

Statistics GCSE**Paper 2**

Edexcel Higher - 2026

Higher Tier

Variant 2

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.

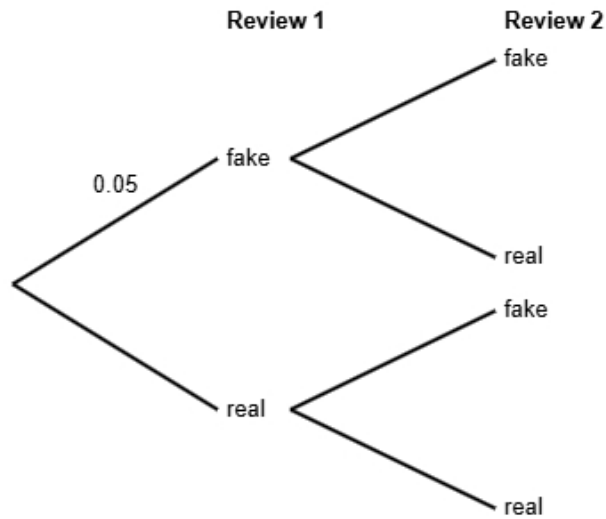
Copyright

- 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
- This paper and more are available on our site with questions that change subtly after each attempt.
- Good luck!

- 1 Research suggests that 5% of online product reviews are fake.
All other reviews are genuine.
Emma is reading two reviews for a product.
She does not know if each review is fake or real.



- (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 Emma's reviews are real.

(2 marks)

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

(c) Emma states that the probability that exactly one reviews is fake is less than 10%.
Find out whether or not Emma is correct.

(3 marks)

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

Select **one** box.

- The probability that exactly one reviews is fake is less than 10%, so Emma is not correct.
- The probability that exactly one reviews is fake is more than 10%, so Emma is not correct.
- The probability that exactly one reviews is fake is more than 10%, so Emma is correct.
- The probability that exactly one reviews is fake is less than 10%, so Emma is correct.

2 The table shows information about apartments for rent in Manchester.

number of rooms	number of apartments
1	225
2	180
3	495
4	120
5 or more	180
Total	1200

A researcher wants to investigate the price of these apartments and takes a stratified sample of 80 apartments according to the number of rooms.

(a) The researcher says the mode of the number of rooms for these apartments is 3.

Explain how the researcher knows this.

(1 mark)

Select **one** box.

3 rooms has the highest frequency.

3 is the middle number.

3 is the difference between the largest and smallest number.

3 apartments has the highest frequency.

(b) Work out the number of apartments in the sample for each number of rooms.

number of rooms	number of apartments in the sample
1	
2	
3	
4	
5 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 80 apartments in the sample should be selected.

(3 marks)

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

- Ensure that all 1200 apartments are included in the sample.
- Use a sampling frame for each strata.
- Generate random numbers, remove repeats or numbers out of range.
- Complete two of the strata.
- Number each of the apartments, and then use the random numbers to select the required amount of apartments.
- Select the first 80 apartments.

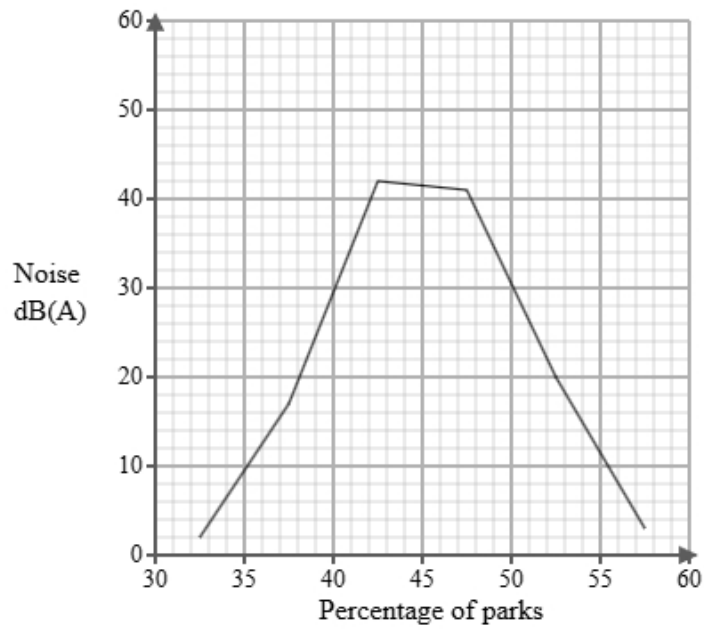
- 3 Luca works for a public health institute. He has been tasked with investigating noise levels in urban parks. Below is a section of the spreadsheet he used to record his findings.

Noise dB(A)	Percentage of parks
$30 < n \leq 35$	8
$35 < n \leq 40$	six
$40 < n \leq 45$	8
$45 < n \leq 50$	111
$50 < n \leq 55$	57
$55 < n \leq 60$	10
Total	100

Luca cleans the data to create the table below.

Noise dB(A)	Percentage of parks
$30 < n \leq 35$	8
$35 < n \leq 40$	6
$40 < n \leq 45$	8
$45 < n \leq 50$	11
$50 < n \leq 55$	57
$55 < n \leq 60$	10
Total	100

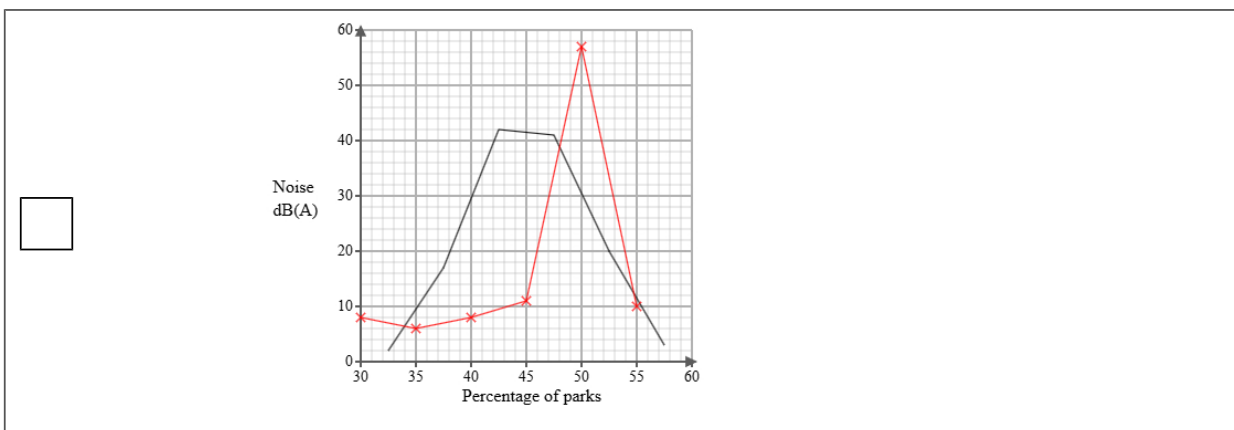
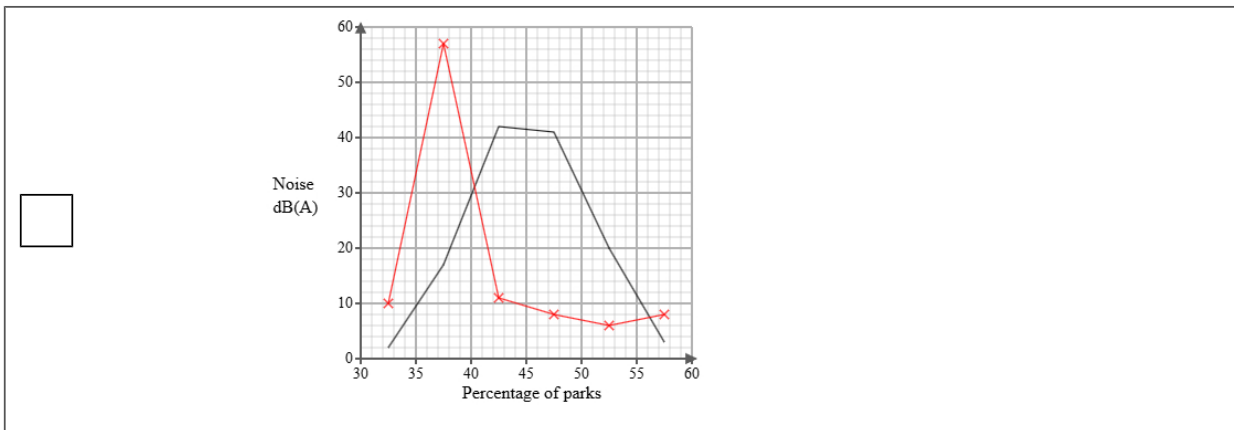
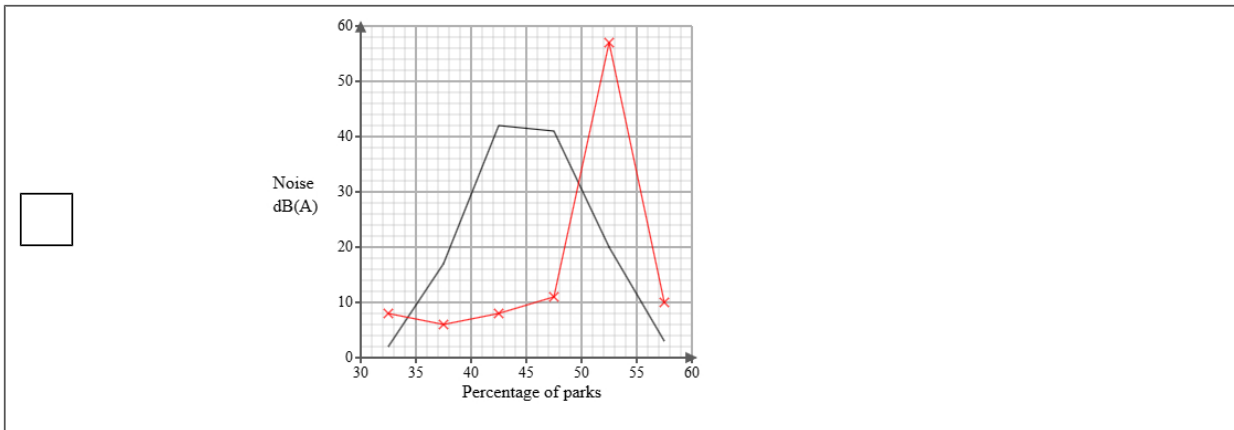
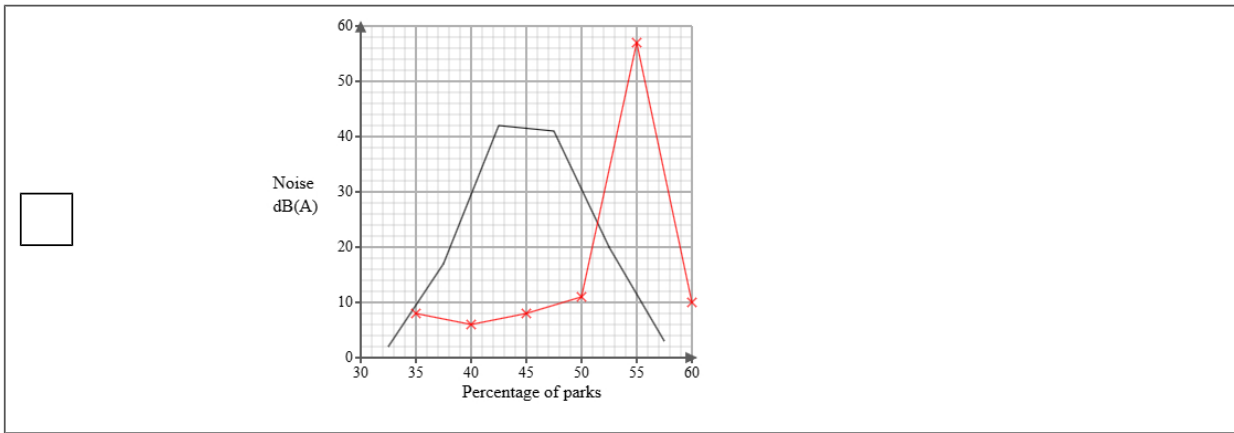
A frequency polygon has been drawn for noise levels in residential streets.



- i) On the same graph, draw the frequency polygon for noise levels in urban parks.
- 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 noise levels in residential streets is positively skewed whereas the distribution of noise levels in urban parks is symmetrical.
- The distribution of noise levels in residential streets is negatively skewed whereas the distribution of noise levels in urban parks is symmetrical.
- This means that in residential areas the noise levels were equally spread out on either side of the median and in urban parks the noise levels were mainly at the upper end of the distribution.
- This means that in residential areas the noise levels were mainly at the upper end of the distribution and in urban parks the noise levels were equally spread out on either side of the median.
- This means that in residential areas the noise levels were mainly at the lower end of the distribution and in urban parks the noise levels were mainly at the upper end of the distribution.
- The distribution of noise levels in residential streets is symmetrical whereas the distribution of noise levels in urban parks is negatively skewed.

- 4 The table shows information about the consumer price index (CPI) and average cost of a book (£) in England for Jan 1995, Jan 2005 and Jan 2015.

	Jan 1995	Jan 2005	Jan 2015
consumer price index	100	116	149
average cost of a book (£)	4.57	5.15	7.35

Describe how the increase in average cost of a book (£) compares with the CPI 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).

- Between Jan 1995 and Jan 2015 the change in price was more than the CPI.
- Between Jan 1995 and Jan 2005 the change in price was less than the CPI.
- $\frac{5.15}{116} \times 100 = 4$ (nearest integer)
- Between Jan 1995 and Jan 2015 the change in price was less than the CPI.
- $\frac{7.35}{149} \times 100 = 5$ (nearest integer)
- $\frac{5.15}{4.57} \times 100 = 113$ (nearest integer)
- Between Jan 1995 and Jan 2005 the change in price was more than the CPI.
- $\frac{7.35}{4.57} \times 100 = 161$ (nearest integer)

5 Tom is investigating how the distance travelled in km, x , affects the resale price (£), y for two types of bicycles, type A and type B.

He found ten bicycles of each type and recorded their distance travelled and resale 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:

Type	Gradient of line of best fit	y-intercept of line of best fit
A	-15	600
B	-10	750

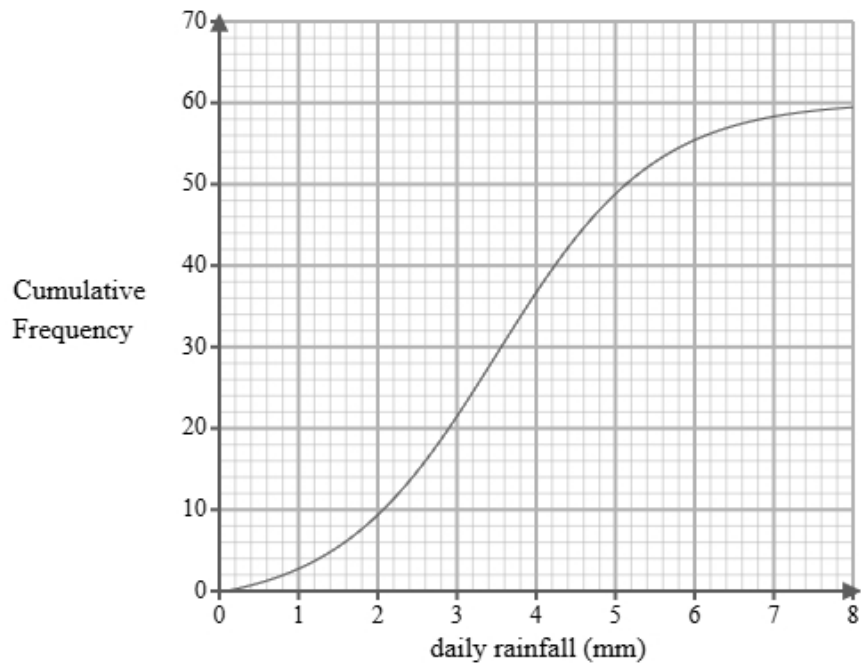
Interpret and compare these results in context.

(5 marks)

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

- Type A reduces in resale price less than Type B.
- Type A has a greater initial resale price.
- Type B reduces in resale price by £10 per km.
- Both bicycles increase in resale price as the distance travelled increase.
- Both bicycles decrease in resale price as the distance travelled increase.
- Type A changes in resale price by £600 per km.
- Type A reduces in resale price by £15 per km.
- Type A reduces in resale price more per km than Type B.
- Type B changes in resale price by £750 per km.
- Type B has a greater initial resale price.

- 6 A researcher measures the daily rainfall, in millimetres, recorded over 60 days in a small town. 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	4.1	5.9

<input type="checkbox"/>	Lower quartile	Median	Upper quartile
	1.8	4.1	5.3

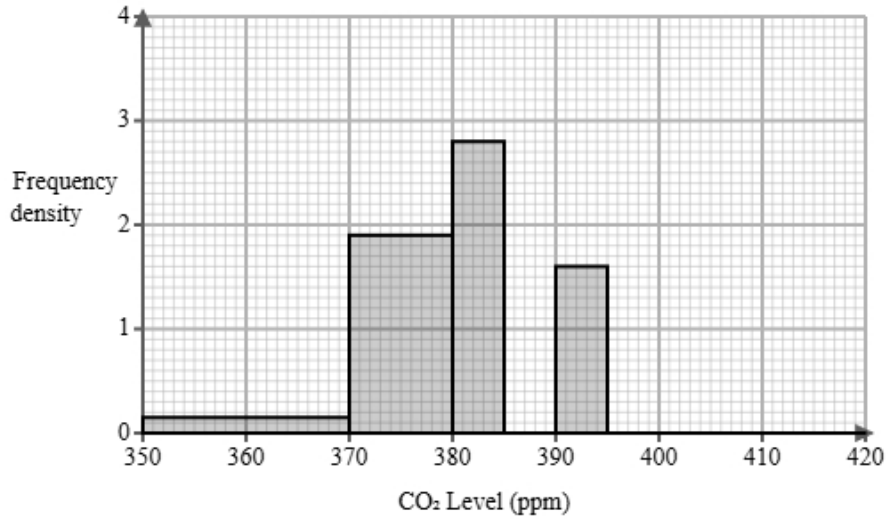
<input type="checkbox"/>	Lower quartile	Median	Upper quartile
	3.1	3.6	4.1

<input type="checkbox"/>	Lower quartile	Median	Upper quartile
	2.5	3.6	4.6

7 The concentration of carbon dioxide in the air is recorded in parts per million (ppm).

A scientist is analysing the CO₂ levels in the atmosphere of 60 urban areas during a 12-hour night period with minimal traffic.

The partially completed histogram and grouped frequency table provide details about these CO₂ concentrations.



CO ₂ Level c (ppm)	Frequency
$350 < c \leq 370$	3
$370 < c \leq 380$	19
$380 < c \leq 385$	
$385 < c \leq 390$	6
$390 < c \leq 395$	
$395 < c \leq 420$	10

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

(2 marks)

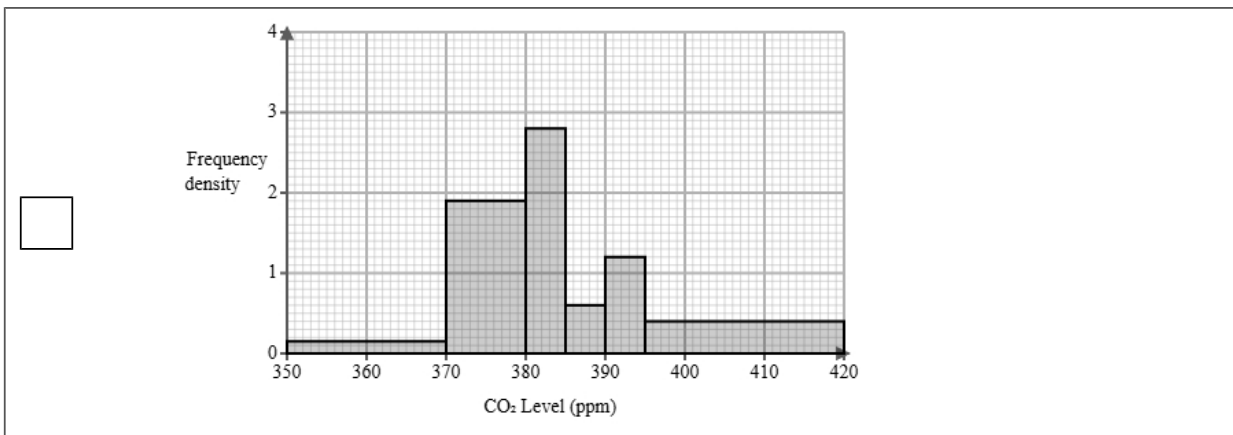
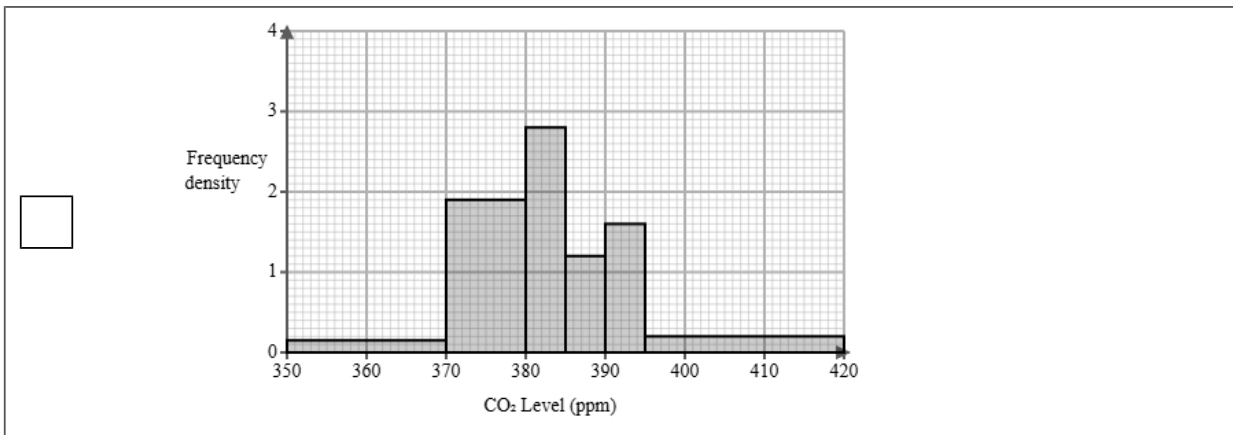
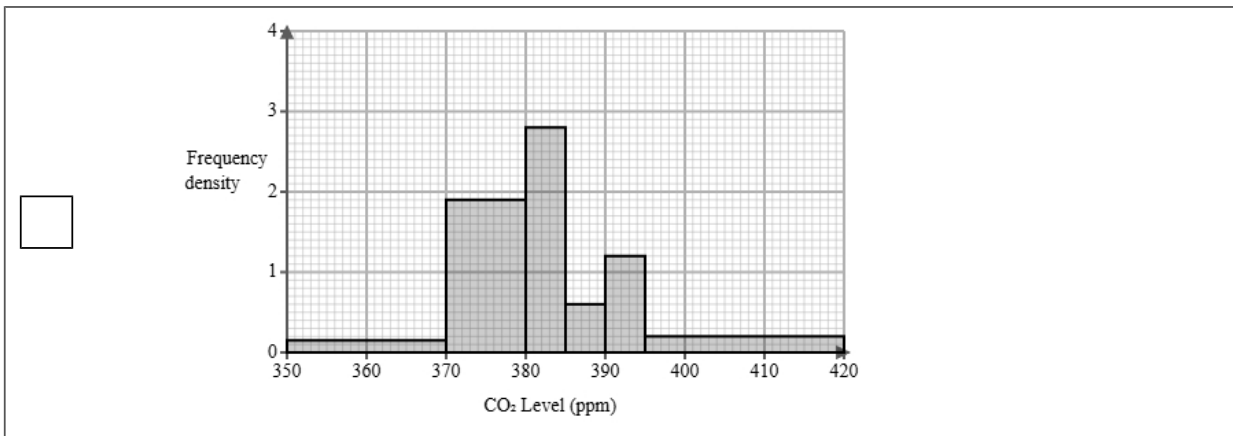
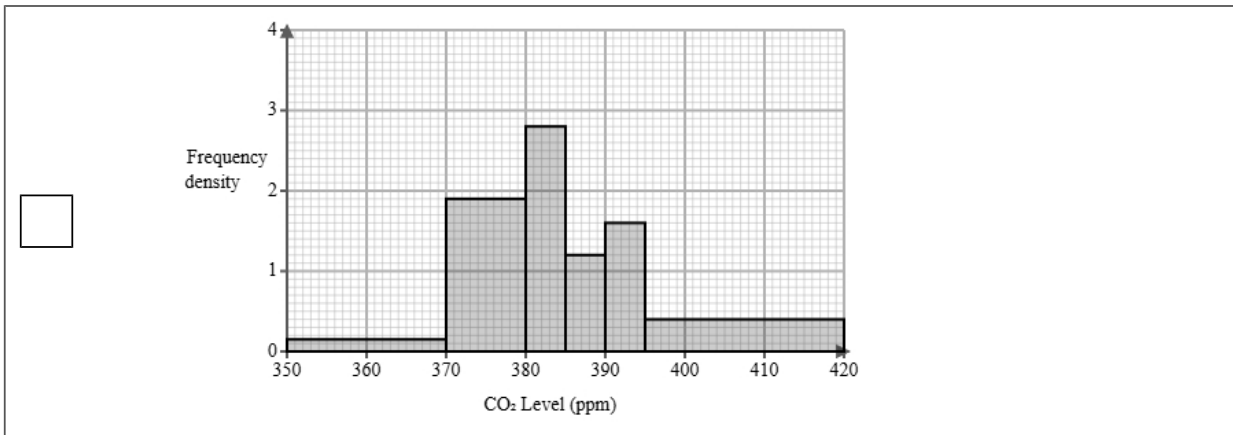
Use the frequency density formula
 $\text{frequency} = \text{frequency density} \times \text{class width}$

CO ,Level c (ppm)	Frequency
$350 < c \leq 370$	3
$370 < c \leq 380$	19
$380 < c \leq 385$	_____
$385 < c \leq 390$	6
$390 < c \leq 395$	_____
$395 < c \leq 420$	10

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

(2 marks)

Select the correct answer.



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

$$\sum c = 23091$$

$$\sum c^2 = 8891399$$

$$n = 60$$

Explain whether or not there may be any outliers in the scientist'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.

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

It is possible that there is an outlier as the lower outlier limit is within the group $350 < c \leq 370$ and the upper outlier limit is within the group $395 < c \leq 420$.

(d) A different scientist is analysing the CO₂ levels in the atmosphere of industrial areas during a 12-hour day period.

They find the following summary statistics for the data.

mean = 389.45

median = 412

standard deviation = 11.5

Calculate and interpret the skew for the industrial areas.

You must round your answer to 2 decimal places.

(3 marks)

Skew = _____

Select **one** box.

- The skew shows there is a negative correlation.
- There is a negative skew showing the data is not normally distributed.
- The skew shows that the three averages are equal.

(e) Find the class interval that contains the 35th percentile.

(1 mark)

Select **one** box.

- $350 < c \leq 370$
- $385 < c \leq 390$
- $380 < c \leq 385$
- $370 < c \leq 380$
- $390 < c \leq 395$

- 8 Leo 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 ²
Arundel	171	3	5	-2	4
Brighton	178	6	7	-1	1
Cambridge	179	7	6	1	1
Derby	176	5	3	2	4
Enfield	168	1	2		
Farnham	169	2	1		
Gloucester	180	8	8		
Hastings	173	4	4		

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

Suggest a diagram that Leo could draw.

(1 mark)

Select **one** box.

- Histogram
- Bar chart
- Pie chart
- Scatter 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).

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

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

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

- 9 Luca works for a public health institute. He has been tasked with investigating noise levels in urban parks. Below is a section of the spreadsheet he used to record his findings.

Noise dB(A)	Percentage of parks
$30 < n \leq 35$	8
$35 < n \leq 40$	six
$40 < n \leq 45$	8
$45 < n \leq 50$	111
$50 < n \leq 55$	57
$55 < n \leq 60$	10
Total	100

Luca cleans the data to create the table below.

Noise dB(A)	Percentage of parks
$30 < n \leq 35$	8
$35 < n \leq 40$	6
$40 < n \leq 45$	8
$45 < n \leq 50$	11
$50 < n \leq 55$	57
$55 < n \leq 60$	10
Total	100

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

(1 mark)

Select **one** box.

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

- (b) Use linear interpolation to work out an estimate of the median noise 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

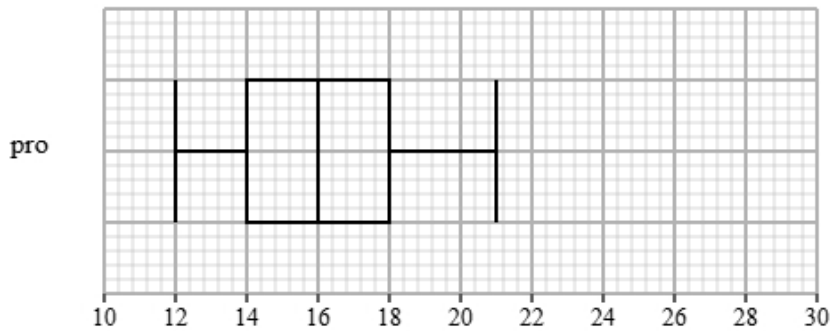
width of group = w

_____ dB(A)

10 Liam recorded the completion times for pro and beginner runners in a 5K race.

Both groups ran the same course.

The box plot presents data on the completion times for the pro runners.



The table gives information about the completion times for the beginner runners.

Least tall	Lower quartile	Median	Upper quartile	Most tall
17	19	20	24	30

Compare the two distributions of completion times.

Give three comparisons and interpret one of these comparisons.

(4 marks)

Select **one** box.

- The median is bigger.
- The median completion times for pro runners is greater than beginner runners.
- The median completion times for pro runners is lower than beginner runners.
- The median completion times for pro and beginner runners are equal.

Select **one** box.

- The IQR is bigger.
- The IQR for the completion times of the pro runners is greater than beginner runners.
- The IQR for the completion times of the pro and beginner runners are equal.
- The IQR for the completion times of the pro runners is lower than beginner runners.

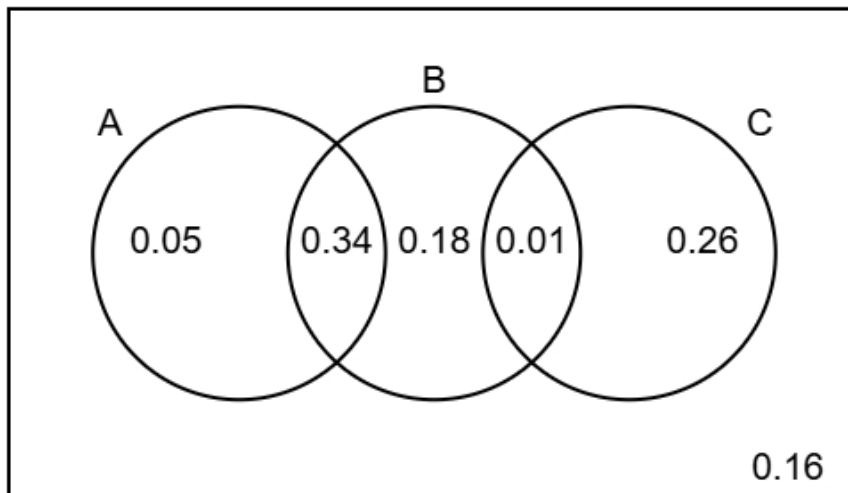
Select **one** box.

- The skews for the completion times of the pro and beginner runners are both positive.
- The skew for the completion times of the pro runners is symmetrical and the skew for the beginner runners is positive.
- The skews for the completion times of the pro and beginner runners are both symmetrical.
- The skew for the completion times of the pro runners is symmetrical and the skew for the beginner runners is negative.

Select **one** box.

- The times for the pro runners are more spread out than the beginner runners.
- The pro runners are on average slower than the beginner runners.
- The pro runners are on average faster than the beginner runners.
- The pro runners are more skewed than beginner runners.

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 lowest total probability.
- B and C are mutually exclusive
- A and C are mutually exclusive
- because they do not overlap.
- A and B are mutually exclusive
- because they only overlap once.

(b) Find $P(B)$

(1 mark)

We are looking for the probabilities inside B.

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

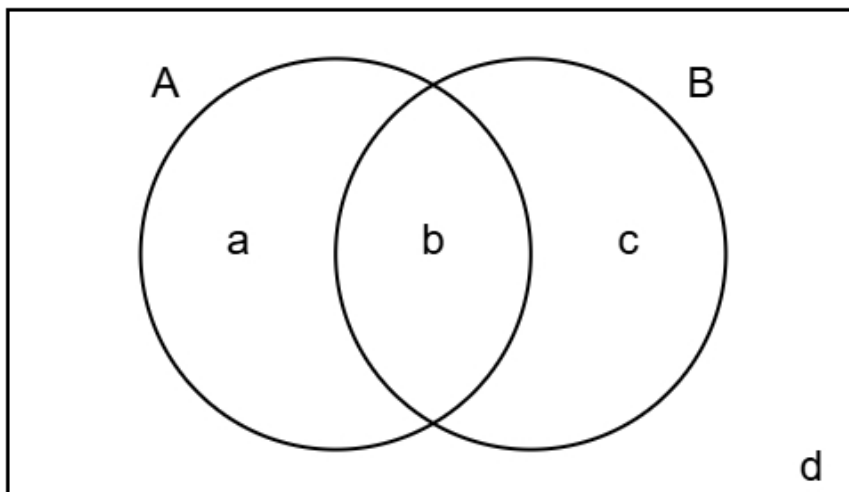
(2 marks)

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 Japan between 2012 and 2018

24 26 28 27 29 31 33

The table gives a summary of the amount, in millions, of tourists who visited USA between 2012 and 2018

Mean	Standard Deviation	Largest Amount
29	2	32

Compare the amount of tourists in Japan and USA between 2012 and 2018

You may use:

$$24^2 + 26^2 + 28^2 + 27^2 + 29^2 + 31^2 + 33^2 = 5656$$

(5 marks)

Find the mean for Japan

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

Find the standard deviation for Japan

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

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

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

13 The reaction times of a group of adult drivers have a mean of 1.4 seconds and a standard deviation of 0.3 seconds.

(a) Mark is adult driver with a standardised score of 0.

Find Mark's reaction time.

(1 mark)

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

_____ seconds

(b) Priya and Sarah are both adult drivers in the group.

Priya's standardised score for reaction time is 0.8 seconds.

Sarah's standardised score for reaction time is -1.2 seconds.

Priya had a slower reaction time than Sarah.

How much slower is Priya?

(3 marks)

Rearrange the formula to make the value the subject

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

_____ seconds

(c) Mina takes a sample of 5 adult drivers from the group.

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

(i) Discuss the appropriateness of using 1.4 seconds as the mean in the calculation of the standardised score,

(ii) Discuss the appropriateness of using 0.3 seconds 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 1.4 seconds as the mean is appropriate
- because the sample mean will be close to the population mean.
- because the sample mean will be smaller than the population mean.
- Using 1.4 seconds as the mean is not appropriate

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

- Using 0.3 seconds as the standard deviation is not appropriate
- Using 0.3 seconds as the standard deviation is 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.