

Political Data Science

Spring 2013

Tuesdays and Thursdays 1:15-3:05 pm

School of Education Building Rm 128 (Location change likely, TBA)

Instructors

Prof. Adam Bonica (Office Hours: TBA)

Prof. Clayton Nall (Office Hours: Tuesday/Thursday 3:05pm-4:30pm)

Course Description

Few fields have been left untouched by the revolution combining computing power and statistics, and the scientific study of politics has been one of the most noted aspects of these changes. The 2012 Obama campaign developed a massive database containing the political behavioral data of tens of millions of voters, and similar databases are now widely used in political science. Through the popular media, quantitative methods in the social sciences have gained wide attention.

We have many tools at our disposal, but what questions about politics are we interested in and able to answer? We will begin answering this question by reviewing basic probability and statistics using examples from politics. We will then turn to an in-depth discussion of the basics of causal inference and the limitations of experimental and observational methods in the study of social phenomena, culminating in tutorials on the use of matching and linear regression for causal inference. The course will then shift to a set of cutting edge quantitative methods related to the “big data” revolution, including network analysis and data visualization techniques.

Because we expect to have a class of not more than about 20 students, we will mix short lectures with in-class group problem solving, seminar discussion, and statistical software training.

Learning Goals

By the end of this class, you will have achieved the following:

- Understand the application of basic probability and statistics to political science research settings.
- Understand and evaluate major approaches to causal inference in experimental and observational social-science research.
- Interpret least-squares regression and use R to do least-squares regression.
- Know how to use principles of data visualization to represent research results effectively.
- Be able to read and understand research based on new, data-intensive methods in politics.

Grading Policy

Grades for all projects, homework, and tests will be assigned on the customary 4-point scale.

The course grade will be based on a combination of short problem sets, quizzes, class participation, and a final group project.

Problem Sets (5) (35%)

Quizzes (4) (20%)

Class Participation (including class attendance, participation in group work, and questions on class e-mail list) (20%)

Final Group Project: Analysis with Large Data Sets (25%)

Five problem sets will be assigned throughout the quarter. We encourage you to work on these problem sets in groups of up to three students. You are expected to write up problem sets individually, *in your own words*, and report at the top of your homework your coauthors.

Primarily as a tool to assess class progress, four quizzes, each worth 5% of the grade, will be assigned throughout the quarter. These quizzes will not require as much preparation as a typical midterm or final exam. They will focus on recently covered material.

Class participation accounts for 20% of your grade. You can contribute to your participation grade through in-class contributions, assisting your peers in class exercises, and posting questions and answers on the class e-mail list.

Extra credit opportunities will be offered at the professors' discretion during the quarter.

Readings

Kindle/online versions of these texts are acceptable.

Dunning, Thad. *Natural Experiments in the Social Sciences: A Design-Based Approach*. ISBN: 978-1107698000 (Paperback)

Kellstedt and Whitten. *The Fundamentals of Political Science Research*. ISBN: 978-0521697880 (Paperback)

Recommended:

Teetor, Paul. *R Cookbook*. O'Reilly Press. ISBN: 978-0-596-80915-7

Cleveland, William. *Visualizing Data*.

Other course readings will appear on Coursework or will be distributed in class, as necessary.

Students with Documented Disabilities: Students who may need an academic accommodation based on the impact of a disability must initiate the request with the Office of Accessible Education (OAE). Professional staff will evaluate the request with required documentation, recommend reasonable accommodations, and prepare an Accommodation Letter for faculty dated in the current quarter in which the request is made. Students should contact the OAE as soon as possible since timely notice is needed to coordinate accommodations. The OAE is located at 563 Salvatierra Walk (phone: 723-1066, URL: <http://studentaffairs.stanford.edu/oae>).

Topics and assignments are subject to revision during the course at the professors' discretion.

Week 1

April 2 Course Introduction

What kinds of questions do political scientists ask?

Course overview

Diagnostic test

Readings:

Kellstedt and Whitten, Chs. 1 & 2.

Assignments:

If unfamiliar with R, work through Chapters 1 and 2 of the R Cookbook or a comparable R tutorial.

April 4 Review of Basic Probability and Statistics

Basic probability theory

Random variables

Expected value and variance

Readings:

Kellstedt and Whitten Ch. 6

R:

Basic statistical functions in R

Week 2

April 9 Review of Probability and Statistics (cont.)

Measurement

Principles of sampling

Hypothesis testing and confidence intervals

The normal distribution

Readings:

Kellstedt and Whitten, Chs. 5, 7

R tutorial:

Using R to evaluate properties of estimators

April 11 No Class

Week 3

April 16 Causal Inference

What does it mean for an argument to be “causal”?

Potential outcomes

Causal estimands versus causal estimates
Why experiments are the “gold standard” of causal inference

Readings:

King, Keohane, and Verba, *Designing Social Inquiry* (Ch. 3) [Coursework]
Kellstedt and Whitten, Ch. 3
Jas Sekhon, “Causality.” pp. 1-12. [Coursework]

R:

Practice simulating potential outcomes and experiments in R

April 18 Experiments in the Social Sciences

Field experiments
Lab experiments
Survey experiments
Variations in Survey Design
Review of null hypothesis significance testing, t-tests
Randomization/permutation inference
Handling broken experiments

Readings:

Gerber, Green, and Larimer (2008). “Social Pressure and Vote Turnout.” *APSR* [Coursework]
Olken (2010). “Direct Democracy and Public Goods.” *APSR*.
King et al. (2009). “Public Policy for the Poor?” *The Lancet*. [Coursework]

R:

Analyze experimental data in R

Week 4

April 23 Observational Studies

Natural Experiments

Readings:

Dunning, selected readings TBA
Ho and Imai (2008). “Estimating Causal Effects of Ballot Order from a Randomized Natural Experiment.” *Public Opinion Quarterly*. [Coursework]
Gomez, Hansford, and Krause (2007). “The Republicans Should Pray for Rain: Weather, Turnout, and Voting in U.S. Presidential Elections.” *Journal of Politics* [Coursework]
Devin Caughey and Jasjeet Sekhon (2011). “Elections and the Regression Discontinuity Design: Lessons from Close U.S. House Races, 1942-2008.” *Political Analysis* [Coursework]

R:

Evaluating natural experiments in R

April 25 Analysis of Observational Data for Causal Inference

Subclassification
Matching
Difference-in-Difference

Readings:
Card and Krueger [TBA] [Coursework]

R:
Subclassification and matching
Difference-in-difference examples

Week 5

April 30 Bivariate Analysis and Simple Regression

Review of bivariate correlation
Regression as conditional expectation
Naïve nonparametric averaging
The curse of dimensionality
Least squares regression

Readings:
Kellstedt and Whitten, Ch. 9

R:
Scatterplots
Least squares regression in R
Evaluating fit and regression diagnostics

May 2 Multiple Regression

Regression for Prediction
Regression for Causal Inference

R:
Multiple regression examples
Using multiple regression to analyze experiments

Week 6

May 7 Online Survey Design

Experimental Surveys
Collecting data with Mechanical Turk

Readings:
Berinsky, Lenz, Huber (2012) "Using Mechanical Turk As a Subject Recruitment Tool for Experimental Research" [coursework]

Bailenson, Collins, Iyengar, and Yee (2008), “Facial similarity between voters and candidates causes influence” [coursework]

May 9 Automated Text Analysis

Words as data

Methods for text mining/web scraping

Topic models

Reading:

- Selections from Grimmer [coursework]
- (Optional) Monroe, Colaresi, Quinn (2008), “Fighting Words” [coursework]

R:

Working with the *tm* package in R

Week 7

May 14 Big Data and Political Science

Readings:

King, Pan and Roberts (2013) “How Censorship in China Allows Government Criticism but Silences Collective Expression” [coursework]

Large Datasets and You: A Field Guide, Matthew Blackwell and Maya Sen [coursework]

R:

Working with large datasets with R and MySQL

May 21 Principles of Statistical Graphics

Readings:

- Cleveland (1979), *Visualizing Data*. <http://www.stat.purdue.edu/~wsc/visualizing.html>
- Excerpts from Tufte [coursework]
- Excerpts from Wilkinson, *The Grammar of Graphics*. [coursework]

Week 8

May 23 Measuring Political Preferences

Spatial models of politics

Estimating latent preferences from survey responses and observation data

Using latent measures of political preferences to predict the behavior and choices of voters, donors, and politicians

- Poole (2005) “The Evolving Influence of Psychometrics in Political Science” [coursework]

- Tausanovitch, Chris, and Christopher Warshaw (2012) "Measuring Constituent Policy Preferences in Congress, State Legislatures, and Cities." [coursework]
- "Mapping the Ideological Marketplace", Bonica [coursework]

May 28 Forecasting Elections

Readings:

- Selections from Silver, "The Signal and the Noise" [coursework]
- Selections from Wleizen and Ericson [coursework]

Week 9

May 30 Social Network Analysis

Readings:

- Wasserman and Faust [coursework]
- Selections from Sinclair (2012), "The Social Citizen" [coursework]

Week 10

June 4 Student presentations

June 6 Student presentations