17.802 Quantitative Research Methods II

Spring 2020

MIT

Class Time: T&R 9:30 - 11 AM
Recitation Time: Fridays 9 - 10 AM
Class Room: E51–057
Recitation Room: E51–057

<table>
<thead>
<tr>
<th>Instructor</th>
<th>TA</th>
<th>TA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name: F. Daniel Hidalgo</td>
<td>Stuart Russell</td>
<td>Nicole Wilson</td>
</tr>
<tr>
<td>Email: <a href="mailto:dhidalgo@mit.edu">dhidalgo@mit.edu</a></td>
<td><a href="mailto:sruss@mit.edu">sruss@mit.edu</a></td>
<td><a href="mailto:newilson@mit.edu">newilson@mit.edu</a></td>
</tr>
<tr>
<td>Office Hours:</td>
<td>Thursday 3:00-5:00</td>
<td>Wednesday 3:00-4:30</td>
</tr>
</tbody>
</table>

* Class will not meet on: Feb 18 (Monday schedule), Mar 24 and 26 (spring break).

Purpose and 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 \( R \) (see additional notes on computation 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/mathprefresher/index.html

**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 Thursday after class and due the following Thursday, prior to lecture. Each problem set will be graded on a check +/- scale and will count equally toward the calculation of the final grade. 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.

  - Regardless of the grade you receive, you should go through your returned problem sets and read all the comments made by the TAs. Learning from your own mistakes is often the best way to accumulate knowledge and skills efficiently. Even the very best answers to typical problem sets contain several errors from which you can learn a lot. We will also post detailed example solutions on Stellar for each problem set; make sure to go through them as well.

  - To encourage you to understand the problem sets which you have had trouble with, you are allowed to redo *one* problem set for a regrade. This redo must be submitted within one week of the receipt of the graded problem set (e.g. if the problem set were due 2/14, the grade would be received on 2/21, and the redo would be due 2/28). Solutions *must be in your own language and annotated code* (i.e. do not just copy and paste from the solution set). Moreover, this redo may only be used for a problem set if a good faith effort was made on the problem set originally (as determined by the grader).

- **Quizzes** (15%): Three in-class, closed-book, 30 minute quizzes will take place on Tuesday March 3, Tuesday March 31, and Tuesday April 28 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 around 10 pages in length and look like an empirical journal article minus literature review and lengthy theoretical motivations. That is, the paper should start with a concise statement of your research question, followed by description of data, empirical strategy, results, and conclusions. 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 generally encouraged. However, you should be mindful of the solo-authorship requirement for your second-year paper, if you are a first-year Ph.D. student in political science and you intend to use your project as a basis for your second year paper. 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. To guide your thoughts, we will post a short list of readings that exemplify empirical studies using the main research designs and statistical methods covered in the course. You are encouraged to skim the listed articles to get the sense of what these methods are and whether they will be useful for answering empirical questions of your interest. Once you think you have a promising idea, go ahead and read more on the methods from the full reading list provided at the end of this syllabus. You should also run your ideas by the TAs and instructor during their office hours and after classes/recitations to obtain their reactions.

- **March 19:** Turn in a brief description of your proposed project. By this date you need to have acquired the data you plan to use and completed a descriptive analysis of the data (e.g. simple summary statistics, crosstabs and plots). Your proposal should be no more than 2 pages, with tables and graphs included in an appendix.

- **Late March to early April:** Meet with the instructor to discuss your proposal. We will set up a Doodle poll to assign you to a 20-minute meeting slot. You may be asked to revise the proposal and resubmit within one to two weeks of the meeting.

- **May 5, 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 20:** 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 readings; once you have decided on a focus for your project, you should consider the relevant suggested readings very closely.

**Recitation Sessions**

Recitation sessions will be held in E51–057 on Fridays at 9 AM. Sessions will cover various topics, including review of lecture material, hints on problem set questions, and help with computing issues. The TAs will run the sessions and can give more details. Attendance is very strongly encouraged.
Course Website

You can find the Stellar/LMOD website for this course at:

http://stellar.mit.edu/S/course/17/sp20/17.802/

We will distribute course materials, including readings, lecture slides and problem sets, on this website.

Questions about Course Materials

In this course, we will utilize an online discussion board called Piazza. This is a question-and-answer platform that is easy to use and designed to get you answers to questions quickly. We encourage you to use the Piazza Q&A board when asking questions about lectures, problem sets, and other class materials outside of 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 Android and iOS devices):

piazza.com/mit/spring2020/17802/home

Using Piazza will allow you to see and learn from questions others have. Both the TAs and the instructor will check the board at regular times each day and answer questions posted, but everyone else is also encouraged to contribute to the discussion. Your respectful and constructive participation on the forum will count toward your class participation grade. Finally, please do not email your questions directly to the instructors or TAs (unless they are of a personal nature) — we will not answer them!

Books

- **Required books:** We will read chapters from the following books, which we strongly recommend that you purchase (they are relatively cheap; about $100 total). The books can be purchased at online bookstores (e.g. Amazon) and are generally available in the library.
  
  
  

Additionally, we will assign several book chapters and journal articles as required readings (see the reading list below). We will post either their scanned copies or links to electronic versions on Stellar.

- **Recommended books:** These books and review articles cover particular sections of the course more in depth and are recommended for your reference, particularly if the sections are directly relevant for your final project.


  - Pearl, Judea, Madelyn Glymour, and Nicholas P. Jewell. *Causal Inferene in Statistics: A Primer*


Computation

We teach this course in R, an open-source statistical computing environment that is very widely used in statistics and political science. We assume some familiarity with R and we cannot provide introductions to the programming language. That said, we are happy to point to R resources and provide advice on how to improve your programming skills.

You can download it for free from www.r-project.org. We recommend that you use the RStudio IDE (integrated development environment) to work with R, which can be downloaded for free from http://www.rstudio.com.

The web provides many great tutorials and resources to learn R: This list is a good list to start. A quick nice way to start you off is the R tutorial created by Data Camp: here. R runs on a wide variety of UNIX-based platforms (including Mac OS X), Windows and Linux – you can download and use it even if your computer is 10 years old.

Topics and Readings

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


Sands, Melissa and Daniel de Kadt. 2019. “Local exposure to inequality among the poor increases support for taxing the rich.” *SoxArXiv Papers*.


**Readings: Natural Experiments**


**Readings: Non-technical Overviews or Controversies**


**Readings: Implementation and Practical Guides**

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

- Treatment noncompliance
- Principal stratification
- Local average treatment effects
- Wald estimator and two-stage least squares

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


