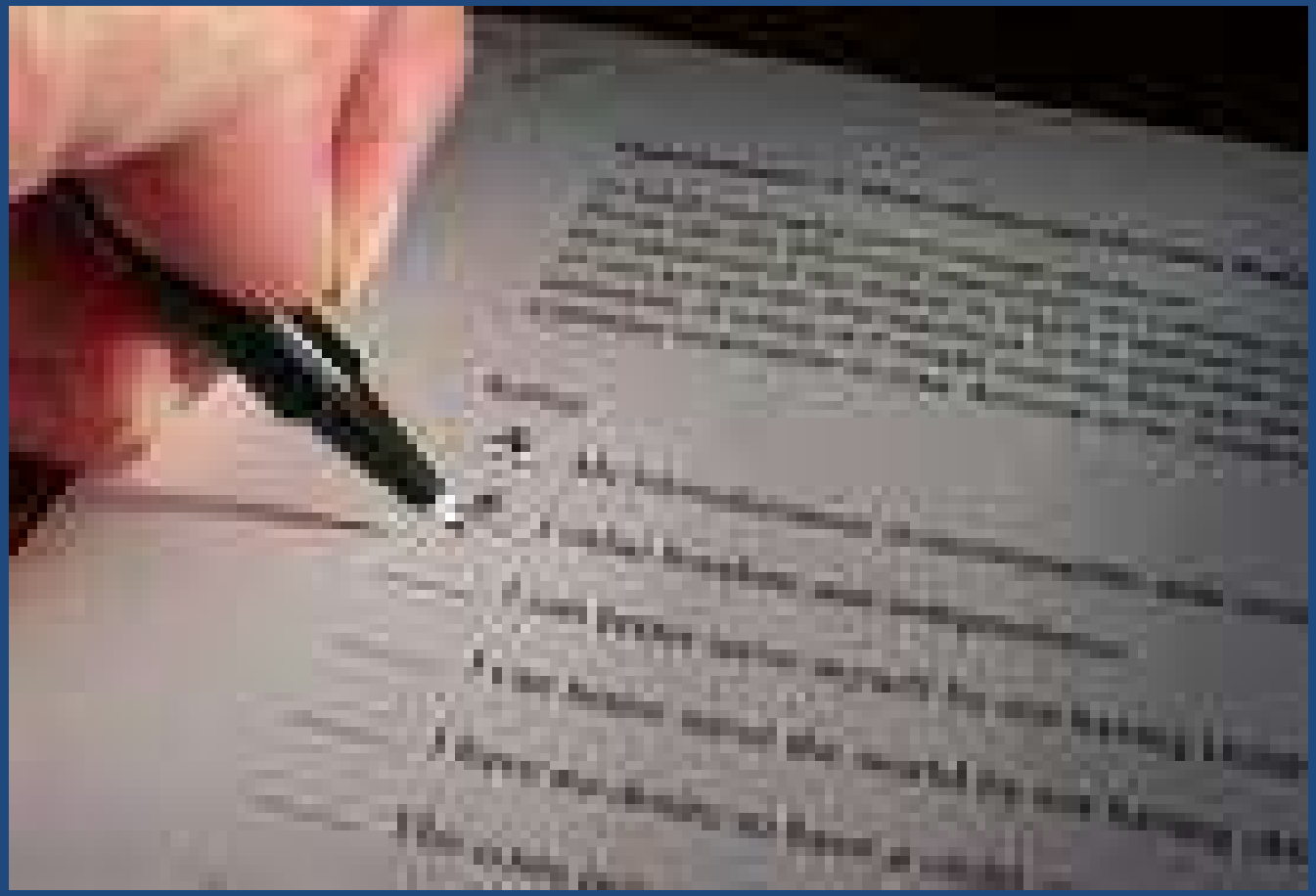
Part III – Gathering Data

Ch. 12 – Sample Surveys

(Designed to ask questions of a small group of people in the hope of learning something about the entire population)



Sampling Techniques

- Census-Include everyone and "sample" the entire population
- Because it is usually not practical to work with an entire population because of time and money, we have to use samples
- Sample-Smaller group of individuals selected from the population

Sampling Techniques

- We want the statistics we compute from the samples to reflect the corresponding parameters accurately
- A sample that does this is said to be representative
- Today we will look at several different techniques used to create representative samples

Simple Random Samples

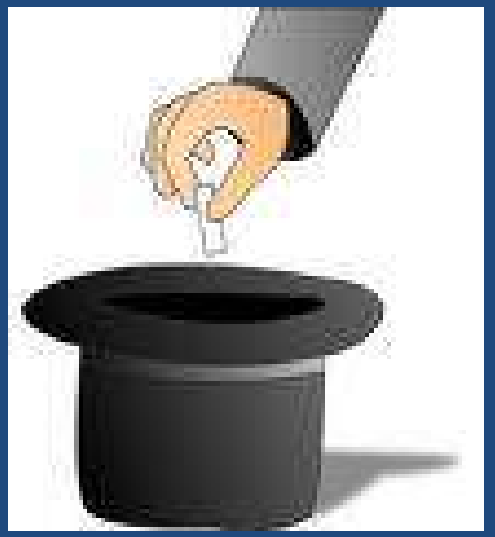
- The most basic way to create a representative sample is by using a simple random sample (SRS)
- In this method, every possible combination of individuals has an equal chance to be selected
- We do this by selecting members of our population at random to be in our sample
- Remember that “at random” doesn’t mean haphazardly – there is structure to a random selection!

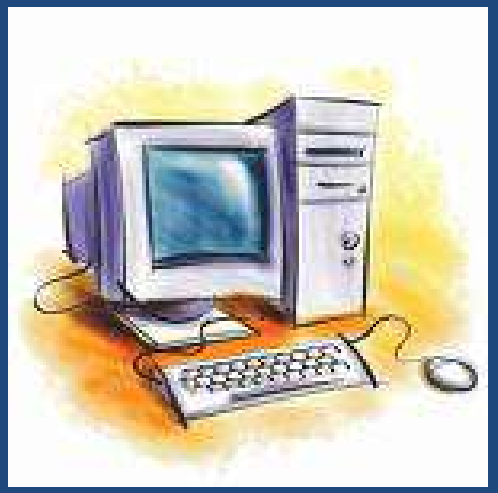
Drawing a Simple Random Sample

- In order to select a sample, we must first decide where the sample will come from
- Remember that our population is the entire group of individuals we want to know about
- The sampling frame is a list of individuals from which the sample will be drawn
  - EX: To draw a random sample of students at a college, we might obtain a list of all registered full time students and sample from that list.
- Ideally, we want our sampling frame to match our population perfectly, but in real life this is not always possible

Drawing a Simple Random Sample

- Once we have our sampling frame, we need to choose the individuals for our sample
- Some ways to do this:
  - Random numbers (from a table or a random number generator)





Drawing a Simple Random Sample

- To select an SRS using random numbers:
  - Label every individual in your sampling frame with a number
  - Select random numbers from a table or random number generator until you have enough individuals for your sample
  - Can we repeat numbers?
    - NO

Example

- Let's draw a simple random sample of 5 students from this class using our calculators
- There are 80 students enrolled in an introductory Statistics class, you are to select a sample of 5. How can you select an SRS of 5 students using these random digits found on the Internet: 05166 29305 77482

Sampling Variability

- Since each random sample will select a different specific set of individuals, they will usually provide us with somewhat different values for the variables we are trying to measure
- These sample-to-sample differences are called sampling variability
- Unless each individual in a population is exactly the same, we can't avoid sampling variability

Sampling Variability

- We can learn about our population by looking at the variability of our samples
  - If the samples show a lot of variability, there is probably a lot of variability in the population
  - If the sample values are all about the same, there is probably not much variability in the population
- Other factors contribute to sampling variability:
  - Sample size
  - Sampling technique

Sample Size

- How big of a sample do you need?
- Depends on what you're estimating.
- You need a large enough sample to get a representative sample
- For a survey that tries to find the proportion of the population falling into a category, you'll usually need several hundred respondents to say anything precise enough to be useful.

Stratified Sampling

- Sometimes alternate methods of sampling are used, either to improve our results or to make our samples easier to collect
- All statistically valid sampling designs have random chance, rather than human choice, as their basis
- When a population contains identifiable groups that are different from each other, stratified sampling is often used
- The goal of this method is to make sure our sample is more likely to be representative of the population and to reduce sampling variability

Stratified Sampling

- In a stratified sample, the population is first divided into *homogeneous* groups called strata, and then a simple random sample is taken from each stratum (group), and the results are combined
- The size of the samples collected in each stratum is usually proportional to the way the group appears in the population
- Examples:
  - Poll of the student body stratified by grade level
  - Poll for funding of football team at University stratified by gender

Cluster Sampling

- The goal of stratified sampling is to improve the results our sampling method produces
- Several other methods may be employed to make the sampling process easier or more practical
- One such method is cluster sampling, which divides the population into smaller existing groups, selects a few of these groups at random for our sample, and then takes a census within each group
- Examples:
  - Classes in a school

Cluster v. Stratified Sampling

- Remember that these two methods have different goals: accuracy (stratified) and efficiency (cluster)
- In a stratified sample, the groups are homogeneous (the individuals within each group are similar, and the groups are different from each other)
- In a cluster sample, the groups are usually similar to each other, each representing the variety of individuals seen in the population
- Stratified samples take an SRS of each group
- Cluster samples randomly select groups and take a census of everyone in group.

Example

- You're trying to find out what freshmen think of the food served on campus. Food services believes that men and women typically have different opinions about the importance of the salad bar. How should you adjust your sampling strategy to allow for this difference?
- I will stratify my sample by drawing an SRS of men and a separate SRS of women

Example

- In trying to find out what freshmen think about the food served on campus, you've considered both an SRS and a stratified sample. You've run into a problem: It's simply too difficult and time consuming to track down the individuals whose names were chosen for your sample. Fortunately freshmen at your school are all housed in 10 freshmen dorms. How could you use this fact to draw a cluster sample? How might that alleviate the problem? What concerns do you have?

Example

- To draw a cluster sample, I would select one or two dorms at random and then try to contact everyone in each selected dorm. I could save time by simply knocking on doors on a given evening and interviewing people. I'd have to assume freshmen were assigned to dorms pretty much at random and that the people I'm able to contact are representative of everyone in the dorm.

Multistage Sampling

- Sometimes two or more sampling methods are combined, called multistage sampling
- For example, to conduct an opinion poll we could stratify by geography to select areas to survey, do a cluster sample to select particular streets, and then conduct a simple random sample within each cluster to choose a few houses on each street

Systematic Samples

- Systematic samples are often used for large populations where choosing an SRS is time consuming, and no natural clusters are evident
- In this type of sample, we choose every kth (10^{th}, 15^{th}, etc.) individual from our sampling frame
- For example, if I want to randomly select 5 staffing agencies from the 100 which are listed in the phone book, I could choose every 20^{th} listing (100/5 = 20)
- Remember that for my method to be valid, I still need to include random chance, so I would use a random number to choose my starting place from the first k (in this case the first 20) individuals

Summary: How to Sample

1) Plan: State what you are trying to find out
2) Population and parameter: Who do you want to know about and what exactly will you be measuring?
3) Sampling plan: Choose the best method for your goal, your population and your resources

- What will your sampling frame be?
- What will your sample size be?
- How will individuals be selected?
- Don't forget to include randomization!

4) Carry out the sample, and look at the results
5) Draw a conclusion

- What can your sample tell you about your population?
- Were there any problems in your sampling process which may have affected your results?

Example

- Conduct your own sample survey to find out how many hours per week students at your school spend watching TV during the school year. How would you design this survey?

Example

- Plan: I want to design a study to find out how many hours of TV students at my school watch
- Population and Parameter: The population studied was students at our school. I obtained a list of all students currently enrolled and used it as the sampling frame. The parameter of interest was the number of TV hours watched per week during the school year, which I attempted to measure by asking students how much TV they watched during the previous week.

- Sampling Plan: I selected a SRS of students from the list. I obtained an alphabetical list of students, assigned each a random digit between 0 and 9 and then selected all students who were assigned a “4” This method generated a sample of 212 students from the population of 2133 students.

- Sampling Practice: This survey was taken over the period Oct 15 to Oct 25th. Surveys were sent to selected students by email with the request that they respond by email as well. Students who could not be reached by email were handed the survey in person.

- Summary and conclusion: of the 212 students surveyed, 110 responded. Its possible that the non-respondents differ in the number of tv hours watched from those who responded. The 110 that responded reported an average of 3.62 hours of tv watching per week. The median was only 2 hours per week.

Homework 12-1

- Pg. 288 # 1, 2, 4, 9, 10, 21
- For 9 and 10 answer parts (a)-(e)
- Do not answer part (f)