

Vocabulary of Statistics

Statistics is the science of collecting, organizing, analyzing, and interpreting 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.

Data consist of information coming from observations, counts, measurements, or responses.

A **population** is the collection of *all* outcomes, responses, measurements, or counts that are of interest.

A **parameter** is a numerical description of a *population* characteristic.

A **sample** is a subset, or part, of a population.

A **statistic** is a numerical description of a *sample* characteristic.

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.

Levels of Measurement

Nominal level of measurement applies to data that consist of names, labels, or categories. There are no implied criteria by which the data can be ordered from smallest to largest.

Ordinal level of measurement applies to data that can be arranged in order; however, differences between data values either cannot be determined or are meaningless.

Interval level of measurement applies to data that can be arranged in order. In addition, differences between data values are meaningful.

Ratio level of measurement applies to data that can be arranged in order. In addition, both differences between data values and ratios of data values are meaningful. Data at the ratio level have a true zero.

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 an **observational study** researchers do not influence the responses. They simply observe. In an **experiment** a researcher deliberately applies a treatment before observing responses.

A researcher observes and measures characteristics of interest of part of a population but does not change existing conditions.

In an experiment, the researcher applies a treatment – **treatment group** – and responses are observed. A **control group** could be part of the same population in which no treatment is applied. The subjects are referred to as **experimental units**. A **placebo** (a fake treatment) can be used in the control group.

Data Collection

Simulation – using math or physical models to reproduce conditions of a situation or process. Simulations allow you to study situations that are impractical or dangerous to create in real life.

Survey – carried out on people by asking questions by Internet, phone, mail, or interviewing. Careful wording is needed so bias is avoided.

Three key elements of a well-designed experiment are **control, randomization, and replication**.

Problems with design can occur. A **confounding variable** occurs when an experimenter cannot tell the difference between the effects of different factors on the variable. The **placebo effect** occurs when a subject reacts favorably to a placebo when in fact the subject has been given a fake treatment.

Use of **blinding** techniques is where the subjects do not know whether or not they received a treatment or a placebo. In a **double-blind experiment**, neither the researcher nor the subject knows if they are receiving a treatment or placebo. This type of experimental design is preferred by researchers.

Randomization is a process of randomly assigning subjects to different treatment groups.

Blocks, randomized block design, matched-pairs design

The number of subjects in an experiment is called the **sample size**.

Replication is the repetition of an experiment under the same or similar conditions.