

# Sampling and Experimentation: Overview of Methods of Data Collection

Everyday, statements appear in newspapers, magazines, and on television that involve the use of data. Statements such as “The average home price is now \$456,000,” “Sixty-three percent of college students eat in restaurants at least four times a week,” and “The average cost of a college education has increased 47% in the past two years” involve collecting and analyzing data. Everything a statistician does relies on data. Accurate data and valid data collection methods are essential to ensure valid results. Incorrect data or data collected with poorly designed methods can lead to incorrect conclusions.

**Data** consists of information derived from measurements, observations, counts, or responses. **Statistics** is a science. It is the science of collecting, organizing, summarizing, analyzing, and interpreting data. There are two main types of sets of data: populations and samples. A **population** is a collection of all of the measurements, observations, counts, or responses that you are interested in analyzing. A **sample** is a subset or part of a population. Statisticians usually deal with samples because working with an entire population is usually difficult, costly, or even impossible.

Characteristics of a population are called **parameters**. The mean of a population is a population parameter. The standard deviation of a population is a population parameter. Measures that identify characteristics of a sample are called **statistics**. (The word “statistics” should not be capitalized. The capitalized version is the name of the science.) The mean of a sample is a sample statistic. The standard deviation of a sample is a sample statistic.

## EXAMPLE:

Classify each of the following as either a sample or a population:

The height of each employee in a company.

The speed of every other car passing a point on a highway during a given hour.

A questionnaire from 350 students in a school of 2000 students.

The area of each state in the United States.

The number of pages contained in each of the encyclopedia in a library.

The height of each employee in a company is data from a population. Data is obtained from every member in the set of company employees. The speed of every other car is data from a sample since only some of the cars are used. Questionnaires from 350 of 2000 students is also a sample. Only a portion of the entire set of students is sampled. The area of each state would be a population since all the states are included. Since each encyclopedia is included in the data set, this is also an example of a population.

The populations listed in the previous example, although fitting the definition of *population*, are quite small. Populations can be very large. A study of tuna fish includes the population of all tuna. This is an extremely large population. Most of the tuna that exist are not even accessible for study. It would be impossible to work with the entire population. Studies using samples of tuna would be more appropriate. A study of all 18-year-old males in the United States would be a population of more than 2,000,000 people. It would be very difficult and very costly to conduct a study using this entire population. Using a sample of 18-year-old males would be much more efficient.

## Census

When measurements are taken for every member of a population it is called a **census**. Most people equate the word census with the tabulation of U.S. population data conducted every 10 years. The last U.S. Census was conducted in 2000. It is the intent of our government to collect data on every member of the U.S. population, which is an impossible task. Many of the people living in the United States are missed and do not appear in the census figures. The U.S. census is, of course, not the only census. Anytime a study is made of an entire population, a census is the result. It is possible to conduct a census of all students taking a specific class in a specific community college. It is also possible, but more difficult, to conduct a census of all students attending a specific community college. It would not be practical to try to conduct a census of all students in all community colleges.

## Sample Survey

It is the intent of a **sample survey** to collect data from a representative portion of a population and to record the results. The way the sample is determined may vary, but the goal is that inferences about the entire population can be drawn from data taken in the sample survey. Political polls are examples of sample surveys. Statements about the entire population of voters are made based on the results of the poll. Surveys may be conducted in person, through the mail, with questionnaires, or on the telephone. No matter how the data is collected, it is important that it be representative of the population. The researcher makes no attempt to alter or change the population. A well-executed sample survey can obtain valid data without the time and expense of conducting a census.

## Experiment

In designed, researcher-controlled studies, called **experiments**, the researcher maintains some control over the subjects of the study. By controlling certain aspects of the experiment, the researcher can determine the effect, if any, changes in the explanatory variable has on the outcome, or response variable. Subjects are usually divided into two groups, the *treatment group* and the *control group*. Medical studies are often experiments. Test groups are divided in two, with one half given certain medication and the other half given some placebo. If both the subjects and the researchers are unaware of which group received the placebo and which group received the medicine, the experiment is called a double-blind study. Testing the effectiveness of artificial sweeteners by asking opinions after giving half the test subjects a beverage sweetened by sweetener A and half of the test subjects the same beverage sweetened by sweetener B and then comparing the results would also be an experiment. In a properly designed experiment, more information is obtainable than from an uncontrolled study.

## Observational Study

**Observational studies** are similar to experiments except that the researcher has no control over which subjects are placed in the control group and which in the treatment group. If a researcher is studying to see whether increased sleep time prior to the test results in better scores on IQ tests, the researcher cannot artificially control the sleep time of the control and treatment groups. Data is obtained from all subjects, and the results are examined for possible relationships. A researcher may observe a group of young children and note which children speak loudly and which speak softly. The researcher then looks for factors that might be related, such as height or weight. Unlike controlled experiments, a researcher conducting an observational study makes no attempt to control any part of the study.

### EXAMPLE:

As a researcher, you are interested to see whether aerobic exercise is more effective in weight loss than strength training. Which of the following two studies would be considered an observational study and which one would be an experiment?

- A. Data was collected by observing people at a gym, noting the kind of exercise they were doing and the resulting amount of weight loss.
- B. Data was collected by observing two groups of people. One group was asked to concentrate on strength training and the other group on aerobic exercise. Both groups were given the same diet to follow.

The first is an example of an observational study. The researcher is not controlling any aspect of the study. The observed data then is analyzed, and conclusions may be drawn from the results.

The second is an example of an experiment. The researcher is controlling the study. The researcher divides the test subjects into two groups, and different treatments are applied to each group.

Both experiments and observational studies are examples of valid statistical studies, but experiments are more useful in establishing cause-and-effect relationships than observational studies.