**Math 200**

Mod 4-5 study guide

**Collecting data** (Mod 4):

* Sampling**:** 
  + **Population** and **sampling frame**
  + **SRS**- Simple Random Sampling provides every member of a population and equal chance of being selected thus creating a representative sample.
    - One of the important – and counterintuitive – consequences of random sampling is that the size of the population doesn’t significantly affect the sample size needed. A random sample of 1,000 will provide roughly the same accuracy for a population of 10,000 and one of 10,000,000!
  + **Systematic**
  + **Cluster**
  + **Stratified** – why might we might prefer stratified over SRS?
  + (Avoid) **Convenience** sampling.
  + **Bias** (sampling that systematically favors a particular group or outcome).

The purpose of **random sampling (selection**) is to produce a representative sample which allows us to generalize results of a study from the sample to the general population from which it was taken.

**Types of studies (Mod 5):**

* + **Experiment**:
    - Manipulates explanatory var., tests cause and effect vs observational: studies properties of population (summary OLI pg 25).
    - Elements of experimental design:
      * **Treatment groups**
      * **(possibly) Control group**
      * **Random selection**
      * **Random assignment**
      * **Blinding**
      * **Direct control** (making things the same for all participants – e.g. listening to the same music for the same length of time or washing hands with the same amount of soap.)
      * **Placebo**

**Random assignment** creates similar groups and mitigates the effects of confounding variables.

**Direct control** allows experimenters to eliminate or reduce the effects of confounding variables by making the actions or uniform across groups.

* Variables associated with studies:
  + **Explanatory**
  + **Response**
  + **Confounding/Lurking**

Experiments are the only consistent way to demonstrate cause and effect. Observational studies can be used to show this but the burden of eliminating confounding variables is often too costly or impractical to overcome.

*Again* . . .

1. When comparing two groups, the size of one group does not have to match the other. It’s desirable to have the same sizes but not essential.
2. Sample size is largely independent of population size. You don’t need to have a large sample in order to make inferences about a large population. The larger the sample, typically the more refined the estimate it provides but a suitably *random* sample of size 1,000 will do a good job of estimating population parameters for populations that are 10,000 as well as 10,000,000.