POLI 171: Making Policy with Data

Winter, 2019

Instructor: Yiqing Xu

- *Email:* yiqingxu@ucsd.edu
- *Office:* SSB 377
- *Office Hours*: Thursday, 3:30-5:30pm

Teaching Assistant: Charles McClean

- *Email:* cmcclean@ad.ucsd.edu
- *Office:* SSB 447
- *Office Hours*: Tuesday, noon-2:00pm

Course Description

This course explores how we can make policy recommendations using data. The overall goal of the course is to provide a survey of the most commonly-used empirical tools for political science and public policy research. Our focus is design-based causal inference, or the use of statistical methods to answer research questions that concern the impact of some cause (e.g., an intervention, a change in institutions, passage of a law, changes in economic conditions, or policies) on a certain outcome (e.g., vote choice, income, election results, levels of violence, political attitudes). We cover a variety of causal inference designs and methods, including experiments, regression, matching, and difference-in-differences. We will analyze the strengths and weaknesses of these methods using applications from the real world.

The objectives of this course include:

1. Introducing an analytical framework of policy evaluation
2. Surveying the most commonly-used research designs for policy making
3. Introducing the most basic (and some of the most important) statistical concepts
4. Providing basic data analytical skills crucial for today's job market and academic research, including basic R programming

This course meets twice a week for the ten weeks of the quarter. 15 of the class days will be devoted to lecture, and the other 5 will be R lab sessions that will take place in our normal classroom. Each meeting will begin with a policy brief that will demonstrate how the methods covered in this course can be used to evaluate real-world policies. You are expected to attend all lectures and lab sessions, and you are strongly encouraged to bring your laptop (with R installed) to the four lab sessions.

Evaluation

- Four Problem Sets (15% each; 60% total). Posted on TritonEd one week before the due date; due at the beginning of lecture. All homeworks must be typed and submitted as a hardcopy.
  - HW1 will consist of basic R programming exercises, and HW2 will cover the potential outcomes framework and simple randomized experiments. You should work independently on both assignments; **do not collaborate with any other students on these assignments.**
  - HW3 and HW4 will consist of slightly more complicated data analysis problems in R. **You may collaborate with a single other student on these assignments**, or if you prefer, you may work independently. If you collaborate, you and your coauthor will turn in a single document, and you will each receive the same grade for the assignment.
  - Late submission will be penalized (a day = 2 points; in other words, if your submission is late for 8 days, you will receive 0 points for that one).
- Final exam (25%). March 21 3:00-4:30 PM. Closed book. Location TBD.
- Class participation using i>Clickers (15%).
- Optional Extra Credit Opportunity (5%). Posted February 5; due March 14.

Participation (with i>Clickers)

You are expected to come to class meetings prepared to discuss central questions, puzzles, and concerns that arise from course readings assigned for that day. Evidence-based research on teaching and learning has documented a strong causal relationship between active participation/discussion and student learning. The risk of large courses like ours is that students miss out on the opportunity to meaningfully discuss course materials, and thus learn less. For this reason, I will use clickers.

1. Official counting period. We will begin experimenting with clickers during the first two weeks, but the "official" counting period will not begin until Week 3 (Jan 22). This should give you time to find a clicker to borrow or purchase.
2. Type of questions. In general, we will ask two types of questions: (1) factual questions and (2) discussion questions. Factual questions focus on a central point from your readings, or a point covered in lectures. Discussion questions ask that you take a stand on a particular problem or issue using course materials as evidence.
3. Grading.
   - Factual questions. One point is given for correct answers, and 0.7 for participating.
   - Discussion questions. You will receive full points (1 point) simply for participating.

- "Grace points." In assessing your grade for this component of the course, you're allowed to miss 10% of all the questions asked throughout the class. This should provide sufficient buffer in case you forget your clicker, or you need to miss class for whatever reason. For example, if we ask 40 questions total over the quarter, and you receive 35 points, you can still earn an "A" for this part — 10% of 40 is 4, and (35 + 4)/40 = 97.5%. If you receive all 40 points, you will get (40 + 4)/40 = 110%. In other words, your grade for this part can be as high as 16.5 points.
- One-time exemption. Throughout this quarter, you will have *one* chance of not participating using your iClicker, either because you cannot physically come to class that day or because you forget to bring your iClicker or it does not function properly – in case that happens, please write to Charlie immediately after class. Your score on that day will be the average score of the entire class. We will not accept any other excuses or complaints.
- Reporting. You will be able to see your iClicker response records on TritonEd throughout the quarter (they may be lagged for a couple of days).
- No iClicker questions will be asked during R lab sessions.

Expectations and Policies

- If you seek a re-grade, you must email the TA within 24 hours of the assignment being returned to the class, and explain – in that email and in detail – why you believe you deserve reconsideration. The TA then has the ability to review the entire assignment, and he has the authority to increase your grade, decrease your grade, or keep the grade unchanged.
- We expect you to attend all lectures and labs.
- There is no prerequisite for this course. However, ECON5/POLI5D and POLI 30 (or the equivalent) is strongly recommended.
- This course covers graduate-level concepts with undergraduate-level math. The focus is on the intuition, and the math will not be particularly difficult. If you passed POLI 30, you should have no problem doing well in this course.
- Prior experience with R programming will be useful, but not required. If you haven't used R before, you should expect a steep learning curve. Although we will cover the basics during our four lab sessions, most of you will master R programming only through doing it yourselves and by learning from each other.
- If you choose to collaborate with another student on HW3 or HW4, we expect that you and your coauthor will collaborate on every portion of the assignment. You should understand (and be prepared to explain) every part of the document that you turn in.
- All lecture slides and assignments will be available on this course's Piazza page.

Books

Required Textbooks

1. Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, Christel M. J. Vermeersch. *Impact Evaluation in Practice.* The World Bank. Available here.

2. Imai, Kosuke. *Quantitative Social Science: An Introduction*, by Kosuke Imai, Princeton University Press. Amazon.

Optional Textbooks

1. Peng, Roger D. *R Programming for Data Science.* A Leanpub Book. For sale here.

2. Angrist, Joshua D. and Jorn-Steffen Pischke. *Mastering Metrics: The Path from Cause to Effect.* Princeton University Press.

3. Angrist, Joshua D. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press. Available as an ebook through Roger.

Piazza

Throughout this class we will use the Piazza online discussion board. This is a question-and-answer platform that is easy to use and designed to get you answers to questions quickly. It supports LaTeX, code formatting, embedding of images, and attaching of files. We encourage you to ask questions on the Piazza forum for clarifications, questions about concepts, or about your projects in addition to attending recitation sessions and office hours. You can sign up to the Piazza course page either directly from the below address (there are also free Piazza apps for the iPhone, iPad, and Android phones):

`https://piazza.com/ucsd/winter2019/poli171`

Using Piazza will allow students to see and learn from other students' questions. Both the TA and the instructor will regularly check the board and answer questions posted, although everyone else is also encouraged to contribute to the discussion. A student's respectful and constructive participation on the forum will count toward his/her class participation grade. *Do not email your questions directly to the instructors or TAs* (unless they are of personal nature) —we will not be answering your questions regarding course materials or problem sets through email.

Computation

The labs and homework assignments of this course will use R, an open-source computing language that is very widely used in statistics and the social sciences.

- R runs on a wide variety of UNIX platforms, Windows and MacOS. R makes programming very easy, has strong graphical capabilities, and also contains canned functions for most commonly-used estimators.

- You can download R for free here. Select the link that matches your machine's platform. If you are using Windows, click on the "install R for the first time" link, and then click on the link for the latest version of R. If you are using MacOS, scroll down and find the version of R that matches your version of MacOS (Unless you are running an old version of OS X, you probably want "R-3.4.3.pkg," the latest version of R. You can check your version of MacOS by clicking on the apple icon and then on the "About This Mac" option.).
- RStudio is an integrated development environment designed for R. It is possible to program in R without using RStudio, but RStudio makes R programming much easier, especially for beginners. You can download RStudio for free here. Choose the RStudio Desktop Open Source License option, and select the installer designed for your platform.
- A nice way to get started is with the two video tutorials provided by Dan Goldstein: Tutorial 1 and Tutorial 2
- Another good resource is the set of tutorials provided by DataCamp
- The web provides many other great tutorials and resources for learning R.
- We will cover the basic tools that you need for the homework assignments in lab sessions held in our normal classroom.

Course Schedule

Introduction

Reading: Gertler, Chapter 1, pp. 3-9

- January 8: Introduction and Course Overview
- January 10: **R Lab 1**: R Basics

Module I: Causality and Potential Outcomes

Reading: Gertler, Chapter 3

- January 15: The Potential Outcomes Framework
- January 17: Omitted Variable Bias and Selection Into Treatment

Module II: Randomized Control Trials (RCT)

Reading: Gertler, Chapter 4, pp. 49-69

- January 22: Randomization and Experiments
  - **HW1 Due**
- January 24: **R Lab 2**: RCT 1
- January 29: Inference and Experiments
- January 31: What Can Go Wrong? Attrition, Noncompliance, and Spillovers
- February 5: Clustering, Blocking and Multiple Treatment Arms
  - **HW2 Due**
- February 7: **R Lab 3**: RCT 2

Module III: Basic Observational Studies

Reading: Gertler, Chapter 7

- February 12: Selection on Observables
- February 14: Regression
- February 19: Matching
- February 21: **R Lab 4**: Regression and Matching

Module IV: Panel Methods

Reading: Gertler, Chapters 6

- February 26: Difference-in-Differences
  - **HW3 Due**
- February 28: Difference-in-Differences
- March 5: Synthetic Control Methods
  - **Extra Credit Assignment Announced**
- March 7: **R Lab 5**: Difference-in-Differences

Advanced Methods and Conclusion

No readings required.

- March 12: Advanced Methods
  - **HW4 Due**
- March 14: Summary and Conclusion
  - **Extra Credit Assignment Due**

Final Exam

- March 31: 3:00-4:30 PM. Closed Book. Location TBD.