2.2 Statistics: Data Presentation and Interpretation: 30-minute Pack

Total Marks

EDEXCEL 9MA0 · Calculator allowed · about 30 minutes

Name: _____ Date: ____ / ____ / _____

Answer ALL questions. Show all your working.

1 House prices, in thousands of pounds, were recorded for two streets. For 20 houses in Elm Street: sum(x) = 4200, sum(x^2) = 920000. For 15 houses in Birch Street: sum(x) = 3300, sum(x^2) = 746000.

(a) Calculate the mean and the standard deviation of house prices in Elm Street. **(4)**

(b) Calculate the mean and the standard deviation of house prices in Birch Street. **(4)**

(c) Using your answers to parts (a) and (b), compare house prices in Elm Street and Birch Street. **(2)**

(Total for Question 1 is 10 marks)

2 A histogram is drawn for the lengths of time, in minutes, that customers wait. The class widths and frequencies are:
Class: 10 - 15, 15 - 17, 17 - 22, 22 - 24
Frequency: 30, 7, 8, 24

(a) Find the frequency density for each class. **(2)**

(b) Estimate the number of customers who waited longer than 16 minutes but not more than 17 minutes. **(2)**

(c) The height of the tallest bar is 6 cm. Find the height of the bar for the class with the smallest frequency density, in cm to 2 significant figures. **(2)**

(Total for Question 2 is 6 marks)

© Revision Library | revisionlibrary.co.uk

3 The masses, in kg, of a sample of lambs have lower quartile 39, median 48, upper quartile 55, smallest value 30 and largest value 76.

(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 3 is 7 marks)

4 A scientist records data on height above sea level, h hundred metres, and the air temperature, T degrees Celsius, for several days. The regression line of T on h is T = 37 - 2.5h. The product moment correlation coefficient is r = -0.33.

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

(b) Use the regression line to estimate T when h = 27, and comment on the reliability of this estimate. **(3)**

(c) Comment on the strength and direction of the linear correlation indicated by r = -0.33. **(2)**

(Total for Question 4 is 6 marks)

© Revision Library | revisionlibrary.co.uk

Mark scheme · 2.2 Statistics: Data Presentation and Interpretation: 30-minute Pack

Question 1

(a) attempt mean = sum(x)/20 **M1**

(a) cao mean = 210 (GBP thousand) **A1**

(a) attempt $\sqrt{\text{sum}(x^2)/20 - \text{mean}^2}$ **M1**

(a) awrt 43.6 (GBP thousand) **A1**

(a) Answer: mean = £210,000, SD = awrt £43,600

(b) attempt mean = sum(x)/15 **M1**

(b) cao mean = 220 (GBP thousand) **A1**

(b) attempt $\sqrt{\text{sum}(x^2)/15 - \text{mean}^2}$ **M1**

(b) awrt 36.5 (GBP thousand) **A1**

(b) Answer: mean = £220,000, SD = awrt £36,500

(c) Birch Street has the higher mean price (£220,000 vs £210,000), so houses there tend to be more expensive, oe **B1**

(c) Elm Street has the higher standard deviation (£43,600 vs £36,500), so house prices there are more variable/spread out, oe **B1**

(c) Answer: Houses in Birch Street are, on average, more expensive (higher mean), but house prices in Elm Street are more variable (higher standard deviation).

Question 2

(a) frequency density = frequency / class width **M1**

(a) values 6, 7/2, 8/5, 12 **A1**

(a) Answer: 6, 7/2, 8/5, 12

(b) proportion of the class 15 - 17 lying in the required interval = (17 - 16)/2 **M1**

(b) 3.5 **A1**

(b) Answer: 3.5

(c) height proportional to frequency density: scale = 6/largest fd **M1**

(c) 0.80 cm **A1**

(c) Answer: 0.80 cm

Question 3

(a) 55 - 39 **M1**

(a) 16 **A1**

(a) Answer: 16

(b) lower limit 39 - 1.5 x 16 = 15 and upper limit 55 + 1.5 x 16 = 79 **M1**

(b) smallest value 30 is not below 15, so is not an outlier **A1**

(b) largest value 76 is not above 79, so is not an outlier **A1**

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

(c) compares median - Q1 = 9 and Q3 - median = 7 **M1**

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

(c) Answer: negative skew

Question 4

(a) each extra hundred metre in h is associated with a change of -2.5 in T, so as h increases T decreases **B1**

(a) Answer: gradient -2.5: T decreases by 2.5 for each unit increase in h

(b) substitutes h = 27 **M1**

(b) T = -30.5 **A1**

(b) The weak correlation makes the prediction unreliable. The observed range is not given, so it is not possible to determine whether this is interpolation. Accept equivalent comments. **B1**

(b) Answer: -30.5 degrees Celsius. The weak correlation makes the prediction unreliable. The observed range is not given, so it is not possible to determine whether this is interpolation.

(c) weak **B1**

(c) negative correlation **B1**

(c) Answer: weak negative linear correlation

© Revision Library | revisionlibrary.co.uk