

Statistics GCSE

Paper 1

Edexcel Higher - 2026

Higher Tier

Variant 5

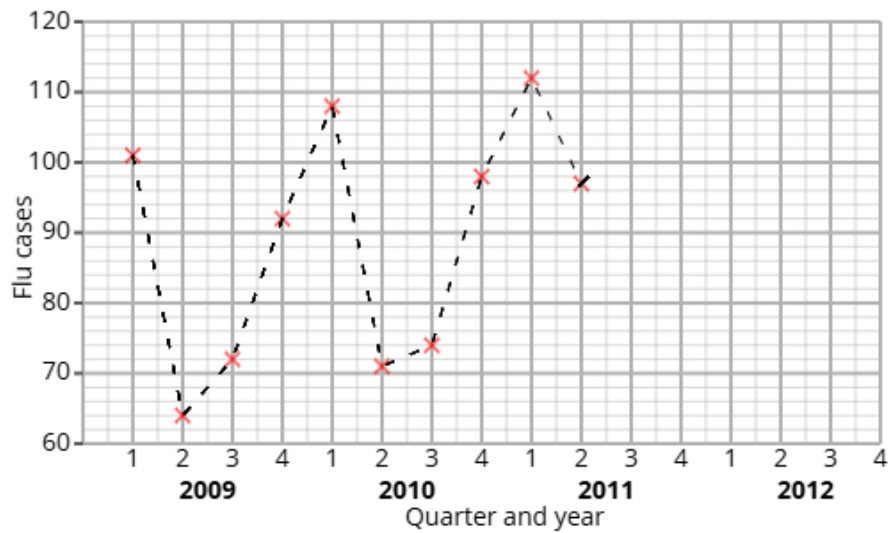
1ST0/1H

Answers

Visit our website for tutorials on each question.

www.statsgcse.com

- 1 The time series graph shows information about the the number of flu cases at a doctor's surgery from 2009 to 2011.



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

82 84 86 86 88 89 95

- (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 have flu in the winter.
- The lowest values are in Q1
- 1 The greatest values are in Q1
- 2 which shows that more people have flu in the winter.

(b) Ethan uses the time series graph to estimate that there were 119 flue cases in Q1 2012

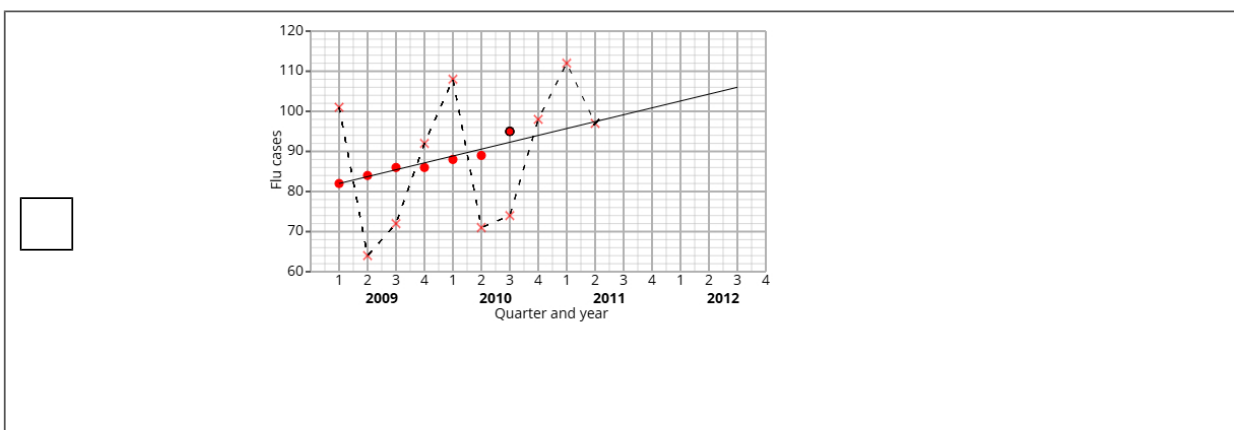
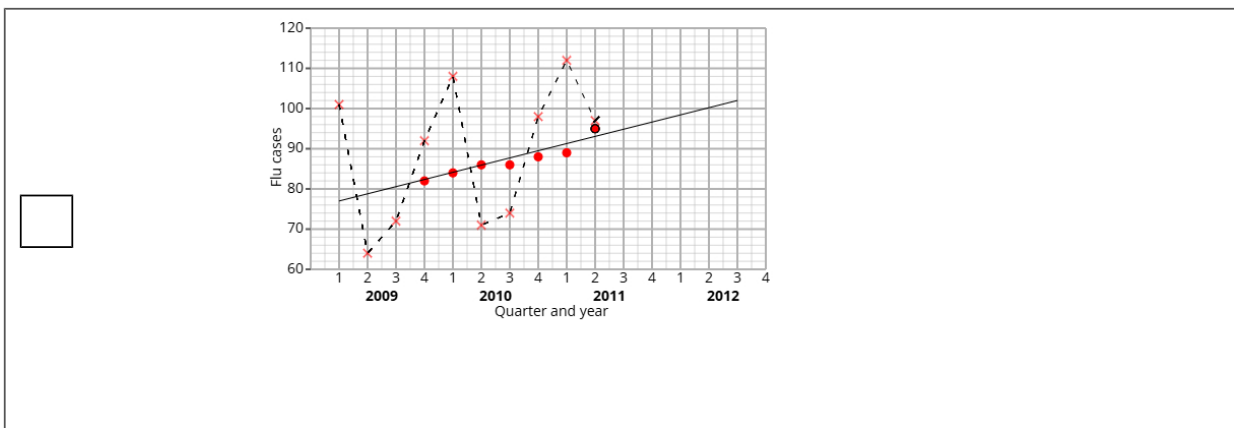
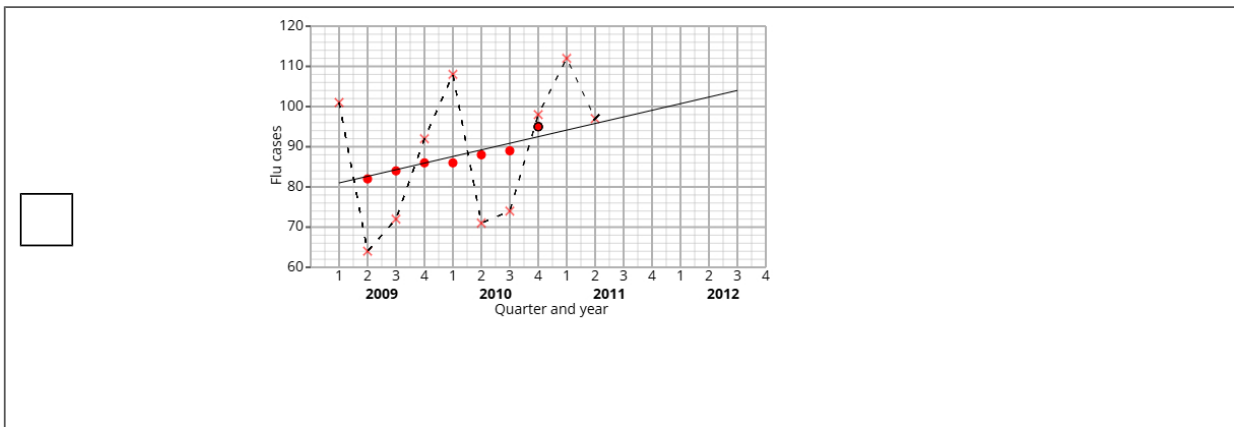
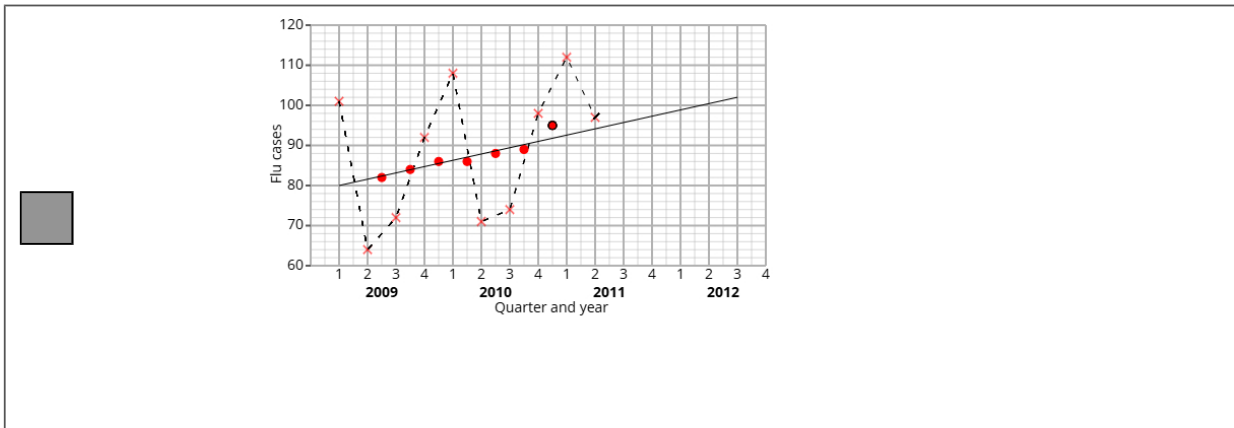
i) Plot the moving averages onto the time series graph and draw a trend line from 2009 to 2011.

ii) Describe the trend.

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

(7 marks)

Select the correct answer.



Select **one** box.

Upward trend

No trend

Downward trend

Flat trend

Select the correct answer.

$99 - \frac{101 + 108 + 112}{20} = 119.3$

$99 - \frac{(80 - 101) + (88 - 108) + (92 - 112)}{3} = 119.3$

$\frac{99 + 80 + 88 + 92}{4} = 119.3$

$\frac{99 + 101 + 108 + 112}{3} = 119.3$

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

(1 mark)

Select **one** box.

A 4-point moving average captures cyclic patterns that repeat every four quarters.

It gives more weight to the middle values.

It shows skew within the data.

A 4-point moving average gives us more data.

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

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

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

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

The game is won when the total is at least 6.

Noah plays the game once.

(a) Complete the sample space diagram.

(2 marks)

Select the correct answer.

<input type="checkbox"/>	3-sided spinner	4-sided spinner				
			1	2	3	4
		1	2	3	3	4
		2	3	3	4	5
3	3	4	5	6		

<input checked="" type="checkbox"/>	3-sided spinner	4-sided spinner				
			1	2	3	4
		1	2	3	4	5
		2	3	4	5	6
3	4	5	6	7		

<input type="checkbox"/>	3-sided spinner	4-sided spinner				
			1	2	3	4
		1	2	3	3	4
		2	3	4	6	8
3	3	6	9	12		

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

(2 marks)

$\frac{1}{4}$

3 Mia organises two language workshops, French A and French B, to help people learn conversational French.

She wants to compare the two workshops to see which improves conversation skills.

The table shows number of participants who passed or failed the conversation test.

	Passed	Failed	Total
French A	24	16	40
French B	25	50	75

(i) Find the relative risk of failing the conversation test having been in French A compared to French B.

(ii) Give an interpretation of your answer to part (i).

(4 marks)

Write your answer as a decimal.

_____ 0.6

Select **one** box.

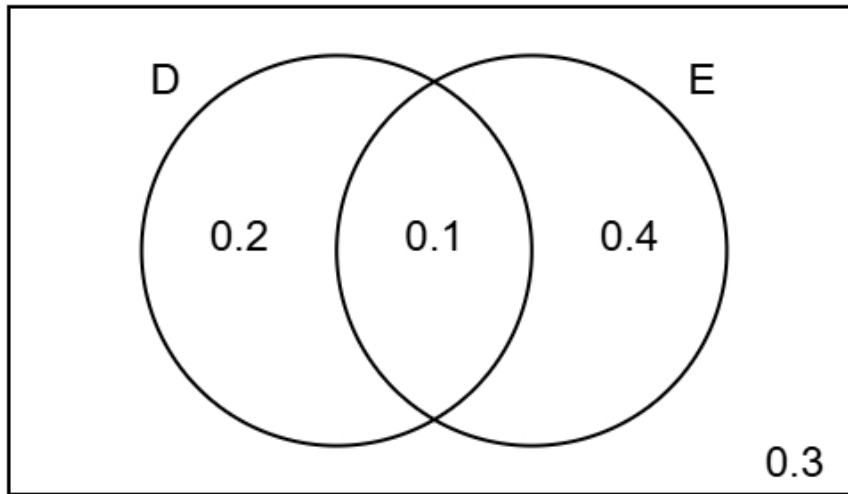
More people failed the conversation test in French A than in French B.

Less people failed the conversation test in French A than in French B.

The risk of failing the conversation test having taken French A is greater than the risk of failing the conversation test having taken French B.

The risk of failing the conversation test having taken French A is lower than the risk of failing the conversation test having taken French B.

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



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

(1 mark)

Leave your answer as a decimal.

_____ 0.5

- (b) Find $P(D \text{ and } E)$

(1 mark)

Leave your answer as a decimal.

_____ 0.1

- (c) Find $P(E | D)$

(2 marks)

$\frac{1}{3}$

(d) Two different events events A and B are independent.

$$P(A) = 0.3$$

$$P(B) = 0.9$$

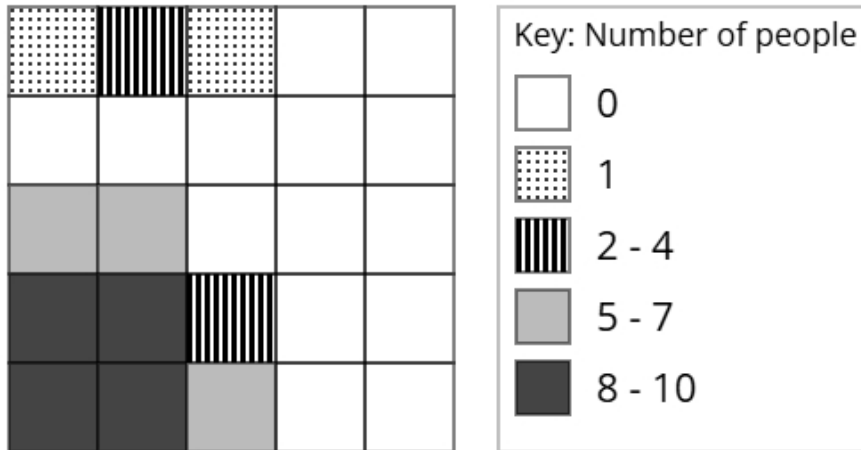
Find $P(A \text{ and } B)$

(2 marks)

Leave your answer as a decimal.

0.27

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



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

(3 marks)

62

- (b) Fatima would like to open a bagel stand in the university campus.

After analysing the data, she decides that she should open the bagel stand in the corner of the university campus shown at the bottom left of the choropleth map.

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

(2 marks)

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

because there were less people at the bottom left of the university campus.

Fatima's comment is not valid

because there were more people at the bottom left of the university campus.

Fatima's comment is valid

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

Assess whether Ethan'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 Friday.

Ethan is correct

Ethan is not correct

because there was a large amount of data collected.

6 A scientist is conducting an experiment to investigate the effect of background music on learning. She plans to use a matched pairs design.

Ben is one of the participants in the study.

As part of the experiment, he completes four learning assessments.

Each test has a different weighting.

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

Test	Weighting	Score
A	2	16
B	4	29
C	1	10
D	3	21

(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).

- In a matched pairs design, the subject repeats the same experiment twice with one variable changing
- 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 variables.
- so that more results can be gathered with less data.

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

(3 marks)

22.1

7 Jordan is trying to estimate the ducks population in a lake.

He first captures and tags 70 ducks, then releases them.

One week later, he catches a second sample of 14 ducks.

Using the Petersen capture-recapture method, he estimates the total number of ducks to be 196.

(a) How many of the 196 ducks in Jordan's second sample were tagged?

(2 marks)

5

(b) Discuss how reliable Jordan'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 Jordan has assumed that no ducks left or entered the lake.

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 ducks that died, which is unlikely.

This means that the estimate will be reliable.

He has also collected a very large sample.

Jordan has assumed that there are only ducks in the lake.

8 A factory supervisor wants to find out if machine operators have been taking unauthorized breaks in the last 4 months.

Jack suggests using the random response technique to ask the machine operators if they had taken any unauthorized breaks.

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

(1 mark)

Select **one** box.

Jack used the random response technique to help respondents feel safe giving honest answers.

Jack believed the random response technique was required for all types of surveys.

Jack thought the random response technique would be cheaper to do.

Jack thought the random response technique would let him choose which answers to include.

(b) The factory supervisor 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 taking any unauthorized 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.
- people are randomly picking a month
- 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.

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

The machine operators are made up from production machine, computer numerical control, and maintenance operators.

They work either full-time or part-time.

The table shows how many machine operators there are in each category

		Role		
		Production Machine	Computer Numerical Control	Maintenance
Employment status	Full-time	75	83	74
	Part-time	73	85	72

The factory supervisor plans to take a stratified sample based on role and employment status and requires a minimum of 25 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 25 people are selected from each stratum.

(2 marks)

158

- 9 Casey is investigating the profits made by two different shops, Horizon Foods and MetroFoods. Casey has obtained the annual percentage profits made by Horizon Foods for the years 2013 to 2017 and the annual percentage profits made by MetroFoods for the years 2014 to 2017.

The table below gives this information.

Year	Percentage profit (%)	
	Horizon Foods	MetroFoods
2013	2.7	
2014	3	0.9
2015	3.4	1.6
2016	3.9	1.3
2017	4.5	5.2

Casey concludes that the average annual percentage profit made by MetroFoods over the 4 years is greater than the average annual percentage profit made by Horizon Foods over the 5 years.

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

You must show your working.

(5 marks)

Geometric mean for Horizon Foods = 3.5% (2 decimal places)

Geometric mean for MetroFoods = 2.24% (2 decimal places)

Select **one** box.

Casey's conclusion is correct.

Casey's conclusion is not correct.

10 Marcus has collected data about the heights, in cm, of cyclists in a school.

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

Percentile	Height (cm)
97.5th	185.5
80th	178.0
60th	174.1
40th	170.9
20th	167.0
5th	161.8
2.5th	159.5

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

(1 mark)

_____ 26 cm

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

Find the probability that their height is between 161.8 cm and 167 cm.

(1 mark)

_____ 15 %

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

(1 mark)

Select **one** box.

The distribution is symmetric.

The data has been measured.

The data is continuous.

The distribution is exponential.

- (d) Marcus claims that the heights of the cyclists can be modelled using a normal distribution with mean 172.5 cm and standard deviation 6.5 cm.

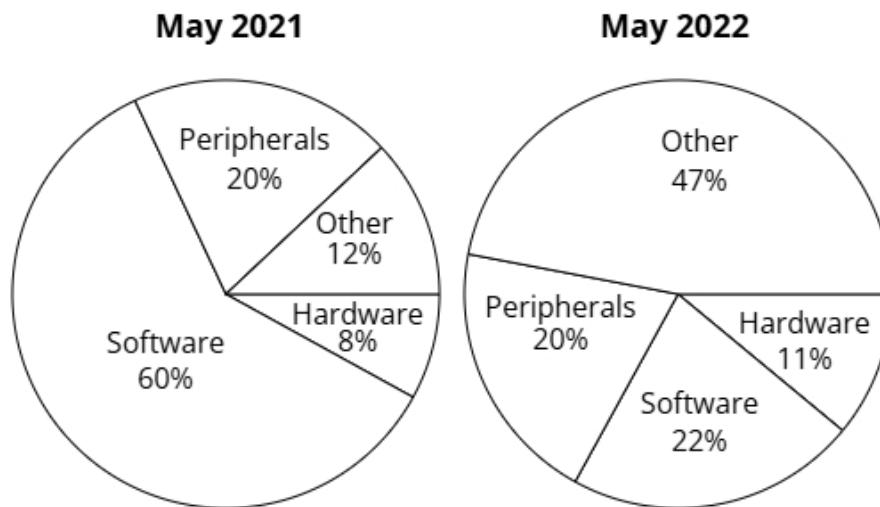
Use the data to assess the validity of Marcus's claim.

(4 marks)

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

- $172.5 + 2 \times 6.5 = 185.5$
- $172.5 - 2 \times 6.5 = 159.5$
- 172.5 is not shown in the table, so is incorrect.
- The distribution is symmetric so $\frac{174.1 + 170.9}{2} = 172.5$
- $172.5 - 3 \times 6.5 = 153$
- $80\% - 20\% = 60\%$ so the claim is not valid.
- $97.5\% - 2.5\% = 95\%$ so the claim is valid.
- $172.5 + 3 \times 6.5 = 192$

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



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 _____ 1.15 (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.

12 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)

_____ 830 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.

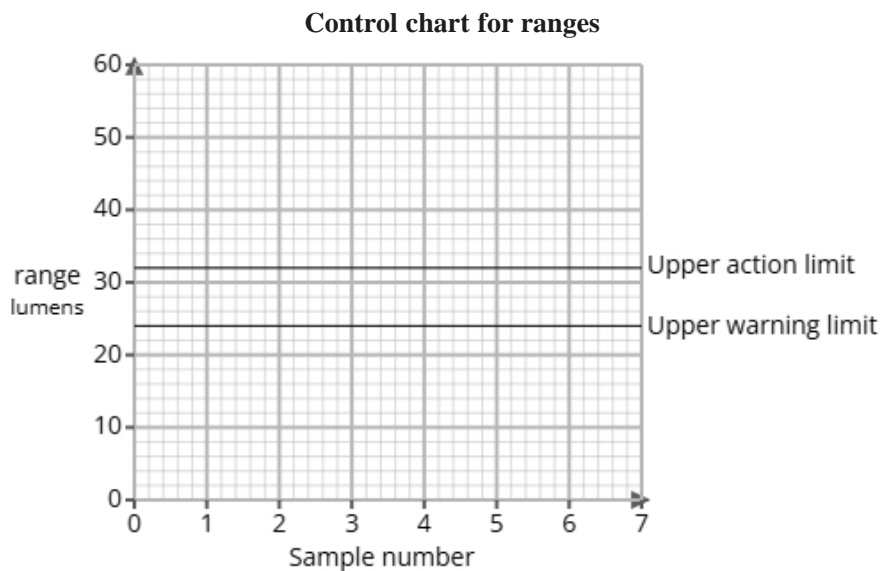
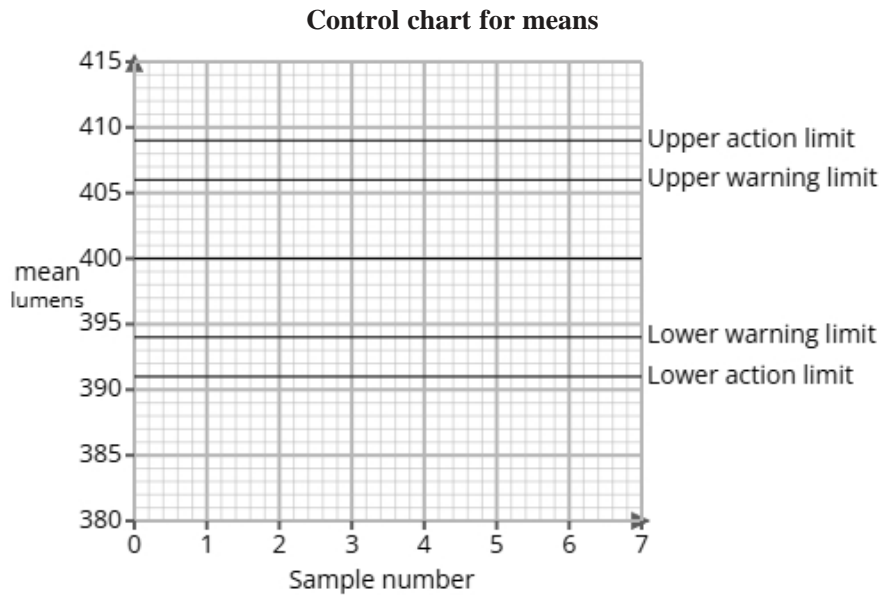
The production process would be stopped less frequently.

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.



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

13 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) 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.

(b) 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)

14 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)

$$\frac{1}{4096}$$

- (c) Calculate the probability, as a fraction, that at least 2 of the spinners land on an 8.

(3 marks)

$$\frac{323}{4096}$$