2.2 Statistics: Data Presentation and Interpretation: 1-hour Pack

Total Marks

EDEXCEL 9MA0 · Calculator allowed · about 60 minutes

Name: _____ Date: ____ / ____ / _____

Answer ALL questions. Show all your working.

1 The table summarises the times, in minutes, taken by 42 people to complete a puzzle.
Time: 10 - 15, 15 - 20, 20 - 25, 25 - 30, 30 - 35
Frequency: 2, 15, 10, 12, 3

Time (minutes)	Frequency
10 - 15	2
15 - 20	15
20 - 25	10
25 - 30	12
30 - 35	3

(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)

© Revision Library | revisionlibrary.co.uk

2 The masses, in kg, of a sample of lambs have lower quartile 33, median 41, upper quartile 52, smallest value 28 and largest value 89.

(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)

---

3 The table summarises the times, in minutes, taken by 30 people to complete a puzzle.
Time: 20 - 30, 30 - 40, 40 - 50, 50 - 60, 60 - 70
Frequency: 6, 8, 9, 2, 5

Time (minutes)	Frequency
20 - 30	6
30 - 40	8
40 - 50	9
50 - 60	2
60 - 70	5

(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 3 is 9 marks)

© Revision Library | revisionlibrary.co.uk

4 A histogram is drawn for the lengths of time, in minutes, that customers wait. The class widths and frequencies are:
Class: 0 - 2, 2 - 7, 7 - 11, 11 - 13
Frequency: 24, 14, 23, 29

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

(b) Estimate the number of customers who waited longer than 4.5 minutes but not more than 7 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 4 is 6 marks)

---

5 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 $T = 33 - 1.5h$. The product moment correlation coefficient is $r = -0.78$.

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

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

(c) Comment on the strength and direction of the linear correlation indicated by $r = -0.78$. **(2)**

(Total for Question 5 is 6 marks)

---

6 The typing speeds, in words per minute, of 11 trainees on a data entry course are: 42, 45, 47, 48, 50, 51, 53, 55, 58, 60, 90.

(a) Calculate the mean and the standard deviation of these 11 typing speeds. **(4)**

(b) Show that the value 90 is an outlier, using the rule that a value is an outlier if it lies more than 2 standard deviations from the mean. **(3)**

(c) Given that, without the outlier, the sum of the remaining 10 typing speeds is 509 and the sum of their squares is 26201, calculate the new mean and new standard deviation for the remaining 10 trainees. **(4)**

(d) Comment on the effect that removing the outlier has had on the mean and on the standard deviation of the typing speeds. **(2)**

(Total for Question 6 is 13 marks)

© Revision Library | revisionlibrary.co.uk

Mark scheme · 2.2 Statistics: Data Presentation and Interpretation: 1-hour Pack

Question 1

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

(a) sum of f x midpoint = 940 **M1**

(a) mean = 940/42 = 22.38 minutes **A1**

(a) Answer: 22.38 minutes

(b) sum f x midpoint^2 = 22212.5 **M1**

(b) variance = 22212.5/42 - (22.3810)^2 **M1**

(b) standard deviation = 5.29 minutes **A1**

(b) Answer: 5.29 minutes

(c) identifies the median class 20 - 25 (cumulative frequency reaches 27 at its upper bound) **M1**

(c) interpolates: 20 + ((21 - 17)/10) x 5 **M1**

(c) median = 22.0 minutes **A1**

(c) Answer: 22.0 minutes

Question 2

(a) 52 - 33 **M1**

(a) 19 **A1**

(a) Answer: 19

(b) lower limit 33 - 1.5 x 19 = 4.5 and upper limit 52 + 1.5 x 19 = 80.5 **M1**

(b) smallest value 28 is not below 4.5, so is not an outlier **A1**

(b) largest value 89 is above 80.5, so is an outlier **A1**

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

(c) compares median - Q1 = 8 and Q3 - median = 11 **M1**

(c) positive skew **A1**

(c) Answer: positive skew

Question 3

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

(a) sum of f x midpoint = 1270 **M1**

(a) mean = 1270/30 = 42.33 minutes **A1**

(a) Answer: 42.33 minutes

(b) sum f x midpoint^2 = 58950 **M1**

(b) variance = 58950/30 - (42.3333)^2 **M1**

(b) standard deviation = 13.1 minutes **A1**

(b) Answer: 13.1 minutes

(c) identifies the median class 40 - 50 (cumulative frequency reaches 23 at its upper bound) **M1**

(c) interpolates: 40 + ((15 - 14)/9) x 10 **M1**

(c) median = 41.1 minutes **A1**

(c) Answer: 41.1 minutes

Question 4

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

(a) values 12, 14/5, 23/4, 29/2 **A1**

(a) Answer: 12, 14/5, 23/4, 29/2

(b) proportion of the class 2 - 7 lying in the required interval = (7 - 4.5)/5 **M1**

(b) 7 **A1**

(b) Answer: 7

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

(c) 1.2 cm **A1**

(c) Answer: 1.2 cm

Question 5

(a) each extra hundred metre in h is associated with a change of -1.5 in T, so as h increases T decreases **B1**

(a) Answer: gradient -1.5: T decreases by 1.5 for each unit increase in h

(b) substitutes h = 16 **M1**

(b) T = 9.00 **A1**

(b) The strong negative correlation supports the prediction, but reliability also depends on whether h = 16 lies within the observed range, which is not given. Accept equivalent comments. **B1**

(b) Answer: 9 degrees Celsius. The strong negative correlation supports the prediction, but reliability also depends on whether h = 16 lies within the observed range, which is not given.

(c) strong **B1**

(c) negative correlation **B1**

(c) Answer: strong negative linear correlation

© Revision Library | revisionlibrary.co.uk

Question 6

(a) attempt sum of the 11 values / 11 **M1**

(a) mean = awrt 54.5 **A1**

(a) attempt $\sqrt{\text{sum } x^2 / 11 - \text{mean}^2}$, using sum of squares of the 11 values **M1**

(a) SD = awrt 12.4 **A1**

(a) Answer: mean = awrt 54.5, SD = awrt 12.4 (words per minute)

(b) attempt mean + 2 x (their standard deviation) **M1**

(b) correct value: 54.4545 + 2(12.3683) = awrt 79.2 **M1**

(b) cso: since 90 > 79.2 (their boundary), 90 is an outlier; conclusion must follow correctly from their boundary **A1**

(b) Answer: 90 is an outlier (given)

(c) attempt new mean = 509/10 **M1**

(c) new mean = 50.9 (cao) **A1**

(c) attempt $\sqrt{26201/10 - (\text{their new mean})^2}$, dependent on a correct new mean **dM1**

(c) new SD = awrt 5.41 **A1**

(c) Answer: new mean = 50.9, new SD = awrt 5.41 (words per minute)

(d) the mean has decreased (from awrt 54.5 to 50.9), since the outlier of 90 was well above the rest of the data and was pulling the mean up, oe **B1**

(d) the standard deviation has decreased substantially (from awrt 12.4 to awrt 5.41), since the outlier was a long way from the mean and contributed a large amount to the spread, oe **B1**

(d) Answer: Both the mean and the standard deviation have decreased: the mean fell from awrt 54.5 to 50.9 because the high outlier was no longer pulling it up, and the standard deviation fell sharply from awrt 12.4 to awrt 5.41 because the outlier had been contributing a large deviation to the spread.

© Revision Library | revisionlibrary.co.uk