

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

Edexcel Higher - 2026

Higher Tier

Variant 1 (same as video)

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!



(b) Find the probability that both of David's test results are accurate.

(2 marks)

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

(c) David states that the probability that exactly one of the tests is false positive is less than 4%.  
Find out whether or not David is correct.

(3 marks)

Find the probability of exactly one of the tests is false positive (there are two outcomes on the tree diagram).

---

---

---

---

Select **one** box.

- The probability that exactly one of the tests is false positive is more than 4%, so David is not correct.
- The probability that exactly one of the tests is false positive is less than 4%, so David is not correct.
- The probability that exactly one of the tests is false positive is more than 4%, so David is correct.
- The probability that exactly one of the tests is false positive is less than 4%, so David is correct.

2 The table shows information about cars for sale in Cambridge.

number of doors	number of cars
2	340
3	180
4	260
5	520
6 or more	500
Total	1800

A researcher wants to investigate the price of these cars and takes a stratified sample of 90 cars according to the number of doors.

(a) The researcher says the mode of the number of doors for these cars is 5.

Explain how the researcher knows this.

(1 mark)

Select **one** box.

- 5 is the middle number.
- 5 doors has the highest frequency.
- 5 is the difference between the largest and smallest number.
- 5 cars has the highest frequency.

(b) Work out the number of cars in the sample for each number of doors.

number of doors	number of cars in the sample
2	
3	
4	
5	
6 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 90 cars in the sample should be selected.

(3 marks)

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

- Ensure that all 1800 cars are included in the sample.
- Number each of the cars, and then use the random numbers to select the required amount of cars.
- Select the first 90 cars.
- Use a sampling frame for each strata.
- Generate random numbers, remove repeats or numbers out of range.
- Complete two of the strata.

- 3 Aisha works for an environmental agency. She has been tasked with investigating air pollution levels near schools.

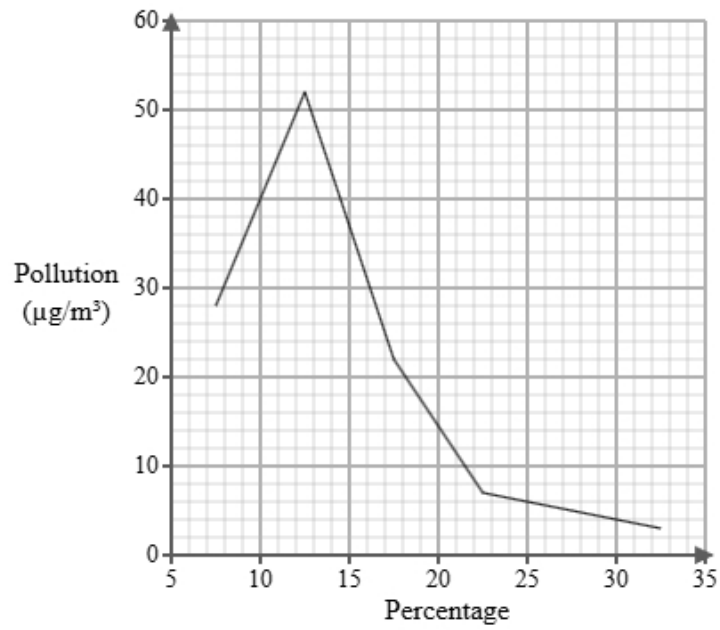
Below is a section of the spreadsheet she used to record her findings.

Pollution ( $\mu\text{g}/\text{m}^3$ )	Percentage
$5 < p \leq 10$	6
$10 < p \leq 15$	eight
$15 < p \leq 20$	8
$20 < p \leq 25$	117
$25 < p \leq 30$	48
$30 < p \leq 35$	13
Total	100

Aisha cleans the data to create the table below.

Pollution ( $\mu\text{g}/\text{m}^3$ )	Percentage
$5 < p \leq 10$	6
$10 < p \leq 15$	8
$15 < p \leq 20$	8
$20 < p \leq 25$	17
$25 < p \leq 30$	48
$30 < p \leq 35$	13
Total	100

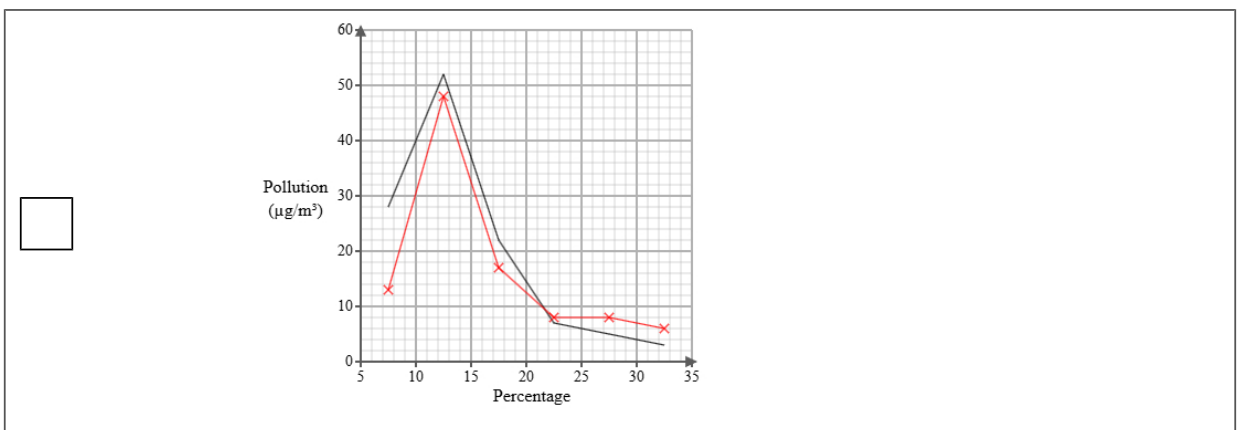
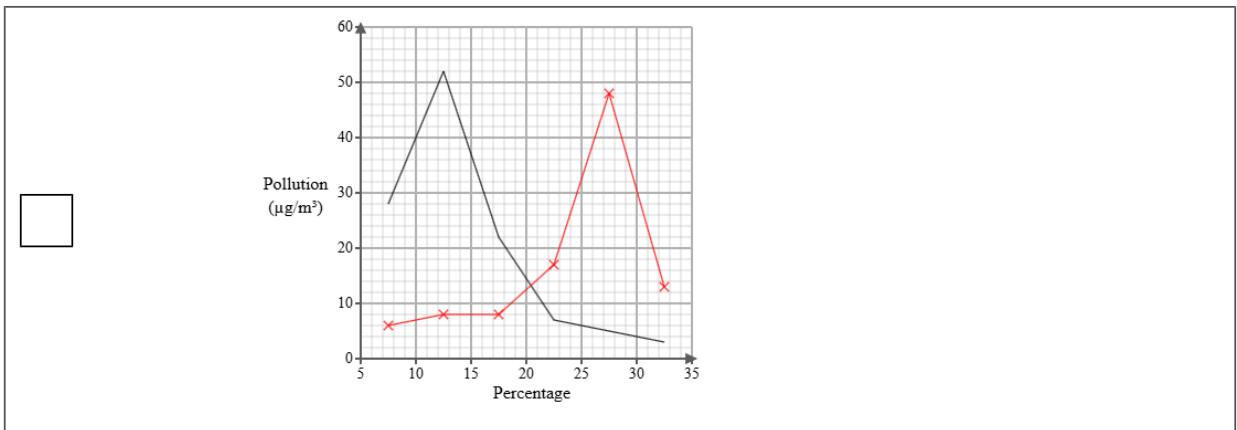
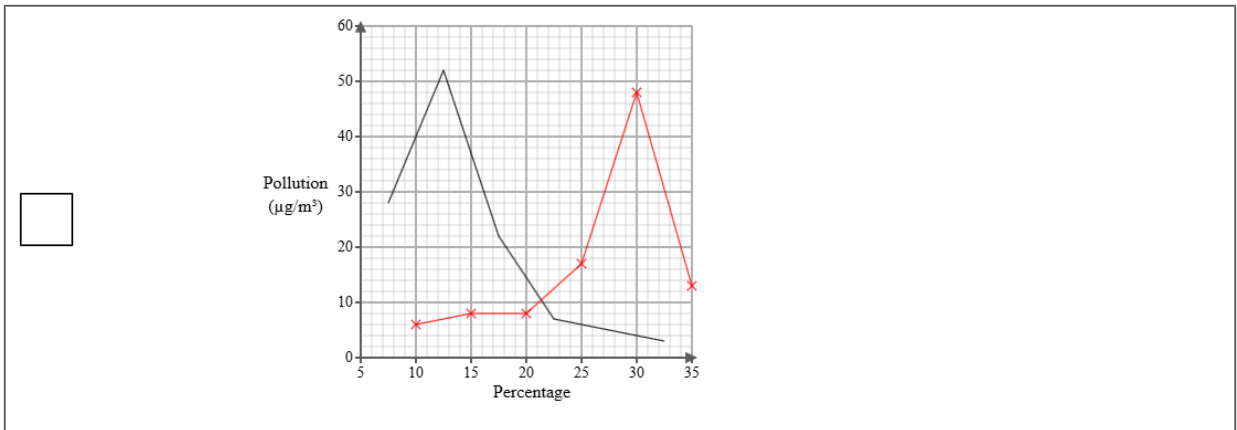
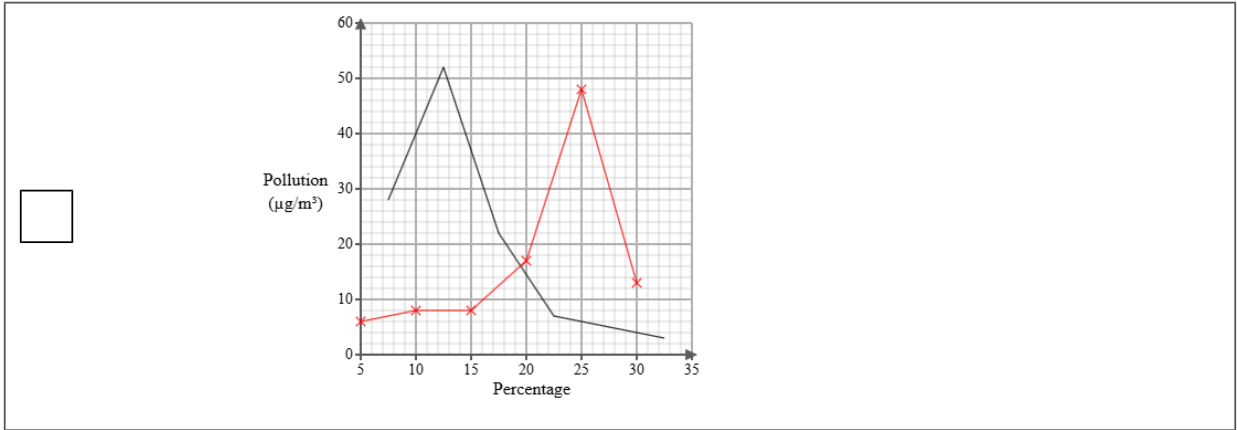
A frequency polygon has been drawn for air pollution levels near parks.



- i) On the same graph, draw the frequency polygon for air pollution levels near schools.
- 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).

- This means that for the parks the pollution is mainly at the upper end of the distribution and for the schools the pollution is mainly at the lower end of the distribution.
- This means that for the parks the pollution is mainly at the lower end of the distribution and for the schools the pollution is mainly at the upper end of the distribution.
- The distribution of air pollution levels near parks is symmetrical whereas the distribution of air pollution levels near schools is positively skewed.
- The distribution of air pollution levels near parks is positively skewed whereas the distribution of air pollution levels near schools is negatively skewed.
- This means that for the parks the pollution is equally spread out on either side of the median and for the schools the pollution is mainly at the upper end of the distribution.
- The distribution of air pollution levels near parks is negatively skewed whereas the distribution of air pollution levels near schools is positively skewed.

- 4 The table shows information about the consumer price index (CPI) and TV Licence price (£) in the United Kingdom for Jan 2000, Jan 2010 and Jan 2020.

	Jan 2000	Jan 2010	Jan 2020
consumer price index	100	123	153
TV Licence price (£)	104	145.5	157.5

Describe how the increase in TV Licence price (£) compares with the CPI over the ten years to Jan 2010 and over the twenty years to Jan 2020.

(5 marks)

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

$\frac{145.5}{123} \times 100 = 118$  (nearest integer)

$\frac{157.5}{153} \times 100 = 105$  (nearest integer)

Between Jan 2000 and Jan 2010 the change in price was less than the CPI.

Between Jan 2000 and Jan 2020 the change in price was more than the CPI.

Between Jan 2000 and Jan 2010 the change in price was more than the CPI.

Between Jan 2000 and Jan 2020 the change in price was less than the CPI.

$\frac{157.5}{104} \times 100 = 151$  (nearest integer)

$\frac{145.5}{104} \times 100 = 140$  (nearest integer)

5 Sarah is investigating how the age in years,  $x$ , affects the resale price (£),  $y$  for two types of laptops, model A and model B.

She found ten laptops 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
A	-120	1500
B	-95	2000

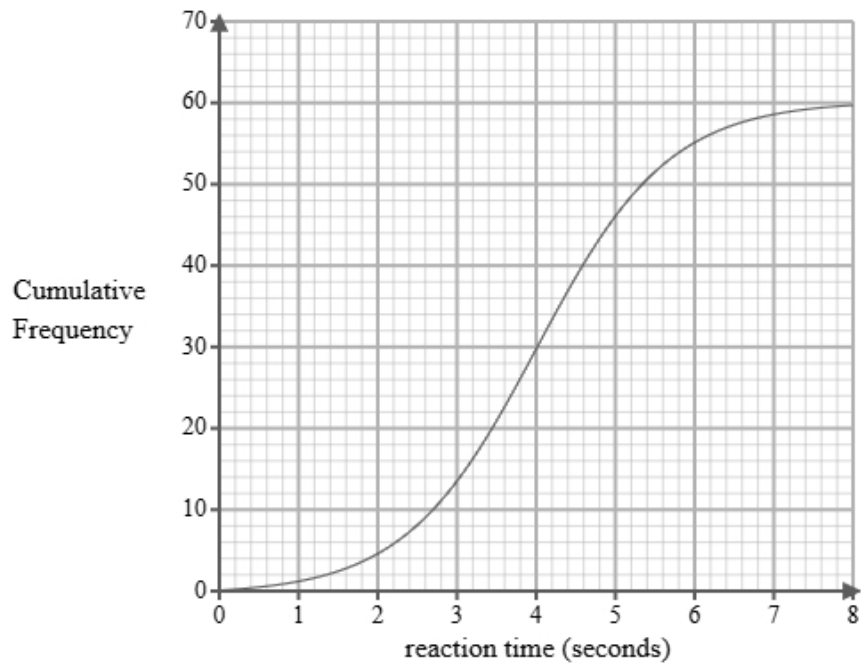
Interpret and compare these results in context.

(5 marks)

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

- Model A reduces in resale price more per year than Model B.
- Model A has a greater initial resale price.
- Model A changes in resale price by £1500 per year.
- Model B changes in resale price by £2000 per year.
- Model A reduces in resale price less than Model B.
- Both laptops increase in resale price as the age increase.
- Model A reduces in resale price by £120 per year.
- Both laptops decrease in resale price as the age increase.
- Model B has a greater initial resale price.
- Model B reduces in resale price by £95 per year.

- 6 A researcher measures the reaction times, in seconds, of 60 students completing a computer task. 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
	2.1	4.6	6.5

<input type="checkbox"/>	Lower quartile	Median	Upper quartile
	3.1	4	4.9

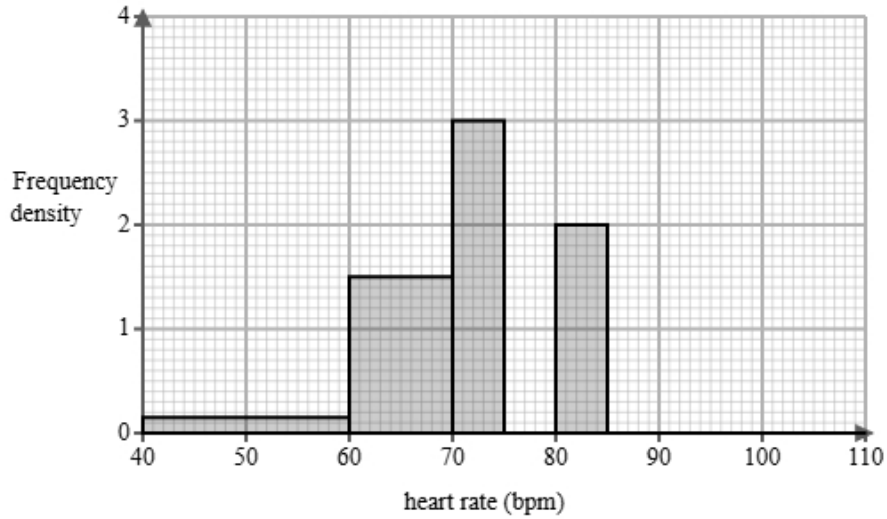
<input type="checkbox"/>	Lower quartile	Median	Upper quartile
	2.6	4.6	5.6

<input type="checkbox"/>	Lower quartile	Median	Upper quartile
	3.9	4	4.5

7 The heart rate is recorded in beats per minute (bpm).

A fitness instructor is analysing the resting heart rates of a group of 60 adult women after 12 hours without physical activity.

The partially completed histogram and grouped frequency table provide details about their heart rate measurements.



heart rate $h$ (bpm)	Frequency
$40 < h \leq 60$	3
$60 < h \leq 70$	15
$70 < h \leq 75$	
$75 < h \leq 80$	12
$80 < h \leq 85$	
$85 < h \leq 110$	5

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

(2 marks)

Use the frequency density formula

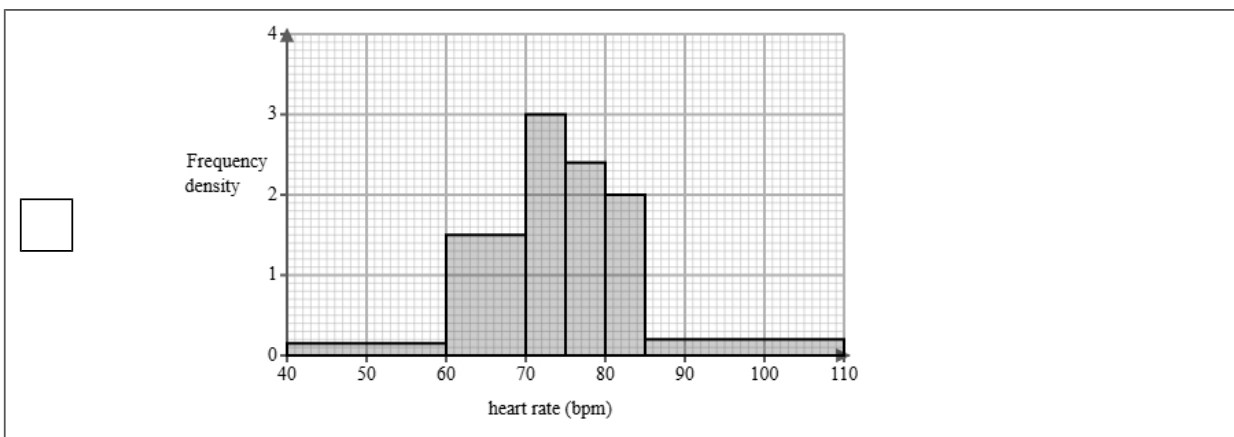
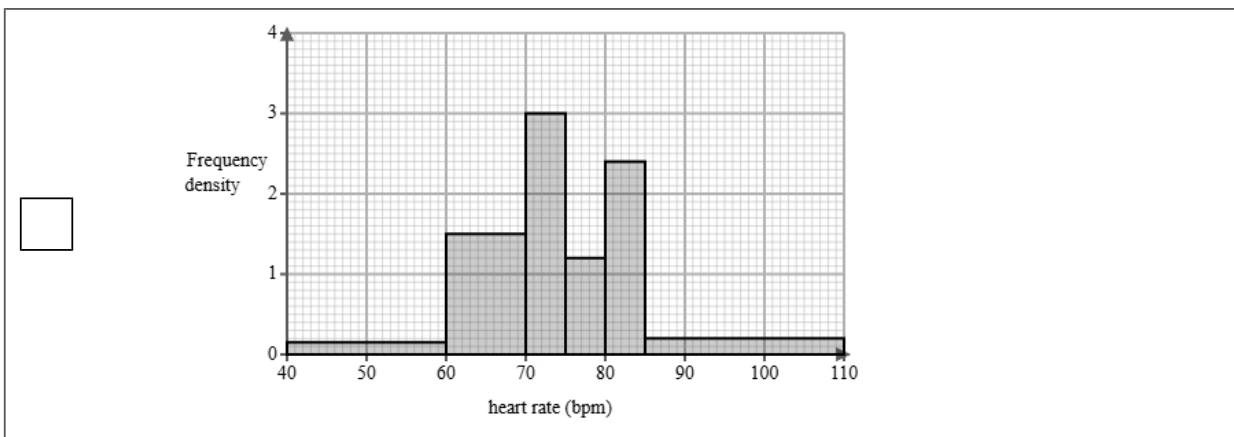
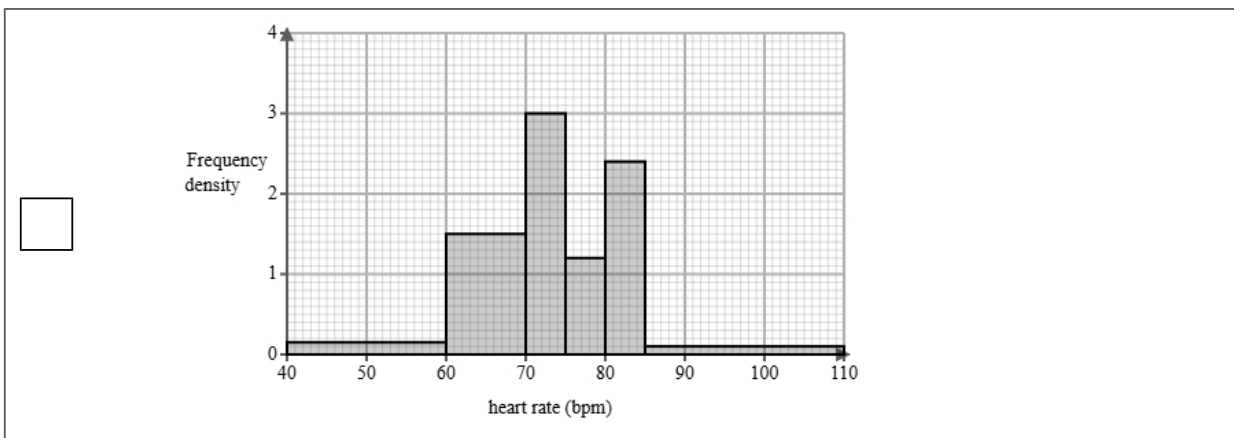
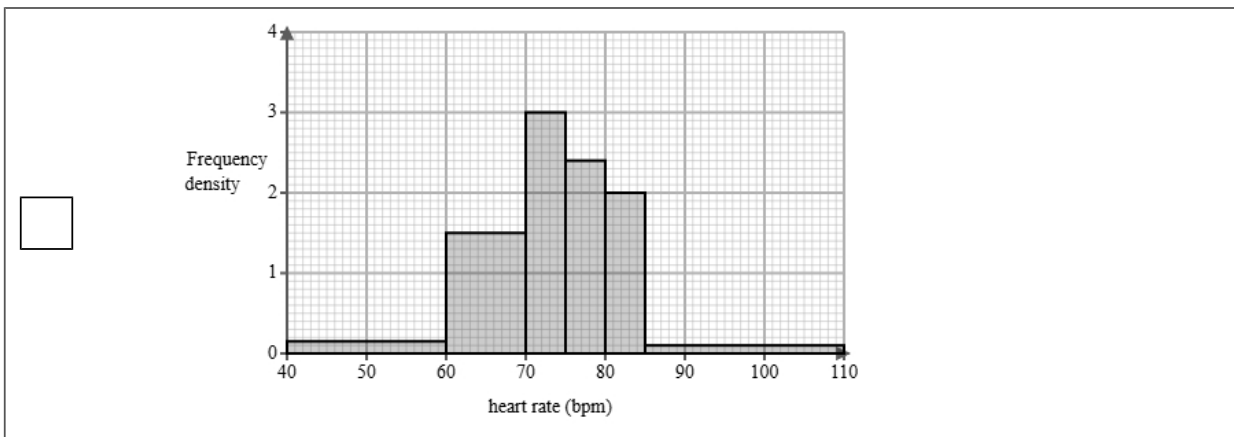
frequency = frequency density  $\times$  class width

heart rate $h$ (bpm)	Frequency
$40 < h \leq 60$	3
$60 < h \leq 70$	15
$70 < h \leq 75$	_____
$75 < h \leq 80$	12
$80 < h \leq 85$	_____
$85 < h \leq 110$	5

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

(2 marks)

Select the correct answer.



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

$$\sum h = 4467$$

$$\sum h^2 = 337697$$

$$n = 60$$

Explain whether or not there may be any outliers in the fitness instructor'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  $40 < h \leq 60$  and the upper outlier limit is within the group  $85 < h \leq 110$ .

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

- (d) A different fitness instructor is analysing the resting heart rates of a group of adult men after 12 hours without physical activity.

They find the following summary statistics for the data.

mean = 79.56

median = 83

standard deviation = 11.2

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 15th percentile.

(1 mark)

Select **one** box.

- $60 < h \leq 70$
- $70 < h \leq 75$
- $40 < h \leq 60$
- $75 < h \leq 80$
- $80 < h \leq 85$

- 8 Jake 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>
Albion	168	5	6	-1	1
Bristol	162	2	1	1	1
Canterbury	171	7	8	-1	1
Durham	165	4	4	0	0
Exeter	170	6	5		
Fulham	164	3	2		
Guildford	173	8	7		
Harrow	161	1	3		

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

Suggest a diagram that Jake could draw.

(1 mark)

Select **one** box.

- Histogram
- Pie chart
- Scatter diagram
- Bar chart

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

Oliver suggests that Jake 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
- Jake used the correct method, Oliver's suggestion is not appropriate.
- Spearman's rank correlation is influenced by outliers.
- Oliver's suggestion is more appropriate than Jake's method.
- PMCC measures linear correlation.
- PMCC compares bivariate data

- 9 Aisha works for an environmental agency. She has been tasked with investigating air pollution levels near schools.

Below is a section of the spreadsheet she used to record her findings.

Pollution ( $\mu\text{g}/\text{m}^3$ )	Percentage
$5 < p \leq 10$	6
$10 < p \leq 15$	eight
$15 < p \leq 20$	8
$20 < p \leq 25$	117
$25 < p \leq 30$	48
$30 < p \leq 35$	13
Total	100

Aisha cleans the data to create the table below.

Pollution ( $\mu\text{g}/\text{m}^3$ )	Percentage
$5 < p \leq 10$	6
$10 < p \leq 15$	8
$15 < p \leq 20$	8
$20 < p \leq 25$	17
$25 < p \leq 30$	48
$30 < p \leq 35$	13
Total	100

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

(1 mark)

Select **one** box.

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

- (b) Use linear interpolation to work out an estimate of the median air pollution level.  
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

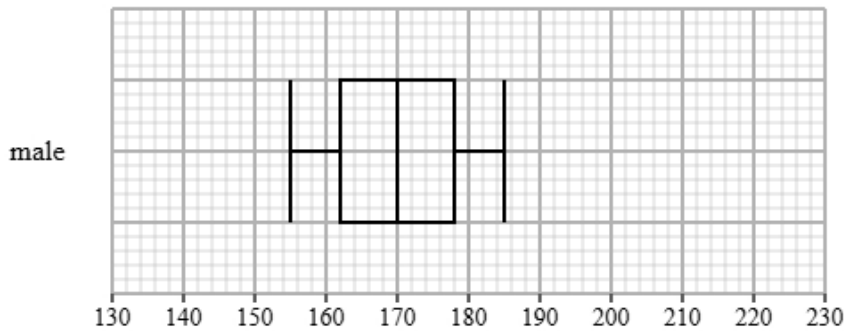
width of group = w

\_\_\_\_\_  $\mu\text{g}/\text{m}^3$

10 Sophie recorded the heights of male and female students in a school.

Both groups were measured using the same method.

The box plot shows information about the heights for the male students.



The table gives information about the heights for the female students.

Least tall	Lower quartile	Median	Upper quartile	Most tall
150	158	162	172	190

Compare the two distributions of heights.

Give three comparisons and interpret one of these comparisons.

(4 marks)

Select **one** box.

- The median is bigger.
- The median heights for male and female students are equal.
- The median heights for male students is greater than female students.
- The median heights for male students is lower than female students.

Select **one** box.

- The IQR is bigger.
- The IQR for the heights of the male and female students are equal.
- The IQR for the heights of the male students is greater than female students.
- The IQR for the heights of the male students is lower than female students.

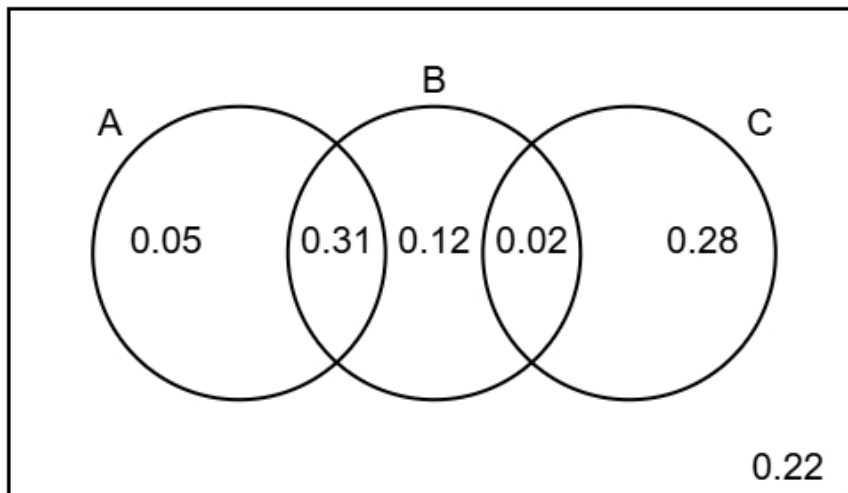
Select **one** box.

- The skews for the heights of the male and female students are both positive.
- The skew for the heights of the male students is symmetrical and the skew for the female students is negative.
- The skew for the heights of the male students is symmetrical and the skew for the female students is positive.
- The skews for the heights of the male and female students are both symmetrical.

Select **one** box.

- The heights for the male students are less spread out than the female students.
- The male students are more skewed than female students.
- The male students are on average taller than the female students.
- The male students are on average shorter than the female 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 B are mutually exclusive
- because they only intersect once.
- B and C are mutually exclusive
- because they do not intersect.
- A and C are mutually exclusive

(b) Find  $P(B)$

(1 mark)

We are looking for the probabilities inside B.

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

(2 marks)

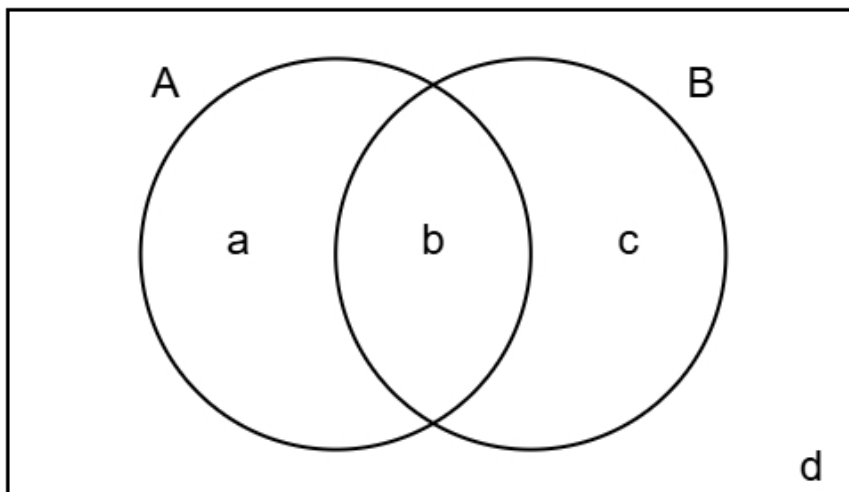
We are looking for the probabilities inside A or C.

\_\_\_\_\_

(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 Spain between 2010 and 2016

56 59 61 58 62 65 67

The table gives a summary of the amount, in millions, of tourists who visited France between 2010 and 2016

Mean	Standard Deviation	Largest Amount
60	7	69

Compare the amount of tourists in Spain and France between 2010 and 2016

You may use:

$$56^2 + 59^2 + 61^2 + 58^2 + 62^2 + 65^2 + 67^2 = 26260$$

(5 marks)

Find the mean for Spain

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

Find the standard deviation for Spain

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

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

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

**13** The ages of a group of boys have a mean of 5.6 years and a standard deviation of 0.8 years.

(a) Liam is boy with a standardised score of 0.

Find Liam's age.

(1 mark)

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

\_\_\_\_\_ years

(b) Daniel and Alex are both boys in the group.

Daniel's standardised score for age is 0.9 years.

Alex's standardised score for age is -0.5 years.

Daniel is older than Alex.

How much older is Daniel?

(3 marks)

Rearrange the formula to make the value the subject

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

\_\_\_\_\_ years

(c) Arjun takes a sample of 4 boys from the group.

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

(i) Discuss the appropriateness of using 5.6 years as the mean in the calculation of the standardised score,

(ii) Discuss the appropriateness of using 0.8 years 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 5.6 years as the mean is not appropriate

Using 5.6 years 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 0.8 years as the standard deviation is not appropriate

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

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

Using 0.8 years as the standard deviation is appropriate