Can We Say What Caused the Effect?

ALISON KLEFFNER

Explanatory vs. Response Variables

Association vs. Causation

- Impact of Confounding Variables

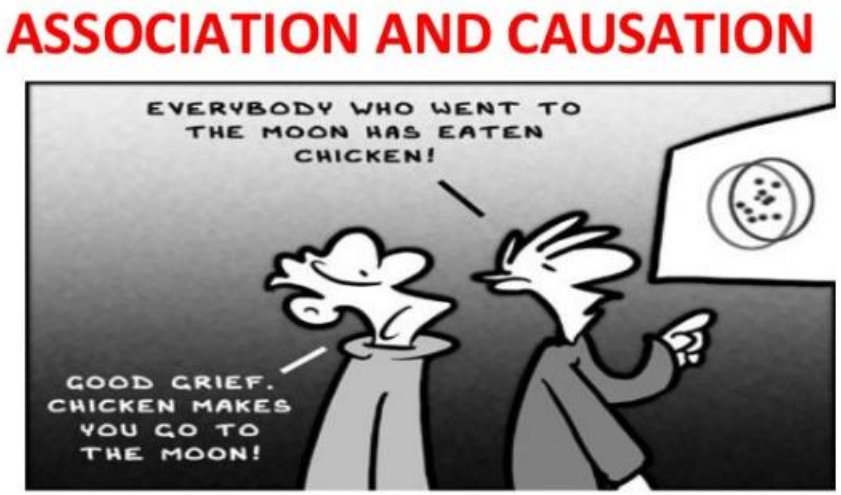
Observational vs. Experimental Studies

- Random Assignment
- What conclusions can we draw?

Schedule

Association vs Causation

Introduction



Can we say what caused the effect?

- What other things outside our study could affect our results?
- What must we keep in mind when we collect our data?

Focus on research questions that compare two groups:

- Are children who used night lights as infants more likely to need glasses than those who didn't use night lights?

Objectives

Determine when there is an association between our two variables.

Determine when we can conclude the outcome of one variable **causes** an outcome of the other.

- Does smoking **cause** lung cancer?

What other things outside our study could affect our results?

What must we keep in mind when we collect data?

Terminology

Explanatory Variable (Independent, Predictor):

- Variable we think is “explaining” the change in the response variable.
- This is the variable researchers manipulate.

Response Variable (Dependent):

- Variable that is being changed by explanatory variable.

Example: We may think that if you **eat an apple every day** you have to **go to the doctor** less.

Choose the Explanatory and Response

- Choose the explanatory and response variable:
  - Smoking and Lung Cancer
  - Heart Disease and Diet
  - Exam Score and Hours Studied
  - Vitamin C and Colds
  - Hair Color and Eye Color
- Sometimes there is a clear distinction between explanatory and response variables and sometimes there isn't.

Association vs. Causation

Concepts

Association: if one variable gives you valuable information about the other

- Hours of studying before exam perhaps gives information about the score you will achieve
- How many apples you ate perhaps gives information about how many times you will go to the doctor

Cause and Effect:

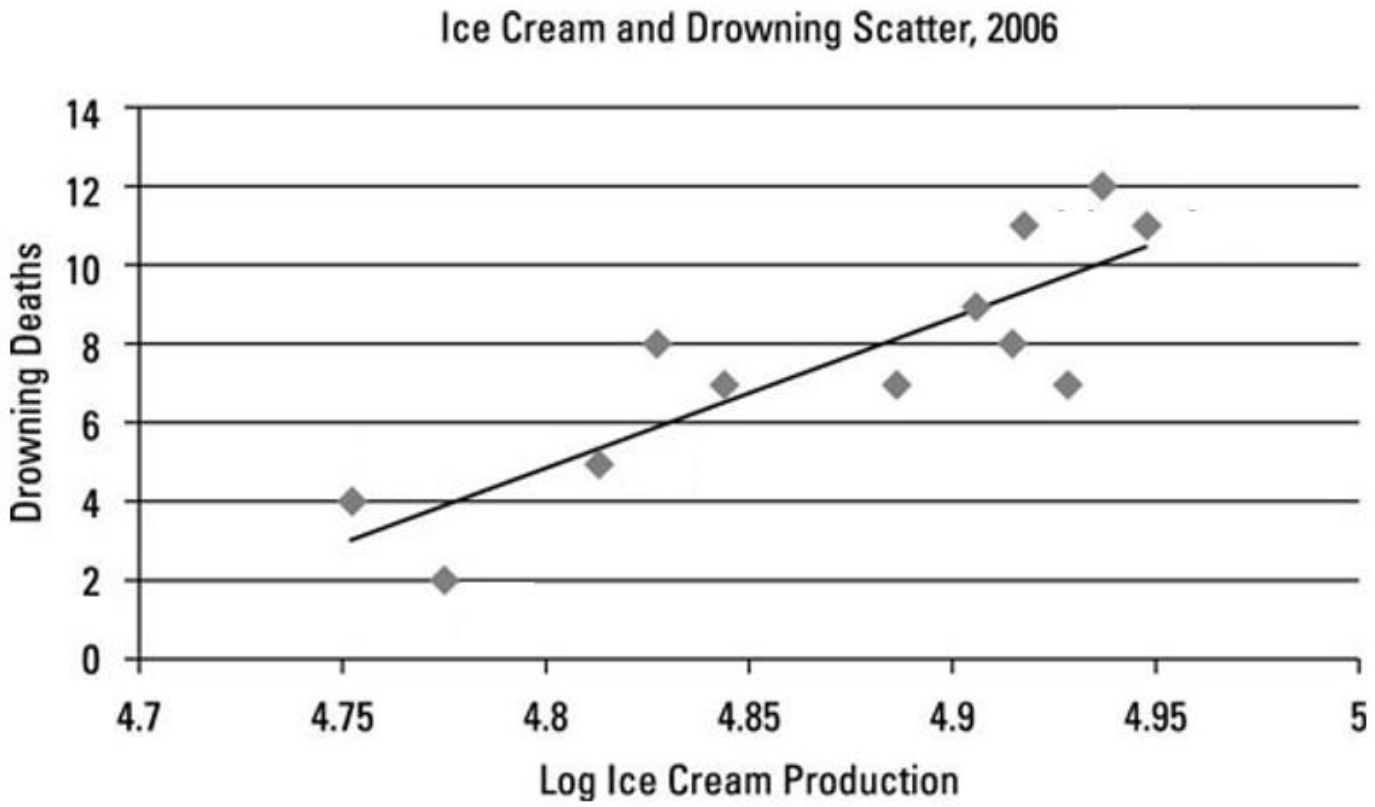
- Can we conclude cause and effect from our study? Is our explanatory *causing* the change in the response?

Confounding variable: A variable related to both the explanatory and response variable in such a way that its effects cannot be separated from the effects of the explanatory

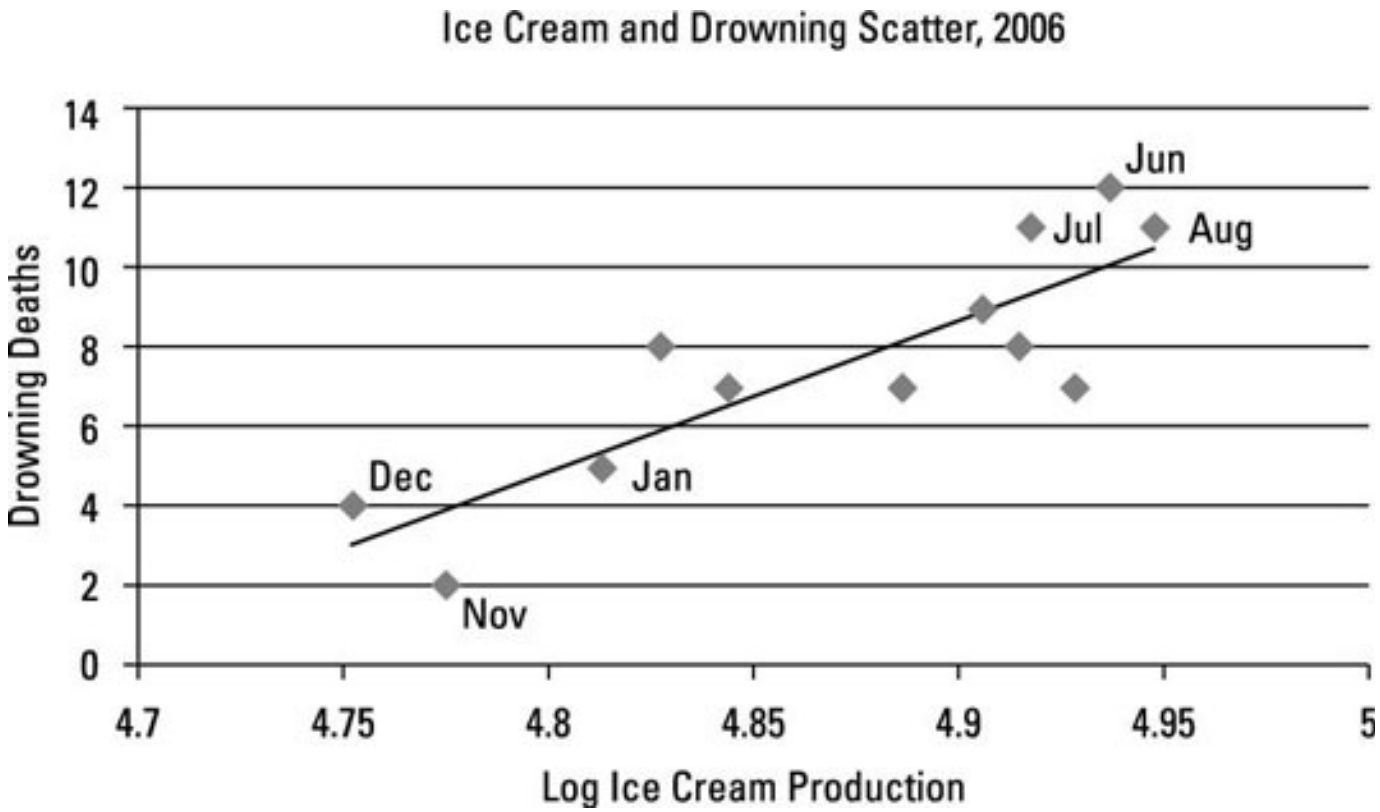
- Causes Bias

Association <u>does not</u> justify causation (because of possible confounding variables)

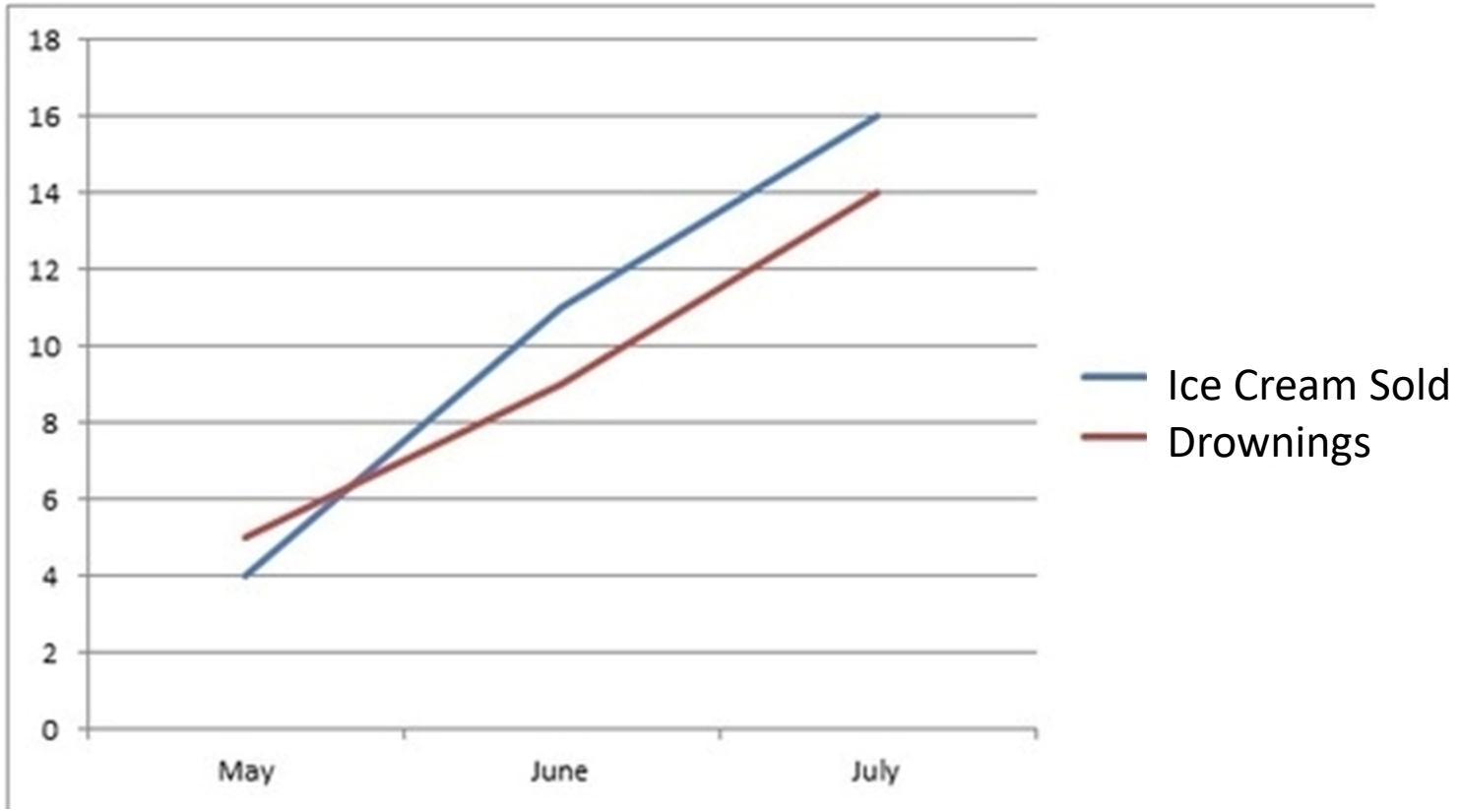
Amount of ice cream sold and deaths by drownings (Moore, 1993)

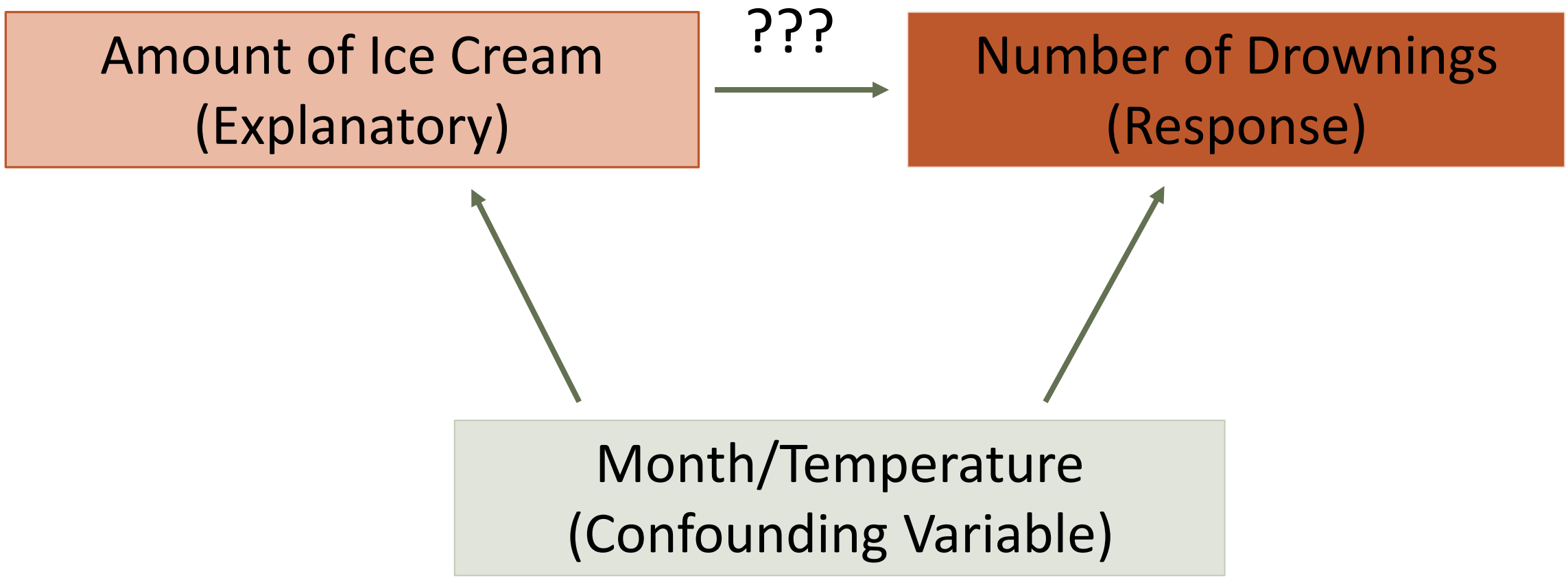


Amount of ice cream sold and deaths by drownings (Moore, 1993)



Confounding Variable – Month/Temp




Amount of Ice Cream
(Explanatory)
???
Number of Drownings
(Response)
Month/Temperature
(Confounding Variable)

Scenario

Researchers found that countries that have more TV's per 1,000 people have a higher average lifespan.

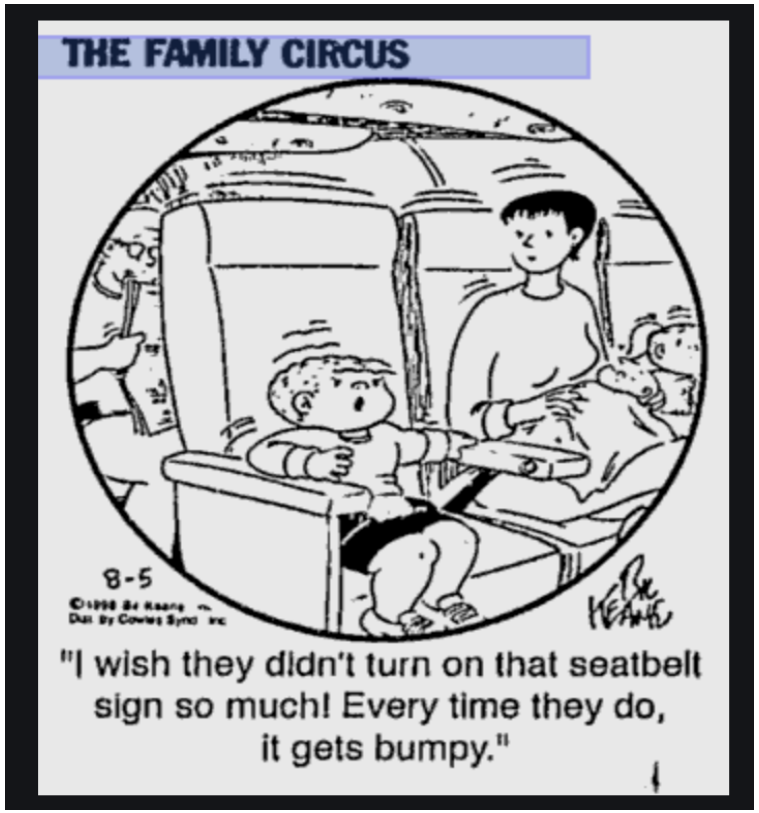
Claim-

Explanatory variable-

Response variable-

Potential confounding variable-

What's the confounding variable here?



Your Turn!

- Think of a topic!
- After you've found a topic, come up with a research question.
- Be creative!

Claim -

Then identify the following:

- Explanatory variable-
- Response variable-
- Potential confounding variable-
  - Explain Bias

Resource: Google Spurious Correlations

Take Away

- Identify which explanatory and response variables.
- Determine if there is an association between the explanatory and response variables.
- Identify potential confounding variables.
  - How do they provide an alternate explanation of the association between the explanatory and response variable?

Introduction

How can we **take a sample** so that we can **reduce the risk** of having confounding variables?

Confounding variable add **bias** to our study

- Additional factor we're not accounting for that affects our results.

Big picture

- Observational vs. Experimental Studies
- Random Sampling vs. Random Assignment

Sampling

Definitions:

Population: entire collection of observational units we are interested in

- Relates to Parameter

Sample: A subset of the population

- Relates to Statistic

- **Goal: Generalization**
  - Not only to describe the sample but to generalize characteristics of the sample to a much larger population

Sampling

Representative sample: A sample is representative if it is similar to or represents the whole population from which it has been drawn

- *Example: Would data from this class be representative of all TCU students?*

Bias: when the statistic consistently overestimates or underestimates the population parameter of interest.

- *Example: Would asking TCU students enrolled in 8 am classes if they are morning people be representative of all TCU students?*

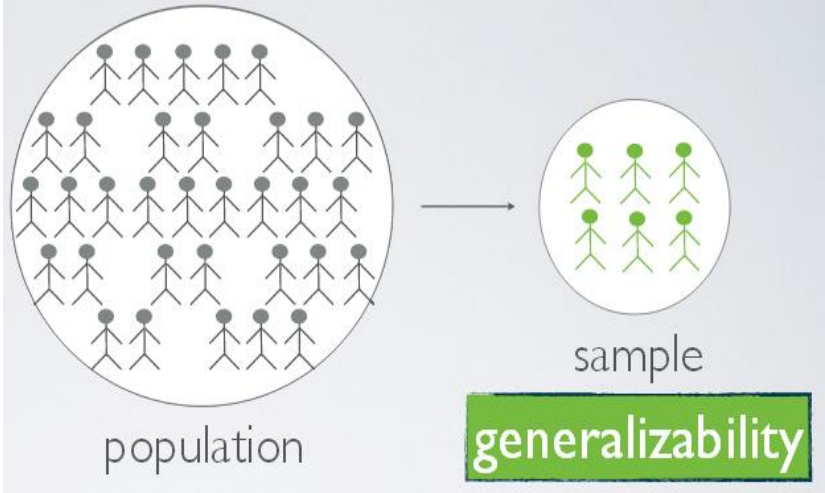
How to Sample?

1. Convenience Sampling

- Not a very good way to select a sample
- Essentially just go out and pick a set observational units.
- Relies on human judgment.

2. Simple Random Sample

- Every population member has an equal probability of being chosen
- **Allows us to generalize our results to the entire population**
- How does it work?
  - A simple random sample is when you choose individually randomly from the sampling frame and each individual has the same chance of being selected.
  - A **sampling frame** is a list of all the individual in the population of interest



Experimental vs Observational Studies

Have a nice trip

An area of research in biomechanics and gerontology concerns falls and fall-related injuries, especially for elderly people. Recent studies have focused on how individuals respond to large postural disturbances (e.g., tripping, induced slips). One question is whether subjects can be instructed to improve their recovery from such disturbances.

Researchers want to compare two recovery strategies:

- Lowering (quickly stepping down with front leg and then raising back leg over the object)
- Elevating (lifting front leg over the object).

24 subjects have agreed to participate in such a study:

- 8 between the ages of 21-40
- 16 between the ages of 41-60

Identify the following:

- Explanatory Variable
- Response Variable
- Potential Confounding Variable

Observational Study

Groups you compare are “just there.”

- Data observed and collected on subjects

Examples:

- Customer review surveys
- Teacher evaluations
- Smokers vs nonsmokers and lung cancer

Experiments

You create the groups by what you choose to do to the people or the experimental units.

Assigning conditions to be compared

- *Manipulation* of groups

Examples

- Assign one group of people to a diet another group to an exercise program, then measure weight loss on each group
- Students are assigned to listen to classical music or country music before an exam. Then the two groups scores on the exam are compared

Example

Researchers want to set up an experiment OR observational study in order to answer the research question "Does exercise help prevent against colds and the flu?"

How could this study be an example of an **observational** study?

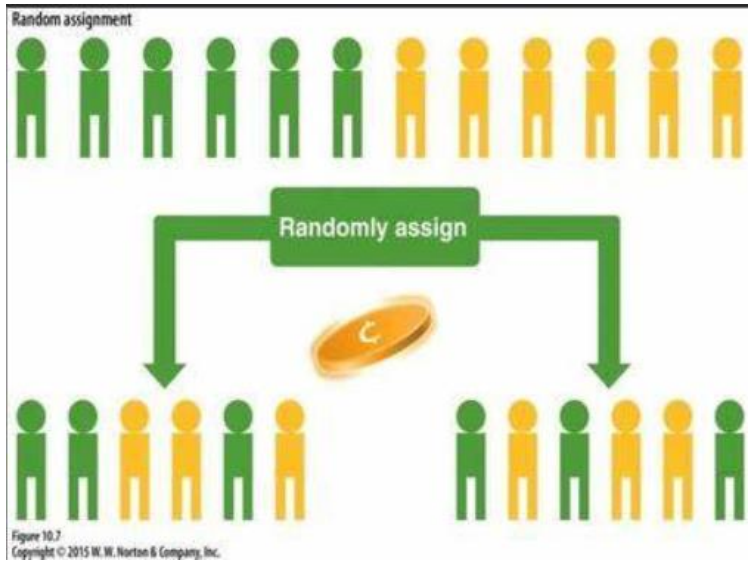
What would we have to do in order to make this an **experimental** study?

Randomized Experiment

Use a method of chance in order to make the assignments.

- Random assignment allows us to balance out confounding variables

Random Assignment: produce similar groups so we can say there is a cause-and-effect relationship between explanatory and response variables.



Have a nice trip

- Researchers want to compare two recovery strategies:
  - Lowering
  - Elevating
- 24 subjects have agreed to participate in such a study:
  - 8 between the ages of 21-40
  - 16 between the ages of 41-60

Possibility 1: Assign the 8 people between the ages of 21-40 to use the elevating strategy and the 16 people between the ages of 41-60 to use the lowering strategy.

Possibility 2: Assign 4 people between the ages of 21-40 and 8 people between the ages of 41-60 to each group. Show how the proportion of people between the ages of 41-60 in each group is the same.

Have a nice trip

- Researchers want to compare two recovery strategies:
  - Lowering
  - Elevating
- 24 subjects have agreed to participate in such a study:
  - 8 between the ages of 21-40
  - 16 between the ages of 41-60

Random Assignment!

- Because there will always be more potential confounding variables which could be distributed unevenly between the groups being compared

Suppose we put each person's name on a slip, put those slips in a hat and mix them up thoroughly, and then randomly draw out 12 slips for names of people to assign to the elevating strategy.

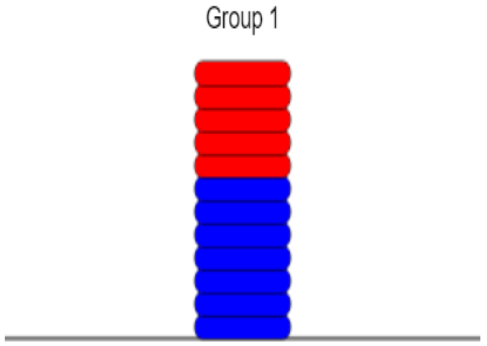
- What proportion of this group (elevating strategy) do you expect will be between the ages of 41-60?
- Do you think we will always get an 8/8 split?

Have a nice trip

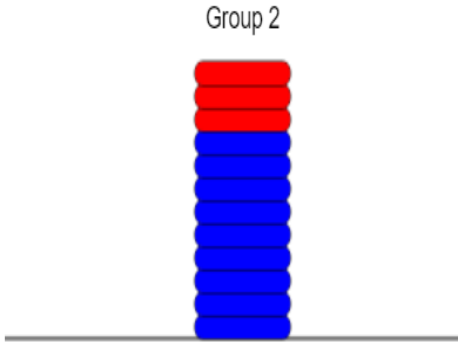
- Researchers want to compare two recovery strategies:
  - Lowering
  - Elevating
- 24 subjects have agreed to participate in such a study:
  - 8 between the ages of 21-40
  - 16 between the ages of 41-60

Key Idea: Randomly assigning experimental units to groups tends to balance out all other variables between the groups. Any variables that could have an effect on the response should be equalized between the two groups and therefore should not be confounding.

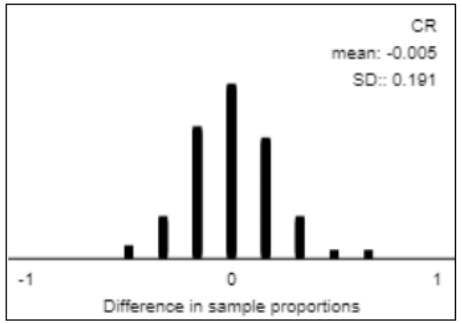
Simulation: Random Assignment between the two groups, simulated 100 times



Proportion 41-60: 0.5833



Proportion 41-60: 0.75



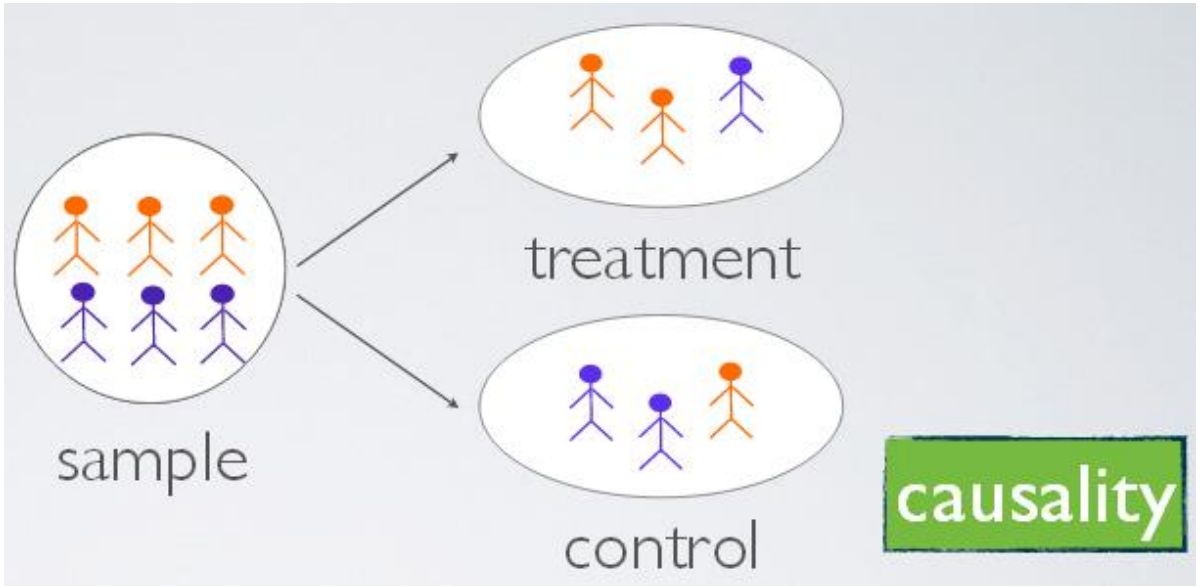
Assignment of Explanatory Variable

RANDOM ASSIGNMENT

Experimental Studies balance out confounding variables

- Manipulate groups to analyze

Determine Causation



NO RANDOM ASSIGNMENT

Observational Studies

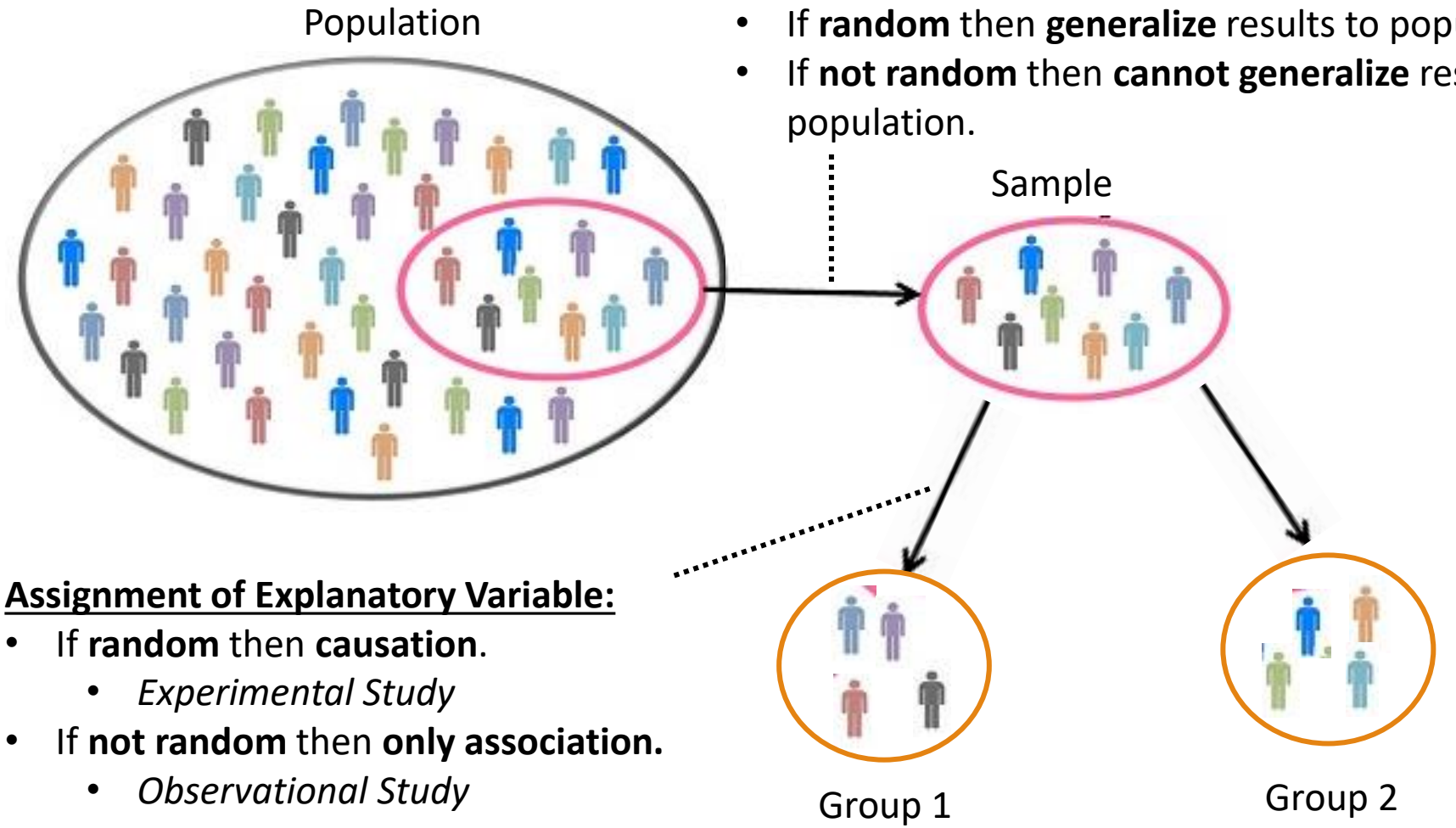
- Analyzing data that we do not change, simply given the data
- A study may be observation because it is easier, or because perhaps it is not possible or ethical to assign groups

Cannot Determine Causation, only potentially association

Random vs. Random

Sampling:

- If **random** then **generalize** results to population.
- If **not random** then **cannot generalize** results to population.



Assignment of Explanatory Variable:

- If **random** then **causation**.
  - *Experimental Study*
- If **not random** then **only association.**
  - *Observational Study*

Example

An article about handwriting appeared in the October 11, 2006 issue of the *Washington Post*. The article mentioned that among student who took the essay portion of the SAT exam in 2005-2006, those who wrote in cursive style score significantly higher on the essay , on average, than students who used printed block letters.

- Identify the explanatory and response variables in this study
- Is this an experimental or observation study? How do you know?
- If this is an observational study, how would you turn this into an experiment? Is this possible/ethical?

Review

Explanatory vs. Response Variables

Association vs. Causation

- Impact of Confounding Variables

Generating a Sample

Observational vs. Experimental Studies

- Random Assignment
- What conclusions can we draw?