17.806: Quantitative Research Methods IV

Spring 2013

Instructors: Danny Hidalgo & Teppei Yamamoto
TA: Chad Hazlett

Department of Political Science
MIT

1 Contact Information

<table>
<thead>
<tr>
<th>Danny</th>
<th>Teppie</th>
<th>Chad</th>
</tr>
</thead>
<tbody>
<tr>
<td>Office: E53-402</td>
<td>E53-401</td>
<td>E40-446</td>
</tr>
<tr>
<td>Email: <a href="mailto:dhidalgo@mit.edu">dhidalgo@mit.edu</a></td>
<td><a href="mailto:teppei@mit.edu">teppei@mit.edu</a></td>
<td><a href="mailto:hazlett@mit.edu">hazlett@mit.edu</a></td>
</tr>
<tr>
<td>URL:</td>
<td><a href="http://web.mit.edu/teppei/www">http://web.mit.edu/teppei/www</a></td>
<td></td>
</tr>
</tbody>
</table>

2 Logistics

- Lectures: Mondays and Wednesdays 3:30 – 5:00pm, E53-438
- Recitations: TBA
- Danny’s office hours: Make an appointment.
- Teppie’s office hours: Make an appointment
- Chad’s office hours: Mondays 1:30 – 3:00, E40-446

3 Course Description

This course is the fourth and final course in the quantitative methods sequence at the MIT political science department. The course covers various advanced topics in applied statistics, including those that have only recently been developed in the methodological literature and are yet to be widely applied in political science. The topics for this year are organized into three broad areas: (1) Advanced causal inference, where we build on the basic materials covered earlier in the sequence (17.802) and study more advanced topics; (2) statistical learning, where we provide an overview of machine learning, one of the most active subfields in applied statistics in the past decade; and (3) Bayesian inference and statistical computing, where we extend the model-based inference techniques covered in the previous course of the sequence (17.804) and study more technically sophisticated materials as well as applications in political science.
4 Prerequisites

There are three prerequisites for this course:

1. Mathematics: basic calculus and linear algebra.
3. Statistical computing: familiarity with at least one statistical software.

For 1 and 3, refer to this year’s math camp materials to see the minimum you need to know; see https://stellar.mit.edu/S/project/mathprefresher/

5 Course Requirements

The final grades are based on the following items:

• **Problem sets** (30%): A total of four problem sets will be given throughout the semester. Problem sets will contain analytical, computational, and data analysis questions. Each problem set will be counted equally toward the calculation of the final grade. The following instructions will apply to all problem sets unless otherwise noted.

  – Neither late submission nor electronic submission will be accepted, except for special circumstances which must be discussed with us prior to the deadlines.

  – Working in groups is encouraged, but each student must submit their own writeup of the solutions. In particular, you should not copy someone else’s answers or computer code. We also ask you to write down the names of the other students with whom you solved the problems together on the first sheet of your solutions.

  – 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.

• **Final project** (50%): 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.

  We strongly encourage you to co-author a paper with another student in the class. By co-authoring you will (1) learn how to effectively collaborate with someone else on your research, which is very important in political science where most cutting-edge research is collaborative (see any recent issue of *APSR* or *AJPS!* and (2) more likely have a good, potentially publishable paper (multiple brains are usually better than one).

  Unless you already have a concrete research project suitable for this course (e.g., from your dissertation project), we recommend that you start with replicating the results in a published article and then improve the original analysis using the methods learned in this course (or elsewhere), methodologically or substantively. Oftentimes, gathering an original dataset is too time-consuming and not suitable for a course project.

  Students are expected to adhere to the following deadlines:
– March 11: Turn in a one-page 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). After submission, schedule a meeting with us to discuss your proposal.

– April 17 and 22: Students will present interim reports on their projects in front of the class. Each presentation should last about 10 minutes and will be followed by a short Q&A session. Students should prepare electronic slides to accompany their presentation. Performance on this presentation will be counted towards your total final grade (see below).

– May 15: Paper due. Please hand in one printed copy of your paper by 5pm, and also email electronic copies to us by then. Your final paper should have all the proper format of an academic journal article (except extensive literature review and theory sections), including a title, abstract, introductory and concluding sections, tables and/or figures with appropriate captions, and references with a coherent citation style. You will be penalized if any of these elements is missing from your submitted manuscript. You should use this project as an exercise to write a good scientific paper. We recommend that you closely follow the advice given in this article:


• Participation in Applied Paper Sessions (10%): Throughout the semester, we will have four applied paper sessions, in which we discuss journal articles and working papers which apply the methods covered in the lectures to empirical problems in political science and related fields. For each paper (or set of papers in some cases) marked as “required,” one student will deliver a 15-minute oral report which will walk the rest of the class through its content and provide comments on its merits and weaknesses, followed by a class discussion. The other students must read the required papers in advance and submit short written comments on each of the papers by the previous day. These comments should be no longer than a few sentences for each paper, optionally followed by a list of questions for class discussion. Although the written comments will not be graded, your participation in the class discussion will count towards the participation grade.

• Midterm Project Presentation (10%)

In addition, there will be required readings for each lecture which students must complete in advance in order to enhance their understanding. The lectures and applied paper sessions will also have optional readings, which are listed in the course schedule below.

6 Course Website

You can find the Stellar website for this course at:

http://stellar.mit.edu/S/course/17/sp13/17.806/

We will distribute course materials, including readings, lecture slides and problem sets, on this website.
7 Questions about Course Materials

In this course, we will utilize an online discussion board called Piazza. Below is an official blurb from the Piazza team:

Piazza is a question-and-answer platform specifically designed to get you answers fast. They support LaTeX, code formatting, embedding of images, and attaching of files. The quicker you begin asking questions on Piazza (rather than via individual emails to a classmate or one of us), the quicker you’ll benefit from the collective knowledge of your classmates and instructors. We encourage you to ask questions when you’re struggling to understand a concept ... See this New York Times article to learn more about their founder’s story: http://www.nytimes.com/2011/07/04/technology/04piazza.html

In addition to recitation sessions and office hours, please use the Piazza Q & A board when asking questions about lectures, problem sets, and other course materials. You can access the Piazza course page either directly from the below address or the link posted on the Stellar course website: https://piazza.com/mit/spring2013/17806

Using Piazza will allow students to see other students’ questions and learn from them. 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!

8 Recitation Sessions

Recitation sections will be held during the two weeks each problem set is available. Time and location will be announced after the first week of class. The purpose of these sessions will be to clarify theoretical material and assist with computing issues, particularly as needed to complete each problem set. Attendance is strongly encouraged.

9 Notes on Computing

In this course we use [R](#), an open-source statistical computing environment that is very widely used in statistics and political science. (If you are already well versed in another statistical software, you are free to use it, but you will be on your own.) Problem sets will contain computing and/or data analysis exercises which can be solved with R but often require going beyond canned functions and writing your own program.

In addition to the materials from the department’s math prefresher (see above), there are many resources for R targeted at both introductory and advanced levels, including:

- Fox, John and Sanford Weisberg. 2010. *An R Companion to Applied Regression*. Sage Publications. (focused on regression analysis)
- For specific questions about R, searching the CRAN website with appropriate keywords will often yield satisfactory results.
10 Books

The course has no required or recommended textbooks. All the reading materials are listed in the course schedule below and will be made available electronically.

11 Course Schedule

Part I: Advanced Causal Inference

1. Complex Experiments (2/11)
   
   **Required:**

   **Optional:**

2. Causal Inference with Interference between Units (2/13)
   
   **Required:**

   **Optional:**
3. Multiple Comparisons (2/19)

**Required:**


**Optional:**


4. Applied paper session (2/20)

**Required:**


5. Causal Diagrams (2/25)

**Required:**


**Optional:**


6. Partial Identification (2/27)

**Required:**

*Optional:*

• The rest of Manski, 2007.


7. Causal Mediation (3/4)

*Required:*


*Optional:*


8. Causal Attribution (3/6)

*Required:*


*Optional:*


9. Experimental Approaches for Measurement (3/11)

*Required*  


Optional:


10. Applied paper session (3/13)
Required: Controversy on Suicide Terrorism


Required: Voter Registration and Turnout


Optional: Ecological Inference


Optional: Causal Mediation Analysis Applications


Part II: Statistical Learning

1. Introduction to Learning and Regularization (3/18)

   Required:

   Optional:

2. Model assessment and selection, Classification Trees (3/20)

   - Pg. 305 - 313 in *Elements of Statistical Learning*.

   Optional:

3. Learning Algorithms (4/1)

   Required:

   Optional:
   - Sections 6.4-6.4.2 in Bishop, Christopher M. 2006. *Pattern Recognition and Machine Learning*. Springer.
   - Two tougher but very good background pieces on SVM, kernels, etc.
4. Ensemble Learning (4/3)

   Required:

   Optional:

5. Applied paper session (4/8)

   Required:


Midterm Project Presentations (4/17, 4/22)

Part III: Bayesian Inference and Statistical Computing

1. Advanced Simulation Algorithms (4/24)

   Required:

   Optional:

2. Discrete Choice Analysis (4/29)

Required:


Optional:


3. Bayesian Measurement Techniques (5/1)

Required:


Optional:


4. Applied paper session (5/6)

Required: EM Algorithm Application


Required: Ideal Point Estimation


Optional: Measuring Democracy


Optional: More Ideal Point Estimation


Optional: Bayesian Approaches to Ecological Inference


5. Multi-level Regression and Post-Stratification: Guest lecture by Chris Warshaw (5/8)

6. Missing Data: Guest lecture by James Honaker (5/13)

Bonus Sessions

1. Web Scraping (5/15)