Statistics, Day 1

Objectives: To identify a statistical question

To recognize the roll of randomness in statistical studies

To distinguish between surveys, observational studies and experiments

Statistics is the study of data. It includes collecting data, organizing data, and drawing conclusions about data.

A statistical question is one for which you don't expect to get a single answer. Instead, you expect to get a variety of different answers, and you are interested in the distribution and tendency of those answers.

For example, "How tall are you?" is not a statistical question. But "How tall are the students in our school?" is a statistical question.

The data to answer a statistical question may be collected by a survey, observational study, or experiment. In a survey, the data is reported directly by the subjects. In an observational study, researchers observe and/or measure the data of interest. In an experiment, the researcher manipulates something in order to observe and measure a response.

Example: A student wants to know if the students at his high school prefer Coke or Pepsi. Describe how this data could be collected in a survey? An observational study?

An experiment?

What might be some of the advantages and disadvantages of each type of study?

The populationin a statistical study is the entire group that we want information about. A sample is the part of the population that we actually collect the information about. In order for a survey to be unbiased, the sample must be selected randomly.

For example, for the statistical question “How tall are the students in our school?" the population is all of the students in our school. It would take a long time to measure every student in the school, so we might choose to measure a smaller random sample of all the students.

Example: Which of the following methods could be used to produce a random sample of 30 students to survey?

1. Put all of the students names in a hat and randomly choose 30 names
2. Put all of the teachers names in a hat and randomly select a teacher and measure the students in her class
3. Label all of the students in the school with a number and use a calculator to generate 30 random numbers.
4. Pick the first 30 students on an alphabetical list of all students.
5. Go to the gym and choose 30 students who are practicing basketball.

A census is when everyone in the population is sampled. A census is frequently time-consuming and expensive. In the US, a census to count the population is required every 10 years.

Unlike a survey, the subjects of an experiment are usually not selected at random. Instead, for an experiment, the treatments must be assigned randomly.

Example: A biology teacher wants to know if his students learn as well watching a virtual lab as when doing an actual lab. He allows each student to decide which lab to participate in. Both the virtual lab and the actual lab take one class period, and the next day each student takes a test on what they have learned. The teacher finds that the students doing the actual lab do better on the test. Can he conclude that the virtual lab is not as effective as the real one? Explain.

Example: How could you improve the experiment above?

In general, surveys and observational studies can show association (or correlation) but only randomized experiments can prove causation.