**BASIC STATISTICS**

**MEAN:**Arithmetic average of a range of values or quantities, computed by dividing the total of all values by the number of values.

MEAN=(N1+N2+………+Nn)/Total Number of observations

**MEDIAN:**

The median is the point at which there are an equal number of data points whose values lie above and below the median value. In a numerical data set, median refers to the original midpoint and not mean.

**For even number of observations:**

Median is calculated by arranging dataset in ascending order. If the total number of items in dataset is even, then find out average of the two middle numbers to find the median.

**For odd number of observations:**

If the total number of items in dataset is odd, then the middle number is said to be the median of the dataset.

**For example:**

To find the average salary of group of people, median should be used over mean. As lower outlier and higher outlier may cause variations in mean value of the dataset.

**MODE:**

Mode refers to the value that gets repeated most number of times in a dataset. There can be more than one mode for a dataset which makes it ambiguous to make conclusions at times. The mode is used when the most common item, characteristic or value of a data set is required.

**STANDARD DEVIATION:**

In a simple terms, it can be stated that standard deviation refers to how values in a dataset are spread from the mean.

**Formula to calculate standard deviation is **

In simple terms:

1.Calculate mean for the dataset.

2.From each number subtract the mean and square the result.

3.Find out the mean for squared differences.

4.Finally take the square root of that mean to get the standard deviation.

**VARIANCE:**

Variance is the method to find how much the numbers in dataset varies and spreads from the mean of the data. Variance is square of standard deviation. If the variance is zero it means that the data items are identical. At the same time, high variance refers to the fact that data items are very much spread out from the mean as well as among themselves.

**COVARIANCE:**

Covariance is a measure of correlation. Correlation refers to how strongly one variable is related to another variable. Covariance refers to how two random variables in a dataset will change together. Positive covariance means that two variables are Positively related whereas negative covariance means that they are inversely related.



X:independent variable

Y:dependent variable

n: number of items in the dataset

x(bar)-mean of independent variable x

y(bar):mean of dependent variable y