

7.3

The Effects of Outliers on Average

≠ Outlier - Valor Extremo



Quick Review

A number in a data set that is very different from the other numbers is an outlier.

In this set of data: 35, 37, 39, 42, 82,
the outlier is 82 because it is much greater than the other numbers in the set.

The mean and median may be affected by removing the outliers.

For example, for the data set above: 35, 37, 39, 42, 82

To find the mean, add then divide:

$$(35 + 37 + 39 + 42 + 82) \div 5 = 235 \div 5 \\ = 47$$

The numbers are arranged in order, so the median is the third number: 39

With the outlier:

The mean is 47.

The median is 39.

Remove the outlier.

The data set is then: 35, 37, 39, 42

$$\text{The new mean is: } (235 - 82) \div 4 = 153 \div 4 \\ = 38.25$$

The new median is the mean of 37 and 39: $(37 + 39) \div 2 = 38$

Without the outlier:

The mean is 38.25.

The median is 38.

An outlier may result from an error in measurement or in recording.

In these cases, the outlier should be ignored when calculating averages.

Sometimes it is important to include outliers when calculating averages.

Practice

1. Identify any outliers in each data set.

a) 10, 20, 35, 35, 15, 95 _____

b) 3, 5, 8, 3, 2, 8, 5, 7, 4 _____

c) 10, 55, 61, 48, 60, 54, 97 _____

2. For each data set in question 1:

- Order the data from least to greatest.
- Calculate the mean.
- Calculate the median.
- Remove the outliers if they exist, then calculate the mean and median again.

a) From least to greatest: _____

Mean with outlier: _____

Median with outlier: _____

Mean without outlier: _____

Median without outlier: _____

b) From least to greatest: _____

Mean with outlier: _____

Median with outlier: _____

Mean without outlier: _____

Median without outlier: _____

c) From least to greatest: _____

Mean with outlier: _____

Median with outlier: _____

Mean without outlier: _____

Median without outlier: _____

3. During one week in February, the daily snowfalls in Kingston were:

5 cm, 4 cm, 21 cm, 6 cm, 3 cm, 7 cm, 3 cm

Calculate the mean, median, and mode of the data.

Mean: _____

Median: _____

Mode: _____

4. Use the data in question 3.
Calculate the mean, median, and mode without the outlier.

Data without the outlier: _____

Mean: _____

Median: _____

Mode: _____

5. Samia has these scores on her math quizzes:

55, 89, 78, 99, 85, 83, 82, 87, 80, 78

For the mid-term report, Samia can choose between:

- using the highest average of all 10 quiz scores or
- the highest average of those scores without the outliers.

What should Samia's choice be? Justify your answer.

For all 10 scores:

Mean = _____

= _____

Arrange the 10 scores in order: _____

Median = _____

= _____

Mode = _____

The outliers are: _____

The scores without the outliers are: _____

Mean = _____

= _____

Median = _____

= _____

Mode = _____

