

**Statistics GCSE****Paper 1**

Edexcel Higher - 2026

Higher Tier

Variant 4

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.

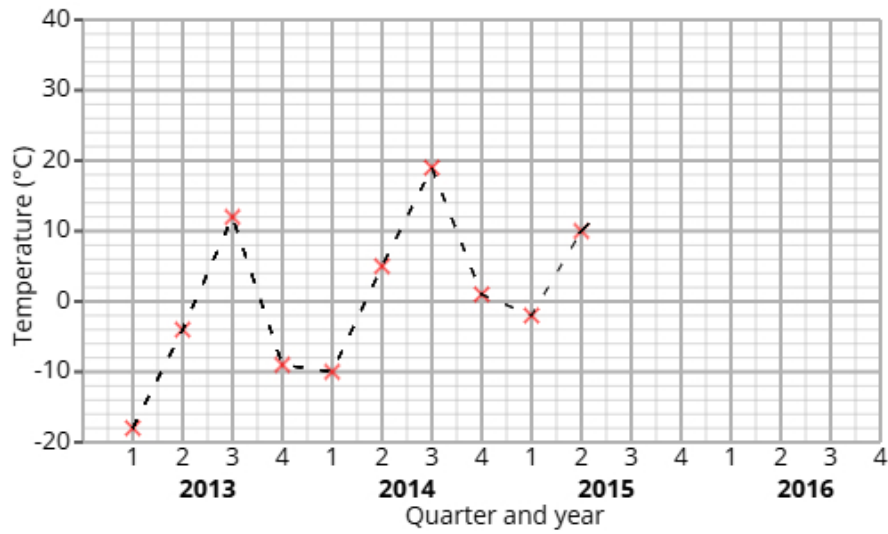
**Copyright**

- 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 The time series graph shows information about the the temperature of Eastern Galm from 2013 to 2015.



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

-5      -3      -1      1      4      6      7

(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 the lowest temperature is in the summer.
- which shows that the highest temperature is in the summer.
- The lowest values are in Q2
- The greatest values are in Q3

(b) Chloe uses the time series graph to estimate that there was a temperature of  $6^{\circ}\text{C}$  in Q1 2016

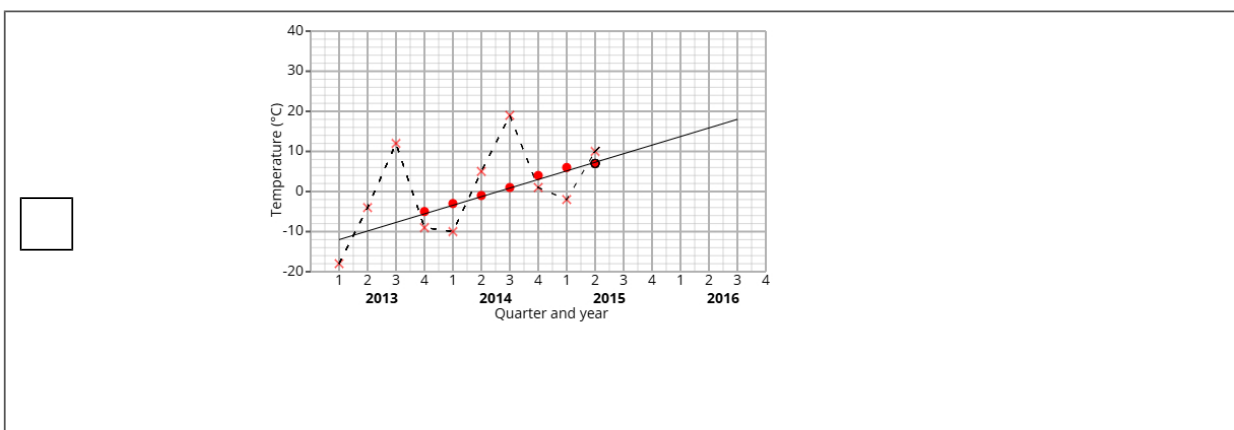
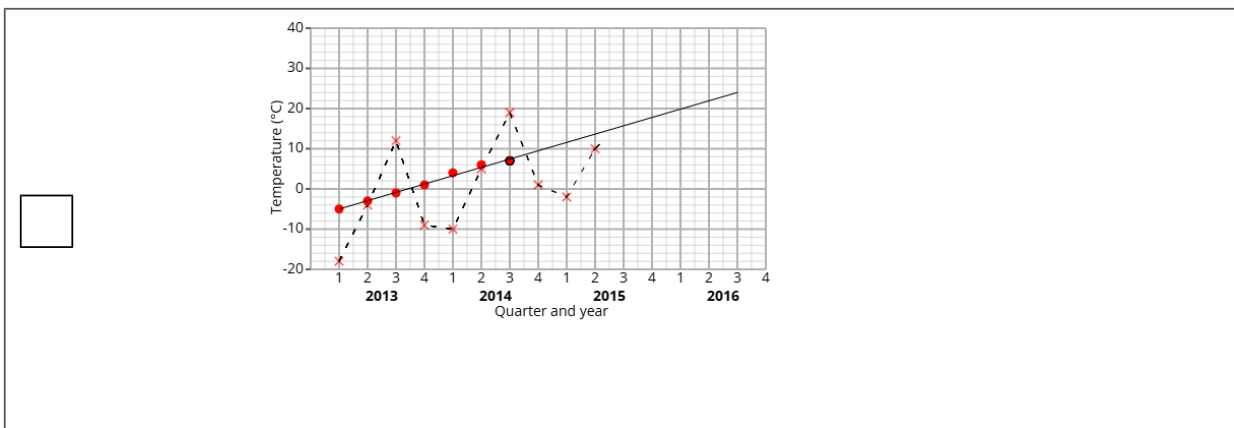
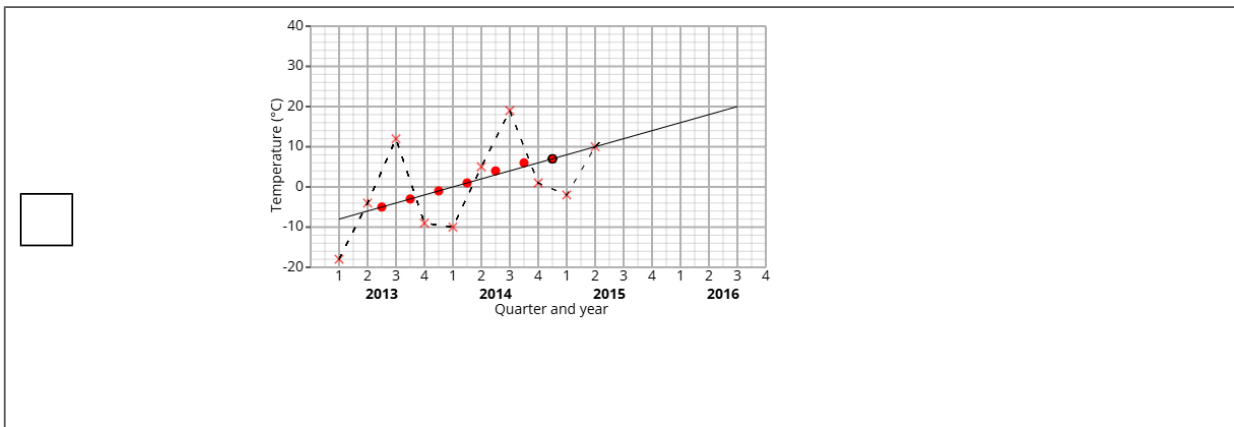
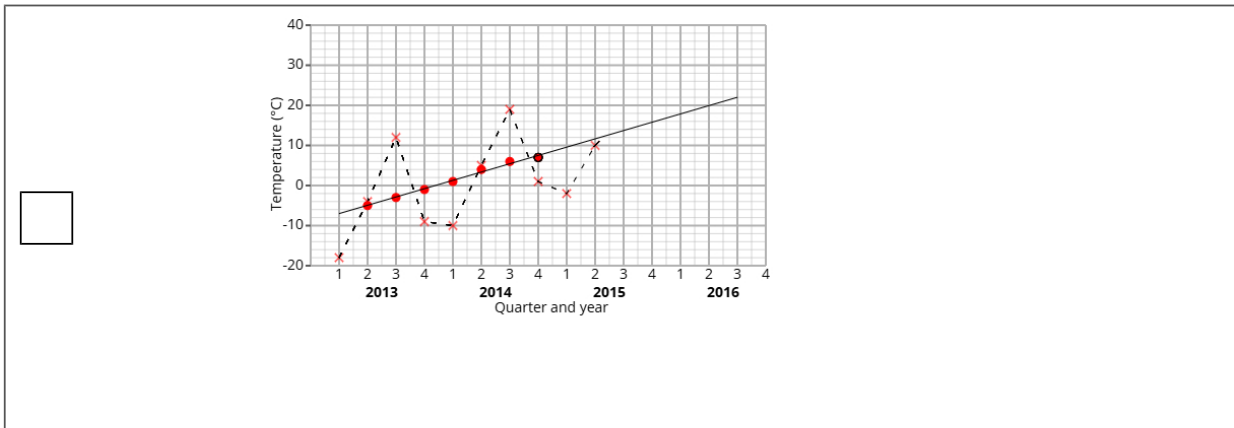
i) Plot the moving averages onto the time series graph and draw a trend line from 2013 to 2015.

ii) Describe the trend.

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

(7 marks)

Select the correct answer.



Select **one** box.

- Upward trend
- Downward trend
- No trend
- Flat trend

Select the correct answer.

$16 - \frac{(-8 - -18) + (0 - -10) + (8 - -2)}{3} = 6$

$\frac{16 + -8 + 0 + 8}{4} = 6$

$16 - \frac{-18 + -10 + -2}{20} = 6$

$\frac{16 + -18 + -10 + -2}{3} = 6$

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

(1 mark)

Select **one** box.

- It shows skew within the data.
- It shows correlation.
- A 4-point moving average gives us more data.
- A 4-point moving average captures cyclic patterns that repeat every four quarters.

2 A fair 5-sided spinner is numbered 1, 2, 3, 4, 5.

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
5-sided spinner	1	2	3	
	2	3		
	3			
	4			
	5			

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

Sofia plays the game once.

(a) Complete the sample space diagram.

(2 marks)

Select the correct answer.

		3-sided spinner		
		1	2	3
5-sided spinner	1	2	3	3
	2	3	3	4
	3	3	4	5
	4	4	5	6
	5	5	6	7

		3-sided spinner		
		1	2	3
5-sided spinner	1	2	3	3
	2	3	4	6
	3	3	6	9
	4	4	8	12
	5	5	10	15

		3-sided spinner		
		1	2	3
5-sided spinner	1	2	3	4
	2	3	4	5
	3	4	5	6
	4	5	6	7
	5	6	7	8

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

(2 marks)

Find all the numbers in the table that are 6 or larger

Put this number as the numerator and total amount of numbers as the denominator

$$\text{probability} = \frac{\text{6 or larger}}{\text{total outcomes}}$$

- 3 Noah organises two coding bootcamps, Bootcamp 1 and Bootcamp 2, to teach basic Python programming. He wants to compare the two bootcamps to see which teaches Python better.

The table shows number of participants who passed and failed the coding assessment.

	Passed	Failed	Total
Bootcamp 1	24	28	52
Bootcamp 2	15	50	65

(i) Find the relative risk of failing the coding assessment having been in Bootcamp 1 compared to Bootcamp 2.

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

(4 marks)

Write your answer as a decimal.

\_\_\_\_\_

Select **one** box.

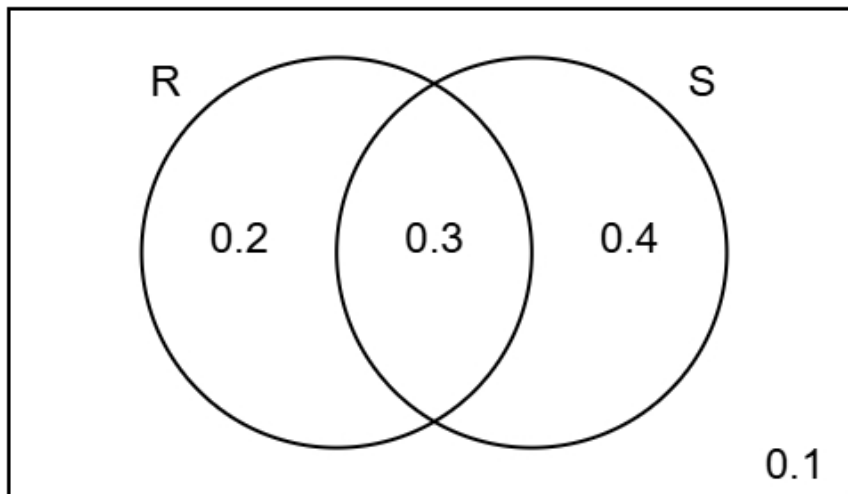
More people failed the coding assessment in Bootcamp 1 than in Bootcamp 2.

The risk of failing the coding assessment having taken Bootcamp 1 is lower than the risk of failing the coding assessment having taken Bootcamp 2.

Less people failed the coding assessment in Bootcamp 1 than in Bootcamp 2.

The risk of failing the coding assessment having taken Bootcamp 1 is greater than the risk of failing the coding assessment having taken Bootcamp 2.

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



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

(1 mark)

Add the probabilities in the circle marked S together

Leave your answer as a decimal.

\_\_\_\_\_

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

(1 mark)

$P(R \text{ and } S)$  is shown in the overlap of the Venn diagram

Leave your answer as a decimal.

\_\_\_\_\_

(c) Find  $P(S | R)$

(2 marks)

Use the formula to find  $P(S | R)$

$$P(S | R) = \frac{P(R \text{ and } S)}{P(R)}$$

(d) Two different events events D and E are independent.

$$P(D) = 0.2$$

$$P(E) = 0.6$$

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

(2 marks)

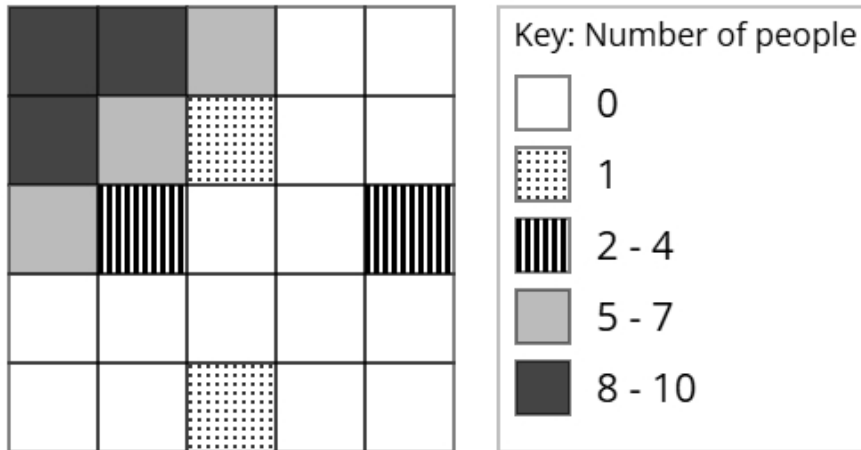
Use the formula for independent events

$$P(D \text{ and } E) = P(D) \times P(E)$$

Leave your answer as a decimal.

\_\_\_\_\_

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



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

(3 marks)

Find the midpoints for the groups.

Multiply each key with amount of squares and add them up.

(b) Daniel would like to open a bubble tea booth in the airport terminal.

After analysing the data, he decides that he should open the bubble tea booth in the corner of the airport terminal shown at the top left of the choropleth map.

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

(2 marks)

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

- because there were less people at the top left of the airport terminal.
- Daniel's comment is valid
- Daniel's comment is not valid
- because there were more people at the top left of the airport terminal.

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

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

(1 mark)

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

- Fatima is not correct
- because the data was only collected on one Monday.
- Fatima is correct
- because there was a large amount of data collected.

6 A scientist is conducting an experiment to investigate whether meditation improves focus. She plans to use a matched pairs design.

Priya is one of the participants in the study.

As part of the experiment, she completes four focus-related tasks.

Each test has a different weighting.

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

Test	Weighting	Score
A	4	30
B	1	12
C	2	20
D	3	27

(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, subjects are matched in pairs so that each pair is as alike as possible
- In a matched pairs design, the subject repeats the same experiment twice with one variable changing
- 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 Priya's four tests.

(3 marks)

Use the formula

$$\text{weighted mean} = \frac{\sum w \times s}{\sum w}$$

7 Mei is trying to estimate the squirrels population in a park.  
She first captures and tags 40 squirrels, then releases them.  
One week later, she catches a second sample of 16 squirrels.  
Using the Petersen capture-recapture method, she estimates the total number of squirrels to be 320.

(a) How many of the 320 squirrels in Mei's second sample were tagged?

(2 marks)

Use the capture recapture formula (you will need to rearrange it)

$$N = \frac{M \times n}{m}$$

where:

N = estimated total population

M = number of individuals originally marked (first capture)

n = total number of individuals captured in the second sample

m = number of marked individuals recaptured in the second sample

(b) Discuss how reliable Mei'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).

- These assumptions mean that the estimate may not be reliable, because the assumptions are unlikely to hold.
- She has also collected a very large sample.
- She has also assumed that there were no squirrels that died, which is unlikely.
- This means that the estimate will be reliable.
- Mei has assumed that there are only squirrels in the park.
- Mei has assumed that no squirrels left or entered the park.

**8** A retail store manager wants to find out whether employees have shared their employee discount with friends or family in the last 5 months.

Isla suggests using the random response technique to ask the employees have shared their employee discount.

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

(1 mark)

Select **one** box.

- Isla used the random response technique to use a more fun and engaging way to collect responses.
- Isla used the random response technique to reduce the risk of people lying on a sensitive question.
- Isla thought the random response technique would avoid using consent forms.
- Isla assumed the random response technique would automatically correct errors in the data.

(b) The retail store manager 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 shared your employee discount with friends or family?

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).

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

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

The employees are made up from stockroom, sales, and customer service assistants.

They work either full-time or part-time.

The table shows how many employees there are in each category

		Role		
		Stockroom	Sales	Customer Service
Employment status	Full-time	44	32	40
	Part-time	28	28	22

The retail store manager plans to take a stratified sample based on role and employment status and requires a minimum of 5 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 5 people are selected from each stratum.

(2 marks)

Find the smallest group

Decimals are rounded up, so find the smallest sample for this group before rounding it to 5

Use the stratified sampling formula to find  $n$  (the total number required in the sample)

- 9 Morgan is investigating the profits made by two different shops, Summit Stores and ValleyMart. Morgan has obtained the annual percentage profits made by Summit Stores for the years 2017 to 2021 and the annual percentage profits made by ValleyMart for the years 2018 to 2021.

The table below gives this information.

Year	Percentage profit (%)	
	Summit Stores	ValleyMart
2017	1.1	
2018	1.5	1.5
2019	1.9	2
2020	2.6	1.8
2021	3.3	6.8

Morgan concludes that the average annual percentage profit made by ValleyMart over the 4 years is greater than the average annual percentage profit made by Summit Stores over the 5 years.

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

You must show your working.

(5 marks)

$$\text{geometric mean} = \sqrt[n]{\text{value}_1 \times \text{value}_2 \times \dots \times \text{value}_n}$$

You will need to convert all the percentage increases into multipliers

Select **one** box.

Morgan's conclusion is not correct.

Morgan's conclusion is correct.

**10** Sofia has collected data about the heights, in cm, of swimmers in a school.

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

Percentile	Height (cm)
97.5th	190.5
80th	181.8
60th	177.4
40th	173.6
20th	169.2
5th	163.2
2.5th	160.5

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

(1 mark)

The 2.5th to 97.5th interpercentile range is the distance between those percentiles

\_\_\_\_\_ cm

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

Find the probability that their height is between 163.2 cm and 190.5 cm.

(1 mark)

Find the percentiles for 163.2 cm and 190.5 cm

The probability will be the distance between the percentiles

\_\_\_\_\_ %

- (c) Give a reason why it is appropriate for Sofia 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 data is continuous.
- The data has been measured.

- (d) Sofia claims that the heights of the swimmers can be modelled using a normal distribution with mean 175.5 cm and standard deviation 7.5 cm.

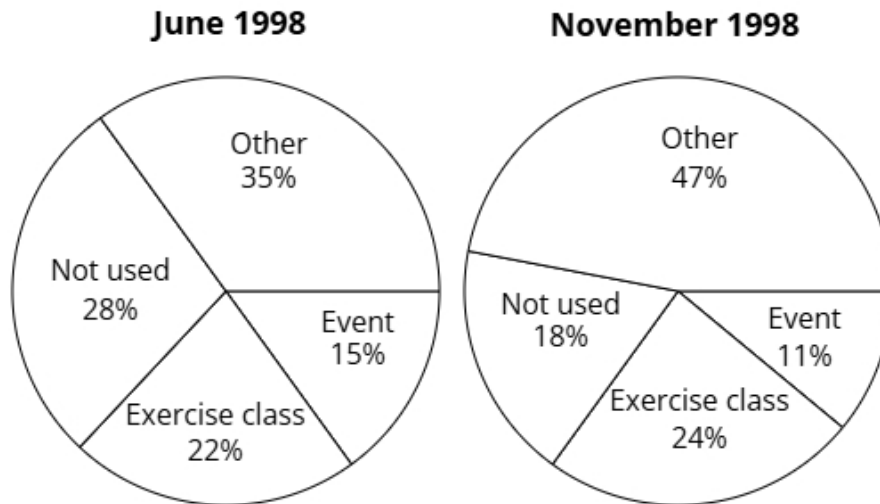
Use the data to assess the validity of Sofia's claim.

(4 marks)

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

- 175.5 is not shown in the table, so is incorrect.
- $175.5 + 3 \times 7.5 = 198$
- $175.5 + 2 \times 7.5 = 190.5$
- $97.5\% - 2.5\% = 95\%$  so the claim is valid.
- $175.5 - 2 \times 7.5 = 160.5$
- $80\% - 20\% = 60\%$  so the claim is not valid.
- $175.5 - 3 \times 7.5 = 153$
- The distribution is symmetric so  $\frac{177.4 + 173.6}{2} = 175.5$

11 The pie charts show the uses for a village hall in June 1998 and November 1998.



In June 1998 the total number of people using the village hall was 36000 (nearest thousand).

In November 1998 the total number of people using the village hall was 43000 (nearest thousand).

Joe wants to use the totals to draw pie charts.

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

(5 marks)

Select **one** box.

- Joe can use a box plot.
- Joe can use comparative pie charts.
- Joe can use dual pie charts.
- Joe can use 3D pie charts.

Select **one** box.

$\frac{43000^2}{36000^2}$

$\sqrt{\frac{43000}{36000}}$

$\left(\frac{43000}{36000}\right)^2$

$\frac{43000}{36000}$

The radius of the June 1998 pie chart will be \_\_\_\_\_ (2 d.p.) larger than the November 1998 pie chart.

Select **one** box.

Because the totals differ across these pie charts, the overall area of each will reflect that difference, while the proportions remain the same.

With areas inversely related to frequency, the charts become easier to interpret.

Each sector's size will shift depending on the total amount represented in the chart.

Compared to other pie chart formats, this method is much faster.

12 A company applies a protective coating to metal parts.

The coating has a target thickness of 0.15 mm.

The company uses quality assurance to monitor the thickness of the coating on each part.

Samples of the parts are taken from the production line at regular intervals and the mean thickness of the coating in each sample is found.

The sample means should be normally distributed with a mean of 0.15 mm and a standard deviation of 0.01 mm.

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

(2 marks)

Use the formula for the upper action limit

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

\_\_\_\_\_ mm

(b) The upper action limit will be set closer to the target thickness of 0.15 mm.

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

(1 mark)

Select **one** box.

The production process would be stopped less frequently.

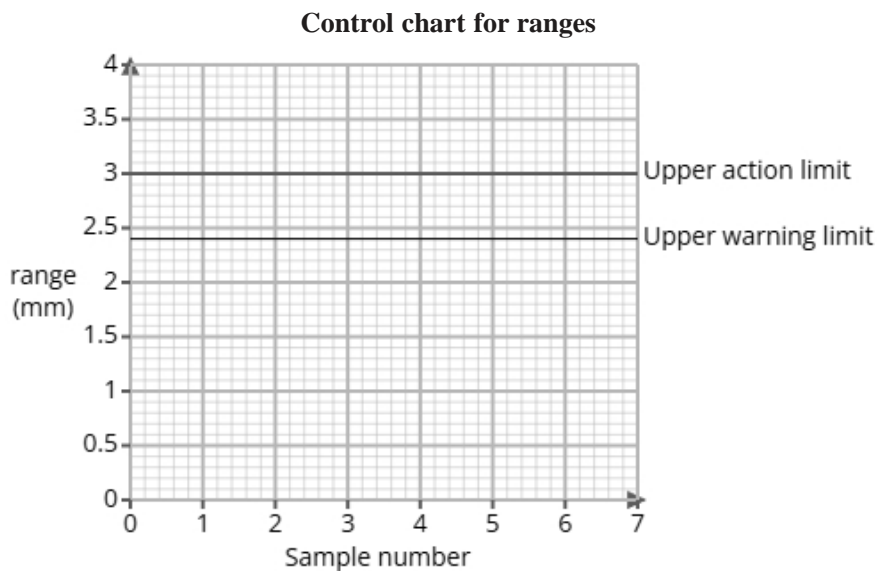
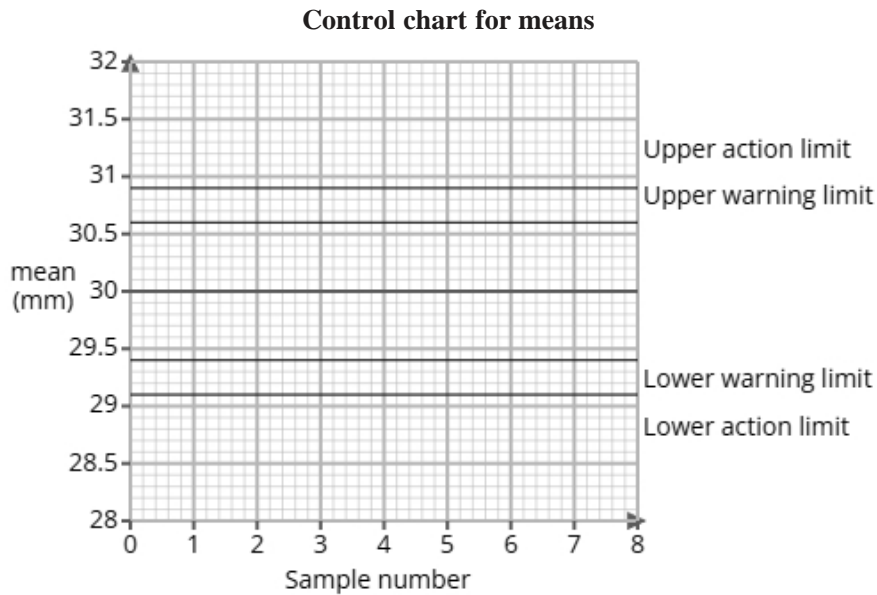
This is unrelated to the production process so will have no effect.

The production process would be stopped more frequently.

The production process will work faster.

(c) The company also applies a protective coating to plastic parts and uses quality assurance to monitor the thickness of the coating.

Here are the control charts for the sample means and for the sample ranges of the thicknesses of each plastic part's coating.



A sample is taken and is found to have a mean of 29.3 mm and a range of 3.1 mm.

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

**13** A study took place in Canada to find if there was a relationship between hours worked and calories burned of remote workers.

The researchers found the equations of the regression lines for the relationship between hours worked ( $x$  hours) and calories burned ( $y$  kcal) for junior employees and senior employees the first quarter (Q1) and the second quarter (Q2).

The table below gives the equations of the regression lines.

	Q1	Q2
junior employees	$y = -30x + 2200$	$y = -25x + 2300$
senior employees	$y = -40x + 2400$	$y = -35x + 2500$

- (a) Compare the relationships between hours worked and calories burned in junior and senior employees. Include in your comparisons reference to whether it is Q1 or Q2.

(3 marks)

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

- As hours worked increased, the calories burned decreased more rapidly in Q1 compared to Q2.
- As hours worked increased, the calories burned decreased more rapidly in Q2 compared to Q1.
- For all remote workers, an increase in hours worked led to a increase in calories burned.
- The calories burned of junior employees decreased more per hours worked than senior employees.
- For all remote workers, an increase in hours worked led to a decrease in calories burned.
- The calories burned of senior employees decreased more per hours worked than junior employees.

(b) The researchers would like to use a normal distribution as a model for the calories burned of junior employees in Q1.

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 increases, a normal distribution could be a suitable model.

If the histogram has a skew, a normal distribution could be a suitable model.

If the histogram is symmetrical, a normal distribution could be a suitable model.

If the histogram decreases, a normal distribution could be a suitable model.

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

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

and 63% of data is within 2 standard deviations from the mean, a normal distribution could be a suitable model.

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

and 95% of data is within 2 standard deviations from the mean, a normal distribution could be a suitable model.

**14** Each visitor to a website has a 10% chance of making a purchase.

A sample of 4 visitors are randomly selected, and the number of purchases are recorded.

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

(2 marks)

Select *two* boxes.

There are only two possible outcomes, purchase or not purchase.

The purchases have a high value.

The probability increases with every purchase.

The chance of a purchase remains constant.

(b) Calculate the probability, as a fraction, that all 3 of the visitors purchased an item.

(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 visitors purchased an item.

(3 marks)

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

---

---

---

---