

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

Edexcel Higher - 2026

Higher Tier

Variant 3

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.

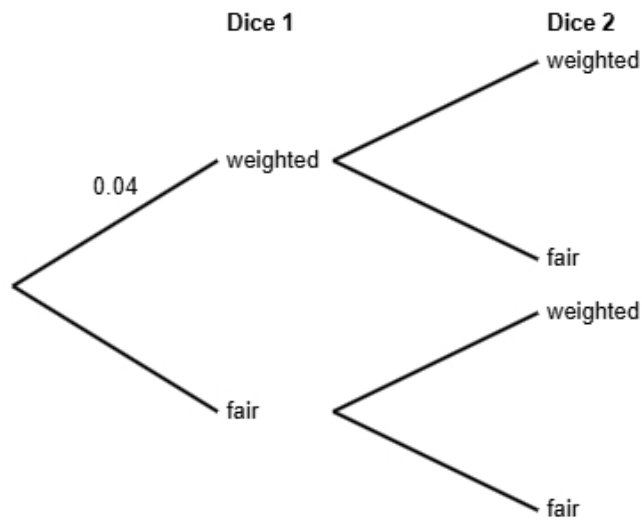
**Copyright**

- 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 company found that 4% of its dice are slightly weighted and do not roll fairly.  
The rest of the dice are perfectly balanced.  
Jack picks two dice from a random batch.  
He does not know if each die is weighted or fair.



- (a) Complete the probability tree diagram.

(2 marks)

The branches for each stage must add up to 1.  
Each test is independent so will have the same probabilities.

(b) Find the probability that both of Jack's dice are fair.

(2 marks)

You will need to find  $P(\text{fair})$  AND  $P(\text{fair})$ .  
Remember, AND means  $\times$  in probability.

(c) Jack states that the probability that exactly one dice is weighted is less than 8%.  
Find out whether or not Jack is correct.

(3 marks)

Find the probability of exactly one dice is weighted (there are two outcomes on the tree diagram).

Select **one** box.

- The probability that exactly one dice is weighted is more than 8%, so Jack is not correct.
- The probability that exactly one dice is weighted is less than 8%, so Jack is not correct.
- The probability that exactly one dice is weighted is more than 8%, so Jack is correct.
- The probability that exactly one dice is weighted is less than 8%, so Jack is correct.

2 The table shows information about laptops for sale in London.

screen size (inches)	number of laptops
13	660
15	540
17	240
19	150
20 or more	210
Total	1800

A researcher wants to investigate the price of these laptops and takes a stratified sample of 60 laptops according to the screen size (inches).

(a) The researcher says the mode of the screen size (inches) for these laptops is 13.

Explain how the researcher knows this.

(1 mark)

Select **one** box.

- 13 is the middle number.
- 13 inches has the highest frequency.
- 13 is the difference between the largest and smallest number.
- 13 laptops has the highest frequency.

(b) Work out the number of laptops in the sample for each screen size.

screen size (inches)	number of laptops in the sample
13	
15	
17	
19	
20 or more	

(3 marks)

Find the divisor for the stratified sample:  $\frac{\text{total}}{\text{sample size}}$

Divide each frequency by this number to find the required sample in each group

(c) Describe how the 60 laptops in the sample should be selected.

(3 marks)

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

- Ensure that all 1800 laptops are included in the sample.
- Complete two of the strata.
- Select the first 60 laptops.
- Generate random numbers, remove repeats or numbers out of range.
- Number each of the laptops, and then use the random numbers to select the required amount of laptops.
- Use a sampling frame for each strata.

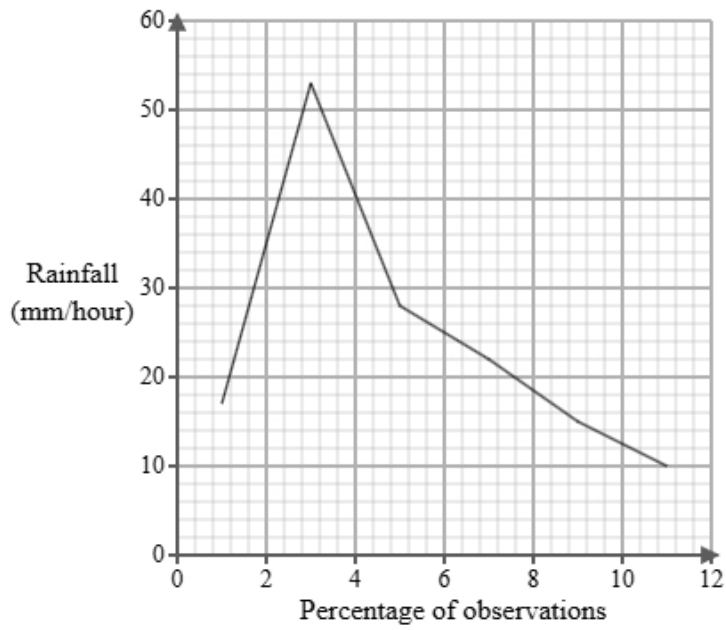
- 3 Priya works for a meteorological office. She has been tasked with investigating rainfall intensity in June. Below is a section of the spreadsheet she used to record her findings.

Rainfall (mm/hour)	Percentage of observations
$0 < r \leq 2$	5
$2 < r \leq 4$	eight
$4 < r \leq 6$	7
$6 < r \leq 8$	120
$8 < r \leq 10$	47
$10 < r \leq 12$	13
Total	100

Priya cleans the data to create the table below.

Rainfall (mm/hour)	Percentage of observations
$0 < r \leq 2$	5
$2 < r \leq 4$	8
$4 < r \leq 6$	7
$6 < r \leq 8$	20
$8 < r \leq 10$	47
$10 < r \leq 12$	13
Total	100

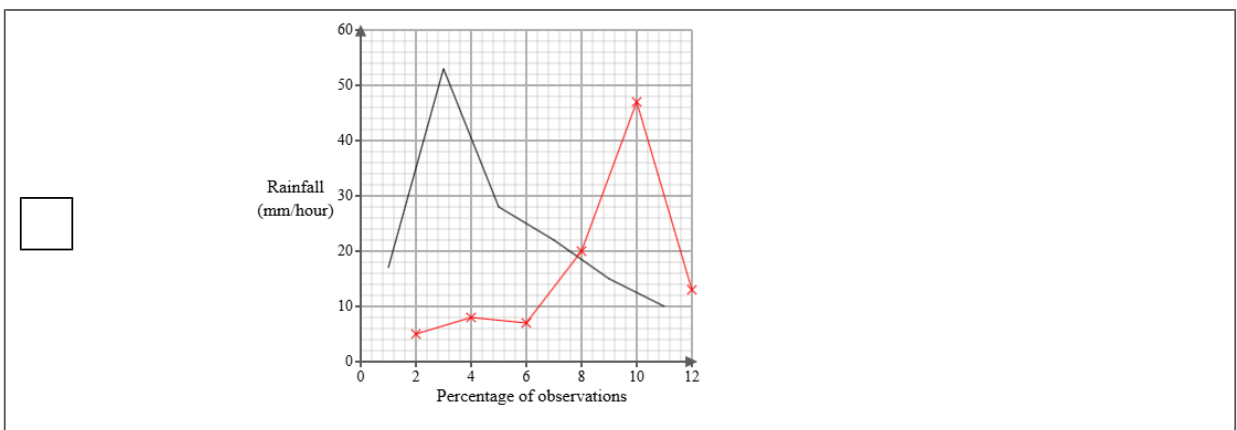
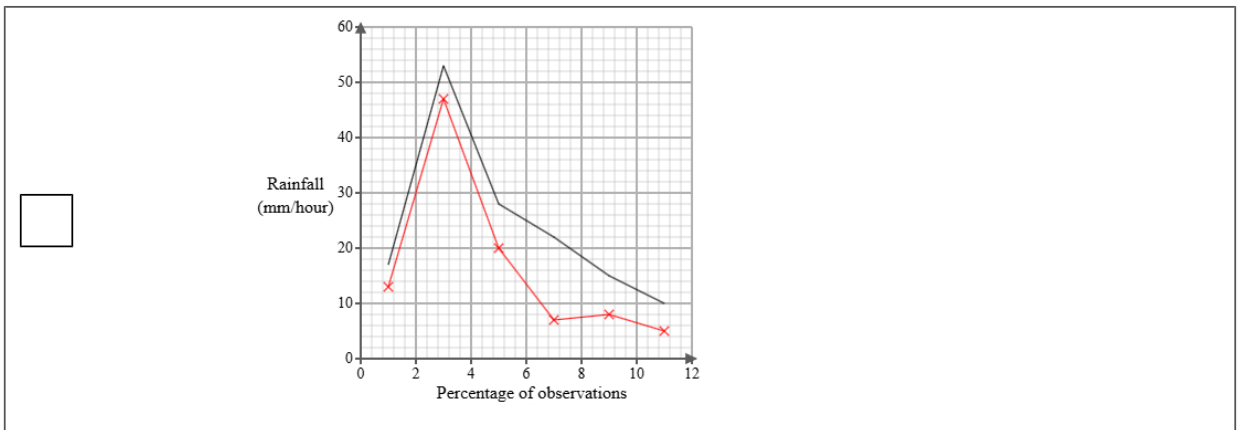
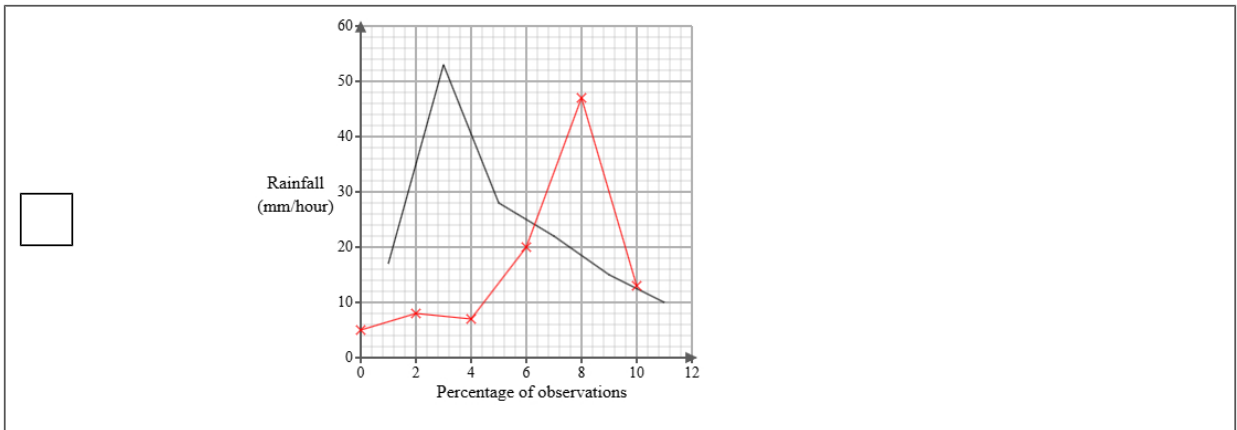
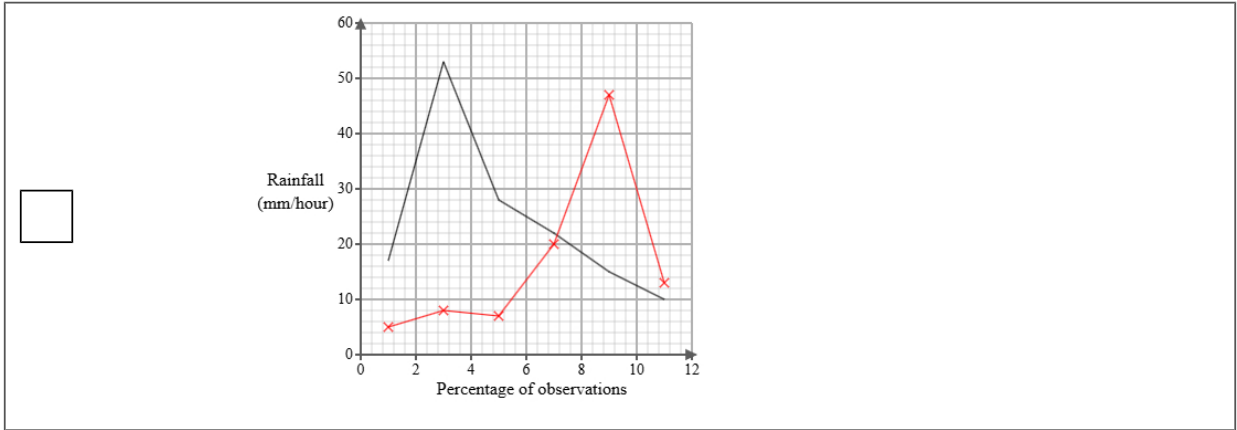
A frequency polygon has been drawn for rainfall intensity in August.



- i) On the same graph, draw the frequency polygon for rainfall intensity in June.
- ii) Using the two frequency polygons, compare the skew of the distributions and explain what your comparison means in context.

(4 marks)

Select the correct answer.



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

- The distribution of rainfall intensity in August is negatively skewed whereas the distribution of rainfall intensity in June is positively skewed.
- This means that in August the rainfall is mainly at the upper end of the distribution and in June the rainfall is mainly at the lower end of the distribution.
- This means that in August the rainfall is mainly at the lower end of the distribution and in June the rainfall is mainly at the upper end of the distribution.
- The distribution of rainfall intensity in August is symmetrical whereas the distribution of rainfall intensity in June is positively skewed.
- This means that in August the rainfall is equally spread out on either side of the median and in June the rainfall is mainly at the upper end of the distribution.
- The distribution of rainfall intensity in August is positively skewed whereas the distribution of rainfall intensity in June is negatively skewed.

- 4 The table shows information about the retail price index (RPI) and NHS Prescription Charges (£) in England for Jan 1995, Jan 2005 and Jan 2015.

	Jan 1995	Jan 2005	Jan 2015
retail price index	100	121	173
NHS Prescription Charges (£)	5.25	6.5	8.05

Describe how the increase in NHS Prescription Charges (£) compares with the RPI over the ten years to Jan 2005 and over the twenty years to Jan 2015.

(5 marks)

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

$\frac{6.5}{121} \times 100 = 5$  (nearest integer)

$\frac{8.05}{173} \times 100 = 5$  (nearest integer)

Between Jan 1995 and Jan 2015 the change in price was less than the RPI.

Between Jan 1995 and Jan 2015 the change in price was more than the RPI.

Between Jan 1995 and Jan 2005 the change in price was less than the RPI.

$\frac{6.5}{5.25} \times 100 = 124$  (nearest integer)

$\frac{8.05}{5.25} \times 100 = 153$  (nearest integer)

Between Jan 1995 and Jan 2005 the change in price was more than the RPI.

5 Aisha is investigating how the age in years,  $x$ , affects the resale price (£),  $y$  for two types of smartphones, model X and model Y.

She found ten smartphones of each type and recorded their age and resale price and plotted each on scatter diagrams.

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

Here are the results:

Model	Gradient of line of best fit	y-intercept of line of best fit
X	-80	900
Y	-65	1100

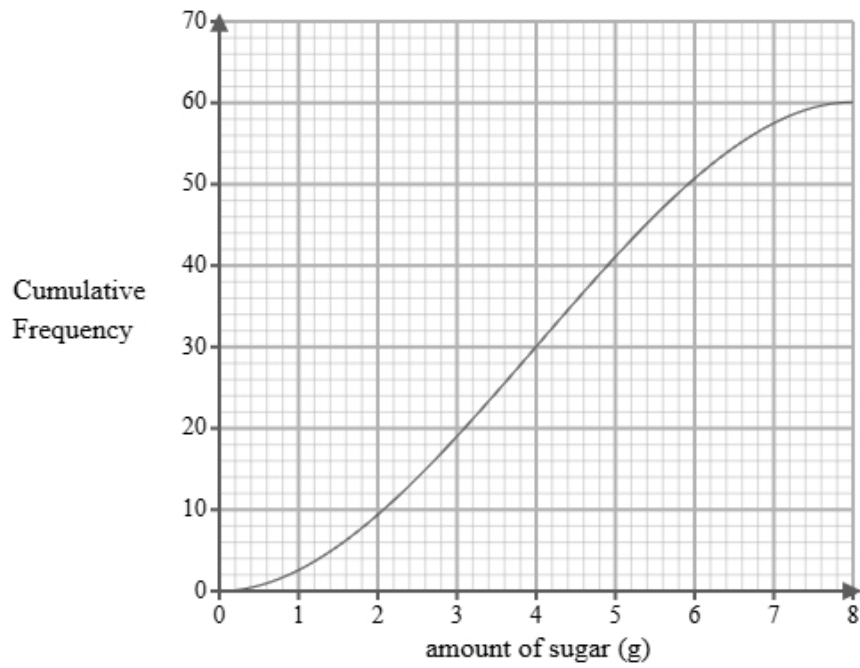
Interpret and compare these results in context.

(5 marks)

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

- Model X reduces in resale price by £80 per year.
- Model X has a greater initial resale price.
- Model X changes in resale price by £900 per year.
- Model Y has a greater initial resale price.
- Model Y changes in resale price by £1100 per year.
- Both smartphones decrease in resale price as the age increase.
- Model X reduces in resale price less than Model Y.
- Model X reduces in resale price more per year than Model Y.
- Model Y reduces in resale price by £65 per year.
- Both smartphones increase in resale price as the age increase.

- 6 A researcher measures the amount of sugar, in grams, in 60 different breakfast cereals.  
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)

Select the correct answer.

<input type="checkbox"/>	Lower quartile	Median	Upper quartile
	1.4	4.7	6.8

<input type="checkbox"/>	Lower quartile	Median	Upper quartile
	3.8	4	4.7

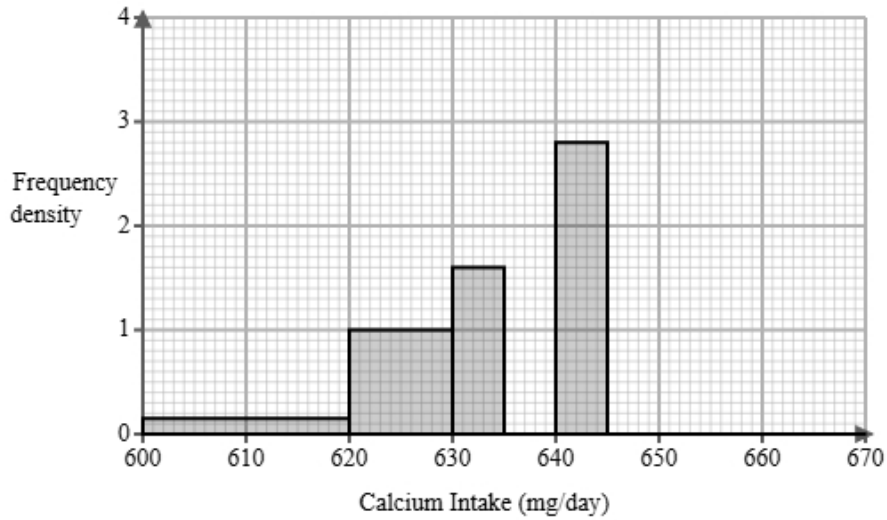
<input type="checkbox"/>	Lower quartile	Median	Upper quartile
	1.9	4.7	6.1

<input type="checkbox"/>	Lower quartile	Median	Upper quartile
	2.6	4	5.4

7 The calcium intake is recorded in milligrams per day (mg/day).

A dietitian is analysing the daily calcium intake of 60 adult women who have fasted for 12 hours.

The partially completed histogram and grouped frequency table provide details about the calcium intake levels recorded.



Calcium Intake $c$ (mg/day)	Frequency
$600 < c \leq 620$	3
$620 < c \leq 630$	10
$630 < c \leq 635$	
$635 < c \leq 640$	15
$640 < c \leq 645$	
$645 < c \leq 670$	10

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

(2 marks)

Use the frequency density formula

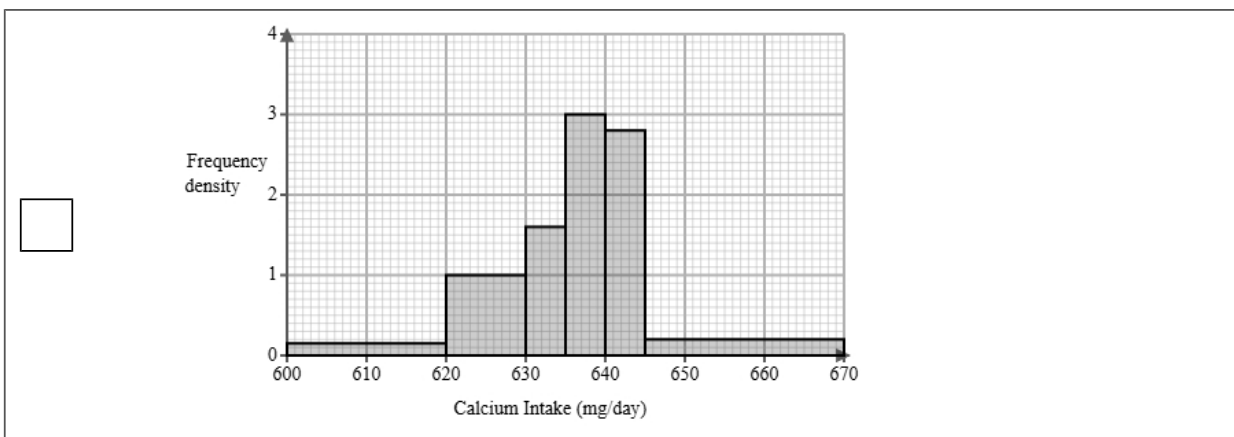
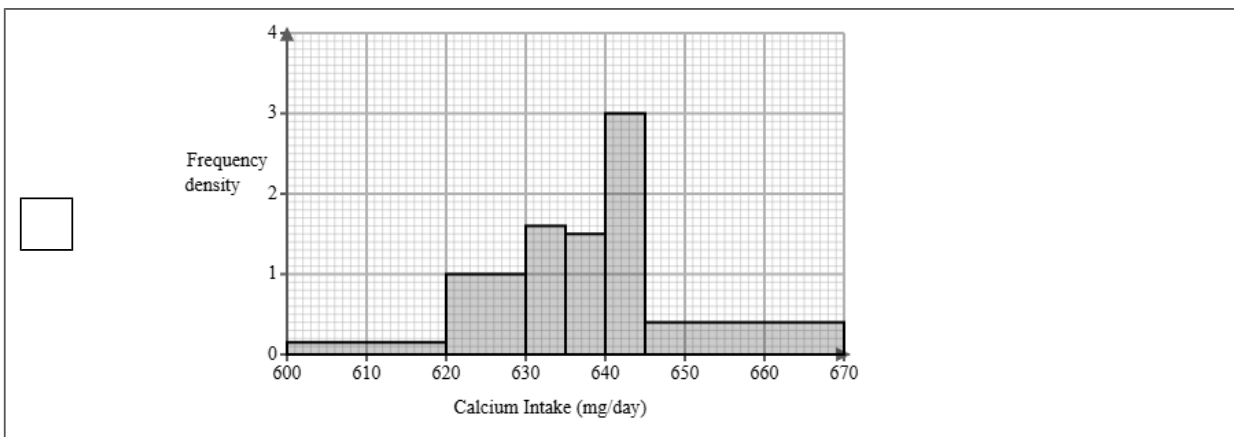
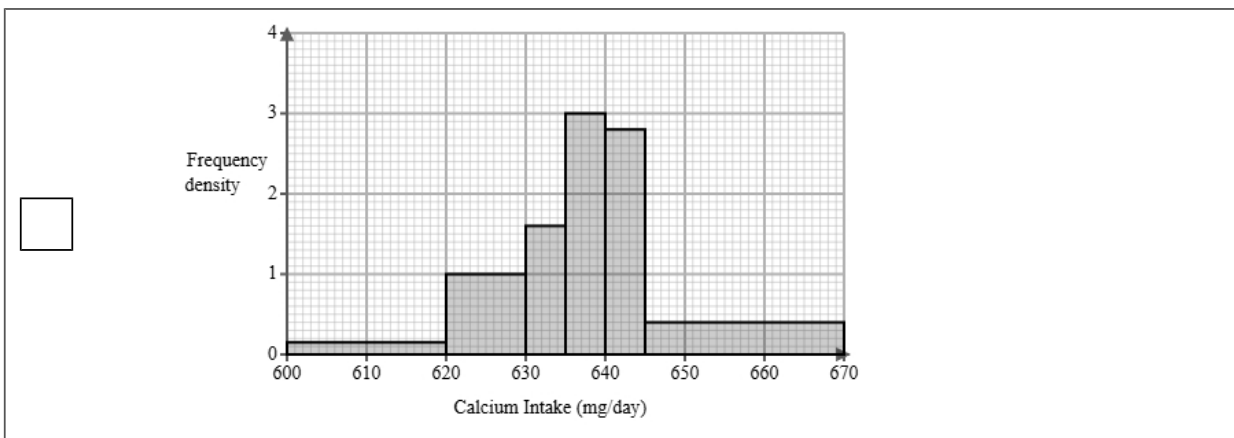
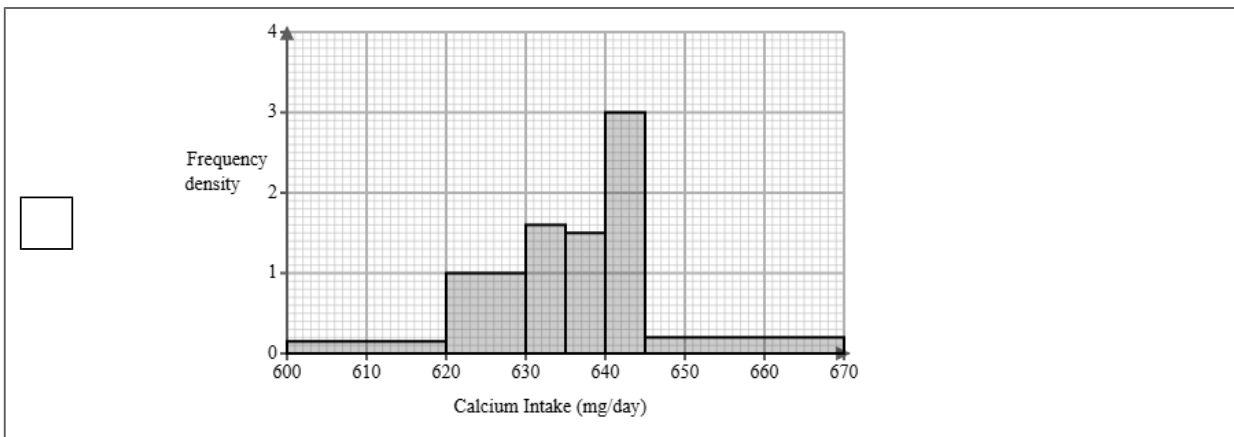
$$\text{frequency} = \text{frequency density} \times \text{class width}$$

Calcium Intake $c$ (mg/day)	Frequency
$600 < c \leq 620$	3
$620 < c \leq 630$	10
$630 < c \leq 635$	_____
$635 < c \leq 640$	15
$640 < c \leq 645$	_____
$645 < c \leq 670$	10

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

(2 marks)

Select the correct answer.



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

$$\sum c = 38218$$

$$\sum c^2 = 24348590$$

$$n = 60$$

Explain whether or not there may be any outliers in the dietitian'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)

mean = \_\_\_\_\_

standard deviation = \_\_\_\_\_

lower outlier limit = \_\_\_\_\_

upper outlier limit = \_\_\_\_\_

Select **one** box.

There are outliers because the limits are within the data.

It is possible that there is an outlier as the lower outlier limit is within the group  $600 < c \leq 620$  and the upper outlier limit is within the group  $645 < c \leq 670$ .

There are no outliers because the limits are outside the ranges of the data.

- (d) A different dietitian is analysing the daily calcium intake of adult men who have fasted for 12 hours. They find the following summary statistics for the data.

mean = 649.54

median = 680

standard deviation = 11.3

Calculate and interpret the skew for the men.

You must round your answer to 2 decimal places.

(3 marks)

Skew = \_\_\_\_\_

Select **one** box.

There is a negative skew showing the data is not normally distributed.

The skew shows there is a negative correlation.

The skew shows that the three averages are equal.

- (e) Find the class interval that contains the 20th percentile.

(1 mark)

Select **one** box.

$620 < c \leq 630$

$630 < c \leq 635$

$600 < c \leq 620$

$635 < c \leq 640$

$640 < c \leq 645$

- 8 Hannah 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>
Ashford	179	7	8	-1	1
Banbury	180	8	7	1	1
Chelmsford	170	3	1	2	4
Crawley	168	1	3	-2	4
Dover	176	6	6		
Epsom	169	2	2		
Fareham	173	4	5		
Gravesend	175	5	4		

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

Suggest a diagram that Hannah could draw.

(1 mark)

Select **one** box.

Histogram

Pie chart

Scatter diagram

Cumulative frequency diagram

(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)

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

- Anomalous data would decrease the correlation.
- As the mean height of players increases, the position of the team in the league is lower.
- As the mean height of players increases, the position of the team in the league is higher.
- Anomalous data would increase the correlation.

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

Harry suggests that Hannah 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)

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

- Spearman's rank correlation is used for ranked data
- Hannah used the correct method, Harry's suggestion is not appropriate.
- PMCC measures linear correlation.
- Harry's suggestion is more appropriate than Hannah's method.
- PMCC compares bivariate data
- Spearman's rank correlation is influenced by outliers.

- 9 Priya works for a meteorological office. She has been tasked with investigating rainfall intensity in June. Below is a section of the spreadsheet she used to record her findings.

Rainfall (mm/hour)	Percentage of observations
$0 < r \leq 2$	5
$2 < r \leq 4$	eight
$4 < r \leq 6$	7
$6 < r \leq 8$	120
$8 < r \leq 10$	47
$10 < r \leq 12$	13
Total	100

Priya cleans the data to create the table below.

Rainfall (mm/hour)	Percentage of observations
$0 < r \leq 2$	5
$2 < r \leq 4$	8
$4 < r \leq 6$	7
$6 < r \leq 8$	20
$8 < r \leq 10$	47
$10 < r \leq 12$	13
Total	100

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

(1 mark)

Select **one** box.

- Priya must have more data.
- Data must be in the same format before it can be used
- Priya needs the raw data.

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

(3 marks)

Find the group where the 50th value is in

You can use the interpolation formula to find the median

$$\text{estimated median} = L + \frac{\frac{n}{2} - F}{f} \times w$$

lower boundary = L

number of values = n

cumulative frequency before group = F

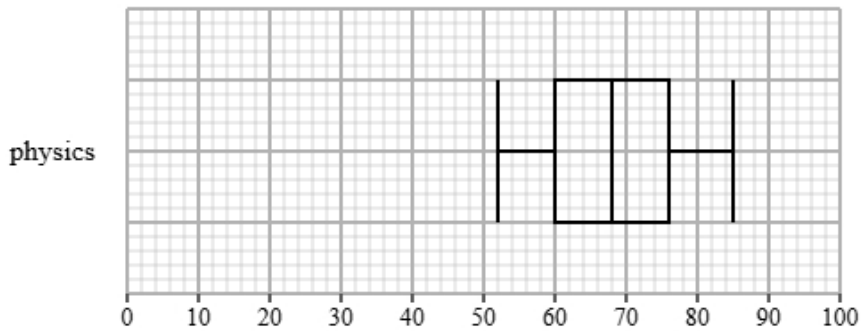
frequency of group = f

width of group = w

\_\_\_\_\_ mm/hour

10 Olivia collected the marks for physics and business students from a statistics exam.  
Both groups took the same exam.

The box plot presents data on the marks for the physics students.



The table gives information about the marks for the business students.

Least tall	Lower quartile	Median	Upper quartile	Most tall
52	58	60	72	85

Compare the two distributions of marks.

Give three comparisons and interpret one of these comparisons.

(4 marks)

Select **one** box.

- The median is bigger.
- The median marks for physics students is greater than business students.
- The median marks for physics and business students are equal.
- The median marks for physics students is lower than business students.

Select **one** box.

- The IQR is bigger.
- The IQR for the marks of the physics and business students are equal.
- The IQR for the marks of the physics students is greater than business students.
- The IQR for the marks of the physics students is lower than business students.

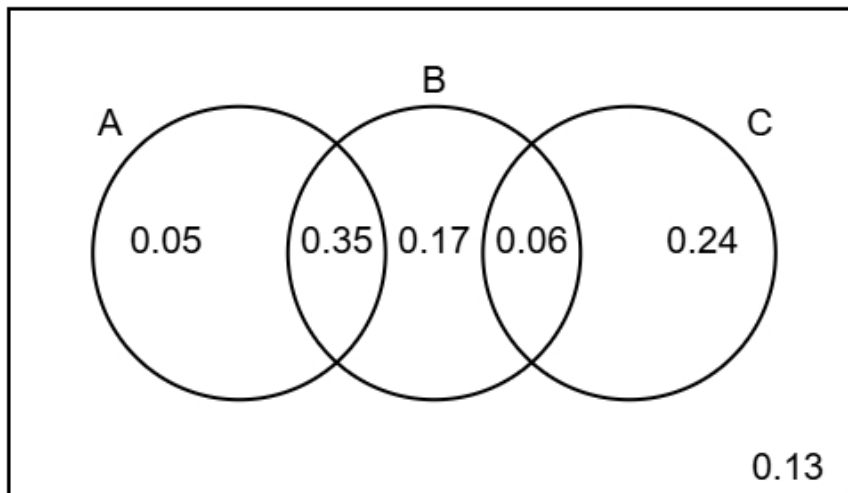
Select **one** box.

- The skews for the marks of the physics and business students are both positive.
- The skews for the marks of the physics and business students are both symmetrical.
- The skew for the marks of the physics students is symmetrical and the skew for the business students is negative.
- The skew for the marks of the physics students is symmetrical and the skew for the business students is positive.

Select **one** box.

- The marks for the physics students are less spread out than the business students.
- The physics students are more skewed than business students.
- The physics students are on average did better on the statistics test than the business students.
- The physics students are on average did worse on the statistics test than the business students.

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)

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

- because they have the highest total probability.
- A and C are mutually exclusive
- A and B are mutually exclusive
- because they only intersect once.
- B and C are mutually exclusive
- because they do not intersect.

(b) Find  $P(B)$

(1 mark)

We are looking for the probabilities inside B.

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

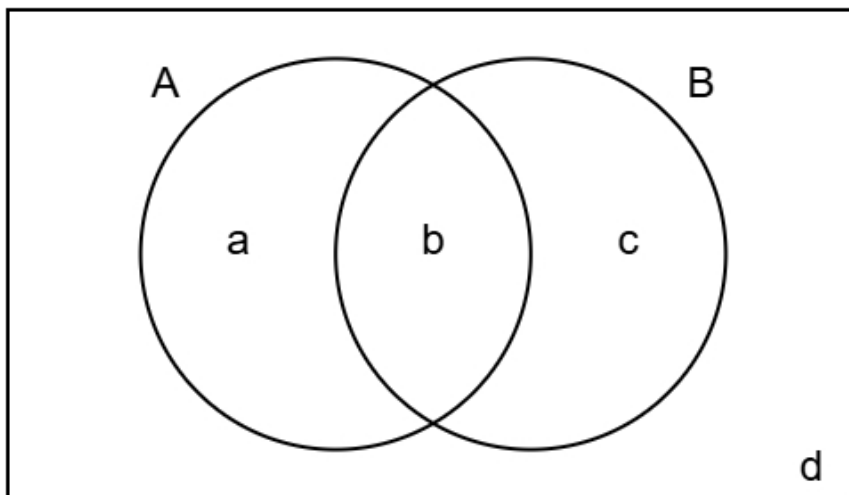
(2 marks)

We are looking for the probabilities inside A or C.

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

(2 marks)

Combine the probabilities from C into either B or the outside area.



**a** = \_\_\_\_\_ **b** = \_\_\_\_\_

**c** = \_\_\_\_\_ **d** = \_\_\_\_\_

12 The figures below show the amount, in millions, of tourists who visited Italy between 2011 and 2017

45 48 50 47 52 55 57

The table gives a summary of the amount, in millions, of tourists who visited Spain between 2011 and 2017

Mean	Standard Deviation	Largest Amount
50	5	55

Compare the amount of tourists in Italy and Spain between 2011 and 2017

You may use:

$$45^2 + 48^2 + 50^2 + 47^2 + 52^2 + 55^2 + 57^2 = 18016$$

(5 marks)

Find the mean for Italy

$$\text{mean} = \frac{\text{sum}}{\text{amount}}$$

Find the standard deviation for Italy

$$\text{standard deviation} = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

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

- The amount of tourists varied less for Italy than Spain because the standard deviation was smaller for Italy.
- The amount of tourists varied more for Italy than Spain because the standard deviation was greater for Italy.
- On average less tourists visited Italy than Spain because the mean was smaller for Italy.
- On average more tourists visited Italy than Spain because the mean was greater for Italy.

**13** The heights of a group of ten-year-old girls have a mean of 137.5 cm and a standard deviation of 4.9 cm.

(a) Sophie is ten-year-old girl with a standardised score of 0.

Find Sophie's height.

(1 mark)

A standardized score of 0 indicates that the value is equal to the mean of the distribution.

\_\_\_\_\_ cm

(b) Amelia and Ruby are both ten-year-old girls in the group.

Amelia's standardised score for height is 1.1 cm.

Ruby's standardised score for height is -0.3 cm.

Amelia is taller than Ruby.

How much taller is Amelia?

(3 marks)

Rearrange the formula to make the value the subject

$$\text{Standardised score} = \frac{\text{value} - \text{mean}}{\text{standard deviation}}$$

\_\_\_\_\_ cm

(c) Fatima takes a sample of 3 ten-year-old girls from the group.

She wants to calculate the standardised score for the sample mean of their ages.

(i) Discuss the appropriateness of using 137.5 cm as the mean in the calculation of the standardised score,

(ii) Discuss the appropriateness of using 4.9 cm as the standard deviation in the calculation of the standardised score.

(4 marks)

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

Using 137.5 cm as the mean is not appropriate

Using 137.5 cm as the mean is appropriate

because the sample mean will be smaller than the population mean.

because the sample mean will be close to the population mean.

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

Using 4.9 cm as the standard deviation is appropriate

Using 4.9 cm as the standard deviation is not appropriate

because the sample standard deviation will be close to the population standard deviation.

because the sample mean will be more closely distributed than the individual values.