STATISTICAL SAMPLING

Module One
Lesson Three

This presentation is based on material and graphs from OpenStax and is copyrighted by OpenStax and Georgia Highlands College
Sampling

- Is the process used to select a subset of the larger population (called the sample)
- Is very practical and useful since it would take a lot of time and money to study an entire population (The government tries to accomplish this every 10 years --- it’s called the CENSUS)
- The sample is examined, information is interpreted and applied to the entire population.
  - *In presidential elections, opinion poll samples of 1000-2000 people are taken. The opinion poll is supposed to represent the view of the people in the entire country*
  - *Manufacturers of canned carbonated drinks take samples to determine if a 16-ounce can contains 16-ounces of carbonated drink. The sample is supposed to ensure that all 16-ounces cans contain the correct amount.*
Sampling requirements

- Care must be taken in choosing the sample so that the population is represented well and the results of the statistical study will be meaningful.

- All samples should be:
  - Representative---which means the sample has the same relevant characteristics of the defined population, the sample does not favor one group of the population over another
  - Random ---selected by “chance”, no rhyme or reason to the selection

- In this lesson, we will identify five sampling techniques that can help to ensure the randomness and the representativeness of the selected sample
  - Random
  - Systematic
  - Convenience
  - Stratified
  - Cluster
  - We will consider these as individual sampling techniques; however, in most research studies, researchers use a combination of two or more of these techniques.
Statistical Sampling Techniques

- **RANDOM** (Simple Random Sample)
  - Each sample of the same size has an equal opportunity of being chosen.
  - Initially, each member of the population has an equal chance of being selected for the sample.

Examples:
- Drawing names from a hat
- Random Number Generator
Statistical Sampling Techniques

- **SYSTEMATIC**
  - Select a random starting point and then select every $n^{th}$ subject in the population
  - Simple to use so is used often

- Look for “th” in the description
Statistical Sampling Techniques

- **CONVENIENCE**
  - Use results that are readily available
  - Use “subjects” that are easily accessible
  - **CAUTION:** could result in a non-representative or biased sample

- Examples:
  - family members
  - students in classroom
  - mall shoppers
Stratified Sampling Techniques

- **STRATIFIED**
  - Divide the population into at least two different groups (strata) with common characteristic(s), then draw **SOME** subjects from each group
  - Results in a more representative sample
  - Helps preserve certain characteristics of the population
Statistical Sampling Techniques

- **CLUSTER**
  - Divide the population into groups (*clusters*), randomly select some of the groups, and then collect data from **ALL** members of the selected groups
  - Used extensively by government and private research organizations

- Examples:
  - Exit polls
Determine the statistical sampling technique used (random, systematic, convenience, stratified, or cluster)
Scenario: A study is done to determine the average tuition that San Jose State undergraduate students pay per semester. Each student in the following samples is asked how much tuition he or she paid for the Fall semester.

- A sample of 100 undergraduate San Jose State students is taken by organizing the students’ names by classification (freshman, sophomore, junior, or senior), and then selecting 25 students from each classification
  - STRATIFIED — groups, some selected from each group

- A random number generator is used to select a student from the alphabetical listing of all undergraduate students in the Fall semester. Starting with that student, every 50th student is chosen until 75 students are included in the sample.
  - SYSTEMATIC ---”th”
Scenario: A study is done to determine the average tuition that San Jose State undergraduate students pay per semester. Each student in the following samples is asked how much tuition he or she paid for the Fall semester.

- A completely random method is used to select 75 students. Each undergraduate student in the fall semester has the same probability of being chosen at any stage of the sampling process.
  - RANDOM

- The freshman, sophomore, junior, and senior years are numbered one, two, three, and four, respectively. A random number generator is used to pick two of those years. All students in those two years are in the sample.
  - CLUSTER—groups, select groups—all in each group used

- An administrative assistant is asked to stand in front of the library one Wednesday and to ask the first 100 undergraduate students he encounters what they paid for tuition the Fall semester. Those 100 students are the sample.
  - CONVENIENCE—easy access
A soccer coach selects six players from a group of boys aged eight to ten, seven players from a group of boys aged 11 to 12, and three players from a group of boys aged 13 to 14 to form a recreational soccer team
  - STRATIFIED

A pollster interviews all human resource personnel in five different high tech companies
  - CLUSTER

A high school educational researcher interviews 50 high school female teachers and 50 high school male teachers
  - STRATIFIED
A medical researcher interviews every third cancer patient from a list of cancer patients at a local hospital
  - **SYSTEMATIC**

A high school counselor uses a computer to generate 50 random numbers and then picks students whose names correspond to the numbers
  - **RANDOM**

A student interviews classmates in his algebra class to determine how many pairs of jeans a student owns on the average.
  - **CONVENIENCE**