

Section 1-3
Collecting Sample Data

DEFINITIONS

- **Observational Study:** observing and measuring specific characteristics without attempting to *modify* the subjects being studied
- **Experiment:** apply some *treatment* and then observe its effects on the subjects

TYPES OF OBSERVATIONAL STUDIES

- In a **cross-sectional study**, data are observed, measured, and collected at one time.
- In a **retrospective** (or **case-control**) **study**, data are collected from the past by going back in time (through examination of records, interviews, etc.).
- In a **prospective** (or **longitudinal** or **cohort**) **study**, data are collected in the future from groups (called *cohorts*) sharing common factors.

DESIGNING AN EXPERIMENT OR OBSERVATIONAL STUDY

1. Identify your objective
2. Collect sample data
3. Use a Random Procedure that avoids bias
4. Analyze the data and form conclusions

CONFOUNDING

Confounding occurs in an experiment when the experimenter is not able to distinguish between the effects of different factors.

Try to plan the experiment so that confounding does not occur.

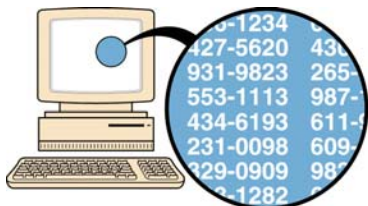
SAMPLING

Sampling refers to the method used to select the subjects being studied. The methods of sampling we will study are:

- Random and Simple Random
- Systematic
- Convenience
- Stratified
- Cluster

RANDOM SAMPLING

In a **random sample** members of the population are selected in such a way that each **individual member** has an equal chance of being selected.



SIMPLE RANDOM SAMPLING

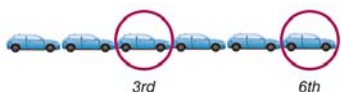
A **simple random sample** of size n is selected in such a way that every possible **sample of the same size n** has the same chance of being chosen.

EXAMPLE

Suppose there is a class with 5 men in one row and 5 women in the other. I can get a random sample of size 5 by flipping a coin. If heads, I select the row of women. If tails, I select the row of men. Each student has an equal chance of being selected, and the sample is clearly random. However, it is not an simple random sample of size 5 because not every group of 5 students is equally likely to be selected. With the above method, I cannot get 2 men and 3 women.

SYSTEMATIC SAMPLING

Systematic sampling: randomly select some starting point and then select every k^{th} element.



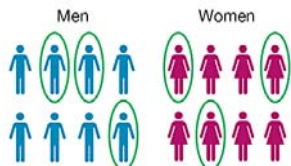
CONVENIENCE SAMPLING

Convenience Sampling: use results that are readily available.



STRATIFIED SAMPLING

Stratified Sampling: subdivide the population into subgroups that share the same characteristic, then draw a sample from each stratum.



CLUSTER SAMPLING

Cluster Sampling: subdivide the population into sections (clusters), randomly select some of the clusters, then choose **ALL** the members from the selected clusters.



ERROR

- A **sampling error (or random sampling error):** occurs when the sample as been selected with a random method, but there is a discrepancy between a sample and the true population result; such an error results from chance sample fluctuations.
- A **nonsampling error** is the result of human error, including such factors as wrong data entries, computing errors, questions with biased wording, false data provided by respondents, forming biased conclusions, or applying statistical methods that are not appropriate for the circumstances.
- A **nonrandom sampling error** is the result of using a sampling method that is not random, such as using a convenience sample or a voluntary response sample.
