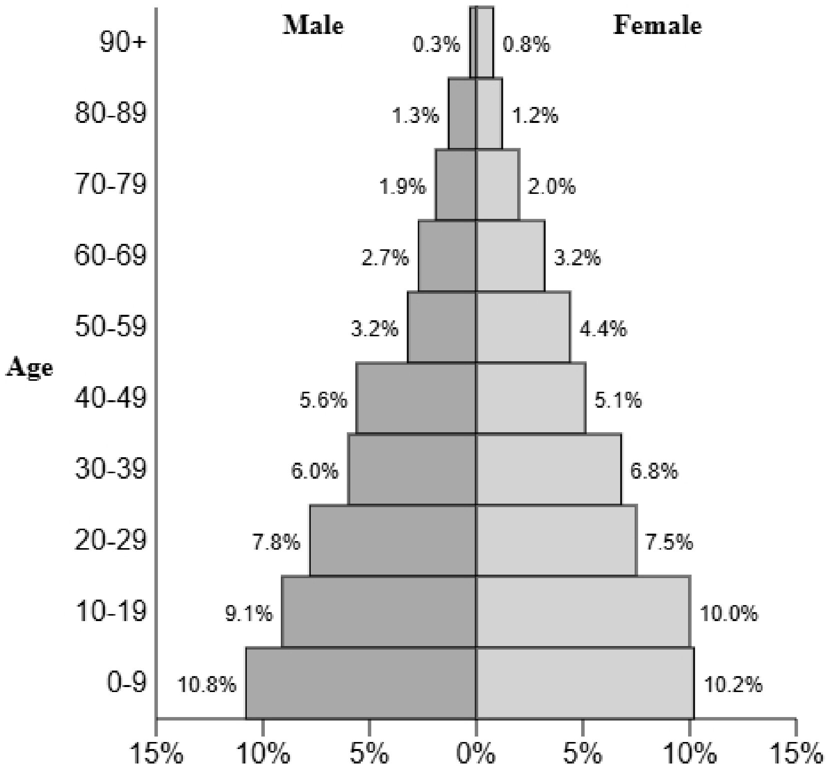
(c) Amelia is starting a yoga class.
She suspects that over half of her customers will be female.
Explain how she used the information from the table to come to this conclusion.

(1 mark)

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

- ☐ but the percentage for males was higher at 19%
- ☐ but the percentage for males was lower at 11%
- ☐ The percentage of females who did yoga or pilates was 11%
- ☐ The percentage of females who did yoga or pilates was 19%

3 The population pyramid below shows the percentage of males and females in each age group for the town Elderleigh.



(a) Write down the percentage of females in the age group 30-39.

(1 mark)

Find 30-39 on the population pyramid and read off the number on the right-hand side (females).

__________%

(b) Find the age group for males that has 3.2% of the population.

(1 mark)

Select ***one*** *box.*

- ☐ 20-29
- ☐ 40-49
- ☐ 50-59
- ☐ 30-39

(c) Find the age group that has 19.1% of the population.

(1 mark)

Select ***one*** *box.*

- ☐ 30-39
- ☐ 10-19
- ☐ 20-29
- ☐ 40-49

(d) Compare the percentage of the population aged 20-49 between males and females.

(1 mark)

Select ***one*** *box.*

- ☐ There are more males.
- ☐ There are more females.
- ☐ They are both the same.

(e) Give a reason why the sum of all the percentages is 99.9% and not 100%.

(1 mark)

*Select **one** box.*

- ☐ It is incomplete.
- ☐ There are gaps in the ages.
- ☐ The figures are wrong.
- ☐ They have been rounded.

4 A theatre sells four different ticket types at different prices.
The tickets for a weekend have been recorded in this table.

Ticket price	£35	£45	£75	£105	Total
Amount sold	1092	728	494	286	2600

The owner of the theatre wants to open up a new theatre in a different city.
She uses the data to predict how the tickets will sell in the new location.

Ticket price	£35	£45	£75	£105
Predicted proportion	42%	28%	19%	11%

(a) The owner want to improve her predictions.
Explain how she can do this.

(1 mark)

*Select **one** box.*

- ☐ Record data over more days.
- ☐ Rely more on the tickets she feels are popular.
- ☐ Draw a pie chart from the data.
- ☐ Only collect data on one of the days.

(b) The owner opens up the new theatre.
The table below shows the tickets sales over the opening weekend at the new theatre.

Ticket price	£35	£45	£75	£105
True proportion	33%	25%	23%	19%

Compare the sales of £105 tickets at the new location compared to the original location and suggest a reason for any difference.

(2 marks)

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

☐ There was a higher proportion of £105 tickets at the new location.

☐ This could have been because at an opening weekend people may be happier to spend more on a ticket.

☐ There was a higher amount of £105 tickets at the new location.

☐ This could have been because the new location was more popular.

5 Aisha owns a food truck that sells sandwiches.
Customers can choose either chicken or tuna sandwich **and** either a iced tea or a orange juice drink.
She records her sales in a two-way table, but the table is incomplete.

	iced tea	orange juice	Total
chicken sandwich		46	96
tuna sandwich	49	32	
Total	99		177

(a) Complete the two way table.

(2 marks)

	iced tea	orange juice	Total
chicken		46	96
tuna	49	32	
Total	99		177

(b) Find the probability a customer chosen at random:

i) ordered a chicken sandwich.

ii) ordered a tuna sandwich and a iced tea.

iii) did **not** order a iced tea.

(4 marks)

You can use the formula:

$$\text{Probability} = \frac{\text{Desired frequency}}{\text{Total frequency}}$$

The total frequency is found at the bottom-right of the table.

For part iii, we are looking at **not** orange juice, which means they must have picked iced tea.

Probability of chicken sandwich.

Probability of tuna sandwich and a iced tea.

Probability of **not** a iced tea.

(c) Aisha is ordering drinks for the following month.
Explain, with a reason, if Aisha should order more bottles of iced tea than orange juice.

(2 marks)

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

- ☐ No, Aisha should buy less bottles of orange juice than iced tea
- ☐ Yes, Aisha should buy more bottles of iced tea than orange juice
- ☐ because she sold more iced tea than orange juice.
- ☐ because she sold less orange juice than iced tea.

6 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.

7 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) 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.

(c) 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.

(d) 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.

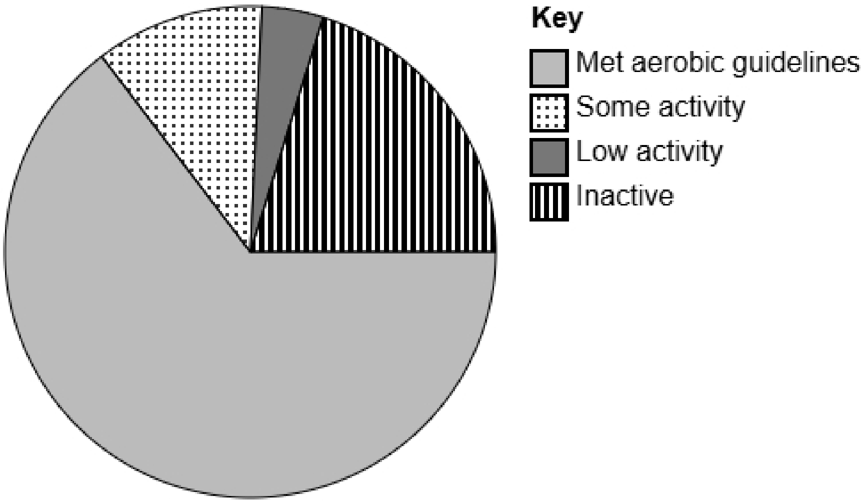
☐ The salary of teachers may not truly reflect the hard work they do.

☐ One value cannot represent many.

☐ Small sample size.

☐ Teachers may not want to give their salary.

8 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.

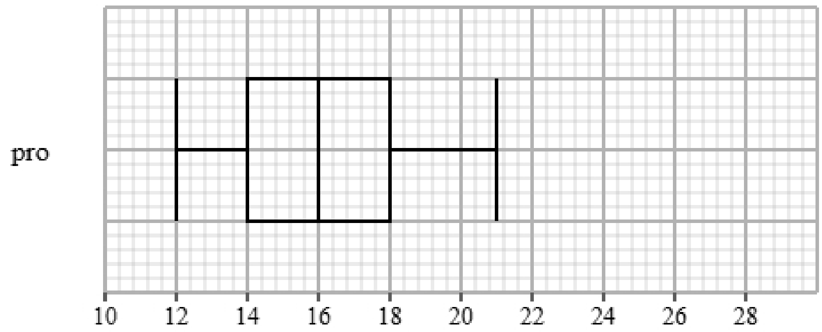
(b) The population in England in 2021 was estimated to be 56 million.
Calculate an estimate for the number of people in the UK in 2021 who 'Met aerobic guidelines'.
Round your answer to the nearest million.

(2 marks)

Start by measuring the angle with a protractor. Next find the amount. Remember that there are 360 degrees in a circle.

__________ million

9 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.

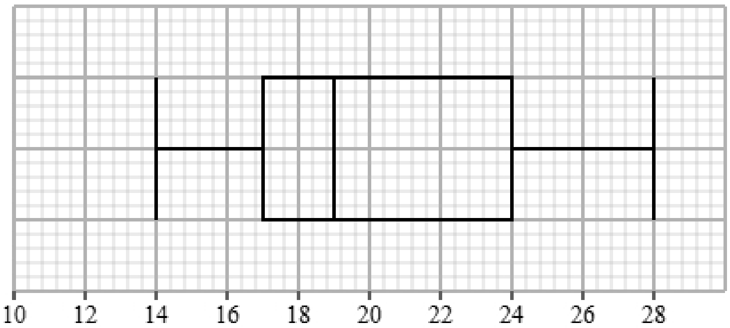
Least tall	Lower quartile	Median	Upper quartile	Most tall
17	19	20	24	30

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

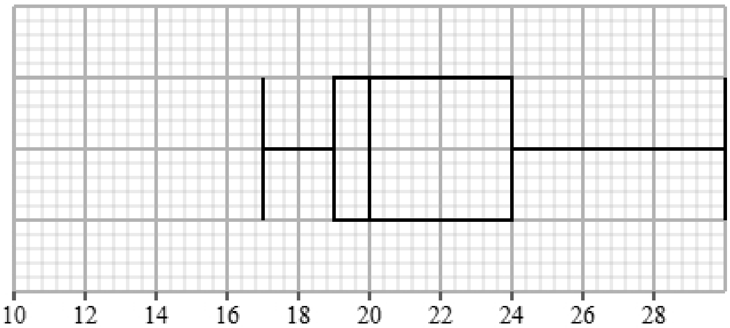
(2 marks)

Select the correct answer.

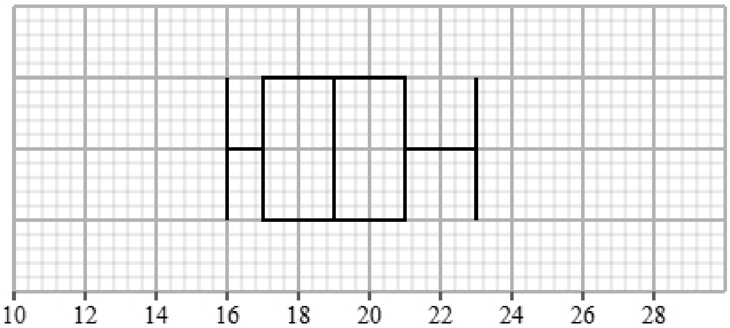
☐ beginner



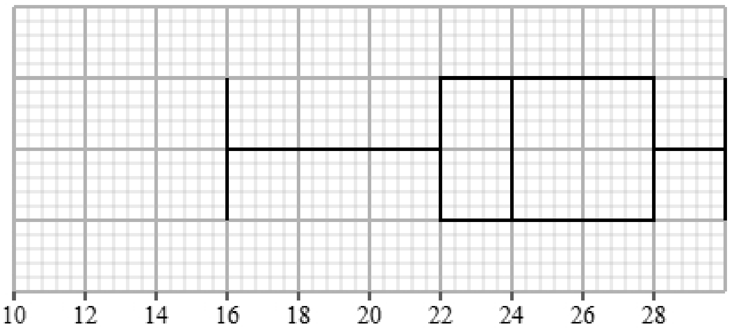
☐ beginner



☐ 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.*

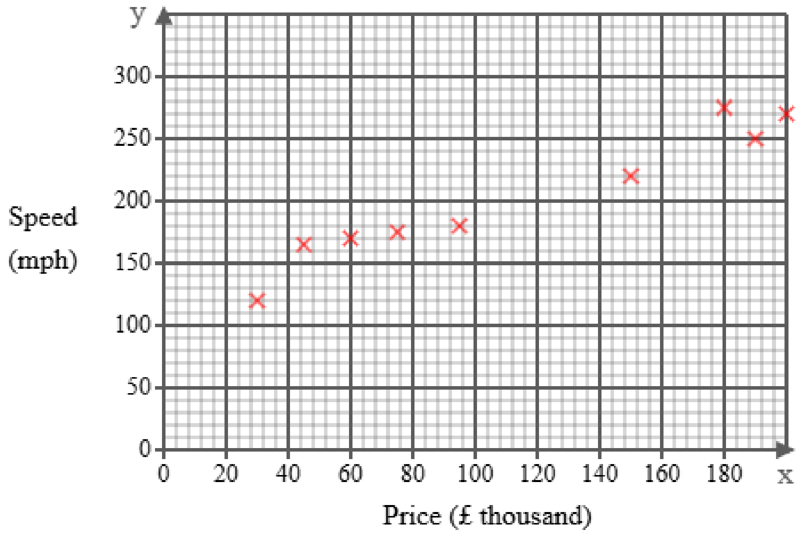
☐ The times for the pro runners are more spread out than the beginner runners.

☐ 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.

10 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.

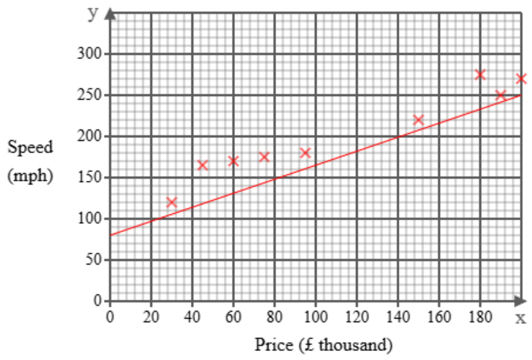


(a) Draw a line of best fit on the scatter diagram.

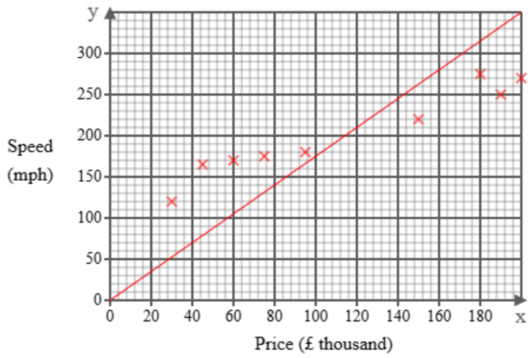
(1 mark)

Select the correct answer.

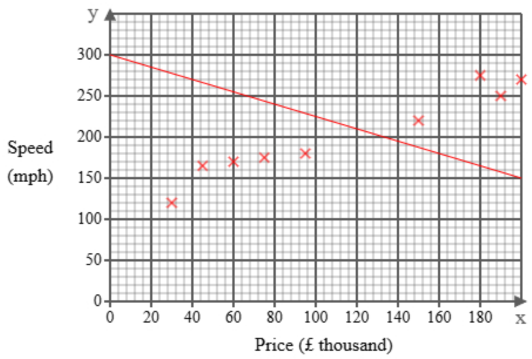
☐



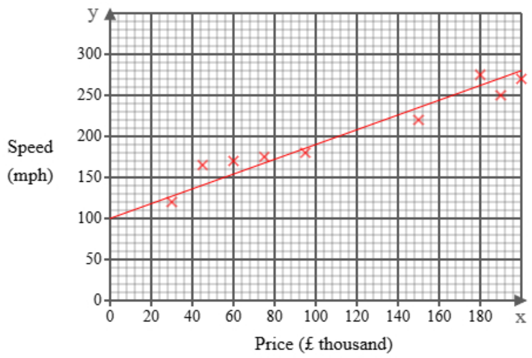
☐



☐



☐



(b) 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.

(c) 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

11 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) 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.

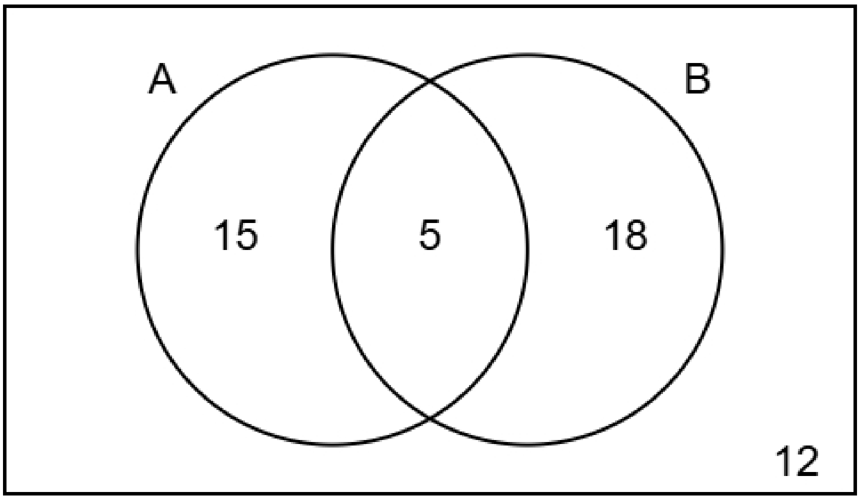
(c) Liam wants to work out the average weight of the 180 dogs at the pet shelter.
He decides to use Table B.
Calculate the average weight of the 180 dogs at the pet shelter, giving your answer to 1 decimal place.

(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

12 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):

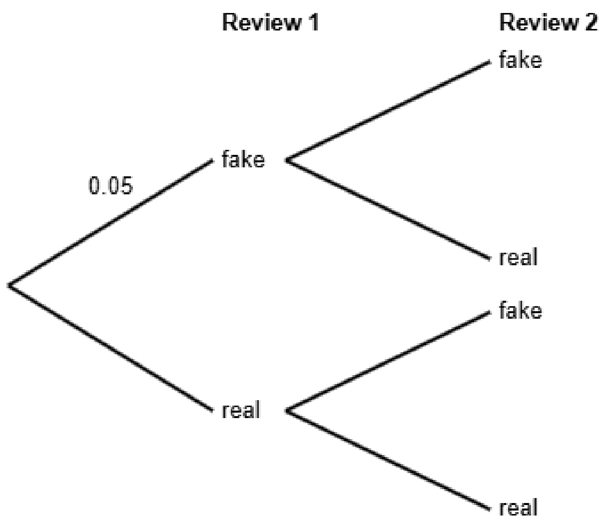
$$P(B) = \frac{\text{amount in B}}{\text{total amount}}$$

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

*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

13 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.*

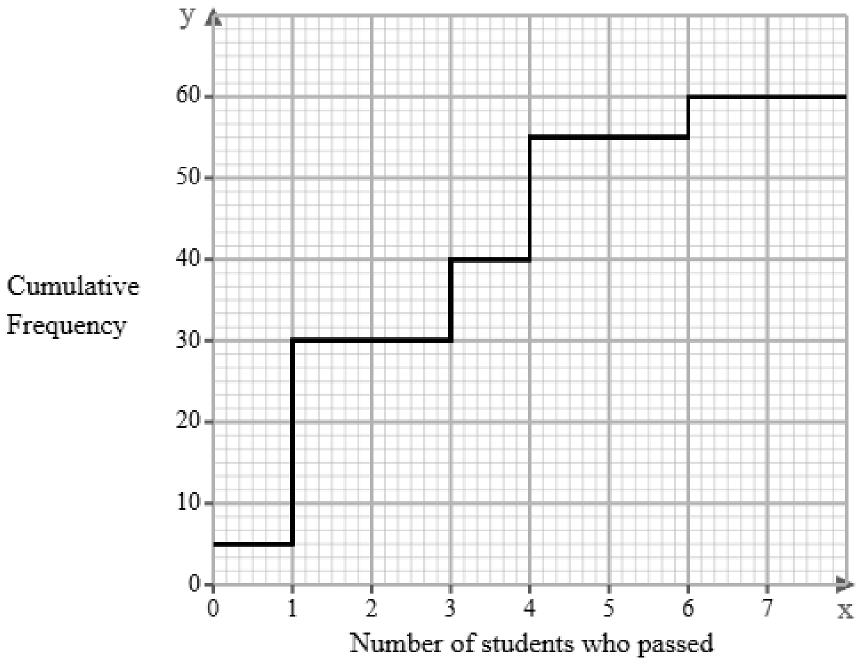
☐ The probability that exactly one reviews is fake is less than 10%, so Emma is not correct.

☐ 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.

14 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 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: __________

(c) 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.

____

(d) 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.