HIGHER

2.3 Data Cleaning and Outliers (Higher): 1-hour Pack

Total Marks

EDEXCEL 1ST0 · Calculator allowed · about 60 minutes

Name: _____ Date: ____ / ____ / _____

Answer ALL questions. Show all your working.

1 A sample of 6 measurements is 50, 49, 54, 119, 48, 55. The value 119 is an outlier.

(a) Work out the mean and the median of all 6 values (mean to 1 decimal place if needed).

(b) Work out the mean and the median without the outlier.

(c) Which average is changed less by removing the outlier?

(Total for Question 1 is 4 marks)

---

2 Here are the shoe sizes recorded for 9 people. Valid values are whole numbers from 3 to 12 inclusive.

10, 5, 10, 8, 4, 1, 4, -1, 12

(a) Write down the invalid values.

(b) Work out the mean of all 9 values, to 1 decimal place if needed.

(c) Work out the mean after the invalid values are removed, to 1 decimal place if needed.

(Total for Question 2 is 5 marks)

© Revision Library | revisionlibrary.co.uk

3 A sample of 10 measurements is 32, 34, 36, 30, 99, 29, 32, 34, 32, 33. The value 99 is an outlier.

(a) Work out the mean and the median of all 10 values (mean to 1 decimal place if needed).

(b) Work out the mean and the median without the outlier.

(c) Which average is changed less by removing the outlier?

(Total for Question 3 is 4 marks)

---

4 Here are the heights of pupils in centimetres recorded for 8 people. Valid values are whole numbers from 130 to 190 inclusive.

175, 157, 43, 130, 164, 149, 146, 999

(a) Write down the invalid values.

(b) Work out the mean of all 8 values, to 1 decimal place if needed.

(c) Work out the mean after the invalid values are removed, to 1 decimal place if needed.

(Total for Question 4 is 5 marks)

© Revision Library | revisionlibrary.co.uk

5 For the heights, in centimetres, of a group of people, the lower quartile is 23 and the upper quartile is 35. The smallest value is 11 and the largest value is 62. A value is an outlier if it is more than 1.5 x interquartile range below the lower quartile or above the upper quartile.

(a) Work out the interquartile range.

(b) Work out the limits beyond which a value is an outlier.

(c) State which of the smallest and the largest value, if any, are outliers.

(Total for Question 5 is 4 marks)

---

6 Here are the number of hours of sleep per night recorded for 9 people. Valid values are whole numbers from 3 to 12 inclusive.

7, 4, 12, 1, 5, 1, 10, 7, 12

(a) Write down the invalid values.

(b) Work out the mean of all 9 values, to 1 decimal place if needed.

(c) Work out the mean after the invalid values are removed, to 1 decimal place if needed.

(Total for Question 6 is 5 marks)

© Revision Library | revisionlibrary.co.uk

7 For the marks of a group of people in a test, the lower quartile is 25 and the upper quartile is 43. The smallest value is 9 and the largest value is 79. A value is an outlier if it is more than 1.5 x interquartile range below the lower quartile or above the upper quartile.

(a) Work out the interquartile range.

(b) Work out the limits beyond which a value is an outlier.

(c) State which of the smallest and the largest value, if any, are outliers.

(Total for Question 7 is 4 marks)

---

8 A sample of 8 measurements is 20, 18, 16, 21, 15, 19, 25, 57. The value 57 is an outlier.

(a) Work out the mean and the median of all 8 values (mean to 1 decimal place if needed).

(b) Work out the mean and the median without the outlier.

(c) Which average is changed less by removing the outlier?

(Total for Question 8 is 4 marks)

---

9 For the masses, in kilograms, of a group of people, the lower quartile is 26 and the upper quartile is 43. The smallest value is 10 and the largest value is 65. A value is an outlier if it is more than 1.5 x interquartile range below the lower quartile or above the upper quartile.

(a) Work out the interquartile range.

(b) Work out the limits beyond which a value is an outlier.

(c) State which of the smallest and the largest value, if any, are outliers.

(Total for Question 9 is 4 marks)

© Revision Library | revisionlibrary.co.uk

10 A sample of 8 measurements is 54, 52, 57, 57, 17, 55, 56, 51. The value 17 is an outlier.

(a) Work out the mean and the median of all 8 values (mean to 1 decimal place if needed).

(b) Work out the mean and the median without the outlier.

(c) Which average is changed less by removing the outlier?

(Total for Question 10 is 4 marks)

---

11 Here are the ages of club members in years recorded for 8 people. Valid values are whole numbers from 11 to 17 inclusive.

16, 999, 12, 12, 17, 11, -1, 13

(a) Write down the invalid values.

(b) Work out the mean of all 8 values, to 1 decimal place if needed.

(c) Work out the mean after the invalid values are removed, to 1 decimal place if needed.

(Total for Question 11 is 5 marks)

© Revision Library | revisionlibrary.co.uk

12 A sample of 10 measurements is 34, 93, 38, 35, 37, 41, 43, 35, 35, 37. The value 93 is an outlier.

(a) Work out the mean and the median of all 10 values (mean to 1 decimal place if needed).

(b) Work out the mean and the median without the outlier.

(c) Which average is changed less by removing the outlier?

(Total for Question 12 is 4 marks)

---

13 Here are the heights of pupils in centimetres recorded for 10 people. Valid values are whole numbers from 130 to 190 inclusive.

149, 43, 178, 168, 145, 132, 135, 177, 131, 133

(a) Write down the invalid values.

(b) Work out the mean of all 10 values, to 1 decimal place if needed.

(c) Work out the mean after the invalid values are removed, to 1 decimal place if needed.

(Total for Question 13 is 5 marks)

© Revision Library | revisionlibrary.co.uk

Mark scheme · 2.3 Data Cleaning and Outliers (Higher): 1-hour Pack

Question 1

mean 62.5, median 52	**B1**
removes 119 and recalculates	**M1**
mean 51.2, median 50	**A1**
the median	**B1**

Answer: (a) mean 62.5, median 52; (b) mean 51.2, median 50; (c) median

Question 2

invalid values 1, -1	**B1**
sum of all values divided by 9	**M1**
5.9 cao	**A1**
sum of valid values divided by 7	**M1**
7.6 cao	**A1**

Answer: (a) 1, -1; (b) 5.9; (c) 7.6

Question 3

mean 39.1, median 32.5	**B1**
removes 99 and recalculates	**M1**
mean 32.4, median 32	**A1**
the median	**B1**

Answer: (a) mean 39.1, median 32.5; (b) mean 32.4, median 32; (c) median

Question 4

invalid values 43, 999	**B1**
sum of all values divided by 8	**M1**
245.4 cao	**A1**
sum of valid values divided by 6	**M1**
153.5 cao	**A1**

Answer: (a) 43, 999; (b) 245.4; (c) 153.5

Question 5

12	**B1**
23 - 1.5 x 12 or 35 + 1.5 x 12	**M1**
5 and 53 cao	**A1**
the largest only	**B1**

Answer: (a) 12; (b) 5 and 53; (c) the largest only

Question 6

invalid values 1, 1	**B1**
sum of all values divided by 9	**M1**
6.6 cao	**A1**
sum of valid values divided by 7	**M1**
8.1 cao	**A1**

Answer: (a) 1, 1; (b) 6.6; (c) 8.1

Question 7

18	**B1**
25 - 1.5 x 18 or 43 + 1.5 x 18	**M1**
-2 and 70 cao	**A1**
the largest only	**B1**

Answer: (a) 18; (b) -2 and 70; (c) the largest only

Question 8

mean 23.9, median 19.5	**B1**
removes 57 and recalculates	**M1**
mean 19.1, median 19	**A1**
the median	**B1**

Answer: (a) mean 23.9, median 19.5; (b) mean 19.1, median 19; (c) median

Question 9

17	**B1**
26 - 1.5 x 17 or 43 + 1.5 x 17	**M1**
0.5 and 68.5 cao	**A1**
neither	**B1**

Answer: (a) 17; (b) 0.5 and 68.5; (c) neither

Question 10

mean 49.9, median 54.5	**B1**
removes 17 and recalculates	**M1**
mean 54.6, median 55	**A1**
the median	**B1**

Answer: (a) mean 49.9, median 54.5; (b) mean 54.6, median 55; (c) median

Question 11

invalid values 999, -1	**B1**
sum of all values divided by 8	**M1**
134.9 cao	**A1**
sum of valid values divided by 6	**M1**
13.5 cao	**A1**

Answer: (a) 999, -1; (b) 134.9; (c) 13.5

Question 12

mean 42.8, median 37	**B1**
removes 93 and recalculates	**M1**
mean 37.2, median 37	**A1**
the median	**B1**

Answer: (a) mean 42.8, median 37; (b) mean 37.2, median 37; (c) median

Question 13

invalid values 43	**B1**
sum of all values divided by 10	**M1**
139.1 cao	**A1**
sum of valid values divided by 9	**M1**
149.8 cao	**A1**

Answer: (a) 43; (b) 139.1; (c) 149.8

© Revision Library | revisionlibrary.co.uk