

MEASURES OF CENTRAL TENDENCY

MEASURES OF CENTRAL TENDENCY

The measures of central tendency are numbers that locate the “center” of a set of data. The three most common measures of center are **mean**, **median** and **mode**.

MEAN

The **mean** is the arithmetic average of a data set. To find the mean, add all of the numbers in a data set, then divide by the number of values in the set.

Example 1 Find the mean of this set of data: 34, 31, 37, 44, 38, 34, 42, 34, 43 and 41.

- $34 + 31 + 37 + 44 + 38 + 34 + 42 + 34 + 43 + 41 = 378$
- $378 \div 10 = 37.8$

MEDIAN

The **median** is the middle number in a set of data that is arranged in numerical order. If there are an even number of values, the median is the mean of the two middle numbers.

Example 2 Find the median of this set of data: 34, 31, 37, 44, 34, 42, 38, 34, 43, and 41.

- Arrange the data in order: 31, 34, 34, 34, 37, 38, 41, 42, 43, 44.
- Find the middle value(s): 37 and 38.
- Since there are two middle values, find their mean: $37 + 38 = 75$, $75 \div 2 = 37.5$. Therefore, the median of this data set is 37.5.

Example 3 Find the median of this set of data: 92, 82, 80, 92, 78, 75, 95, 77, and 77.

- Arrange the data in order: 75, 77, 77, 78, 80, 82, 92, 92, and 95.
- Find the middle value(s): 80. Therefore, the median of this data set is 80.

MODE

The **mode** is the value in the data set that occurs most often. Data sets may have more than one mode or no mode at all.

Example 4 Find the mode of this set of data: 34, 31, 37, 44, 34, 42, 34, 43, 44, and 41.

- The mode of this data set is 34 because there are three 34's and none of the other numbers occur more than once or twice.

Example 5 Find the mode of this set of data: 92, 82, 80, 92, 78, 75, 95, 77, and 77.

- The modes of this set of data are 77 and 92 since there are two of each of these and only one of each of the other numbers. This data set is said to be bimodal because it has two modes.

RANGE

The **range** of a set of data is the difference between the highest value and the lowest value.

Example 6 Find the range of this set of data: 114, 109, 131, 96, 140, and 128.

- The highest value is 140 and the lowest value is 96.
- $140 - 96 = 44$.
- The range of this set of data is 44.

Problems

Find the mean, median, mode, and range of each set of data.

1. 29, 28, 34, 30, 33, 26, and 34.
2. 25, 34, 27, 25, 31, and 30.
3. 80, 89, 79, 84, 95, 79, 78, 89, 76, 82, 76, 92, 89, 81, and 123.
4. 116, 104, 101, 111, 100, 107, 113, 118, 113, 101, 108, 109, 105, 103, and 91.

Answers

1. mean: 30.57
median: 30
mode: 34
range: 8
2. mean: 30.33
median: 28.5
mode: 25
range: 9
3. mean: 86.13
median: 82
mode: 89
range: 47
4. mean: 106.6
median: 107
mode(s): 101 and 113
range: 27

OUTLIERS

An **outlier** is a number in a set of data that is much larger or much smaller than the other numbers in the set. Not all data sets have outliers.

To determine if a number is an outlier:

1. Find the median of the data.
2. Find the median of the lower half and the upper half of the data; these are called **lower** and **upper quartiles**, or **Q_1** and **Q_3** .
3. Find the **interquartile range**, by subtracting the *lower quartile* from the *upper quartile*.
4. Calculate $1.5 \times$ (Interquartile Range).

If any points are *below* $Q_1 - 1.5(\text{IQR})$ or *above* $Q_3 + 1.5(\text{IQR})$, they are outliers.

Example 1

Find the outlier of this set of data: 88, 90, 96, 93, 87, 72, 85, and 94.

- Order the data from least to greatest: 72, 85, 87, 88, 90, 93, 94, 96
- Find the median of the data: 72, 85, 87, **88, 90**, 93, 94, 96: median = 89
- Find the median of the lower and upper half of the data: 72, **85, 87**, 88, 90, **93, 94**, 96
lower median(Q_1) = 86, upper median(Q_3) = 93.5
- Calculate the interquartile range (IQR): $93.5 - 86 = 7.5$
- Find $1.5 \times$ (IQR): $1.5 \times 7.5 = 11.25$
- $Q_1 - 1.5(\text{IQR})$: $86 - 11.25 = 74.75$, $Q_3 + 1.5(\text{IQR})$: $93.5 + 11.25 = 104.75$
- Any number in the data set smaller than 74.75 or larger than 104.75 is an outlier. So, 72 is an outlier.

Example 2

Find the outlier of this set of data: 67, 54, 53, 76, 64, 59, 60, 72, 105, 44, and 66.

- Order the data from least to greatest: 44, 53, 54, 59, 60, 64, 66, 67, 72, 76, 105
- Find the median of the data: 44, 53, 54, 59, 60, **64**, 66, 67, 72, 76, 105
- Find the median of the lower and upper half of the data: 44, 53, **54**, 59, 60, **64**, 66, 67, **72**, 76, 105
 $Q_1 = 54$, $Q_3 = 72$
- Calculate the interquartile range (IQR): $72 - 54 = 18$
- Find $1.5 \times$ (IQR): $1.5 \times 18 = 27$
- $Q_1 - 1.5(\text{IQR})$: $54 - 27 = 27$, $Q_3 + 1.5(\text{IQR})$: $72 + 27 = 99$
- Any number in the data set smaller than 27 or larger than 99 is an outlier. So, 105 is an outlier.

Problems

Find the outlier (if it exists) for each set of data.

1. 70, 77, 75, 68, 98, 70, 72, and 71.
2. 14, 22, 17, 61, 20, 16, and 15.
3. 1376, 1645, 1783, 1455, 3754, 1790, 1384, 1643, 1492, and 1776.
4. 62, 65, 115, 51, 55, 14, 79, 85, 55, 72, 78, 83, 91, and 76.

Answers

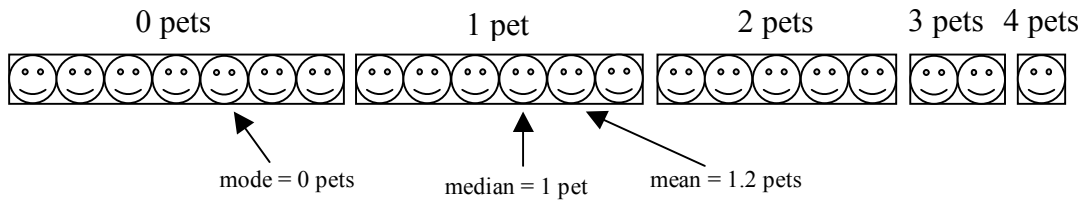
- | | | | |
|-------------------|---------------------|---------------------|------------------|
| 1. median = 71.5 | 2. median = 17 | 3. median = 1644 | 4. median = 74 |
| $Q_1 = 70$ | $Q_1 = 15$ | $Q_1 = 1455$ | $Q_1 = 55$ |
| $Q_3 = 76$ | $Q_3 = 22$ | $Q_3 = 1783$ | $Q_3 = 83$ |
| IQR = 6 | IQR = 7 | IQR = 328 | IQR = 28 |
| $6(1.5) = 9$ | $7(1.5) = 10.5$ | $328(1.5) = 492$ | $28(1.5) = 42$ |
| $Q_1 - 9 = 61$ | $Q_1 - 10.5 = 4.5$ | $Q_1 - 492 = 963$ | $Q_1 - 42 = 13$ |
| $Q_3 + 9 = 85$ | $Q_3 + 10.5 = 32.5$ | $Q_3 + 492 = 2246$ | $Q_3 + 42 = 126$ |
| 98 is an outlier. | 61 is an outlier. | 3754 is an outlier. | No outlier. |

RESISTANCE

Some measures of central tendency, namely median and mode, are said to be *resistant* to outliers, meaning that the presence of an outlier does not change the value much, if any. The mean is *not resistant* to outliers, so the value of the mean of a data set changes more drastically when an outlier is presented.

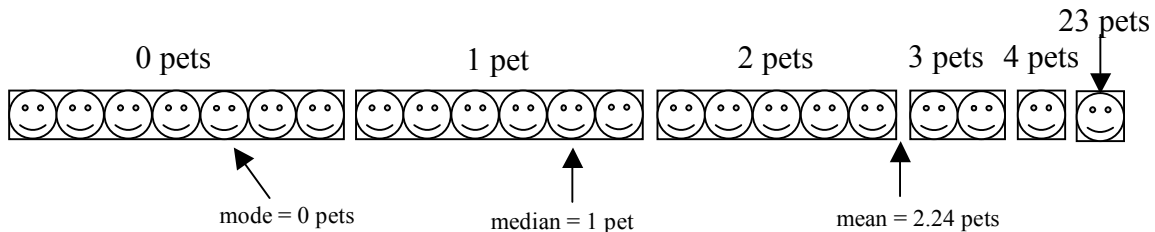
Example 3

Suppose we were to take a survey of a class of students and ask them how many pets they have at home. The results are shown below. ☺ = 1 student



Mode: 0 pets, Median: 1 pet, Mean: 1.2 pets

After the survey was taken a student came into class late. When asked how many pets he had he said that he lived on a farm and had 23 animals. The class wondered how this would affect the measures of central tendency.



Mode: 0 pets, Median: 1 pet, Mean: 2.24 pets

Notice, that the mode and the median did not change as much as the mean. Compare the two displays and look to see how the mean has changed—it went from 1.2 pets to over 2.2 pets. In this case one outlier nearly doubled the mean, thus distorting the fact that almost two-thirds of the class has at most one pet. This example why the mode and median are generally said to be *resistant to outliers*, and the mean is *not resistant to outliers*.

The notion of resistance is really only applicable to large data sets. Given a small set of data, any new piece of data could potentially have a drastic effect of any of the measures of central tendency. Therefore, we will not be doing any problems involving resistance, but it is important to understand the concept. Box-and-whisker plots, which are discussed in a later section, are based on medians because the median is not resistant to outliers.