

Statistics GCSE

Paper 1

Edexcel Higher - 2026

Higher Tier

Variant 3

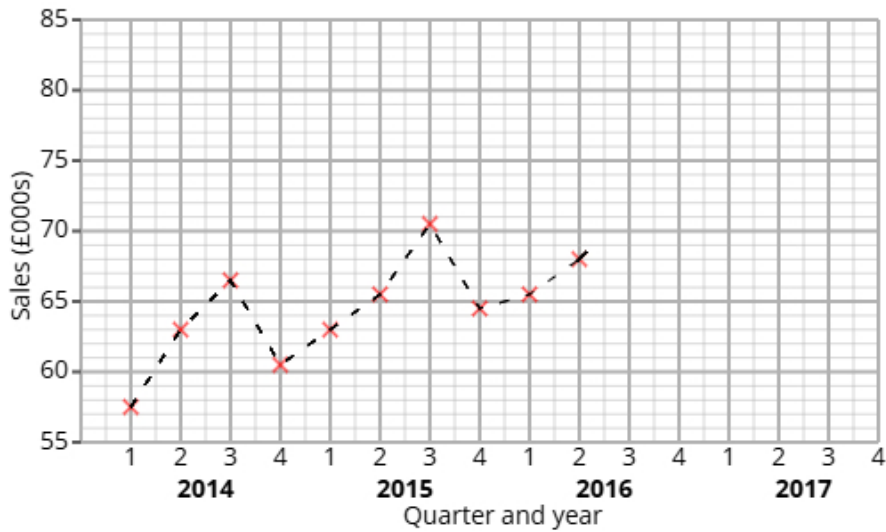
1ST0/1H

Answers

Visit our website for tutorials on each question.

www.statsgcse.com

1 The time series graph shows information about the sales at a clothes shop from 2014 to 2016.



Ava calculates the 4-point moving averages from the time series graph, which are shown below.

62 63 64 65 66 67 67

(a) Identify and interpret in context one example of seasonality displayed in the time series graph.

(2 marks)

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

- which shows that less people buy clothes in the summer.
- 1 The greatest values are in Q3
- 2 which shows that more people buy clothes in the summer.
- The lowest values are in Q2

(b) Ava uses the time series graph to estimate that there was £70000 of sales in Q1 2017

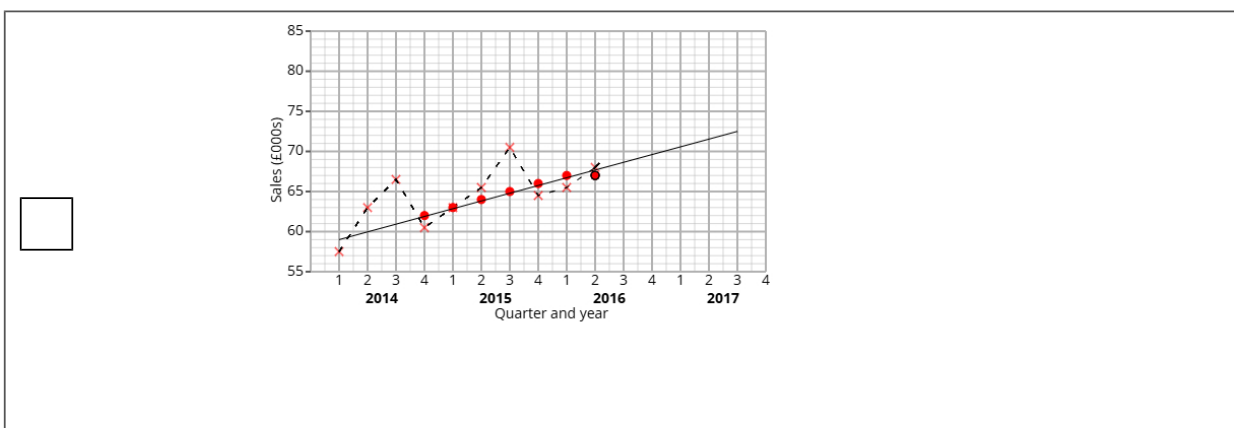
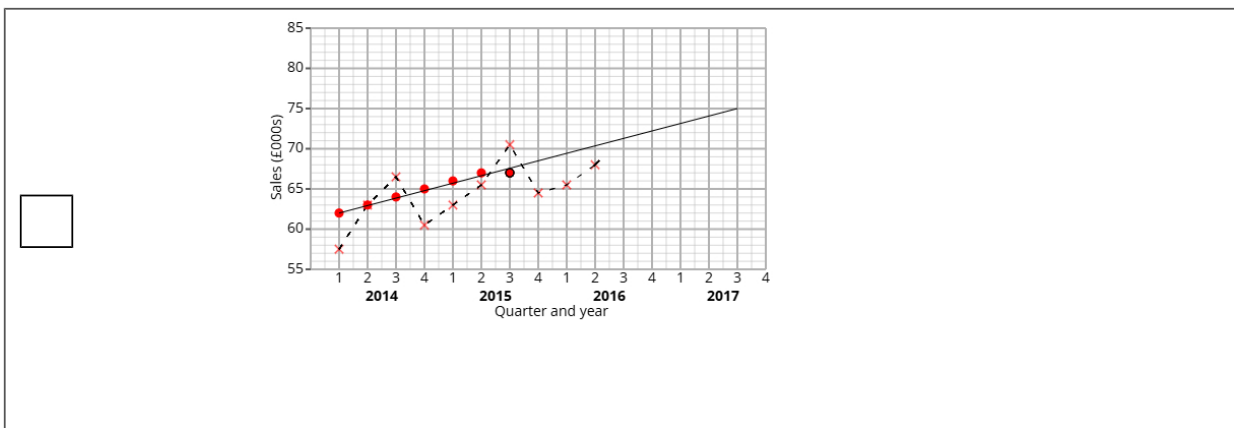
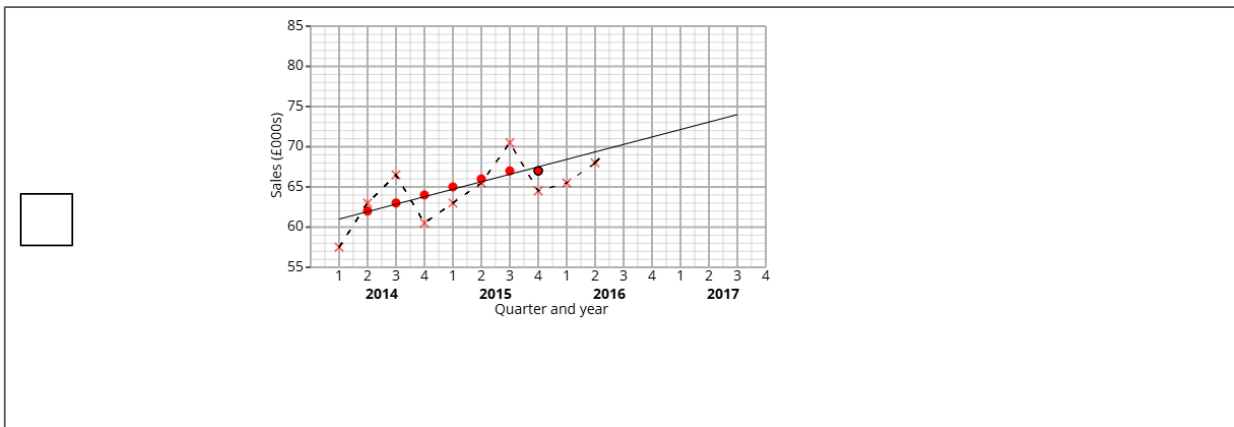
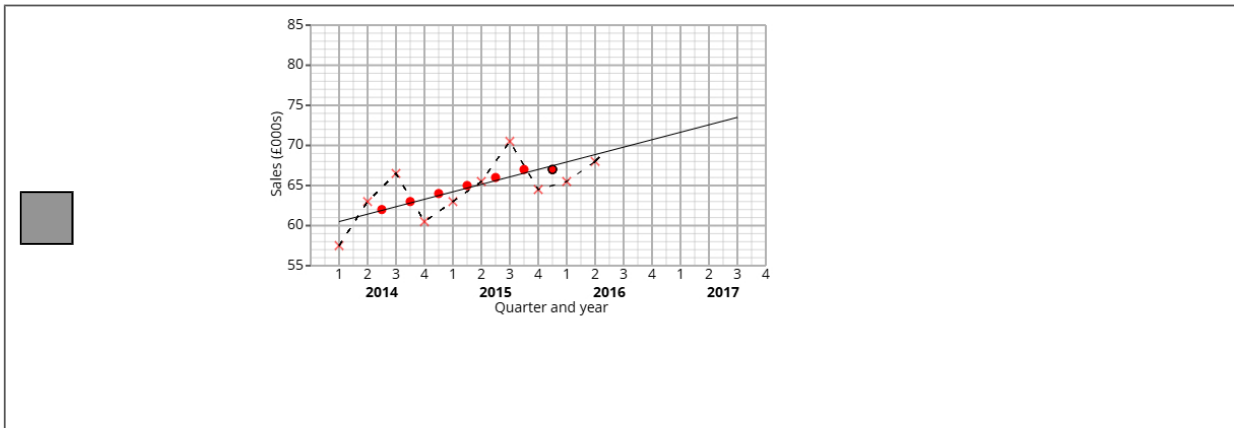
i) Plot the moving averages onto the time series graph and draw a trend line from 2014 to 2016.

ii) Describe the trend.

iii) By using the average seasonal effect for Q1, show that Ava's estimate is reasonable.

(7 marks)

Select the correct answer.



Select **one** box.

No trend

Upward trend

Downward trend

Flat trend

Select the correct answer.

$71.5 - \frac{57.5 + 63 + 65.5}{20} = 69.3$

$71.5 - \frac{(60.5 - 57.5) + (64 - 63) + (68 - 65.5)}{3} = 69.3$

$\frac{71.5 + 60.5 + 64 + 68}{4} = 69.3$

$\frac{71.5 + 57.5 + 63 + 65.5}{3} = 69.3$

(c) Explain why a 4-point moving average is appropriate.

(1 mark)

Select **one** box.

The pattern in the data repeats every four quarters.

It shows correlation.

4-points is more detailed than annual data.

It shows skew within the data.

2 A fair 4-sided spinner is numbered 1, 2, 3, 4.

A fair 3-sided spinner is numbered 1, 2, 3.

The spinners are used to play a game. Both spinners are spun and the total score is recorded.

| | | | | |
|-----------------|---|-----------------|---|---|
| | | 3-sided spinner | | |
| | | 1 | 2 | 3 |
| 4-sided spinner | 1 | 2 | 3 | |
| | 2 | 3 | | |
| | 3 | | | |
| | 4 | | | |

The game is won when the total is at least 5.

Kai plays the game once.

(a) Complete the sample space diagram.

(2 marks)

Select the correct answer.

3-sided spinner

| | | | |
|----------|----------|----------|----------|
| | 1 | 2 | 3 |
| 1 | 2 | 3 | 5 |
| 2 | 3 | 5 | 6 |
| 3 | 5 | 6 | 7 |
| 4 | 6 | 7 | 8 |

4-sided spinner

3-sided spinner

| | | | |
|----------|----------|----------|----------|
| | 1 | 2 | 3 |
| 1 | 2 | 3 | 4 |
| 2 | 3 | 4 | 5 |
| 3 | 4 | 5 | 6 |
| 4 | 5 | 6 | 7 |

4-sided spinner

3-sided spinner

| | | | |
|----------|----------|----------|----------|
| | 1 | 2 | 3 |
| 1 | 2 | 3 | 3 |
| 2 | 3 | 4 | 6 |
| 3 | 3 | 6 | 9 |
| 4 | 4 | 8 | 12 |

4-sided spinner

(b) Find the probability that Kai wins the game.

(2 marks)

$$\frac{1}{2}$$

- 3 Sara organises two painting classes, Class Red and Class Blue, to help people learn watercolour techniques. She wants to compare the two classes to see which improves skills more.

The table shows number of participants who passed or failed the painting challenge.

| | Passed | Failed | Total |
|------------|--------|--------|-------|
| Class Red | 24 | 8 | 32 |
| Class Blue | 35 | 25 | 60 |

- (i) Find the relative risk of failing the painting challenge having been in Class Red compared to Class Blue.
(ii) Give an interpretation of your answer to part (i).

(4 marks)

Write your answer as a decimal.

0.6

Select **one** box.

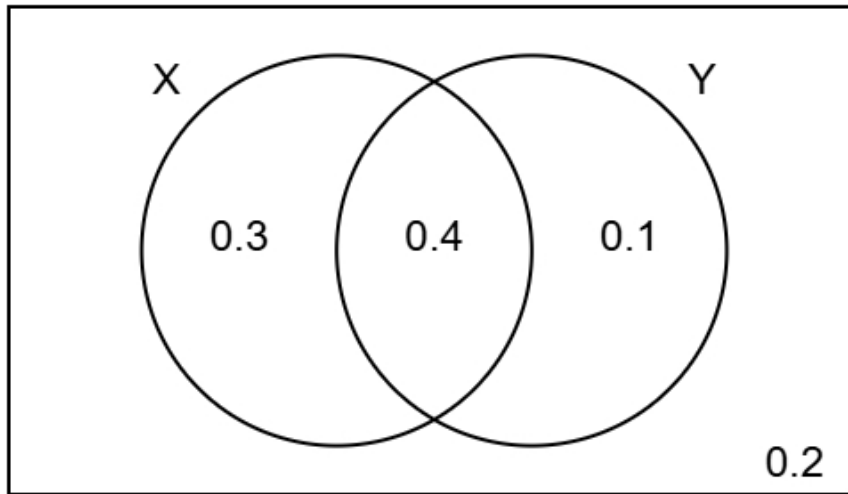
More people failed the painting challenge in Class Red than in Class Blue.

The risk of failing the painting challenge having taken Class Red is greater than the risk of failing the painting challenge having taken Class Blue.

The risk of failing the painting challenge having taken Class Red is lower than the risk of failing the painting challenge having taken Class Blue.

Less people failed the painting challenge in Class Red than in Class Blue.

- 4 The Venn diagram shows information about the probabilities of two events occurring.
The events are labelled as X and Y.



- (a) Find the probability of event Y happening.

(1 mark)

Leave your answer as a decimal.

_____ 0.5

- (b) Find $P(X \text{ and } Y)$

(1 mark)

Leave your answer as a decimal.

_____ 0.4

- (c) Find $P(Y | X)$

(2 marks)

$\frac{4}{7}$

(d) Two different events events R and S are independent.

$$P(R) = 0.4$$

$$P(S) = 0.8$$

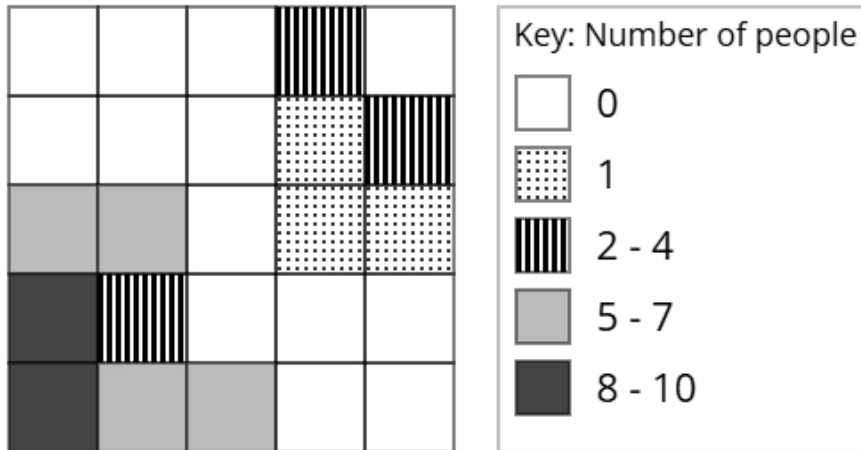
Find $P(R \text{ and } S)$

(2 marks)

Leave your answer as a decimal.

0.32

- 5 The choropleth map below represents a town square that has been divided into 25 squares of equal area. Aisha has collected data about the popularity of different parts of the town square. The number of people recorded in each square on one Thursday morning is shown.



- (a) Calculate an estimate of the total number of people that were recorded on Thursday.

(3 marks)

54

- (b) Aisha would like to open a flower stall in the town square.

After analysing the data, she decides that she should open the flower stall in the corner of the town square shown at the bottom left of the choropleth map.

Using the information in the choropleth map, assess the validity of Aisha's conclusion.

(2 marks)

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

- because there were less people at the bottom left of the town square.
- Aisha's comment is not valid
- because there were more people at the bottom left of the town square.
- Aisha's comment is valid

(c) Daniel argues that the method used by Aisha to collect the data is not appropriate for reaching a reliable conclusion.

Assess whether Daniel's argument is correct and give a reason.

(1 mark)

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

because the data was only collected on one Thursday.

Daniel is not correct

Daniel is correct

because there was a large amount of data collected.

6 A scientist is conducting an experiment to investigate the impact of sleep on reaction time. She plans to use a matched pairs design.

Marco is one of the participants in the study.

As part of the experiment, he takes four timed response tests.

Each test has a different weighting.

The table below shows the weightings and Marco's scores for each test.

| Test | Weighting | Score |
|------|-----------|-------|
| A | 3 | 22 |
| B | 2 | 19 |
| C | 4 | 27 |
| D | 1 | 14 |

(a) Explain the concept of matched pairs in an experimental design.

(2 marks)

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

- so that more results can be gathered with less data.
- 1 In a matched pairs design, subjects are matched in pairs so that each pair is as alike as possible
- 2 so that the experimenter can control for other factors.
- In a matched pairs design, the subject repeats the same experiment twice with one variable changing

(b) Calculate the weighted mean score for Marco's four tests.

(3 marks)

22.6

7 Carlos is trying to estimate the deer population in a deer park.

He first captures and tags 50 deer, then releases them.

One week later, he catches a second sample of 10 deer.

Using the Petersen capture-recapture method, he estimates the total number of deer to be 100.

(a) How many of the 100 deer in Carlos's second sample were tagged?

(2 marks)

5

(b) Discuss how reliable Carlos's estimate is by considering the assumptions required for using the Petersen capture-recapture method.

(3 marks)

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

1 Carlos has assumed that he selected a random sample, which is unlikely.

3 These assumptions mean that the estimate may not be reliable, because the assumptions are unlikely to hold.

2 He has also assumed that there were no deer that died, which is unlikely.

This means that the estimate will be reliable.

He has also collected a very large sample.

Carlos has assumed that there are only deer in the deer park.

8 A company team leader wants to find out if software developers have taken extended lunch breaks in the last 6 months.

Harry suggests using the random response technique to ask the developers if they have taken extended lunch breaks.

(a) Explain why Harry has suggested using the random response technique for this situation.

(1 mark)

Select **one** box.

Harry believed the random response technique would make the data easier to graph.

Harry used the random response technique to randomly select which participants could skip the question.

Harry thought the random response technique would speed up how quickly participants finished the survey.

Harry chose the random response technique to encourage honesty in responses to a potentially embarrassing topic.

(b) The company team leader uses the random response question below:

Think of a month of the year.

If the month is after June, do not read the question and tick box A.

If the month is June or before, answer the question truthfully.

Have you taken extended lunch breaks?

If yes, tick box A. If no, tick box B.

A B

State if this question is appropriate and give reasons for your answer.

(2 marks)

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

- This question is appropriate because
- so no one will know if they were forced to answer A.
- 2 the month people will pick is not random
- 1 This question is not appropriate because
- 3 and there is no time frame in the question.
- people are randomly picking a month

(c) The final questionnaire will be distributed to a sample of developers.

The developers are made up from those who work with front-end, back-end, and full-stack.

They work either full-time or part-time.

The table shows how many developers there are in each category

| | | Type | | |
|-------------------|-----------|-----------|----------|------------|
| | | Front-End | Back-End | Full-Stack |
| Employment status | Full-time | 88 | 72 | 68 |
| | Part-time | 33 | 78 | 24 |

The company team leader plans to take a stratified sample based on type and employment status and requires a minimum of 15 individuals from each stratum.

If the calculated sample size for a particular stratum is a decimal, he will round it to the nearest integer.

Determine the smallest total sample size that ensures at least 15 people are selected from each stratum.

(2 marks)

220

- 9 Taylor is investigating the profits made by two different shops, Oak & Co and Pine & Co.
Taylor has obtained the annual percentage profits made by Oak & Co for the years 2014 to 2018 and the annual percentage profits made by Pine & Co for the years 2015 to 2018.

The table below gives this information.

| Year | Percentage profit (%) | |
|------|-----------------------|-----------|
| | Oak & Co | Pine & Co |
| 2014 | 2.5 | |
| 2015 | 2.8 | 0.6 |
| 2016 | 3 | 1.4 |
| 2017 | 3.6 | 1.2 |
| 2018 | 4 | 4.8 |

Taylor concludes that the average annual percentage profit made by Pine & Co over the 4 years is greater than the average annual percentage profit made by Oak & Co over the 5 years.

By using appropriate geometric means, assess Taylor's conclusion.

You must show your working.

(5 marks)

Geometric mean for Oak & Co = 3.18% (2 decimal places)

Geometric mean for Pine & Co = 1.99% (2 decimal places)

Select **one** box.

Taylor's conclusion is correct.

Taylor's conclusion is not correct.

10 Daniel has collected data about the heights, in cm, of gymnasts in a school.

The table gives some of the percentiles of Daniel's data.

| Percentile | Height (cm) |
|------------|-------------|
| 97.5th | 171.5 |
| 80th | 165.1 |
| 60th | 161.9 |
| 40th | 159.1 |
| 20th | 155.9 |
| 5th | 151.5 |
| 2.5th | 149.5 |

(a) Find the 2.5th to 97.5th interpercentile range.

(1 mark)

_____ 22 cm

(b) One of the gymnasts from the sample is selected at random.

Find the probability that their height is between 151.5 cm and 165.1 cm.

(1 mark)

_____ 75 %

(c) Give a reason why it is appropriate for Daniel to use the mean and the standard deviation to summarise this data

(1 mark)

Select **one** box.

The distribution is negative.

The distribution is symmetric.

The distribution is positive.

The data is continuous.

- (d) Daniel claims that the heights of the gymnasts can be modelled using a normal distribution with mean 160.5 cm and standard deviation 5.5 cm.

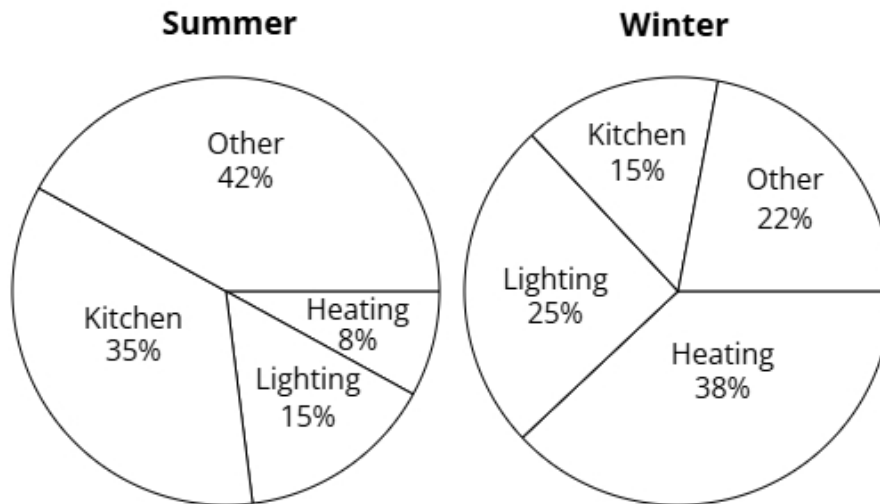
Use the data to assess the validity of Daniel's claim.

(4 marks)

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

- $160.5 + 2 \times 5.5 = 171.5$
- $97.5\% - 2.5\% = 95\%$ so the claim is valid.
- $160.5 - 3 \times 5.5 = 144$
- The distribution is symmetric so $\frac{161.9 + 159.1}{2} = 160.5$
- $160.5 + 3 \times 5.5 = 177$
- $80\% - 20\% = 60\%$ so the claim is not valid.
- 160.5 is not shown in the table, so is incorrect.
- $160.5 - 2 \times 5.5 = 149.5$

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



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.

12 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)

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.

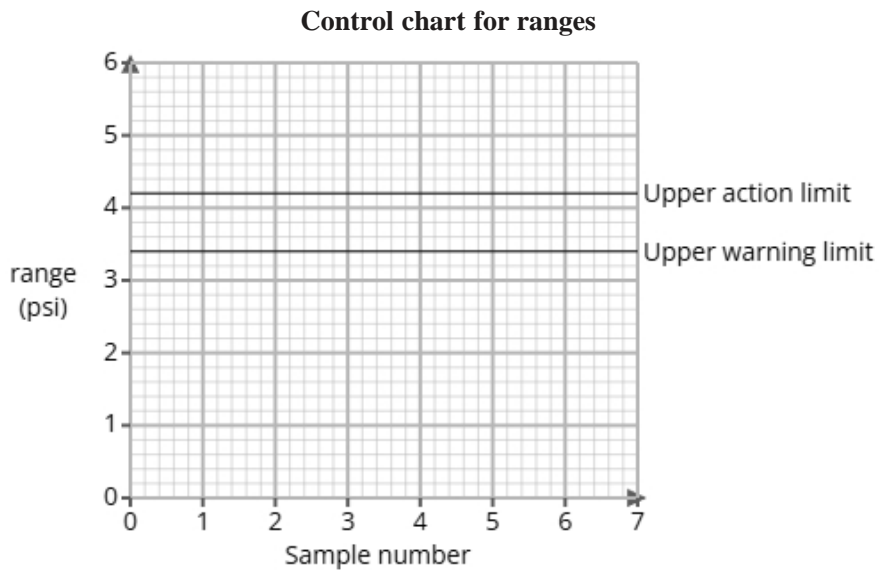
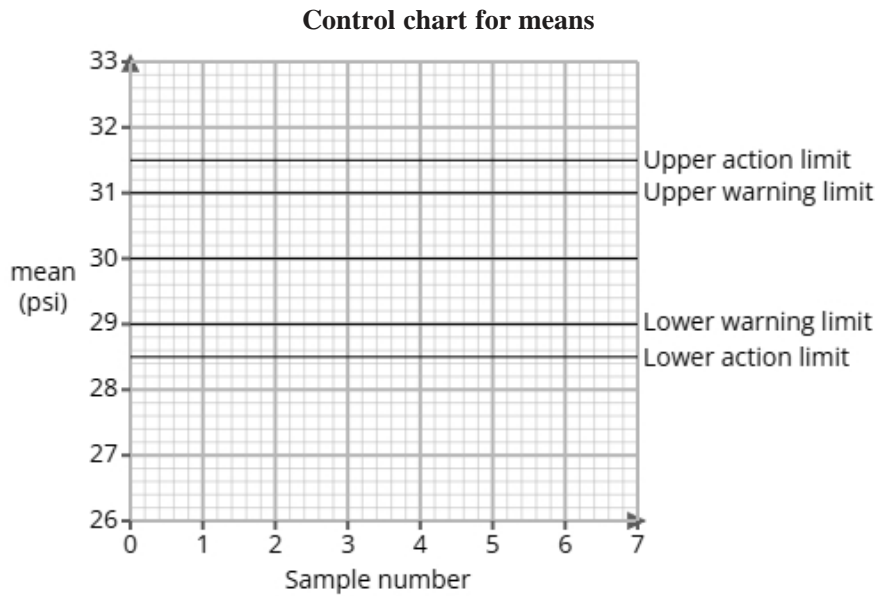
The production process would reset more frequently.

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.



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

13 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) 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.

(b) 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).

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.

If the mean, median and mode are equal

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

14 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)

$$\frac{1}{8000}$$

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

(3 marks)

$$\frac{29}{4000}$$