Some Terminology

Population: Universe. The total category under consideration. It is the data which we have not completely examined but to which our conclusions refer.

Sample: That portion of the population that is available, or is to be made available, for analysis.

Parameter: A characteristic of a population. E.g., the population mean, μ.

Statistic: A measure derived from the sample data. E.g., the sample mean, $\bar{X}$.

Statistical Inference: The process of using sample statistics to draw conclusions about population parameters. For instance, using $\bar{X}$ (based on a sample of, say, n=1000) to draw conclusions about μ (population of, say, 240 million).

Descriptive Statistics: Those statistics that summarize a sample of numerical data in terms of averages and other measures for the purpose of description. This includes the presentation of data in the form of graphs, charts, and tables. Descriptive statistics are not concerned with the theory and methodology for drawing inferences that extend beyond the particular set of data examined.

Primary data: Data compiled by the researcher. Secondary data: Data compiled or published elsewhere, e.g., Statistical Abstracts, census data.

Qualitative data result in categorical responses. Quantitative data result in numerical responses, and may be discrete or continuous. Discrete data arise from a counting process. Continuous data arise from a measuring process.

Probability Sample: A sample collected in such a way that every element in the population has a known chance of being selected.

Simple Random Sample: A sample collected in such a way that every element in the population has an equal chance of being selected.

The median is the data value such that half of the observations are larger than it and half are smaller.

The mode is the value of the data that occurs with the greatest frequency.

Probability: The word probability is actually undefined, but the probability of an event can be explained as the proportion of times, under identical circumstances, that the event can be expected to occur. It is the event's long-run frequency of occurrence.

Objective probabilities are long-run frequencies of occurrence, as above.

Subjective probabilities measure the strengths of personal beliefs. Objective probability refers to stochastic processes.

Stochastic process: A repetitive process which generates outcomes (called events) that are not identical, and not individually predictable with certainty, but that may be described in terms of relative frequencies.

Random variable: That which is observed as the result of a stochastic process. A random variable takes on (usually numerical) values. Associated with each value is a probability that the value will occur.

A Discrete random variable can take on only specified, distinct values.

A Continuous random variable can take on any value within an interval.

A probability distribution for a discrete random variable is a mutually exclusive listing of all possible numerical outcomes for that random variable, such that a particular probability of occurrence is associated with each outcome. Some discrete probability distributions: Binomial distribution, Hypergeometric distribution, Poisson distribution.

Probability density function: A continuous probability distribution. The probability is interpreted as "area under the curve." Some continuous probability distributions: Normal distribution, Standard Normal (*Z*) distribution, Student's *t* distribution, Chi-square (χ^2) distribution, *F* distribution.

Simple probability: P(A). The probability that an event (say, A) will occur. Also called a marginal probability.

Joint probability: P(A and B). P($A \cap B$). The probability of events A and B occurring together.

Conditional probability: P(A|B), read "the probability of A given B." The probability that event A will occur given event B has occurred.

Mutually Exclusive events: Two events are mutually exclusive if they cannot occur together. E.g., male or female; heads or tails.

Independent events: Events A and B are independent if knowledge of the occurrence of one has no effect on the probability that the other will occur.
Example: P(Blue eyes | Male) = P(Blue eyes)