This version: January 5, 2015

Political Science 271
Advanced Statistical Applications

Winter Quarter 2015
SSB 353, Tuesday 6-7:30PM, Thursday 3:30-5PM

Molly Roberts
SSB 339
meroberts@ucsd.edu
Office Hours: Tuesday, 1–3 PM

Prerequisites

Political Science 270 (or equivalent)

Overview

This course is the second course in the quantitative research methods sequence at the UCSD Political Science department. Building on Political Science 270, this course teaches advanced statistical tools for empirical political science. In the first half of the course, we will focus on techniques for model-based inference, with a specific focus on generalized linear models. We will cover the basics of the fundamental statistical principles underlying these models (e.g., maximum likelihood theory) as well as a variety of estimation techniques. In the second half of the course, we will focus on design-based inference, causal inference, and matching methods. Time permitting, we will cover special topics, including measurement, text analysis, and missing data. The ultimate goal of this course is to provide students with adequate methodological skills for conducting cutting-edge empirical research in their own fields of substantive interest.

Assessment

There are no written exams in the class, and your grade will be based on a combination of:

- **Homeworks (50%):** Seven problem sets will be given throughout the quarter, skewed heavily toward the beginning of the quarter. 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.

- Late submission will not be accepted unless you ask for special permission from the instructor in advance. Problem set write-ups should be turned in in hard copy, a separate copy of the problem set write-up and code will be turned in electronically.

- Working in groups is encouraged for conceptual and sometimes technical discussion, but each student must submit their own writeup of the solutions that shows their independent work on the assignment. In particular, you should not copy someone elses 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 (40%)**: The final project will be a poster and short research memo which typically applies a method learned in this course to an empirical problem of your substantive interest. The memo should outline a research paper that could potentially be written after the class has been completed.

  - I encourage you to work with another student on your poster and memo. 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 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). Oftentimes, the most time-consuming part of a research project is data collection (which is not the focus of this course) and using data someone has already archived for their publication and made publicly available gets around this problem.

- Students are expected to adhere to the following deadlines:

  - January 27: 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). Meet with the instructor to discuss your proposal during her office hours. You may be asked to revise and resubmit the proposal in two weeks from the meeting.

  - March 12: Poster session: Class time will be spent in a poster session on this day where students present the results of their paper and comment on one anothers' work. You can incorporate the feedback given in the poster session into the research memo.

  - March 18: Memo due. Please turn in one printed copy of your memo by the end of the day, and email electronic copies to the instructor.

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

Academic Honesty and Plagiarism

All of your graded work must be done by you. If you are unfamiliar with the Universitys policy on academic integrity, please see http://students.ucsd.edu/academics/academic-integrity/policy.html.

Syllabus and Plan

The syllabus will be updated periodically throughout the course, so that we can keep with the cadence of the class. I will post to Piazza when such updates are made.

Reading and Textbooks

We will read chapters from these books throughout the course. We recommend that you purchase the King and Cameron and Trivedi books. The others we will only read a few chapters from, and will be available on electronic reserve.

- Cameron, Colin and Pravin Trivedi. *Microeconometrics: Methods and Applications.* Cambridge University Press, 2005.

- Efron, Bradley and Robert J. Tibshirani. 1994. *An Introduction to the Bootstrap.* Chapman and Hall/CRC.

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

- Hastie, Trevor, Robert Tibshirani and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer-Verlag, 2009. (available online at: `http://statweb.stanford.edu/~tibs/ElemStatLearn/`.)

- Jackman, Simon. 2009. *Bayesian Analysis for the Social Sciences.* Wiley.

- James, Gareth, Daniela Witten, Trever Hastie and Robert Tibshirani. *An Introduction to Statistical Learning with Applications to R.* Springer-Verlag, 2013. (available online at: `http://www-bcf.usc.edu/~gareth/ISL/`.)

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

- Murphy, Kevin. *Machine Learning: A Probabilistic Perspective.* MIT Press, 2012.

Piazza

We will be using Piazza for general discussion and questions and answers throughout the class. Piazza allows students to see other students' questions and learn from them as well as answer them. Your respectful and thoughtful participation in the discussion forum will count toward your participation grade. Please do not e-mail the instructor with questions (post them on Piazza!) unless they are personal in nature. I will check the Piazza forum daily to provide my own answers and contributions.

Software

We will be using `R` an open-source statistical package. You can download it from the web here:

`http://cran.r-project.org/`

COURSE SCHEDULE

1 Jan 6: Course Intro and Inference

Reading

      Chapter 1, Section 1.1, King

      Chapter 2, Section 2.1, Cameron and Trivedi

      Chapter 2, Section 2.1, James, Witten, Hastie and Tibshirani

2 Jan 8: Linear Regression Reframed

Reading

      Chapter 1, Section 1.2-1.3, King

      OPTIONAL: Chapter 4, Cameron and Trivedi (for the Econometrics perspective)

      OPTIONAL: Chapter 2, Hastie, Tibshirani, and Friedman (for the prediction perspective)

3 Jan 13: Probability Distributions

Reading

      Chapter 3, King

      Chapter 2, Murphy

4 Jan 15: Introduction to Maximum Likelihood

Reading

      Chapter 2, King

5 Jan 20: Maximum Likelihood

Reading

      Chapter 4, 4.1-4.3, King

      Chapter 5, Cameron and Trivedi

6 Jan 22: Simulation/Monte Carlo Methods

Reading

Chapter 6, Efron & Tibshirani.

King, Gary, Mike Tomz and Jason Wittenberg, 2000, Making the Most out of Statistical Analysis: Improving the Interpretation and Presentation. American Journal of Political Science, 44(2), pp.341355.

OPTIONAL: Chapter 11, Cameron and Trivedi

7 Jan 27: Properties of Maximum Likelihood Estimators

Reading

Chapter 7, 7.2-7.4, Cameron and Trivedi

Chapter 4, 4.4-4.8, King

Buse, A. 1982. The Likelihood Ratio, Wald, and Lagrange Multiplier Tests: An Expository Note. The American Statistician, 36(3), 153157.

8 Jan 29: Binary Dependent Variables

Reading

Chapter 5, 5.1-5.3, King

Chapter 14, Cameron and Trivedi

Section 4.3, James, Witten, Hastie and Tibshirani

9 Feb 3: Binary Dependent Variables

Reading

Chapter 5, 5.4-5.6, King

Chapter 15, Cameron and Trivedi

10 Feb 5: Event Count Models

Reading

Chapter 5, 5.7-5.9, King

Chapter 6, 6.1-6.3, Gelman & Hill

Chapter 20, Cameron and Trivedi

11 Feb 10: Event Count Models

Reading

Chapter 5, 5.7-5.9, King

Chapter 6, 6.1-6.3, Gelman & Hill

Chapter 20, Cameron and Trivedi

12 Feb 12: Checking Model Fit

Reading

Chapter 11, Gelman & Hill

Chapter 21, Cameron and Trivedi

Chapter 7, Hastie, Tibshirani, and Friedman

Chapter 5, James, Witten, Hastie and Tibshirani

13 Feb 17: Causal Inference

Reading

Chapter 9 & 10, Gelman & Hill

Chapter 3, Cameron and Trivedi

Andrew Gelman and Guido Imbens. "Why ask Why? Forward Causal Inference and Reverse Causal Questions."

14 Feb 19: Design-based Inference

Reading

Gary King and Langche Zeng. The Dangers of Extreme Counterfacturals. Political Analysis, 2006.

Kosuke Imai, Gary King, and Elizabeth Stuart. Misunderstandings among Experimentalists and Observationalists: Balance Test Fallacies in Causal Inference. Journal of the Royal Statistical Society, Series A Vol. 171, Part 2 (2008): Pp. 1-22

15 Feb 24: Matching

Reading

> Ho, Daniel, Kosuke Imai, Gary King, and Elizabeth Stuart. 2007. Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference, Political Analysis 15: 199236. Copy at http://j.mp/jPupwz

16 Feb 26-10: Special Topics

May include depending on time and demand:

> Time-series
>
> Prediction & Measurement
>
> Missing Data
>
> Text as Data

17 Mar 12: Poster Session