ECON 340
Economic Research Methods

Div Bhagia

Lecture 5
Research Questions and Data

Discussion: Chapter 1, Stock and Watson

1. Does reducing class size improve elementary school education?

- Using data on 420 California school districts in 1999, we find that *students in districts with small class sizes tend to perform better on standardized tests.*
- Can we conclude that small class sizes → better test scores?
- What could be some *confounding* factors?

Discussion: Chapter 1, Stock and Watson

2. Is there racial discrimination in the market for home loans?

- Researchers at the Federal Reserve Bank found that 28% of black applicants are denied mortgages, while only 9% of white applicants are denied.
- Does this indicate there is racial bias in mortgage lending?
- What could be some *confounding* factors?

Discussion: Chapter 1, Stock and Watson

3. Does healthcare spending improve health outcomes?

   - What are the two challenges identified in the reading when trying to answer this question using cross-country data on healthcare expenditures and mortality rates?

Discussion: Chapter 1, Stock and Watson

4. By how much will US GDP grow next year?

   - How is this question different from the first three?

Causal Questions

- Causality: specific action leads to specific, measurable consequences
- The first three questions we discussed are causal
- More examples:
  - Does sleep affect productivity?
  - Will the Fed's interest rate hike lower inflation?
  - Do higher minimum wages decrease employment?

Prediction

- Prediction: using the information on some variables to predict the value of another variable
- You do not need to know a causal relationship to make a good prediction
- A good way to “predict” whether it is raining is to observe whether pedestrians are using umbrellas, but using an umbrella does not cause it to rain.
- Forecasting: predictions about the future

Where are we headed?

- Conceptual framework we will build up to in this class—multiple regression model
- Multiple regression model can be used to answer both types of questions
- The multiple regression model is very useful because it gives us a mathematical way to *quantify how a change in one variable affects another variable, holding other things constant*.
- You will be utilizing this model for your research project

Multiple Regression Model

For example, using the multiple regression model, we can answer questions such as:

> What effect does a change in class size have on test scores, holding constant or *controlling* for student characteristics (such as family income)?
>
> What effect does your race have on your chances of having a mortgage application granted, holding constant other factors such as your ability to repay the loan?

Dependent vs Independent Variable

- The outcome variable is often called the *dependent variable*
- The variable(s) that affect the dependent variable are called *independent variable*(s)
- Other variables that might confound the effect of an independent variable on the dependent variable are called *control variables*

Your Research Project

- Pick a question that can be answered using one of the datasets compiled for this class or an external dataset
- The question should be *well-defined* and *feasible* using the data you picked

Not well-defined: *Are smaller classes better?*

Well-defined: *Does a smaller class size lead to better scores on standardized tests?*

Your Research Project

Say, you picked this question: *Does a smaller class size lead to better scores on standardized tests?*

- Identify your dependent variable
- Identify your independent variable
- Identify some other variables in the dataset that are potential confounders, which will be your control variables.

If all the needed variables are available in your dataset, your question is feasible.

Types of Data

Experimental versus observational (mostly what we will use)

- **Cross-Sectional:** many entities, single time period
- **Time Series:** single entity, multiple time periods
- **Panel/Longitudinal Data:** multiple entities, multiple time periods

Examples?

Establishing Causality

- As we have said, establishing causality is hard.
- Think about the following question:

  Does the use of electronic devices inhibit classroom learning?
- Say, I give you data from all classes held at CSUF in the last five years. The data contains average grades for each class and whether the instructor allowed electronic devices in the classroom.
- Could you answer the above question? If yes, how?

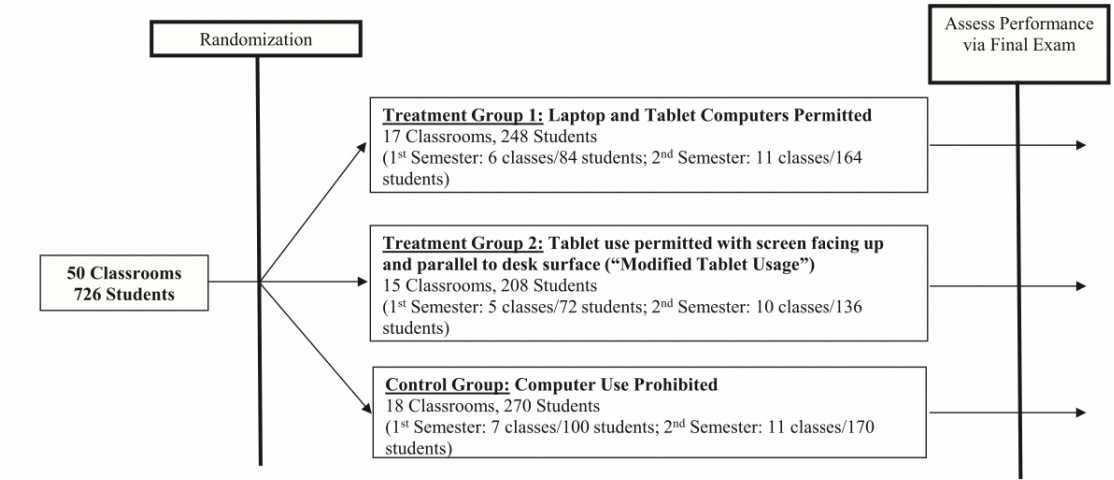
Establishing Causality

- We could compare grades for classes in which devices are permitted vs. others
- But what if better instructors don't care about students using devices?
- This would lead us to find little impact of device use, as classes where devices are allowed also happen to be taught by better instructors
- Here, the instructor quality is an example of a *confounding* factor

Establishing Causality

- How to establish causality?
  - Gold standard: experiments/randomized controlled trials (RCTs)
  - Assign randomly to *treatment* or *control* group
- Carter, Greenberg, and Walker (2017): randomly allowed students to access their laptop and tablet computers during an introductory economics course at the United States Military Academy at West Point

Carter, Greenberg, and Walker (2017)



Why does randomization help?

- Which classes allow electronic devices or not is random
- There should be no differences in instructor effectiveness because the classes were chosen randomly
- RCTs are widely employed in clinical research, e.g., experimental drug trials, other treatments
- Increasingly used in economics and other social sciences, e.g., PROGRESSA, free deworming program in Kenya (Miguel and Kremer, 2004), microcredits (Banerjee et al., 2013)

Aside: Experiments in Development Economics

- Development economists were early adopters of experiments in economics
- The 2019 Nobel prize in economics was awarded to Abhijit Banerjee, Esther Duflo, and Michael Kremer “for their experimental approach to alleviating global poverty.”
- (Duflo is the youngest person and the second woman to win the award.)
- If you are intrigued, check out Poor Economics by Banerjee and Duflo

But, we can't all run experiments...

- Hard to conduct experiments for a lot of Economics questions
- Economists have developed a whole toolkit of *quasi-experimental* methods. This is the heart of Econometrics.
- Starting point: multiple regression model that allows us to *control* for variables
- Focus on being able to come as close as possible to an idealized experiment

Another aside

- The 2021 Nobel prize in economics was awarded to David Card, Joshua D. Angrist, and Guido W. Imbens for their contributions to answering causal questions using *natural experiments*
- In the last two decades, *quasi-experimental* methods have been used to quantify the labor market effects of minimum wages, immigration, and education, amongst other questions.
- We will talk more about this towards the end of the semester.

Things To Do Next

- For the rest of this week and next week, we will learn how to prepare and analyze data in R
- Install R and R Studio on your computer if you haven't (how to handout on Canvas)
- Download the dataset "caschool.csv" from our Dropbox data folder and save it in a new folder called "Econ340_R" on your laptop (don't forget the location of the folder).
- Bring your laptop to the next class. Alternatively, you can use the computers in the lab. In the latter case, try to be here 5 minutes before class to set up.