Chap 7 Notes: Obtaining Data: Random Sampling and Randomized Experiments

J. Harner A. Billings

Department of Statistics
West Virginia University

Stat 211 Fall 2007

Outline

Chap.7 Notes: Obtaining Data:Random Sampling and Randomized Experiments

- Sect 7.1: Obtaining Data: Random Sampling
- Sec 7.2: The Randomized Controlled Experiment

Obtaining Data: Random sampling

Population:

The entire collection of objects or people that one desires to study, usually to make inference about the population, or to estimate a population parameter.

Sample:

A subset of the population.

- Two types of populations:
  1. **Real population:**
     Real population consists of a finite number of members (actual objects) from which data can be collected.

  Example:

  a. Population of all college students in US.

  b. Population of all legally registered automobiles.

  2. **Conceptual population:**
     Conceptual population exists only as a concept, and can be thought of as having an infinite number of members.

  Example:

  a. Tosses of a fair coin.

Population Parameter:

A number that describe some characteristics of a population,
e.g., μ: the theoretical mean of the population.

Example:

a. Mean age of the population of college students.
b. Mean value of an infinite number of tosses of a fair 6 - sided die.

Sample Statistics:

A number that describes some characteristics of a sample,
e.g., $\bar{X}$: the mean of the sample.

Example:

a. Mean age of a (random) sample of 1000 college students
b. Mean value of 500 tosses of a fair 6 - sided die.

- In order to draw conclusions about a population, or to estimate a population parameter, we need a sample that is representative of the population.
- Even though we do not have information about the entire population, the data contained in the sample can be used to make inferences about the entire population.

Example:

We can use the sample mean $\bar{X}$ (calculated from the random sample of 1000 college students) to estimate the (unknown) value of the mean age μ for the entire population of college students.

Random Sample:

A sample chosen from a population in such a fashion that each member of the population has an equal probability of being selected.

- Use of a randomization method to obtain a sample helps to ensure:
  1. **Validity** - Using appropriate statistical methods, information that is accurate and trustworthy can be extracted from the data.
  2. **Believability** - The fact that a valid statistical inference or estimate can be trusted or relied on.

(Contd.)

3. **Prediction of Accuracy** - An estimate of the likelihood that a statistical inference or estimate made from randomly obtained data is accurate or correct.

- When a randomization method is used in sampling from a population the sampling is called **Probability sampling**.

Randomized Controlled Experiment:

- A controlled experiment in which subjects are randomly assigned to the treatment and groups.

- The purpose is to estimate the effect of a treatment when it is applied to a subject, by comparing the outcomes produced by the treatment to the outcomes produced when no treatment (or a standard treatment) is used. The subjects that receive the treatment is called the **treatment group**.
  The subjects that receive no treatment (or a standard treatment) is called the **control group**.

- Before the experiment begins, we want the treatment group and the control group to be as similar as possible - so that any observed difference between the groups at the end of the experiment can be attributed solely to the treatment.

- To ensure that two groups are as similar as possible (before the experiment), randomly assign the subjects in the experiment to the control and treatment groups.

Example:

Pharmaceutical Study of effectiveness of new drug to control hypertension. Randomly assign half of subjects to treatment group (new drug) and the other half of subjects to control group (placebo).

Bias

Systematic error in the quantity being estimated, usually resulting from a poorly designed study or experiment.

Example:

Estimate frequency of college students' drinking by collecting data from members of the "Young Female Republicans" club.