STA 320:
Design and Analysis of Causal Studies

Dr. Kari Lock Morgan and Dr. Fan Li
Department of Statistical Science
Duke University

Office Hours and Support

- Dr. Kari Lock Morgan, kari@stat.duke.edu
  - Office Hours: M 3-4pm, W 3-4pm, F 1-3pm in Old Chem 216
- Dr. Fan Li, fli@stat.duke.edu
  - Office Hours: M 10:30-1130am, W 2-3pm in Old Chem 122
- TA: Wenjing Shi, wenjing.shi@duke.edu
  - Office Hours: tbd in Old Chem 211A
- Statistics Education Center
  - Sun, Mon, Tues, Wed, Thurs 4 – 9 pm in Old Chem 211A

Reading

- Causal Inference for Statistics and Social Sciences
  - By Guido W. Imbens and Donald B. Rubin
  - Not yet published
  - Draft pdf available on Sakai (do not share)
- Will distribute other relevant papers

Assessment

- Most weeks will either have a homework due or a quiz
- One midterm exam: 3/5
- Final project
- Grading
  - Homework: 20%
  - Quizzes: 20%
  - Midterm: 30%
  - Final project: 30%

Course Website

- stat.duke.edu/courses/Spring14/sta320.01/
- Class materials will be posted here
- Textbook, relevant papers, etc. will be posted on Sakai

Causality and Potential Outcomes

Causality?

- What is causality???
- In this class we will learn…
  - a formal framework for causal effects
  - how to design studies to estimate causal effects
  - how to analyze data to estimate causal effects

Rubin Causal Model

- In the class, we'll use the Rubin Causal Model framework for causal effects
- Key points:
  - Causality is tied to an action (intervention)
  - Causal effect as a comparison of potential outcomes
  - Assignment mechanism

Causal Question

- What is the effect of the treatment on the outcome?
- If the opposite treatment had been received, how would the outcome differ?
- Example:
  - treatment: choosing organic produce
  - outcome: get cancer? (yes or no)
  - Question: Does choosing organic produce decrease risk of cancer?

Treatment

- For most of this course, we will consider only two possible treatments:
  - Active treatment ("treatment")
  - Control treatment ("control") – often just not getting the treatment
- Two treatments (including the alternative to specified treatment) must be well defined

Causal Questions

- What is the effect of studying on test scores?
- Does texting while driving cause accidents?
- Does exercising in the morning give you more energy during the day?
- Do students learn better in smaller classes?
- Did the hook-up happen because alcohol was involved?
- Come up with your own!

Intervention

- Causality is tied to an action (or manipulation, treatment, or intervention)
- "no causation without manipulation" – manipulation need not be performed, but should be theoretically possible
- Treatments must be plausible as a (perhaps hypothetical) intervention
- Gender? Age? Race?

Not Clearly Defined Causal Questions

- Are parents more conservative then their children because they are older?
- What is the causal effect of majoring in statistical science on future income?
- Did she get hired because she is female?
- Can you make these into well-defined causal questions?

Potential Outcomes

- **Key question: what *would* have happened, under the opposite treatment?**
- A potential outcome is the value of the outcome variable for a given value of the treatment
- Outcome variable: Y
- Y(treatment): outcome under treatment
- Y(control): outcome under control

Potential Outcomes

- Y(organic) = cancer or no cancer
- Y(non-organic) = cancer or no cancer
- Possibilities:

Y(organic) = no cancer, Y(non-organic) = cancer
Y(organic) = Y(non-organic) = cancer
Y(organic) = Y(non-organic) = no cancer
Y(organic) = cancer, Y(non-organic) = no cancer

- Formulate the potential outcomes for your example

Causal Effect

- The causal effect is the comparison of the potential outcome under treatment to the potential outcome under control
- For quantitative outcomes, we often take a difference:

Causal Effect = Y(treatment) – Y(control)

Causal Effect

- Possibility 1:

Y(organic) = no cancer, Y(non-organic) = cancer

Causal effect for this individual: choosing organic produce prevents cancer, which he otherwise would have gotten eating non-organic

- Possibility 2:

Y(organic) = cancer, Y(non-organic) = cancer

Causal effect for this individual: he would get cancer regardless, so no causal effect

Causal Effect

- Y = test score
  - Y(study) – Y(don't study)
  - Example: Y(study) = 90, Y(don't study) = 60
  - Causal effect = 90 – 60 = 30. Studying causes a 30 point gain in test score for this unit.
- Y = accident (y/n)
  - Y(texting) vs Y(not texting)
- Y = energy during the day
  - Y(morning exercise) – Y(no morning exercise)
- Y = hook-up (y/n)
  - Y(alcohol) vs Y(no alcohol)

Units

- These are unit-level causal effect
- Unit: the person, place, or thing upon which the treatment will operate, *at a particular point in time*
- Note: the same person at different points in time = different units
- The causal effect will probably not be the same for each unit

Causal Effects

- The definition of a causal effect does not depend on which treatment is observed
- Causal effects, in their definition, do not relate to probability distribution for subjects "who got different treatments", or to coefficients of models
- Not a before-after comparison. Potential outcomes are a "what-if" comparison.
- Potential outcomes may depend on other variables (not just the treatment)

Causal Effects

- **Fundamental problem in estimating causal effects: at most one potential outcome observed for each unit**
- The other potential outcome lies in an unobserved counterfactual land… what *would* have happened, under a different treatment

For treated units:
Y(treatment) = observed, Y(control) = ???

For control units:
Y(treatment) = ???, Y(control) = observed

Estimation

- For the definition of a causal effect: compare potential outcomes for a single unit
- However, in reality, we can only see *one* potential outcome for each unit
- For estimation of a causal effect we will need to consider multiple units, some who have been exposed to the treatment, and some to the control

Multiple Units

- Need to compare similar units, some exposed to active treatment, and some to control
- Can be same units at different points in time, or different units at the same point in time
- In this class we will focus primarily on comparing different (but similar) units at the same point in time

Summary

- Causality is tied to an action (treatment)
- Potential outcomes represent the outcome for each unit under treatment and control
- A causal effect compares the potential outcome under treatment to the potential outcome under control for each unit
- In reality, only one potential outcome observed for each unit, so need multiple units to estimate causal effects