

Arithmetic Average: See Mean

Deviation from the Mean: The deviation from the mean for a data value is the difference between the value and the mean. The deviation from the mean can be positive, negative, or zero.

Example: For the data on the line plot, the numbers in **red** are the deviations from the mean. The mean of the data is 3.

| | | | | | | | | | |
|---|----|---|---|----|---|----|--|---|----|
| X | -2 | | | | | | | | |
| X | -2 | | X | +0 | X | +1 | | X | +3 |
| | 1 | 2 | 3 | 4 | 5 | 6 | | | |

Note: Adding all the deviations from the mean, positive and negative, must result in zero, since the mean represents a balance point for these deviations—the point at which the excesses and deficits are perfectly balanced.

Mean: The mean of a data set is the arithmetic average of the data set, which is obtained by adding all the values in the data set, then dividing by the number of values in the set.

Example: Let's say you are writing about the World Wide Widget Co. and the salaries of its nine employees.

- The CEO makes \$100,000 per year,
- Two managers make \$50,000 per year,
- Four factory workers make \$15,000 each, and
- Two trainees make \$9,000 per year.

So you **add** \$100,000 + \$50,000 + \$50,000 + \$15,000 + \$15,000 + \$15,000 + \$15,000 + \$9,000 + \$9,000 (all the values in the set of data), which gives you \$278,000. Then **divide** that total by 9 (the number of values in the set of data)

That gives you the **mean**, which is \$30,889.

Note 1: The mean is very sensitive to extreme values in the data set

Note 2: An alternative explanation for the mean: The mean of a set of data from a single variable is the amount that each person or item would have if the total amount were shared equally among all persons or items.

Measures of Central Tendency: A measure of central tendency is a value that represents the data set. The mean, median, and mode are examples of measures of central tendency. Although all measures of central tendency represent the data set, they are not necessarily the same value.

- Note:*
- The mean is a good representation of the center when the data is close and there is no outlier
 - The median is a good representation of the center where there is an outlier and you don't want the outlier to affect the measure of center.
 - The mode is the only usually measure of central tendency for categorical data.

Median: The median of a set numbers is the value for which half the numbers are larger and half are smaller. If there are two middle numbers, the median is the arithmetic mean of the two middle numbers.

Example 1: Set of ordered data 1, 2, 3, 4, 5
The median number is 3 because 50% of the numbers are smaller and 50% of the numbers are larger.

Example 2: Set of ordered data 1, 2, 3, 4, 5, 6
The median number is 3.5 (add the two middle numbers and divide the sum by two).

Note 1: In finding the median the data **must** be ordered. We can say that the median is the value in the center of an ordered list of the data.

Note 2: Unlike the mean, the median is relatively unaffected by extreme values in the data and is therefore a resistant statistic.

Midrange: The midrange is the arithmetic average of the minimum and maximum values of the data set.

Note: The formula for midrange is ***midrange*** = (***minimum*** + ***maximum***)/2

Example: For the data set {10, 15, 20, 25, 30, 35, 40, 45, 50}
The minimum = 10 and the maximum = 50
Thus the midrange = $(10 + 50)/2 = 30$.

Mode: The mode is the most frequently occurring value in a data set.

Note 1: There can be more than one mode. If a data set has two modes it is called bimodal. There can be no mode if each number appears with equal frequency.

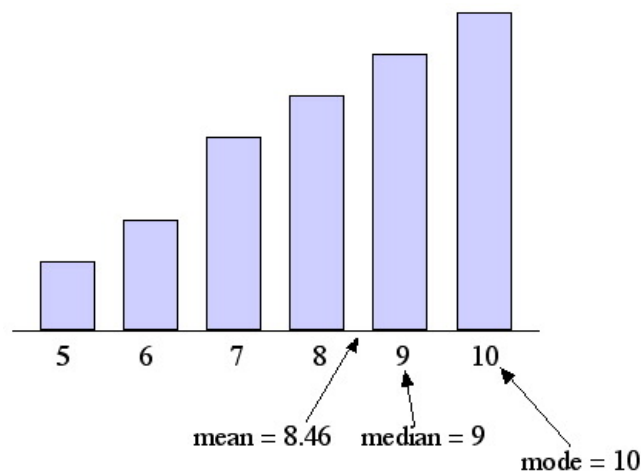
Example 1: Data Set 1: {14, 14, 14, 14, 15, 15, 16, 16, 16}
Mode = 14 because it occurred most often.

Example 2: Data Set 2: {15, 15, 15, 16, 16, 16, 17, 17, 17}
Has no mode because all numbers occurred equally

Example 3: Data Set 3: {blue, blue, red, red, red, red, green, green, green, green, white, orange, yellow, purple, purple, pink}
This data set is bimodal since it has 2 modes: red and green.

Note 2: The mode is the only measure of central tendency that applies to categorical data.

Illustration: For a distribution that is skewed left we can see that the
mean < median < mode



Resistant Statistic: A statistic that is relatively unaffected by extreme values or outliers in the data set.

Note 1: The mean and standard deviation are not resistant statistics since they are greatly influenced by extreme values.

Note 2: The median and interquartile range are resistant statistics since they are relatively unaffected by extreme values in the data