POLI 171 Making Policy with Data

Spring 2017

Professor Yiqing Xu

Time: Tuesday/Thursday 9:30-10:50AM

Location: WLH 2111

Instructor: Yiqing Xu

- *Email*: yiqingxu@ucsd.edu
- *Office*: SSB 377
- *Office Hours:* Thursday 11:00AM-noon
  Friday 2:30-3:30PM

Teaching Assistant: Charles McClean

- *Email*: cmcclean@ad.ucsd.edu
- *Office*: SSB 346
- *Office Hours*: Tuesday 11:00AM-2:30PM

Course Description

This class explores how we can make policy recommendations using data. The overall goal of course is to provide a survey of most commonly used empirical tools for political science and public policy research. Our focus is *design-based causal inference*, that is, to use 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 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, difference-in-differences, and regression discontinuity designs. We will analyze the strengths and weaknesses of these methods using applications from the real world.

The objectives of this course include:

1. Introduce an analytical framework of policy evaluation
2. Survey the most commonly used research designs for policy making
3. Introduce the most basic (and some of the most important) statistical concepts

4. Provide basic data analytical skills crucial for today's job market and academic research, including basic R programming

Requirements and Grading

- Four Problem Sets (40%)

This is a course that is designed to help you develop skills in understanding and applying statistical methods. You can only learn statistics by doing statistics. The assignments consist of analytical problems, computer simulations, and data analysis.

- ✓ Every two weeks or so, we will have an in-class lab session to help you with programing and the problem sets.
- ✓ We encourage students to work together on the problem sets (2 people maximum), but each student needs to write his/her solutions. We also ask you to write the names or your co-worker on each assignment. *This rule will be strictly enforced.*
- ✓ No late homework will be accepted.
- ✓ All sufficiently attempted homework (i.e., a typed and well organized write-up with all problems attempted) will be graded on a (✓−, ✓, ✓+) scale.

- Midterm (20%) and Final (30%) Exams.

In class, closed book. They will include both analytical and conceptual questions on basic statistics, summarizing statistical findings from research articles, policy evaluation and research design.

- In-class quizzes and participation (10%).

Quizzes in lecture will count toward your grade. Participation will be graded on quality of the contributions and knowledge of the weekly readings.

Course Website

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 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 and iPad):

https://piazza.com/ucsd/spring2017/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.

Books

1. Angrist, Joshua D. and Jörn-Steffen Pischke. *Mastering 'Metrics: The Path from Cause to Effect*. Princeton University Press.
2. Peng, Roger D. *R Programming for Data Science*. A Leanpub Book (For sale at http://leanpub.com/rprogramming).
3. (Optional) Imai, Kosuke. *Quantitative Social Science: An Introduction*. Princeton University Press.
4. (Optional) Abedin, Jaynal and Hrishi V. Mittal. *R Graphs Cookbook*, Second Edition. PACKT Publishing.

Computation

We teach the course in R, which is an open-source computing language that is very widely used in statistics. You can download it for free from www.r-project.org. The web provides many great tutorials and resources to learn R. A list of these is provided at

http://scs.math.yorku.ca/index.php/R:_Getting_started_with_R.

A nice way to start you off are the two video tutorials provided by Dan Goldstein:

http://www.dangoldstein.com/flash/Rtutorial1/Rtutorial1.html

and http://www.dangoldstein.com/flash/Rtutorial2/Rtutorial2.html

Another good resource is tutorials provided by DataCamp (https://www.datacamp.com/).

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.

Besides Roger Peng's *R Programming for Data Science*, you can also check out one of the following free tutorials. All three tutorials cover similar material, just pick the one you like best:

1. W. J. Owen. The R Guide. https://cran.r-project.org/doc/contrib/Owen-TheRGuide.pdf
2. W. N. Venables and D. M. Smith. An Introduction to R. https://cran.r-project.org/doc/manuals/R-intro.pdf
3. J. Verzani. Simple R. https://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf

If you are very familiar with another statistical software (such as STATA, SPSS, or SAS), you may use it for the course as well, but we can only support R in class.

Academic Integrity

Students are expected to maintain the highest standards of academic integrity. Students are encouraged to work together on the assignments, but you always need to write your own

solutions, and we ask that you make a solo effort at all the problems before consulting others. We also ask that you write the names of your co-worker on your assignments—again, this rule will be strictly enforced and we will keep a record. Cheating, plagiarism and other forms of academic dishonesty will not be tolerated and will be subject to disciplinary action consistent with University rules and regulations. Students are expected to familiarize themselves with University regulations regarding plagiarism and academic dishonesty. For more information, please visit: https://academicintegrity.ucsd.edu/

Tentative Course Schedule

- **April 4, 6:** Course Introduction and Introduction to R
  Readings: Peng, "R Nuts and Bolts," pp. 12-22

- **April 11, 13:** Randomized Trials
  Readings: Angrist and Pischke, Chapter 1.

- **April 18, 20:** Random Variables and Sampling Distribution
  Readings: Peng, "Simulation," pp. 123-130.

- **April 25:** Making Inference
  Readings: Angrist and Pischke, Chapter 1 Appendix.

- **April 27:** Midterm Review

- **May 4:** Blocked Experiment and Matching
  Readings: Imbens, Guido W. "Matching Methods in Practice: Three Examples"

- **May 9, 11, 16:** Regression Recap
  Readings: Angrist and Pischke, Chapter 2.

- **May 18, 23:** Regression Discontinuity Design

  Readings: Angrist and Pischke Chapter 4.

- **May 25, 30 & June 1:** Difference-in-Differences Design

  Readings: Angrist and Pischke Chapter 5.

- **June 6:** (Optional Topic) The Synthetic Control Method

- **June 8:** Summary and Final Review

Exams

- **May 2:** Midterm Exam
- **June 15:** Final Exam

Lab Sessions

(Tentative Schedule)

- **April 6:** R refresher
- **April 20:** Randomized Trials
- **May 16:** Matching and Regression
- **June 1:** Regression Discontinuity and Difference-in-Differences