MODULE 1. FUNDAMENTAL CONCEPTS

Module 1.2: Populations vs. Samples

Learning Objectives

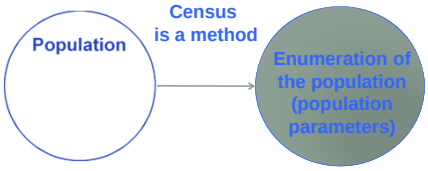
- Differentiate between a sample and population
- Explain the advantages of using samples in public health research
- Describe the types of bias in public health research
- Differentiate between a parameter and statistic
- Differentiate between common sampling methods

Population

- **Population**: The entire collection of individuals or measurements about which information is desired.
- A **population** can really be anything. Floridians, counties, all students of USF, computers, cars or hospitals in US could all be considered populations, depending on the research question. Although most populations being researched are large, they do not necessarily have to be.

Population vs. Census

Census: a complete count or enumeration of the population. Every member of the population is included!

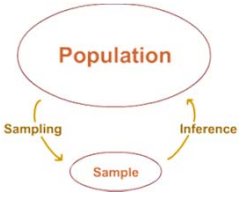


Population vs. Census

- In a census we examine each and every member of the population. **Why would we not use a census for all studies?**
  1) Not accurate
  2) Very expensive
  3) Perhaps impossible
  4) If using destructive sampling, you would destroy population
     - Breaking strength of soda bottles
     - Lifetime of flashlight batteries
     - Safety ratings of cars

Population vs. Sample

- **Sample**: A subset of the population selected for study. Not every element (e.g., individual) in the population is included in the sample.
  - We use the sample to make inferences about the population.
  - But having the right sample size is not enough to ensure good results. We want a sample that is representative of the population.



Population vs. Sample (cont'd)

Example: Suppose we want to find out how many books an ordinary American reads annually.

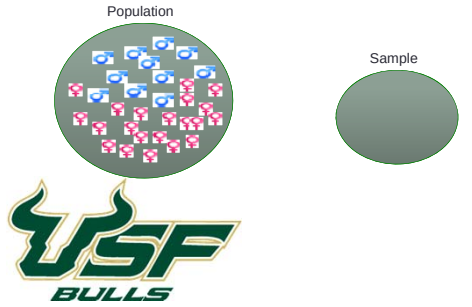
- **Solution**: We ask 2000 college students to keep track of what they read over the year, then check back with them after a year has gone by. Suppose we find the mean number of books read is 12, and then conclude that the average American reads 12 books per year.

Does this sample accurately represent the population?

Population vs. Sample (cont'd)

- The above result may not be accurate because a majority of college students are between 18-25 years old, and are required by their instructors to read textbooks and novels.
- This is a poor representation of the average American.
- A good sample would contain people of different ages, from all walks of life, and from different regions of the country.
- To acquire such a balanced, representative sample we would need to sample randomly so that every American has an equal chance or probability of being in the sample.

Population vs. Sample (cont'd)



Bias Due to Unrepresentative Samples

- A good sample is **representative**. This means that each sample point represents the attributes of a known number of population elements.
- Bias often occurs when the survey sample does not accurately represent the population. The bias that results from an unrepresentative sample is called **selection bias**.
- There are also other sources of bias: **Non-response bias**, **measurement bias, response bias**

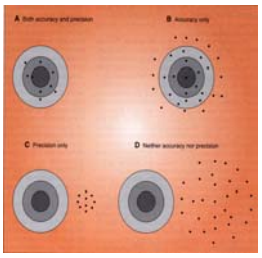
Bias Due to Unrepresentative Samples (cont'd)

- **Non-response bias** occurs when certain groups in the population are *under-represented* due to *low* rates of participation.
- **Measurement bias** occurs when there are errors in the measurement technique
- **Response bias** occurs when respondents *purposely* give inaccurate or false answers
  - can occur when respondents *lie* to avoid embarrassment or if they want to influence the results
  - can also occur when there are *leading questions*

Precision and Accuracy

Precision

- The degree to which the value of a measurement has nearly the same value when measured several times (repeatability)



Accuracy

- The degree to which the measurement actually represents what it was intended to represent
- How close is the measurement to the truth?

Parameter vs. Statistics

Parameter	Statistic
• Any statistical characteristic of a population. • Population mean, population median, population standard deviation are examples of parameters. • Describes the distribution of a population • Parameters are fixed and usually unknown	• Any summary characteristic of a sample. • Sample mean, sample median, sample standard deviation are some examples of statistics. • Statistic describes the distribution of a sample • Value of a statistic is known and varies for different samples • Are used for making inference on parameter (about the population)

Review – Definition of Terms

Population - The entire group of people of interest from whom the researcher needs to obtain information.

Element (sampling unit) - one unit from a population upon which a measurement is taken.

Sampling - The selection of a subset of the population

Sampling Frame - Listing of elements of a population from which a sample is chosen

Census - A polling of the entire population

Survey - A polling of the sample

Sampling Methods

How can we help to ensure that our samples are accurate representations of the population for which we are trying to make factual statements about?

Answer: A sample is claimed to be representative of a population from which it is drawn if every element of the population has equal chance of being included in the sample.

Sampling Methods (cont'd)

These are special types of study designs which are used to obtain data from a representative sample that can be used to estimate the value of a population parameter(s).

Sampling design refers to the method used to choose the sample from the population.

Common sampling methods

1. Simple random
2. Convenience
3. Systematic
4. Cluster
5. Stratified

Overview of sampling methods

Overview of sampling methods

Steps in Sampling Process

1. Define the population
2. Identify the sampling frame
3. Select a sampling design or procedure
4. Determine the sample size
5. Draw the sample

Types of Sampling Methods

Probability Sample	Non-probability Sample
All elements (e.g., individuals, hospitals, etc.) have an equal chance of being included in the sample (Random)	Elements do NOT have an equal chance of being included in the sample (non-random)
Examples Simple Random Sample Systematic Sample Stratified Sample Cluster Sample	Examples Convenience Sample (e.g., snowball, quota)

Simple Random Sample (SRS)

- Consist of individuals selected from a well-defined population in such a way that every individual has an equal (and known) chance of being selected
- **Purpose:** To avoid **BIAS**

Stratified Random Sample

- A population is divided into homogeneous groups called strata
- Then a simple random sample is drawn from each stratum.
- The proportion of individuals in the overall sample that are sampled from each stratum is the same as the stratum's proportional representation in the population.
- **Example**: Suppose we were to take a stratified random sample of 100 USF students. Since students are divided by gender, gender can be our strata. Then randomly select 50 females and randomly select 50 males.

Systematic Random Sample

- **Systematic sampling** is a method of selecting a probability-based sample from a **directory** or **list**. This method is at times more efficient than simple random sampling.
- Select sample by following a systematic approach
- Randomly select where to begin by using

Sampling interval (SI) = population list size (N) divided by a pre-determined sample size (n)

Systematic Random Sample

- **Example**: Suppose we want to do a systematic random sample of College of Public Health students.
- **Steps**:
  1. Get a list of students' names alphabetically from Registrar's Office
  2. There are approximately 2000 students in the College and if we want a random sample of 100 students, then decide where to begin the systematic sampling. To determine the starting point, 2000/100 = 20.
  3. Select a number between 1 and 20 at random, say that number is 5. The fifth student will be the first student chosen, then choose every 20th student from there. That is, the next sample is the 25th , 45th and so on.

Cluster Sample

- **Cluster sampling** is a method by which the population is divided into **groups** (**clusters**), any of which can be considered a representative sample.
- These clusters are mini-populations and therefore are heterogeneous.
- Once clusters are established a random draw is done to select one (or more) clusters to represent the population.

Probability Sampling Methods

Sampling Method	Advantage	Disadvantage
Simple Random	• Unbiased • Easy to do	• May not be representative • Less precise statistics • Must have sampling frame
Stratified	• More unbiased than Simple Random • More precise statistics • Cost reduced if strata already exist	• Difficult to do if you must divide stratum • More complicated mathematically • Need sampling frame
Systematic	• Unbiased • Ensures sample is representative • Efficient/cheaper	• Less precise statistics • More complicated mathematically • Can be confounded by trend or cycle
Cluster	• Unbiased • Reduced Cost • No sampling frame needed	• My not be representative • Complicated mathematically

Review

- Representative samples provide information that can be used to describe the entire population.
- Representative samples are obtained through Randomization and Replication.
  - Randomization minimizes BIAS (increases accuracy)
  - Replication minimizes random sampling variation (increases precision)
- Complex sampling designs can address problems that arise in studies of heterogeneous populations

EXAMPLES

What type of sampling design?

The Educational Testing Service (ETS) needed a sample of colleges. ETS first divided all colleges into groups of similar types (small public, small private, etc.) Then they randomly selected 3 colleges from each group.

Stratified random sample

What type of sampling design?

A local restaurant manager wants to survey customers about the service they receive. Each night the manager randomly chooses a number between 1 & 10. He then gives a survey to that customer, and to every 10^{th} customer after them, to fill it out before they leave.

Systematic random sampling