**Statistics Introduction**

**What is Statistics?**

It is Method which involves **Collection, Organizing, Analyzing and Interpretation** of Data/numbers into useful information for taking a decision.

**Collection:** Data accumulation

**Organizing:** Ordering/sorting data

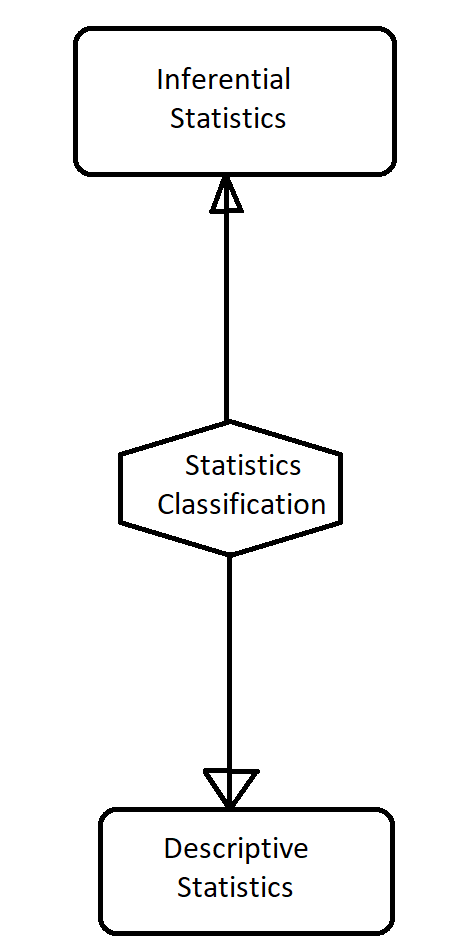
**Analyzing:** Looking closely/various dimensions of ordered Data

**Interpretation:** Understanding/making sense of the Data after analysis.

**Why Statistics?**

1. Better and more informed Decision making.
2. Presenting the data in a proper way
3. Reliable forecast about a business activity
4. Helps in efficient processes and outcome

**Classification of Statistics?**



**Descriptive Statistics :** That which describes the Data with presenting, organizing and Summarizing the data

**Inferential Statistics :** Arriving at a conclusion with observed data as a sample.

## Types of Descriptive Statistics :

There are two kinds of descriptive statistics that social scientists use:

[Measures of central tendency](https://www.thoughtco.com/measures-of-central-tendency-3026706) capture general trends within the data and are calculated and expressed as the mean, median, and mode. A mean tells scientists the mathematical average of all of a data set, such as the average age at first marriage; the median represents the middle of the data distribution, like the age that sits in the middle of the range of ages at which people first marry; and, the mode might be the most common age at which people first marry.

Measures of spread describe how the data are distributed and relate to each other, including:

* The range, the entire range of values present in a data set
* The frequency distribution, which defines how many times a particular value occurs within a data set
* Quartiles, subgroups formed within a data set when all values are divided into four equal parts across the range
* [Mean absolute deviation](https://www.thoughtco.com/what-is-the-mean-absolute-deviation-4120569), the average of how much each value deviates from the mean
* [Variance](https://www.thoughtco.com/variance-and-standard-deviation-3026711), which illustrates how much of a spread exists in the data
* Standard deviation, which illustrates the spread of data relative to the mean

Measures of spread are often visually represented in tables, pie and bar charts, and histograms to aid in the understanding of the trends within the data.

## Types of Inferential Statistics :

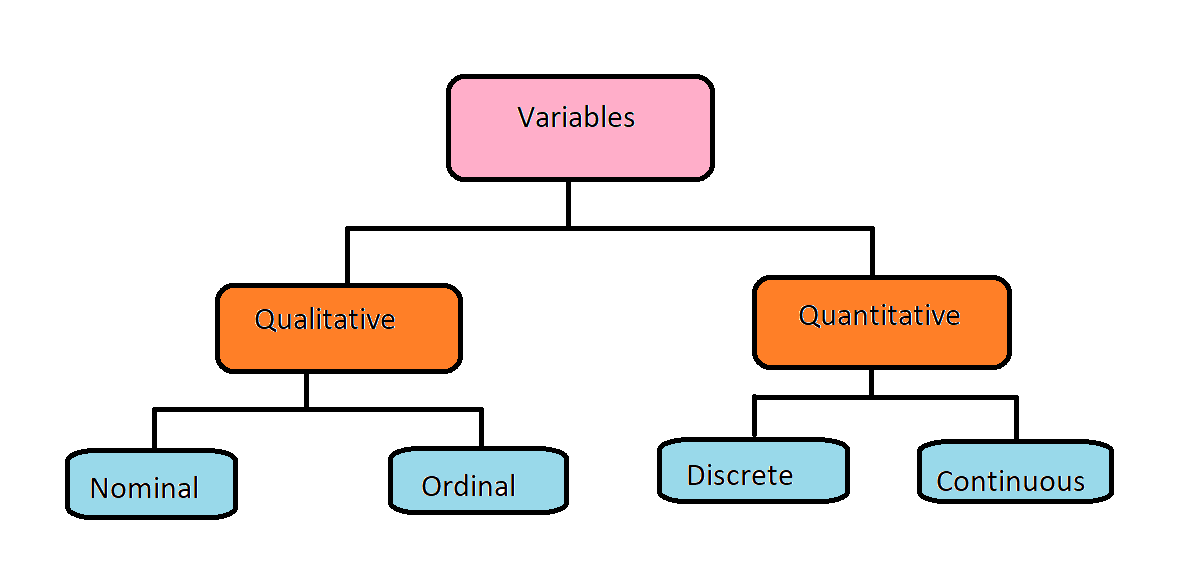
Inferential statistics are produced through complex mathematical calculations that allow scientists to infer trends about a larger population based on a study of a sample taken from it. Scientists use inferential statistics to examine the relationships between variables within a sample and then make generalizations or predictions about how those variables will relate to a larger population.

It is usually impossible to examine each member of the population individually. So scientists choose a representative subset of the population, called a statistical sample, and from this analysis, they are able to say something about the population from which the sample came. There are two major divisions of inferential statistics:

* A confidence interval gives a range of values for an unknown parameter of the population by measuring a statistical sample. This is expressed in terms of an interval and the degree of confidence that the parameter is within the interval.
* Tests of significance or [hypothesis testing](https://www.thoughtco.com/introduction-to-hypothesis-testing-3126336) where scientists make a claim about the population by analyzing a statistical sample. By design, there is some uncertainty in this process. This can be expressed in terms of a level of significance.

Techniques that social scientists use to examine the relationships between variables, and thereby to create inferential statistics, include [linear regression analyses](https://www.thoughtco.com/linear-regression-analysis-3026704), logistic regression analyses, [ANOVA](https://www.thoughtco.com/analysis-of-variance-anova-3026693), [correlation analyses](https://www.thoughtco.com/what-is-correlation-analysis-3026696), [structural equation modeling](https://www.thoughtco.com/structural-equation-modeling-3026709), and survival analysis. When conducting research using inferential statistics, scientists conduct a test of significance to determine whether they can generalize their results to a larger population. Common tests of significance include the [chi-square](https://www.thoughtco.com/chi-square-statistic-formula-and-usage-3126280) and [t-test](https://www.thoughtco.com/sample-t-test-confidence-interval-example-4022456). These tell scientists the probability that the results of their analysis of the sample are representative of the population as a whole.

## Variables :



**Qualitative :** Non-Numerical, More descriptive than prescriptive.

**Quantitative :** Numerical, More prescriptive than descriptive.

**Nominal :** Unstructured/unsorted Data describing some entity

**Ordinal :** Ordered /Structured which has a structured sequence

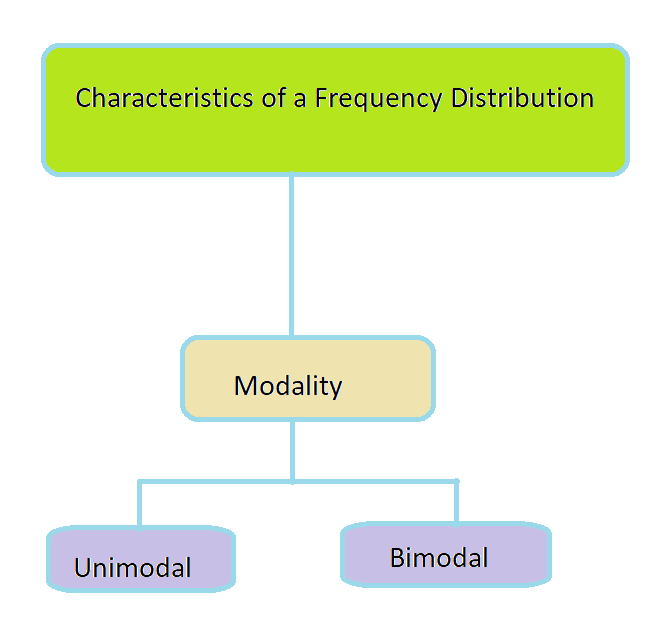
**Discrete :** Countable, something which occurs every once in a while.

**Continuous :**  Measurable, something which occurs regularly.

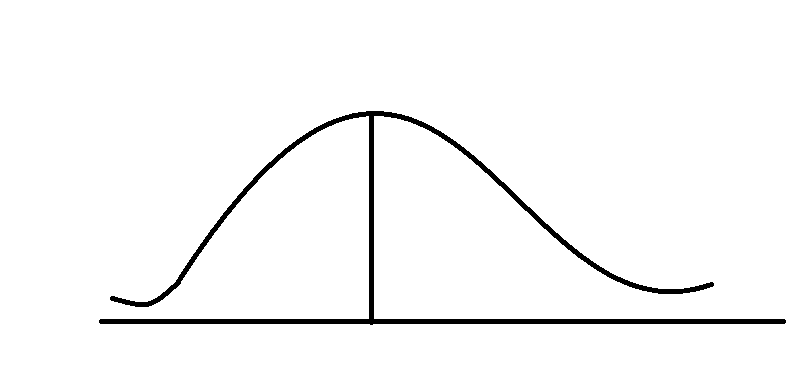
## Dependent and Independent Variables :

## Output of any system is a Dependent Variable and the Input is an Independent Variable

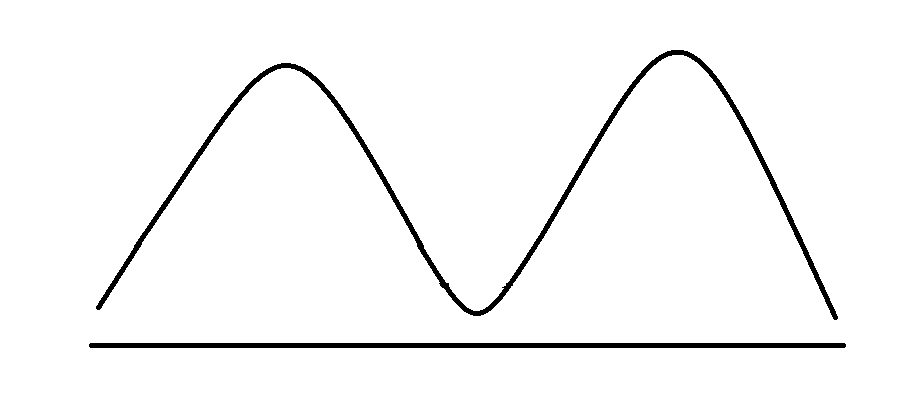
**Modality :**

****

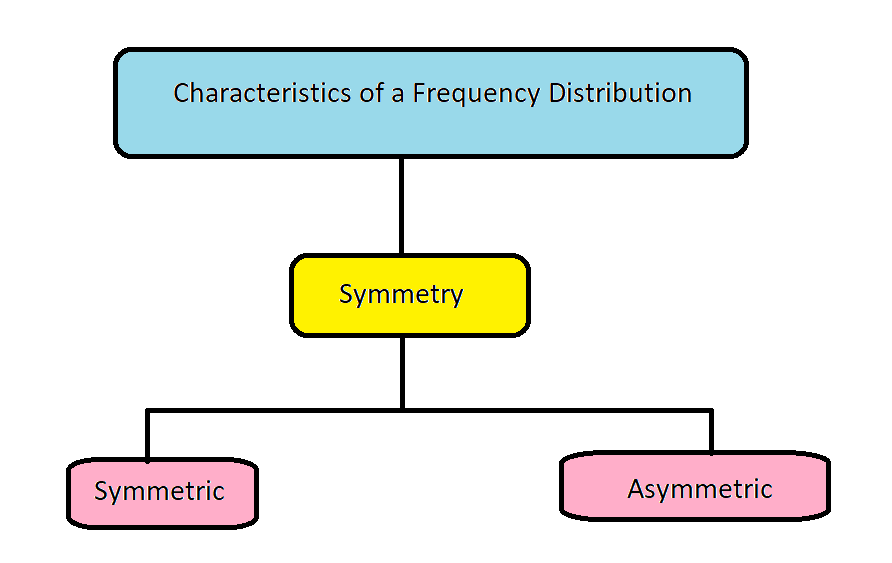
**Unimodal :**

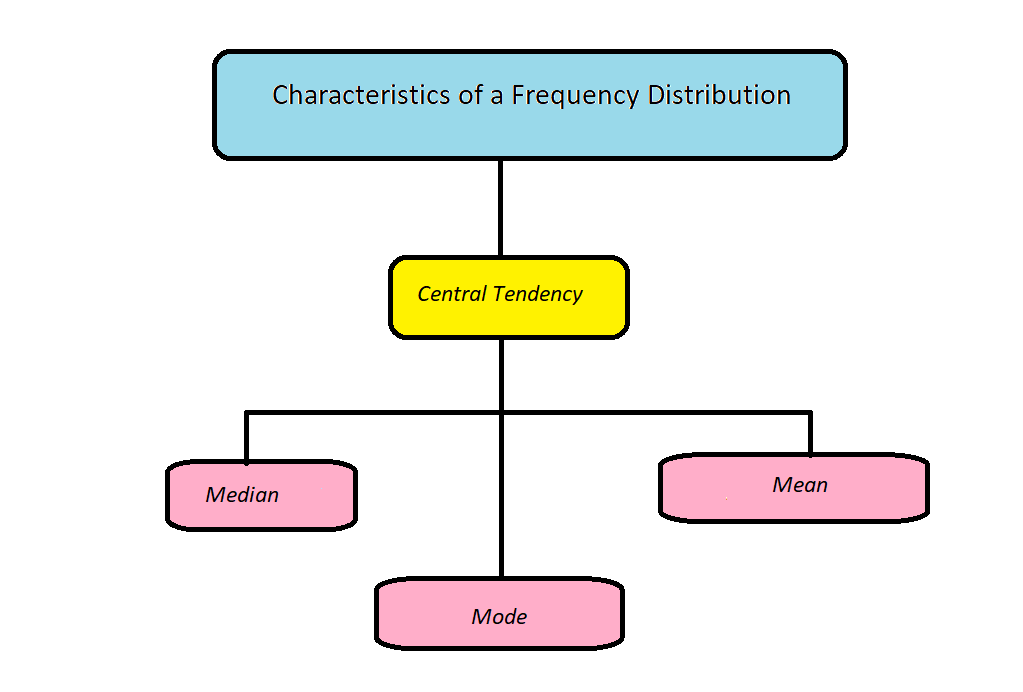
****

**Bimodal :**

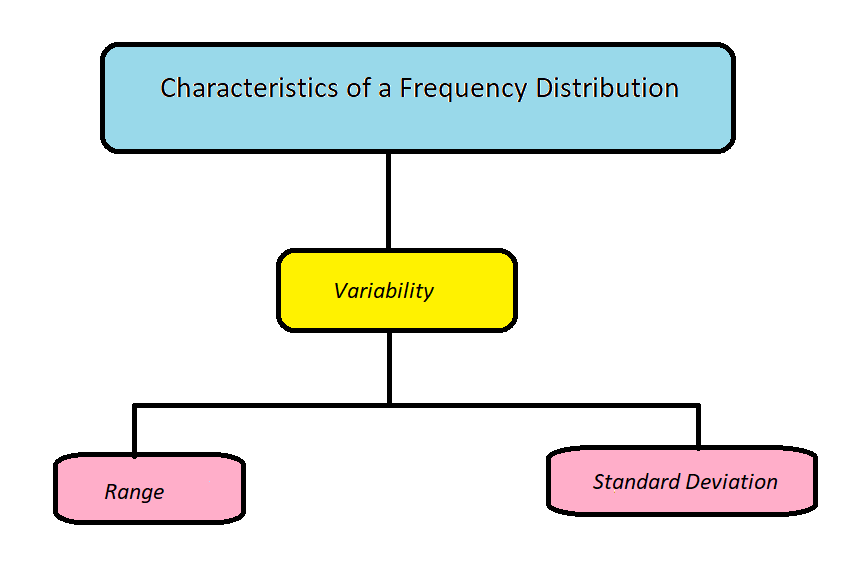
****

**Symmetry :**



**Central Tendency : **

**Variability :**

****

**Central Tendency:**

It is a single value around which the data is distributed.

**Mean: Median: Mode:**

Average Value is **Mean**, Mid Value after arranging all the value in order is **Median** and the most frequently occurring value is the **Mode**