S1 Representation of Data: exam questions

10 questions · 33 marks · full mark schemes

Name _____ Date _____ Score / 33

- Histograms (frequency density), stem-and-leaf, box plots, cumulative frequency, outliers, skewness; comparing distributions.
- Try every question before you look at the mark schemes. They start after the questions: M marks for method, A marks for accuracy, B marks for a correct result on its own.
- Online, with more practice: alevelmathrevision.com/alevel-frontend/units/s1/representation-questions

1. **2 marks · warm-up**

Explain the key difference between a bar chart and a histogram in terms of what they represent.

2. **3 marks · warm-up**

For a data set, the lower quartile is 15 and the upper quartile is 33. An outlier is defined as any value outside the limits $Q_1 - 1.5 \times \text{IQR}$ and $Q_3 + 1.5 \times \text{IQR}$. Find the lower and upper limits for outliers.

3. **2 marks · standard**

Give one reason why it might be appropriate to remove an outlier from a data set (cleaning the data), and one reason why it might be inappropriate.

4. 3 marks · standard

On a box plot, an outlier is marked with a cross. The lower whisker extends to 15, the box spans from 25 to 40, and the upper whisker extends to 55. An outlier is marked at 65. Explain how the whisker ends are typically determined when outliers are present.

5. 3 marks · standard

If the data set representing salaries has a median of £45,000 and an IQR of £15,000, and a new policy increases everyone's salary by 10%, what will be the new median and the new IQR to display on an updated box plot?

6. 4 marks · standard

The following values are recorded: $4, 18, 22, 23, 24, 26, 28, 50$. Using the rule $Q_1 - 1.5 \times \text{IQR}$ and $Q_3 + 1.5 \times \text{IQR}$, identify any outliers in the data set.

7. 4 marks · harder

A histogram is used to represent a data set. The class $40 \leq t < 50$ has a frequency of 30. Assuming the data is uniformly distributed within the class, estimate the number of values that lie between 42 and 47.

8. 4 marks · harder

Two box plots, A and B, represent the test scores of two different classes. Box plot A has a median of 60 and an IQR of 15. Box plot B has a median of 70 and an IQR of 10. Compare the two distributions in context.

9. 4 marks · harder

A data set of $n = 50$ has $\sum x = 2000$ and $\sum x^2 = 83200$. An outlier is defined as being more than 2 standard deviations from the mean. Find the acceptable range of values for this data set.

10. 4 marks · harder

A frequency polygon is drawn on top of a histogram. The points of the polygon are located at $(15, 4)$, $(25, 10)$, and $(40, 6)$. What were the class intervals and frequencies for these three groups?

Answers and worked solutions

Question 1

2 marks

1. B1 for stating that a bar chart represents discrete or categorical data, while a histogram represents continuous grouped data.
2. B1 for stating that in a bar chart, the height is proportional to frequency, whereas in a histogram, the area of the bar is proportional to frequency.

Question 2

3 marks

1. M1 for calculating the interquartile range (IQR) $= 33 - 15 = 18$.
2. M1 for calculating the outlier boundaries $15 - 1.5(18)$ and $33 + 1.5(18)$.
3. A1 for stating the limits as -12 and 60.

Question 3

2 marks

1. B1 for stating it is appropriate if the outlier is clearly an error in recording or measuring.
2. B1 for stating it is inappropriate if the outlier is a genuine, legitimate observation that reflects natural variation.

Question 4

3 marks

1. M1 for stating that outliers are identified using a specific rule (e.g., $Q_3 + 1.5 \times \text{IQR}$).
2. A1 for explaining that the upper whisker is drawn to the highest data value that is *not* an outlier (in this case, 55).
3. A1 for explaining that the lower whisker is drawn to the lowest data value that is *not* an outlier (in this case, 15).

Question 5

3 marks

1. M1 for recognising that multiplying data by a scale factor scales both the median and the IQR by the same factor.
2. A1 for calculating the new median: $45000 \times 1.10 = £49,500$.
3. A1 for calculating the new IQR: $15000 \times 1.10 = £16,500$.

Question 6

4 marks

1. M1 for finding the quartiles: $n = 8$. $Q_1 = \frac{18+22}{2} = 20$, $Q_3 = \frac{26+28}{2} = 27$.
2. M1 for calculating $\text{IQR} = 27 - 20 = 7$.
3. M1 for calculating the boundaries: Lower $= 20 - 1.5(7) = 9.5$, Upper $= 27 + 1.5(7) = 37.5$.
4. A1 for identifying the outliers: 4 (since $4 < 9.5$) and 50 (since $50 > 37.5$).

Question 7

4 marks

1. M1 for identifying the class width as 10.
2. M1 for calculating the frequency density: $\frac{30}{10} = 3$.
3. M1 for finding the width of the required interval: $47 - 42 = 5$.
4. A1 for calculating the estimated frequency: $5 \times 3 = 15$ values.

Question 8

4 marks

1. M1 for comparing the medians: Class B has a higher median score ($70 > 60$).
2. A1 for interpreting the median: On average, students in Class B scored higher than students in Class A.
3. M1 for comparing the IQRs: Class A has a larger IQR ($15 > 10$).
4. A1 for interpreting the IQR: The middle 50% of scores in Class A are more spread out (more variable) than those in Class B.

Question 9

4 marks

1. M1 for calculating the mean: $\bar{x} = \frac{2000}{50} = 40$.
2. M1 for calculating the variance: $\frac{83200}{50} - 40^2 = 1664 - 1600 = 64$.
3. M1 for finding the standard deviation: $\sqrt{64} = 8$.
4. A1 for calculating the acceptable range: $40 - 2(8)$ to $40 + 2(8)$, which is 24 to 56.

Question 10

4 marks

1. M1 for recognising that the x-coordinates of the points represent the class midpoints, and the y-coordinates represent the frequency densities.
2. M1 for deducing the class boundaries from the midpoints: assuming contiguous classes, the boundaries are $10 - 20$, $20 - 30$, and $30 - 50$.
3. M1 for identifying the class widths: 10, 10, and 20 respectively.
4. A1 for calculating frequencies by multiplying width by height (frequency density): Frequencies are $10 \times 4 = 40$, $10 \times 10 = 100$, and $20 \times 6 = 120$.

More representation of data practice with instant feedback: alevelmathrevision.com/alevel-frontend/units/s1/representation-questions

Our own original questions, written and checked by A Level Math Revision. Independent, not affiliated with Pearson Edexcel.