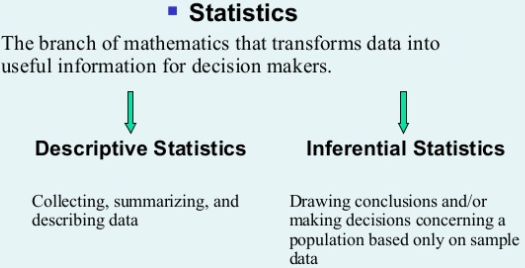
**Statistics**

A branch of mathematics taking and transforming numbers into useful information for decision makers. It is the process of **collecting data, analyzing data, interpreting data and presenting data.** It is a way to get information from data. Learning it is important as knowledge of statistics allows you to make a better sense of use of numbers.

**Types**:

Statistics can be classified into two types:

1. **Descriptive** Statistics
2. **Inferential** Statistics



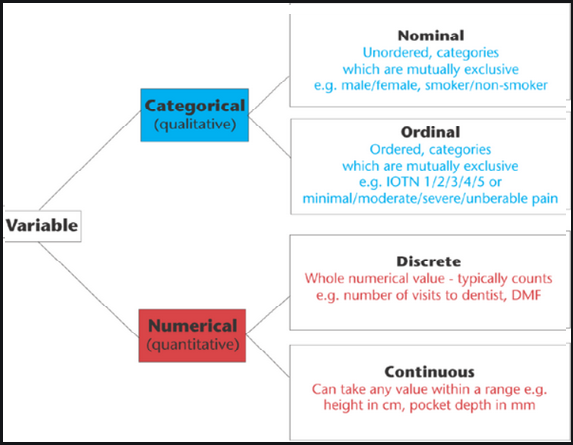
**Descriptive statistics** present, summarize and organize characteristics of a data set(sample). A data set is a collection of responses or observations from a [sample or entire population](https://www.scribbr.com/methodology/population-vs-sample/).

In [quantitative research](https://www.scribbr.com/methodology/quantitative-research/), after [collecting data](https://www.scribbr.com/methodology/data-collection/), the first step of data analysis is to describe characteristics of the responses, such as the average of one variable (e.g., age), or the relation between two variables (e.g., age and creativity).

The next step is**inferential statistics**, which are tools that help you decide whether your data confirms or refutes your hypothesis and whether it is generalizable to a larger population..Drawing conclusions about a population based on data observed in a sample. use the data to learn about the [population](https://en.wikipedia.org/wiki/Statistical_population) that the sample of data is thought to represent

**Variables in Statistics:**

Statistical variables (data) can be classified in two ways, **quantitative** and **qualitative**

****

1. **Quantitative:**

Qualitative variables are variables that are **categorical** in nature, or that don’t have any numerical representation. They are further classified as

* 1. **Ordinal:**

Ordinal variables are those that can be arranged in **orders** or during the arrangement where order matters.

* 1. **Nominal**

Nominal variables are those variables where order doesn’t matter at all.

1. **Quantitative** **variables** :

**Quantitative**variables are variables that have some **numerical** representation and they contain some information numerically. Height, Age, Weight are the types that come under this category.They are further classfied as:

* 1. **Discrete**:

Variables whose values are **countable**(whole number)

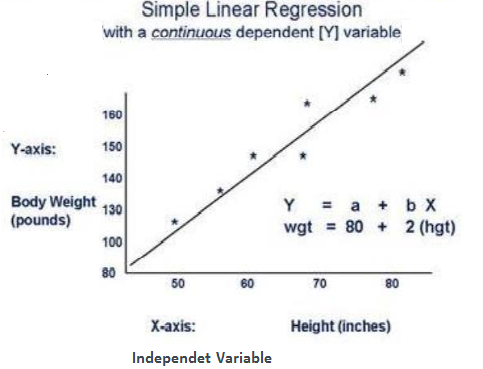
e.g: Number of Students : 12

* 1. **Continuous**:

Variables whose values are **measurable**(continuous in nature - decimal)

e.g: Height : 175.5 cm.

**Variables  -  Dependent  &  Independent**

****

1. **Dependent  or** **Target** **or** **Class** **or** **Predeicted** **variable:**

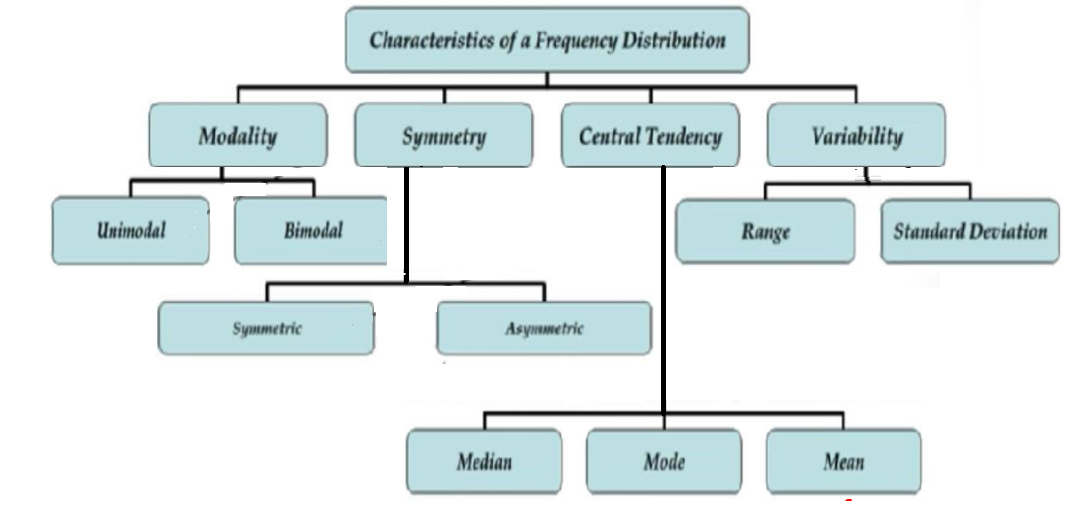
The dependent variables are the types of variables that are completely **dependent** on the independent variables (Feature/Column). The dependent variables are named as such because they are the values that are predicted or assumed by the predictor / independent variables.

1. **Independent  or** **Predictor variable:**

The independent variables are called as such because **independent** variables **predict** or forecast the values of the dependent variable in the model.

**Characteristics of frequency distribution  (Summarizing data):**

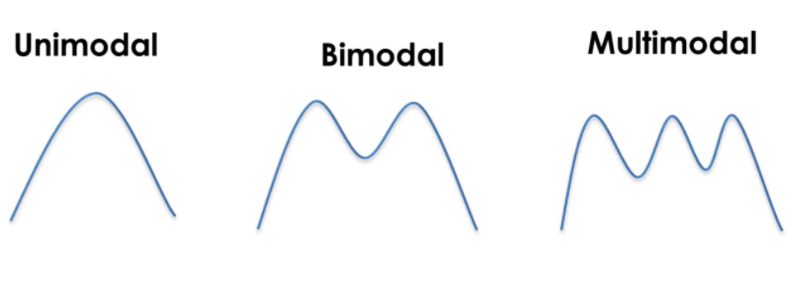
Frequency distribution in [statistics](https://www.toppr.com/guides/business-mathematics-and-statistics/statistical-description-of-data/introduction-to-statistics/) provides the information of the [number](https://www.toppr.com/guides/maths/can-you-see-the-pattern/pattern-in-figures-and-numbers/) of occurrences (frequency) of distinct values distributed within a given period of time or interval, in a list, table, or [graphical representation](https://www.toppr.com/guides/business-economics-cs/descriptive-statistics/graphic-presentation-of-data/)

****

1. **Modality:**

The **mode** or modal value is the number that appears **most frequently** in a data set. A set of numbers may have one mode, more than one mode, or no mode at all.

 Unimodal has single peak(mode=1),Bimodal has two peak(mode=2,there are two values that occured most frequently)

****

1. **Symmetry:**

Symmetry can be classified as **symmetric** and **asymmetric**

**Symmetric:**

Symmetrical distribution is a situation in which the values of variables occur at regular frequencies and the **mean, median, and mode** occur at **same** **points**.   Symmetric curve is **bell shaped curve**

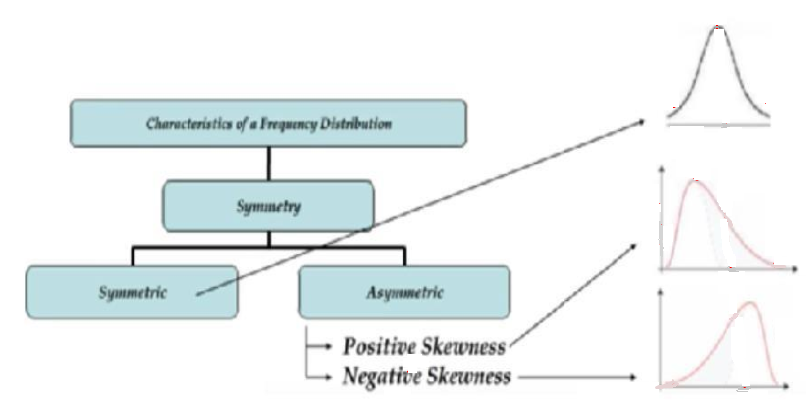
**Asymmetric:**

Asymmetrical distribution is a situation in which the values of variables occur at **irregular** frequencies and the **mean, median, and mode** occur at **different** points. An asymmetric distribution exhibits **[skewness](https://www.investopedia.com/terms/s/skewness.asp)**.   Two types of skewness are :

1. **Positive skewness**or**right-skewed:**

Tail on the right side of the distribution is longer than left

1. **Negative** **skewness**  or  **left-skewed:**

Tail on the left side of the distribution is longer than right****

1. **Central tendency:**

 A  **central tendency** (or **measure of central tendency**) is a central or typical value for a [probability distribution](https://en.wikipedia.org/wiki/Probability_distribution). It may also be called a **center** around which data is distributed.

The common measures of central tendency are **Mean**(average),   **mode**(most frequent value),   **Medain**(middle value in a sorted list of numbers in case of odd amount of numbers and average of middle values in a sorted list of even amount of numbers )

1. **Variability**:

  The difference being exhibited by data points within a data set, as related to each other or as related to the mean. This can be expressed through the range, variance or standard deviation(tells how each value is distributed from mean) of a data set.

1. **Range**:

Difference between highest and lowest value.

1. **Standard deviation:**

Measure of the amount of variation or dispersion of a set of values.  A low standard deviation indicates that the values tend to be close to the mean

