**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 on the first day of the semester.

* 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.*

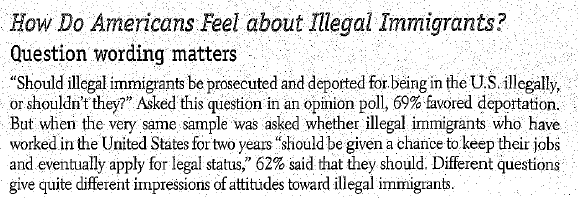
**Sampling Errors** - The margin of error tells us how much sampling variability to expect (based upon probability laws) and we can control it by choosing the size of our random sample. It does not tell us about sampling errors -- mistakes made in the process of taking a sample that could lead to false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

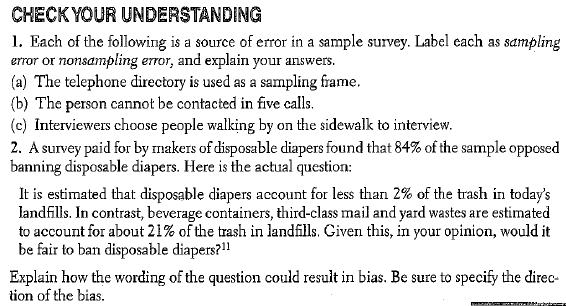
* **Bad sampling methods** - voluntary response samples, convenience samples
* **Undercoverage** - when some groups of the population are left out of the process of choosing the sample.

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

**Example:**





HW: 23-35 odd