Syllabus

POLI 273 Causal Inference (Spring 2017)

Instructor: Professor Yiqing Xu

**Time & Room**
- Class: TH 3:00-4:20PM
- Recitation: F 10:00AM
- Room: SSB 104

**Office**
- Yiqing Xu
  - Room: SSB 377
  - Email: yiqingxu@ucsd.edu
  - Office Hours: Friday 1:30-3:30PM (please email first) or by appointment

- John Seungmin Kuk
  - Room: SSB 353
  - Email: jskuk@ucsd.edu
  - Office Hours: Monday 11AM–noon, Friday 9–10AM or by appointment

**Course Description**

This is the third course in the quantitative political methodology sequence at the Political Science Department at UCSD. The 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, matching, regression, fixed effects models, difference-in-differences, synthetic control methods, instrumental variable estimation, regression discontinuity designs, simple machine learning tools, and, if time permits, quantile regressions and sensitivity analysis. We will analyze the strengths and weaknesses of these methods using applications from various social sciences with a special focus on political science.

The class is open to qualified students from other departments and undergraduates, but priority will be given to graduate students in the Political Science Department.

**Prerequisites**

A willingness to work hard on possibly unfamiliar material is key. In addition to introductory statistics and probability, we assume that you have a reasonably good knowledge of linear regression (meaning that you probably should have taken at least one graduate class on this subject, such as POLI204B). Knowledge of the maximum likelihood method is useful, but not required—in fact, we are thinking about teaching POLI 273 prior to POLI 271 in the near future.

Students are also expected to be reasonably proficient in the statistical software R (you may use other software packages that you are very familiar with, but we can only support R). If you have some background in other programming languages, you can R fairly quickly. We will give you a chance to get familiar with R in the first few weeks.
Class Requirements

Grades will be based on a weighted average of your weekly assignments, the midterm and final projects, as well as your participation.

**Reading (for your own intellectual satisfactory)**

The syllabus lists the required readings for every week. This required reading should be completed prior to lecture in a given week. Students are expected to read the material carefully. You may even find it helpful to read the material multiple times.

**7 Problem Sets (50%)**

This is a methodological course, developing skills in understanding and applying statistical methods. You can only learn statistics by doing statistics and therefore homework for this course is extensive, including weekly assignments. The assignments consist of analytical problems, computer simulations, and data analysis.

- They will usually be assigned on **Tuesday** and due the following **Tuesday**, prior to lecture.
- 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.
- Only six highest-scored problem sets will be counted, but all seven need to be finished.

We encourage students 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-workers on your assignments [this rule will be strictly enforced and we will keep a record].

**Midterm (20%)**

The in-class, closed-book midterm will take place on May 9 (Tuesday) during the regular class time. There will be no assignment in the week before the midterm.

**Student Project (30%)**

A student project should be a short empirical paper (in the form of a blog post) that applies methods learned in this class to a research question of their choice. It needs to meet the following requirements:

I. **Data and methods (15%)**

   - Students can either collect their own data related to an empirical problem of their own interest, or use replication data from a published paper.
   - Replication works are encouraged: students are expected to go beyond the original analysis in some significant way both substantively and methodologically, for example, by either (1) collecting additional data or (2) applying techniques learned in the course to make significant improvements on the original paper.
   - At least one method taught in the class should be used.

III. **Presentation (5%)**

   - You will be presenting your work on June 8.
   - Bonus points will be given to best projects based on a secret polling among class members

II. **“Paper” (Blog Post) (10%)** – No more than 1500 words. It’s a blog post, so you don’t really have to follow any particular order. But if you don’t have much experience writing a blog, there’s some hint:

   - An interesting title
Collaboration: We encourage you to collaborate with another student (a group should not consist of more than 2 students). Note that most cutting-edge research is collaborative (see any recent issue of *APSR* or *AJPS*), and collaboration is more likely result in a good, potentially publishable paper (multiple brains are usually better than one). We expect higher quality with a coauthored paper. Same page limit applies.

Deadlines: Please be aware of the following deadlines. Late submission will be penalized (1% for each day).

- **May 16 (Project description):** By this date, you should email the instructor and TA what topic you plan to work on and how you plan to collect/obtain data (1 page).
- **June 8 (Project presentation):** Present your work in class.
- **June 13 (Blog post due):** By this date, you should email the instructor and TA your blog post. It should summarize the theoretical/empirical contributions, methods, and main results (figures and tables). The blog post should be sent to TA by midnight.

Recitation Sections

Weekly recitation sections will be held on Friday. The section will cover a review of the theoretical material and also provide help with computing issues. The TA will run the sections and can give more detail.

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 L\TeX, 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/poli273

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

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://wiki.math.yorku.ca/index.php/R:_Getting_started. A nice way to start you off are the two video tutorials provided by Dan Goldstein here and also here. Another good resource is the set of tutorials provided by DataCamp. 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.

To refresh your knowledge about R, you can check out one of the following free tutorials. All three tutorials cover similar material, just pick the one you like best:

2. W. N. Venables and D. M. Smith. An Introduction to R.
3. J. Verzani. Simple R.

For advanced R programmers, The R Inferno is highly recommended (it’s hilarious).

If you are very familiar with another statistical software package you may use that for the course at your own risk. We can only support R. Political science graduate students are strongly recommended to use R.

Office Hours and Availability

My office hours are by appointment and John will hold office hours at the listed times.

Books

Required Books

We will read chapters from the following textbooks:


Useful Summary Articles

The following papers summarize the main methods learned in this course. They are dense and detailed and you might not understand all of the details the first time you read through them. However, if you plan to conduct applied empirical work that involves causal inference, you should revisit these again and again as reference.


Optional Books for Your Future Reference

The following books are optional but may be useful for deeper understanding for some of the course topics and for your future reference.
• **Causal Inference**

• **Experiments**

• **Matching**

• **Panel Data**

**Tentative Course Outline**

Below is a preliminary schedule of course topics. Notice that required readings are marked with a (⋆).

1 **Introduction (Week 1)**

• Review of Statistical Concepts Useful for Causal Inference
  – Random Variables, Measures of Location and Dispersion
  – Inference and Properties of Estimators
  – Conditional Mean Function

• Overview, Course Requirements, Course Outline

2 **The Potential Outcome Framework (Week 1)**

• Counterfactual Responses and the Fundamental Identification Problem

• Estimands and Assignment Mechanisms

• Heterogeneity and Selection

**Readings**

• Angrist and Pischke: Chapter 1. (⋆)

• Morgan and Winship: Chapter 1-2. (⋆)


3 Randomized Experiments (Week 2)

3.1 Theory

• Identification of Causal Effects under Randomization
• Implementation, Estimation, Diagnostics, Blocking
• Attrition and Other Threats to Validity

Readings

• Angrist and Pischke: Chapter 2. (∗)

3.2 Statistical Inference

• Variance estimation under the Neyman model.
• Inference for clustered designs.
• Randomization inference.
• The bootstrap.
• Power analysis.

Readings

• Angrist and Pischke: Chapter 8. (∗)
• Rosenbaum, Paul R. 2009. *Design of Observational Studies*: pgs. 29-49
3.3 Applications

Readings: Experiments


Readings: Natural Experiments

• Hyde, Susan D. 2007. The Observer Effect in International Politics: Evidence from a Natural Experiment. *World Politics* 60(1): 37-63. (⋆)


Readings: Review Articles


• List, John A., and Steven Levitt. 2006. What Do Laboratory Experiments Tell Us About the Real World? University of Chicago and NBER.


Readings: Methodological Guides


4 Selection on Observables (Weeks 3 and 4)

4.1 Theory

• Identification under Selection on Observables

• Subclassification

Readings

• Morgan and Winship: Chapter 3. (⋆)


• Rosenbaum, P. R. 1984. The Consequences of Adjustment for a Concomitant Variable That Has Been Affected by the Treatment. Journal of the Royal Statistical Society. Series a (General), 147(5), 656-666.

4.2 Regression Recap

• Identification with Regression

• Non-parametric Regression

Readings

• Angrist and Pischke: Chapter 3. (⋆)

• Morgan and Winship: Chapter 5. (⋆)


4.3 Matching Methods

- Covariate Matching, Balance Checks
- Properties of Matching Estimators
- Inference

Readings: Theory

- Morgan and Winship: Chapter 4. (⋆)
- Rubin: Chapters 3 to 5.

Readings: Applications


4.4 Propensity Score and Weighting Methods

- Identification, Propensity Score Estimation, Matching on the Propensity Score, Weighting on the Propensity Score, Reweighting methods

Readings: Propensity Score Methods Theory
• Morgan and Winship: Chapter 3. (∗)
• Rubin: Chapters 10, 11 and 14 (all with Paul R. Rosenbaum).


Readings: Propensity Score Methods Applications


4.5 Summary

• Can Non-Experimental Method Recover Causal Effects?

Readings: Comparison of Experimental and Non-experimental Methods


5 Cross-Sectional Research Designs (Weeks 5 and 6)

5.1 Instrumental Variables

• Identification: Using Exogenous Variation in Treatment Intake Given by Instruments

• Imperfect Compliance in Randomized Studies
• Wald Estimator, Local Average Treatment Effects, 2SLS

Readings: Instrumental Variable Theory

• Angrist and Pischke: Chapter 4 (∗)
• Morgan and Winship: Chapter 7

Readings: Instrumental Variable Critiques


Readings: Instrumental Variable Applications

• Angrist and Krueger. 2001 Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments

5.2 The Regression Discontinuity Design

• Sharp and Fuzzy Designs, Identification, Estimation, Falsification Checks

Readings: RDD Theory

• Angrist and Pischke: Chapter 6 (∗)

**Readings:** RDD Applications


6 Longitudinal Research Designs (Weeks 7 and 8)

6.1 Difference-in-Differences Estimators

• Identification, Estimation, Falsification tests

**Readings:** DID Theory

• Angrist and Pischke: Chapter 5.2-5.4 (∗)

**Readings:** DID Applications


6.2 Fixed Effects and Random Effects Models

• Fixed Effects and Random Effects Estimation

**Readings:** Panel Methods Theory

• Angrist and Pischke: Chapter 5.1 (∗)

Readings: Panel Methods Applications


6.3 Synthetic Control Methods

Readings


7 Machine Learning for Causal Inference (Week 9)

• Double Selection

• Heterogeneous Treatment Effects

• Predictive Models and Causal Inference

Readings


8 Additional Topics

8.1 Sensitivity Analysis

• Nonparametric Bounds, Formal sensitivity tests

Readings

• Morgan and Winship: Chapter 6 (∗)


8.2 External Validity

• Sample Selection, Generalizability, Randomization Bias

Readings


8.3 Attrition

- Missing Data, Attrition, Sample Selection, Truncation by Death

Readings


8.4 Distributional Effects

- Quantile Regression
- Nonlinear difference-in-difference

Readings

- Angrist and Pischke: Chapter 7 (∗)

Acknowledgment

I learned tremulously from my teachers and friends. This course is built upon, or borrows from, course materials prepared by Josh Angrist (MIT 14.387), Jens Hainmueller (MIT 17.802), Chad Hazlett (UCLA PS200D), Danny Hildago (MIT 17.802), In Song Kim (MIT 17.804), Gary King (Harvard Gov2001), Guido Imbens (Harvard Econ2140).

May, 2017