statsGCSE

Scaffolded version

Statistics GCSE

Paper 2

2025

Edexcel Foundation

Variant 4 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 Ms. Rivera asked the students in her art class which colour is their favourite from the choices Blue, Red, Green and Yellow.
Their answers are shown below:

Blue Green Blue Red Yellow Blue
Blue Red Green Yellow Blue Red Blue
Blue Blue Green Blue Yellow Red Red

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

(2 marks)

Select the correct answer.

☐

Colour	Tally	Frequency
Blue	𝍸 𝍸 ||	12
Red	𝍸 ||	7
Green	||||	4
Yellow	||||	4

☐

Colour	Tally	Frequency
Blue	𝍸 ||||	9
Red	𝍸	5
Green	|||	3
Yellow	|||	3

☐

Colour	Tally	Frequency
Blue	𝍸 𝍸	10
Red	𝍸 |	6
Green	||||	4
Yellow	||||	4

☐

Colour	Tally	Frequency
Blue	𝍸 𝍸 |	11
Red	𝍸 |	6
Green	|||	3
Yellow	|||	3

(b) Find the number of students in Ms. Rivera's art class.

(1 mark)

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

(c) Compare the amount of students who chose Blue with the amount who chose Yellow.

(1 mark)

*Select **one** box.*

- ☐ The students selected Yellow and Blue in equal amounts.
- ☐ More students selected Yellow than Blue.
- ☐ More students selected Blue than Yellow.

(d) Ms. Rivera 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 quantitative.
- ☐ The data is qualitative.
- ☐ The data is numeric.
- ☐ This will give the most popular.

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

(1 mark)

*Select **one** box.*

- ☐ A tally chart shows more data.
- ☐ It is more reliable.
- ☐ There is more data.
- ☐ The data is more organised.

2 The table shows the results of a survey into Reading Habits.
For example, 95% of all respondents said "Read books" but only 92% of 16-24 year olds said this.
Some respondents participated in more than one option.

Reading Habit	All People %	Gender		Age			
		Male	Female	18-26	27-36	37-56	57+
Read books	95%	94%	96%	92%	94%	96%	98%
Read newspapers	21%	20%	22%	4%	8%	28%	44%
Read online articles	54%	56%	52%	62%	58%	52%	44%
Read magazines	14%	11%	17%	13%	16%	17%	10%
Listened to audiobooks	12%	11%	13%	14%	13%	11%	10%
Visited the library	91%	90%	92%	84%	89%	96%	95%
Read e-books	37%	33%	41%	51%	45%	29%	23%
Subscribed to book services	7%	6%	8%	10%	8%	6%	4%
None of the above	6%	7%	5%	9%	7%	5%	3%

(a) A person who selected **Visited the library** 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.

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

(b) Describe the relationship in the data between reading e-books and age.

(1 mark)

*Select **one** box.*

- [ ] The proportion of people reading e-books decreases with age.
- [ ] The proportion of people reading e-books is highest at 57+.
- [ ] The proportion of people reading e-books is the same at all ages.
- [ ] The proportion of people reading e-books increases with age.

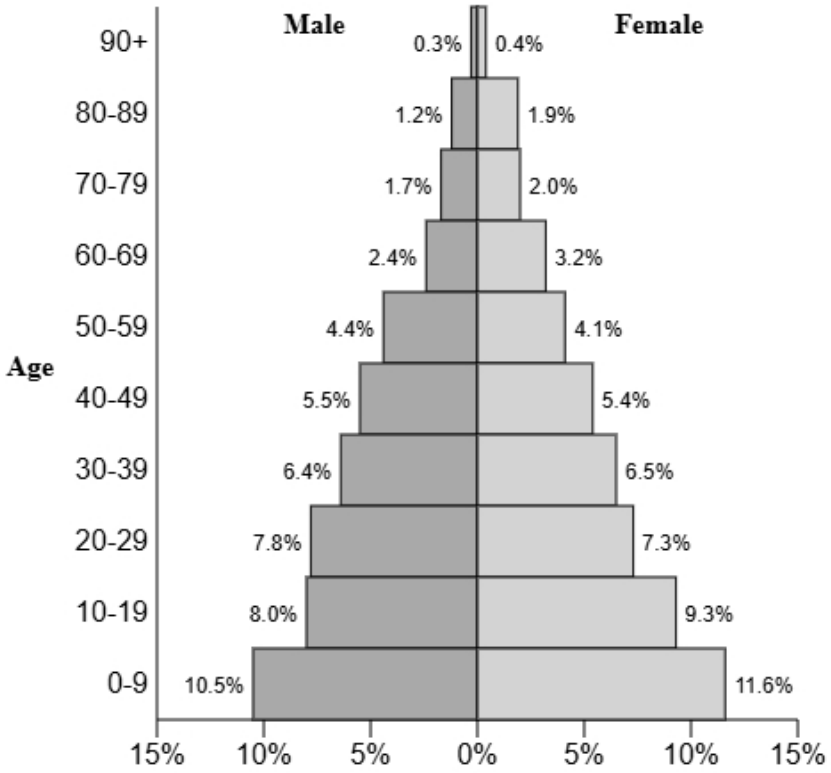
(c) Emily is starting a new library.
She suspects that over half of her customers will be female.
Explain how she used the information from the table to come to this conclusion.

(1 mark)

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

- ☐ but the percentage for males was lower at 92%
- ☐ but the percentage for males was lower at 90%
- ☐ The percentage of females who visited the library was 90%
- ☐ The percentage of females who visited the library was 92%

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



(a) Write down the percentage of females in the age group 20-29.

(1 mark)

Find 20-29 on the population pyramid and read off the number on the right-hand side (females).

_____ %

(b) Find the age group for males that has 6.4% of the population.

(1 mark)

*Select **one** box.*

- ☐ 60-69
- ☐ 30-39
- ☐ 40-49
- ☐ 50-59

(c) Find the age group that has 17.3% of the population.

(1 mark)

*Select **one** box.*

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

(d) Compare the percentage of the population aged 50-79 between males and females.

(1 mark)

*Select **one** box.*

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

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

(1 mark)

*Select **one** box.*

- ☐ Some of the epopulation may not have been counted.
- ☐ They are percentages not amounts.
- ☐ The figures are wrong.
- ☐ The figures have been rounded.

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

Ticket price	£5	£10	£20	£35	Total
Amount sold	290	120	75	15	500

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

Ticket price	£5	£10	£20	£35
Predicted proportion	58%	24%	15%	3%

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

(1 mark)

*Select **one** box.*

- ☐ Use her common sense.
- ☐ Only collect data on one of the days.
- ☐ Draw a pie chart from the data.
- ☐ Record data over more days.

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

Ticket price	£5	£10	£20	£35
True proportion	51%	21%	19%	9%

Compare the sales of £35 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).*

- ☐ This could have been because at an opening weekend people may be happier to spend more on a ticket.
- ☐ There was a higher proportion of £35 tickets at the new location.
- ☐ There was a higher amount of £35 tickets at the new location.
- ☐ This could have been because the new location was more popular.

5 Emily owns a food truck that sells bagels.
Customers can choose either ham or cheese bagel **and** either a coffee or a hot chocolate drink.
She records her sales in a two-way table, but the table is incomplete.

	coffee	hot chocolate	Total
ham bagel		45	87
cheese bagel	33	38	
Total	75		158

(a) Complete the two way table.

(2 marks)

	coffee	hot chocolate	Total
ham		45	87
cheese	33	38	
Total	75		158

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

i) ordered a ham bagel.

ii) ordered a cheese bagel and a coffee.

iii) did **not** order a coffee.

(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** hot chocolate, which means they must have picked coffee.

Probability of ham bagel.

Probability of cheese bagel and a coffee.

Probability of **not** a coffee.

(c) Emily is ordering drinks for the following month.
Explain, with a reason, if Emily should order more bottles of coffee than hot chocolate.

(2 marks)

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

- ☐ because she sold less coffee than hot chocolate.
- ☐ Yes, Emily should buy more bottles of hot chocolate than coffee
- ☐ No, Emily should buy less bottles of coffee than hot chocolate
- ☐ because she sold more hot chocolate than coffee.

6 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.

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

(2 marks)

The outlier is _____ hours

*Select **one** box.*

- ☐ This value 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.

(c) 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).*

- ☐ 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.
- ☐ The mean is less.
- ☐ The mean is greater.

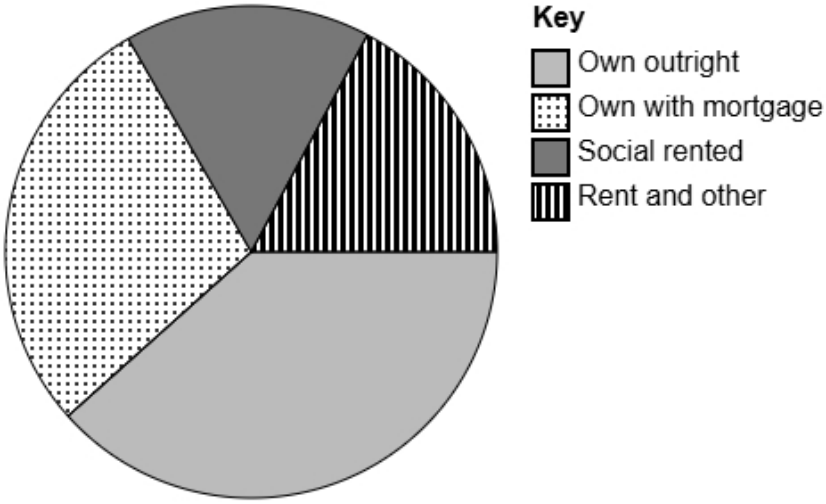
(d) After calculating the mean of the nine values without the outlier, Liam uses this mean in a report to describe all the employees in the UK.
Describe two things that could affect the reliability of her conclusions.

(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.

8 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.

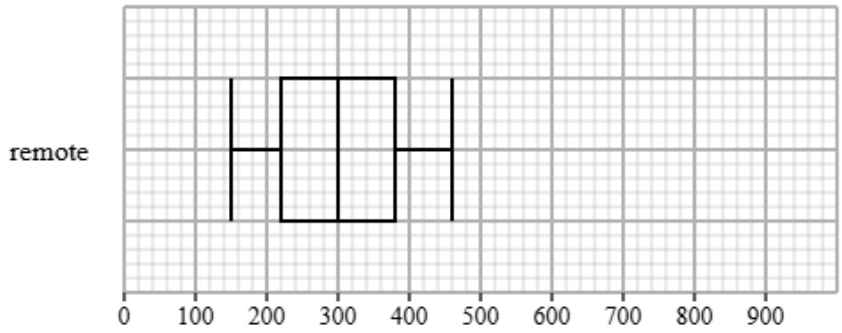
(b) In 2021 there was an estimated 130 thousand households.
Calculate an estimate for the number of people in the UK in 2021 who's highest level of qualification was 'Own outright'.
Round your answer to the nearest thousand.

(2 marks)

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

_____ thousand

9 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.

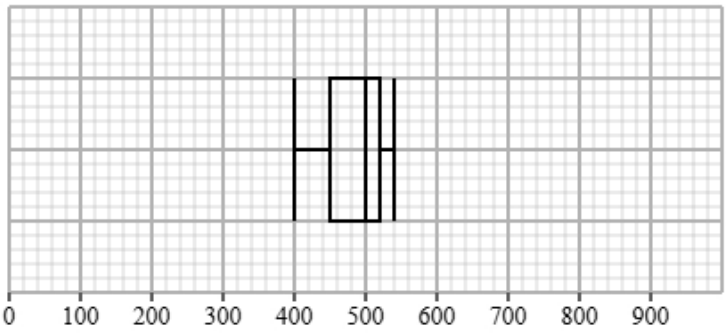
Least tall	Lower quartile	Median	Upper quartile	Most tall
300	550	800	850	900

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

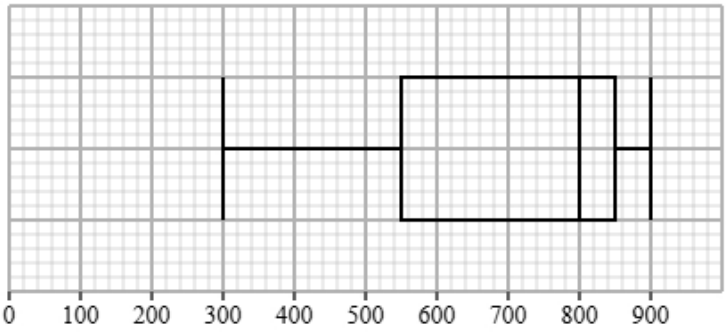
(2 marks)

Select the correct answer.

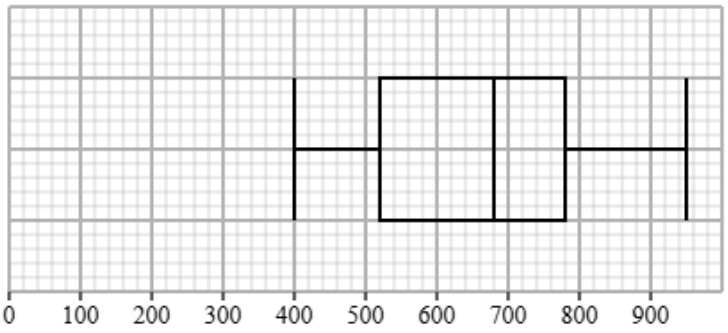
☐ on-site



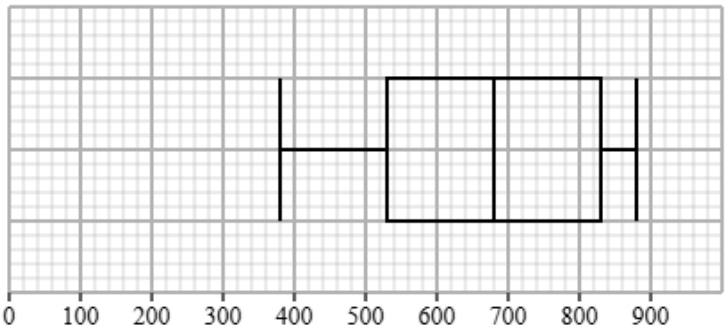
☐ on-site



☐ on-site



☐ on-site



(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.

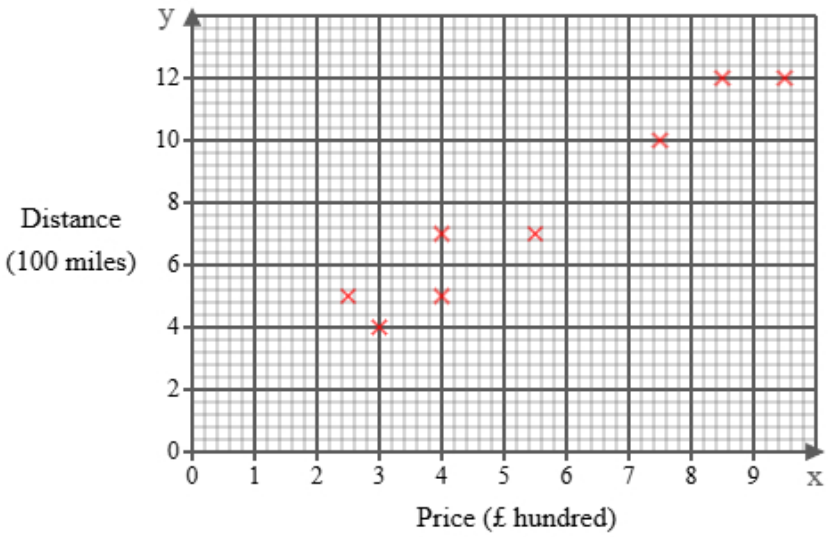
*Select **one** box.*

- ☐ 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.

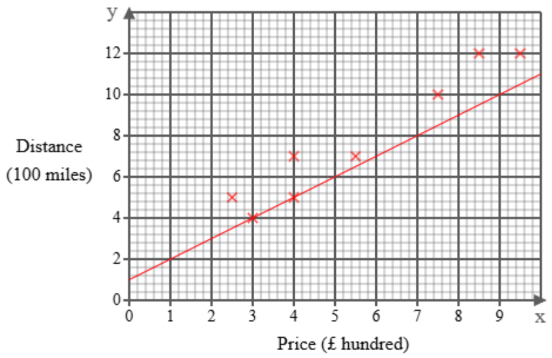
10 Jamie collected data on 11 flights, recording the distance (in hundreds of miles) and the price (in hundreds of pounds) of each flight. He represented his findings in the scatter diagram below.

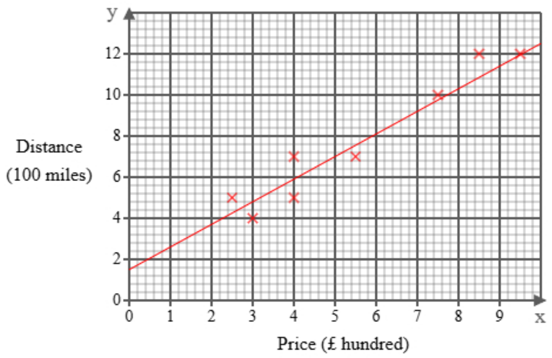


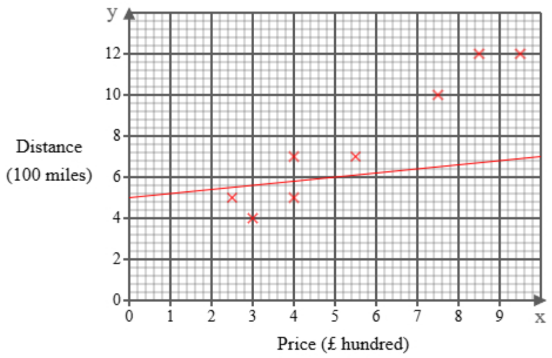
(a) Draw a line of best fit on the scatter diagram.

(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
- ☐ The correlation is negative and
- ☐ weak
- ☐ 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.

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

11 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

Time (t ms)	Frequency
$170 < t \leqslant 200$	0
$200 < t \leqslant 230$	31
$230 < t \leqslant 260$	67
$260 < t \leqslant 290$	22
$290 < t \leqslant 210$	0

Table B

Time (t ms)	Frequency
$200 < t \leqslant 220$	14
$220 < t \leqslant 240$	40
$240 < t \leqslant 260$	44
$260 < t \leqslant 280$	20
$280 < t \leqslant 300$	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) 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.

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

He decides to use Table B.

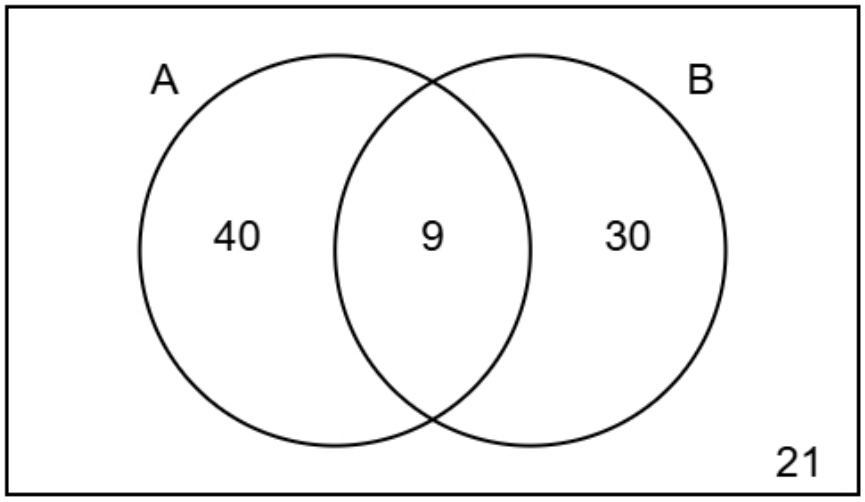
Calculate the average reaction times of the 120 people taking a driving test, giving your answer to 1 decimal place.

(3 marks)

Add midpoint and ft columns onto the table.
Then find the sums of the f and ft.
Find the mean by $\frac{\sum ft}{\sum f}$

_____ ms

12 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)

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

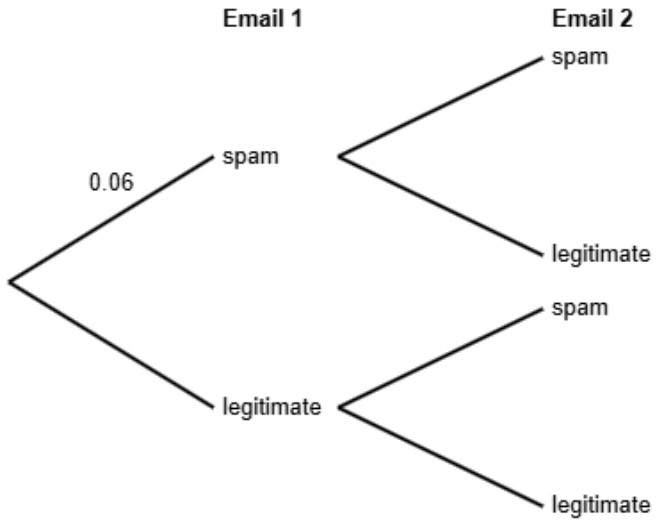
$$P(B) = \frac{\text{amount in B}}{\text{total amount}}$$

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

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

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

13 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)

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 Maria's emails are legitimate.

(2 marks)

You will need to find P(legitimate) AND P(legitimate).
Remember, AND means × in probability.

(c) Maria states that the probability that exactly one email is spam is less than 12%
Find out whether or not Maria is correct.

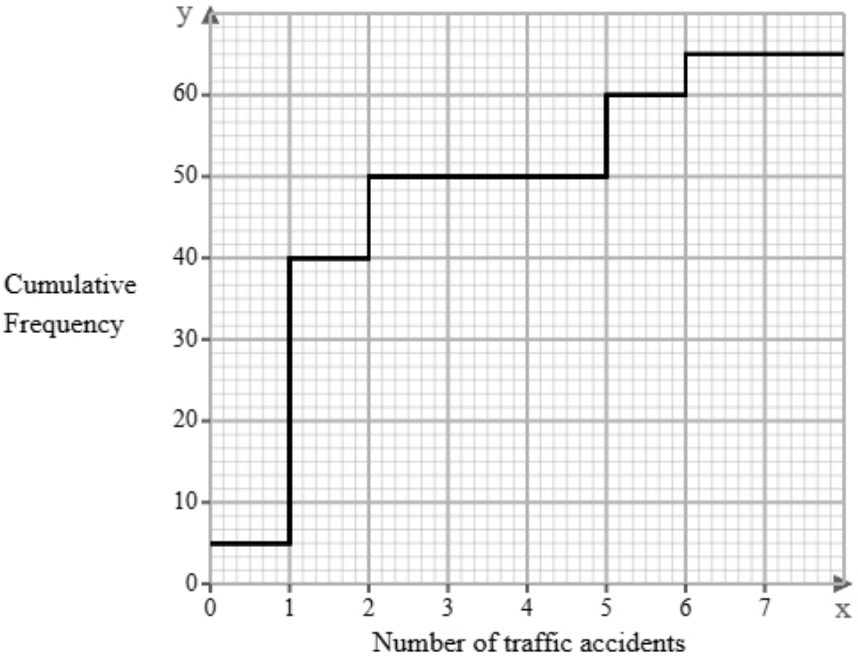
(3 marks)

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

*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.

14 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 number of days where there were:
i) exactly 4 traffic accidents.
ii) more than 4 traffic accidents.

(3 marks)

The frequency is shown by how much the graph 'goes up' at each point.
Remember, the overall frequency is 65.

i) Exactly 4 traffic accidents: _____

ii) More than 4 traffic accidents: _____

(c) In 60 days fewer than x traffic accidents were reported.
Find the value of x

(1 mark)

Draw a line across from 60 on the graph and see where all the 'jumps up' to this line are under.

_____

(d) 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.