Session 01

Data Collection Methods

To derive conclusions from data, we need to know how the data were collected; that is, we need to know the method(s) of data collection.

Methods of Data Collection

There are four main methods of data collection.

- **Census**. A census is a study that obtains data from every member of a population. In most studies, a census is not practical, because of the cost and/or time required.

- **Sample survey**. A sample survey is a study that obtains data from a subset of a population, in order to estimate population attributes.

- **Experiment**. An experiment is a controlled study in which the researcher attempts to understand cause-and-effect relationships. The study is "controlled" in the sense that the researcher controls (1) how subjects are assigned to groups and (2) which treatments each group   receives.

  In the analysis phase, the researcher compares group scores on some dependent variable. Based on the analysis, the researcher draws a conclusion about whether the treatment (independent variable) had a causal effect on the dependent variable.

- **Observational study**. Like experiments, observational studies attempt to understand cause-and-effect relationships. However, unlike experiments, the researcher is not able to control (1) how subjects are assigned to groups and/or (2) which treatments each group receives.

Data Collection Methods: Pros and Cons

Each method of data collection has advantages and disadvantages.

- **Resources**. When the population is large, a sample survey has a big resource advantage over a census. A well-designed sample survey can provide very precise estimates of population parameters - quicker, cheaper, and with less manpower than a census.

- **Generalizability**. Generalizability refers to the appropriateness of applying findings from a study to a larger population. Generalizability requires random selection. If participants in a study are randomly selected from a larger population, it is appropriate to generalize study results to the larger   population;   if   not,   it   is   not   appropriate   to   generalize.

  Observational studies do not feature random selection; so it is not appropriate to generalize from the results of an observational study to a larger population.

- **Causal inference**. Cause-and-effect relationships can be teased out when subjects are randomly assigned to groups. Therefore, experiments, which allow the researcher to control assignment of subjects to treatment groups, are the best method for investigating causal relationships.