

Statistics GCSE

Paper 2

2025

Edexcel Higher

Variant 3

1ST0/1H

# Answers

Visit our website for tutorials on each question.

[www.statsgcse.com](http://www.statsgcse.com)

- 1 At a community sports club, 70% of members are male and 30% are female. Sophie and Marcus plan to conduct a feedback survey.

Sophie decides to use simple random sampling to select 100 participants. She uses the club's registration list as a sampling frame, assigning a number to each participant. She then generates 100 random numbers and selects her sample accordingly.

Marcus decides to use quota sampling to collect a sample of 100 participants. He plans to approach people at the club until 70 male and 30 female have been interviewed.

- (a) Give two reasons why Sophie's method may **not** produce a sample of 100 participants.

(2 marks)

Select *two* boxes.

- The random numbers may not be whole numbers.
- The selected person may not want to participate.
- The method may produce more people than is required.
- The person may have joined.
- Some of the random numbers may be repeated.

- (b) Give **two** advantages of quota sampling.

(2 marks)

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

- Quota sampling is random
- Quota sampling is quick
- and will give more accurate results.
- and allows for comparison between male and female members.

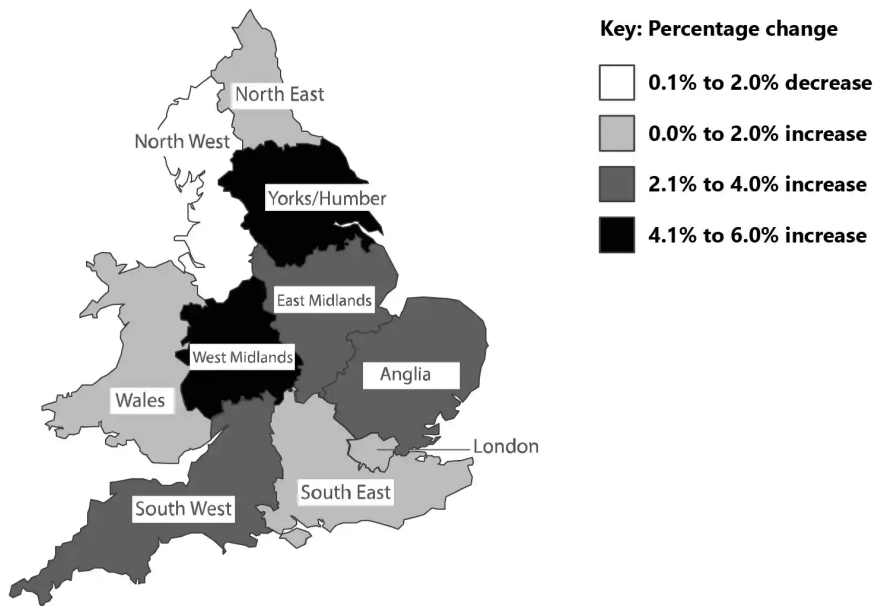
(c) Explain why the quota sample used by Marcus is not a random sample.

(1 mark)

Select **one** box.

- All people needed to answer anonymously.
- The sample was selected from a large population.
- Marcus is picking the people.
- Random sampling must use a random number generator.

2 The map below shows the percentage change in domestic tourist visits across different regions of England and Wales between 2012 and 2013.



(a) Write down the percentage change in domestic tourist visits between 2012 and 2013 in West Midlands.

(1 mark)

Select **one** box.

- 4.1% to 6.0% increase
- 0.0% to 2.0% increase
- 0.1% to 2.0% decrease
- 2.1% to 4.0% increase

(b) There are 10 regions shown.  
Find the number of regions that the domestic tourist visits **decreased**.

(1 mark)

Make sure to check the key carefully.  
Decreases are shown as white, whilst increases are shaded.

(c) Erin states that domestic tourism in England and Wales increased overall between 2012 and 2013.

Explain why this may not be the case.

(1 mark)

Select **one** box.

- The map is not detailed enough.
- More regions decreased than increased.
- The map does not show the increases and decreases as amounts.
- More regions increased than decreased.

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

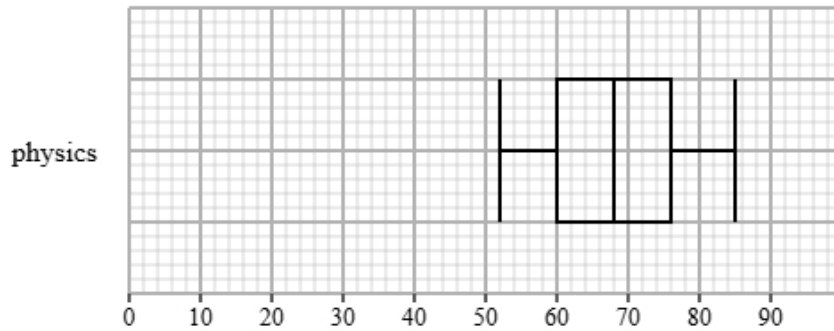
(1 mark)

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

C \_ \_ \_ \_ P L E T H

choropleth

- 3 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.



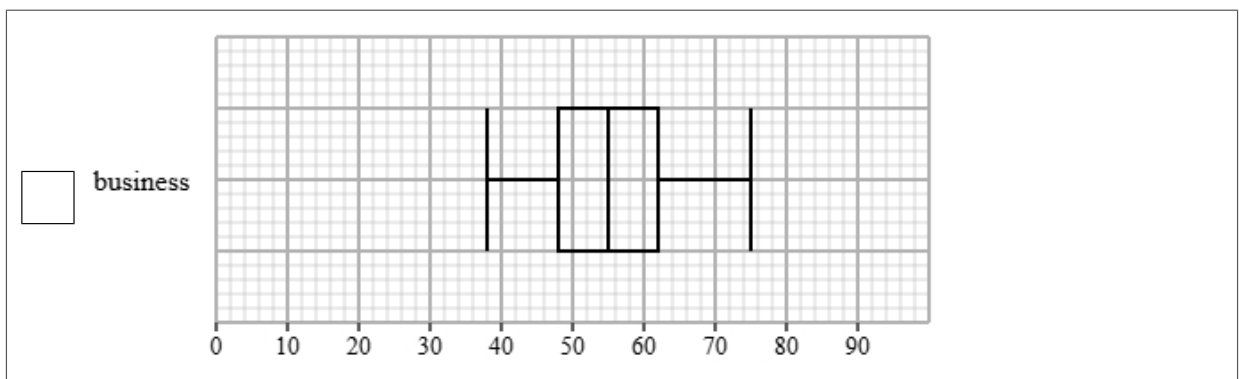
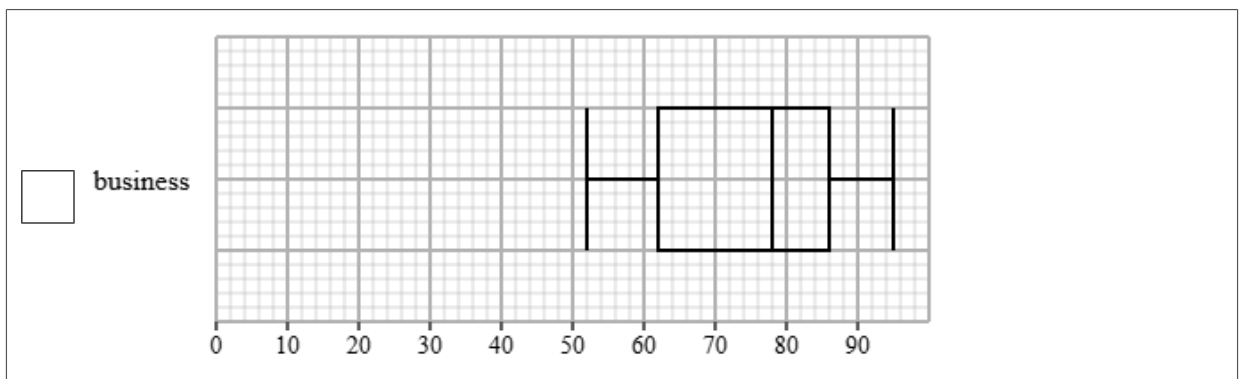
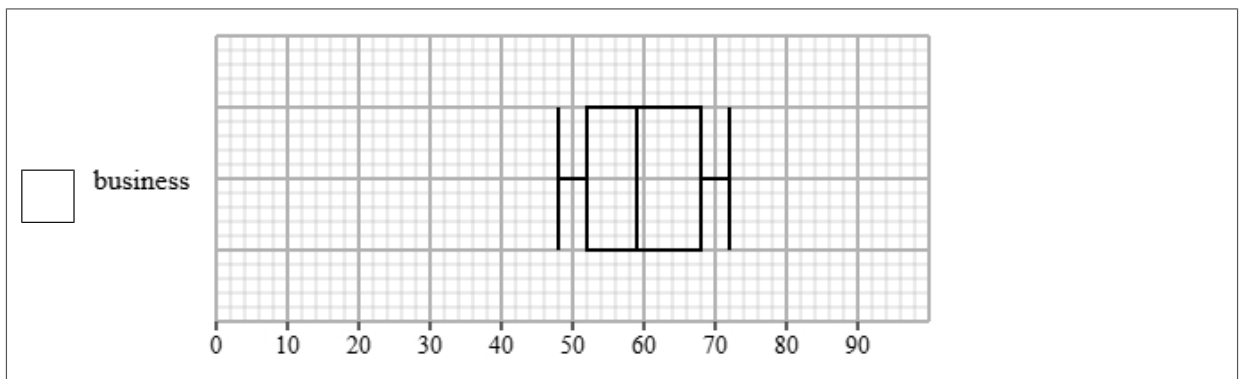
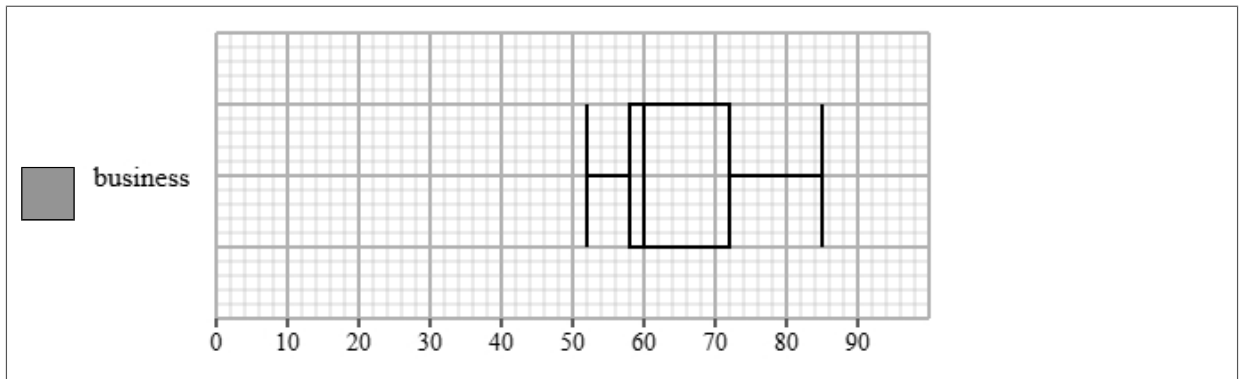
The table gives information about the marks for the business students.

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.



(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.

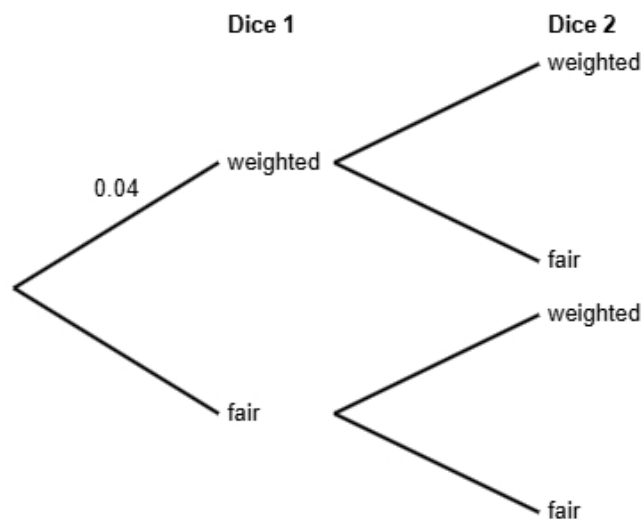
Select **one** box.

- 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.

- 4 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  $P(\text{fair})$  AND  $P(\text{fair})$ .  
Remember, AND means  $\times$  in probability.

0.9216

(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.

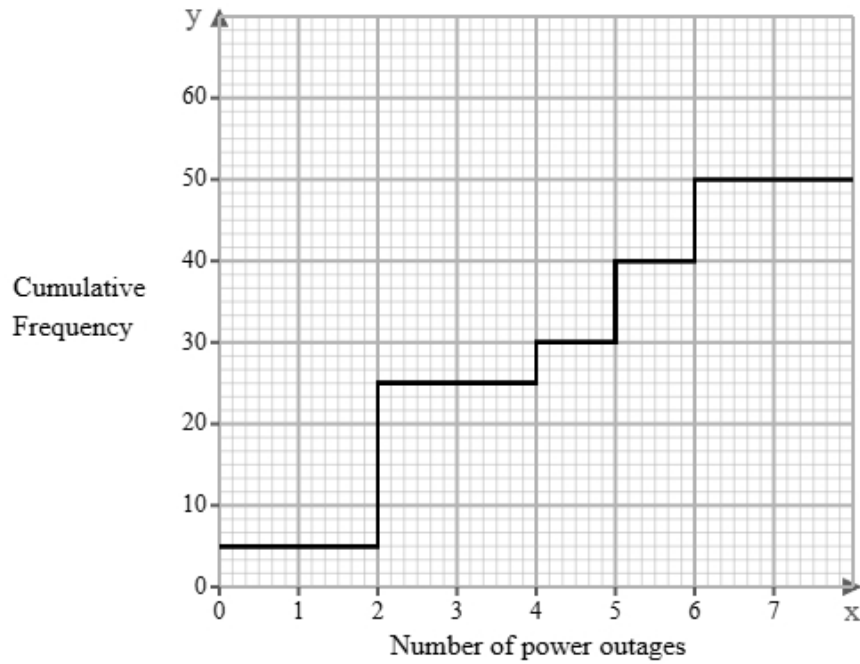
The probability that exactly one dice is weighted is more than 8%, so Jack is not correct.

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.

- 5 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 mode of the number of power outages.

(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.

2

(c) 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: 0

ii) More than 3 power outages: 25

(d) 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.

6

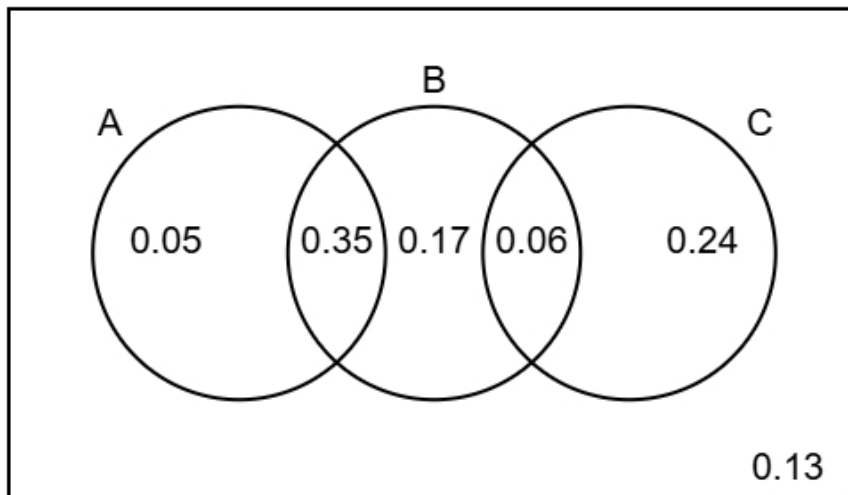
- (e) Nigel believes the interquartile range of number of power outages is 8.  
Explain why the interquartile range for this data cannot be 8.

(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.

6 The Venn diagram illustrates the probabilities associated with events A, B, and C.



(a) Identify the **two** events that are mutually exclusive, giving a reason for your answer. (2 marks)

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

- because they have the highest total probability.
- 1 A and C are mutually exclusive
- A and B are mutually exclusive
- because they only intersect once.
- B and C are mutually exclusive
- 2 because they do not intersect.

(b) Find P(B)

(1 mark)

We are looking for the probabilities inside B.

0.58

(c) Find  $P(A \text{ or } C)$

(2 marks)

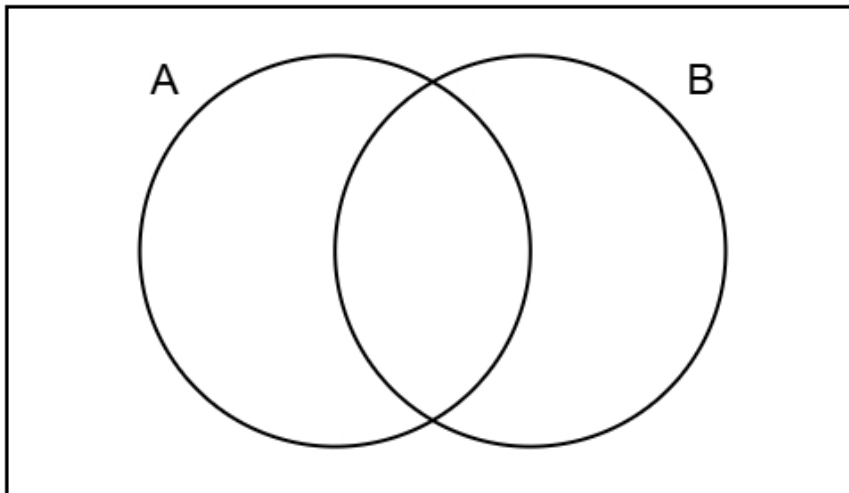
We are looking for the probabilities inside A or C.

0.7

(d) Complete the Venn diagram to show **only** the probabilities for events A and B.

(2 marks)

Combine the probabilities from C into either B or the outside area.



7 Aiden is researching customer ratings and critic ratings for 10 restaurants.

(a) Suggest a diagram that Aiden could draw to determine if there is a relationship between customer ratings and critic ratings for the 10 restaurants.

(1 mark)

Select *one* box.

Line graph

Scatter diagram

Bar chart

Cumulative frequency graph

- (b) i) Calculate Spearman's rank correlation coefficient from the table.  
ii) Interpret your answer in the context of Aiden's research. You will need to refer to the effects of any outliers.

(5 marks)

Complete the d (difference) and d squared columns then use the formula

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

n is the number of rows in the table (given in the question).

Restaurant	Average Customer Rating (out of 5)	Customer Rank	Critic Rank	d	d <sup>2</sup>
Bella's Bistro	4.8	1	3	-2	4
Caffe Vino	2.9	8	7	1	1
Downtown Grill	4.4	3	2	1	1
El Camino	3.2	7	8	-1	1
Fireside Diner	3.7	6	6	0	0
Grand Oak Tavern	4.5	2	5	-3	9
Harbor View Restaurant	4.1	4	4	0	0
Joy's Kitchen	1.8	10	9	1	1
Olive Garden Grill	3.9	5	6	-1	1
The Golden Spoon	2.7	9	10	-1	1

---



---



---



---

Select *two* boxes.

- As customer ratings increases, critic ratings increases.
- The presence of outliers would decrease the correlation.
- As customer ratings increases, critic ratings decreases.
- The presence of outliers would increase the correlation.

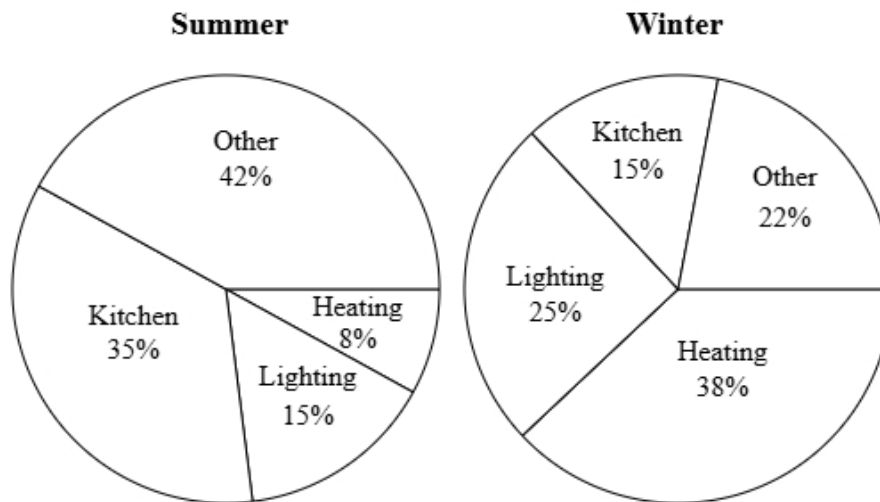
(c) Discuss whether Aiden should have used Pearson's product moment correlation coefficient instead of Spearman's rank correlation coefficient to measure the correlation.

(3 marks)

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

- therefore either would be suitable.
- 1** Aiden should not use PMCC because
- 3** and Spearman's rank correlation is used for ranked data.
- Aiden should use PMCC because
- 2** the PMCC only measures linear correlation
- the PMCC also measures correlation

8 The pie charts show the electricity usage in a house in Summer and Winter.



(a) Explain why the pie charts do not show a decrease in kitchen electricity usage in Winter compared to Summer.

(1 mark)

Select **one** box.

- The pie charts are both the same size, so the total will be equal.
- Pie charts only show positive trends.
- Summer could have had a larger total budget than Winter.
- Pie charts are inaccurate.

(b) The electricity usage in the kitchen in Summer was 224.735 kWh.  
Find the electricity usage used for lighting in Summer.

(2 marks)

The electricity usage in the kitchen in Summer was 224.735 kWh which is shown by 35% on the pie chart.

$$35\% = 224.735 \text{ kWh}$$

Find 1%

Then find the electricity usage used for lighting (15%).

96.315 kWh

- (c) In Summer the total electricity usage was 640 kWh (nearest ten).  
In Winter the total electricity usage was 830 kWh (nearest ten).

Liam wants to use the totals to draw pie charts.

Explain, giving reasons, how Liam can use the totals to draw these pie charts.

(5 marks)

Select **one** box.

- Liam can use comparative pie charts.
- Liam can use a cumulative frequency diagram.
- Liam can use dual pie charts.
- Liam can use flex pie charts.

Select **one** box.

- $\sqrt{\frac{830}{640}}$
- $\left(\frac{830}{640}\right)^2$
- $\frac{830^2}{640^2}$
- $\frac{830}{640}$

The radius of the Summer pie chart will be 1.14 (2 d.p.) larger than the Winter pie chart.

Select **one** box.

- The areas are now scaled inversely to frequency, which enhances clarity.
- Since the pie charts have varying totals, their overall sizes will correspond to those totals while keeping the proportions the same.
- The share of each sector will adjust based on the chart's total.
- This technique is quicker than using conventional pie charts.

9 A company manufactures car tires.

The tires have a target pressure of 32 psi.

The company uses quality assurance to monitor the pressure of each tire.

Samples of the tires are taken from the production line at regular intervals and the mean pressure of the tires in each sample is found.

The sample means should be normally distributed with a mean of 32 psi and a standard deviation of 0.8 psi.

(a) Find the upper action limit for the sample means for the tires.

(2 marks)

Use the formula for the upper action limit

$$\text{Upper action limit} = \mu + 3\sigma$$

\_\_\_\_\_ 34.4 psi

(b) The upper action limit will be set closer to the target pressure of 32 psi.

Describe the effect this will have on the frequency of production process stoppages.

(1 mark)

Select **one** box.

The production process would reset more frequently.

The production process would reset less frequently.

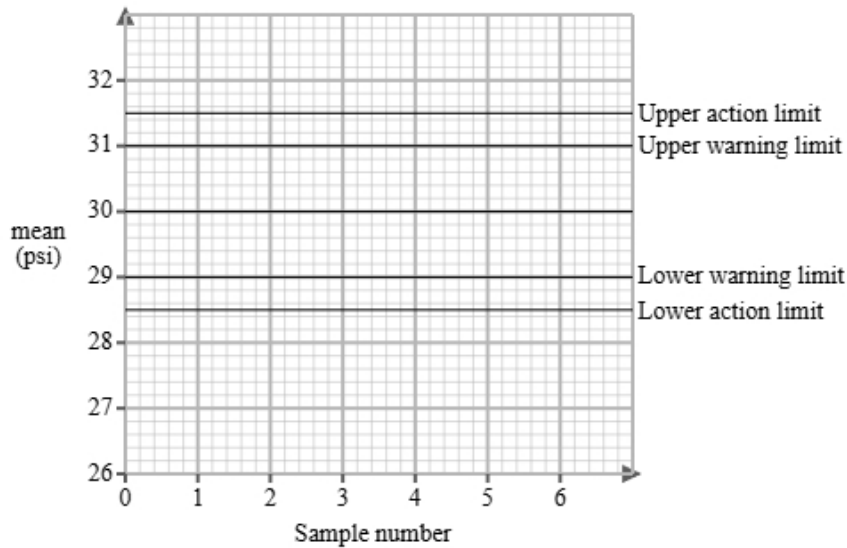
This is unrelated to the production process so will have no effect.

The production process will work faster.

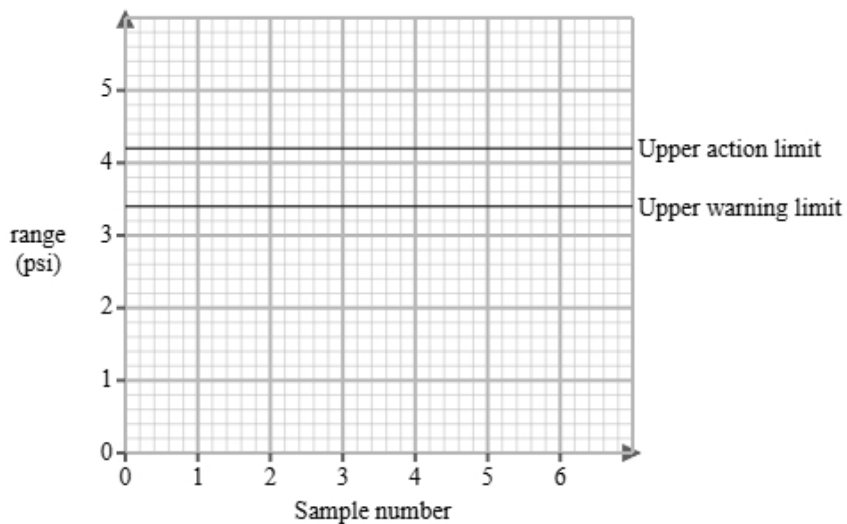
(c) The company also manufactures bike tires and uses quality assurance to monitor the pressure of each tire.

Here are the control charts for the sample means and for the sample ranges of the pressures of the tires.

**Control chart for means**



**Control chart for ranges**



A sample is taken and is found to have a mean of 31.1 psi and a range of 3.2 psi. Use the sample mean and range to determine what action, if any, needs to be taken.

(2 marks)

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

- 2** and the mean is within the upper warning limit.
- 1** The production process should continue because the range is within the warning limit
- although the mean is outside the upper warning limit but not the action limit.
- The production process should be stopped as the range is outside the upper action limit

**10** The heights of a group of ten-year-old girls have a mean of 137.5 cm and a standard deviation of 4.9 cm.

**(a)** Sophie is ten-year-old girl with a standardised score of 0.  
Find Sophie's height.

(1 mark)

A standardized score of 0 indicates that the value is equal to the mean of the distribution.

137.5 cm

**(b)** Amelia and Ruby are both ten-year-old girls in the group.  
Amelia's standardised score for height is 1.1 cm.  
Ruby's standardised score for height is -0.3 cm.  
Amelia is taller than Ruby.  
How much taller is Amelia?

(3 marks)

Rearrange the formula to make the value the subject

$$\text{Standardised score} = \frac{\text{value} - \text{mean}}{\text{standard deviation}}$$

6.86 cm

(c) Fatima takes a sample of 3 ten-year-old girls from the group.

She wants to calculate the standardised score for the sample mean of their ages.

(i) Discuss the appropriateness of using 137.5 cm as the mean in the calculation of the standardised score,

(ii) Discuss the appropriateness of using 4.9 cm as the standard deviation in the calculation of the standardised score.

(4 marks)

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

Using 137.5 cm as the mean is not appropriate

1 Using 137.5 cm as the mean is appropriate

because the sample mean will be smaller than the population mean.

2 because the sample mean will be close to the population mean.

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

Using 4.9 cm as the standard deviation is appropriate

1 Using 4.9 cm as the standard deviation is not appropriate

because the sample standard deviation will be close to the population standard deviation.

2 because the sample mean will be more closely distributed than the individual values.

**11** A study took place in Brazil to find if there was a relationship between caffeine intake and reaction time of university students.

The researchers found the equations of the regression lines for the relationship between caffeine intake ( $x$  mg) and reaction time ( $y$  ms) for male students and female students before and after exams.

The table below gives the equations of the regression lines.

	before exams	after exams
male students	$y = -0.2x + 350$	$y = -0.25x + 340$
female students	$y = -0.1x + 330$	$y = -0.15x + 320$

(a) Interpret in context the figure  $-0.15$  in the regression equation for female students after the exams.

(1 mark)

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

- 1** For each 1 mg increase in caffeine intake
- For each 1 extra male student
- 2** the reaction time would decrease by 0.15 ms.
- the reaction time would increase by 0.15 ms.

(b) Compare the relationships between caffeine intake and reaction time in male and female students. Include in your comparisons reference to whether it is before or after exams.

(3 marks)

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

- For all university students, an increase in caffeine intake led to a increase in reaction time.
- The reaction time of male students decreased more per mg of caffeine intake than female students.
- As caffeine intake increased, the reaction time decreased more rapidly after exams compared to before exams.
- The reaction time of female students decreased more per mg of caffeine intake than male students.
- For all university students, an increase in caffeine intake led to a decrease in reaction time.
- As caffeine intake increased, the reaction time decreased more rapidly before exams compared to after exams.

(c) The researchers would like to use a normal distribution as a model for the reaction time of male students before exams.

i) Explain how they could check whether a normal distribution is a suitable model by drawing a histogram.

ii) Explain how they could check whether a normal distribution is a suitable model by calculating the averages and the standard deviation.

(3 marks)

Select **one** box.

If the histogram is linear, a normal distribution could be a suitable model.

If the histogram has a negative skew, a normal distribution could be a suitable model.

If the histogram has a positive skew, a normal distribution could be a suitable model.

If the histogram is symmetrical, a normal distribution could be a suitable model.

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

2 and almost all of data is within 3 standard deviation from the mean, a normal distribution could be a suitable model.

and 95% of data is within 3 standard deviation from the mean, a normal distribution could be a suitable model.

1 If the mean, median and mode are equal

If the mean, median and mode are larger than the standard deviation

**12** A factory produces lightbulbs, 5% of which are defective.

An inspector randomly selects 3 bulbs and the number of defective lightbulbs is recorded.

**(a)** Identify two conditions needed so that a binomial distribution is a suitable model for the number of defective lightbulbs is recorded.

(2 marks)

Select *two* boxes.

Each of the emails are independent.

The chance of getting a defective lightbulb remains constant.

The lightbulbs have the same packaging.

Outliers are removed.

**(b)** Calculate the probability, as a fraction, that all 3 of the lightbulbs are defective.

(2 marks)

Work out  $p^n$  where  $p$  is the given probability and  $n$  is the amount.

---

---

---

---

**(c)** Calculate the probability, as a fraction, that at least 2 of the lightbulbs are defective.

(3 marks)

You can use Pascal's triangle, or your calculator to find  $P(X < 2)$  then take this away from 1.

---

---

---

---