

Brief Explanation of Statistical Descriptive Measures

Centrality, Spread, and Shape

Table of contents

Centrality.....	1
Spread	1
Shape.....	1

There are 3 types of *measures of central tendency*: the mean (or average), median, and mode.

Centrality

The **MEAN** (or **AVERAGE**) is a measure of central tendency that returns the value obtained from adding all values and dividing the result by the number of values.

The **MEDIAN** is a measure of central tendency that returns the middle value in an ordered set of data. If there is an odd number of elements in the data set, then the median is that value for which the number of items preceding that value is equal to the number of items following that value.

In symmetrical distributions, the mean and the median are typically close in value.

In asymmetrical (skewed) distributions, or in samples with extreme outliers, the difference between the mean and the median can be significant.

The mean, contrary to the median, is very sensitive to outliers.

The **MODE** is a measure of central tendency that returns the value(s) in the data set that appear most frequently. In case more than one value is tied for the greatest frequency, the mode is considered the set of all such values.

Spread

There are several types of *measures of spread*: range, variance, standard deviation. The routine computes the standard deviation.

The **STANDARD DEVIATION** measures how spread out numbers are in relation to the mean/average of the numbers.

Shape

There are 2 types of *measures of shape*: skewness and kurtosis.

SKEWNESS indicates how symmetrical the data set is.

In perfectly symmetrical distributions, skewness will be zero.

In asymmetrical distributions:

1. When the skewness is negative, the data are left-skewed and the median is lower than the mean.
2. When the skewness is positive, the data are right-skewed and the median is greater than the mean.

KURTOSIS indicates how heavy your data set is about its mean compared to its tails.

In perfectly symmetrical distributions, kurtosis will be approximately 3.

In asymmetrical distributions:

1. When kurtosis is high (> 3), data will be clustered much more about the mean, and the tails will be relatively lighter.
2. When kurtosis is low (kurtosis < 3), data will not have a pronounced peak about the mean, and will consequently have heavier tails.