

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

Edexcel Higher - 2026

Higher Tier

Variant 5

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.

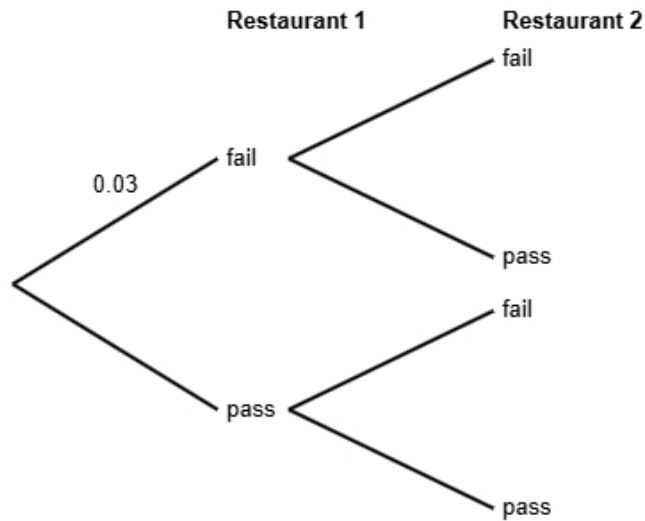
**Copyright**

- 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 Health inspections show that 3% of restaurants in a city fail basic hygiene standards.  
The rest pass the required health standards.  
Emma is a food safety officer and will be inspecting two different restaurants.  
She does not know if each restaurant will fail or pass.



- (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 restaurants have an outcome of pass.

(2 marks)

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

(c) Emma states that the probability that exactly one restaurant outcome is fail is less than 6%.  
Find out whether or not Emma is correct.

(3 marks)

Find the probability of exactly one restaurant outcome is fail (there are two outcomes on the tree diagram).

---

---

---

---

Select **one** box.

- The probability that exactly one restaurant outcome is fail is less than 6%, so Emma is correct.
- The probability that exactly one restaurant outcome is fail is more than 6%, so Emma is not correct.
- The probability that exactly one restaurant outcome is fail is more than 6%, so Emma is correct.
- The probability that exactly one restaurant outcome is fail is less than 6%, so Emma is not correct.

2 The table shows information about office spaces in Leeds.

number of desks	number of offices
1	120
2	220
3	740
4	400
5 or more	220
Total	1700

A researcher wants to investigate the price of these office spaces and takes a stratified sample of 85 office spaces according to the number of desks.

(a) The researcher says the mode of the number of desks for these office spaces is 3.

Explain how the researcher knows this.

(1 mark)

Select **one** box.

3 office spaces has the highest frequency.

3 is the middle number.

3 is the difference between the largest and smallest number.

3 desks has the highest frequency.

(b) Work out the number of offices in the sample for each number of desks.

number of desks	number of offices 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 85 office spaces in the sample should be selected.

(3 marks)

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

- Use a sampling frame for each strata.
- Number each of the office spaces, and then use the random numbers to select the required amount of office spaces.
- Complete two of the strata.
- Select the first 85 office spaces.
- Generate random numbers, remove repeats or numbers out of range.
- Ensure that all 1700 office spaces are included in the sample.

- 3 Mei works for a marine conservation group. She has been tasked with investigating plastic waste collected by divers.

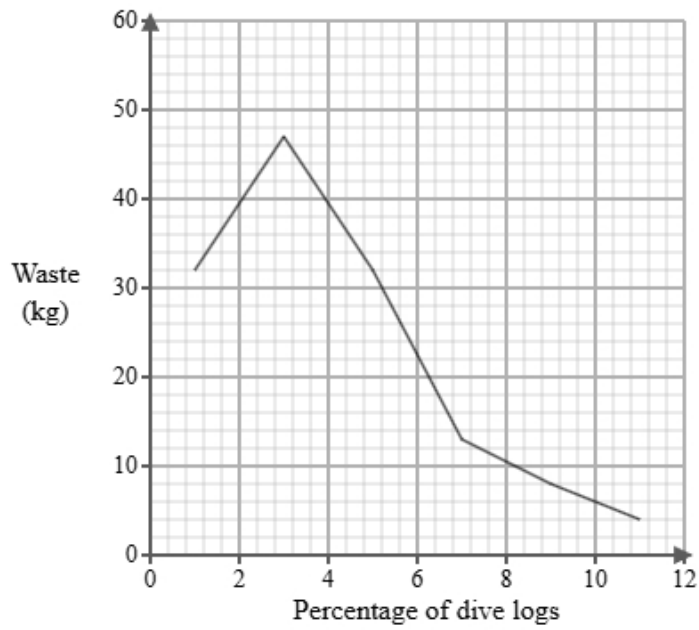
Below is a section of the spreadsheet she used to record her findings.

Waste (kg)	Percentage of dive logs
$0 < p \leq 2$	7
$2 < p \leq 4$	seven
$4 < p \leq 6$	6
$6 < p \leq 8$	118
$8 < p \leq 10$	47
$10 < p \leq 12$	15
Total	100

Mei cleans the data to create the table below.

Waste (kg)	Percentage of dive logs
$0 < p \leq 2$	7
$2 < p \leq 4$	7
$4 < p \leq 6$	6
$6 < p \leq 8$	18
$8 < p \leq 10$	47
$10 < p \leq 12$	15
Total	100

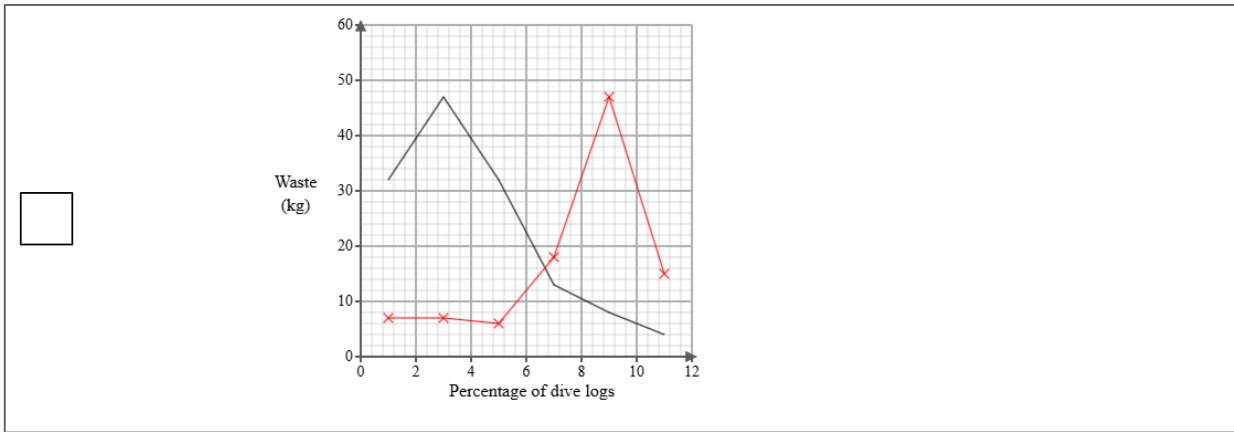
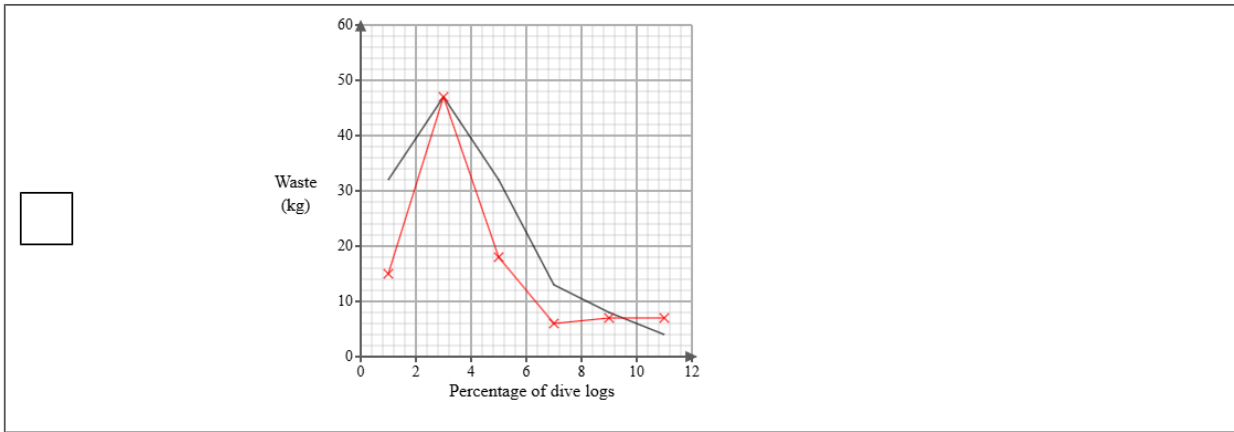
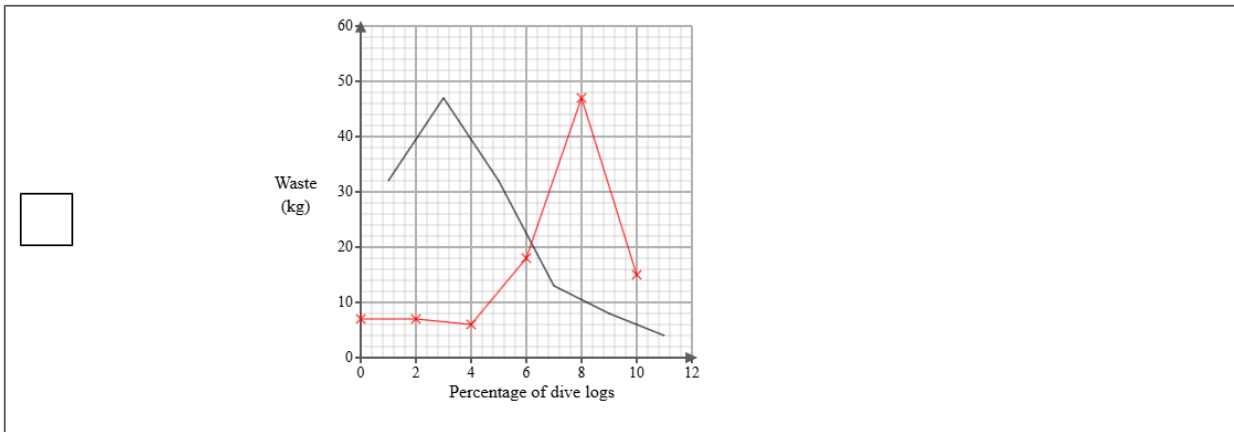
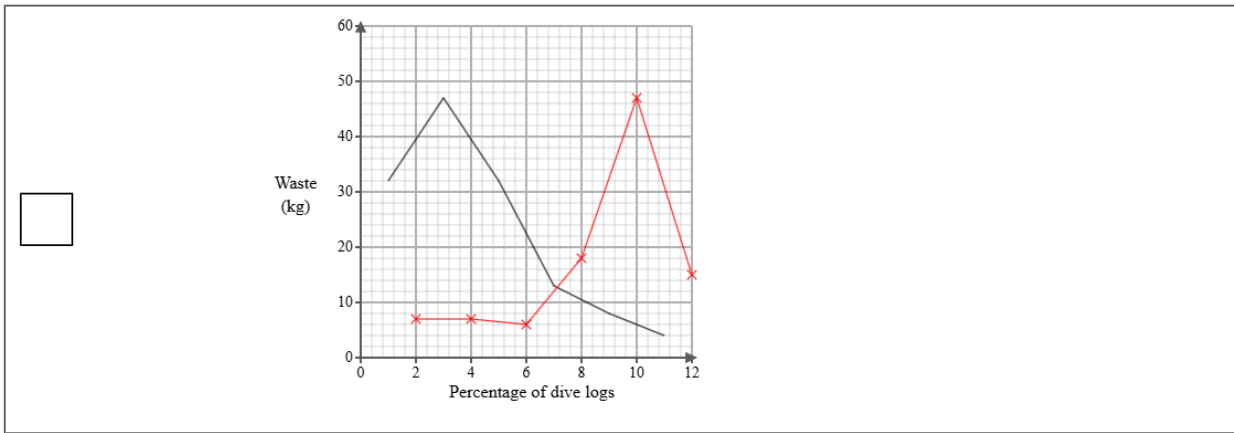
A frequency polygon has been drawn for metal waste collected by divers.



- i) On the same graph, draw the frequency polygon for plastic waste collected by divers.
- 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 metal waste collected by divers is negatively skewed whereas the distribution of plastic waste collected by divers is positively skewed.
- The distribution of metal waste collected by divers is positively skewed whereas the distribution of plastic waste collected by divers is negatively skewed.
- This means that the metal waste collected by divers was mainly at the upper end of the distribution and the plastic waste collected by divers was mainly at the lower end of the distribution.
- The distribution of metal waste collected by divers is symmetrical whereas the distribution of plastic waste collected by divers is positively skewed.
- This means that the metal waste collected by divers was equally spread out on either side of the median and the plastic waste collected by divers was mainly at the upper end of the distribution.
- This means that the metal waste collected by divers was mainly at the lower end of the distribution and the plastic waste collected by divers was mainly at the upper end of the distribution.

- 4 The table shows information about the consumer price index (CPI) and average cost of bread (pence) in the United Kingdom for Jan 1990, Jan 2000 and Jan 2010.

	Jan 1990	Jan 2000	Jan 2010
consumer price index	100	130	160
average cost of bread (pence)	65	70	110

Describe how the increase in average cost of bread (pence) compares with the CPI over the ten years to Jan 2000 and over the twenty years to Jan 2010.

(5 marks)

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

- $\frac{70}{130} \times 100 = 54$  (nearest integer)
- Between Jan 1990 and Jan 2000 the change in price was less than the CPI.
- $\frac{110}{160} \times 100 = 69$  (nearest integer)
- Between Jan 1990 and Jan 2010 the change in price was less than the CPI.
- $\frac{110}{65} \times 100 = 169$  (nearest integer)
- Between Jan 1990 and Jan 2000 the change in price was more than the CPI.
- $\frac{70}{65} \times 100 = 108$  (nearest integer)
- Between Jan 1990 and Jan 2010 the change in price was more than the CPI.

5 Emma is investigating how the engine hours used,  $x$ , affects the resale price (£),  $y$  for two types of boats, brand A and brand B.

She found ten boats of each type and recorded their engine hours used 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:

Brand	Gradient of line of best fit	y-intercept of line of best fit
A	-200	30000
B	-150	38000

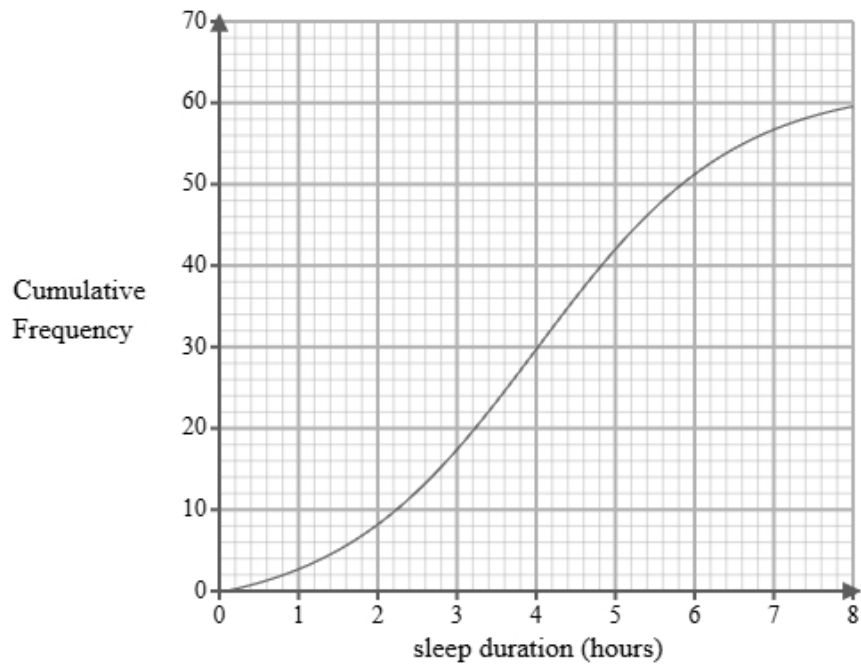
Interpret and compare these results in context.

(5 marks)

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

- Brand A reduces in resale price by £200 per engine hour.
- Brand A has a greater initial resale price.
- Brand A changes in resale price by £30000 per engine hour.
- Both boats decrease in resale price as the engine hours used increase.
- Brand B reduces in resale price by £150 per engine hour.
- Brand B changes in resale price by £38000 per engine hour.
- Both boats increase in resale price as the engine hours used increase.
- Brand A reduces in resale price less than Brand B.
- Brand A reduces in resale price more per engine hour than Brand B.
- Brand B has a greater initial resale price.

- 6 A researcher measures the number of hours that 60 students sleep on a specific Saturday. 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
	1.5	4.8	6.5

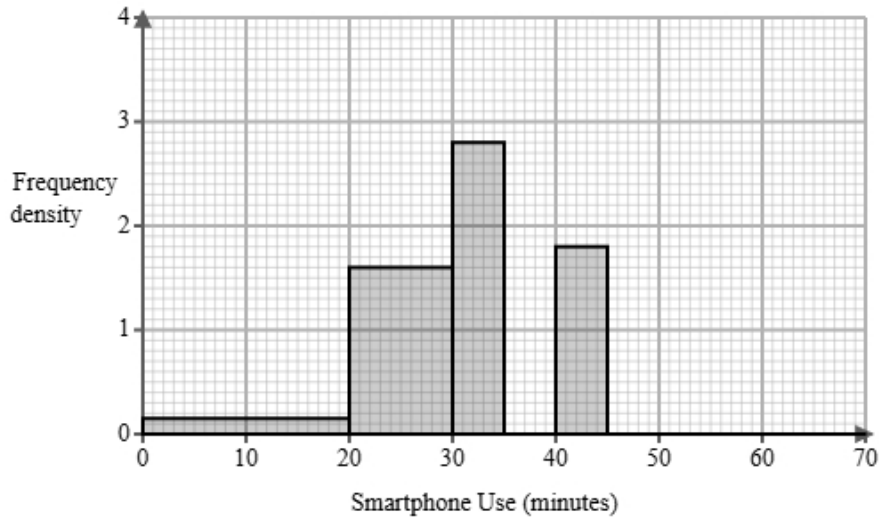
<input type="checkbox"/>	Lower quartile	Median	Upper quartile
	3	4	5.2

<input type="checkbox"/>	Lower quartile	Median	Upper quartile
	2.8	4	5.3

<input type="checkbox"/>	Lower quartile	Median	Upper quartile
	2	4.8	5.9

7 The time spent using a smartphone is recorded in minutes.

A behavioural researcher is analysing the smartphone usage of 60 adult females over a 1-hour lunch break. The partially completed histogram and grouped frequency table provide details about the recorded usage durations.



Smartphone Use $u$ (minutes)	Frequency
$0 < u \leq 20$	3
$20 < u \leq 30$	16
$30 < u \leq 35$	
$35 < u \leq 40$	13
$40 < u \leq 45$	
$45 < u \leq 70$	5

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

(2 marks)

Use the frequency density formula

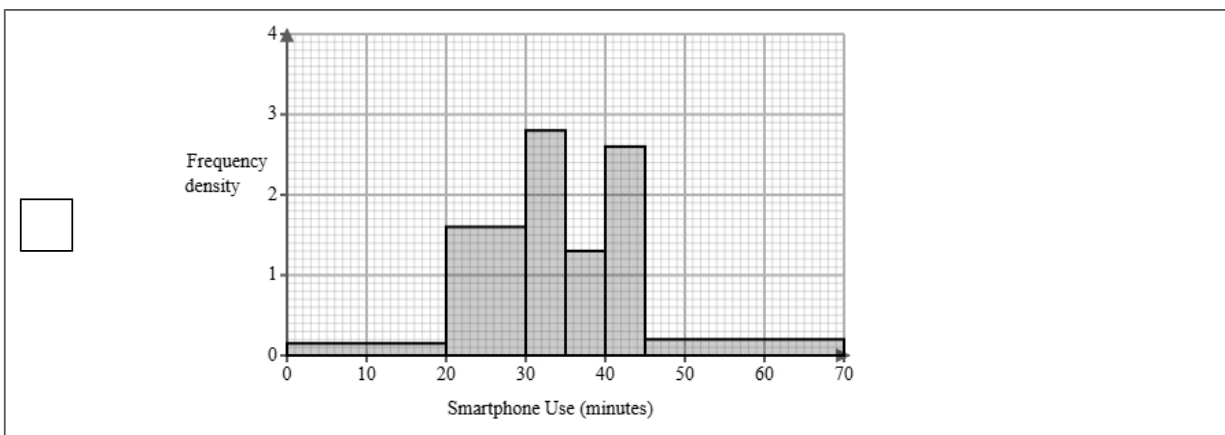
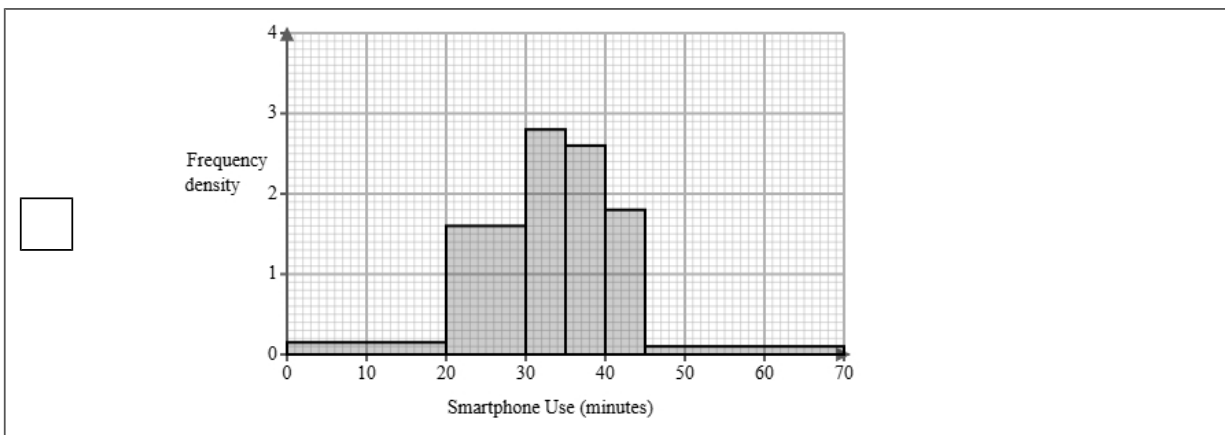
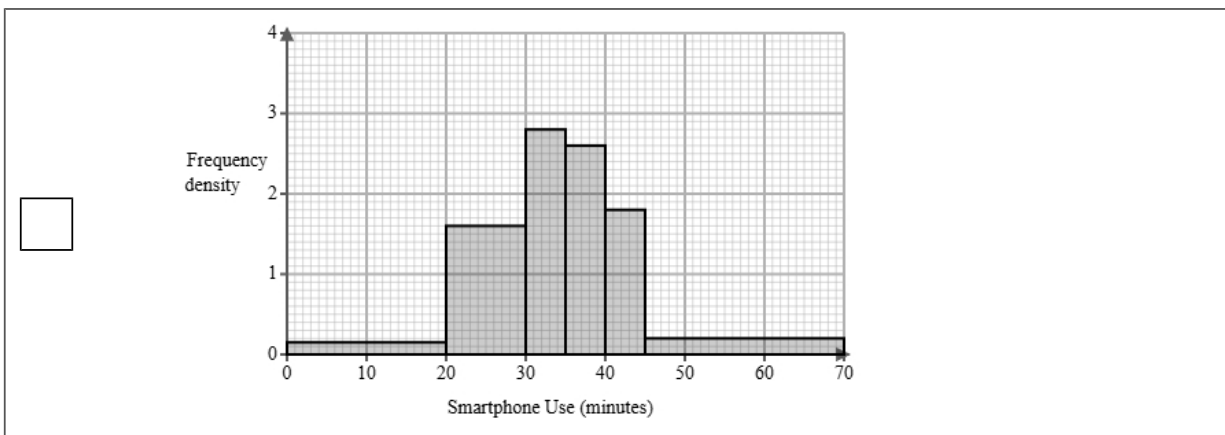
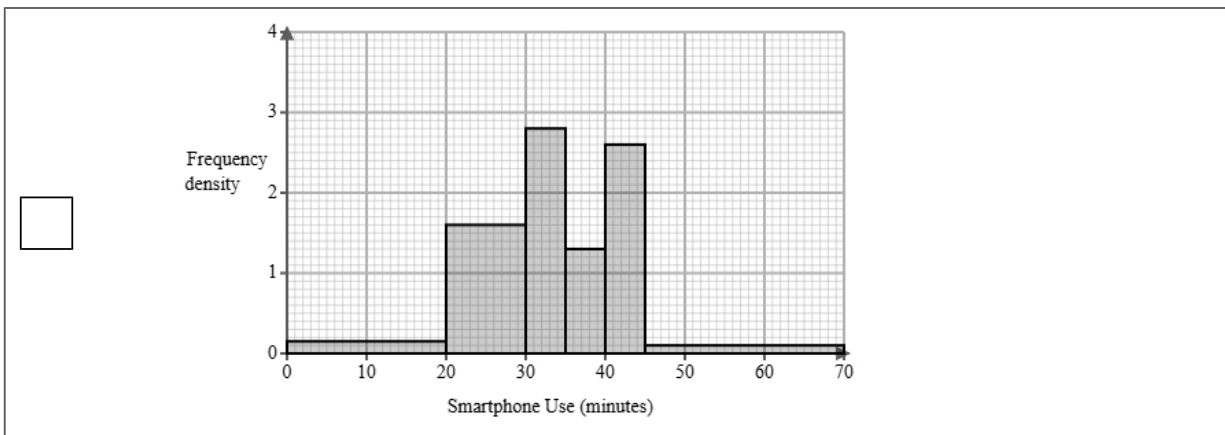
$$\text{frequency} = \text{frequency density} \times \text{class width}$$

Smartphone Use $u$ (minutes)	Frequency
$0 < u \leq 20$	3
$20 < u \leq 30$	16
$30 < u \leq 35$	_____
$35 < u \leq 40$	13
$40 < u \leq 45$	_____
$45 < u \leq 70$	5

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

(2 marks)

Select the correct answer.



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

$$\sum u = 2072$$

$$\sum u^2 = 76144$$

$$n = 60$$

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

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  $0 < u \leq 20$  and the upper outlier limit is within the group  $45 < u \leq 70$ .

- (d) A different behavioural researcher is analysing the smartphone usage of adult males over a 1-hour lunch break.

They find the following summary statistics for the data.

mean = 28.54

median = 34

standard deviation = 11.1

Calculate and interpret the skew for the males.

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 10th percentile.

(1 mark)

Select **one** box.

- $35 < u \leq 40$
- $30 < u \leq 35$
- $0 < u \leq 20$
- $20 < u \leq 30$
- $40 < u \leq 45$

- 8 Maya 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>
Barnet	185	8	5	3	9
Basingstoke	175	3	4	-1	1
Carlisle	178	4	3	1	1
Darlington	182	6	7	-1	1
Exeter	183	7	8		
Fareham	170	1	2		
Gloucester	180	5	6		
Harrogate	173	2	1		

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

Suggest a diagram that Maya could draw.

(1 mark)

Select **one** box.

- Line graph
- Scatter diagram
- Venn 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).

- 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) Maya used Spearman's rank correlation coefficient to analyse the data.

Jack suggests that Maya 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).

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

- 9 Mei works for a marine conservation group. She has been tasked with investigating plastic waste collected by divers.

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

Waste (kg)	Percentage of dive logs
$0 < p \leq 2$	7
$2 < p \leq 4$	seven
$4 < p \leq 6$	6
$6 < p \leq 8$	118
$8 < p \leq 10$	47
$10 < p \leq 12$	15
Total	100

Mei cleans the data to create the table below.

Waste (kg)	Percentage of dive logs
$0 < p \leq 2$	7
$2 < p \leq 4$	7
$4 < p \leq 6$	6
$6 < p \leq 8$	18
$8 < p \leq 10$	47
$10 < p \leq 12$	15
Total	100

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

(1 mark)

Select **one** box.

- The data should be spread evenly between each group.
- It should show frequencies not percentages.
- One of the percentages says seven not 7.

(b) Use linear interpolation to work out an estimate of the median waste.  
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

\_\_\_\_\_ kg

10 Isla collected the satisfaction scores for online and in-store customers.

Both types of shopping experiences were rated by the same number of customers.

The box plot presents data on the satisfaction scores for the online customers.



The table gives information about the satisfaction scores for the in-store customers.

Least tall	Lower quartile	Median	Upper quartile	Most tall
40	60	80	85	90

Compare the two distributions of satisfaction scores.

Give three comparisons and interpret one of these comparisons.

(4 marks)

Select **one** box.

- The median is bigger.
- The median satisfaction scores for online customers is greater than in-store customers.
- The median satisfaction scores for online and in-store customers are equal.
- The median satisfaction scores for online customers is lower than in-store customers.

Select **one** box.

- The IQR is bigger.
- The IQR for the satisfaction scores of the online customers is greater than in-store customers.
- The IQR for the satisfaction scores of the online and in-store customers are equal.
- The IQR for the satisfaction scores of the online customers is lower than in-store customers.

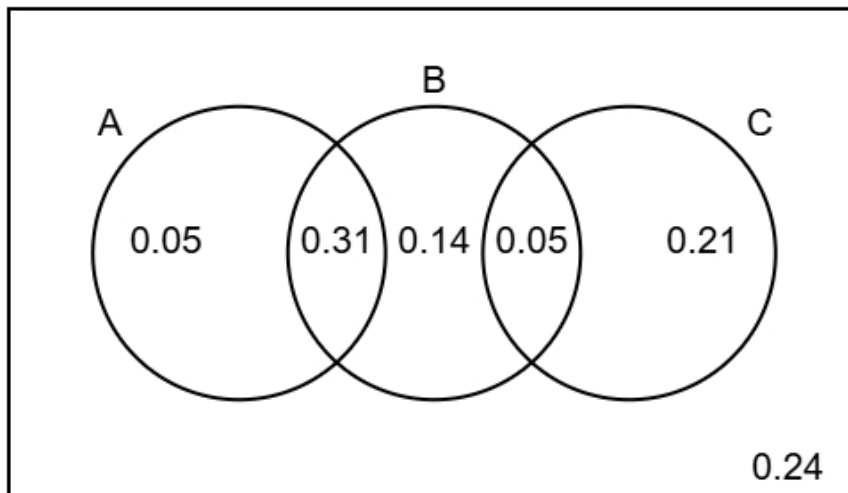
Select **one** box.

- The skews for the satisfaction scores of the online and in-store customers are both positive.
- The skews for the satisfaction scores of the online and in-store customers are both symmetrical.
- The skew for the satisfaction scores of the online customers is symmetrical and the skew for the in-store customers is negative.
- The skew for the satisfaction scores of the online customers is symmetrical and the skew for the in-store customers is positive.

Select **one** box.

- The scores for the online customers are less spread out than the in-store customers.
- The online customers on average have higher satisfaction scores than the in-store customers.
- The online customers are more skewed than in-store customers.
- The online customers on average have lower satisfaction scores than the in-store customers.

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
- B and C are mutually exclusive
- A and B are mutually exclusive
- because they only intersect once.
- because they do not intersect.

(b) Find  $P(B)$

(1 mark)

We are looking for the probabilities inside B.

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

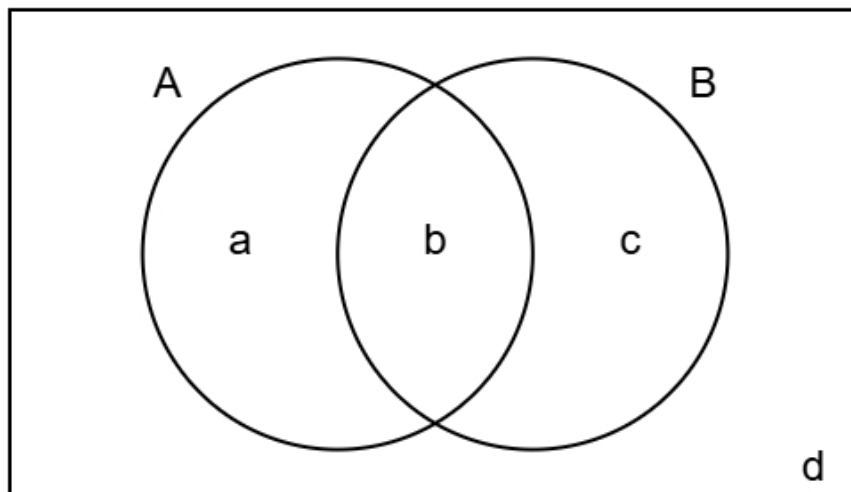
(2 marks)

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 Australia between 2013 and 2019

18 20 22 21 23 25 27

The table gives a summary of the amount, in millions, of tourists who visited Japan between 2013 and 2019

Mean	Standard Deviation	Largest Amount
22	3	25

Compare the amount of tourists in Australia and Japan between 2013 and 2019

You may use:

$$18^2 + 20^2 + 22^2 + 21^2 + 23^2 + 25^2 + 27^2 = 3532$$

(5 marks)

Find the mean for Australia

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

Find the standard deviation for Australia

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

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

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

**13** The running speeds of a group of athletes have a mean of 14.5 km/h and a standard deviation of 1.8 km/h.

(a) Olivia is athlete with a standardised score of 0.

Find Olivia's running speed.

(1 mark)

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

\_\_\_\_\_ km/h

(b) Aisha and Mia are both athletes in the group.

Aisha's standardised score for running speed is 1.3 km/h.

Mia's standardised score for running speed is -0.7 km/h.

Aisha was faster than Mia.

How much faster is Aisha?

(3 marks)

Rearrange the formula to make the value the subject

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

\_\_\_\_\_ km/h

(c) Emily takes a sample of 5 athletes from the group.

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

(i) Discuss the appropriateness of using 14.5 km/h as the mean in the calculation of the standardised score,

(ii) Discuss the appropriateness of using 1.8 km/h 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).

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

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

Using 14.5 km/h as the mean is appropriate

Using 14.5 km/h as the mean is not appropriate

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

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

Using 1.8 km/h as the standard deviation is appropriate

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

Using 1.8 km/h as the standard deviation is not appropriate