2.6 Statistics: Data Presentation and Interpretation Depth: 30-minute Pack

EDEXCEL 9MA0 · Calculator allowed · about 30 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 masses, in kg, of a sample of lambs have lower quartile 35, median 43, upper quartile 53, smallest value 25 and largest value 92.

(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 1 is 7 marks)

© Revision Library | revisionlibrary.co.uk

2 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 = 6 - 1.5x. The product moment correlation coefficient is r = -0.85.

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

(b) Use the regression line to estimate V when x = 16, and comment on the reliability of this estimate. **(3)**

(c) Comment on the strength and direction of the linear correlation indicated by r = -0.85. **(2)**

(Total for Question 2 is 6 marks)

© Revision Library | revisionlibrary.co.uk

3 The masses, in kg, of a sample of parcels processed at a sorting depot are summarised in the grouped frequency table below. One frequency is missing.

Mass, m (kg): $0 < m \leq 10$, frequency 8
Mass, m (kg): $10 < m \leq 20$, frequency 22
Mass, m (kg): $20 < m \leq 30$, frequency f (unknown)
Mass, m (kg): $30 < m \leq 50$, frequency 18
Mass, m (kg): $50 < m \leq 80$, frequency 12

It is known that the frequency density for the class $20 < m \leq 30$ is 2.6.

(a) Find the value of f. **(2)**

(b) Calculate the frequency density for each of the other four classes, and state, with a reason, which class would be represented by the tallest bar if a histogram were drawn for this data. **(3)**

(c) Estimate the number of parcels with mass greater than 30 kg. **(1)**

(Total for Question 3 is 6 marks)

---

4 The table shows the distribution of the number of hours, h, spent revising per week by 60 sixth-form students during study leave.

Hours, h: $0 \leq h < 5$, midpoint 2.5, frequency 6
Hours, h: $5 \leq h < 10$, midpoint 7.5, frequency 14
Hours, h: $10 \leq h < 15$, midpoint 12.5, frequency 20
Hours, h: $15 \leq h < 20$, midpoint 17.5, frequency 12
Hours, h: $20 \leq h < 25$, midpoint 22.5, frequency 8

A coding $y = (x - 12.5)/5$ is used, where x is the class midpoint.

(a) Complete the coded midpoint values, y, for each class. **(2)**

(b) Using the coded data, calculate an estimate for the mean number of hours spent revising per week. **(3)**

(c) Calculate an estimate for the standard deviation of the number of hours spent revising per week. **(4)**

(Total for Question 4 is 9 marks)

© Revision Library | revisionlibrary.co.uk

Mark scheme · 2.6 Statistics: Data Presentation and Interpretation Depth: 30-minute Pack

Question 1

(a) 53 - 35 **M1**

(a) 18 **A1**

(a) Answer: 18

(b) lower limit 35 - 1.5 x 18 = 8 and upper limit 53 + 1.5 x 18 = 80 **M1**

(b) smallest value 25 is not below 8, so is not an outlier **A1**

(b) largest value 92 is above 80, so is an outlier **A1**

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

(c) compares median - Q1 = 8 and Q3 - median = 10 **M1**

(c) positive skew **A1**

(c) Answer: positive skew

Question 2

(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 = 16 **M1**

(b) V = -18.0 **A1**

(b) reliable only if 16 is within the range of the data (interpolation); extrapolation is unreliable **B1**

(b) Answer: -18.0

(c) strong **B1**

(c) negative correlation **B1**

(c) Answer: strong negative linear correlation

Question 3

(a) f = 2.6 x 10 (class width 10) **M1**

(a) f = 26 **A1**

(a) Answer: f = 26

(b) attempt at frequency density = frequency/class width for at least two classes **M1**

(b) all four correct: 0.8, 2.2, 0.9, 0.4 **A1**

(b) correctly identifies $20 < m \leq 30$ as the tallest bar, since it has the highest frequency density (2.6, higher than all of 0.8, 2.2, 0.9, 0.4) **B1**

(b) Answer: Frequency densities: 0.8, 2.2, 0.9, 0.4; the tallest bar is for $20 < m \leq 30$ (frequency density 2.6)

(c) 18 + 12 = 30 parcels **B1**

(c) Answer: 30 parcels

Question 4

(a) three of the five coded values correct **B1**

(a) all five coded values correct: -2, -1, 0, 1, 2 **B1**

(a) Answer: y values: -2, -1, 0, 1, 2

(b) sum(fy) = 2 found (using -12-14+0+12+16) **M1**

(b) mean of y = 2/60 used, then mean of x = 12.5 + 5 x mean of y **M1**

(b) mean = awrt 12.7 hours (exact 38/3) **A1**

(b) Answer: Mean = 12.7 hours (3sf)

(c) sum(fy^2) = 82 found **M1**

(c) variance of y = sum(fy^2)/60 - (mean of y)^2 used **M1**

(c) variance of x = 25 x variance of y (or equivalent unscaling of the coding) **M1**

(c) standard deviation = awrt 5.84 hours **A1**

(c) Answer: Standard deviation = 5.84 hours (3sf)

© Revision Library | revisionlibrary.co.uk