1.3 Introduction to Experimental Design

Basic guideline for planning a statistical study

1. State a hypothesis
2. Identify the individual of interest
3. Specify the variable to measure
4. determine if you will use the entire population or a simple
    1. if you choose a sample, choose sampling method
5. Address ethical concerns before data collection
6. Collect the data
7. use descriptive or inferential statistics to answer your hypothesis

Feature	Descriptive Statistics	Inferential Statistics
Purpose	Summarize and organize data.	Make inferences about a population.
Data Used	Entire dataset or sample.	Representative sample.
Output	Tables, graphs, measures of center/spread.	Probability scores, P-values, confidence intervals.
Focus	"What is the data showing?".	"What can we predict about the population?".

8. Note any concerns about your data collection or analysis

Example:
1-3

Hypothesis & Variables

- *Hypothesis:* Air pollution causes asthma in children who live in urban settings
- *Individuals:* Children in urban settings
- *Variables:* Air pollution and asthma

4-6

Sampling, Ethics & Data Collection

Photo courtesy of US Army Africa.

- Either collect data or use existing dataset
  - Can use a government dataset for population measures
- Can collect data from a sample for estimates
  - Need to choose sampling approach
  - Will need consent if legally found to be "human research"
  - May need consent from parents to collect data about children

7-8

Experiment vs. Observational Study

Experiment

- A *treatment or intervention* is deliberately assigned on the individuals
- The *purpose is to study the possible effect of the treatment or intervention* on the variables measured

Observational Study

- Observations and measurements of individuals are taken
- However, *no treatment or intervention is assigned* by the researcher

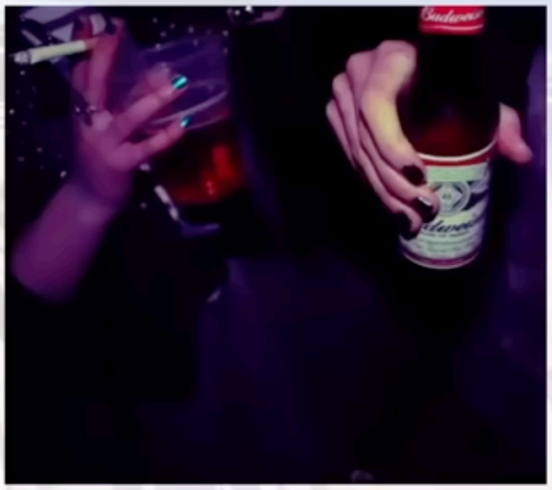
Replication

- Studies must be done rigorously enough to be replicated.
- Replicating the results of observational studies and experiments is necessary for science to progress.

Avoiding Bias in Survey Design

Non-response & Voluntary Response

- If many people refuse your survey, the people who do complete it are likely to have a biased opinion.
- There may be a reason they do not complete your survey that has to do with how they feel about your survey topic.



Truthfulness of Response

- Respondents may lie on purpose
  - If asked a question that is too personal
  - If asked a question too hard to think about
- Respondents may lie inadvertently
  - May not remember if asking about something that happened a long time ago
  - May have “recall bias” influenced by events that have happened since original event

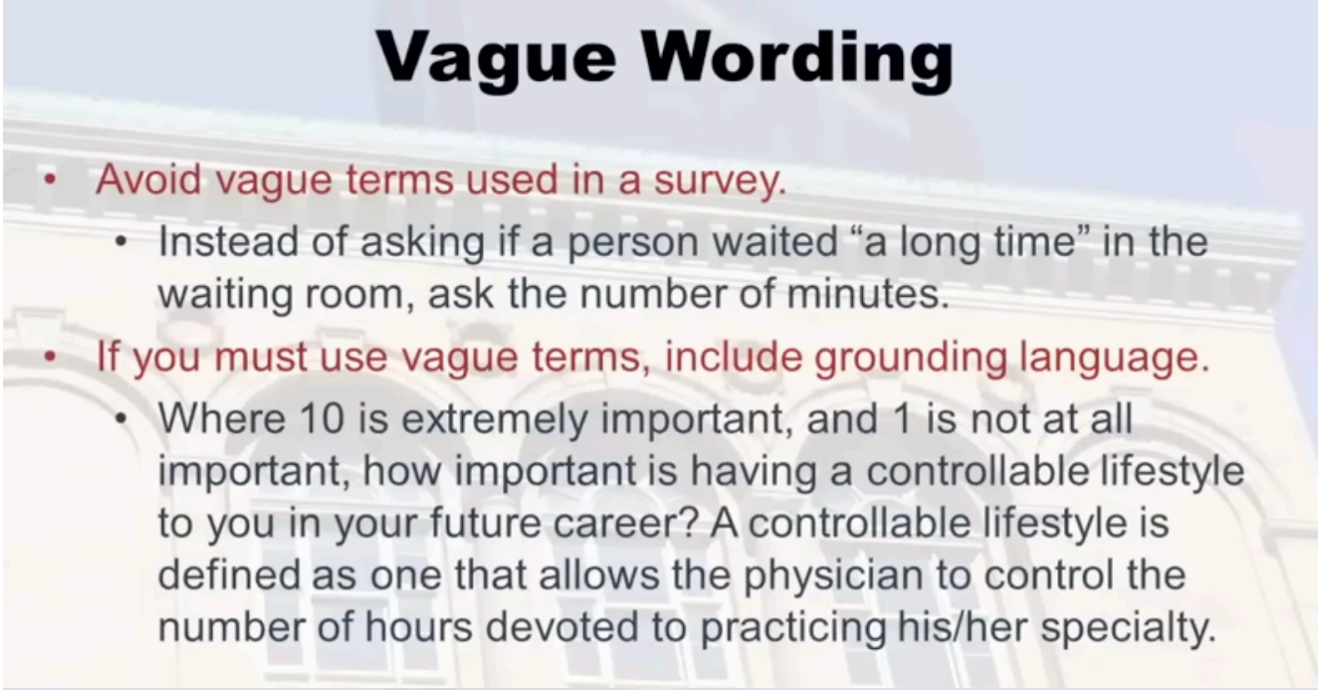
Hidden Bias

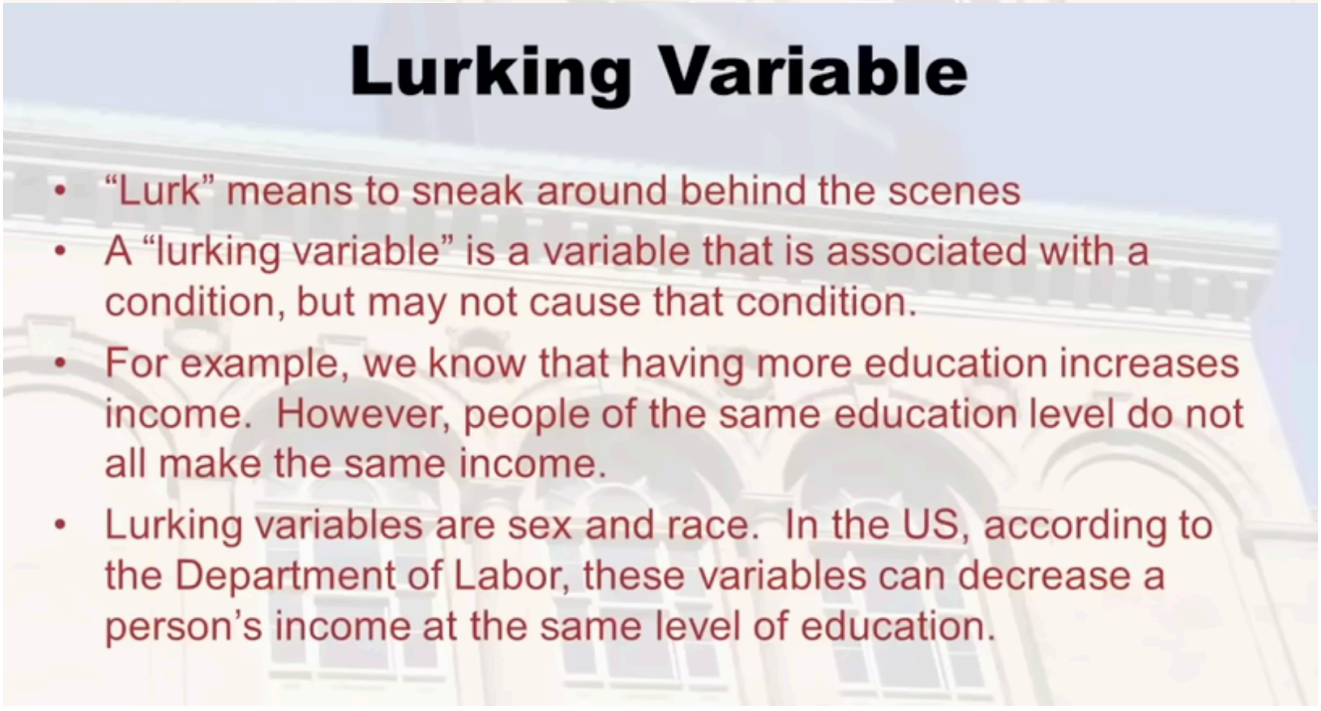
- Question wording may induce a certain response.
  - How long have you been using Software A?
- Order of questions and other wording may induce a certain response.
  - Do you agree with Obamacare?
  - *More people have health insurance than ever before.* Do you agree with Obamacare?
- Scales of questions may not accurately measure responses
  - Do your feelings always fit on a scale of 1 to 5?

Interviewer Influence

Photo by UK Department for International Development.

- This is important with in-person and phone surveys
- Best to have interviewer from same population as research participant
- All verbal and non-verbal influences matter



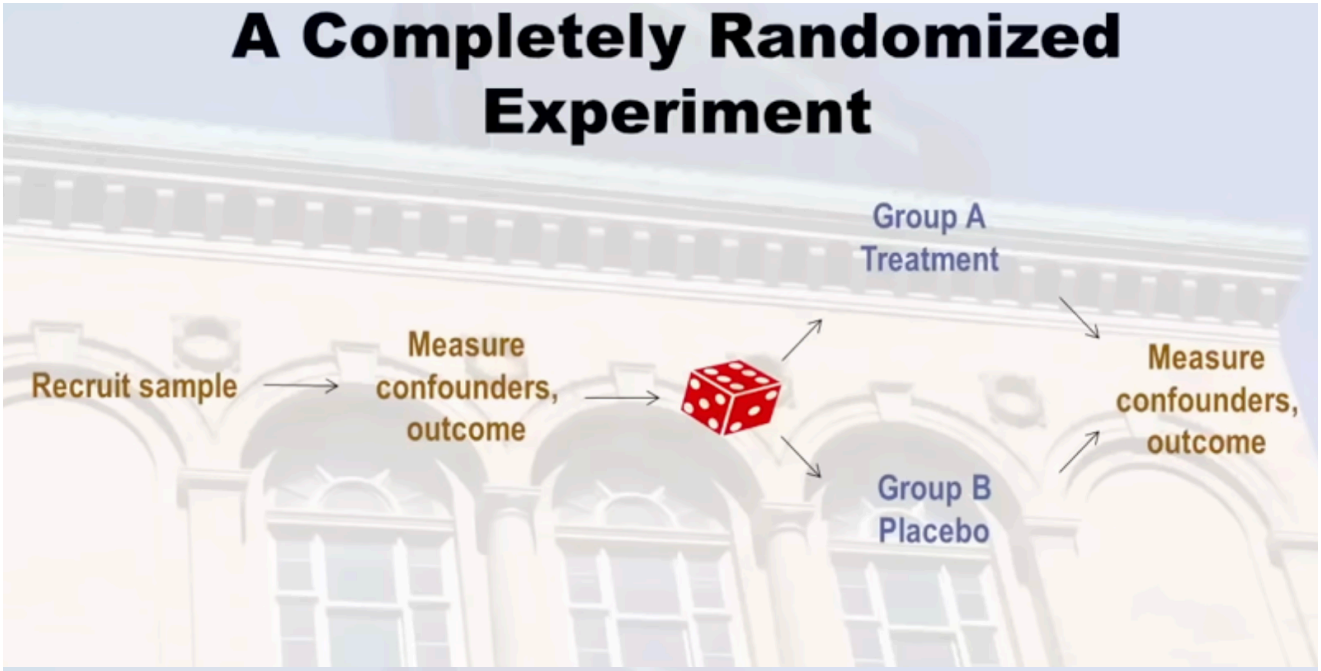


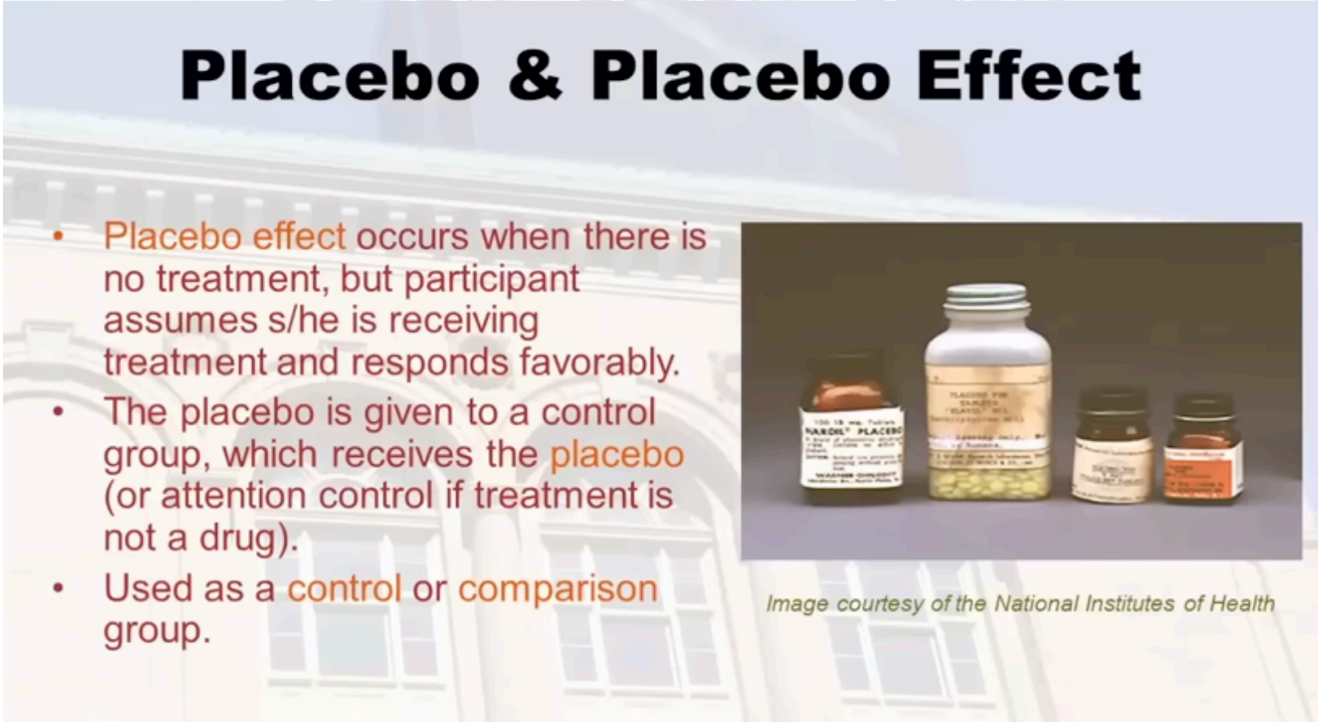
Randomization

-Randomize is used to assign individuals to treatment groups.

-This helps prevent bias in selecting members for each group

it disturbs 'lurking variable'





Blocked Randomization

If you want men and women equal in two randomized groups, create 'block' with two slots - one for a man, and one for a woman.

-Blinding

is where a person (participant, research staff) is deliberately not told of a treatment assignment in a study so she/he is not biased in reporting study information.

ex: A participant is blinded to treatment or placebo

-Double-Blind

means study staff and participant do not know treatment assignment