1. Understanding Chapter 11

A mathematician once wrote "The roulette wheel has neither conscience nor memory."

We enjoy playing games that we perceive as FAIR (as opposed to unfair or biased) and randomness or random outcomes seem fair because no one can guess the outcome before it happens and frequently (but not always) outcomes are equally likely.

Random outcomes actually have a structure that we can study and understand. Randomness doesn't mean careless or totally unpredictable and chaotic, in Statistics, randomness implies that in the "short run" it's not possible to guess an outcome, but in the long run, there is a pattern that can be understood.

In statistics, your textbook is clear that the value of randomness can't be overstated – if you want to collect data that presents a fair and accurate picture of the world, you must have randomness and if you want to generalize from your specific data to the larger population, your ability to generalize will depend on randomness.

2. Randomness and Sample Surveys Chapter 12

In an ideal world, we would study populations and you wouldn't need this statistics class. Studying populations is not practical because they are too large (so it costs a lot of time & money.) So we settle for samples.

The **POPULATION** is the entire set of people (or animals, things) we wish to study.

A **SAMPLE** is a part of the population.

A numerical fact about a sample is a **STATISTIC**.

A numerical fact about a population is a **PARAMETER**.

California October 7th Recall results (these can be taken as parameters)

Recall Davis: 55% in favor -44% opposed
Replacement candidates: Arnold 48.7% Cruz 31.7% all others 19.6%

In the days before the actual election (see handout) many SAMPLE SURVEYS (or polls) were being conducted to predict what would happen on October 7th. Sample surveys are designed to ask questions of a much smaller group of people (in this case voters) in order to learn something about the population. The numbers from the sample surveys are statistics.

A little history (p. 223 of your text)

3. BIAS

If a sample is "representative", then a statistic can be a good estimate of the parameter; but if the sample includes or excludes certain people systematically, the sample is BIASED. See examples of non-random samples…results from cnn.com and other places…

4. The solution

RANDOMNESS is a statistician's friend. The best way to insure representativeness (and protection from bias) in a sample is to make sure that the selection process is RANDOM. The process is called

RANDOMIZATION and it effective "mixes" the population such that a sample selected from the population will reflect all the features of your population.

Randomness serves another purpose – it assists us in making generalizations about the population.

5. SAMPLE SIZE

One question that arises is "how large of a sample do we need to insure unbiasedness?" What is interesting is that "size matters" but fraction does not. We will return to this in later chapters.

6. SAMPLE DESIGNS

a. Simple random sample (SRS): every object in the population has an equal chance of getting into the sample with each draw. In practice this is drawing at random without replacement (because it would not make sense to select the same person or measure the same animal/thing twice).

b. Not every sample is a simple random sample; other sampling schemes include CLUSTER SAMPLING (selecting groups and then randomly sampling within groups) and STRATIFIED RANDOM SAMPLING (stratifying sample on the basis of a characteristic, like gender – forcing a sample to be 50%-50% male-female)

c. MULTISTAGE SAMPLES – in practice, survey organizations use a combination of simple random samples, cluster samples, and stratified samples.

Note: these methods can be applied to things other than households. Examples might be estimating the corn harvest, sampling firms on hiring expectations, etc.

7. What can go wrong?

A. Voluntary Response Bias --- certain types of people choose to respond to a survey
B. Convenience Sampling – when a researcher/pollster just includes people or things just because they are nearby or easy to talk to.
C. Bad sampling frame – when a researcher/pollster has an incomplete list of the population
D. Non-response bias --- people selected for the survey don't bother to answer you
E. Response bias --- people selected for the survey answer, but they lie to you or they are manipulated by the way you asked the question
F. Wording of question --- phrasing may not be neutral (e.g. a loaded question).