Section 1.2

Sampling

Learning Objectives

At the end of this lecture, the student should be able to:

- Define “sampling frame” and “sampling error”.
- Give one example of how to do simple random sampling, and one example of how to do systematic sampling.
- Explain one reason to choose stratified sampling over other approaches.
- State two differences between cluster sampling and convenience sampling.
- Give an example of a national survey that uses multi-stage sampling

Outline

- Sampling Definitions
- Simple Random Sampling
- Stratified Sampling
- Systematic Sampling
- Convenience & Multi-stage Sampling

Sampling Definitions

Terms you Need to Know

Concepts in Sampling

- What is a “sample”?
- Sampling frames, and errors in representing sampling frames
- Summary of definitions presented

Photo by Intel Free Press

Sampling and Samples

Photo by Jayal Aheram

- We take a sample of the population because we want to do “inferential statistics”
  - We want to infer from the sample to the population
- Reasons not to measure the whole population
  - Impractical
  - Unnecessary

Sampling Frame

- List of individuals from which a sample is actually selected.
- “List” may be a physical, concrete list
  - List of students enrolled at a nursing college
- May be a theoretical list not made up yet
  - List of patients who will present to the Emergency Department today

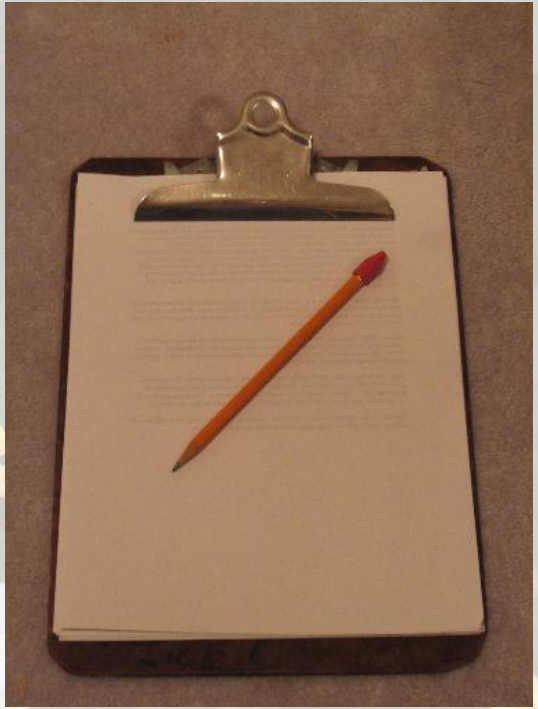


Photo by Schzmo

Sampling Frame

Sampling frame is the part of the population from which you want to draw a sample.

Therefore, you want everyone from your sampling frame to have a chance of being selected for your sample.

Undercoverage

What is it?

- Omitting population members from the sampling frame

How can it happen?

- List of nursing students may not include everyone for administrative reasons
- People who present to the Emergency Department at night might be different than those in the day

Errors in Statistics

Fact-of-Life Error

- Sampling error
  - The population mean will probably be different from your sample mean
  - The population percentage will probably be different from your sample percentage

Errors in Statistics

Fact-of-Life Error

- Sampling error
  - The population mean will probably be different from your sample mean
  - The population percentage will probably be different from your sample percentage

Error you want to Avoid

- Non-sampling error
  - Using a bad list.
  - Make sure that you pay careful attention that everyone in the population who is supposed to be represented in your sampling frame is in there!

Causes of Error

Fact-of-Life Error

- Sampling error – caused by the fact that, regardless of what you do, your sample will not perfectly represent the population.

Causes of Error

Fact-of-Life Error

- Sampling error – caused by the fact that, regardless of what you do, your sample will not perfectly represent the population.

Error you want to Avoid

- Non-sampling error – caused by poor sample design, sloppy data collection, inaccurate measurement instruments, bias in data collection, other problems introduced by the researcher.

Simulation

- A simulation is defined as a “numerical facsimile or representation of a real-world phenomenon.”
- It is a essentially working through a pretend situation to see how it would come out in the case it was real.
- That is why this course includes many simulations, or real-life examples.

Concepts in Sampling

- It is important to do your best to avoid non-sampling error
- This is achieved by making sure you do not have undercoverage when sampling from your sampling frame.

Simple Random Sampling

What it means

Simple Random Sampling

- What is simple random sampling?
- Two methods of randomly sampling from a list
- Limits of simple random sampling

Photo courtesy of Felix O

Definition & Example

Definition

- "A simple random sample of *n* measurements from a population is a subset of the population selected in such a manner that every sample of size *n* from the population has an equal chance of being selected."

Definition & Example

Definition

- "A simple random sample of *n* measurements from a population is a subset of the population selected in such a manner that every sample of size *n* from the population has an equal chance of being selected."

Example

- You have a list of the *population* of students in a class.
- You want to take a *sample* of 5 (n=5).
- If you take a simple random sample (SRS) from the class list, it means all the different possible groups of 5 students you could pick from the list has an equal chance of being the sample (group) you actually pick.

One Method of SRS

- Number all of the individuals in the population with a unique number.
  - Like student ID number
- Put all the student ID numbers in a place from which you can draw randomly without looking (like a hat)
- Draw 5 ID's and use those students as your sample.

Photograph by US Army

Another Method of SRS

Photograph by US Navy

- Generate a list of random numbers as long as the list of the population.
- Randomly assign these numbers to the population in the list.
- Take the first 5 numbers (whoever gets assigned 1 through 5).

SRS Means Equal Chance of Being Selected

Photograph by Joe Mabel

- *First method:* old-fashioned “hat”
- *Second method:* Electronic “hat”
- *In both methods*, all members of the population had an equal probability of being selected into the sample

Limits of Simple Random Sampling

- You need a list
  - You don’t know who will present at the Emergency Department that day, how do you sample?
  - Okay when a list is available.

Limits of Simple Random Sampling

- You need a list
  - You don’t know who will present at the Emergency Department that day, how do you sample?
  - Okay when a list is available.
- You need a *good* list
  - Otherwise, you risk undercoverage
  - What if part-time students were not on the list?
  - Non-sampling error

Simple Random Sampling

- Characteristics of SRS
- Two methods of randomly sampling from a list
- Limits of SRS

Photograph by ShakataGaNai

Stratified Sampling

What it means

Stratified Sampling

- What is stratified sampling?
- Steps in stratified sampling
- Examples in stratified sampling
- Limitations of stratified sampling

Photo courtesy of Angie of Sawara, Chia-ken, Japan.

What is Stratified Sampling?

- First, the list is divided into groups, or strata.
- This is a way to make it so that there are certain proportions of groups in the final sample.
- Next, simple random sampling (SRS) takes place for each of the strata

Photograph by Wonderlane

Steps in Stratified Sampling

1. Divide entire population into distinct subgroups called strata.
2. The strata are based on a specific characteristic, such as age, income, education level, and so on.
3. All members of a stratum share this specific characteristic.
4. Draw an SRS from each stratum.

Examples of Stratified Sampling

- In a high school, sampling so many students from each of the grades (freshman, sophomore, junior, senior)
- In hospitals, sampling so many patients or providers from departments (different intensive care units)

Photo courtesy of Undergrounded199.

Limitations of Stratified Sampling

- Oversampling one group means your summary statistic is unbalanced

Limitations of Stratified Sampling

- Oversampling one group means your summary statistic is unbalanced
- It is not possible to do without a list beforehand (like with SRS)

Limitations of Stratified Sampling

- Oversampling one group means your summary statistic is unbalanced
- It is not possible to do without a list beforehand (like with SRS)
- It also is hard because you have to split the list into groups (“strata”) then SRS from the strata

Stratified Sampling

Photograph by Sciondriver

- Stratified = taken from groups.
- Several steps are involved.
- Useful if necessary to make all strata equal, or to sample from groups that are small in the large population

Systematic Sampling

What it means

Systematic Sampling

- Systematic sampling can be done with or without a list!
- Systematic sampling is best described through the steps one takes to do it

Photo courtesy of Immanuel Giel.

Steps in Systematic Sampling

1. Arrange all individuals of the population in a particular order.
2. Pick a random individual as a start.
3. Then take every *kth* member of the population in the sample.
   - "*kth*" means "every so many".

Examples of Systematic Sampling from a List

- Take out a list of classes available next semester.
- Pick a random number that is small – like 3. Go to the third class.
- Pick another random number – like 5. Pick every 5th class after that.

Photo courtesy of Hobvias Sudoneighm

Characteristics of Systematic Sampling

- You *cannot* do this when there is a pattern to the data (boy/girl/boy/girl)

Characteristics of Systematic Sampling

- You *cannot* do this when there is a pattern to the data (boy/girl/boy/girl)
- You *can* do systematic sampling in a clinical setting, where you do not know who is going to come in that day

Photos courtesy of Nils Simon and US Army.

Systematic Sampling

Photograph by US Navy

- Systematic sampling is easy to do with or without a list.
- Just pick a random starting point, then pick every *k*th individual.

Cluster Sampling

What it means

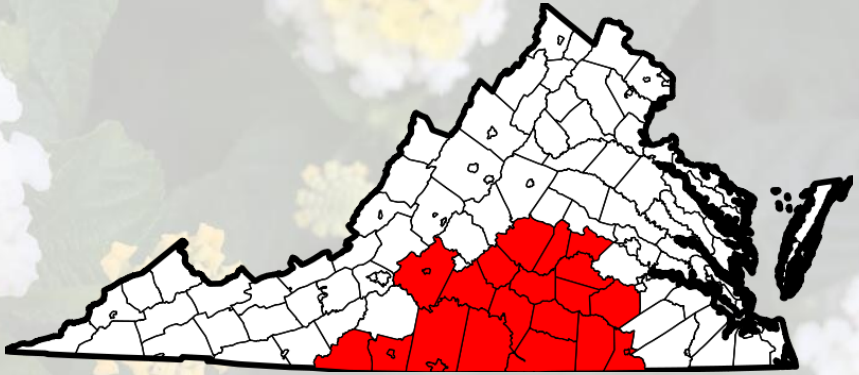
Cluster Sampling

- Why use cluster sampling when you could use stratified, systematic or simple random sampling?
- Because the problem is in a particular geographic location

Photo by Simisa.

Why use Cluster Sampling?

- The problem is localized to a particular location
- In cluster sampling, we begin by dividing the map in geographic areas.
- Then we randomly pick clusters, or areas, from the map. We take all the people in the cluster.



Map by Dbenbenn and Makeemlighter.

Problems with Cluster Sampling

- Sometimes, the people located in a cluster are all similar in a way that makes the problem hard to study.

Photo courtesy of Hansueli Krapf

Problems with Cluster Sampling

- Sometimes, the people located in a cluster are all similar in a way that makes the problem hard to study.
- If cancer rates are high all over the clusters, it's hard to see if a geographic location is causing higher rates.

Photo courtesy of Hansueli Krapf

Cluster Sampling

Photograph by Born Londoner

- Cluster sampling is used when geography is important in sampling.
- The map is divided into areas, and all the people in a particular area are sampled.
- Biased toward type of people living in the area.

Convenience & Multi-Stage Sampling

What it means

Convenience Sampling

- Convenience sampling can be used under low risk circumstances
- What ice cream is the best from the restaurant next to the hospital?
- However, often results are not reliable

Photo by Managementboy.

What is Convenience Sampling?

- Using results or data that are conveniently or readily obtained.
- Can be useful if not a lot of resources allocated to the study.
- Use an already-assembled group for surveys.
- Ask patients in the waiting room to fill out a survey, or students in a class.

What are the Problems with Convenience Sampling?

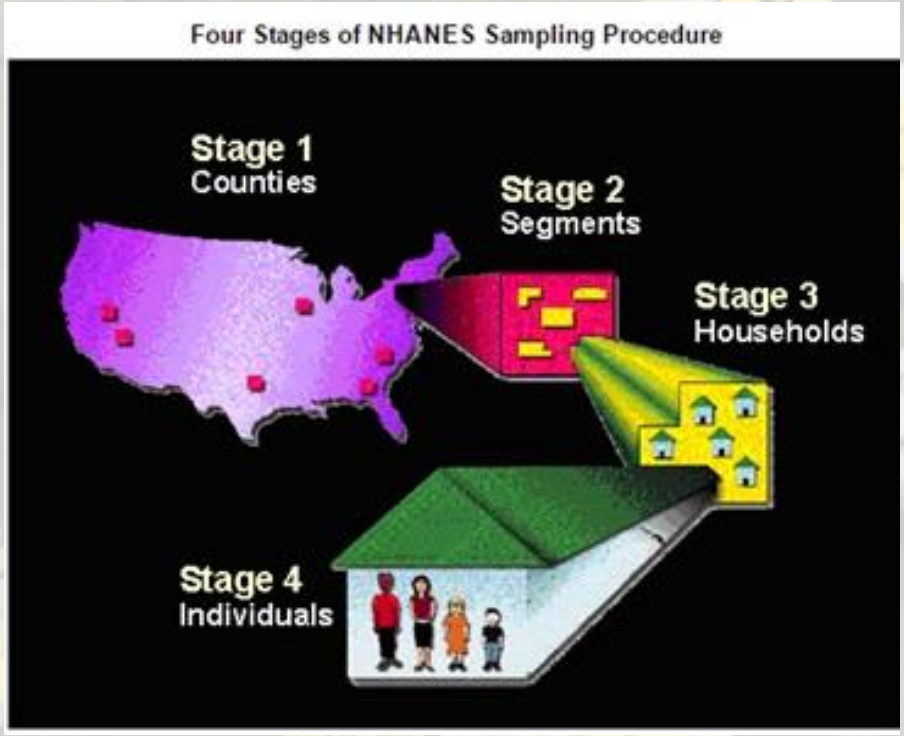
- There is a bias in every group.
- Often miss important subpopulations (what stratified sampling addresses).
- Results can be severely biased

Photo by Fanch Gaume.

Multi-stage Sampling

- Combination of sampling strategies layered in stages.
- Example:
  - Stage 1: Cluster sample of states (two census regions)
  - Stage 2: Simple random sample of counties (from each state)
  - Stage 3: Stratified sample of schools (urban/rural)
  - Stage 4: Stratified sample of classrooms

National Health and Nutrition Examination Survey (NHANES)



Courtesy of the Centers for Disease Control and Prevention.

Convenience & Multi-Stage Sampling

- Avoid using convenience sampling unless the question is low risk
- Use if the only type of sampling possible under the circumstances
- Also used when resources are low
- Multi-stage sampling usually used in large, governmental studies.

Photograph by FBI Buffalo Field Office

Conclusion

- Sampling Definitions
- Simple Random Sampling
- Stratified Sampling
- Systematic Sampling
- Cluster Sampling
- Convenience & Multi-stage Sampling