Other Sampling Designs

AP Statistics
Probability Sampling

- Uses random selection and gives each member of the population a known chance (greater than zero) of being selected.
- This the general framework for all defensible sampling. Ex. SRS
Stratified Random Sample

- A probability sampling method.
- More accurate than SRS.
- Divide the population into subgroups of similar individuals, called *strata*. Then choose a separate SRS within each stratum and combine the SRS's into a full sample.
Examples

- Ex. Freshmen, Sophomores, Juniors and Seniors at CRHS (*proportionate* - *similar fractions*).
- Ex. White, Black, Asian, Hispanic (*disproportionate* - *different fractions*).
Systematic Random Sampling

- Also a probability sampling method.
- May be more precise than SRS. It is assumed the units in the population are randomly ordered.
- Select a random starting point and select every nth individual.
Systematic Random Sample

- **Advantage**: Easy. Only have to randomly select one number to determine where to start.
- **Disadvantage**: Not to be used with a population that is repetitive or cyclical in nature.
Multi-stage Sampling Design

- Also a probability sampling method.
- Choose the sample in stages starting with larger groupings and working down to smaller groupings. Used when populations are large and widely spread out.
Multi-stage Sampling Design

- **Ex. National Multi-stage Sample:**
  - **Stage 1** - SRS from 3000 US counties
  - **Stage 2** - SRS of townships within each county chosen
  - **Stage 3** - SRS of blocks within each township chosen
  - **Stage 4** - SRS of households within each block
Multi-stage Sampling Design

Advantages:
- Don't need a national list of households.
- Lists of counties and townships are easily available.
- Sample of households are clustered and easy to reach by an interviewer.
Cluster Sample

- Population is broken down into subgroups that are small images of the target population, then an SRS is taken from each subgroup.
- Used when the population is dispersed over a wide geographic region.
- Divide the population into clusters (usually along geographic boundaries).
- Randomly sample clusters.
- Measure all units within sampled clusters.
Sources of Bias in Sampling

- **Undercoverage** - occurs when groups in the process are left out (census and homeless).

- **Non-response** - an individual chosen for the sample can't be contacted or refuses to answer.

- **Response Bias** - for whatever reason the subject doesn't tell the truth. (Illegal or unpopular behavior, faulty memory about past events).

- **Wording effects** - questions shouldn't be confusing or leading.

- **Sampling Bias** – bias which is a result of poor sampling techniques.

Cautions about Sampling

- Inference about a population - No two samples are likely to give the same response. Need to give a margin of error.
- *Larger samples are typically more reliable (if bias is eliminated).
Homework

• Textbook: 5.20, 5.21, 5.24-5.26, 5.28, 5.30