ST 517 Note Outline 1: Data Collection

Notes for Lecture 1.1: Introduction

Introduction
In this outline we discuss where data comes from–the basics of data collection through surveys and experiments. As we will see throughout this course, if the data is collected in an incorrect way it will greatly impact our analysis.

We will start with some basic terminology, just to set a common language.

Key Terminology

- _____: All individuals of interest

- _____: Recording information about all individuals in a population

  Problem:

  Solution:

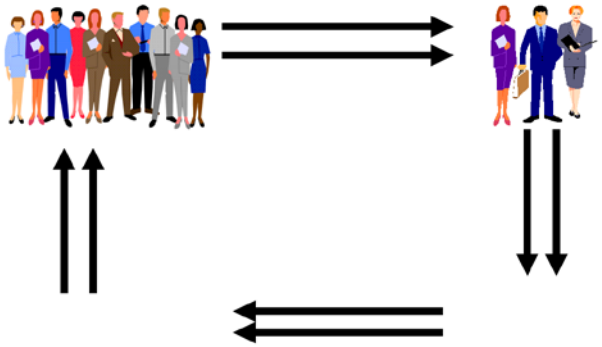
- _____: Subgroup of population from which we collect information
  - Notation: $\boldsymbol{n}$ =

- _____: Characteristic of the individuals we want to learn about
- _____: A summary of a variable for the entire population

- _____: A summary of a variable for a sample

- _____: Process of using sample information to make conclusions about the population

The Basic Paradigm of Statistics



Example: What are homes like in Wake County NC? A researcher selects 200 homes in Wake County to examine.

Population –

Sample –

Statistic –

Parameter –

The researcher lives in a Raleigh subdivision. He walks to each home in his subdivision and asks his neighbors if they have a fireplace.

- Is there a problem with this?

Notes for Lecture 1.2: Types of Samples

Problems with Samples

- **Biased samples:** More likely to produce some outcomes than others

- **Convenience samples:** Samples that are easy to take

- **Volunteer response sample:** Self-selected sample of people who responded to a general appeal

Avoid Biased Samples

- **Simple random sample (SRS):** A sample taken in such a way that every set of *n* items has an equal chance of being chosen
- Uses a chance mechanism

- How to take a Simple Random Sample:
  1. Compile a numbered list of the units in the population
  2. Use a computer, calculator, or table to pick items from the list

- **Sampling frame**: List of individuals from which we choose our sample
  - **Example:** If we want to take a sample of students in a statistics class, we might use the course roster as a sampling frame.

- Advantages of SRS:

  - **Unbiased** – Preferences of person taking sample does not come into play

  - Statistics that result have a predictable long run pattern

Other Types of Random Samples:

<u>Stratified Samples</u>

- Population is divided into groups (_____) and random samples are taken within those groups

- **Example:** We might want to sample university students by randomly selecting 25 graduate students and 25 undergraduate students.
- Advantages:

- Disadvantage:

<u>Cluster Samples</u>

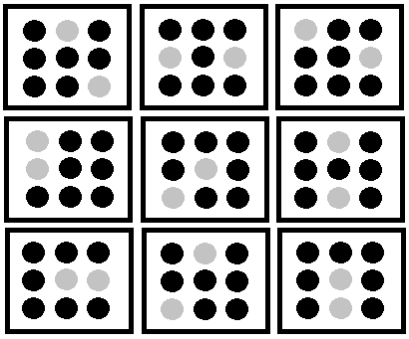
- Population divided into naturally occurring groups (_____)

  Randomly select _____ of subjects and talk with all subjects in that group

- **Example:** We may sample high school students by randomly selecting 10 high schools and talking with all students within that school.
- Advantages:

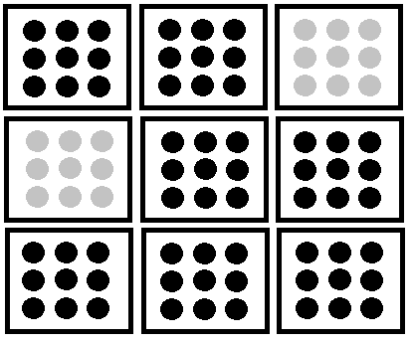
- Disadvantage:

Stratified Sampling



E.g. Randomly select 2 from each strata

Cluster Sampling



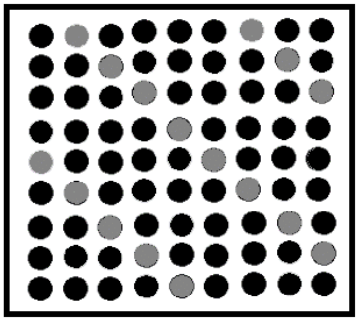
E.g. Randomly select 2 clusters, select all within

<u>Systematic Samples</u>

- Select every k^{th} item or individual from the sampling frame

- Advantage:

- Disadvantage:



E.g. Randomly select 2nd individual to start; then select every 5th individual

<u>Multi–stage samples</u>

- **Example:** Combine stratified with cluster samples by randomly selecting 30 high schools in urban areas and 30 in rural areas then talking to all students at those schools.

Notes for Lecture 1.3: Problems with Samples and Surveys

Sampling Errors and Bias

- Some possible errors and sources of bias arise because of the sampling process
- Use of bad sampling methods tends to lead to a type of bias generally called…
    - _____: Only a particular subset of people are selected or volunteer to be in the sample
- _____: Sampling frame does not include all of the population

Non-sampling Errors and Bias

- Some possible errors and sources of bias are not due to the process of taking a sample, but they can still cause problems in studies
- Data entry or processing errors (e.g. mistakes in typing in or copying data from its original source, problems with reading in or merging data, problems with transforming or categorizing data)
- _____: Some part of the population may not respond or refuses to participate
- _____: Responses given to the questions differ from the truth

Asking Questions

- Some questions are sensitive

- *Question wording* – Set up of the question can make a big difference

    - **Example:** CNN/USA Today Poll:
        1. Would you favor or oppose a new U.S. space program that would send astronauts to the moon?
        2. Would you favor or oppose the U.S. government spending billions of dollars to send astronauts to the moon?

- *Order of the questions* – Order of the questions can also make a difference in the response

Under-coverage, response, or non–response?

- *Under-coverage* – Part of the population could not be selected

- *Non–response* – Selected but don't have a response to all or part of the survey

- *Response bias* – Have a response but it is not a good reflection of the truth

Important points about sampling and surveys

- It is important to pay attention to the sampling method used when considering the results of a survey
- If the sample is not random, proceed with extreme caution!
    - You may not be able to make any conclusions about the full population
    - Instead, you have to think about what restricted/other population the sample is **representative** of

Example: Shopping cart or basket?

Weird! You may have heard the theory of shopping for groceries with a basket, rather than a full-size shopping cart, as a trick to limit spending, especially on impulse purchases. Logically, this makes sense: With less space to carry groceries, there'd seem to be less chance for making bad decisions. But a new study shows that shoppers gathering groceries in baskets are more likely to make unhealthy, wasteful purchases. Why might this be? The research of a group of European professors indicates that, oddly enough, the answer has something to do with how the basket shopper must flex his or her arm carrying the groceries. According to the study, in the Journal of Marketing Research: "We demonstrate that arm flexor contraction makes individuals more likely to choose immediately pleasing options."

That's another way of saying "instant gratification." The tension and strain on the arm (and presumably, back and shoulders as well) makes shoppers more likely to pick up "vice products" such as candy and soda, apparently as some sort of unconscious counterbalance to the hassles of carrying a shopping basket. When pushing a shopping cart on wheels, there is no "arm flexor contraction" necessary.

In a study the researcher simply followed around random shoppers in a supermarket, noting who was shopping with a cart, which had a basket, and what they brought to the register. What the data indicated is that: "The odds of purchasing vice products at the cashier for a basket shopper is 6.84 times the odds of purchasing vices for a cart shopper, all other things being equal."

Does the article above present convincing evidence that using a basket will cause shoppers to buy "vice" items?

Important Points

- The shopping cart example involved an
    - The researcher did not deliberately assign some people to use a basket and other to cart, rather, she simply observed people's shopping
- Observational studies are vulnerable to the presence of
    - Variables that may influence the response but that often are not studied explicitly
- In general, evidence of a relationship between two variables

- _____: impose a difference in the explanatory variables and try to determine if there is a difference in the outcome variable

Key Terminology

- _____: Individuals being measured

- _____: Measures the outcome of interest in a study

- _____: Variable we think explains changes in the response variable

- _____: Specific regimen or procedure assigned to subjects; different levels of an explanatory variable

Notes for Lecture 1.5: Designing Experiments

Basic Principles of Experimental Design

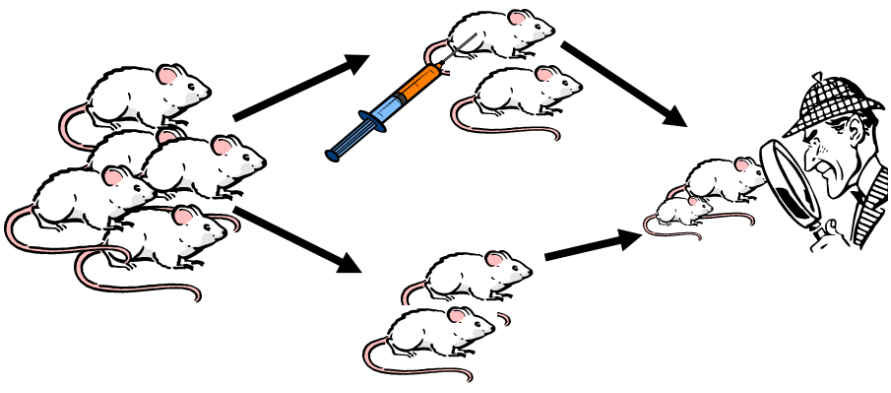
- _____: Subjects are assigned to the different levels of the explanatory variable (i.e. different treatments) by a random mechanism.
    - What it does:

- _____: Absence of treatment; some baseline or reference condition that is used for comparison.
    - What it does:

Notes:

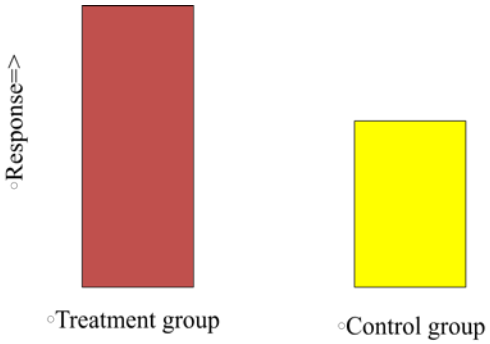
- Well-designed experiments include some type of _____
    - Group of subjects that have the same sources of variability as those receiving the treatment.
- The subjects in the control group are…
    - Similar to those in the treatment group
    - Treated identically to those in the treatment group

- Sometimes, they are given something called a _____, which seems like the treatment but is not

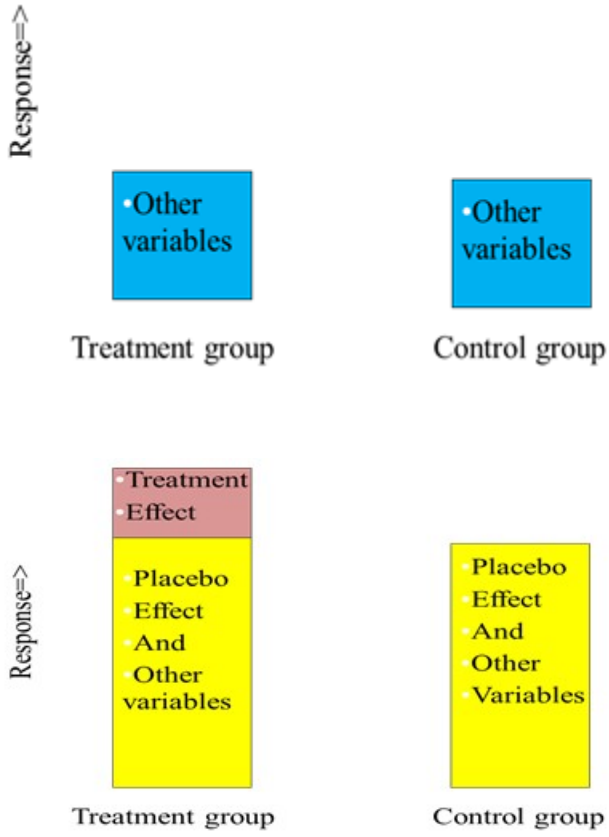


Notes, continued:

- Comparing Groups



- **<u>Placebo Effect:</u>** Tendency to react to a drug regardless of its actual physical function



Example: A 'Magic' Pain-Killing Bracelet That Isn't

A controversial bracelet that supposedly sends out special electrical waves to relieve pain provides no more relief than an ordinary bracelet, a new study finds. A lot of the people who wore the "ionized" bracelets reported relief from their pain during the four weeks of the study, says a report in the Mayo Clinic Proceedings. However, so did a lot of the people who wore the ordinary bracelets.

In the study, Dr. Robert Bratton and colleagues from the department of family medicine at the Mayo Clinic in Jacksonville, Fla., recruited 610 adults who reported having persistent pain. Half wore the ionized bracelets, half the ordinary bracelets. They gave assessments of their pain after 1, 3, 7, 14, 21 and 28 days, reporting not only on the area of greatest pain but also their overall feeling. The results were strikingly similar for both groups. For example, 75.1% of those in the placebo group reported an improved maximum pain score after seven days, compared to 75.7% in the ionized bracelet group. Bratton's conclusion: It's all a placebo effect and "based on the study results, you may be just as well off wearing a rubber band around your wrist and saving the money spent on the bracelet."

Identify the following characteristics of this experiment:

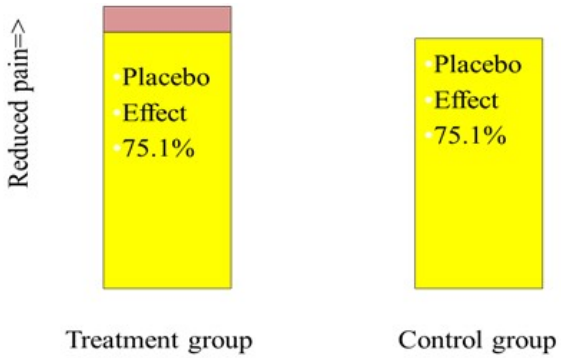
The type of bracelet worn was the:

The reported amount of pain experienced was the:

The 610 adults with persistent pain were the:

What was the treatment?

What was the placebo?



Important Points

- Placebos are not just sugar pills

- The placebo effect occurs

  That is why you need the control group for comparison—it allows you to measure and account for the placebo effect

Notes for Lecture 1.6: Problems with Experiments

Problems with Experiments

- In addition to the placebo effect, other issues can arise in experiments
- Bias of the Subjects

- **Hawthorne Effect**: People act differently when they know they are in an experiment

- Researcher Bias

Solutions

- _____: Someone involved in the experiment (either subjects or experimenter/technician) are not aware of which group a subject is assigned to

- _____: Neither subjects nor the experimenter/technician are aware of which group a subjects is assigned to

Other problems

- Non response /dropout – some subjects do not complete the study

- Non-adherence – some subjects do not follow the study protocol

- Generalization/Realism – can we say the results would hold true for the whole population?

Notes for Lecture 1.7: Types of Experimental Designs

- _____: Each subject receives one treatment, without taking other variables into consideration. This is the simplest type of experimental design.

- _____: Each unit receives two treatments. The units could be:
  1) A single subject (each subject serves as their own control)
  2) Two subjects that have been matched together (one receives the treatment and the other receives the control)

- _____: Subjects are divided into similar groups called blocks, and each treatment is applied in each block.

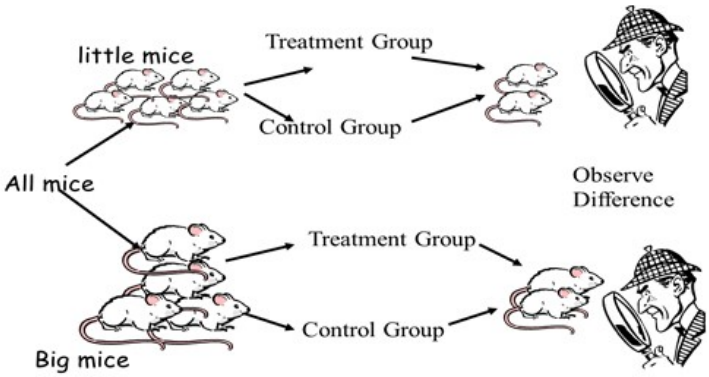
- _____: A special case of a block design.

  Blocks =

  Units =

  Design =

- Block Design: Visually



- Advantage of matched and block designs:

Important points about experiments

- The key to a good experimental design is that steps are taken to reduce…
    - Variation in the response that is due to variables you are not interested in studying (this increases your ability to detect a treatment effect)
    - The potential for bias (this increases your ability to say that any effect is due to the treatment and not other variables)
- A well-designed, randomized experiment (not an observational study) is the best way to collect data to establish cause and effect relationships