

**Statistics GCSE****Paper 2**

2025

Edexcel Higher

Variant 5

1ST0/1H

**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.

**Copyright**

- Sharing this PDF is strictly forbidden unless you have the author's permission.
- This paper is only authorised to be used by the person who has bought it.
- You may not store this online or any shared area such as an intranet.
- Please contact us if you have any queries.

**Advice**

- You can get support for all these questions at our website: [www.statsgcse.com](http://www.statsgcse.com)
- This paper and more are available on our site with questions that change subtly after each attempt.
- Good luck!

- 1 At a career fair, 90% of attendees are recent graduates and 10% are professionals. Ella and Nathan plan to conduct a feedback survey.

Ella decides to use simple random sampling to select 100 participants.

She uses the list of registered attendees as a sampling frame, assigning a number to each participant.

She then generates 100 random numbers and selects her sample accordingly.

Nathan decides to use quota sampling to collect a sample of 100 participants.

He plans to stand near one of the entrances until 90 recent graduates and 10 professionals have been interviewed.

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

(2 marks)

Select **two** boxes.

- The registered attendees list may not be up to date.
- The selected person may not want to participate.
- The numbers may not have been random.
- Some of the random numbers may be repeated.
- The career fair may be busy.

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

(2 marks)

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

- Quota sampling uses the full population
- and does not require a sample frame.
- and will split the sample frame into recent graduates and professionals.
- Quota sampling is cheap

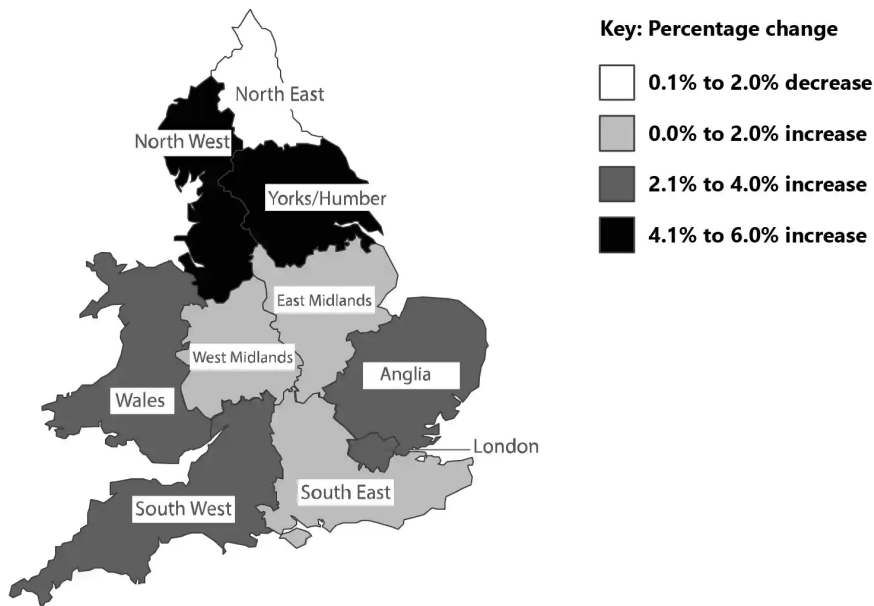
(c) Explain why the quota sample used by Nathan is not a random sample.

(1 mark)

Select **one** box.

- He made sure the sample was balanced.
- He made sure the sample was balanced.
- Not every attendee has an equal chance of being selected.
- All people needed to answer anonymously.

2 The map below shows the percentage change in domestic tourist visits across different regions of England and Wales between 2014 and 2015.



(a) Write down the percentage change in domestic tourist visits between 2014 and 2015 in North West.

(1 mark)

Select **one** box.

- 0.1% to 2.0% decrease
- 0.0% to 2.0% increase
- 2.1% to 4.0% increase
- 4.1% to 6.0% 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) Hadley states that domestic tourism in England and Wales increased overall between 2014 and 2015.

Explain why this may not be the case.

(1 mark)

Select **one** box.

- Overall there was no change, because some increased and others decreased.
- We would need a pie chart.
- The map only shows percentages not numbers.
- We do not know any data after 2015.

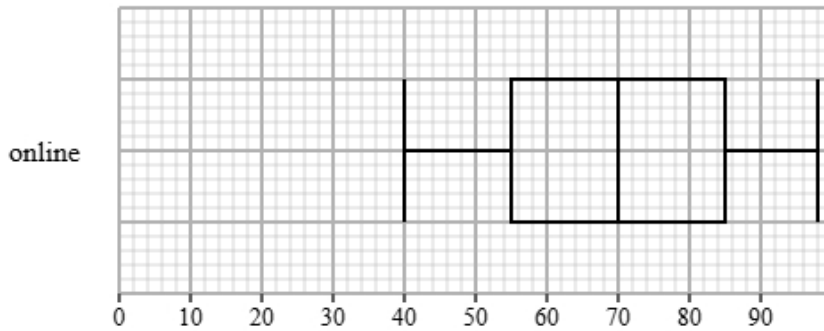
(d) State the statistical name for the type of map shown.

(1 mark)

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

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

- 3 Isla collected the satisfaction scores for online and in-store customers. Both types of shopping experiences were rated by the same number of customers. The box plot presents data on the satisfaction scores for the online customers.



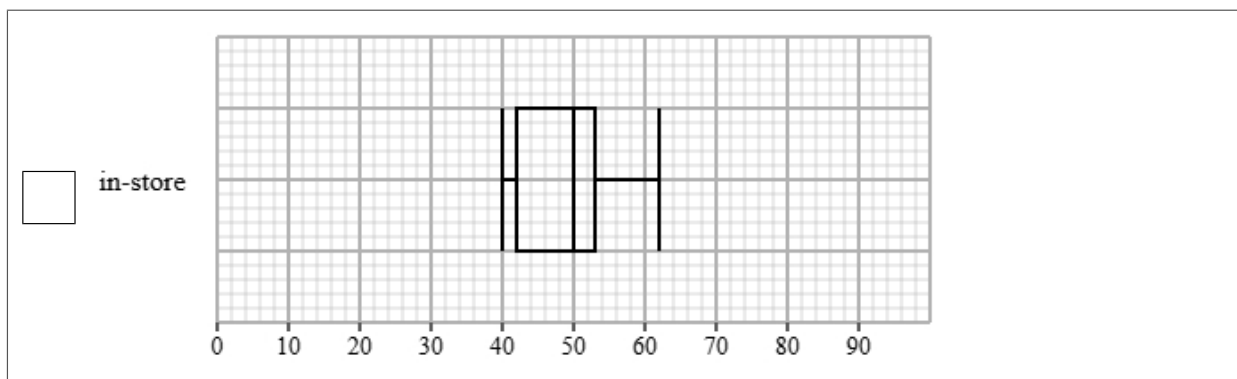
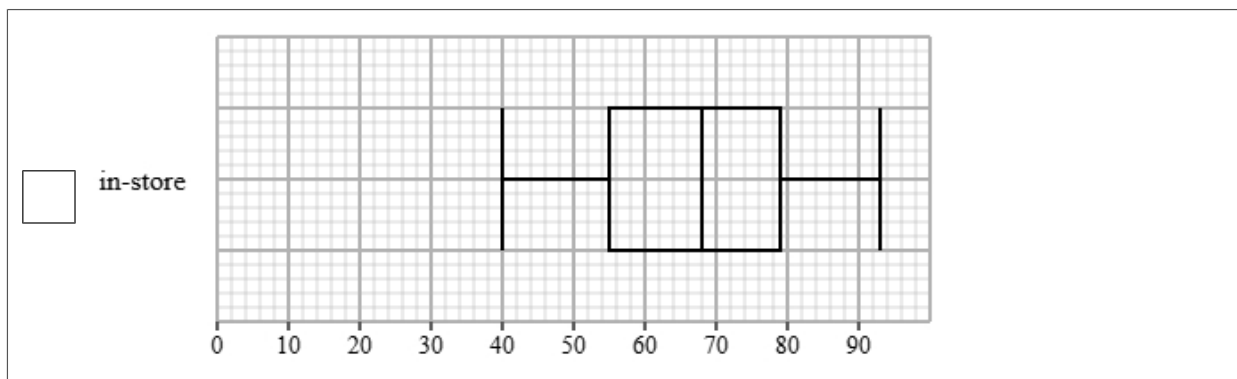
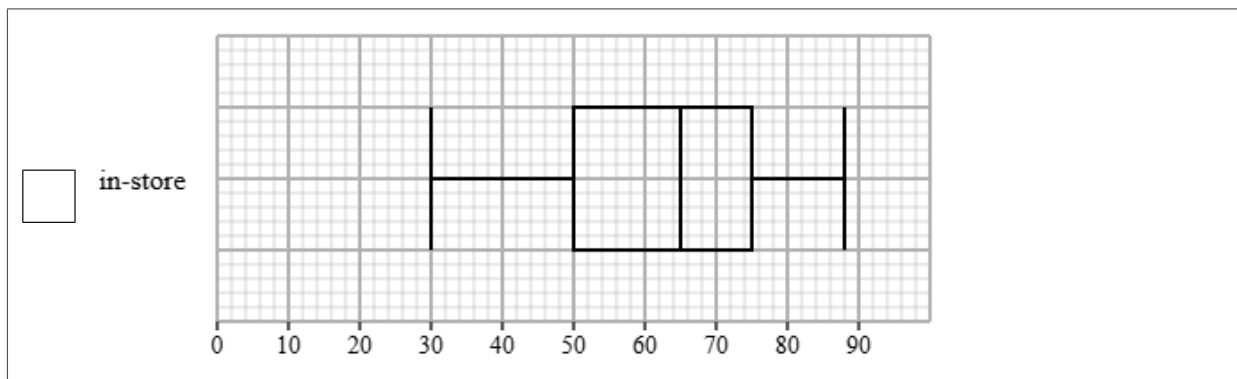
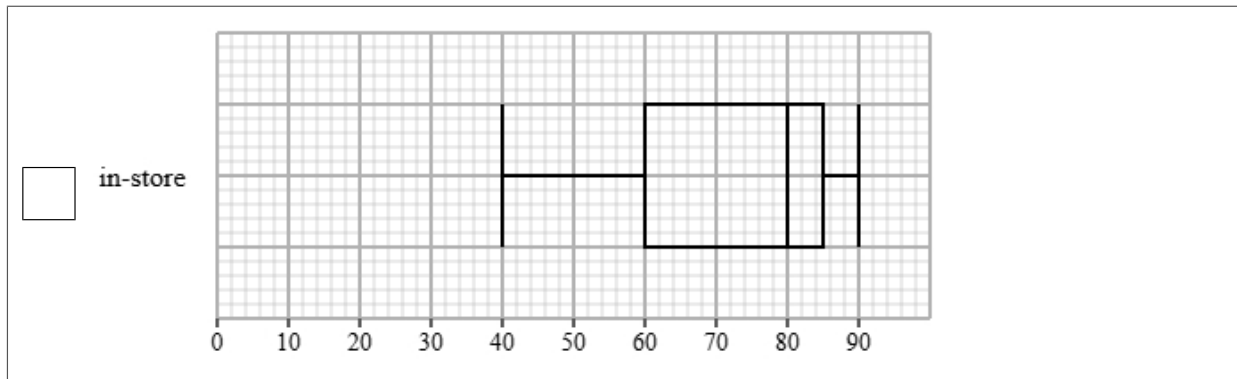
The table gives information about the satisfaction scores for the in-store customers.

Least tall	Lower quartile	Median	Upper quartile	Most tall
40	60	80	85	90

- (a) Draw a box plot for the satisfaction scores for the in-store customers.

(2 marks)

Select the correct answer.



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

(4 marks)

Select **one** box.

- The median is bigger.
- The median satisfaction scores for online customers is greater than in-store customers.
- The median satisfaction scores for online and in-store customers are equal.
- The median satisfaction scores for online customers is lower than in-store customers.

Select **one** box.

- The IQR is bigger.
- The IQR for the satisfaction scores of the online customers is greater than in-store customers.
- The IQR for the satisfaction scores of the online and in-store customers are equal.
- The IQR for the satisfaction scores of the online customers is lower than in-store customers.

Select **one** box.

- The skews for the satisfaction scores of the online and in-store customers are both positive.
- The skews for the satisfaction scores of the online and in-store customers are both symmetrical.
- The skew for the satisfaction scores of the online customers is symmetrical and the skew for the in-store customers is negative.
- The skew for the satisfaction scores of the online customers is symmetrical and the skew for the in-store customers is positive.

Select **one** box.

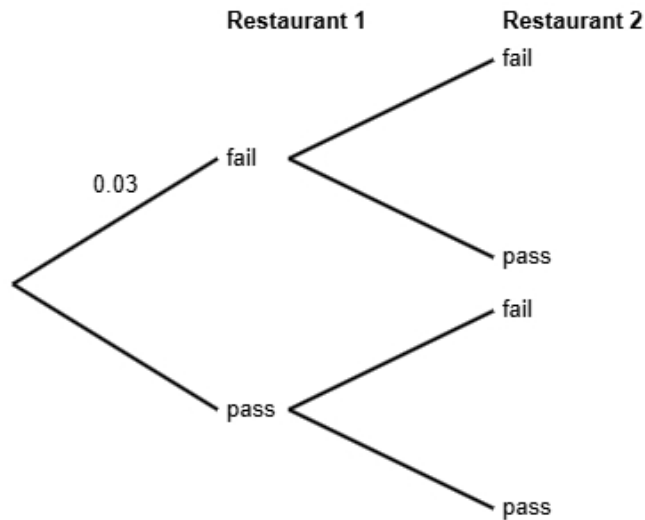
- The scores for the online customers are less spread out than the in-store customers.
- The online customers on average have higher satisfaction scores than the in-store customers.
- The online customers are more skewed than in-store customers.
- The online customers on average have lower satisfaction scores than the in-store customers.

4 Health inspections show that 3% of restaurants in a city fail basic hygiene standards.

The rest pass the required health standards.

Emma is a food safety officer and will be inspecting two different restaurants.

She does not know if each restaurant will fail or pass.



(a) Complete the probability tree diagram.

(2 marks)

The branches for each stage must add up to 1.  
Each test is independent so will have the same probabilities.

(b) Find the probability that both of Emma's restaurants have an outcome of pass.

(2 marks)

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

(c) Emma states that the probability that exactly one restaurant outcome is fail is less than 6%

Find out whether or not Emma is correct.

(3 marks)

Find the probability of exactly one restaurant outcome is fail (there are two outcomes on the tree diagram).

---

---

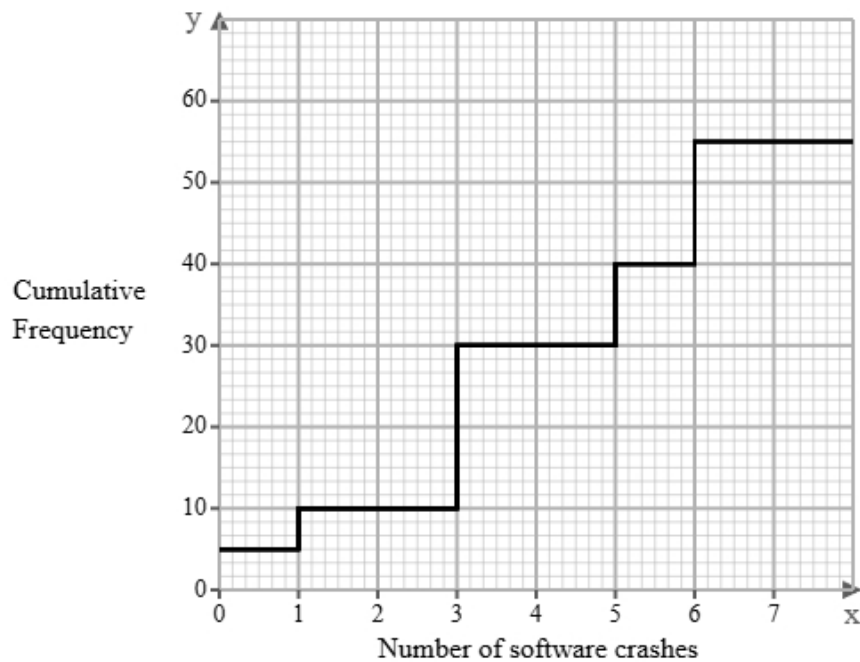
---

---

Select **one** box.

- The probability that exactly one restaurant outcome is fail is less than 6%, so Emma is correct.
- The probability that exactly one restaurant outcome is fail is more than 6%, so Emma is not correct.
- The probability that exactly one restaurant outcome is fail is more than 6%, so Emma is correct.
- The probability that exactly one restaurant outcome is fail is less than 6%, so Emma is not correct.

- 5 The cumulative frequency step polygon shows information about number of software crashes experienced by a company's servers over 55 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 software crashes experienced by a company's servers is continuous.
- Because number of software crashes experienced by a company's servers is discrete.
- Because number of software crashes experienced by a company's servers is quantitative.
- Because number of software crashes experienced by a company's servers is qualitative.

**(b)** Find the mode of the number of software crashes.

(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.

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

- i) exactly 4 software crashes.
- ii) more than 4 software crashes.

(3 marks)

The frequency is shown by how much the graph 'goes up' at each point.  
Remember, the overall frequency is 55.

i) Exactly 4 software crashes: \_\_\_\_\_

ii) More than 4 software crashes: \_\_\_\_\_

**(d)** In 40 days fewer than  $x$  software crashes were experienced.

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.

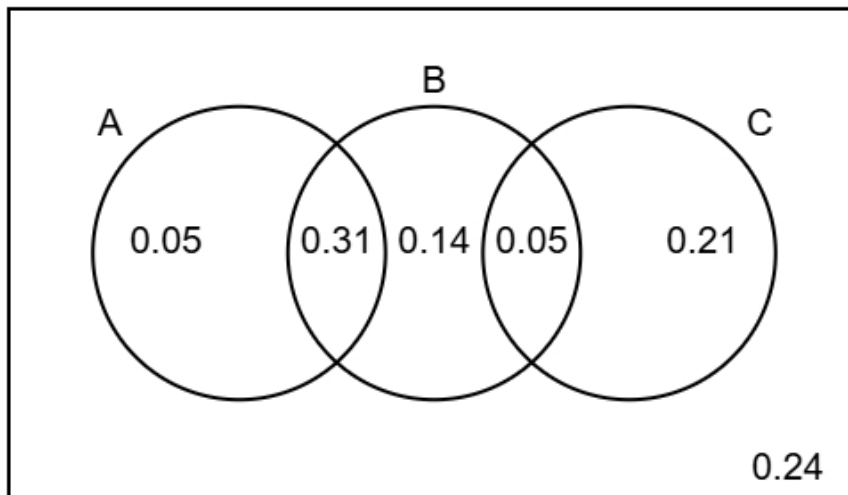
- (e) Delia believes the interquartile range of number of software crashes is 8.  
Explain why the interquartile range for this data cannot be 8.

(1 mark)

Select *one* box.

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

(b) Find  $P(B)$

(1 mark)

We are looking for the probabilities inside B.

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

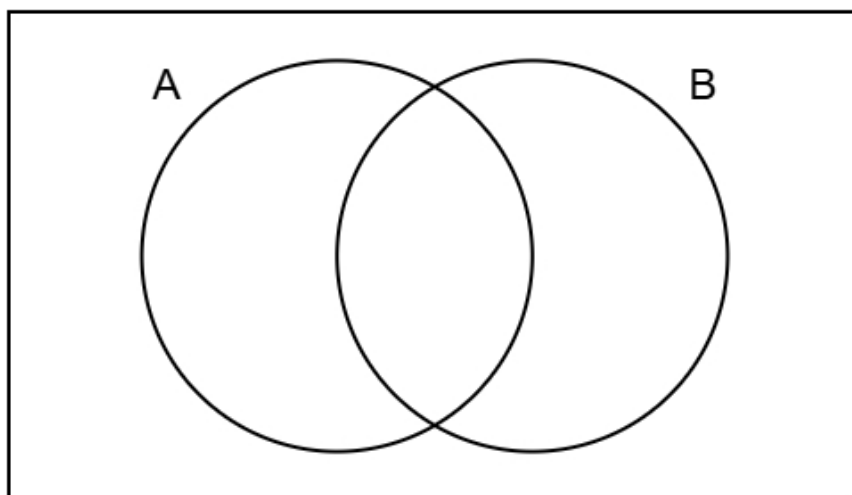
(2 marks)

We are looking for the probabilities inside A or C.

(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 Sebastian is researching average daily number of cups sold and the coffee quality rating for 10 coffee blends at a coffee shop.

(a) Suggest a diagram that Sebastian could draw to determine if there is a relationship between average daily number of cups sold and the coffee quality rating for the 10 coffee blends at a coffee shop.

(1 mark)

*Select one box.*

Bar chart

Scatter diagram

Box plot

Line graph

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

Coffee Blend	Average Daily Number of Cups Sold	Cups Sold Rank	Coffee Quality Rank	d	d <sup>2</sup>
Caramel Delight	92	1	2	-1	1
Caramel Macchiato	21	8	7	1	1
Choco Hazelnut	72	3	3	0	0
Coconut Crunch	28	7	8	-1	1
Dark Roast Supreme	35	6	6		
Hazelnut Dream	89	2	5		
Mocha Fusion	61	5	4		
Mocha Java	2	10	9		
Vanilla Bean Bliss	63	4	6		
Espresso Royale	9	9	10		

---



---



---



---

Select **two** boxes.

- As average daily number of cups sold increases, the coffee quality rating decreases.
- The presence of outliers would decrease the correlation.
- As average daily number of cups sold increases, the coffee quality rating increases.
- The presence of outliers would increase the correlation.

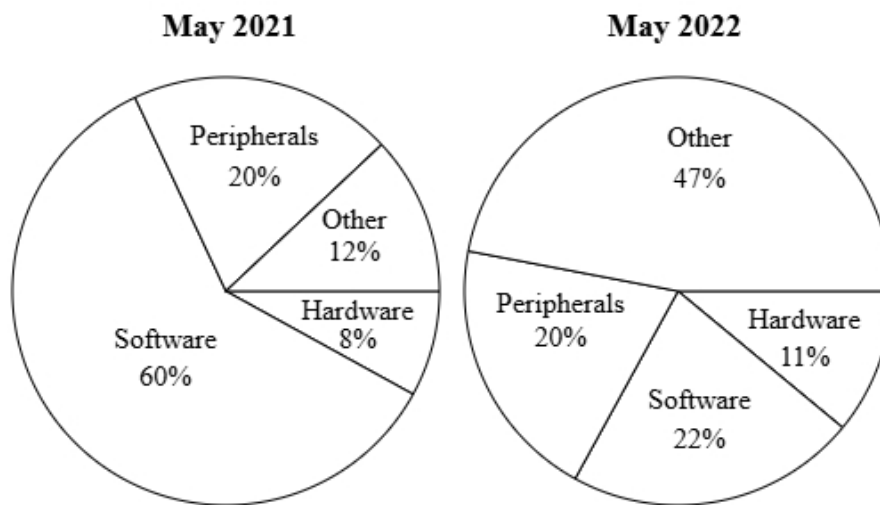
(c) Discuss whether Sebastian should have used Pearson's product moment correlation coefficient instead of Spearman's rank correlation coefficient to measure the correlation.

(3 marks)

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

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

8 The pie charts show the types of sales at a computer store in May 2021 and May 2022.



(a) Explain why the pie charts do not show an increase in hardware sales in May 2022 compared to May 2021.

(1 mark)

Select *one* box.

- Pie charts cannot be used for comparison.
- Pie charts only show positive trends.
- May 2021 could have had a larger total sales than May 2022.
- Pie charts are inaccurate.

- (b) The sales for software in May 2021 was £59160.  
Find the sales for hardware in May 2021.

(2 marks)

The sales for software in May 2021 was £59160 which is shown by 60% on the pie chart.

$$60\% = \text{£}59160$$

Find 1%

Then find the sales for hardware (8%).

£ \_\_\_\_\_

- (c) In May 2021 the total sales was £99000 (nearest thousand).  
In May 2022 the total sales was £132000 (nearest thousand).

Maria wants to use the totals to draw pie charts.

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

(5 marks)

Select **one** box.

- Maria can use dual pie charts.
- Maria can use comparative pie charts.
- Maria can use a line graph.
- Maria can use Venn pie charts.

Select **one** box.

- $\sqrt{\frac{132000}{99000}}$
- $\left(\frac{132000}{99000}\right)^2$
- $\frac{132000^2}{99000^2}$
- $\frac{132000}{99000}$

The radius of the May 2021 pie chart will be \_\_\_\_\_ (2 d.p.) larger than the  
May 2022 pie chart.

Select **one** box.

- Areas are now made inversely proportional to frequency, improving the readability of the charts.
- The total values in these pie charts vary, so the size of each chart will show those differences while keeping sector proportions the same.
- The sector proportions will be influenced by the total value of each chart
- This method is more efficient than other pie chart techniques.

9 A company manufactures LED light bulbs.

The bulbs have a target brightness of 800 lumens.

The company uses quality assurance to monitor the brightness of each bulb.

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

The sample means should be normally distributed with a mean of 800 lumens and a standard deviation of 10 lumens.

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

(2 marks)

Use the formula for the upper action limit

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

\_\_\_\_\_ lumens

(b) The upper action limit will be set closer to the target brightness of 800 lumens.

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

(1 mark)

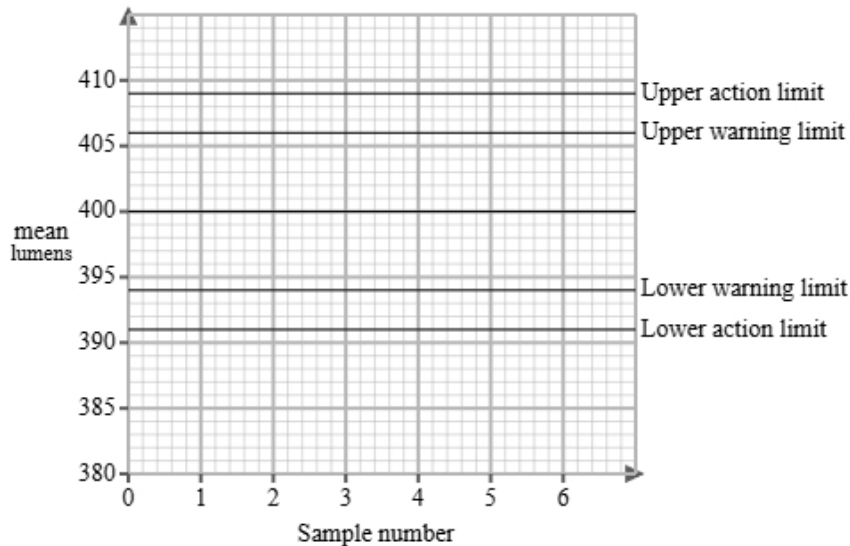
Select *one* box.

- 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.
- The production process will work faster.

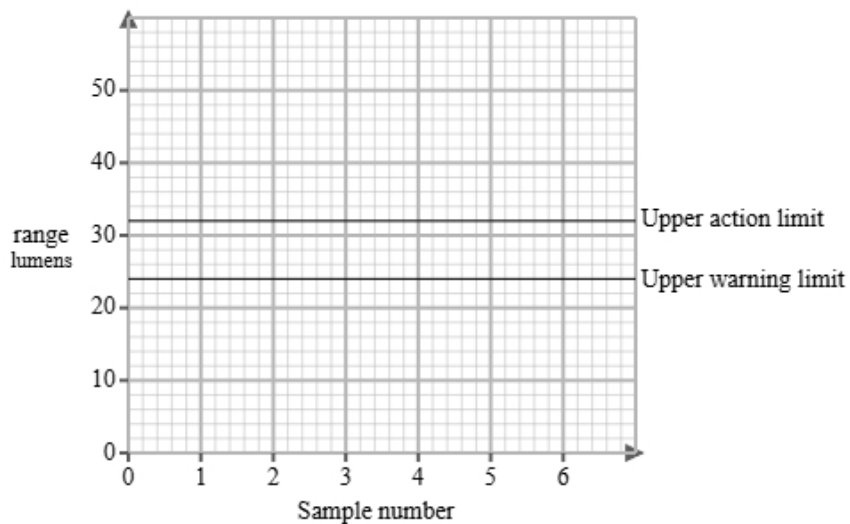
(c) The company also manufactures LED light strips and uses quality assurance to monitor the brightness of each strip.

Here are the control charts for the sample means and for the sample ranges of the brightness of each of the strips.

**Control chart for means**



**Control chart for ranges**



A sample is taken and is found to have a mean of 408 lumens and a range of 32 lumens.

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).

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

**10** The running speeds of a group of athletes have a mean of 14.5 km/h and a standard deviation of 1.8 km/h.

**(a)** Olivia is athlete with a standardised score of 0.  
Find Olivia's running speed.

(1 mark)

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

\_\_\_\_\_ km/h

**(b)** Aisha and Mia are both athletes in the group.  
Aisha's standardised score for running speed is 1.3 km/h.  
Mia's standardised score for running speed is -0.7 km/h.  
Aisha was faster than Mia.  
How much faster is Aisha?

(3 marks)

Rearrange the formula to make the value the subject

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

\_\_\_\_\_ km/h

(c) Emily takes a sample of 5 athletes from the group.

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

(i) Discuss the appropriateness of using 14.5 km/h as the mean in the calculation of the standardised score,

(ii) Discuss the appropriateness of using 1.8 km/h 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).

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

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

Using 14.5 km/h as the mean is appropriate

Using 14.5 km/h as the mean is not appropriate

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

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

Using 1.8 km/h as the standard deviation is appropriate

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

Using 1.8 km/h as the standard deviation is not appropriate

**11** A study took place in Australia to find if there was a relationship between age and number of books read of public library members.

The researchers found the equations of the regression lines for the relationship between age ( $x$  years) and number of books read ( $y$  books) for urban members and rural members who prefer fiction and non-fiction.

The table below gives the equations of the regression lines.

	prefer fiction	prefer non-fiction
urban members	$y = -0.2x + 25$	$y = -0.1x + 20$
rural members	$y = -0.3x + 30$	$y = -0.15x + 22$

**(a)** Interpret in context the figure  $-0.15$  in the regression equation for rural members who prefer non-fiction.

(1 mark)

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

- For each 1 year increase in age
- the number of books read would decrease by 0.15.
- For each 1 extra rural member
- the number of books read would increase by 0.15.

(b) Compare the relationships between age and number of books read in urban and rural members. Include in your comparisons reference to whether they prefer fiction or non-fiction.

(3 marks)

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

- For all public library members, an increase in age led to a increase in number of books read.
- The number of books read of rural members decreased more per years of age than urban members.
- For all public library members, an increase in age led to a decrease in number of books read.
- The number of books read of urban members decreased more per years of age than rural members.
- As age increased, the number of books read decreased more rapidly in members who prefer fiction compared to prefer non-fiction.
- As age increased, the number of books read decreased more rapidly in members who prefer non-fiction compared to prefer fiction.

(c) The researchers would like to use a normal distribution as a model for the number of books read of urban members who prefer fiction.

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 has no skew, a normal distribution could be a suitable model.

If the histogram shows a linear relationship, 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.

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

and 63% of data is within 1 standard deviation from the mean, a normal distribution could be a suitable model.

and 95% of data is within 1 standard deviation from the mean, a normal distribution could be a suitable model.

If the skewness is large (calculated from the mean, median and standard deviation)

If the skewness is 0 (calculated from the mean, median and standard deviation)

**12** A spinner has 8 equal sections, numbered 1 to 8.  
It is spun 4 times.  
The number of times it lands on a 8 is recorded.

**(a)** Identify two conditions needed so that a binomial distribution is a suitable model for the number of times an 8 was recorded.

(2 marks)

Select *two* boxes.

- The spinner is biased.
- There are only two possible outcomes, 8 or not 8.
- All other numbers are also recorded.
- The chance of getting an 8 remains constant.

**(b)** Calculate the probability, as a fraction, that all 4 of the spinners land on an 8.

(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 spinners land on an 8.

(3 marks)

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

---

---

---

---