Political Science 151B: Data Analysis for Political Science

Political science is largely an empirical discipline. That is, most of us studying politics do so because we are motivated by real-world political events, either historical, current, or even events yet to happen. We want to know why these events happen and how to make sense of them. Political science tries to answer these questions in a rigorous way. Data analysis is thus a critical component of political science, serving two important purposes: (1) providing numerical descriptions or summaries of political phenomena, facilitating comparisons across time, countries, states, people, etc; (2) testing theories, models and hypothesis about politics.

For the most part, the “data” we will be looking at in this class is quantitative data. We will use computers to manage, inspect and describe quantitative data. We will see how statistical methods let us perform inference: that is, using the data available for analysis to make claims that apply to broader settings. As we shall see, inference is a critical part of the “science” part of political science, since it speaks directly to whether we can validly understand particular political events as specific instances of more general political processes. Inference is necessarily uncertain, and we will use probability when summarizing the results of a data analysis in the form of inferences.

All this is to say that this class involves a little math. Or, as I like to put it, this class is “techie for fuzzies”: an introduction to the way political scientists use the tools of statistics to rigorously understand political events. I assume virtually zero mathematical background on your part, either because you didn’t take math in high school or because you’ve forgotten it. I assume no prior background with using computers for data analysis. The classes will have a heavy “show-and-tell” feel to them, where I will use statistical software to do data analysis.

Assessment: we will have a midterm exam, and a series of short homework assignments where I ask you to analyze some data on the computer and say reasonably sensible things about what you find. I will give you the option of a final exam, or a final research paper (about 10-15 pages) where you tackle a political science research question of your own. This would be excellent preparation for students thinking of writing theses in the Political Science Department, or looking to do research assistance for professors in the Department.

Text: I have ordered Alan Agresti and Barbara Finley’s Statistical Methods for the Social Sciences. This should be available from the bookstore.
**Contacting me:** I am a very reliable e-mail correspondent: try me at jackman@stanford.edu. My office hours are Wednesday, 2-4pm in Encina West (the side facing Hoover), Rm 408.

---

**Tentative Schedule**

1. **Week One, Tuesday, March 30.** Administrative preliminaries. Discuss syllabus, big picture. Introduction to statistical methodology; distinction between description and inference (Text, Chapter 1). Rectangular data files. Types of variables; levels of measurement (Text, Chapter 2).

2. **Week One, Thursday, April 1.** Introduction to Random Sampling, Section 2.2; Sampling and Nonsampling Variability, Section 2.3 (biased sampling, missing data); Other Sampling Methods, Section 2.4 (stratified sampling, cluster sampling, multi-stage).

3. **Week Two, Tuesday, April 6.** Descriptive Statistics. Histograms, bar graphs, skew, the sample mean, outliers, the sample median, quartiles, the mode, variance, standard deviation, interquartile range, box plots. Chapter 3.

4. **Week Two, Thursday, April 8.** Descriptive Statistics, continued. Chapter 3.

5. **Week Three, Tuesday, April 13.** Descriptive Statistics. Chapter 3.

6. **Week Three, Thursday, April 15.** Probability distributions. Chapter 4.

7. **Week Four, Tuesday, April 20.** Sampling distributions. Central Limit Theorem. Chapters 4 and 5.

8. **Week Four, Thursday, April 22.** Confidence interval for the mean and proportions, Chapter 5.

9. **Week Five, Tuesday, April 27.** Large sample hypothesis testing for the mean, proportions. Type I and Type II error. Small sample inference via the t-distribution. Chapter 6.

10. **Week Five, Thursday, April 29.** Small sample inference for proportions via the binomial distribution; Chapter 6.

11. **Week Six, Tuesday, May 4.** Comparison of Proportions, Fisher’s exact test for small samples. Comparing Dependent Samples. Chapter 7.


13. **Week Seven, Tuesday, May 11.** No class.

14. **Week Seven, Thursday, May 13.** Linear Regression and Correlation. Chapter 9.
15. **Week Eight, Tuesday, May 18.**

16. **Week Eight, Thursday, May 20.**

17. **Week Nine, Tuesday, May 25.**

18. **Week Nine, Thursday, May 27.**

19. **Week Ten, Tuesday, June 1.**

20. **Week Ten, Thursday, June 3.** No class, end quarter period.