

Randomizing - Random Samples

Simple Random Sampling

- For simple random sampling,
 - all samples of a set size are considered to be equally likely
 - all units have the same chance of being selected for a sample
 - all pairs of units have the same chance of being selected to belong to a sample
 - and so on ...

Simple Random Sample

- Create a list of the units of the population
- Number the units in the list
- Use a random number table or random number generator to choose units from the numbered list, one at a time until you have as many as desired/needed

Sampling Frame

Sampling Frame

- Telephone Directory
- Electoral Register
- U.S. Department of Agriculture
- U.S. Environmental Protection Agency
- Area Sampling Frame

Stratification

- **Strata are subgroups in the population that do not overlap and that cover the entire sampling frame**
- **Stratified random sample is a simple random sample within a subgroup**

Choosing a Stratified Random Sample

- Divide the units of the sampling frame into non-overlapping subgroups
- Choose a simple random sample from each subgroup

Why stratify?

- **Convenience** - easier to sample a smaller, more compact group than a large group
- **Coverage** - provides data from each subgroup
 - **Note:** a simple random sample from a frame does not guarantee coverage from each stratum

Why stratify?

- **Convenience** - easier to sample a smaller, more compact group than a large group
- **Coverage** - provides data from each subgroup
- **Precision** - stratification gives estimates that are closer to the values for the entire population than simple random samples

Why stratify?

Precision – stratification gives estimates that are closer to the values for the entire population than simple random samples

This is the fundamental reason for stratification!

Examples of Strata

- In a population
 - Men
 - Women
- In textbooks - by discipline
- In physical objects
 - Size
 - Color

Examples of Strata

- You have taken surveys, what strata have you observed for a population
 - Sex

Examples of Strata

- You have taken surveys, what strata have you observed for a population
 - Sex
 - Ethnicity
 - Marital status
 - Age
 - Location (ex. country, urban/rural)
 - education

Creating the Strata

- Make the strata as different as possible
- Choose the relative sample sizes in stratified random samples that are proportional to the stratum sizes
 - Example - Given a population for which 7.5% are male and 92.5% are female, a sample of 1000 people stratified on gender should contain 75 males and 925 females

Stratification - Discussion

- An administrator wants to estimate the average amount of time high school students spend traveling to school. The plan is to stratify the students according to grade level and then take a simple random sample from each grade. **What is potentially good/bad about this plan for stratification?**

Stratification

- The strata should be as unique as possible
 - Older students tend to drive themselves to school vs. younger students who may take the bus
- Stratify based on distance from home to school
 - Travel distance has greatest effect on travel time

Stratification - Discussion

- Your assignment is to estimate the mean number of hours per week spent studying by students in your class. **Discuss how you would set up a stratified random sampling plan to accomplish this task.**

Stratification

- Ideally, want students who study a lot in one strata and students who do not study a lot in another strata
 - Impossible
- Look for variables that affect study time
 - Taking science/math
 - Year of study (fresh/soph/jr/sr)
 - Academic program

Practical Problem with Random/Stratified Samples

- Choosing sample units one at a time
 - Costly
 - Time consuming
 - A problem or impossible if good frame not available

Practical Problem with Random/Stratified Samples

- Choosing sample units one at a time
 - Costly
 - Time consuming
 - A problem or impossible if good frame not available
- Possible solution - work with groups rather than individuals

Cluster Samples

- A cluster sample is a simple random sample of clusters/groups/sets/collections of units
- Taking simple random samples of clusters/groups/sets/collections of units in a population produces a cluster sample

Cluster Samples

- Why create a cluster sample
 - Your need only a list of the clusters rather than a list of the units in the population
 - Easier to obtain information on individuals grouped by cluster rather than on individuals one at a time

Choosing a Cluster Sample

- Create a numbered list of the clusters in the population
- Choose a random sample of clusters
- Obtain data on each individual in each cluster in the simple random sample

Two-Stage Samples

- For a two-stage sample, we first take a cluster sample and then take a simple random sample from each cluster

Two-Stage Samples

- For a two-stage sample, we first take
 - a cluster sampleand then take
 - a simple random sample from each cluster

Two-Stage Samples

- For a two-stage sample, we first take
 - a cluster sample *(Stage 1)*and then take
 - a simple random sample from each cluster *(Stage 2)* **

Cluster vs. Two-Stage Samples

- For a cluster sample
 - take a random sample to determine the groups of individuals/units

Cluster vs. Two-Stage Samples

- For a cluster sample
 - take a random sample to determine the groups of individuals/units
- For a two-stage sample
 - take a random sample to determine the groups
 - THEN take a random sample to determine the individuals/units

Two-Stage Samples

- Useful when
 - easier to list clusters than individuals
 - but still *relatively* easy to sample individuals in clusters

Systematic Samples with Random Start

- “count off”
- pick one number, k , at random
- take every k th member of the population
 - starting with a random unit between 1 and k , inclusive

Steps in Systematic Sampling

- Known Population Size

- Determine population size, N
- Determine sample size, n
- Compute N/n and round *down* to nearest integer - this is k
- Randomly select number between 1 and k , *inclusive* - let this be p
- Sample individuals numbered
 $p, p+k, p+2k, p+3k, p+4k, \dots, p+(n-1)k$

Stop & Shop

- You are hired by *Stop & Shop* to conduct a customer satisfaction survey.

How might we conduct this survey?

Stop & Shop

You are hired by *Stop & Shop* to conduct a customer satisfaction survey.

- Choose desired sample size
 - How do we choose k ?

Stop & Shop

You are hired by *Stop & Shop* to conduct a customer satisfaction survey.

- Choose desired sample size
 - How do we choose k ?
 - ◉ Food for thought: If we want a sample 200 people and we survey every 30th person, how many *Stop & Shop* customers will we need?

Systematic Sample with Random Start - Groups

- "count off" to divide population into groups of size desired for the sample
- Use a chance method to choose one of the groups for the sample

Systematic Sample with Random Start - Groups

- "count off" to divide population into groups of size desired for the sample
- Use a chance method to choose one of the groups for the sample
- **Note: A systematic Sample with Random Start is not a type of simple random sample since not all possible groups have a chance of being in the sample**

Choosing a Systematic Sample with Random Start

- Suppose we have a directory of names of the 3500 day students at FSC and we want to create a sample of 70

Choosing a Systematic Sample with Random Start

- Suppose we have a directory of names of the 3500 day students at FSC and we want to create a sample of 70
 - Since $3500/70=50$, we select a random start between 1 and 50 (for example, 29) and then take every 50th unit thereafter

Multistage Samples

- An extension of two-stage sampling
- For example
 - Start with a simple random sample of one strata
 - Take a simple random sample of another strata
 - Take a simple random sample of a third strata

Multistage Samples

- You are hired to do a chocolate chip cookie study
 - Take a simple random sample of supermarkets
 - Take a simple random sample of bags of cookies from each supermarket
 - Take a simple random sample of cookies from each bag

Systematic Sample

- Suppose 200 people are waiting in line for tickets to a rock concert. You are working for the school paper and want to interview a sample of the people in line. Show how to select a systematic sample of 5% of the people in line, and then 20% of the people in line.

Systematic Sample

- Take 5% of 200 people: 10 people
- Determine the cont number:
 $200/10 = 20$
- Choose a random start between persons 1 and 20, inclusive, and choose every 20th person after that

Systematic Sample

- Take 20% of 200 people: 40 people
- Determine the cont number:
 $200/40 = 5$
- Choose a random start between persons 1 and 5, inclusive, and choose every 5th person after that

Systematic Sample

- **Q** - There are 3500 day-students at FSC. You want to sample 500 names. How do you use an alphabetical listing to take a systematic sample with random start?

Systematic Sample

- Determine the number for “counting off”: $3500/500 = 7$
- Choose a random number between 1 and 7, inclusive, at which to start and include every 7th person after that

Summary of Sampling Methods

- **Simple Random Sample (SRS) -** number the units; use a random number generator/table to select the numbers for the units to include in the sample
- **Stratified Random Sample -** divide population into strata; take a random sample from each strata

Summary of Sampling Methods

- **Cluster Sample** - select clusters at random then use all units in the selected clusters as the sample
- **Two-Stage Sample** - select clusters at random and then select units at random from each cluster

Summary of Sampling Methods

- **Systematic Sample with Random Start** - "count off" units by a number k determined by the population size and desired sample size, and select one of the counting numbers between 1 and k , inclusive, at which to start and take each k^{th} unit after that