

POL 573: Quantitative Analysis III

Fall 2010

Kosuke Imai

This course is the second course in applied statistical methods for social scientists. Building on the materials we 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

Office: Corwin Hall 036

Office Phone: 258-6601

Email: kimai@Princeton.Edu

URL: <http://imai.princeton.edu>

2 Logistics

- Lectures: Mondays and Wednesdays (Robertson 029) 10:30–11:50am
- Precepts (taught by Michael Donnelly mdonnell@Princeton.EDU): Friday 3-4:30, Robertson 008.
- Kosuke's Office hours: 2:30 – 4:30 Tuesdays, but stop by anytime (or make an appointment by email)
- Michael's Office hours: 1:40 – 3:40 Thursdays or by appointment, Bobst 303

3 Questions about the Course Materials

In addition to precepts and office hours, please use the *discussion board* at Blackboard when asking questions about lectures, problem sets, and other course materials. This allows all students to benefit from the discussion and help each other understand the materials. Students are also encouraged to answer any questions that are posted by other students.

4 Prerequisites

There are three prerequisites for this course.

1. Mathematics at the level of the math camp and POL 502

2. Probability and statistics at the level of POL 571 and POL 572
3. Statistical computing and programming at the level of POL 572

5 Course Requirements

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

- **Problem sets** (40%): There will be approximately bi-weekly problem sets. Although you are allowed to discuss the problem sets with others, you should not copy someone else's computer code or answers. In particular, sharing a paper or electronic copy of your code and answers with other students is strictly prohibited.
- **Final project** (60%): The final project must be a collaborative project with another student in this class. 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 both). Be aware of the following key deadlines. Neither electronic nor late submission is allowed.
 - *November 8 (Data acquisition, descriptive analysis)*: By this date, you should find your coauthor and acquire the data to be analyzed (e.g., request the data from the original authors). Turn in a brief summary of your descriptive analysis of the data after familiarizing yourself with the data set by computing simple statistics and creating some graphs. Meet with the instructor to get feedback for your memo.
 - *December 6, 2010 (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. Meet with the instructor to get feedback for your write-up.
 - *January 7, 2011 (Poster session)*: By this date, you should finish all of your empirical analyses and create all the tables and figures for presenting the results. Make a poster, which concisely summarizes the substantive and methodological motivations/contributions, the methods you used, and the results you obtained. The poster will serve as an outline for the report to be written. You will get feedback on your poster from faculty members and other students, which you can then incorporate into your final report.
 - *January 18, 2011 (Final deadline)*: Please turn in one double-sided paper copy of your final report to Kosuke's Corwin Mailbox by 4pm.

6 Statistical and Word Processing 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.

For word processing software, I recommend L^AT_EX (together with BibTeX for bibliography) for writing a scientific paper rather than Microsoft Word. The program is efficient and the output is beautiful. It works nicely with R and can easily produce a poster too.

7 Final Project

The final project is the most important requirement of this course. The goal is to conduct a project that can be *eventually* developed into a high-quality publishable paper. I encourage you to keep working on your project after this course, and I am happy to continue to provide guidance along the way. In particular, you may consider applying for the summer political methodology meeting using the paper that comes out of your final project. In the past, papers based on final projects for this course appeared in refereed journals and won the best graduate student poster award at the political methodology summer meeting. Here are a few important things to note when conducting the final project for this course. PLEASE BE SURE TO TAKE THE FOLLOWING GUIDELINES SERIOUSLY AND FOLLOW THEM WHENEVER YOU CAN.

- **Use of tables and figures.** Tables and graphs are important because they report the *main* results of your paper. You should carefully decide which results are central to your study and only use tables and figures to present these results. Tables and figures should speak for themselves; readers should be able to understand them without referring to the main text. If people browse through your paper for five minutes only looking at the abstract, tables, and figures, they should understand the main contributions of your paper. Use detailed captions to help achieve this goal.

Use common sense to make sure that readers who are unfamiliar with your research can immediately understand the results; too many numbers in a table and too many lines in a figure, for example, can confuse readers. Often, figures are more effective conveying the empirical results than tables. See the following paper for some good examples.

Johnathan P. Kastellec and Eduardo Leoni. Using graphs instead of tables to improve the presentation of empirical results in political science. *Perspectives on Politics*, 5(4):755–771, 2007.

- **Writing of the paper.** You should NOT start writing your paper until you finish your empirical analyses and finalize tables and figures with informative captions. Once this is done, then you can draft the paper in the following order; the title, the abstract, the introduction, and the rest of the paper.

The title should be informative, conveying what your paper is about. Avoid a catchy but uninformative title. The abstract should be about 150 words, presenting the motivation (i.e., the problem you are going to solve or the question you are going to answer) as well as your substantive and/or methodological contributions (i.e., your proposed method or new empirical findings). Carefully draft each sentence in the abstract to efficiently convey all the important information about your paper. With the 150 word limit, you do not have any sentences to waste.

The introduction of the paper should simply elaborate each sentence in the abstract. That is, use a couple of paragraphs to expand what you write with one sentence in the abstract. The writing should be done in the top-down manner. The introduction should start with a brief discussion of the motivation of your paper immediately followed by a concise summary of the

main contributions. After that, you can further discuss the ways in which your methods or empirical findings contribute to the relevant literature. Avoid a boring “literature review” section whenever possible. Instead, review the literature in light of your results by discussing how your new methods or empirical results differ from the existing ones. I recommend that you lay out the structure of the paper first (i.e., sections and subsections) so that you have the entire picture of the paper while drafting the introduction.

Finally, the rest of the paper should naturally follow from the introduction. Each paragraph in the introduction becomes a section of the paper where you give further details about your methods and empirical results. Use subsections to effectively navigate readers so that they do not lose the big picture. Again, each paragraph should be written in the top-down manner; the first sentence should give the main message and the rest of the paragraph should elaborate it whereas the last sentence should provide a smooth transition to the next paragraph. After drafting the entire paper, read it over a couple of times and polish your writing.

There are many books and articles explaining how to write a good scientific paper, but here is one written by a political scientist.

Gary King. Publication, publication. *PS: Political Science and Politics*, XXXIX(1):119–125, January 2006.

8 Poster Session

We will use the poster session to get feedback from others in an efficient manner. It is a great way to get reactions in a one-on-one settings. Faculty and other graduate students will be invited to the poster session.

Before starting to make your poster, you should finish drafting the title, the abstract, and the introduction as well as creating tables and figures with detailed captions. A poster is essentially a set of presentation slides, structured in the same manner as the paper. Tables and figures will play an essential role. Use bullet points rather than full sentences; you will be verbally explaining them to others rather than having them read every sentence on your poster.

During the poster session, pay attention to whether or not people are immediately understanding your tables and figures, whether the structure of the poster is working, and if there are any significant issues you have not noticed. Such feedback should help you finalize your analysis and draft your paper.

- **Making a poster.** Both \LaTeX and PowerPoint are popular, but we encourage students to use \LaTeX . Various \LaTeX poster templates are available online. For example, the following website <http://nwg.me.uk/docs/posters/> has two different versions, one for landscape and one for portrait. It may also be helpful to look at the posters presented at previous PolMeth conferences. Those can be found at the following website <http://polmeth.wustl.edu/media.php>
- **Printing a poster.** We recommend that students take advantage of a discount available for class projects at the Print Services office. Instructions can be found at the following website http://www.princeton.edu/~oitprint/pm_largeformat.html Another on-campus location where a poster can be printed is in PRISM, which has the necessary information at this website <http://www.princeton.edu/prism/info/poster-printing/> Posters can also be printed at Kinko’s/FedEx Office, etc. but these tend to be more expensive than the above on-campus options.

- **Setting up a poster.** Both easels and poster board will be provided along with push pins, and so you only need to bring the poster to the session. Arrive enough in advance to set up your poster. Use push pins to attach the poster to the board and place them on an easel.
- **Presenting a poster.** The poster session will be approximately two hours long. During the first hour, one author should present the poster and the other should visit other posters. They should switch their roles during the second hour. Presenting a poster is very similar to giving a seminar talk, but the difference is that you can more freely interact with your audience. Encourage your audience to ask questions, point out strengths and weaknesses, and provide alternative ways to approach the problem. If you are visiting someone else's poster, try to do the same to them.

9 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.

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, 2002.

4. Regression and Hierarchical Modeling

Andrew Gelman and Jennifer Hill. *Data Analysis Using Regression and Multi-level/Hierarchical Models*. Cambridge University Press, Cambridge, 2007.

5. Causal Inference and Research Design

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, 2002.

10 Course Outline

We will cover the following topics in the order they are listed (and as time permits!).

1. Discrete Choice Models (continued from POL 572)
 - (a) Multinomial choice models
 - (b) Sample selection models
2. Applied Regression Modeling with Cross-section Data
 - (a) Event count models
 - (b) Generalized linear models
3. Causal Inference with Cross-Sectional Data
 - (a) Matching/Weighting methods
 - (b) Multiple imputation for missing data
4. Statistical Inference with Longitudinal Data
 - (a) Linear mixed effects models
 - (b) Generalized linear mixed effects models
 - (c) Multilevel/Hierarchical models
5. Survival Data Analysis
 - (a) Basic concepts
 - (b) Parametric and semi-parametric regression models
 - (c) Competing risks models