Basic definitions

Statistics: It is a collection of methods for planning experiments, obtaining data, and then organizing, summarizing, presenting, analyzing, interpreting, and drawing conclusions.

Variable: It is a characteristic or attribute that can assume different values.

Random variable: It is a variable whose values are determined by chance.

Population: All subjects possessing a common characteristic that is being studied.

Sample: It is a subset of the population.

Parameter: It is a characteristic or measure obtained from a population.

Statistic: It is a characteristic or measure obtained from a sample. *Not to be confused with Statistics*

Descriptive statistics: It is a collection, organization, summarization and presentation of the data.

Inferential statistics: It is generalizing from samples to populations using probabilities. Performing Hypothesis Testing, determining relationships between variables, and making predictions.

Qualitative variables: Variables which assume non-numerical values. Ex: Type of person -> Man, woman; Indian Seasons -> Summer, autumn, winter, Spring; etc.

Quantitative variables: Variables which assume numerical values. Ex: Temperature during the day, stock prices of company day by day, etc.

Discrete variables: Variables which assume a finite or countable number of possible values. Usually obtained by counting.

Continuous variables: Variables which assume an infinite number of possible values. Usually obtained by measurement.

Nominal Level: It is a level of measurement which classifies data into mutually exclusive, all inclusive categories in which no order or ranking can be imposed on the data.

Ordinal Level: It is a level of measurement which classifies the data into categories that can be ranked. Differences between the ranks do not exist.

Interval Level: It is a level of measurement which classifies the data that can be ranked and differences are meaningful. However there is no meaningful zero, so ratios are meaningless.

Ratio Level: It is a level of measurement which classifies the data that can be ranked, differences are meaningful, and there is a true zero. True ratios exists between the different units of measure.

Random sampling: It is a sampling in which the data is collected using chance methods or random numbers.

Systematic sampling: It is a sampling in which data in obtained by seleting every kth object.

Convenience sampling: It is a sampling in which data is readily available is used.

Stratified sampling: It is a sampling in which data is divided into groups (called *Strata*) according to some characteristic. Each of these strata is then sampled using one of the other sampling techniques.

Cluster sampling: It is a sampling in which the population is divided into groups (usually geographically). Some of these groups are randomly selected and then all of the elements in those groups are selected.