statsGCSE

Scaffolded version

Statistics GCSE

Paper 2

2025

Edexcel Foundation

Variant 3 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 Mrs Amina asked the children in her youth club which sport is their favourite from the choices Football, Basketball, Tennis and Swimming.
Their answers are shown below:

Basketball Tennis Football Basketball Swimming Football
Football Basketball Tennis Swimming Football Basketball Football
Football Football Tennis Football Swimming Basketball Basketball

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

(2 marks)

Select the correct answer.

☐

Sport	Tally	Frequency
Football	𝍸 𝍸 |	11
Basketball	𝍸 |||	8
Tennis	||||	4
Swimming	||||	4

☐

Sport	Tally	Frequency
Football	𝍸 ||||	9
Basketball	𝍸 ||	7
Tennis	|||	3
Swimming	|||	3

☐

Sport	Tally	Frequency
Football	𝍸 |||	8
Basketball	𝍸 |	6
Tennis	|||	3
Swimming	|||	3

☐

Sport	Tally	Frequency
Football	𝍸 ||||	9
Basketball	𝍸 ||	7
Tennis	||||	4
Swimming	||||	4

(b) Find the number of children in Mrs Amina's youth club.

(1 mark)

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

(c) Compare the amount of children who chose Football with the amount who chose Swimming.

(1 mark)

*Select **one** box.*

- ☐ The children selected Swimming and Football in equal amounts.
- ☐ More children selected Football than Swimming.
- ☐ More children selected Swimming than Football.

(d) Mrs Amina 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 not numeric.
- ☐ The data is quantitative.
- ☐ The data is numeric.
- ☐ 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 is quicker to display the data.
- ☐ It is more reliable.
- ☐ It improve the accuracy of the data.
- ☐ It is easier to analyse the data.

2 The table shows the results of a survey into Internet Usage.
For example, 95% of all respondents said "Used internet daily" but only 99% of 16-24 year olds said this.
Some respondents participated in more than one option.

Internet Usage	All People %	Gender		Age			
		Male	Female	18-26	27-36	37-56	57+
Used internet daily	95%	94%	96%	99%	97%	95%	89%
Shopped online	76%	72%	80%	69%	81%	82%	72%
Used social media	94%	92%	96%	99%	98%	95%	84%
Watched streaming services	98%	99%	97%	99%	98%	98%	96%
Worked remotely	57%	58%	56%	79%	62%	49%	38%
Used online banking	97%	96%	98%	99%	98%	96%	95%
Played online games	17%	23%	11%	29%	24%	11%	4%
Attended virtual events	54%	56%	52%	67%	58%	47%	44%
None of the above	1%	2%	1%	2%	2%	1%	1%

(a) A person who selected **Used online banking** 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 attending virtual events and age.

(1 mark)

*Select **one** box.*

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

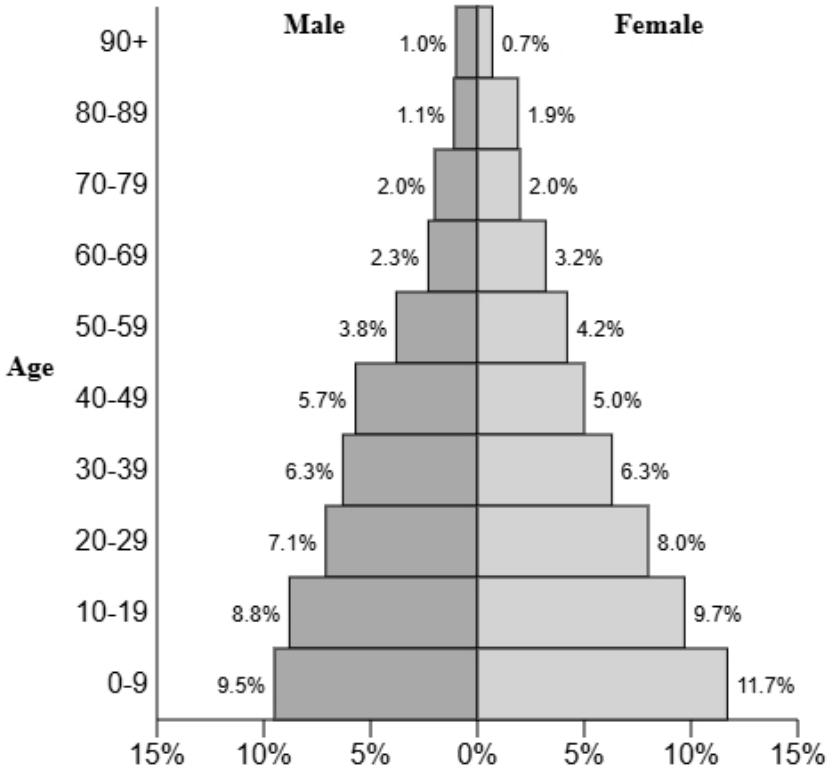
(c) Isla is starting a monthly virtual event.
She suspects that over half of her customers will be male.
Explain how she used the information from the table to come to this conclusion.

(1 mark)

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

- ☐ but the percentage for males was lower at 52%
- ☐ The percentage of females who attended virtual events was 56%
- ☐ The percentage of females who attended virtual events was 52%
- ☐ but the percentage for males was higher at 56%

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



(a) Write down the percentage of females in the age group 10-19.

(1 mark)

Find 10-19 on the population pyramid and read off the number on the right-hand side (females).

_____ %

(b) Find the age group for males that has 7.1% of the population.

(1 mark)

*Select **one** box.*

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

(c) Find the age group that has 12.6% of the population.

(1 mark)

*Select **one** box.*

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

(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 100.3% and not 100%.

(1 mark)

*Select **one** box.*

- ☐ Some of the epopulation may not have been counted.
- ☐ The figures have been rounded.
- ☐ The figures are wrong.
- ☐ There are gaps in the ages.

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

Ticket price	£2	£3	£5	£10	Total
Amount sold	68	56	42	34	200

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

Ticket price	£2	£3	£5	£10
Predicted proportion	34%	28%	21%	17%

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

(1 mark)

*Select **one** box.*

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

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

Ticket price	£2	£3	£5	£10
True proportion	18%	25%	26%	31%

Compare the sales of £10 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 £10 tickets at the new location.
- ☐ There was a higher amount of £10 tickets at the new location.
- ☐ This could have been because at an opening weekend people may be happier to spend more on a ticket.
- ☐ This could have been because the new location was more popular.

5 Noah owns a food truck that sells wraps.
Customers can choose either chicken or veggie wrap **and** either a apple juice or a water drink.
He records his sales in a two-way table, but the table is incomplete.

	apple juice	water	Total
chicken wrap		38	87
veggie wrap	43	42	
Total	92		172

(a) Complete the two way table.

(2 marks)

	apple juice	water	Total
chicken		38	87
veggie	43	42	
Total	92		172

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

i) ordered a chicken wrap.

ii) ordered a veggie wrap and a apple juice.

iii) did **not** order a apple juice.

(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** water, which means they must have picked apple juice.

Probability of chicken wrap.

Probability of veggie wrap and a apple juice.

Probability of **not** a apple juice.

(c) Noah is ordering drinks for the following month.
Explain, with a reason, if Noah should order more bottles of apple juice than water.

(2 marks)

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

- ☐ Yes, Noah should buy more bottles of apple juice than water
- ☐ because he sold more apple juice than water.
- ☐ No, Noah should buy less bottles of water than apple juice
- ☐ because he sold less water than apple juice.

6 A sports club is thinking about upgrading its facilities.
Jack is carrying out a survey to see what all club members think about the improvements.
Jack thinks that he should take a sample rather than a census.

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

(2 marks)

*Select **two** boxes.*

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

(b) Jack has decided to use the register of members as a sampling frame.
Explain what a sampling frame is.

(1 mark)

*Select **one** box.*

- ☐ The tally chart or table used in the survey.
- ☐ The whole group.
- ☐ A list of all the members in the sample.
- ☐ A list of all the members in the population.

(c) Jack has decided to use the register of members as a sampling frame.
State one problem Jack may have using the register of members as a sampling frame.

(1 mark)

*Select **one** box.*

- ☐ The class registers would also include dates of birth.
- ☐ The staff will not be included.
- ☐ There will be too many names.
- ☐ Bias.

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

(2 marks)

*Select **two** boxes.*

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

7 James is a transport analyst working for a London borough council. He is studying commute times of residents using public transport.

He takes a simple random sample of 8 residents and asks them how long their commute took the previous day.

The commute time (hours) of the 8 people are listed:

42	39	39	120
38	43	39	40

James believes that one of the values is an outlier.

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

(1 mark)

*Select **one** box.*

- ☐ Each individual in the population has an equal chance of being chosen.
- ☐ The researcher selects individuals based on their judgment of who best represents the population.
- ☐ The sample is created by selecting every nth individual from a list.
- ☐ The same individuals can be chosen multiple times to ensure fairness.

(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.*

- ☐ The value is far higher than the other pieces of data.
- ☐ The value is located at the lower end of the data.
- ☐ This value occurs less than the others.
- ☐ This value is lower than the mean.

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

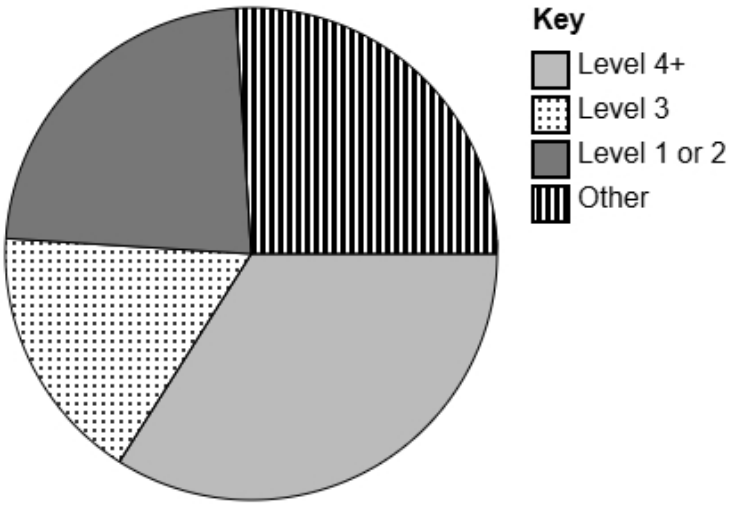
(d) After calculating the mean of the seven values without the outlier, James uses this mean in a report to describe all the residents in the borough.
Describe two things that could affect the reliability of her conclusions.

(2 marks)

*Select **two** boxes.*

- ☐ The mean will not measure the range of the data.
- ☐ One value cannot represent many.
- ☐ Does not include all of the data.
- ☐ Small sample size.
- ☐ The commute time may not truly reflect the difficulties that residents have in getting to work.

8 The accurately drawn pie chart shows information about the highest level of qualification in England and Wales in 2021.



(a) Explain how you can tell that most people's highest level of qualification was level 4+ in England and Wales in 2021 using the pie chart.

(1 mark)

*Select **one** box.*

☐ 'Level 4+' is at the bottom of the pie chart.

☐ 'Level 4+' is the first value in the key.

☐ 'Level 4+' is the most positive response.

☐ 'Level 4+' has the largest sector.

(b) The population in the England and Wales in 2021 was estimated to be 60 million.
Calculate an estimate for the number of people in the UK in 2021 who's highest level of qualification was 'Level 4+'.
Round your answer to the nearest million.

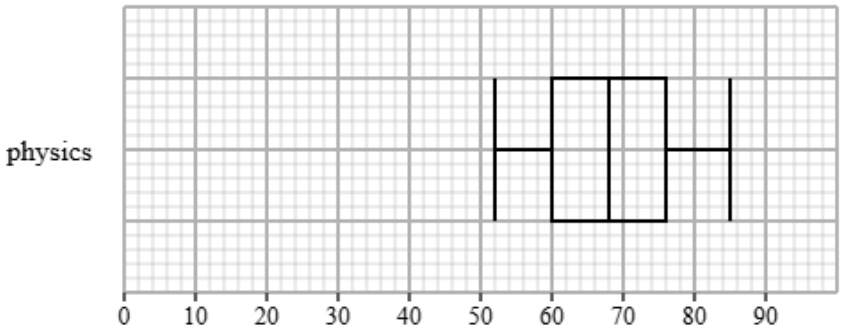
(2 marks)

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

_____ million

9 Olivia collected the marks for physics and business students from a statistics exam. Both groups took the same exam.

The box plot presents data on the marks for the physics students.



The table gives information about the marks for the business students.

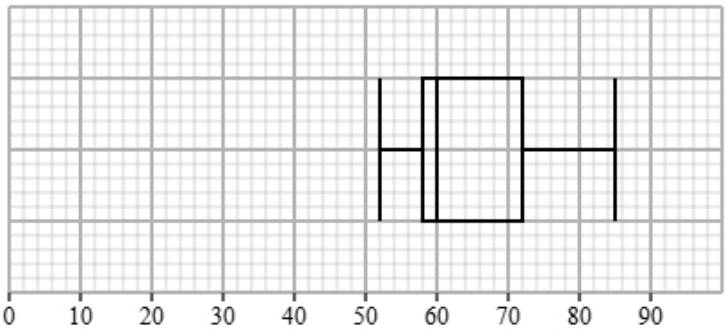
Least tall	Lower quartile	Median	Upper quartile	Most tall
52	58	60	72	85

(a) Draw a box plot for the marks for the business students.

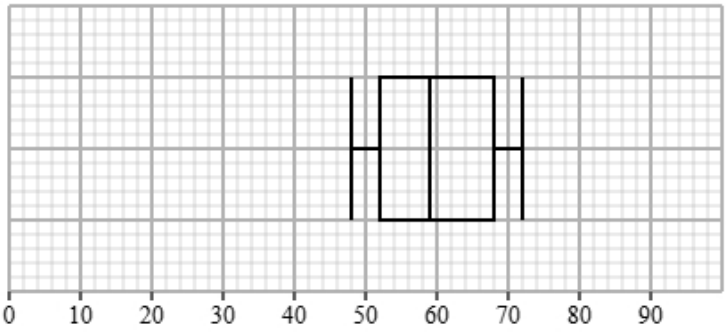
(2 marks)

Select the correct answer.

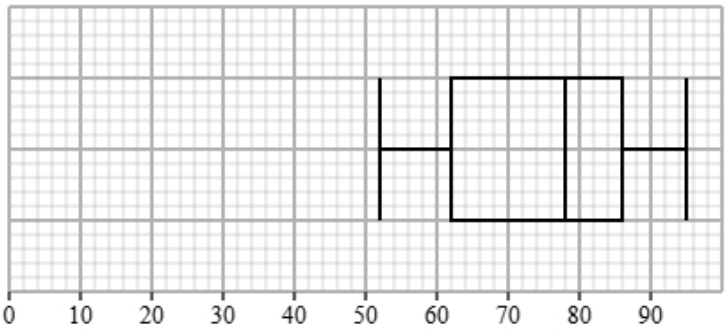
☐ business



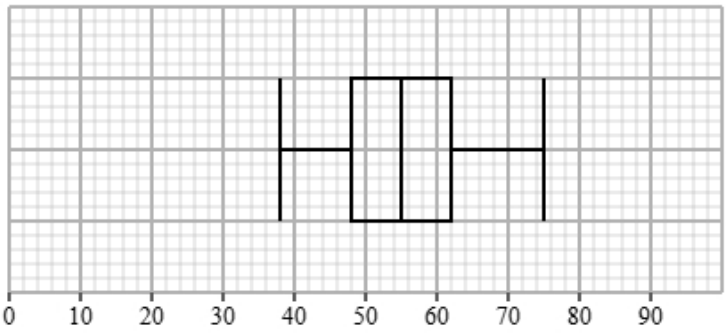
☐ business



☐ business



☐ business



(b) Compare the two distributions of marks.
Give three comparisons and interpret one of these comparisons.

(4 marks)

*Select **one** box.*

- ☐ The median is bigger.
- ☐ The median marks for physics students is greater than business students.
- ☐ The median marks for physics and business students are equal.
- ☐ The median marks for physics students is lower than business students.

*Select **one** box.*

- ☐ The IQR is bigger.
- ☐ The IQR for the marks of the physics and business students are equal.
- ☐ The IQR for the marks of the physics students is greater than business students.
- ☐ The IQR for the marks of the physics students is lower than business students.

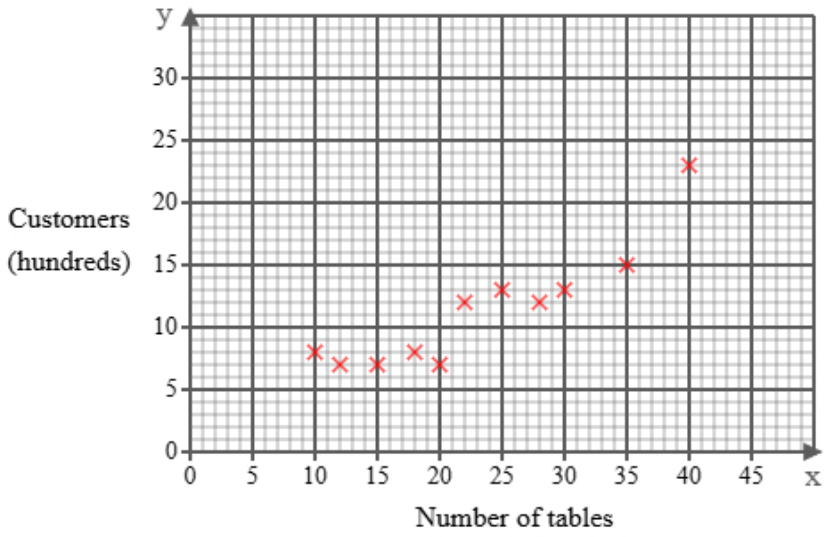
*Select **one** box.*

- ☐ The skews for the marks of the physics and business students are both positive.
- ☐ The skews for the marks of the physics and business students are both symmetrical.
- ☐ The skew for the marks of the physics students is symmetrical and the skew for the business students is negative.
- ☐ The skew for the marks of the physics students is symmetrical and the skew for the business students is positive.

*Select **one** box.*

- ☐ The marks for the physics students are less spread out than the business students.
- ☐ The physics students are more skewed than business students.
- ☐ The physics students are on average did better on the statistics test than the business students.
- ☐ The physics students are on average did worse on the statistics test than the business students.

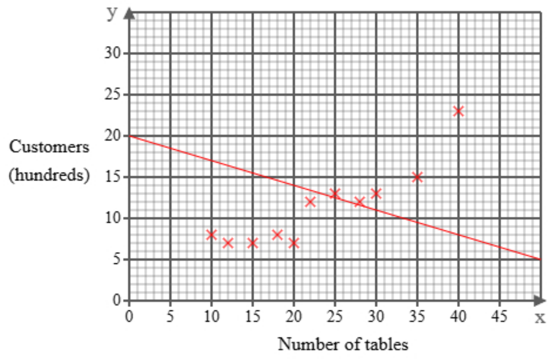
10 Noah collected data on 11 restaurants, recording their number of tables and the average number of customers they serve per week (in hundreds). He represented his findings in the scatter diagram below.

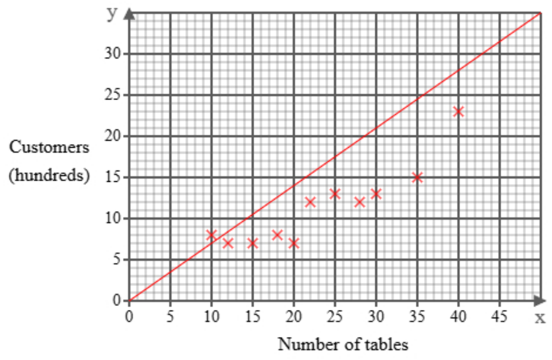


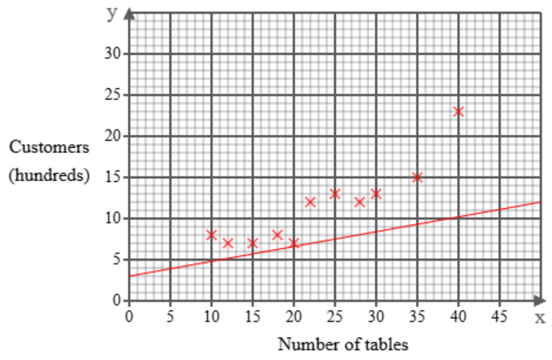
(a) Draw a line of best fit on the scatter diagram.

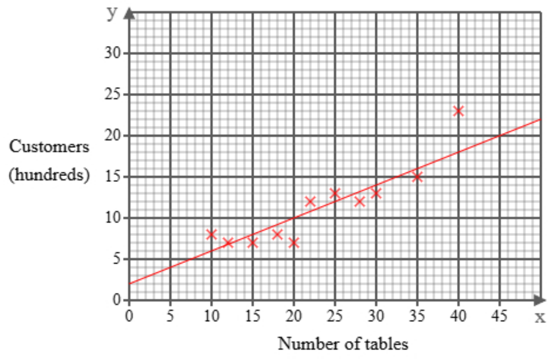
(1 mark)

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

*Select **one** box.*

- ☐ A restaurant that has lots of tables will have a high number of customers.
- ☐ A restaurant that has lots of tables will have a low number of customers.
- ☐ As the tables increases the number of customers increases.
- ☐ As the tables increases the number of customers decreases.

(c) A new restaurant will be opening soon with 75 tables.
Noah is planning on using the line of best fit on the scatter diagram to predict the average number of customers they serve per week.

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

11 Emma investigates the monthly electricity bills (in pounds) of 250 households.
The charges range from £38 to £112.
Emma considers using one of the two possible grouped frequency tables for the results, Table A or Table B, shown below.

Table A

Charge (£ p)	Frequency
$0 < p \leqslant 30$	0
$30 < p \leqslant 60$	45
$60 < p \leqslant 90$	169
$90 < p \leqslant 120$	36
$120 < p \leqslant 150$	0

Table B

Charge (£ p)	Frequency
$30 < p \leqslant 50$	18
$50 < p \leqslant 70$	79
$70 < p \leqslant 90$	116
$90 < p \leqslant 110$	33
$110 < p \leqslant 130$	4

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

(2 marks)

*Select **two** boxes.*

- ☐ Grouped data helps avoid using averages.
- ☐ Grouped data is easier to read.
- ☐ Grouped data provides a more accurate reflection of trends.
- ☐ Grouped data can help to spot patterns in the data.
- ☐ Grouped data eliminates the need to use any statistical methods.

(b) Emma feels that Table B gives more detail than Table A about the results.
Assess the appropriateness of Emma's claim.

(2 marks)

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

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

(c) Emma wants to work out the average monthly electricity bills of the 250 households.

She decides to use Table B.

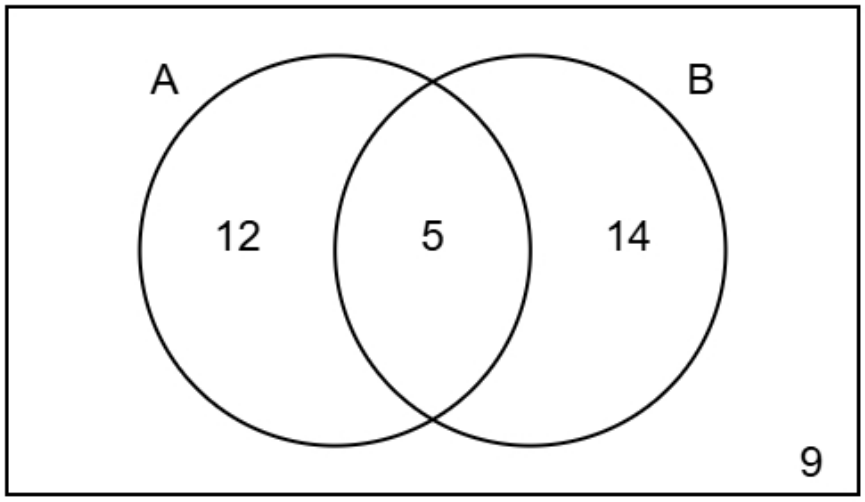
Calculate the average monthly electricity bills of the 250 households, giving your answer to 1 decimal place.

(3 marks)

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

_____ pounds

12 The Venn diagram shows information about 40 cars sold in a dealership in 2022.
A is the event that the car is an electric vehicle.
B is the event that the car is priced above £30,000.
The numbers in the Venn diagram indicate the number of cars.



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

(1 mark)

*Select **one** box.*

- ☐ The number of cars sold in a dealership but are **not** priced above £30,000
- ☐ The number of cars sold in a dealership **and** are priced above £30,000
- ☐ The number of cars sold in a dealership **or** are priced above £30,000
- ☐ The number of cars that are **not** electric **or** priced above £30,000

(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):

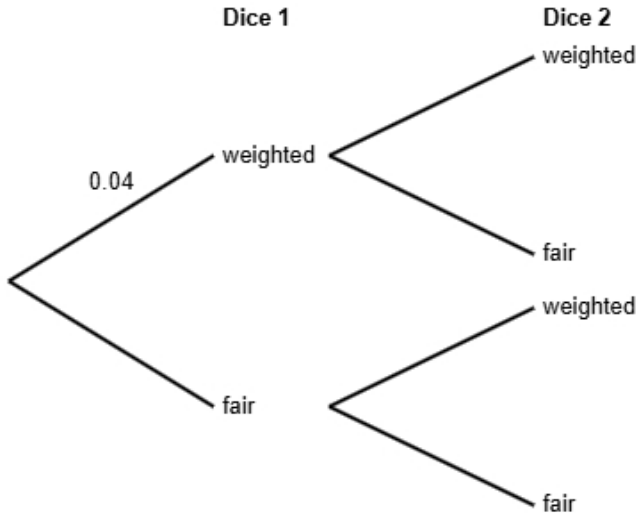
$$P(B) = \frac{\text{amount in B}}{\text{total amount}}$$

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

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

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

13 A company found that 4% of its dice are slightly weighted and do not roll fairly.
The rest of the dice are perfectly balanced.
Jack picks two dice from a random batch.
He does not know if each die is weighted or fair.



(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 Jack's dice are fair.

(2 marks)

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

(c) Jack states that the probability that exactly one dice is weighted is less than 8%
Find out whether or not Jack is correct.

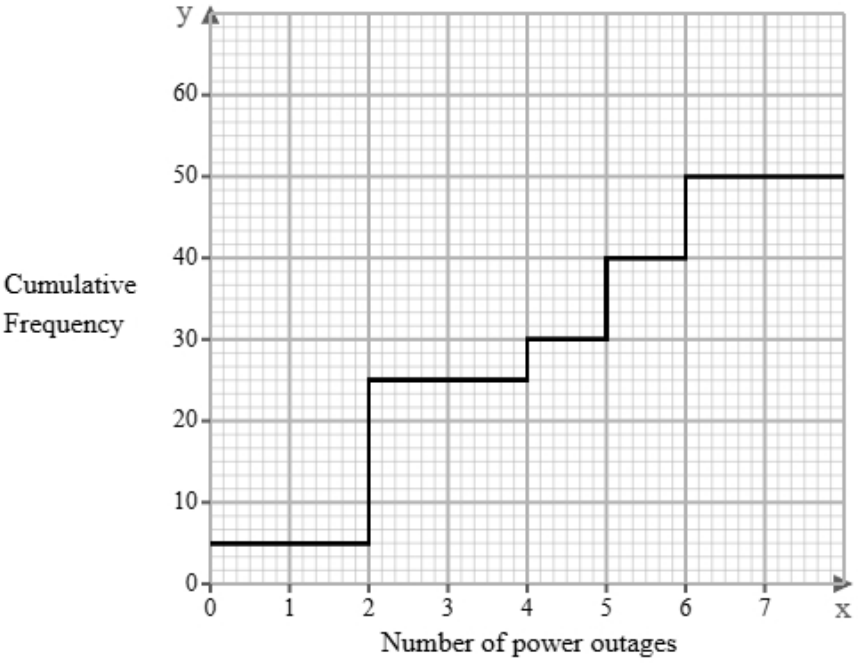
(3 marks)

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

*Select **one** box.*

- ☐ The probability that exactly one dice is weighted is more than 8%, so Jack is not correct.
- ☐ The probability that exactly one dice is weighted is less than 8%, so Jack is not correct.
- ☐ The probability that exactly one dice is weighted is more than 8%, so Jack is correct.
- ☐ The probability that exactly one dice is weighted is less than 8%, so Jack is correct.

14 The cumulative frequency step polygon shows information about number of power outages recorded in a small town over 50 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 power outages recorded in a small town is continuous.

☐ Because number of power outages recorded in a small town is discrete.

☐ Because number of power outages recorded in a small town is quantitative.

☐ Because number of power outages recorded in a small town is qualitative.

(b) Find the number of days where there were:
i) exactly 3 power outages.
ii) more than 3 power outages.

(3 marks)

The frequency is shown by how much the graph 'goes up' at each point.
Remember, the overall frequency is 50.

i) Exactly 3 power outages: _____

ii) More than 3 power outages: _____

(c) In 40 days fewer than x power outages were recorded.
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) Nigel believes the interquartile range of number of power outages is 8.
Explain why the interquartile range for this data cannot be 8.

(1 mark)

*Select **one** box.*

- ☐ 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 6, so the IQR must be less than 6.
- ☐ The range is 7, so the IQR must be more than 7.