

## Basic Terminology:

**Statistics** is the study of how to collect, organize, analyze, and interpret numerical information from data.

**Individuals** are the people or objects included in the study.

A **variable** is the characteristic of the individual to be measured or observed.

A **quantitative variable** has a value or numerical measurement for which operations such as addition or averaging make sense.

A **qualitative variable** describes an individual by placing the individual into a category or group such as male or female.

In **population data**, the variable is from *every* individual of interest.

In **sample data**, the variable is from *only some* of the individuals of interest.

<u>Level of Measurement</u>	<u>Suitable Calculation</u>
Nominal	We can put the data into categories
Ordinal	We can order the data from smallest to largest or “worst” to “best”. Each data can be compared with another data value.
Interval	We can order the data and also take the differences between data values. At this level, it makes sense to compare the differences between data values. For instance, we can say that one data value is 5 more than another or 12 less than another data value.
Ratio	We can order the data, take differences, and also find the ratio between data values. For instance, it makes sense to say that one data value is twice as large as another.

**Descriptive statistics** involves methods of organizing, picturing, and summarizing information from samples or populations.

**Inferential statistics** involves methods of using information from a sample to draw conclusions regarding the population.

A **simple random sample** of  $n$  measurements from a population is a subset of the population selected in a manner such that:

- a) every sample of size  $n$  from the population has an equal chance of being selected and
- b) every member of the population has an equal chance of being included in the sample.

In a **census**, measurements or observations from the *entire* population are used.

In a **sample**, measurements or observations from a *representative part* of the population should be used.

In an **observational study**, observations and measurements of individuals are conducted in a way that doesn't change the response or the variable being measured.

In an **experiment**, a *treatment* is deliberately imposed on the individuals in order to observe a possible change in the response or variable being measured.

There is a **control group**. This group received a dummy treatment enabling the researchers to control for the placebo effect. In general, a control group is used to account for the influence of other known or unknown variables that might be an underlying cause of a change in response in the experimental group. Such variables are called **lurking** or **confounding variables**.

**Randomization** is used to assign individuals to the two treatment groups. This helps prevent bias in selecting members for each group.

**Replication** of the experiment on many patients reduces the possibility that the differences in pain relief for the two groups occurred by chance alone.

A **Pareto chart** is a bar graph in which the height represents *frequency* of an event. In addition, the bars are arranged from left to right according to decreasing height.