

Statistics GCSE

Paper 2

2025

Edexcel Higher

Variant 2

1ST0/1H

Answers

Visit our website for tutorials on each question.

www.statsgcse.com

- 1 In a library, 60% of members are adults and 40% are children.
Priya and Thomas plan to conduct a feedback survey.

Priya decides to use simple random sampling to select 100 participants.
She uses the library membership database as a sampling frame, assigning a number to each participant.
She then generates 100 random numbers and selects her sample accordingly.

Thomas decides to use quota sampling to collect a sample of 100 participants.
He plans to sit at the library's exit until 60 adults and 40 children have been interviewed.

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

(2 marks)

Select **two** boxes.

- The random numbers may not be whole numbers.
 The numbers may not have been mixed properly.
 Some people may have left the library.
 Some of the random numbers may be repeated.
 The method may produce more people than is required.

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

(2 marks)

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

- 1 Quota sampling is cheap
 Quota sampling is the standard method
 and will allow results to be published.
 2 and allows for comparison between adults and children.

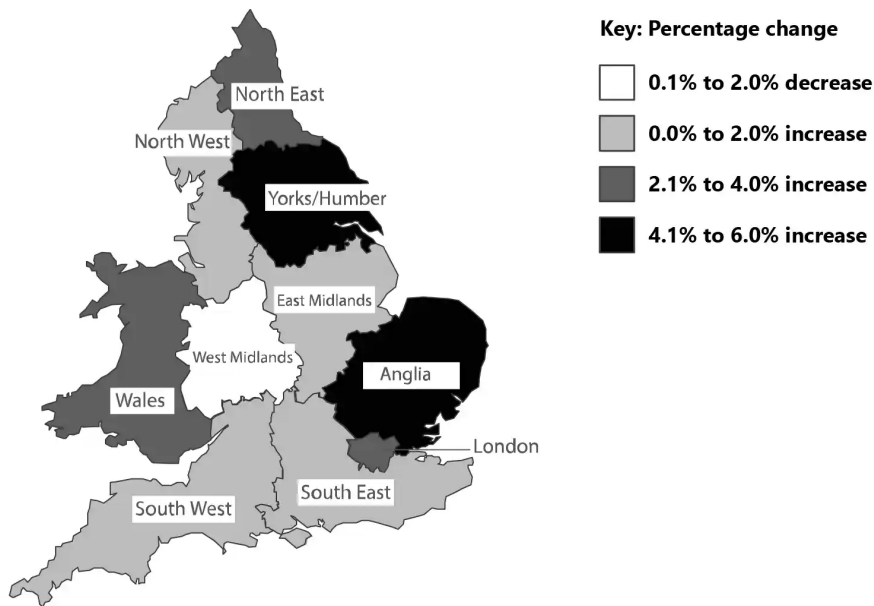
(c) Explain why the quota sample used by Thomas is not a random sample.

(1 mark)

Select **one** box.

- The sample was selected from a large population.
- The participants volunteered to take part.
- Only the people at the library that day are considered.
- He made sure the sample was balanced.

2 The map below shows the percentage change in domestic tourist visits across different regions of England and Wales between 2011 and 2012.



(a) Write down the percentage change in domestic tourist visits between 2011 and 2012 in South East.

(1 mark)

Select **one** box.

- 0.1% to 2.0% decrease
- 4.1% to 6.0% increase
- 2.1% to 4.0% increase
- 0.0% to 2.0% 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) Priscilla states that domestic tourism in England and Wales increased overall between 2011 and 2012.

Explain why this may not be the case.

(1 mark)

Select **one** box.

- The map does not show Scotland or Northern Ireland.
- More regions increased than decreased.
- More regions decreased than increased.
- We only know the percentages, not the amounts.

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

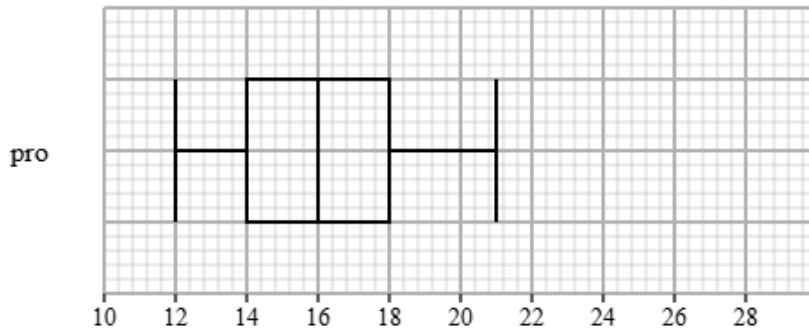
(1 mark)

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

C _ _ _ _ P L E T H

choropleth

- 3 Liam recorded the completion times for pro and beginner runners in a 5K race. Both groups ran the same course. The box plot presents data on the completion times for the pro runners.



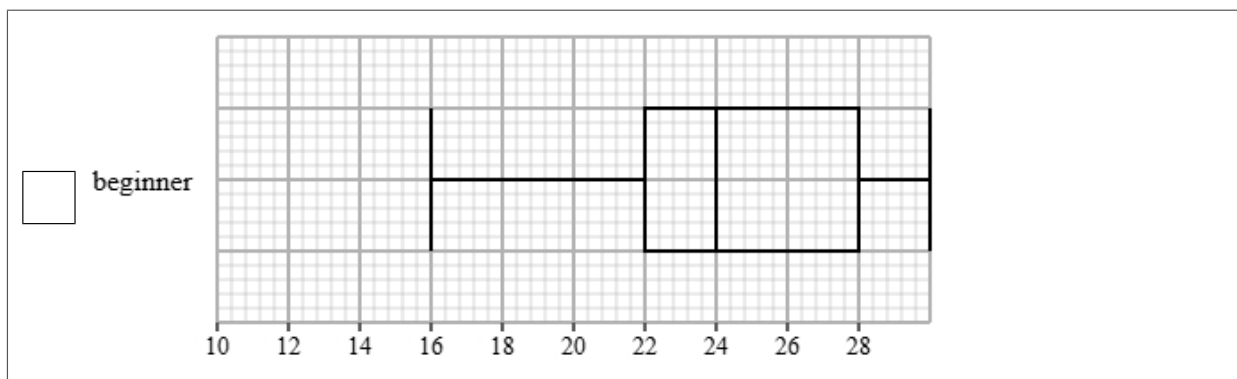
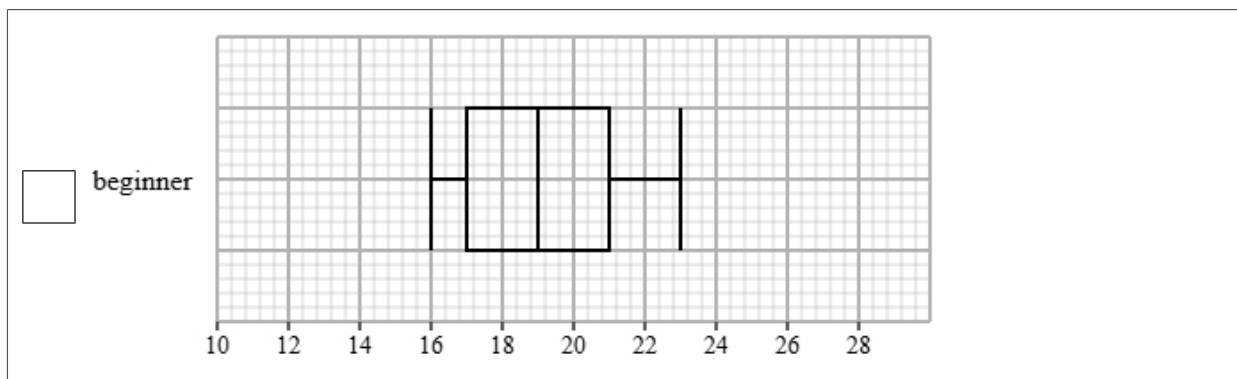
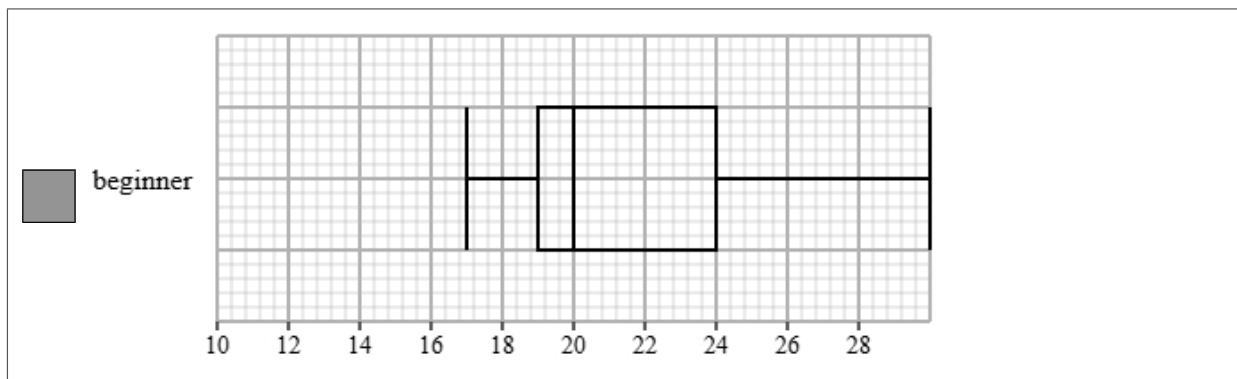
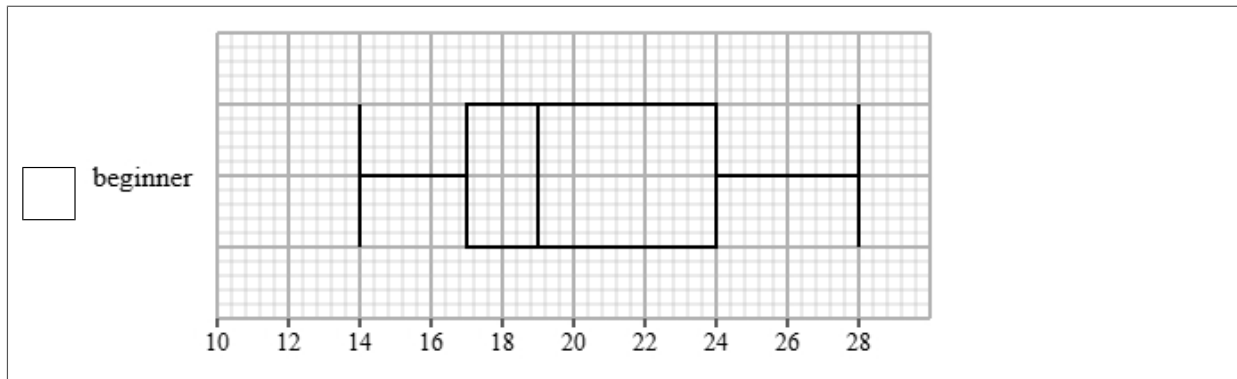
The table gives information about the completion times for the beginner runners.

Least tall	Lower quartile	Median	Upper quartile	Most tall
17	19	20	24	30

- (a) Draw a box plot for the completion times for the beginner runners.

(2 marks)

Select the correct answer.



(b) Compare the two distributions of completion times.

Give three comparisons and interpret one of these comparisons.

(4 marks)

Select **one** box.

- The median is bigger.
- The median completion times for pro runners is greater than beginner runners.
- The median completion times for pro runners is lower than beginner runners.
- The median completion times for pro and beginner runners are equal.

Select **one** box.

- The IQR is bigger.
- The IQR for the completion times of the pro runners is greater than beginner runners.
- The IQR for the completion times of the pro and beginner runners are equal.
- The IQR for the completion times of the pro runners is lower than beginner runners.

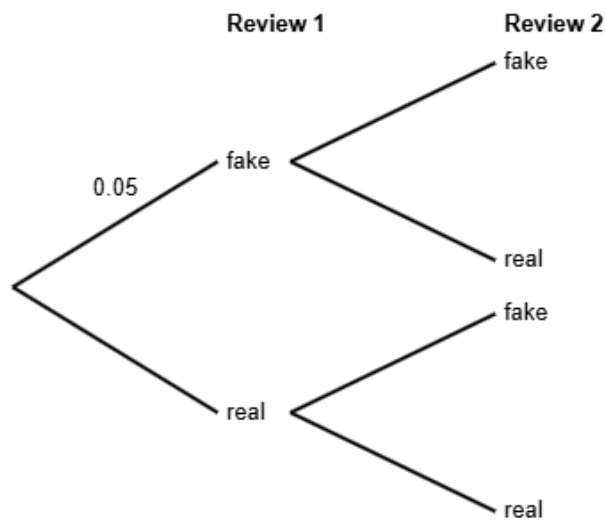
Select **one** box.

- The skews for the completion times of the pro and beginner runners are both positive.
- The skew for the completion times of the pro runners is symmetrical and the skew for the beginner runners is positive.
- The skews for the completion times of the pro and beginner runners are both symmetrical.
- The skew for the completion times of the pro runners is symmetrical and the skew for the beginner runners is negative.

Select **one** box.

- The times for the pro runners are more spread out than the beginner runners.
- The pro runners are on average slower than the beginner runners.
- The pro runners are on average faster than the beginner runners.
- The pro runners are more skewed than beginner runners.

- 4 Research suggests that 5% of online product reviews are fake.
All other reviews are genuine.
Emma is reading two reviews for a product.
She does not know if each review is fake or real.



- (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 reviews are real.

(2 marks)

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

0.9025

(c) Emma states that the probability that exactly one reviews is fake is less than 10%.
Find out whether or not Emma is correct.

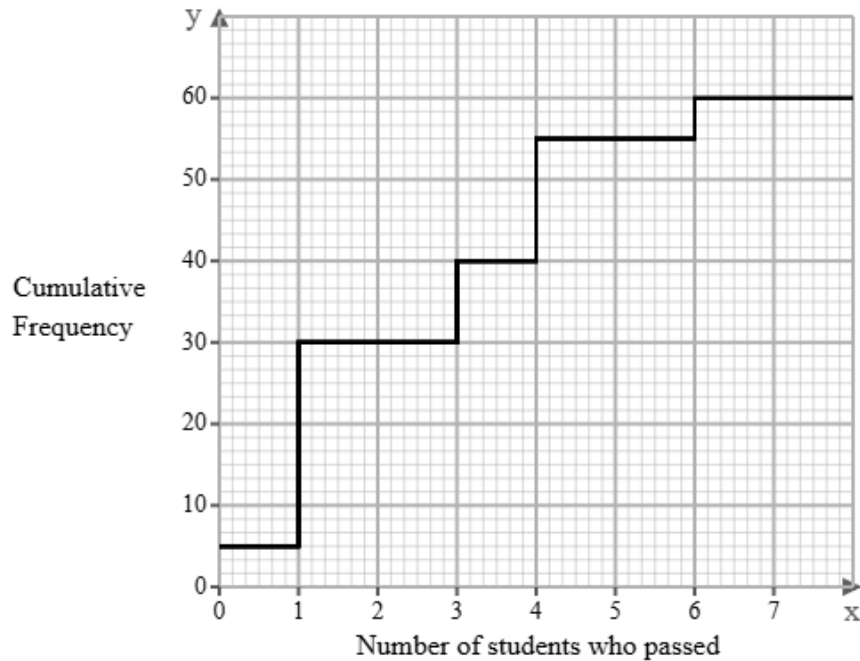
(3 marks)

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

Select **one** box.

- The probability that exactly one reviews is fake is less than 10%, so Emma is not correct.
- The probability that exactly one reviews is fake is more than 10%, so Emma is not correct.
- The probability that exactly one reviews is fake is more than 10%, so Emma is correct.
- The probability that exactly one reviews is fake is less than 10%, so Emma is correct.

- 5 The cumulative frequency step polygon shows information about the number of students who passed a daily maths quiz over 60 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 students who passed a daily maths quiz is continuous.
- Because the number of students who passed a daily maths quiz is qualitative.
- Because the number of students who passed a daily maths quiz is quantitative.
- Because the number of students who passed a daily maths quiz is discrete.

(b) Find the mode of the number of students who passed a daily maths quiz.

(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.

1

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

i) exactly 5 students who passed.

ii) more than 5 students who passed.

(3 marks)

The frequency is shown by how much the graph 'goes up' at each point.
Remember, the overall frequency is 60.

i) Exactly 5 students who passed: 0

ii) More than 5 students who passed: 5

(d) In 40 days fewer than x students passed.

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.

4

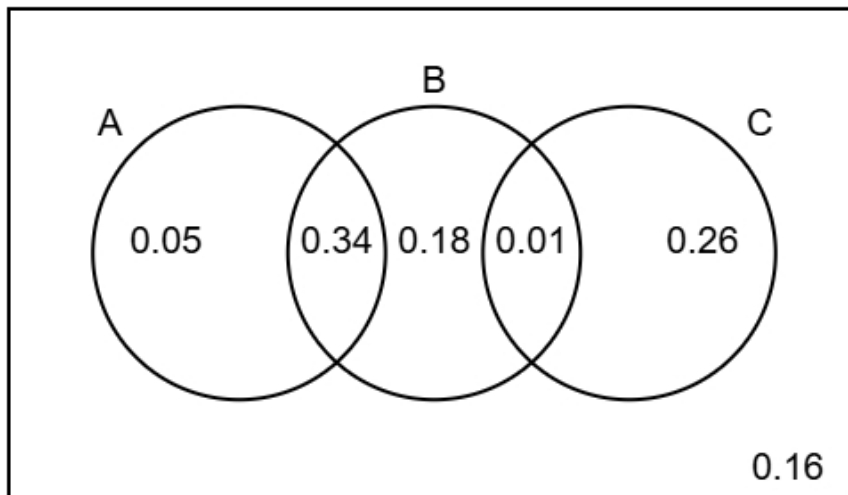
- (e) Peter believes the interquartile range of the number of students who passed is 8.
Explain why the interquartile range for this data cannot be 8.

(1 mark)

Select **one** box.

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

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 lowest total probability.
- B and C are mutually exclusive
- 1** A and C are mutually exclusive
- 2** because they do not overlap.
- A and B are mutually exclusive
- because they only overlap once.

(b) Find $P(B)$

(1 mark)

We are looking for the probabilities inside B.

0.53

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

(2 marks)

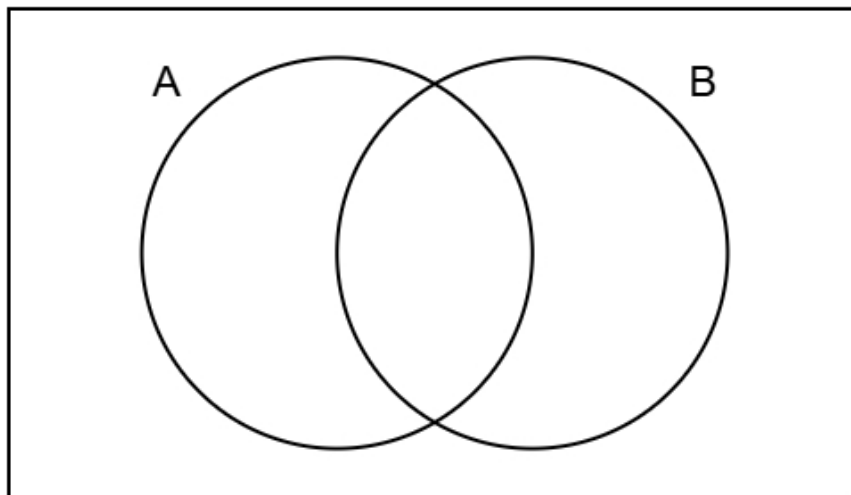
We are looking for the probabilities inside A or C.

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 Lucas is researching customer spend and average number of daily customers for 10 shops.

(a) Suggest a diagram that Lucas could draw to determine if there is a relationship between customer spend and average number of daily customers for the 10 shops.

(1 mark)

Select *one* box.

Scatter diagram

Pie chart

Box plot

Cumulative frequency graph

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

Shop	Daily Customers	Daily Customers Rank	Customer Spend Rank	d	d ²
Appleton Books	84	8	2	6	36
Baker's Treat	136	3	10	-7	49
Caramel Cafe	121	5	4	1	1
Daisy's Boutique	68	9	3	6	36
Eco Gadgets	57	10	1	9	81
Fashion Hub	125	4	5	-1	1
Gourmet Delights	142	2	9	-7	49
Happy Hands	96	6	7	-1	1
Posh Pets	92	7	6	1	1
Silver Spoon	150	1	8	-7	49

Select *two* boxes.

- The presence of outliers would decrease the correlation.
- As customer spend increases, average number of daily customers increases.
- As customer spend increases, average number of daily customers decreases.
- The presence of outliers would increase the correlation.

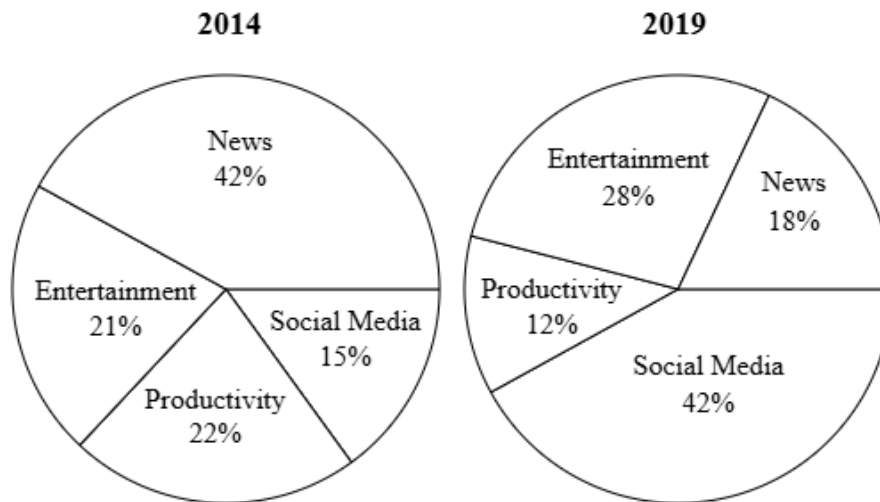
- (c) Discuss whether Lucas should have used Pearson's product moment correlation coefficient instead of Spearman's rank correlation coefficient to measure the correlation.

(3 marks)

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

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

8 The pie charts show a country's mobile phone app downloads in 2014 and 2019.



(a) Explain why the pie charts do not show a decrease in news apps downloads in 2019 compared to 2014.

(1 mark)

Select **one** box.

- We do not know the total budget spend for 2014 or 2019, only the proportions.
- Pie charts are inaccurate.
- Pie charts only display trends over time.
- There may be many other business factors that have changed between the two dates.

- (b) The number of news app downloads in 2014 was 1922130.
Find the number of social media app downloads in 2014.

(2 marks)

The number of news app downloads in 2014 was 1922130 which is shown by 42% on the pie chart.

$$42\% = 1922130$$

Find 1%

Then find the number of social media app downloads (15%).

686475

- (c) In 2014 the total number of app downloads was 4577000 (nearest thousand).
In 2019 the total number of app downloads was 55345000 (nearest thousand).

Santi wants to use the totals to draw pie charts.

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

(5 marks)

Select **one** box.

- Santi can use dual pie charts.
- Santi can use comparative pie charts.
- Santi can use a stem and leaf diagram.
- Santi can use merged pie charts.

Select **one** box.

- $\frac{55345000^2}{4577000^2}$
- $\sqrt{\frac{55345000}{4577000}}$
- $\left(\frac{55345000}{4577000}\right)^2$
- $\frac{55345000}{4577000}$

The radius of the 2014 pie chart will be 3.48 (2 d.p.) larger than the 2019 pie chart.

Select **one** box.

- This approach is faster than traditional pie chart methods.
- The areas will now decrease as the frequency increases, making the charts more intuitive.
- Each sector's proportion will vary depending on the total value of its respective chart.
- These pie charts have different totals, so their areas will reflect those totals while maintaining the same proportions within each chart.

9 A company produces chocolate chip cookies.

The cookies have a target mass of 60 g.

The company uses quality assurance to monitor the mass of each cookie.

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

The sample means should be normally distributed with a mean of 60 g and a standard deviation of 1.2 g.

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

(2 marks)

Use the formula for the upper action limit

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

63.6 g

(b) The upper action limit will be set closer to the target mass of 60 g.

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

(1 mark)

Select **one** box.

The production process will work faster.

The production process would reset more frequently.

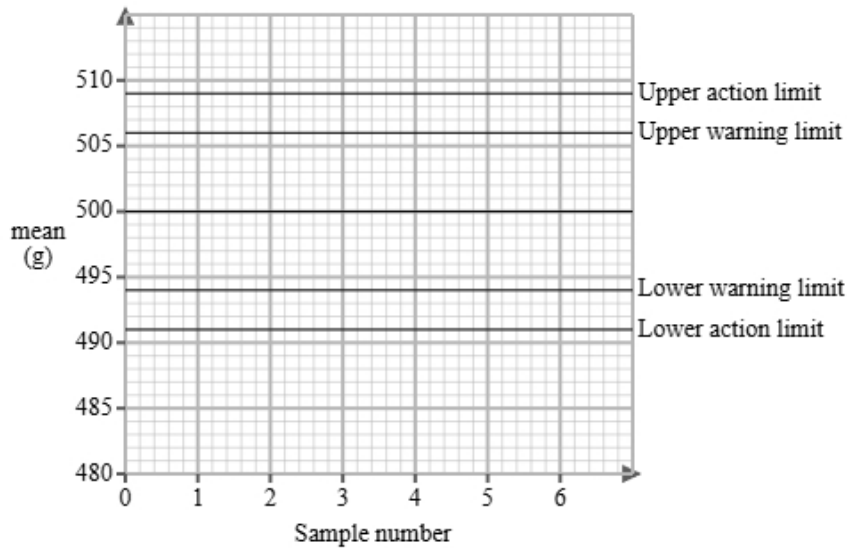
The production process would reset less frequently.

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

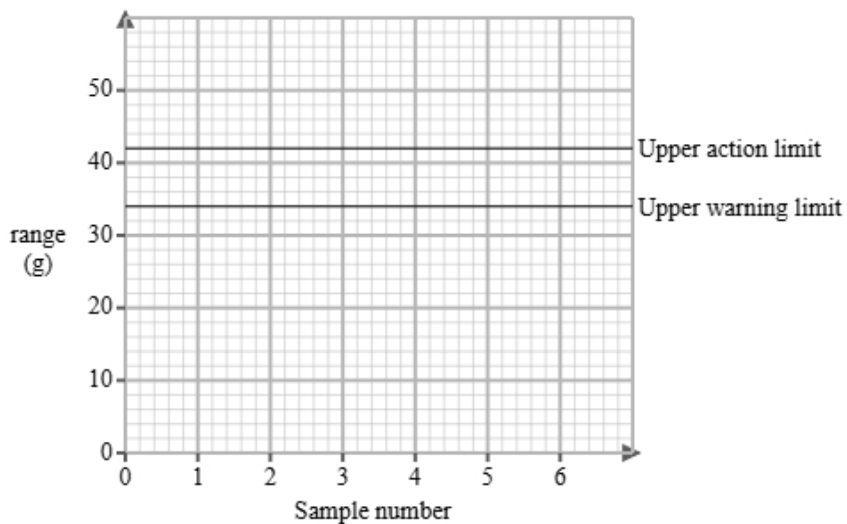
(c) The company also produces sponge cake and uses quality assurance to monitor the mass of each cake.

Here are the control charts for the sample means and for the sample ranges of the masses of the cakes.

Control chart for means



Control chart for ranges



A sample is taken and is found to have a mean of 496 g and a range of 31 g.

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

10 The reaction times of a group of adult drivers have a mean of 1.4 seconds and a standard deviation of 0.3 seconds.

(a) Mark is adult driver with a standardised score of 0.
Find Mark's reaction time.

(1 mark)

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

1.4 seconds

(b) Priya and Sarah are both adult drivers in the group.
Priya's standardised score for reaction time is 0.8 seconds.
Sarah's standardised score for reaction time is -1.2 seconds.
Priya had a slower reaction time than Sarah.
How much slower is Priya?

(3 marks)

Rearrange the formula to make the value the subject

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

0.6 seconds

(c) Mina takes a sample of 5 adult drivers from the group.

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

(i) Discuss the appropriateness of using 1.4 seconds as the mean in the calculation of the standardised score,

(ii) Discuss the appropriateness of using 0.3 seconds 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).

- 1 Using 1.4 seconds as the mean is appropriate
- 2 because the sample mean will be close to the population mean.
- because the sample mean will be smaller than the population mean.
- Using 1.4 seconds as the mean is not appropriate

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

- 1 Using 0.3 seconds as the standard deviation is not appropriate
- Using 0.3 seconds as the standard deviation is 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 Japan to find if there was a relationship between screen time and sleep duration of teenagers.

The researchers found the equations of the regression lines for the relationship between screen time (x hours) and sleep duration (y hours) for male teenagers and female teenagers on school nights and weekend nights.

The table below gives the equations of the regression lines.

	school nights	weekend nights
male teenagers	$y = -0.4x + 7.5$	$y = -0.2x + 8.2$
female teenagers	$y = -0.3x + 7.8$	$y = -0.15x + 8.5$

(a) Interpret in context the figure -0.15 in the regression equation for female teenagers on weekend nights.

(1 mark)

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

- 2** the sleep duration would decrease by 0.15 hours.
- 1** For each 1 hour increase in screen time
- For each 1 extra male teenager
- the sleep duration would increase by 0.15 hours.

(b) Compare the relationships between screen time and sleep duration in male and female teenagers. Include in your comparisons reference to whether it is a school night or weekend night.

(3 marks)

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

- As screen time increased, the sleep duration decreased more rapidly on weekend nights compared to school nights.
- For all teenagers, an increase in screen time led to a decrease in sleep duration.
- For all teenagers, an increase in screen time led to an increase in sleep duration.
- The sleep duration of female teenagers decreased more per hours of screen time than male teenagers.
- As screen time increased, the sleep duration decreased more rapidly on school nights compared to weekend nights.
- The sleep duration of male teenagers decreased more per hours of screen time than female teenagers.

(c) The researchers would like to use a normal distribution as a model for the sleep duration of male teenagers on school nights.

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 shows different heights, a normal distribution could be a suitable model.

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

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

1 If the mean, median and mode are equal

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

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

2 and 68% of data is within 1 standard deviation from the mean, a normal distribution could be a suitable model.

12 A fair coin is tossed 4 times.
The number of heads obtained is recorded.

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

(2 marks)

Select *two* boxes.

- The coin is biased.
- Each of the emails are independent.
- The chance of getting a head remains constant.
- A different coin is used for each trial.

(b) Calculate the probability, as a fraction, that all 4 of the coins land on heads.

(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 coins land on heads.

(3 marks)

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