**What are Descriptive Statistics?**

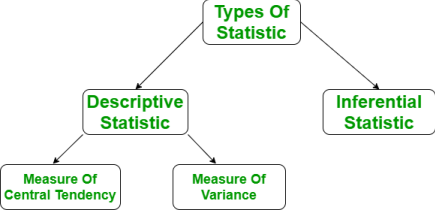
In Descriptive statistics, we are describing our data with the help of various representative methods using charts, graphs, tables, excel files, etc. In descriptive statistics, we describe our data in some manner and present it in a meaningful way so that it can be easily understood. Most of the time it is performed on small data sets and this analysis helps us a lot to predict some future trends based on the current findings. Some measures that are used to describe a data set are measures of central tendency and measures of variability or dispersion.

**Types of Descriptive Statistics**

• Measures of Central Tendency

• Measure of Variability

• Measures of Frequency Distribution



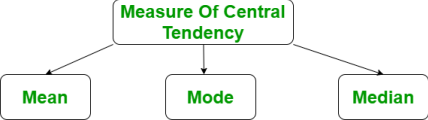
Measures of Central Tendency

It represents the whole set of data by a single value. It gives us the location of the central points. There are three main measures of central tendency:

• Mean

• Mode

• Median



***Mean***

It is the sum of observations divided by the total number of observations. It is also defined as average which is the sum divided by count.



where,

• x = Observations

• n = number of terms

***Mode***

It is the value that has the highest frequency in the given data set. The data set may have no mode if the frequency of all data points is the same. Also, we can have more than one mode if we encounter two or more data points having the same frequency.

***Median***

It is the middle value of the data set. It splits the data into two halves. If the number of elements in the data set is odd then the center element is the median and if it is even then the median would be the average of two central elements.

**Measure of Variability**

Measures of variability are also termed measures of dispersion as it helps to gain insights about the dispersion or the spread of the observations at hand. Some of the measures which are used to calculate the measures of dispersion in the observations of the variables are as follows:

• Range

• Variance

• Standard deviation

***Range***

The range describes the difference between the largest and smallest data point in our data set. The bigger the range, the more the spread of data and vice versa.

*Range = Largest data value – smallest data value*

***Variance***

It is defined as an average squared deviation from the mean. It is calculated by finding the difference between every data point and the average which is also known as the mean, squaring them, adding all of them, and then dividing by the number of data points present in our data set.



where,

• **x** -> Observation under consideration

• **N** -> number of terms

• **mu** -> Mean

***Standard Deviation***

It is defined as the square root of the variance. It is calculated by finding the Mean, then subtracting each number from the Mean which is also known as the average, and squaring the result. Adding all the values and then dividing by the no of terms followed by the square root.



where,

• x = Observation under consideration

• N = number of terms

• mu = Mean

Measures of Frequency Distribution

Measures of frequency distribution help us gain valuable insights into the distribution and the characteristics of the dataset. Measures like,

• Count

• Frequency

• Relative Frequency

• Cumulative Frequency

are used to analyze the dataset on the basis of measures of frequency distribution.

**Descriptive Statistics v/s Inferential Statistics** Generally, there are two types of statistics that are used to deal with the data when the requirements of the analyst are different.

The main difference lies in the final requirements only if the person just wants to extract meaningful insights from the data at hand then the domain of statistics that will be used by him is known as Descriptive Statistics but if we would like to use the observations to predict the future data let’s say in the time series related dataset then our objective contains the process of inference as well and hence it is also known as Inferential Statistics.

We can summarize this as when we would like to make some predictions/inferences based on some dataset at hand then those statistical methods are known as inferential statistics.