Preliminaries

Overview and Class Goals

This course is designed to introduce students to several broad areas of modern statistics and to illustrate how ideas and methods from each of these areas can be fruitfully applied to political science. The course is not designed to provide a comprehensive treatment of any of these topics. Instead, it is designed to introduce students to key ideas and to provide enough background knowledge so that students can pursue a more detailed study of some topics on their own.

Prerequisites

Students should have taken GOV 2000 and GOV 2001 or the equivalent courses. Some students without this background may still be able to do very well in this course. Students who have not taken courses similar to GOV 2000 and GOV 2001 should talk to the instructors to get a sense of how prepared they are for the course.

Class Requirements

Students are expected to attend all meetings. The course is formally structured as a lecture, followed by structured discussion and presentation. We (the instructors) would prefer to have the course operate as an ongoing conversation among class participants. Many of the topics are either open areas of research and/or somewhat controversial and thus are well-suited to detailed discussion. We encourage students to ask questions and spark discussion. We are certainly open to students suggesting directions for the class, and suggesting readings for the various weeks.

Assessment will be based on presentation and participation (20% of final grade), homework(s) (20% of final grade) and a final project (60% of final grade). The final project has a ‘prospectus’ component (4 pages) which are to be handed in to the instructors the class before the Thanksgiving holiday period begins (November 22). The prospectus should lay out the paper: the idea (perhaps the substantive question of interest), the techniques to be used, where the data will be found, the results expected. The final projects are designed to allow students to develop a deeper understanding of one or more of the topics introduced in the course. Possible final projects include (but are not limited to):
• Writing a paper that applies some of the methods discussed in class to a substantive question with real data.
• Building an original dataset from textual data and conducting some preliminary analysis.
• Developing a new model or procedure.
• Proving a conjecture put forth by the instructors.

Final projects need to be approved by both instructors by December 13.

We will not give incompletes in this course.

Each week, two students will present readings from the syllabus in an APSA-style presentation: 12–15 minutes, with slides. Students should be sure to critique the works, thinking about its strengths, weaknesses, and how it could/should be applied in political science. The class meets on Thursdays: students should email the instructors by Sunday should they face particular problems understanding/interpreting the papers under study.

Computation

We will use the R computing language in this course. The R language is completely open-source, so you can download it for free from [http://www.r-project.org/](http://www.r-project.org/). R will also be installed at HMDC and on the FAS computing cluster.

We hope to have classes on Python also, which will be introduced in our weeks on text analysis. Prior knowledge of Python is not necessary for this course.

Course Website

The course website is located at the following URL: [http://isites.harvard.edu/k98475](http://isites.harvard.edu/k98475). This site will provide homework assignments, datasets, and supplementary materials.

Office Hours and Availability

Professors Blackwell and Spirling will have office hours by appointment this semester. If you have questions about the course material, computational issues, or other course-related issues please do not hesitate to set up an appointment with the Professors.

Required Books


Optional Books


---

**Preliminary Schedule**

The course outline below is a preliminary schedule. We will adjust the amount of time spent on each topic to meet student interest and may well cut some of the topics below to make room for additional material. Students should let the instructors know what they are most interested in.

The “Required Reading” should be completed prior to lecture in a given week. “Optional Reading” is directly related to the lecture for the week in question but is not required. The “Optional Reading” typically provides more detail on some of the themes topics covered in lecture. The “Background Reading” provides supplementary material that may be helpful but is not necessarily directly related to what is covered in lecture.

1. **Sept 11. Bayesian Analysis I: Philosophy and Approach**

   **Topics covered**
   - nature of probability and uncertainty
   - Bayesian setup: priors and posterior inference
   - Hypothesis testing in Bayesian framework

   **Required Reading**

   **Presentation** Brad Efron “Why Isn’t Everyone a Bayesian?” The American Statistician, Vol. 40, No. 1 (Feb., 1986) [include following discussion of Efron’s article]

   **Optional Reading**

   Gill, Ch 1–7
2 Sept 18. Bayesian Analysis II: Estimation via MCMC

Topics covered

- Monte Carlo methods/integration
- sampling approaches
- Markov Chain Monte Carlo
- convergence and practical concerns

Required Reading

Gill Ch 8, 9


Optional Reading


Topics covered

- Potential outcomes framework
- Selection on observables and the back-door criterion
- Regression, matching, weighting, and doubly robust estimators

Required Reading

Hernán & Robins, Ch. 1, 7.

Imbens & Rubin, Ch. 4 (skim), 6-7.

Presentation Angrist & Pischke, Ch 1-3.3.


Optional Reading

Angrist & Pischke, Ch 3.3-3.4.
4  Oct 2. Causal Inference II: Instrumental Variables

Topics covered

- Instrumental variables
- Local average treatment effects and compliance

Required Reading

Angrist and Pischke: Chapter 4


Optional Reading


5  Oct 9. Roll Call Analysis

Topics Covered

- summarizing data: clustering, PCA
- item response theory: Rasch model and others
- spatial voting
- likelihood-based approaches: NOMINATE
- Bayesian approaches, identification
- practical problems in roll call analysis: party discipline

Required Reading


Topics covered

- Sharp and fuzzy regression discontinuity designs
- Sorting at the discontinuity

Required Reading

Angrist and Pischke: Chapter 6


Optional Reading


7  Oct 23. Causal Inference IV: Repeated Measurements over Time

Topics covered

- Random and fixed effects
- Differences-in-differences estimators

Required Reading

Angrist & Pischke: Chapter 5


Optional Reading


8 Oct 30. Strategic Interaction in Political Science

Topics covered

- strategic models
- comparative statics analysis
- strategic selection bias
- belief updating

Required Reading


Optional Reading


9 Nov 6. Texts as Data

Topics covered

- n-gram Representations
- The “Bag of Words” Assumption
- Stemming
- The Vector Space Model and Term Weighting
- Unsupervised, Supervised, and Semi-Supervised Learning

Required Reading


Optional Reading
Manning and Schütze. 1999. *Foundations of Statistical Natural Language Processing*, Chapters 4, 5, 6, 14, 15, and 16.

Background Reading

Thomas, Matt; Bo Pang; and Lillian Lee. 2006. “Get Out the Vote: Determining Support or Opposition from Congressional Floor- Debate Transcripts”.

10 Nov 13. Text II: Regular Expressions and Parsing
Guest lecture/hands on session.

11 Nov 20. Text III: Using Python for Web Scraping
Guest lecture/hands on session

Required Reading
Bird, Steven; Ewan Klein; and Edward Loper. 2007. *Natural Language Processing in Python.*
Chapters 2 and 3.

12 Dec 4. Experiments (Enos)
Topics covered
• Randomization Inference
• Experimental Design
• Experimental Artifacts

Required Reading


Optional Reading

13 Dec 11. Student Presentations of Paper Topics: Group Critique

September, 2014