Section 4.1 (Part 2) pp. 224-229

1. Inference for Sampling - The purpose of a sample is to give us information about a larger population. This process is called *inference* because we *infer* information about the population from what we *know* about the sample.

Inference from convenience samples or voluntary response samples would be misleading because these methods are biased. They most likely are not representative of the population of interest.

The first reason to rely on *random sampling* is to eliminate bias in selecting samples from the list of all available individuals. Even if we do this, it is unlikely that the results of a random sample are exactly the same as the entire population. Properly designed samples avoid systematic bias but their results are rarely exactly correct and we expect results to vary from sample to sample.

Let's think about the **Hyena Lab** we conducted.

- How did you sample the population?
- Did each sample yield the same results?
- How much did the proportions vary from sample to sample?

What we observed in the Hyena Lab was the fact that the results of random sampling do not change haphazardly from sample to sample. Because we are using *chance*, the results obey the laws of *probability* that govern chance behavior. **The second reason** to use *random sampling* is that the laws of probability allow trustworthy inference about the population.

It should also be noted at this point that *larger random samples* give better information about the population than smaller samples.

Definition: A **sampling frame** is the list of individuals from which a sample is drawn.

2. Sample Surveys: What Can Go Wrong?

- Random sampling eliminates bias in choosing a sample.
- Even a large sample will give a result that differs from the truth about the population.
- There will be a *"sampling variability"* that is described by the margin of error that comes with most poll results.
- Good sampling technique includes the art of reducing all sources of error.
- There are two main sources of errors in sample surveys: *sampling errors* and *nonsampling errors.*

Sampling Errors - The margin of error tells us how much sampling variability to expect (based upon probability laws) and we can control it by choosing the size of our random sample. It does not tell us about sampling errors -- mistakes made in the process of taking a sample that could lead to false information about the population.

- **Bad sampling methods** - voluntary response samples, convenience samples
- **Undercoverage** - when some groups of the population are left out of the process of choosing the sample.

  Example of Undercoverage: collecting data for a high school study but neglecting to allow freshman students to participate.

Nonsampling Errors – Have nothing to do with choosing a sample; Nonsampling errors are those that can plague even a census.

- **Nonresponse** - Nonresponse occurs when an individual chosen for the sample cannot be contacted or refuses to participate.
  - Voluntary response sample versus nonresponse; These are not the same thing! Nonresponse can occur after a sample has been selected. In a voluntary response sample, every individual has opted to take part, so there won't be any nonresponse.

- **Response Bias (Incorrect Response)** - Often times respondents "remember things that never happened" or answer questions in a manner that is politically correct. Good interviewing technique or survey implementation can help to reduce this.
    Example of Response Bias (Incorrect Response): People may be embarrassed by question being asked, so people may lie about age, income, drug use when responding to survey; may be influenced by race or gender of interviewer when responding.

  - The **wording of questions** is the most important influence on the answers given to a sample survey.

Example: How Do Americans Feel about Illegal Immigrants? Question Wording Matters
"Should illegal immigrants be prosecuted and deported for being in the U.S. illegally, or shouldn't they?" Asked this question in an opinion poll, 69% favored deportation. But when the very same sample was asked whether illegal immigrants who have worked in the United States for two years "should be given a chance to keep their jobs and eventually apply for legal status," 62% said that they should. **Different questions give quite different impressions of attitudes toward illegal immigrants.**

Check Your Understanding
1. Each of the following is a possible source of error in a sample survey. Label each as sampling error or nonsampling error and explain your answers.

a) The telephone directory is used as a sampling frame.

b) The person cannot be contacted in five calls.

c) Interviewers choose people walking by on the sidewalk to interview.

2. A survey paid for by makers of disposable diapers found that 84% of the sample opposed banning disposable diapers. Here is the actual question:
> **It is estimated that disposable diapers account for less than 2% of the trash in today's landfills. In contrast, beverage containers, third-class mail, and yard wastes are estimated to account for about 21% of the trash in landfills. Given this, in your opinion would it be fair to ban disposable diapers?**

Explain how the wording of the questions could result in bias. Be sure to specify the direction of the bias.