

**Statistics GCSE****Paper 2**

Edexcel Higher - 2026

Higher Tier

Variant 4

1ST0/2H

**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.

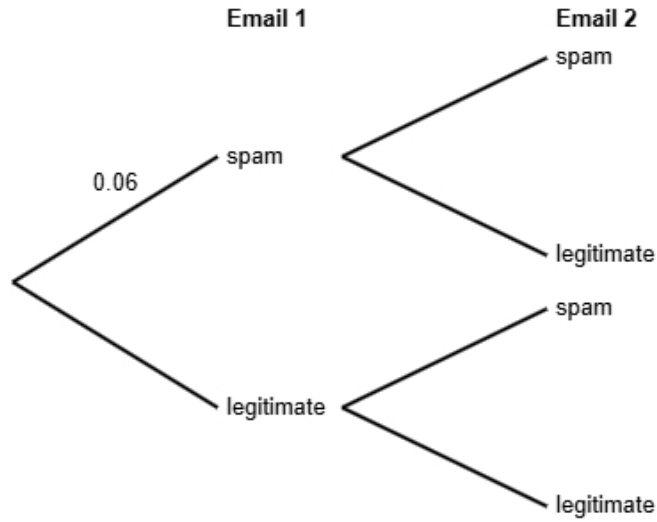
**Copyright**

- 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** A study shows that 6% of emails received by a certain email provider are spam.  
 All other emails are legitimate.  
 Maria receives two emails in her inbox.  
 She does not know if each email is spam or legitimate.



- (a)** Complete the probability tree diagram.

(2 marks)

- (b)** Find the probability that both of Maria's emails are legitimate.

(2 marks)

\_\_\_\_\_

- (c) Maria states that the probability that exactly one email is spam is less than 12%  
Find out whether or not Maria is correct.

(3 marks)

2 The table shows information about bicycles for sale in Birmingham.

gear types	number of bicycles
1	175
2	350
3	200
4	450
5 or more	425
Total	1600

A researcher wants to investigate the price of these bicycles and takes a stratified sample of 64 bicycles according to the gear types.

(a) The researcher says the mode of the gear types for these bicycles is 4.

Explain how the researcher knows this.

(1 mark)

---

---

---

---

(b) Work out the number of bicycles in the sample for each gear type.

gear types	number of bicycles in the sample
1	
2	
3	
4	
5 or more	

(3 marks)

(c) Describe how the 64 bicycles in the sample should be selected.

(3 marks)

---

---

---

---

---

---

---

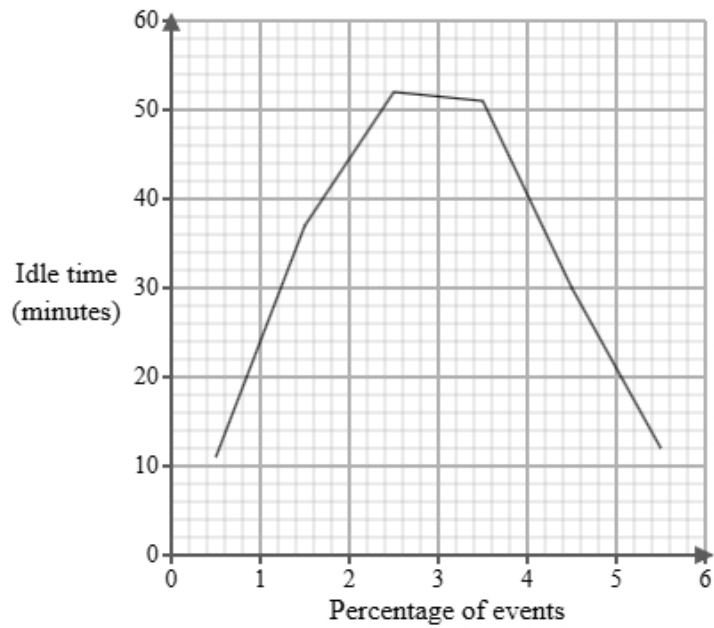
- 3 Tom works for a logistics company. He has been tasked with investigating delivery vehicle idle times. Below is a section of the spreadsheet he used to record his findings.

Idle time (Minutes)	Percentage of events
$0 < d \leq 1$	6
$1 < d \leq 2$	four
$2 < d \leq 3$	9
$3 < d \leq 4$	119
$4 < d \leq 5$	47
$5 < d \leq 6$	15
Total	100

Tom cleans the data to create the table below.

Idle time (Minutes)	Percentage of events
$0 < d \leq 1$	6
$1 < d \leq 2$	4
$2 < d \leq 3$	9
$3 < d \leq 4$	19
$4 < d \leq 5$	47
$5 < d \leq 6$	15
Total	100

A frequency polygon has been drawn for delivery vehicle idle times at a rival firm.



- i) On the same graph, draw the frequency polygon for delivery vehicle idle times for Tom's company.
- ii) Using the two frequency polygons, compare the skew of the distributions and explain what your comparison means in context.

(4 marks)

---



---



---



---



---



---



---



---

- 4 The table shows information about the retail price index (RPI) and cinema ticket price (£) in the United Kingdom for Jan 1990, Jan 2000 and Jan 2010.

	<b>Jan 1990</b>	<b>Jan 2000</b>	<b>Jan 2010</b>
retail price index	100	145	177
cinema ticket price (£)	2.81	3.99	5.97

Describe how the increase in cinema ticket price (£) compares with the RPI over the ten years to Jan 2000 and over the twenty years to Jan 2010.

(5 marks)

---

---

---

---

---

---

---

---

---

---

5 Daniel is investigating how the distance from city centre in km,  $x$ , affects the selling price (£),  $y$  for two types of houses, detached houses and semi-detached houses.

He found ten houses of each type and recorded their distance from city centre and selling price and plotted each on scatter diagrams.

He then drew a line of best fit on each diagram and found the gradient and  $y$ -intercept of each line.

Here are the results:

House type	Gradient of line of best fit	$y$ -intercept of line of best fit
Detached houses	-5000	450000
Semi-detached houses	-4200	380000

Interpret and compare these results in context.

(5 marks)

---

---

---

---

---

---

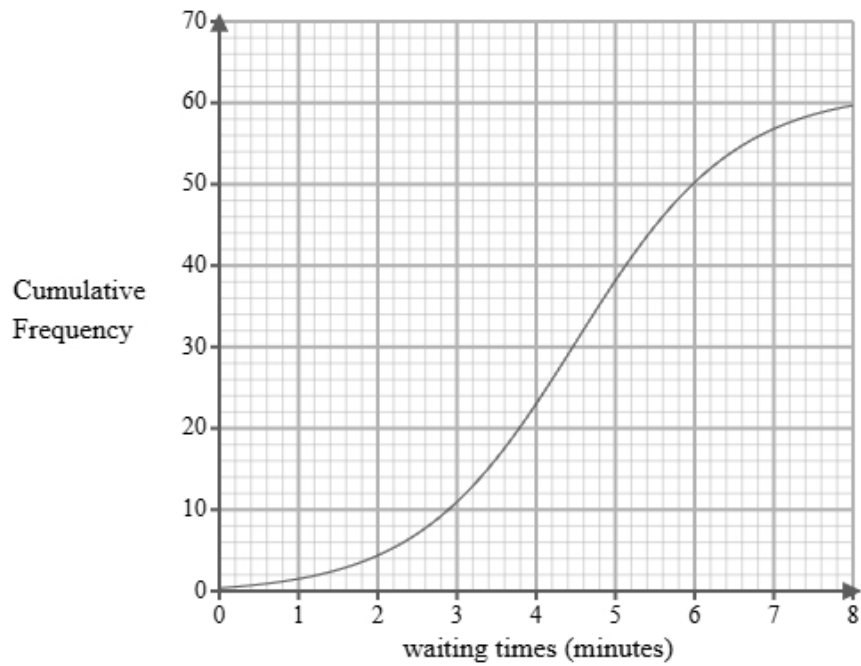
---

---

---

---

- 6 A researcher measures the waiting times, in minutes, of 60 customers at a café.  
A cumulative frequency diagram is drawn from the data.



Complete the table below from the cumulative frequency diagram.

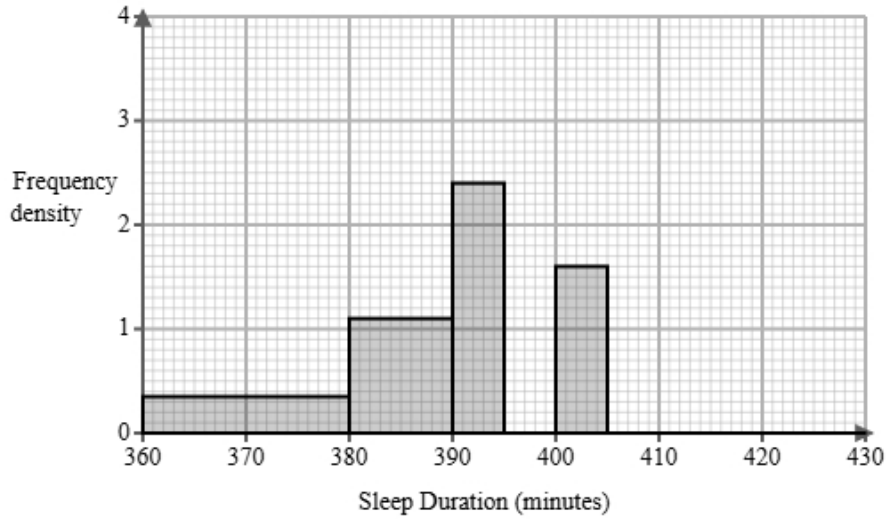
Lower quartile	Median	Upper quartile

(2 marks)

7 The amount of sleep is recorded in minutes.

A sleep researcher is analysing the duration of sleep that 60 adult women get following a 12-hour period of fasting.

The partially completed histogram and grouped frequency table provide details about the sleep durations recorded in the study.



Sleep Duration $s$ (minutes)	Frequency
$360 < s \leq 380$	7
$380 < s \leq 390$	11
$390 < s \leq 395$	
$395 < s \leq 400$	12
$400 < s \leq 405$	
$405 < s \leq 430$	10

(a) Complete the table using the information from the histogram.

(2 marks)

(b) Complete the histogram using the information from the table.

(2 marks)

(c) The sleep researcher finds the following summary statistics for the data.

$$\sum s = 23730$$

$$\sum s^2 = 9393238$$

$$n = 60$$

Explain whether or not there may be any outliers in the sleep researcher's data by calculating the limits for outliers using the mean and standard deviation.

You must round all values to 2 decimal places.

(5 marks)

---

---

---

---

---

---

---

---

---

---

---

(d) A different sleep researcher is analysing the duration of sleep adult men get following a 12-hour period of fasting.

They find the following summary statistics for the data.

mean = 347.54

median = 380

standard deviation = 11.4

Calculate and interpret the skew for the men.

You must round your answer to 2 decimal places.

(3 marks)

---

---

---

---

---

---

(e) Find the class interval that contains the 30th percentile.

(1 mark)

---

- 8 Ethan is researching the final league position of basketball teams in a local league and the mean heights of all the players in each team.

The table below shows the data collected.

Team	Mean Height (cm)	Height Rank	Final Position	d	d <sup>2</sup>
Aylesbury	168	1	1	0	0
Bedford	180	7	7	0	0
Chatham	181	8	8	0	0
Dorking	171	3	3	0	0
Eastleigh	179	6	5		
Farnborough	176	5	6		
Guildford	169	2	4		
Havant	173	4	2		

- (a) Ethan would like to see if there is an association between the final position and the mean value.

Suggest a diagram that Ethan could draw.

(1 mark)

---

(b) i) Calculate Spearman's rank correlation coefficient from the data in the table and leave your answer to 2 decimal places.

ii) Interpret your answer to **part i**, referring to the effects of any anomalous data.

(5 marks)

---

---

---

---

---

---

---

---

---

---

---

(c) Ethan used Spearman's rank correlation coefficient to analyse the data.

Emily suggests that Ethan could have used Pearson's product moment correlation coefficient.

Discuss whether using Pearson's product moment correlation coefficient is appropriate for this data.

(3 marks)

---

---

---

---

---

---

- 9 Tom works for a logistics company. He has been tasked with investigating delivery vehicle idle times. Below is a section of the spreadsheet he used to record his findings.

Idle time (Minutes)	Percentage of events
$0 < d \leq 1$	6
$1 < d \leq 2$	four
$2 < d \leq 3$	9
$3 < d \leq 4$	119
$4 < d \leq 5$	47
$5 < d \leq 6$	15
Total	100

Tom cleans the data to create the table below.

Idle time (Minutes)	Percentage of events
$0 < d \leq 1$	6
$1 < d \leq 2$	4
$2 < d \leq 3$	9
$3 < d \leq 4$	19
$4 < d \leq 5$	47
$5 < d \leq 6$	15
Total	100

- (a) Give a reason Tom cleaned the data.

(1 mark)

---

---

---

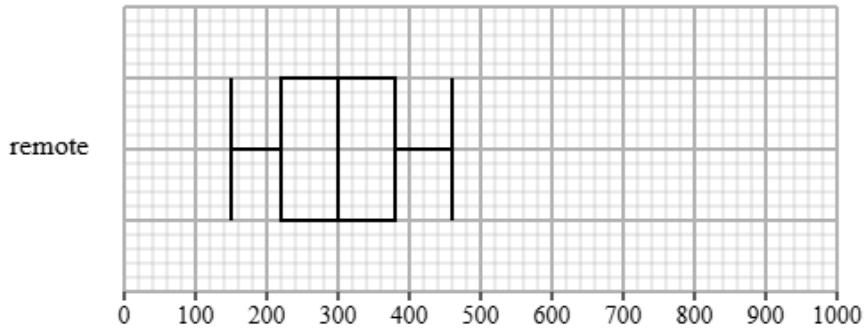
---

(b) Use linear interpolation to work out an estimate of the median idle time.  
Round your answer to one decimal place.

(3 marks)

\_\_\_\_\_ minutes

- 10** Ethan collected the steps for remote and on-site workers in an hour within their day. Both groups recorded their steps over the same period. The box plot presents data on the steps for the remote workers.



The table gives information about the steps for the on-site workers.

Least tall	Lower quartile	Median	Upper quartile	Most tall
300	550	800	850	900

Compare the two distributions of steps.

Give three comparisons and interpret one of these comparisons.

(4 marks)

---



---



---



---



---



---

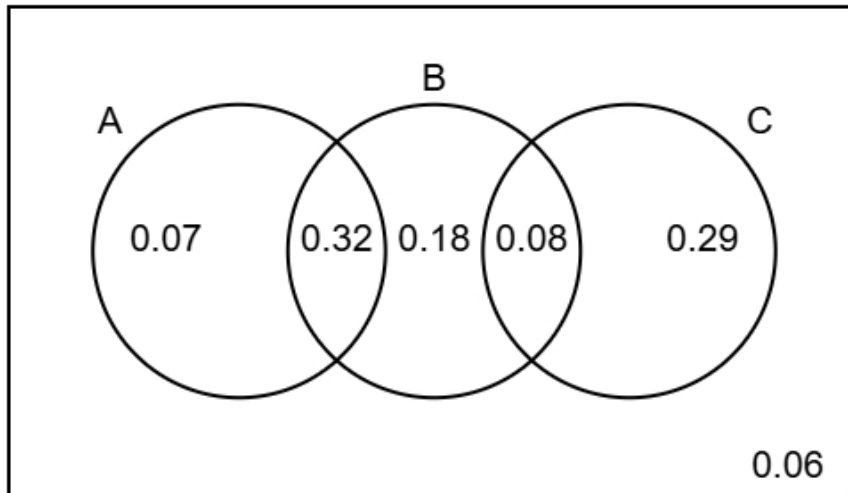


---



---

11 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)

---

---

---

---

---

---

---

---

(b) Find  $P(B)$

(1 mark)

---

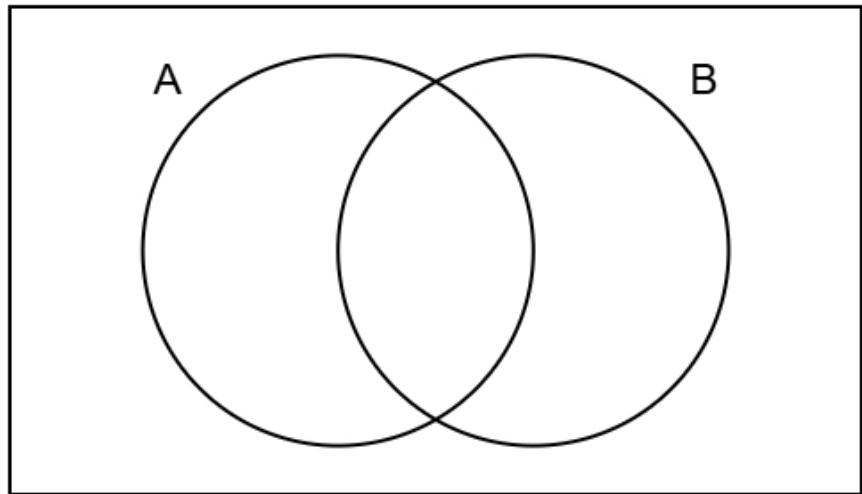
(c) Find  $P(A \text{ or } C)$

(2 marks)

---

(d) Complete the Venn diagram to show **only** the probabilities for events A and B.

(2 marks)



12 The figures below show the amount, in millions, of tourists who visited USA between 2009 and 2015

72 75 78 74 80 83 85

The table gives a summary of the amount, in millions, of tourists who visited Italy between 2009 and 2015

Mean	Standard Deviation	Largest Amount
81	4	83

Compare the amount of tourists in USA and Italy between 2009 and 2015

You may use:

$$72^2 + 75^2 + 78^2 + 74^2 + 80^2 + 83^2 + 85^2 = 42883$$

(5 marks)

---

---

---

---

---

---

---

---

---

---

---

**13** The daily screen times of a group of teenagers have a mean of 3.7 hours and a standard deviation of 1.2 hours.

(a) David is teenager with a standardised score of 0.

Find David's daily screen time.

(1 mark)

\_\_\_\_\_ hours

(b) Ben and Charlie are both teenagers in the group.

Ben's standardised score for daily screen time is 0.6 hours.

Charlie's standardised score for daily screen time is -1.1 hours.

Ben had a higher screen time than Charlie.

How much more screen time is spent by Ben?

(3 marks)

\_\_\_\_\_ hours

