EDEXCEL INTERNATIONAL A LEVEL · S1 STATISTICS 1

S1 Data and Summary Statistics: exam questions

10 questions · 53 marks · full mark schemes

Name _____ Date _____ Score / 53

- Types of data (qualitative/quantitative, discrete/continuous); mean, median, mode, quartiles and percentiles; variance and standard deviation; coding; outliers. (Sampling methods are not in IAL S1.)
- 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/data-questions

1. **4 marks · warm-up**

Class A consists of 15 students who scored a mean of 70 on a test. Class B consists of 25 students who scored a mean of 78 on the same test. Calculate the exact overall mean score for all 40 students combined.

2. **5 marks · warm-up**

A dataset contains 5 numbers: $x, y, 4, 6, 8$. The mean of the dataset is 5 and the variance is 4. Given that $x < y$, find the exact values of x and y.

3. **5 marks · standard**

A dataset consists of n completely identical values, x.

1. Write down the interquartile range (IQR) of this dataset.
2. A new, single data value y is added to the dataset, where $y \neq x$. Assuming n is very large, prove algebraically why the value y will always be classified as an outlier, regardless of how close y is to x.

4. **5 marks · standard**

The annual salaries, in thousands of pounds, for six employees at a small startup are:

$$20,\ 22,\ 25,\ 28,\ 30,\ 120$$

1. Calculate the mean salary.
2. Find the median salary.
3. State which average (mean or median) is the best measure of central tendency for this data set, and give a reason for your choice.

5. 5 marks · standard

A dataset representing the time taken for 40 students to complete a puzzle has a lower quartile of $12\ \textbf{minutes}$ and an upper quartile of $28\ \textbf{minutes}$.

1. Calculate the Interquartile Range (IQR).
2. Calculate the lower and upper boundaries for outliers.
3. One student completed the puzzle in $50\ \textbf{minutes}$. State, with a reason, whether this time is considered an outlier.

6. 5 marks · standard

The weights of packages shipped by a postal service are summarised in a box and whisker plot. The lower quartile is $Q_1 = 15\ \text{kg}$ and the upper quartile is $Q_3 = 45\ \text{kg}$.

1. Calculate the interquartile range (IQR).
2. Determine the upper boundary for outliers in this dataset.
3. A heavy package weighs $95\ \text{kg}$. Determine mathematically whether this package is considered an outlier.

7. 4 marks · harder

The distances travelled by 60 commuters to work are summarised in the frequency table below.

Distance (d km)	Frequency	Cumulative Frequency
$0 \leq d < 10$	4	
$10 \leq d < 20$	11	
$20 \leq d < 30$	25	
$30 \leq d < 40$	16	
$40 \leq d < 50$	4	

1. Copy and complete the cumulative frequency column in the table.
2. State the class interval that contains the median distance.
3. Find an estimate for the mean distance travelled by the commuters.

8. 5 marks · harder

A dataset has a mean of 15 and a standard deviation of 3. A new dataset is formed by taking every value in the original dataset, multiplying it by 2, and then adding 4.

1. Find the mean of the new dataset.
2. Find the standard deviation of the new dataset.
3. Find the variance of the new dataset.
4. A single new data value, which is exactly equal to the original mean (15), is added to the original dataset. State whether the standard deviation of the dataset will increase, decrease, or remain the same. Give a reason for your answer.

9. 6 marks · harder

A dataset consists of five discrete integers: -2, 1, a, b, and 7. The mode of the dataset is 1 and the mean is 3.

1. Deduce the exact values of a and b, assuming $a \leq b$.
2. Calculate the exact variance of the dataset.

10. 9 marks · harder

A dataset of 15 values has a mean of 45 and a standard deviation of 12. During a review of the data, it is discovered that one of the values, 80, was recorded in error and must be removed from the dataset.

1. Find the new mean of the remaining 14 values.
2. Calculate $\sum x^2$ for the original dataset of 15 values.
3. Hence, find the standard deviation of the new dataset with the anomaly removed.

Answers and worked solutions

Question 1

4 marks

Answer: 75

M1 for finding the total sum of scores for Class A: $15 \times 70 = 1050$.

M1 for finding the total sum of scores for Class B: $25 \times 78 = 1950$.

M1 for setting up the combined mean calculation: $\frac{1050+1950}{15+25}$.

A1 for evaluating to the final exact answer: $\frac{3000}{40} = 75$.

Question 2

5 marks

Answer: $x = 2, y = 5$

M1 for using the mean to form an equation: $\frac{x+y+18}{5} = 5 \implies x + y = 7$.

M1 for using the variance formula $\frac{\sum x^2}{n} - \mu^2 = \mathbf{Var}$: $\frac{x^2+y^2+16+36+64}{5} - 25 = 4$.

M1 for simplifying the variance equation: $\frac{x^2+y^2+116}{5} = 29 \implies x^2 + y^2 = 145 - 116 = 29$.

M1 for solving the simultaneous equations by substitution (e.g. $y = 7 - x \implies x^2 + (7-x)^2 = 29$) or using $(x+y)^2 = x^2 + y^2 + 2xy \implies 49 = 29 + 2xy \implies xy = 10$.

A1 for finding the two numbers whose sum is 7 and product is 10: 2 and 5. Since $x < y$, $x = 2$ and $y = 5$.

Question 3

5 marks

Answer: IQR $= 0$; both outlier limits equal x, so any $y \neq x$ is an outlier

1. A1 for stating the IQR is 0 (since $Q3 = x$ and $Q1 = x$, $x - x = 0$).
2. M1 for determining the standard outlier boundaries: $Q1 - 1.5 \times \mathrm{IQR}$ and $Q3 + 1.5 \times \mathrm{IQR}$.
   M1 for substituting $Q1 = x, Q3 = x$, and $\mathrm{IQR} = 0$ to get the interval $[x, x]$.
   B1 for stating that any value must strictly fall within the interval $[x, x]$ to not be an outlier, meaning it must be exactly x.
   B1 for concluding that since $y \neq x$, y falls outside the boundaries and is therefore mathematically classified as an outlier, regardless of its proximity to x. *

Question 4

5 marks

Answer: Mean $= 40.8$ (thousand pounds); median $= 26.5$; the median, since the value 120 is an outlier that distorts the mean

1. Mean $= \frac{20 + 22 + 25 + 28 + 30 + 120}{6} = \frac{245}{6}$ **M1**
   $= 40.8$ (3 s.f.) **A1**
2. Median $= \frac{25 + 28}{2} = 26.5$ **M1 A1**
3. Median: it is not affected by the extreme value 120, whereas the mean is pulled upwards by it **B1**

Question 5

5 marks

Answer: $50 < 52$

1. $IQR = 28 - 12 = 16$ A1
2. Lower boundary $= Q_1 - 1.5 \times IQR = 12 - 1.5(16) = 12 - 24 = -12$ M1
   Upper boundary $= Q_3 + 1.5 \times IQR = 28 + 24 = 52$ A1
3. Since $50 < 52$, it does not exceed the upper boundary. M1
   It is not an outlier. A1

Question 6

5 marks

Answer: IQR $= 30$ kg; upper boundary 90 kg; $95 > 90$, so it is an outlier

1. $IQR = 45 - 15 = 30$. A1

2. Upper boundary
   $= Q_3 + 1.5 \times IQR = 45 + 1.5(30)$ M1
   $= 45 + 45 = 90\ \text{kg}$. A1
3. Since $95 > 90$, the package weight exceeds the upper boundary. M1
   Yes, it is considered an outlier. A1

Question 7

4 marks

Answer: Cumulative frequencies $4, 15, 40, 56, 60$; median class $20 \le d < 30$; mean ≈ 25.8 km

1. Cumulative frequencies: $4, 15, 40, 56, 60$ **B1**
2. Median is the 30th value, which lies in $20 \le d < 30$ **B1**
3. $\sum fx = 4(5) + 11(15) + 25(25) + 16(35) + 4(45) = 1550$ **M1**
4. Mean $= \dfrac{1550}{60} = 25.8$ km (3 s.f.) **A1**

Question 8

5 marks

Answer: Mean $= 34$; s.d. $= 6$; variance $= 36$; the standard deviation decreases

1. New mean $= 2(15) + 4 = 34$ **B1**
2. New s.d. $= 2 \times 3 = 6$ (adding 4 does not change spread) **B1**
3. New variance $= 6^2 = 36$ **B1**
4. Adding a value equal to the mean leaves $\sum(x - \bar{x})^2$ unchanged but increases n **M1**, so the standard deviation decreases **A1**

Question 9

6 marks

Answer: $\frac{74}{5} = 14.8$

1. M1 for deducing that since the mode is 1, at least one of the unknown variables must be 1. Let $a = 1$.
   M1 for using the mean formula: $\frac{-2+1+1+b+7}{5} = 3$.
   A1 for solving for b:
   $\frac{7+b}{5} = 3 \implies 7 + b = 15 \implies b = 8$. Thus $a = 1, b = 8$.
2. M1 for applying the variance formula $\sigma^2 = \frac{\sum(x-\mu)^2}{n}$.
   M1 for calculating the squared deviations:
   $(-2-3)^2 + (1-3)^2 + (1-3)^2 + (8-3)^2 + (7-3)^2 = 25 + 4 + 4 + 25 + 16 = 74$.
   A1 for calculating the exact variance: $\frac{74}{5} = 14.8$.

Question 10

9 marks

Answer: New mean 42.5; $\sum x^2 = 32535$; new s.d. 7.78

1. Original $n = 15$, $\mu = 45$. Original sum $\sum x = 15 \times 45 = 675$. **M1**
   Remove anomaly (80): New sum $= 675 - 80 = 595$. **M1**
   New mean $= \frac{595}{14} = 42.5$. **A1**
2. Using the variance formula:
   $\text{Var}(X) = \sigma^2 = 12^2 = 144$. **M1**
   $144 = \frac{\sum x^2}{15} - 45^2 \implies 144 = \frac{\sum x^2}{15} - 2025$. **M1**
   $\frac{\sum x^2}{15} = 2169 \implies \sum x^2 = 32535$. **A1**
3. Remove anomaly from sum of squares: New $\sum x^2 = 32535 - 80^2 = 32535 - 6400 = 26135$. **A1**
   New Variance $= \frac{26135}{14} - (42.5)^2 = 1866.785\ldots - 1806.25 = 60.535\ldots$ **M1**
   New SD $= \sqrt{60.535\ldots} = 7.7804\ldots \implies 7.78\ (3sf)$. **A1**

More data and summary statistics practice with instant feedback: alevelmathrevision.com/alevel-frontend/units/s1/data-questions
Our own original questions, written and checked by A Level Math Revision. Independent, not affiliated with Pearson Edexcel.