17.802 Quantitative Research Methods II
Spring 2015

Instructor: Teppei Yamamoto
TAs: Daniel de Kadt, James Dunham

Time & Room

Class: Tuesdays and Thursdays, 3:00 – 4:30pm, Room 66–144
Recitation: Fridays, 9:30 – 10:30am, Room E53–438

Office

Teppei Yamamoto
Room: E53–401
Email: teppei@mit.edu
Office Hours: By Appointment

Daniel de Kadt
Room: E53–434
Email: ddekadt@mit.edu
Office Hour: Fridays, 2:30 – 3:30pm

James Dunham
Room: E53–408
Email: jdunham@mit.edu
Office Hour: Mondays, 4:30 – 5:30pm in E53-485

Overview and Class Goals

This is the second course in the quantitative research methods sequence at the MIT political science department. The goal of the four-course sequence is to teach you how to understand and confidently apply a variety of statistical methods and research designs that are essential for political science research.

Building on the first course (17.800) which covered probability, statistics, and linear regression analysis, this second class provides a survey of more advanced empirical tools, with a particular focus on causal inference. We cover a variety of research designs and statistical methods for causal inference, including experiments, matching, regression, panel methods, difference-in-differences, synthetic control methods, instrumental variable estimation, regression discontinuity designs, causal mediation analysis, nonparametric bounds, and sensitivity analysis. We will analyze the strengths and weaknesses of these methods. Applications are drawn from various fields including political science, public policy, economics, and sociology.

The class is open to qualified students from other departments and undergraduates. However, the enrollment will be capped at 30 and priority will be given to graduate students in the political science department in the event of excess demand.

Prerequisites

There are three prerequisites for this course:

2. Probability and statistics covered in 17.800 or an equivalent graduate-level course.
3. Computing: familiarity with at least one statistical software. We will use R in this course (more on this below).

For 1 and 3, we expect the level of background knowledge and skills equivalent to what is covered in the department’s Math Camp and 17.800. For more information about the Math Camp see:

https://stellar.mit.edu/S/project/mathrefresher/index.html
Course Requirements

The final grades are based on the following items:

- **Problem sets** (40%): You can only learn statistics by doing statistics. Therefore, the homework for this course is extensive, including weekly homework assignments. The assignments consist of analytical, computational, and data analysis questions. They will usually be assigned on Tuesday night and due the following Tuesday, prior to lecture. Each problem set will be counted equally toward the calculation of the final grade. All sufficiently attempted assignments (i.e. a typed and well organized write-up with all problems attempted) will be graded on a three-point scale ($\checkmark^{+}, \checkmark, \checkmark^{-}$). The following additional notes will apply to all problem sets unless otherwise noted.

  - No late submission will be accepted, unless you ask for special permission from the instructor in advance of the deadline. (Permission may be granted or not granted, with or without penalty, depending on the specific circumstances.)

  - 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. In particular, you must not simply copy and paste someone else’s answers or computer code. **Violation of this policy will be considered an academic integrity issue and processed accordingly to MIT’s rules and procedures for such violations.** We also ask that you write the names of your co-workers on your assignments.

  - For analytical questions, you should include your intermediate steps, as well as comments on those steps when appropriate. For data analysis questions, include annotated code as part of your answers. All results should be presented so that they can be easily understood.

- **Quizzes** (15%): Three in-class, closed-book quizzes will take place on Thursdays (March 5, April 2 and April 30) during the regular class time.

- **Project** (35%): The final project will be a short research paper which typically applies a method learned in this course to an empirical problem of your substantive interest. The paper should be 5-10 pages in length and focus on the research question, data, empirical strategy, results, and conclusions. Literature reviews, background, lengthy theoretical motivations, etc. should be omitted or may be included as an appendix. You also need to submit a copy of your analysis code. Students are free to choose any topic they want, as long as they have a clear research question that concerns causality. Projects co-authored with another student are **very strongly encouraged** (learning to co-author is essential because nowadays most articles in political science are co-authored). Replication papers are accepted as long as they go beyond the original analysis in some significant way by applying techniques learned in the course.

Students need to meet the following milestones for their project:

  - February to early March: **Start** thinking about possible topics, exploring data sources, and running simple analyses on acquired data sets. Run your ideas by the TAs and instructor during their office hours and after classes/recitations and obtain their reactions.

  - March 12: Turn in a **brief description of your proposed project**. By this date you need to have found your coauthor, acquired the data you plan to use, and completed a descriptive analysis of the data (e.g. simple summary statistics, crosstabs and plots). Schedule a brief meeting with the instructor to discuss your proposal during office hours. You may be asked to revise and resubmit the proposal.

  - May 7 and 12: Students will give **presentations** during the regular class time. Presentations should be approximately 10 minutes in length (determined based on the class size, but time limits will be strictly enforced) and will be oral accompanied by electronic slides, much like presentations at major academic conferences such as APSA and MPSA. Performance will be counted toward the class participation grade (see below).
– May 14: **Paper due.** Turn in the final version of your paper by the end of the day.

- **Participation and presentation** (10%): Students are strongly encouraged to ask questions and actively participate in discussions during lectures and recitation sessions.

In addition, the syllabus lists **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 very carefully. You may even find it helpful to read the material multiple times. The syllabus also lists suggested reading; once you have decided on a focus for your project, you should consider the relevant suggested readings very closely.

**Recitation Sections**
Weekly recitation sections will be held on Fridays, 9:30-10:30am. The section will cover a review of the theoretical material and also provide help with computing issues. The TAs will run the sections and can give more detail.

**Computation**
We teach the course in **R**, which is an open-source statistical computing environment that is very widely used in statistics and political science. You can download it for free from [www.r-project.org](http://www.r-project.org). The web provides many great tutorials and resources to learn R. A list of these is provided [here](http://www.r-project.org). A nice way to start you off are the two video tutorials provided by Dan Goldstein [here](http://www.r-project.org) and [also here](http://www.r-project.org). Another good resource is the set of tutorials provided by DataCamp.

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.

**Course Website**
The course website is located at:


It provides homework assignments, datasets, and supplementary materials.

**Course Forums**
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 or the link posted on the Stellar course website (there are also free Piazza apps for the iPhone and iPad):


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 answer them!

**Schedule**
Please note the following scheduling issues:

- No class on February 17 (Monday schedule).
- No class on March 24 and 26 (Spring Break).
- No class on April 21 (Patriots Day).
Books

Main Books: We will read chapters from the following textbooks.


Optional Books and Summary Articles: The following books and review articles are optional but may prove useful for additional coverage of some of the course topics. We read selected chapters from some of these books, but you need not purchase them.


Course Schedule

Required readings are marked with a (⋆) and are in bold.

1 Introduction

- Overview, course requirements, course outline

2 Statistical Models for Causal Analysis

- Causality as counterfactuals
- Potential outcomes
- The Fundamental Problem of Causal Inference
- Identification and estimation
- Causal estimands
- Interference
- Causal graphs and other causal models
• Sufficient component causes

Readings: Basics
• Morgan and Winship: Chapters 1, 2 and 3. (⋆)
• Angrist and Pischke: Chapter 1. (⋆)

Readings: Potential Outcomes

Readings: Causal Graphs

Readings: Alternative Causal Models

3 Randomized Experiments

3.1 Identification and Estimation
• Identification of Causal Effects under Randomization
• Covariate adjustment
• Blocking
• Practical considerations

Readings: Theory
• Angrist and Pischke: Chapter 2. (⋆)
• Gerber and Green: Chapters 2, 3 and 4. (⋆)

Readings: Field Experiments


**Readings: Natural Experiments**


**Readings: Non-technical Overviews**


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


**Readings: Implementation and Practical Guides**


• MIT Committee on the Use of Humans as Experimental Subjects (COUHES) http://web.mit.edu/committees/couhes/.
3.2 Inference

- Variance estimation under the Neyman model
- Clustered designs
- Randomization inference
- Bootstrap
- Power analysis

Readings: Theory

- Angrist and Pischke: Chapter 8.1 (⋆)

Readings: Application


4 Observational Studies

4.1 Identification

- Selection on observables
- Post-treatment bias
- Subclassification

Readings

- Morgan and Winship: Chapter 4. (⋆)
4.2 Matching and Weighting

- Covariate matching
- Balance checking
- Propensity scores

Readings: Theory

- Morgan and Winship: Chapter 5. (*)


Readings: Applications


4.3 Regression

• OLS as an estimator of causal effects

Readings

• Angrist and Pischke: Chapter 3. (⋆)
• Morgan and Winship: Chapters 6 and 7. (⋆)

4.4 Partial Identification and Sensitivity Analysis

• Nonparametric bounds
• Sensitivity analysis

Readings: Theory

• Morgan and Winship: Chapter 12 (⋆)

Readings: Applications


Readings: Comparison of Experimental and Observational Studies


5 Instrumental Variables

• Structural equation models
• Two-stage least squares
• Treatment noncompliance
• Principal stratification
• Local average treatment effects

Readings: Theory

• Angrist and Pischke: Chapter 4 (+)
• Morgan and Winship: Chapter 9 (+)

Readings: Critiques


Readings: Applications

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

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

Readings: Theory

- Angrist and Pischke: Chapter 6 (*)

Readings: Applications


7 Fixed Effects and Difference in Differences

- Selection on time-invariant unobservables

Readings: Theory

- Angrist and Pischke: Chapter 5 (*)

Readings: Fixed Effects Applications

Readings: Difference in Differences Applications


8 Synthetic Control Methods

Readings


9 Causal Mechanisms

- Direct and indirect effects
- Sequential ignorability
- Sensitivity analysis and research designs

Readings

- **Imai, K., L. Keele, D. Tingley and T. Yamamoto. 2011.** *Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies.* American Political Science Review, 105(4), 765-789. (⋆)