statsGCSE

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.

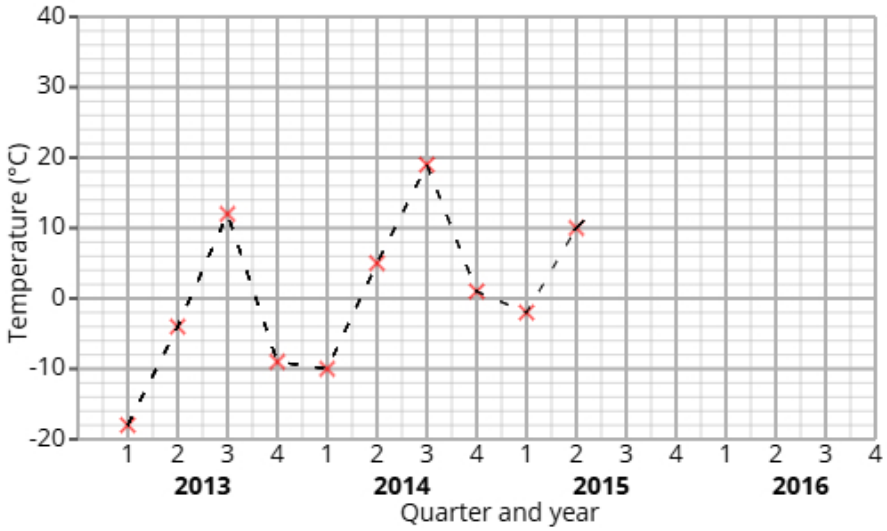
Information

- 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.

Advice

- You can get support for all these questions at our website: 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)

(b) Chloe uses the time series graph to estimate that there was a temperature of 6 °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)

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

(1 mark)

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)

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

(2 marks)

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

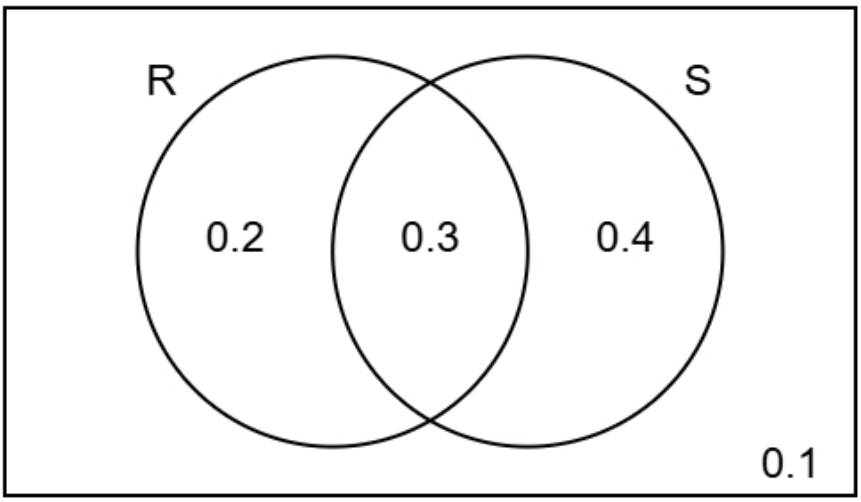
(i) Find the relative risk of failing the coding assessmen having been in Bootcamp 1 compared to Bootcamp 2.

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

(4 marks)

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)

(b) Find P(R and S)

(1 mark)

(c) Find $P(S \mid R)$

(2 marks)

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

$P(D) = 0.2$

$P(E) = 0.6$

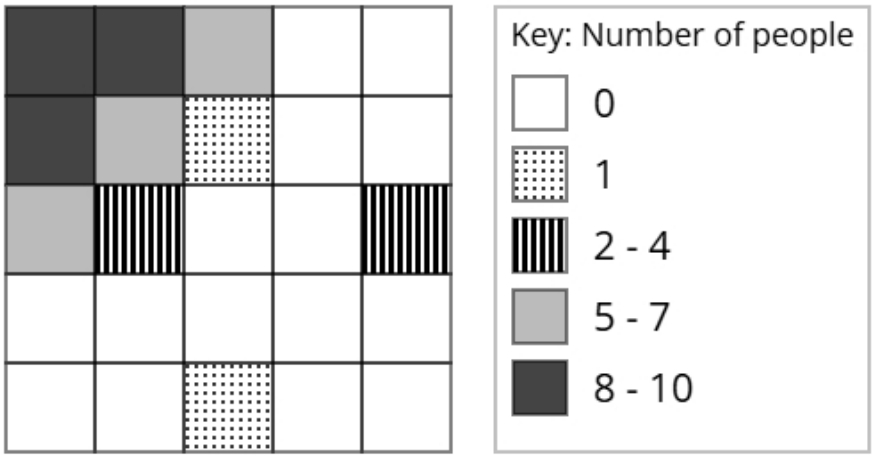
Find $P(D \text{ and } E)$

(2 marks)

5 The choropleth map below represents a 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)

(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)

(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)

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)

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

(3 marks)

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)

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

(3 marks)

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)

(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)

(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)

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)

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)

_____ 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)

_____ %

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

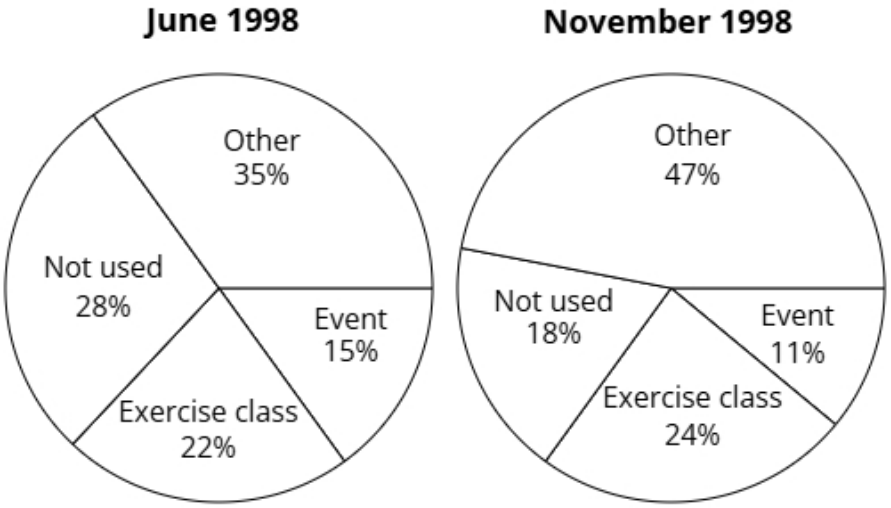
(1 mark)

(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)

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)

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)

_____ mm

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

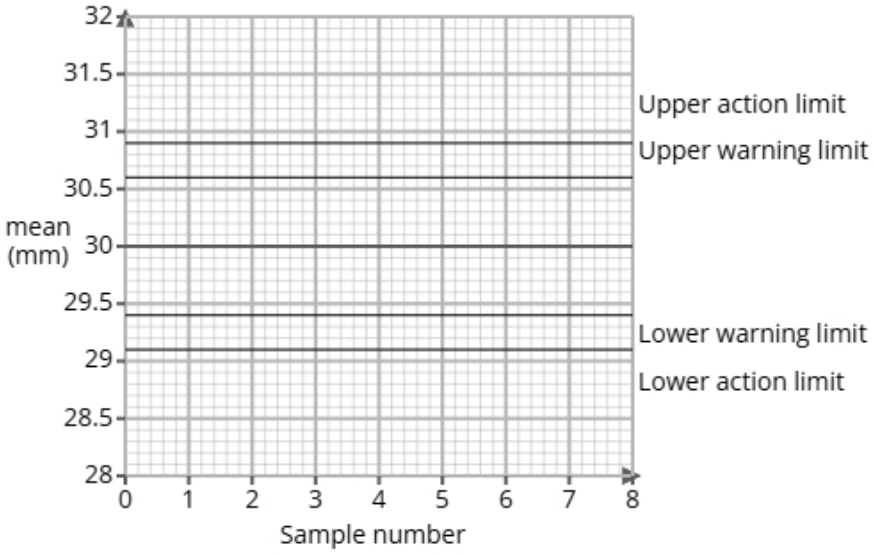
Describe the effect this will have on the frequency of production process stoppages.

(1 mark)

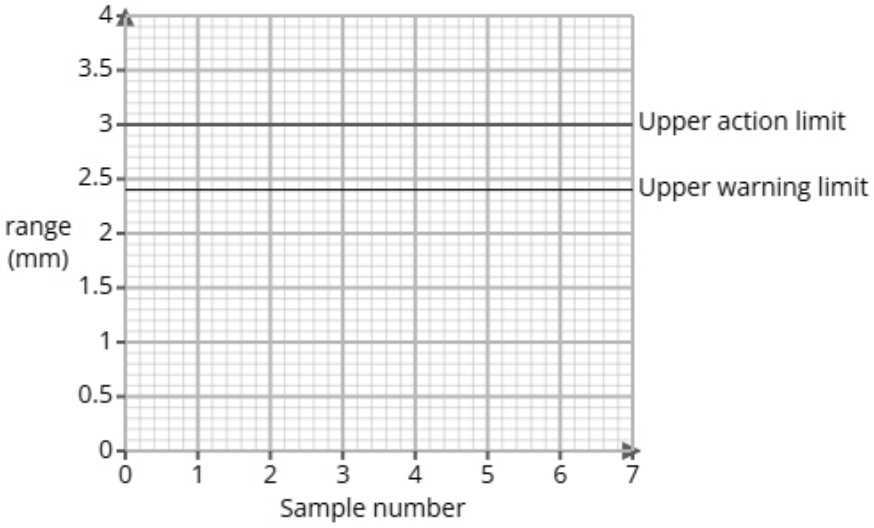
(c) The company also applies a protective coating to plastic parts and uses quality assurance to monitor the thickness of the coating.

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

Control chart for means



Control chart for ranges



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)

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)

(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)

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)

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

(2 marks)

(c) Calculate the probability, as a fraction, that at least 2 of the visitors purchased an item.

(3 marks)