MATH 201: Lecture 2a Handout

Section 1.3 Sampling Principles and Strategies

Name:_____ Date:_____

Learning goals for today

By the end of this lecture, you should be able to:

- Distinguish between a **population** and a **sample**.
- Identify individual **observational units** in a study.
- Recognize anecdotal evidence and common sources of bias.
- Classify sampling strategies (simple random, stratified, cluster, multistage).
- Evaluate scope of inference: generalizability and causality.

Populations and samples

A _____ is the entire group we want to learn information about.

A _____ is a subset of the population that is actually observed.

An _____ is a single individual or object measured in the study.

If we could collect data on everyone in the population we would have a _____.

Example research question: What is the average number of hours of sleep per night for students at CSUCI?

- Population:

- Sample:

- Observational unit:

Example research question: Does a new drug reduce progression of Alzheimeer's disease in people under the age of 65.

- Population:

- Sample:

- Observational unit:

Concerns for data collection

Anecdotal evidence refers to (check all that apply):

☐ information from one or a few individuals
☐ data from a random sample
☐ personal stories or isolated cases
☐ information that may not be representative

Example: "I smoked a pack a day for 30 years and never got sick, so smoking isn't actually that dangerous". Why would we not want to rely on anecdotal evidence? List at least one reason.

Suppose we ask a student who happens to be majoring in nutrition to select several graduates for the study investigating average sleep among CSUCI students. What kind of students do you think she might collect? Do you think her sample would be representative of all graduates?

_____ is where individuals who are easily accessible are more likely to be included in the sample.

Bias:

- is systematic error or deviation from the truth.
- can lead to invalid or incorrect conclusions.
- can result from many parts of a study such as sampling.
- can come in many forms:
  - sampling bias (when sample is not representative of population),
  - researcher bias (researchers' beliefs influence data collection or interpretation),
  - response bias (participants provide inaccurate answers),
  - confirmation bias (focusing on data that confirms existing beliefs while ignoring contradictory evidence)
  - …

You are not expected to learn all forms of bias in this class, but you should be able to identify a potential source of bias in a study and possible effects.

Sampling strategies

Ideally, we want to randomly sample individuals from a population so that our sample is **represenative** of the population. This allows for us to **generalize** our findings beyond the sample to the population. There are four sampling strategies we will learn for trying to get a random sample.

When each case / observational unit in the population has an equally likely chance of being in our sample, we have a _____.

The other three cases involve some grouping of the data into groups. If we first divide the cases into groups (e.g., classes, age groups, etc.) and

- then collect a random sample within each group, we have_____.
- randomly sample groups, and then collect data on all cases within the chosen group, we have a _____.
- randomly sample groups, and then collect a random sample within each chosen group, we have a _____.

Ex: A researcher divides voters into age groups (18–29, 30–49, 50+) and randomly samples from **each** group.

Sampling method:

Ex: A city randomly selects households from a complete list of all households in the city.

Sampling method:

Ex: A national study randomly selects states, then counties within states, then individuals within counties.

Sampling method:

Observational studies and experiments

For each study, indicate whether it is **observational** or an **experiment**.

Ex: Researchers record caffeine intake and sleep duration for 1,200 adults without assigning any treatment.

Type of study:

Ex: Participants are randomly assigned to either take a new allergy medication or a placebo.

Type of study:

Scope of inference

A study randomly samples 800 registered voters from a state and finds that 62% support a proposed policy.

Can the results be *generalized* to all registered voters in the state? Why or why not?

Can the study be used to establish a *causal* relationship? Explain.