POL 573: Quantitative Analysis III

Fall 2016

Kosuke Imai Asya Magazinnik Erik Wang

Department of Politics
Princeton University

This course is the second course in applied statistical methods for social scientists. Building on the materials covered in POL 572 or its equivalent (i.e., linear regression, structural equation modeling, instrumental variables, maximum likelihood estimation, discrete choice models), students will learn a variety of statistical methods including models for longitudinal data and survival data. Unlike traditional courses on applied regression modeling, I will emphasize the connections between these methods and causal inference, which is the primary goal of social science research.

1 Contact Information

Kosuke Imai (Instructor)
Office: Corwin Hall 036
Email: kimai@Princeton.Edu
URL: `http://imai.princeton.edu`

	Asya Magazinnik (Preceptor)	Erik Wang (Preceptor)
Office:	Corwin Hall 025	Corwin Hall 027
Email:	asyam@princeton.edu	haixiaow@princeton.edu

2 Logistics

- Lectures: Mondays and Wednesdays (Corwin 127) 10:30–11:50am
- Precepts: Fridays (Corwin 127) 3:00–4:30pm (starting Sept 25)
- Kosuke's Office Hours: Tuesdays 2:00–3:00pm (sign up by WASS)
- Preceptors' Office Hours:
  Erik: Tuesdays (Corwin 126) 10:15–11:45am
  Asya: Thursdays (Corwin 126) 5:00–6:30pm
- Google Calendar: `http://goo.gl/3SlHn2`

3 Questions and Announcements

In addition to precepts and office hours, please use the **Piazza** Discussion Board at `https://piazza.com/` when asking questions about lectures, problem sets, and other course materials.

This allows all students to benefit from the discussion and to help each other understand the materials. Both students and instructors are encouraged to participate in discussions and answer any questions that are posted.

To join the POL 573 Piazza site, click on "Search Your Classes" from the Piazza homepage. After specifying Princeton University as your school, search for "POL 573: Quantitative Analysis III." You will then be prompted to enter your `princeton.edu` email address to confirm your registration. Piazza can also be accessed from within Blackboard by going to the POL 573 course page and clicking on the link to "Piazza Messageboard." In addition, all class announcements will be made through Piazza. Blackboard will still be used for hosting all class materials.

Some useful tips for Piazza include:

- Piazza has apps available for the iOS and Android platforms. The apps are free downloads and provide complete access to all of Piazza's message board features.
- To insert LaTeX-formatted text in a post, place a double dollar sign ($$) on both ends of the relevant text, or click the fx button in the Details toolbar above your post.
- To add formatted **R** code to a post, click the "pre" button in the Details toolbar above your post. A grey text box will open up where you can paste code from **R**.
- You can classify a post using pre-selected tags, or you can generate your own by prepending a hash (#) to your chosen label. Posts can then be sorted by these tags using the search bar in the left-hand column.

4 Course Requirements

Your final grade is based on the problem sets and the final project:

- **Problem sets** (30%): There will be several problem sets throughout the semester. Each problem set will equally contribute to the final grade and contain both analytical and data analysis questions. The following instructions will apply to all problem sets:
  - *Collaboration.* You are allowed to collaborate with other classmates and receive help from preceptors. However, you must not copy someone else's code or answers and are required to submit your own answer for each problem set.
  - *Submission.* Students should write up their answers using **Rmarkdown**. Submit your **Rmarkdown** file to Blackboard and bring its printed copy to class. Please ensure your code adheres to the Google's **R** Style Guide rules (`https://google-styleguide.googlecode.com/svn/trunk/Rguide.xml`), as style errors will be penalized. Neither late submission nor electronic submission will be accepted unless you obtain a prior approval from the instructor.
- **Final project** (70%): The final project must be a collaborative project with another student in this class (unless you obtain a special permission from me to do an individual project). I strongly recommend that you start with the replication of the empirical results published in the field of your interest (rather than gathering an original data set, which is often too time consuming for a course project). After the replication, your goal is to improve the original analysis either methodologically or substantively (or ideally both). During the entire process, the instructor and preceptor are available to help you with both substantive and technical questions.

  Be aware of the following key deadlines. Late submission will be penalized.

- **September 21 (Project and collaborator identification)** By this date, you should identify your collaborator and project. Upload one paragraph (up to 150 words) description of your project to the Blackboard and bring a copy to the class.
- **October 5 (Data acquisition)** By this date, you should find your coauthor and acquire the data to be analyzed (e.g., by requesting the data from the original authors). Upload your data set and a one page description of the data to Blackboard, and bring a copy of your writeup to the class.
- **November 7 (Descriptive analysis):** By this date, you should finish your descriptive analysis. Turn in a brief summary of your descriptive analysis of the data (up to 3 pages). This memo should not simply list tables of descriptive statistics and graphs of variables. Instead, it should present key aspects of the data, that are central to your substantive and/or methodological questions of interest, in clear and intuitive manners. Meet with the instructor and preceptors to get feedback for your memo.
- **November 30 (Initial analysis and proposed extensions):** By this date, you should finish your initial analysis of the data (e.g., the replication of the original results) and come up with the proposed extensions of the analysis. Turn in a brief summary of your replication and proposed extensions of the original analysis (up to 5 pages). Meet with the instructor to get feedback for your memo.
- **January 10 (Initial write-up):** By this date, you should finish all of your empirical analyses and create all the tables and figures for presenting the results. You must electronically submit an initial write-up of your report to the Blackboard by noon. This write-up should consist of the title, the abstract, the introduction, and the tables and figures with informative captions (up to 10 pages). Over the next few days, you will be asked to comment on the initial write-up by another student in the class so that the author can improve the analysis and writing.
- **January 13 (Poster):** Based on the initial write-up, make a poster, which concisely summarizes the substantive and methodological motivations/contributions, the methods you used, and the results you obtained. The poster should be uploaded to the Blackboard as a pdf file by noon. The size of the poster should be A0 (33.1 $\times$ 46.8 inches).
- **January 16 (Written feedback):**You should electronically submit to the Blackboard your written feedback on the assigned initial write-up by noon. Your written feedback will be graded based on its quality (10% of the course grade).
- **January 17 (Poster session):** The poster session will be held from noon to 2pm in TBD. You will get feedback on your poster from faculty members and other students, which you should then incorporate into your final report. All posters will also be made available online.
- **January 24 (Final paper):** You should write up your final paper reflecting the comments you received on your initial write-up and the poster. A pdf copy of the paper should be submitted to the Blackboard by 4pm. In addition to the electronic submission, you should turn in one double-sided paper copy of your final paper to the preceptor's Corwin Mailbox by the same time. The final paper will be graded based on its overall quality. Specifically, we will look at the novelty of substantive and methodological contributions as well as the effectiveness of presentation and writing. We will be following AJPS guidelines for paper length: no longer than 10,000 words, including the main body of text, notes, references, and the headers of tables and figures (not including the title page, abstract, or supporting information).

5 Statistical Software

In this course, we support a statistical computing environment, called **R**. **R** is available for any platform and without charge at http://www.r-project.org/. We choose **R** for its flexibility and power. However, students may use other statistical software such as STATA for the problem sets and the final project, but at their own risk; that is, we will not be able to answer your software-related questions. Of course, there will be no penalty for using different statistical software. What matters is the analysis you present rather than the software you use.

6 Textbooks

There is no single textbook for this course. However, you may find the relevant parts of the following textbooks useful. Some of these books are reserved at the Firestone library.

1. Political Methodology

   Gary King. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. University of Michigan Press, Ann Arbor, 1998.

2. Probability and Statistics

   Morris H. DeGroot and Mark J. Schervish. *Probability and Statistics*. Addison Wesley, Boston, 3rd edition, 2002.

   David A. Freedman. *Statistical Models: Theory and Practice*. Cambridge University Press, Cambridge, 2005.

   Larry Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer, New York, 2005.

3. Econometrics

   Fumio Hayashi. *Econometrics*. Princeton University Press, Princeton, 2000.

   Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge, MA, 2nd edition, 2010.

4. Regression and Hierarchical Modeling

   Andrew Gelman and Jennifer Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambrdige University Press, Cambridge, 2007.

5. Causal Inference and Research Design

Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* Cambridge University Press, 2015.

Stephen L. Morgan and Christopher Winship. *Counterfactuals and Causal Inference: Methods and Principles for Social Research.* Cambridge University Press, New York, 2007.

Joshua D. Angrist and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press, Princeton, 2009.

6. **R**

John Fox. *An R and S-plus Companion to Applied Regression.* Sage Publications, Thousand Oaks, CA, 2nd edition, 2011.

7 Course Outline

We will cover some of the following topics as time permits!

1. Statistical Modeling with Cross-section Data
   (a) Discrete choice models
   (b) Mixture and sample selection models
2. Causal Inference
   (a) Causality and regression
   (b) Matching and weighting
   (c) Fixed effects and differences-in-differences
3. Statistical Modeling with Longitudinal Data
   (a) Linear mixed effects models
   (b) Generalized linear mixed effects models
   (c) Multilevel/Hierarchical models
   (d) Dynamic models
4. Survival Data Analysis
   (a) Basic concepts
   (b) Parametric and semi-parametric regression models
   (c) Competing risks models