

Statistics GCSE

Paper 2

2025

Edexcel Higher

Variant 1 (same as video)

1ST0/1H

# Answers

Visit our website for tutorials on each question.

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

**1** At a business conference, 60% of attendees are industry professionals and 40% are students.

Jasmine and Leo plan to conduct a feedback survey.

Jasmine decides to use simple random sampling to select 50 participants.

She uses the official attendee list as a sampling frame, assigning a number to each participant.

She then generates 50 random numbers and selects her sample accordingly.

Leo decides to use quota sampling to collect a sample of 50 participants.

He plans to stand at the registration desk until 30 industry professionals and 20 students have been interviewed.

**(a)** Give two reasons why Jasmine's method may **not** produce a sample of 50 participants.

(2 marks)

Select *two* boxes.

The random numbers may not be whole numbers.

The random number may not correspond to a person's number.

The numbers may not have been mixed properly.

The numbers may not have been random.

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).

and will represent the population.

Quota sampling is easy

and will ensure every person has an equal chance of being picked.

Quota sampling is accurate

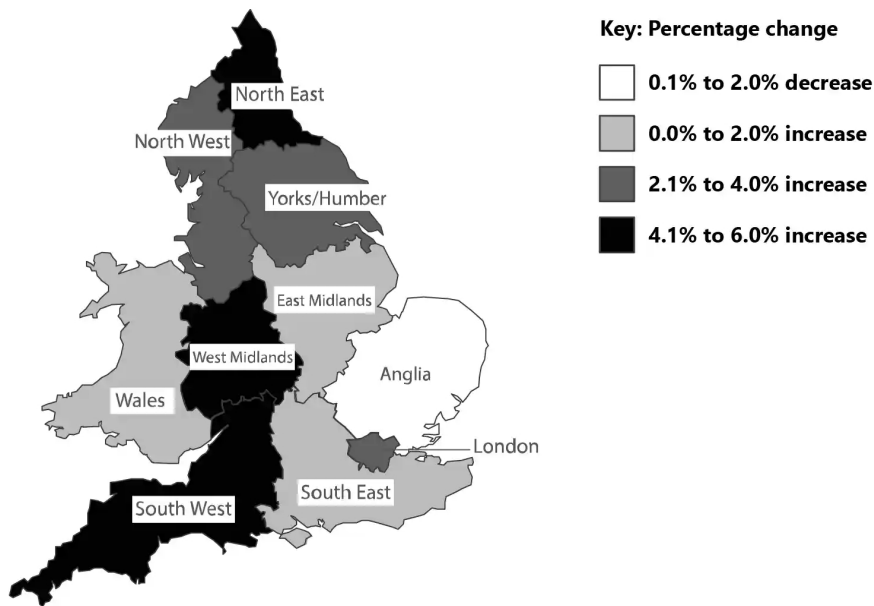
(c) Explain why the quota sample used by Leo is not a random sample.

(1 mark)

Select **one** box.

- Not every attendee has an equal chance of being selected.
- Leo should have only used industry professionals.
- Random sampling must use a random number generator.
- It is biased.

2 The map below shows the percentage change in domestic tourist visits across different regions of England and Wales between 2010 and 2011.



(a) Write down the percentage change in domestic tourist visits between 2010 and 2011 in Wales.

(1 mark)

Select *one* box.

- 0.0% to 2.0% increase
- 4.1% to 6.0% 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 **increased**.

(1 mark)

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

(c) Bryant states that domestic tourism in England and Wales increased overall between 2010 and 2011.

Explain why this may not be the case.

(1 mark)

Select **one** box.

The map does not show Scotland or Northern Ireland.

We only know the percentages, not the amounts.

More regions increased than decreased.

More regions decreased than increased.

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

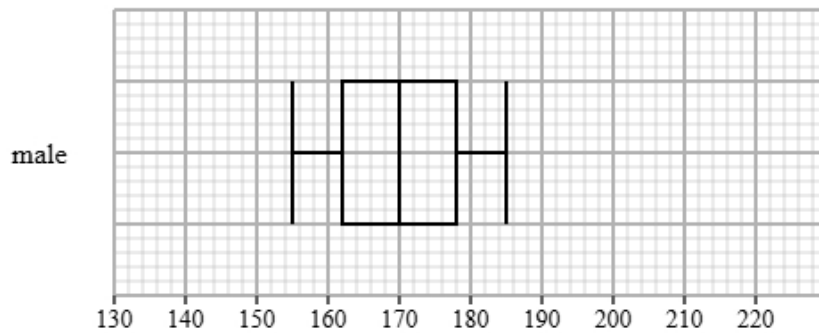
(1 mark)

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

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

choropleth

- 3 Sophie recorded the heights of male and female students in a school.  
Both groups were measured using the same method.  
The box plot shows information about the heights for the male students.



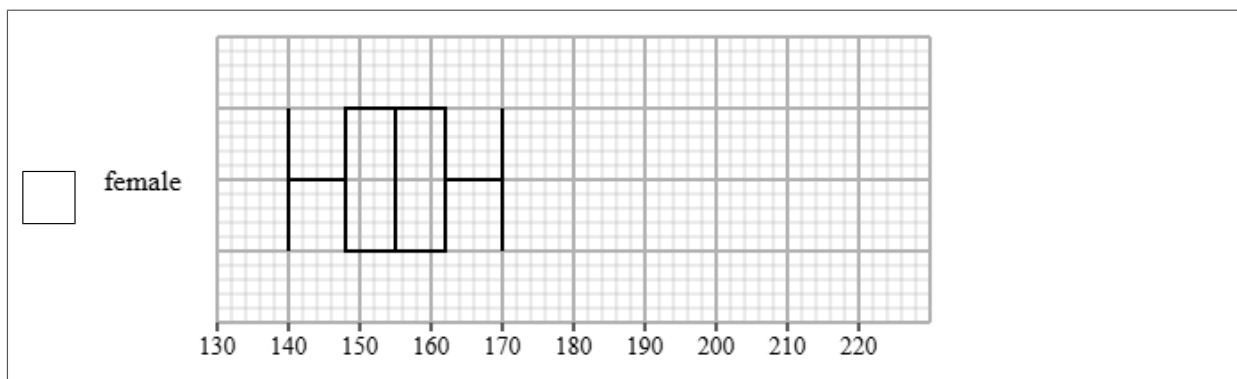
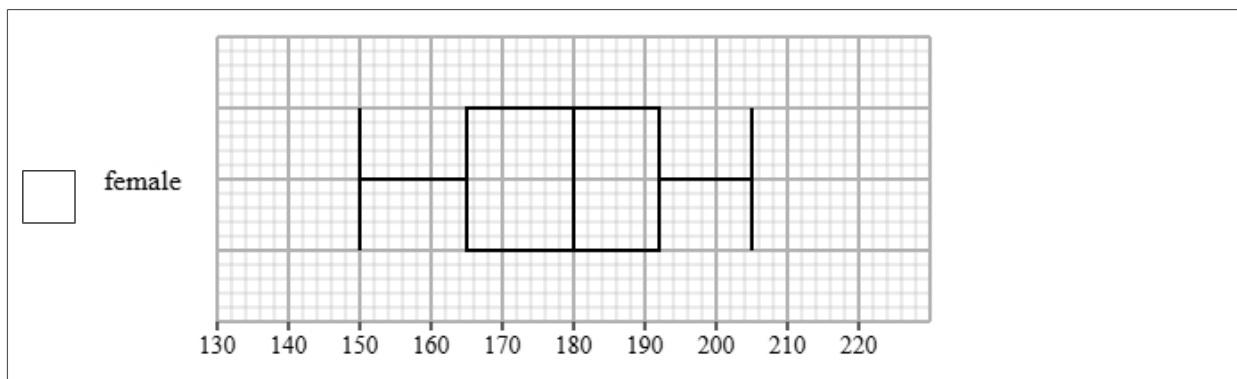
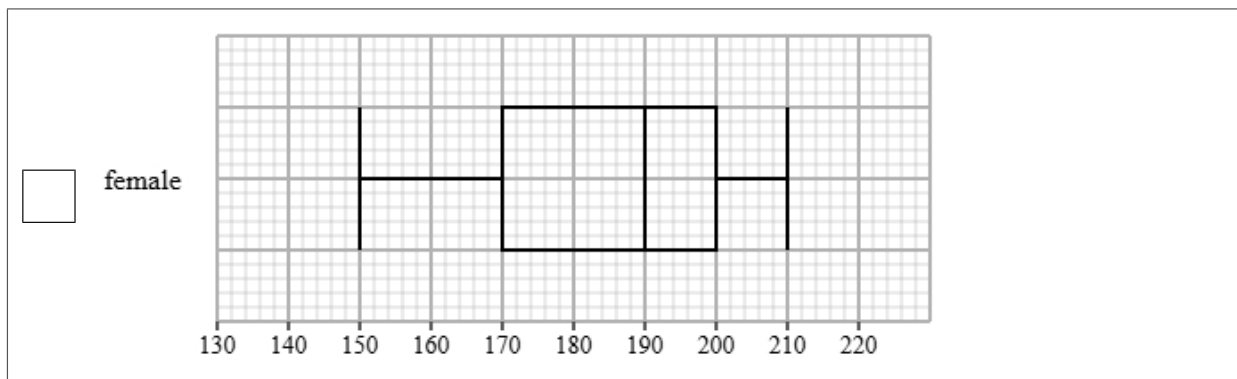
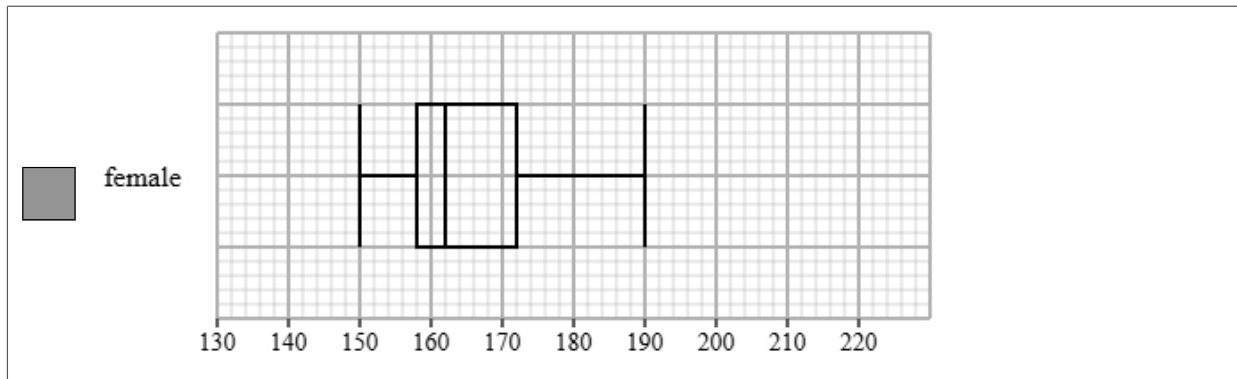
The table gives information about the heights for the female students.

Least tall	Lower quartile	Median	Upper quartile	Most tall
150	158	162	172	190

- (a) Draw a box plot for the heights for the female students.

(2 marks)

Select the correct answer.



(b) Compare the two distributions of heights.

Give three comparisons and interpret one of these comparisons.

(4 marks)

Select **one** box.

- The median is bigger.
- The median heights for male and female students are equal.
- The median heights for male students is greater than female students.
- The median heights for male students is lower than female students.

Select **one** box.

- The IQR is bigger.
- The IQR for the heights of the male and female students are equal.
- The IQR for the heights of the male students is greater than female students.
- The IQR for the heights of the male students is lower than female students.

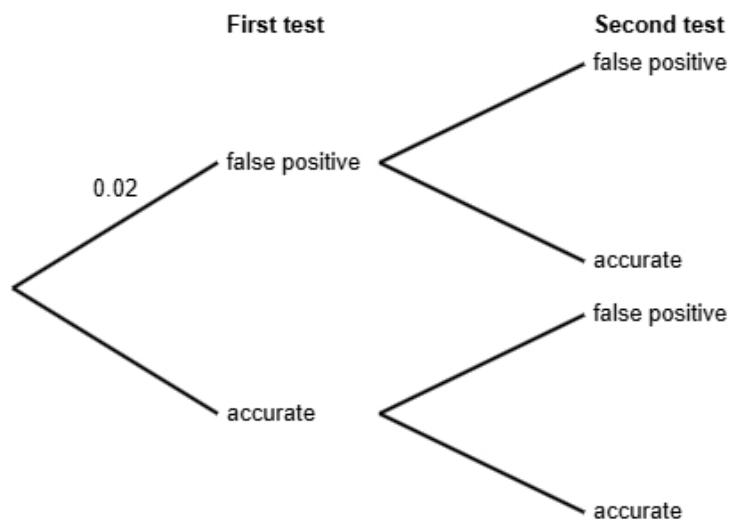
Select **one** box.

- The skews for the heights of the male and female students are both positive.
- The skew for the heights of the male students is symmetrical and the skew for the female students is negative.
- The skew for the heights of the male students is symmetrical and the skew for the female students is positive.
- The skews for the heights of the male and female students are both symmetrical.

Select **one** box.

- The heights for the male students are less spread out than the female students.
- The male students are more skewed than female students.
- The male students are on average taller than the female students.
- The male students are on average shorter than the female students.

- 4 It is estimated that 2% of a certain type of medical test gives a false positive result. The remaining tests provide accurate results. David took two such tests. He does not know if each test result is false positive or accurate.



- (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 David's test results are accurate.

(2 marks)

You will need to find  $P(\text{accurate})$  AND  $P(\text{accurate})$ .  
Remember, AND means  $\times$  in probability.

0.9604

(c) David states that the probability that exactly one of the tests is false positive is less than 4%

Find out whether or not David is correct.

(3 marks)

Find the probability of exactly one of the tests is false positive (there are two outcomes on the tree diagram).

---

---

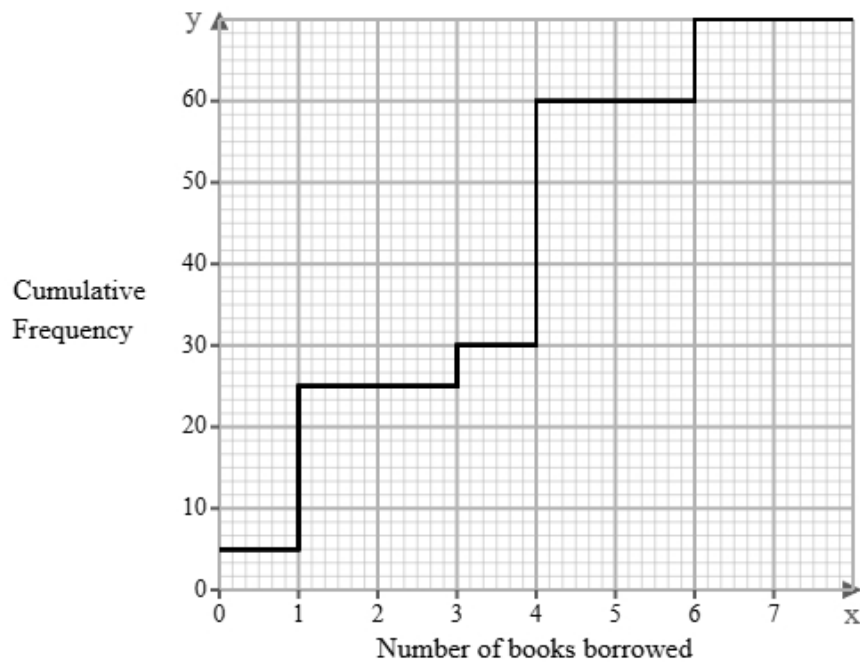
---

---

Select **one** box.

- The probability that exactly one of the tests is false positive is more than 4%, so David is not correct.
- The probability that exactly one of the tests is false positive is less than 4%, so David is not correct.
- The probability that exactly one of the tests is false positive is more than 4%, so David is correct.
- The probability that exactly one of the tests is false positive is less than 4%, so David is correct.

- 5 The cumulative frequency step polygon shows information about the number of books borrowed from a local library over 70 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 books borrowed from a local library is continuous.
- Because the number of books borrowed from a local library is discrete.
- Because the number of books borrowed from a local library is qualitative.
- Because the number of books borrowed from a local library is quantitative.

**(b)** Find the mode of the number of books borrowed.

(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.

\_\_\_\_\_ 4

**(c)** Find the number of days where there were:

- i) exactly 2 books borrowed.
- ii) more than 2 books borrowed.

(3 marks)

The frequency is shown by how much the graph 'goes up' at each point.  
Remember, the overall frequency is 70.

i) Exactly 2 books borrowed: \_\_\_\_\_ 0

ii) More than 2 books borrowed: \_\_\_\_\_ 45

**(d)** In 30 days fewer than  $x$  books were borrowed.

Find the value of  $x$

(1 mark)

Draw a line across from 30 on the graph and see where all the 'jumps up' to this line are under.

\_\_\_\_\_ 4

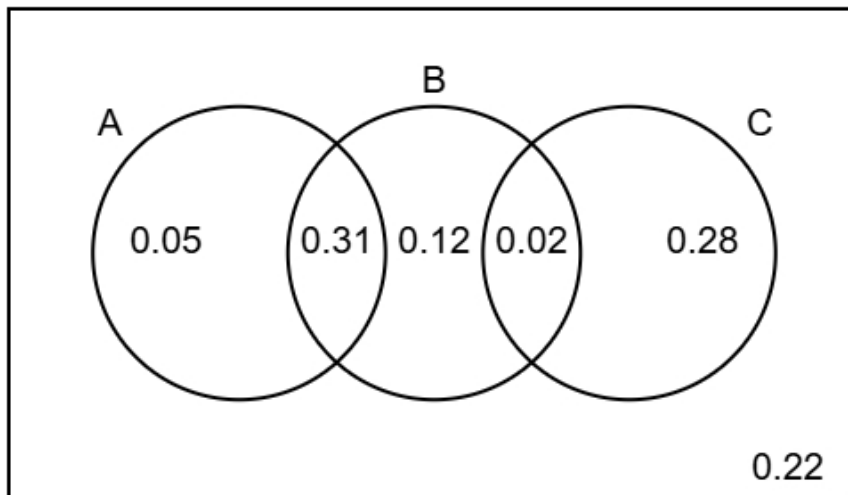
- (e) Laura believes the interquartile range of the number of books borrowed is 8.  
Explain why the interquartile range for this data cannot be 8.

(1 mark)

Select **one** box.

- The range is 6, so the IQR must be more than 6.
- 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 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.
- A and B are mutually exclusive
- because they only intersect once.
- B and C are mutually exclusive
- 2 because they do not intersect.
- 1 A and C are mutually exclusive

(b) Find  $P(B)$

(1 mark)

We are looking for the probabilities inside B.

0.45

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

(2 marks)

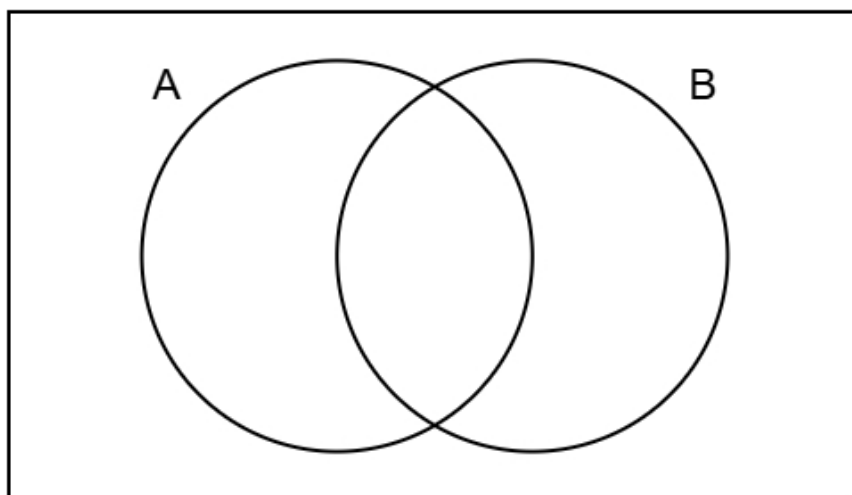
We are looking for the probabilities inside A or C.

0.66

(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 Ethan is researching the average GCSE grades and their rankings in the local newspaper for 10 schools.

(a) Suggest a diagram that Ethan could draw to determine if there is a relationship between the average GCSE grades and their rankings in the local newspaper for the 10 schools.

(1 mark)

*Select one box.*

Bar chart

Scatter diagram

Line graph

Pie chart

- (b) i) Calculate Spearman's rank correlation coefficient from the table.  
ii) Interpret your answer in the context of Ethan'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).

School	Mean GCSE Grade	GCSE Grade Rank	Newspaper Rank	d	d <sup>2</sup>
Brookside High	4.8	7	9	-2	4
Cedar Hill Secondary	3.9	8	5	3	9
Eastbank High	6.9	4	6	-2	4
Greystone Academy	7.2	2	1	1	1
Maple Grove School	7.1	3	3	0	0
Northwood School	3.5	9	8	1	1
Pinecrest School	3.1	10	10	0	0
Riverview School	5.6	5	4	1	1
Silverleaf Academy	5.2	6	7	-1	1
Westfield High	7.9	1	2	-1	1

---



---



---



---

Select *two* boxes.

- As the average GCSE grades increases, their rankings in the local newspaper increases.
- The presence of outliers would decrease the correlation.
- As the average GCSE grades increases, their rankings in the local newspaper decreases.
- The presence of outliers would increase the correlation.

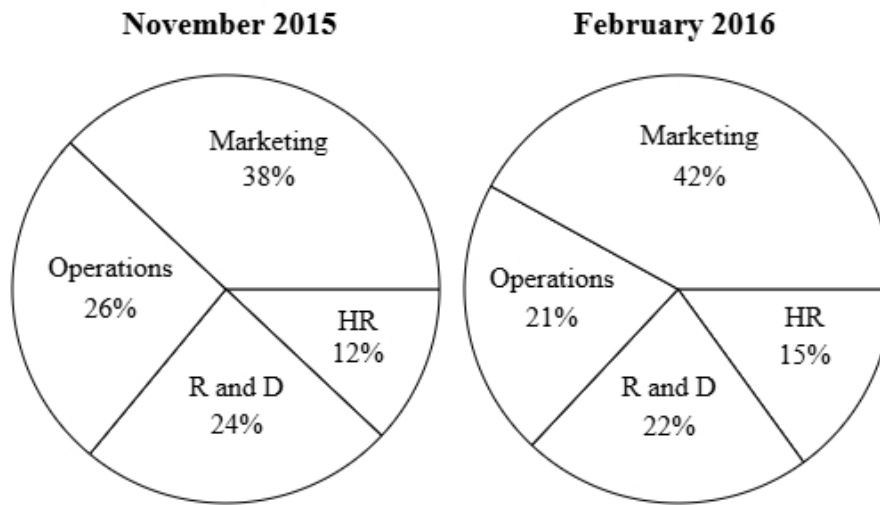
- (c) Discuss whether Ethan should have used Pearson's product moment correlation coefficient instead of Spearman's rank correlation coefficient to measure the correlation.

(3 marks)

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

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

8 The pie charts show the budget allocations for a business in November 2015 and February 2016.



(a) Explain why the pie charts do not show a larger marketing spend in February 2016 compared to November 2015.

(1 mark)

Select **one** box.

- Pie charts cannot be used for comparison.
- There may be many other business factors that have changed between the two dates.
- The pie charts do show a larger marketing spend in February 2016.
- November 2015 could have had a larger total budget than February 2016.

- (b) The marketing budget in November 2015 was £58748.  
Find the R and D budget in November 2015.

(2 marks)

The marketing budget in November 2015 was £58748 which is shown by 38% on the pie chart.

$$38\% = \text{£}58748$$

Find 1%

Then find the R and D budget (24%).

£ 37104

- (c) In November 2015 the total budget was £155000 (nearest thousand).  
In February 2016 the total budget was £185000 (nearest thousand).

Gemma wants to use the totals to draw pie charts.

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

(5 marks)

Select **one** box.

- Gemma can use comparative pie charts.
- Gemma can show the totals instead of the percentages.
- Gemma can use dual pie charts.
- Gemma could use a bar chart.

Select **one** box.

- $\left(\frac{185000}{155000}\right)^2$
- $\frac{185000^2}{155000^2}$
- $\sqrt{\frac{185000}{155000}}$
- $\frac{185000}{155000}$

The radius of the November 2015 pie chart will be 1.09 (2 d.p.) larger than the February 2016 pie chart.

Select **one** box.

- This method is much quicker to do than other types of pie chart.
- The totals for these pie charts are different, so the areas of each pie chart will represent the totals whilst keeping the proportions the same.
- The areas will now be inversely proportional to the frequency, making the pie charts much easier to understand.
- The proportions of each sector will change depending on the total for each pie chart.

9 A company bottles soft drinks.

The bottles have a target volume of 500 ml.

The company uses quality assurance to monitor the volume of soft drink in each bottle.

Samples of the bottles are taken from the production line at regular intervals and the mean volume of soft drink in the bottles in each sample is found.

The sample means should be normally distributed with a mean of 500 ml and a standard deviation of 3 ml.

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

(2 marks)

Use the formula for the upper action limit

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

\_\_\_\_\_ 509 ml

(b) The upper action limit will be set closer to the target volume of 500 ml.

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

(1 mark)

Select **one** box.

The production process will work faster.

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

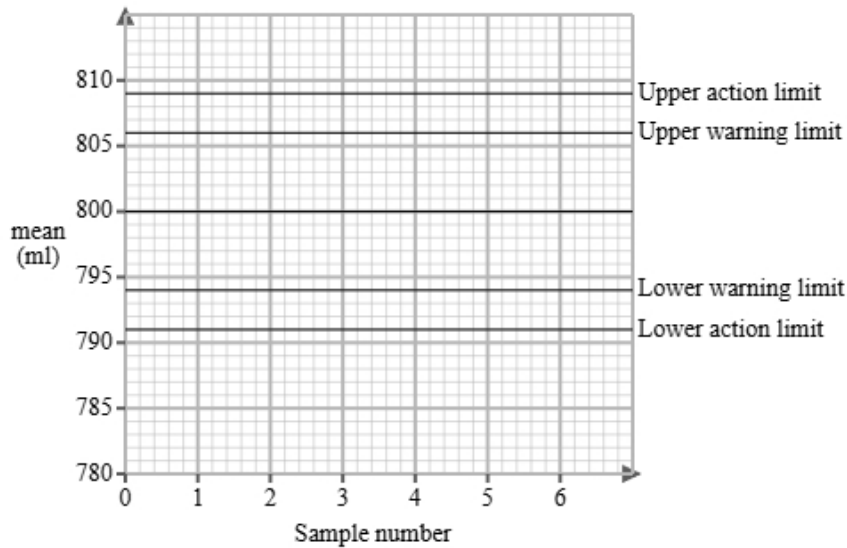
The production process would be stopped less frequently.

The production process would be stopped more frequently.

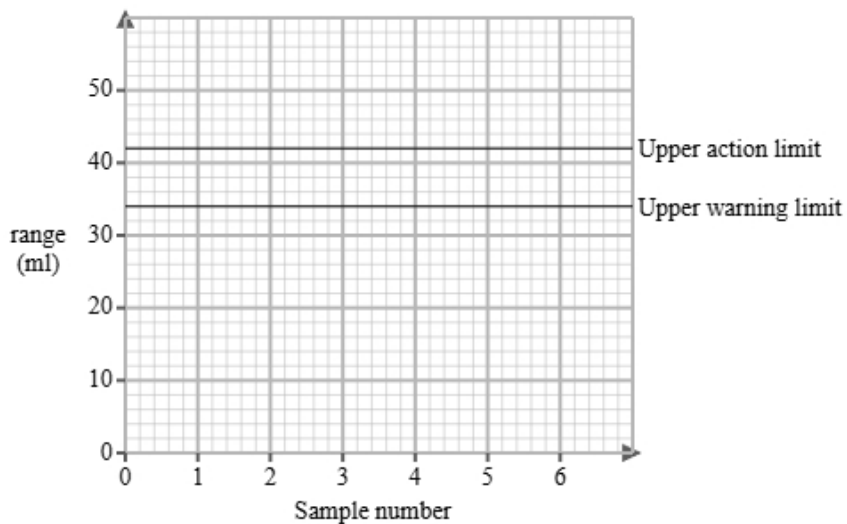
(c) The company also bottles tomato ketchup and uses quality assurance to monitor the volume of tomato ketchup in those bottles.

Here are the control charts for the sample means and for the sample ranges of the volumes of tomato ketchup in the bottles.

**Control chart for means**



**Control chart for ranges**



A sample is taken and is found to have a mean of 807 ml and a range of 43 ml. 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).

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

**10** The ages of a group of boys have a mean of 5.6 years and a standard deviation of 0.8 years.

**(a)** Liam is boy with a standardised score of 0.  
Find Liam's age.

(1 mark)

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

5.6 years

**(b)** Daniel and Alex are both boys in the group.  
Daniel's standardised score for age is 0.9 years.  
Alex's standardised score for age is -0.5 years.  
Daniel is older than Alex.  
How much older is Daniel?

(3 marks)

Rearrange the formula to make the value the subject

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

1.12 years

(c) Arjun takes a sample of 4 boys from the group.

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

(i) Discuss the appropriateness of using 5.6 years as the mean in the calculation of the standardised score,

(ii) Discuss the appropriateness of using 0.8 years 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 5.6 years as the mean is not appropriate

1 Using 5.6 years 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).

1 Using 0.8 years as the standard deviation is not appropriate

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

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

Using 0.8 years as the standard deviation is appropriate

**11** A study took place in Sweden to find if there was a relationship between daily exercise and resting heart rate of middle-aged men.

The researchers found the equations of the regression lines for the relationship between daily exercise ( $x$  minutes) and resting heart rate ( $y$  bpm) for gym members and non-gym members on weekdays and weekends.

The table below gives the equations of the regression lines.

	weekdays	weekends
gym members	$y = -0.3x + 72$	$y = -0.25x + 75$
non-gym members	$y = -0.1x + 78$	$y = -0.05x + 80$

(a) Interpret in context the figure  $-0.05$  in the regression equation for non-gym members at weekends.

(1 mark)

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

- 1** For each 1 minute increase in daily exercise
- For each 1 extra gym member
- 2** the heart rate would decrease by 0.05 bpm.
- the heart rate would increase by 0.05 bpm.

(b) Compare the relationships between daily exercise and resting heart rate in middle-aged men who are gym and non-gym members. Include in your comparisons reference to whether it is a weekday or weekend.

(3 marks)

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

- For all middle-aged men, an increase in daily exercise led to a increase in resting heart rate.
- The resting heart rate of gym members decreased more per minutes of daily exercise than non-gym members.
- As daily exercise increased, the resting heart rate decreased more rapidly on weekends compared to weekdays.
- The resting heart rate of non-gym members decreased more per minutes of daily exercise than gym members.
- As daily exercise increased, the resting heart rate decreased more rapidly on weekdays compared to weekends.
- For all middle-aged men, an increase in daily exercise led to a decrease in resting heart rate.

- (c) The researchers would like to use a normal distribution as a model for the resting heart rate of gym members on weekdays.
- 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 positively skewed, a normal distribution could be a suitable model.
- If the histogram shows an increase, a normal distribution could be a suitable model.
- If the histogram is a bell shape, a normal distribution could be a suitable model.
- If the histogram shows a decrease, a normal distribution could be a suitable model.

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

- 2** and 95% of data is within 2 standard deviations from the mean, a normal distribution could be a suitable model.
- and 68% of data is within 2 standard deviations 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** 20% of the emails a user receives are spam emails.  
Out of 3 incoming emails, the number of spam emails is recorded.

**(a)** Identify two conditions needed so that a binomial distribution is a suitable model for the number of spam emails in the incoming emails.

(2 marks)

Select *two* boxes.

- Each of the emails are independent.
- There are only two possible outcomes, spam or not spam.
- Each email increased the likelihood of a spam email.
- The emails arrive at the same time.

**(b)** Calculate the probability, as a fraction, that all 3 of the incoming emails are spam.

(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 incoming emails are spam.

(3 marks)

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

---

---

---

---