Answers: Introduction to data analysis

Michelle Arnetta

Summary

Answers to questions relating to the guide on introduction to data analysis.

These are the answers to Questions: Introduction to data analysis.

Please attempt the questions before reading these answers!

Q1

1.1. $\overline{x} = \frac{17+91+23+18+27+22}{6} = \frac{198}{6} = 33$

1.2. Ordered data set: $\{17, 18, 22, 23, 27, 91\}$. The list contains 6 elements, which is an even number. Since $\frac{6}{2} = 3$, the two middle elements would be the third and fourth elements. So the median is $\frac{22+23}{2} = 22.5$.

1.3. It is important to find the median of this data set, instead of only finding the mean, because it shows the central tendency of the data more accurately. The mean (33) is considerably higher than the median (22.5). This is because the mean was skewed upwards by an outlier or extremely high value in the data set, 91.

1.4. No values are repeated in the data set, so you would not be able to find a meaningful mode.

1.5. Accept all reasonable answers. Possible examples include "the outlier occurred on a Sunday" or "Cantor's Confectionery offered a special discount on that day".

Q2

2.1.

- $\overline{x} = \frac{14+71+82+39+39+71+71+71+48}{9} = \frac{506}{9} \approx 56.22$
- Ordered data set: $\{14, 39, 39, 48, 71, 71, 71, 71, 82\}$. The data set has 9 elements, which is an odd number. Since $\frac{9}{2} = 4.5$, which rounds up to 5, the middle value is the fifth element in the data set. So the median is 71.

- As shown in the frequency table, the unique value 71 occurs most frequently in the data set, so it is the mode.

Unique Value	Frequency
14	1
39	2
48	1
71	4
82	1

- The minimum value is 14, and the maximum value is 82. So the range is $82 - 14 = 68$.
- The lower half of the data set is $\{14, 39, 39, 48\}$, which has a median of $\frac{39+39}{2} = 39$. The upper half of the data set is $\{71, 71, 71, 82\}$, which has a median of $\frac{71+71}{2} = 71$. So the lower quartile is 39, and the upper quartile is 71. The interquartile range is $71 - 39 = 32$.

2.2.

- $\overline{x} = \frac{919+3293+912+7775}{4} = \frac{12899}{4} = 3224.75$
- Ordered data set: $\{912, 919, 3293, 7775\}$. The data set has 4 elements, which is an even number. Since $\frac{4}{2} = 2$, the two middle elements would be the second and third elements. So the median is $\frac{919+3293}{2} = 2106$.
- No values are repeated in the data set, so a meaningful mode cannot be found.
- The minimum value is 912, and the maximum value is 7775. So the range is $7775 - 912 = 6863$.
- The lower half of the data set is $\{912, 919\}$, which has a median of $\frac{912+919}{2} = 915.5$. The upper half of the data set is $\{3293, 7775\}$, which has a median of $\frac{3293+7775}{2} = 5534$. So the lower quartile is 915.5, and the upper quartile is 5534. The interquartile range is $5534 - 915.5 = 4618.5$.

2.3.

- $\overline{x} = \frac{3+72+800+8+763+8+3+9028+763+39+3}{11} = \frac{11490}{11} \approx 1044.54$

- Ordered data set: $\{3, 3, 3, 8, 8, 39, 72, 763, 763, 800, 9028\}$. The data set has 11 elements, which is an odd number. Since $\frac{11}{2} = 5.5$, which rounds up to 6, the middle value is the sixth element in the data set. So the median is 39.
- As shown in the frequency table, the unique value 3 occurs most frequently in the data set, so it is the mode.

Unique Value	Frequency
3	3
8	2
39	1
72	1
763	2
800	1
9028	1

- The minimum value is 3, and the maximum value is 9028. So the range is $9028 - 3 = 9025$.
- The lower half of the data set is $\{3, 3, 3, 8, 8\}$, which has a median of 3. The upper half of the data set is $\{72, 763, 763, 800, 9028\}$, which has median of 763. So the lower quartile is 3, and the upper quartile is 763. The interquartile range is $763 - 3 = 760$.

Q3

3.1. Pie chart.

3.2. Histogram.

3.3. Scatter plot.

3.4. Line graph.

Q4

4.1. Correlation test.

4.2. Regression analysis.

4.3. PDF, specifically Binomial PDF.

4.4. Confidence interval.

Version history and licensing

v1.0: initial version created 9/25 by Michelle Arnetta as part of a University of St Andrews VIP project.