2.6 Statistics: Data Presentation and Interpretation Depth: 1-hour Pack

EDEXCEL 9MA0 · Calculator allowed · about 60 minutes

Total Marks

Name: _____ Date: ____ / ____ / _____

Answer ALL questions. Show all your working.

A-Level Statistics: Data Presentation and Interpretation - Depth Practice

Original material written for Revision Library, styled on Edexcel AS/A-level Mathematics (9MA0) Statistics content.

This pack extends core data presentation and interpretation skills to full A-level depth: measures of location and spread (including coding), identifying outliers by the quartile rule and the standard deviation rule, box plots and skewness, histograms and frequency density, linear interpolation, correlation and regression (including coding and residuals), and evaluating statistical models in context, in the style of the large data set used across the qualification. All contexts and figures are original.

1 The table summarises the times, in minutes, taken by 57 people to complete a puzzle.
Time: 30 - 40, 40 - 50, 50 - 60, 60 - 70, 70 - 80, 80 - 90
Frequency: 9, 14, 14, 2, 7, 11

(a) Estimate the mean time. **(3)**

(b) Estimate the standard deviation of the times, to 3 significant figures. **(3)**

(c) Use linear interpolation to estimate the median time, to 3 significant figures. **(3)**

(Total for Question 1 is 9 marks)

---

2 The masses, in kg, of a sample of lambs have lower quartile 34, median 39, upper quartile 42, smallest value 31 and largest value 78.

(a) Find the interquartile range. **(2)**

(b) A value is an outlier if it is more than 1.5 x IQR above the upper quartile or below the lower quartile. Determine whether the smallest value and the largest value are outliers. **(3)**

(c) Describe the skewness of the distribution, justifying your answer using the quartiles. **(2)**

(Total for Question 2 is 7 marks)

© Revision Library | revisionlibrary.co.uk

3 A scientist records data on the age of a car, x years, and its value, V thousand pounds, for several days. The regression line of V on x is V = 39 - 1.5x. The product moment correlation coefficient is r = -0.92.

(a) Interpret the value of the gradient of the regression line in context. **(1)**

(b) Use the regression line to estimate V when x = 15, and comment on the reliability of this estimate. **(3)**

(c) Comment on the strength and direction of the linear correlation indicated by r = -0.92. **(2)**

(Total for Question 3 is 6 marks)

---

4 The masses, in kg, of a sample of lambs have lower quartile 31, median 35, upper quartile 37, smallest value 26 and largest value 55.

(a) Find the interquartile range. **(2)**

(b) A value is an outlier if it is more than 1.5 x IQR above the upper quartile or below the lower quartile. Determine whether the smallest value and the largest value are outliers. **(3)**

(c) Describe the skewness of the distribution, justifying your answer using the quartiles. **(2)**

(Total for Question 4 is 7 marks)

---

5 For a sample of 12 UK secondary schools, x represents the average class size and y represents the mean GCSE point score per pupil. The following summary statistics were calculated:

n = 12, Sxx = 248, Syy = 196, Sxy = -152

(a) Calculate the product moment correlation coefficient, r, between x and y. **(3)**

(b) Interpret the value of r found in part (a) in context. **(2)**

(c) The headteacher of one school claims that 'smaller class sizes cause higher GCSE point scores'. Comment on the validity of this claim, referring to the value of r found in part (a). **(2)**

(Total for Question 5 is 7 marks)

© Revision Library | revisionlibrary.co.uk

6 The table below shows the distribution of the ages, in complete years, of 200 members of a gym. Ages are recorded to the nearest whole year, so the class 16-25 has boundaries 15.5 and 25.5, and so on.

Age (years): 16-25, frequency 30
Age (years): 26-35, frequency 54
Age (years): 36-45, frequency f (unknown)
Age (years): 46-55, frequency 38
Age (years): 56-70, frequency 20

(a) Show that the frequency, f, for the class 36-45 is 58. **(1)**

(b) Calculate the frequency density for each class (using class widths based on the true class boundaries), and state, with a reason, which class would be represented by the tallest bar if a histogram were drawn for this data. **(2)**

(c) Use linear interpolation to estimate the median age of the gym members. **(3)**

(d) Estimate the number of gym members aged under 30. **(3)**

(Total for Question 6 is 9 marks)

---

7 The daily maximum temperatures, in degrees Celsius, were recorded for a sample of 31 days in May at two UK weather stations, Aviemore and Camborne, in the style of the large data set. Summary statistics are given below.

Aviemore (n = 31): minimum = 6.1, Q1 = 9.4, median = 11.8, Q3 = 14.2, maximum = 19.5
Camborne (n = 31): minimum = 10.8, Q1 = 12.6, median = 13.9, Q3 = 15.7, maximum = 18.0

(a) Calculate the interquartile range for each location. **(2)**

(b) Using the rule that an outlier is a value more than 1.5 x IQR beyond the nearer quartile, calculate the upper outlier boundary for Aviemore, and state whether the maximum value of 19.5 degrees Celsius is an outlier. **(3)**

(c) Compare and contrast the distributions of maximum daily temperature at the two locations, referring to the median and the interquartile range. **(3)**

(d) Explain one reason, in context, why Aviemore might show a greater spread of daily maximum temperatures than Camborne. **(1)**

(Total for Question 7 is 9 marks)

© Revision Library | revisionlibrary.co.uk

Mark scheme · 2.6 Statistics: Data Presentation and Interpretation Depth: 1-hour Pack

Question 1

(a) uses the midpoint of each class **M1**

(a) sum of f x midpoint = 3305 **M1**

(a) mean = 3305/57 = 57.98 minutes **A1**

(a) Answer: 57.98 minutes

(b) sum f x midpoint^2 = 209025 **M1**

(b) variance = 209025/57 - (57.9825)^2 **M1**

(b) standard deviation = 17.5 minutes **A1**

(b) Answer: 17.5 minutes

(c) identifies the median class 50 - 60 (cumulative frequency reaches 37 at its upper bound) **M1**

(c) interpolates: 50 + ((28.5 - 23)/14) x 10 **M1**

(c) median = 53.9 minutes **A1**

(c) Answer: 53.9 minutes

Question 2

(a) 42 - 34 **M1**

(a) 8 **A1**

(a) Answer: 8

(b) lower limit 34 - 1.5 x 8 = 22 and upper limit 42 + 1.5 x 8 = 54 **M1**

(b) smallest value 31 is not below 22, so is not an outlier **A1**

(b) largest value 78 is above 54, so is an outlier **A1**

(b) Answer: smallest is not an outlier; largest is an outlier

(c) compares median - Q1 = 5 and Q3 - median = 3 **M1**

(c) negative skew (as the larger spread is on that side) **A1**

(c) Answer: negative skew

Question 3

(a) each extra year in x is associated with a change of -1.5 in V, so as x increases V decreases **B1**

(a) Answer: gradient -1.5: V decreases by 1.5 for each unit increase in x

(b) substitutes x = 15 **M1**

(b) V = 16.5 **A1**

(b) reliable only if 15 is within the range of the data (interpolation); extrapolation is unreliable **B1**

(b) Answer: 16.5

(c) strong **B1**

(c) negative correlation **B1**

(c) Answer: strong negative linear correlation

Question 4

(a) 37 - 31 **M1**

(a) 6 **A1**

(a) Answer: 6

(b) lower limit 31 - 1.5 x 6 = 22 and upper limit 37 + 1.5 x 6 = 46 **M1**

(b) smallest value 26 is not below 22, so is not an outlier **A1**

(b) largest value 55 is above 46, so is an outlier **A1**

(b) Answer: smallest is not an outlier; largest is an outlier

(c) compares median - Q1 = 4 and Q3 - median = 2 **M1**

(c) negative skew (as the larger spread is on that side) **A1**

(c) Answer: negative skew

Question 5

(a) correct formula used: r = Sxy / $\sqrt{Sxx \times Syy}$ **M1**

(a) $\sqrt{Sxx \times Syy}$ = $\sqrt{48608}$ = awrt 220 (or 220.5) found **M1**

(a) r = awrt -0.689 cao **A1**

(a) Answer: r = -0.689 (3sf)

(b) recognises a fairly strong negative linear correlation (r close to -1, not exactly -1) **B1**

(b) correct context: as average class size increases, mean GCSE point score per pupil tends to decrease (or vice versa), across these 12 schools **B1**

(b) Answer: There is a fairly strong negative linear correlation: as average class size increases, mean GCSE point score tends to decrease

(c) correlation does not imply causation; r only measures the strength of a linear association between the two variables, not a cause and effect relationship **B1**

(c) valid contextual reasoning, e.g. other factors such as school funding, socio-economic background or prior attainment could explain the association, so the claim is not fully justified by r alone **B1**

(c) Answer: The claim is not fully justified; r shows association, not causation, and other factors could explain the relationship

© Revision Library | revisionlibrary.co.uk

Question 6

(a) cso: 200 - 30 - 54 - 38 - 20 = 58 **B1**

(a) Answer: f = 58

(b) attempt at frequency density = frequency / class width for at least three classes, using correct widths (10, 10, 10, 10, 15) **M1**

(b) all five correct: 3.0, 5.4, 5.8, 3.8, 1.33 (3sf) and correctly identifies 36-45 as the tallest bar (highest frequency density, 5.8) **A1**

(b) Answer: Frequency densities: 3.0, 5.4, 5.8, 3.8, 1.33; the tallest bar is for the 36-45 class

(c) identifies the median class as 35.5-45.5 using cumulative frequencies (30, 84, 142, ...) and n/2 = 100 **M1**

(c) correct interpolation set up: median = 35.5 + [(100-84)/58] x 10 **M1**

(c) median = awrt 38.3 years **A1**

(c) Answer: Median = 38.3 years (3sf)

(d) interpolates within the 25.5-35.5 class: 54 x (4.5/10) attempted **M1**

(d) adds the full first class: 30 + 54 x (4.5/10) **M1**

(d) total = awrt 54 members (54.3) **A1**

(d) Answer: Approximately 54 members (54.3) are aged under 30

Question 7

(a) IQR(Aviemore) = 4.8 **B1**

(a) IQR(Camborne) = 3.1 **B1**

(a) Answer: IQR(Aviemore) = 4.8 degrees Celsius, IQR(Camborne) = 3.1 degrees Celsius

(b) upper boundary = Q3 + 1.5 x IQR = 14.2 + 1.5(4.8) attempted **M1**

(b) upper boundary = 21.4 **A1**

(b) correct conclusion with comparison: since 19.5 < 21.4, the maximum value is not an outlier **A1**

(b) Answer: Upper boundary = 21.4 degrees Celsius; 19.5 is not an outlier

(c) median comparison: Camborne has the higher median (13.9 vs 11.8), so typical maximum temperatures are higher at Camborne **B1**

(c) spread comparison: Aviemore has the larger IQR (4.8 vs 3.1), so maximum temperatures are more variable (less consistent) at Aviemore **B1**

(c) valid contextual link, e.g. Camborne's coastal position gives a milder, more consistent climate, while Aviemore's inland/upland position leads to more variable daily maximum temperatures **B1**

(c) Answer: Camborne is typically warmer (higher median) and more consistent (smaller IQR); Aviemore is cooler on average and more variable

(d) valid contextual reason, e.g. Aviemore is an inland/upland location in the Scottish Highlands, so lacks the moderating maritime influence that keeps Camborne's coastal temperatures more stable, oe **B1**

(d) Answer: Aviemore's inland/upland location lacks the moderating coastal influence that keeps Camborne's temperatures more consistent

© Revision Library | revisionlibrary.co.uk