Planning a study

→ Why is important to spend adequate time planning how to collect data for a study?

→ What are the six steps of the data analysis process?
  a.
  b.
  c.
  d.
  e.
  f.

Sampling

→ How can we ensure that our sample best represents the population we wish to characterize?

→ What happens if the sample does not represent the population?

  *Definition: Selection bias* is introduced when the way that the sample is obtained systematically excludes part of the population

→ Why is selection bias bad?

  *Definition: Measurement or response bias* occurs when the method of observation tends to produce values that systematically differ from the true value in some way

→ Give a few examples of measurement bias.

  *Definition: Nonresponse bias* occurs when responses are not actually obtained from all individuals selected for inclusion in the sample

→ In what settings is nonresponse bias common?
→ How can nonresponse bias be minimized?

→ What is the key issue in sampling?

→ What is chance error due to sampling?

→ Will increasing the sample size reduce the bias in the sample? Explain.

*Example:* Suppose a weight loss clinic is interested in studying the effects of a new diet proposed by one of its researchers. It decides to advertise in the LA Times for participants to come be part of the study.

→ Any sampling bias issues?

*Example:* Suppose a lake is to be studied for toxic emissions from a nearby power plant. The samples that were obtained came from the portion of the lake that was the closest possible location to the plant.

→ Any sampling bias issues?

*Definition:* A *simple random sample* is a sample chosen using a method that ensures that each different possible sample of the desired size has an equal chance of being chosen
Example: Consider our class as the population under study. If we select a sample of size 5, each possible sample of size 5 must have the same chance of being selected.

→ How could we randomly select five members from this class randomly?

→ What type of statistical method(s) could we use to achieve random selection?

→ What is sampling without replacement?

Table Method:

1. Randomly assign id’s to each member in the population (1 - n)
2. Choose a place to start in table (close eyes)
3. Start with the first number (must have the same number of digits as n), this is the first member of the sample.
4. Work left, right, up or down, just stay consistent.
5. Choose the next number (must have the same number of digits as n), this is the second member of the sample.
6. Repeat step 5 until all members are selected. If a number is repeated or not possible move to the next following your algorithm.

Computer Method (see instructions on my MTB web link):

1. Create a column of id’s (1 – n).
2. Randomly select the number of random id’s that you desire and store them in another column.

NOTE: This method can also be used for randomly selecting data from a population if the population data is in stored one column.

Other Sampling Methods

Definition: Stratified random sampling is used when the population can be divided into a set of non-overlapping subgroups, then members are sampled from each subgroup

→ Why would one want to use stratified sampling?
**Definition:** Cluster sampling involves dividing the population into non-overlapping subgroups called clusters, then clusters are selected at random (all individuals)

**Definition:** Systematic Sampling is a procedure that can be used when it is possible to view the population as a list or a sequential arrangement, then one of the first k individuals is sampled and from then on every kth individual is also sampled

**Definition:** Convenience sampling uses an easily available group to form a sample, such as volunteer based studies

→ Why is convenience sampling a bad idea?

**Statistical Studies**

→ Name the two types of statistical studies.

*Example:* Data is collected regarding the race and ethnicity of mothers who give birth to low weight babies compared to normal weight babies

→ What type of study is this an example of?

*Example:* A study is conducted to examine the hair growth (response variable) two different doses (factor) of a new hair growth cream

→ What type of study is this an example of?

→ Which type of study should not draw conclusions about cause and effect?

→ In the previous race/ethnicity and birth weight example could researcher claim that race/ethnicity causes low birth weight babies if the data supported this claim? Why not?
Definition: A confounding variable is one that is related to both group membership and the response variable of interest in the research study

→ Why are confounding variables important to be ruled out?

→ What if the relationship between race/ethnicity and birth weight were really something else?

Definition: An extraneous factor is one that is not of interest in the current study but is thought to affect the response variable

Example: A study was conducted to see if growth of corn plants would be affected by fertilizer: A, B, A+B, or none (control group). The field where the corn is grown is divided into 4 plots. Along the south end of the field is a toxic sewage pipe that runs parallel to the field.

→ Is there an extraneous variable here?

Definition: Blocking creates groups that are as homogenous as possible within the group

→ Which design effectively blocks the field so that the location to the sewage pipe will not cause extraneous variability?

→ Why is your design choice above more effective than the other choice?
Careful! How are we assigning the four treatment groups to the corn within the blocks?

There are four key concepts in experimental design

- **Randomization**, ensures that the experiment does not favor one experimental condition over another
- **Blocking**, use the extraneous factor to create groups that are similar and then use all experimental conditions within each block
- **Direct control**, holds extraneous factors constant so that their effects are not confounded with those of the experimental conditions
- **Replication**, ensures that there is an adequate number of observations for each experimental condition

What are the goals of an experimental design?

What is a control group?

Why is using a control group sometimes important in making a comparison to a treatment group?

*Definition*: A placebo is something that is identical to the treatment received by the treatment group, except it contains no active ingredients

Why is using a placebo group sometimes important in an experiment?
**Definition:** Blinding (aka. masking) is used to ensure that one does not know what treatment is being given, and therefore cannot pass personal judgment as to it’s effectiveness

- **Single blind** refers to an experiment where the participants OR the researchers in a study do not know which treatment was received
- **Double blind** refers to an experiment where the participants and the researchers in a study do not know which treatment was received

Other definitions you should know (read book):

- Census
- Factor
- Treatment
- Confounded
- Undercoverage
- Overcoverage