

Simon Jackman
Spring 2006
jackman@stanford.edu
<http://jackman.stanford.edu>

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* (Prentice-Hall, 1997) and John Verzani's *Using R for Introductory Statistics* (Chapman Hall/CRC, 2005). These should be available from the bookstore. We will make extensive use of the free, open-source software package, R; R is available over the web, such as [this CRAN](#).

Contacting me: I am a very reliable e-mail correspondent: try me at jackman@stanford.edu. My office hours are Thursday, 2-4pm in Encina West (the side facing Hoover), Rm 408.

TENTATIVE SCHEDULE

1. **Week One, Wednesday, April 5.** Administrative preliminaries. Discuss syllabus, big picture (social science, politics, knowledge, inference). 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 Two, Monday, April 10.** 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, Wednesday, April 12.** 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 Three, Monday, April 17.** Descriptive Statistics, continued. Chapter 3.
5. **Week Three, Wednesday, April 19.** Descriptive Statistics. Chapter 3.
6. **Week Four, Monday, April 24.** Probability distributions. Chapter 4.
7. **Week Four, Wednesday, April 26.** Sampling distributions. Central Limit Theorem. Chapters 4 and 5.
8. **Week Five, Monday, May 1.** Confidence interval for the mean and proportions, Chapter 5.
9. **Week Five, Wednesday, May 3.** 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 Six, Monday, May 8.** Small sample inference for proportions via the binomial distribution; Chapter 6.
11. **Week Six, Wednesday, May 10.** Comparison of Proportions, Fisher's exact test for small samples. Comparing Dependent Samples. Chapter 7.

12. **Week Seven, Monday, May 15.** Associations between Categorical Variables. Contingency Tables. Chi-square (χ^2) test of independence. Residuals. Odds ratios as a measure of association in two-by-two tables. Chapter 8.
13. **Week Seven, Wednesday, May 17.** Linear Regression and Correlation. Chapter 9.
14. **Week Eight, Monday, May 22.** Linear Regression and Correlation, continued. Chapter 9.
15. **Week Eight, Wednesday, May 23.** Linear Regression and Correlation, continued. Chapter 9.
16. **Week Nine, Monday, May 29.** No class, Memorial Day.
17. **Week Nine, Wednesday, May 31.** Dealing with confounding: randomized experiments, statistical controls.
18. **Week Ten, Monday, June 5.** A Day for Bayes.
19. **Week Ten, Wednesday, June 7.** Issues in Social and Political Measurement.