POL 204b: Research and Methodology

Fall 2011

MW 9:00-11:00 SSB104 & 139

Professor Scott Desposato
Office: 397 Social Sciences Building
Office Hours: M 1:00-2:00
phone: 858-534-3548 (main office)
email: swd@ucsd.edu
homepage: swd.ucsd.edu
Teaching Assistant Matt Kearney
Office: 323 Social Sciences Building
Office Hours: Thurs 3:00-5:00
email: mkearney@ucsd.edu

Introduction and Course Objectives

This course provides an introduction to the use of quantitative methods in the study of political phenomenon. By the end of this course, you will be able to do everything necessary to conduct basic quantitative research. You will properly and present explain a variety of statistics, both graphical and numerical. You will be equipped to critically analyze others' quantitative work, including their sample framework, methods, and analysis. You will also learn how to use two (or more) statistical software programs, and be able to access and create datasets, prepare them for analysis, and produce results. These skills will be useful in reading, understanding, and conducting research in political science, as well as in many other disciplines.

Reasons to take this course

You have to.

Nearly every first-year PhD student in political science in the United States will be taking a course like this one. Regardless of your field of study, quantitative methods can probably be applied to your research questions of interest. Most of you will use statistics to write course papers, conference papers, Masters' Theses, and Dissertations while at the University of California. Most of you will also use these methods in your professional career. The skills you will learn in this course are absolutely essential for your success in graduate school and academia. They will help you understand the logic of science, read and understand current research, and, ultimately, make it more likely that you will get a good academic job.

If you have absolutely no interest whatsoever in the use of statistics to answer political questions, you should still take this class. You will need the skills in this class (and its sequels) to communicate with and critique the work of other scholars. Also, should you choose to leave academia, solid quantitative skills are very marketable and may help you find employment.

If you are not yet convinced, wait and see. Many anti-quantitative students change their opinions after a few courses on statistics, when they see how powerful these skills can be for answering real and very important questions.

Finally, believe it or not: this stuff is fun.

Preparation

One of the challenges in planning a course like this one is the diversity of the target audience. Some of you already have advanced quantitative training; others haven't seen algebra since high school. My goal is to help everyone in this course become a solid applied statistician.

That said, this course is targeted to an audience with no previous experience in statistics. You should have reasonable math skills, AND be willing to work hard. Calculus is NOT required, but more math is always better. The stronger your math skills, the easier it will be for you to grasp the concepts in this class.

That said, math skills won't guarantee success, and a lack of recent practice won't guarantee failure. Often the most technically skilled students have

difficulty keeping their theoretical feet on the ground, and the least technically skilled students prove the most able to link methods to real questions.

Study Methods

Work hard, work often, and work together.

This course will be challenging. The subject is hard, and we will move fast. The material covered in this course is inherently cumulative. If you do not keep up, you will quickly find yourself too far behind to catch up. Hence you should not expect to be able to blow off this class until the week before the final, unless you are trying to fail (in which case I will be happy to oblige you). I suggest that you review your class notes frequently and try to apply the methods we learn in class immediately, both "by hand" with pencil and paper, and using statistical software.

Working together on assignments will help. If nothing else, misery loves company. But students frequently learn a great deal from each other.

Meeting

The class meets Monday and Wednesday from 9am to 11am. Most of that time will be lecture. On Wednesday, we will sometimes use the second half of class as a computer lab, and sometimes end class early.

Evaluation

- Assignments: 50%

  There will be weekly homework assignments that will include traditional "pencil and paper" assignments, as well as statistical computing exercises. If the assignments are short, these will be fully graded. For longer problem sets, we may grade only a sample of the problems. Half the points will be given for simply completing the assignment, the other half for completing it correctly. I encourage you to work together on all homework assignments, but your final product must be your own.

  In the second half of the course, assignments will include your own data analysis projects and class presentations of your findings.

  Homework assignments will be due at the beginning of each class. Electronic submissions will not be accepted, unless specifically requested in

that format (computer programs).

- Final Exam: 50%

  There will be a cumulative final exam as scheduled by the university.

Late assignments will not be accepted and will receive 0 points.

Policy on Academic Integrity

Students are expected to maintain the highest standards of academic integrity. Cheating, plagiarism and other forms of academic dishonesty will not be tolerated and will be subject to disciplinary action consistent with University rules and regulations. Students are expected to familiarize themselves with University regulations regarding plagiarism and academic dishonesty. If you have any doubt as to what is expected in these regards, please ask.

Additional Information

This syllabus is also posted on the course website. Assignments, updates, and other information will be posted there during the semester.

Materials

Textbooks

The following textbook should be available for sale at the University bookstore:

> Fox, John. *Applied Regression Analysis and Generalized Linear Models*

The Fox book has a number of additional resources available online. There is a website for the book, which includes errata, datasets, and appendicies (some of which are assigned): http://socserv.mcmaster.ca/jfox/Books/Applied-Regression-2E/index.html

To supplement the text, there is a website at UCLA which has computer programs written for Stata, SAS, and SPSS to reproduce the analysis in each chapter: http://www.ats.ucla.edu/stat/examples/ara/

The Fox regression book is fairly challenging and assumes significant prior knowledge. If you are very uncomfortable with the material, you may wish to review a supplemental texts. A basic statistics textbook will provide more examples and intuition, and might be a useful complement to Fox. One that I like is:

De Veaux and Velleman: *Intro Stats*.

This book sometimes comes bundled with "Activstats" which comes with a number of additional multimedia resources and software tools, including a student version of DataDesk (see below).

For basic math skills, two possible supplemental texts are:

Chiang, Alpha, *Fundamental Methods of Mathematical Economics*

or

Simon & Blume, *Mathematics for Economics*, which is usually used in POL270.

There are many, many hypothesis tests out there. A short list of some of the classics is contained in:

Kanji, Gopal K. 2006. *100 Statistical Tests*. Sage Publications: Thousand Oaks.

If you want a basic and intuitive introduction to some of the more advanced methods, consider:

Kennedy, Peter. 2003. *A Guide to Econometrics*. The MIT Press: Cambridge.

This text is widely used in a first-year sequence in "econometrics", which is what economists like to call statistics. Many economics departments will have their first-year PhD students cover the entire text over the course of their first year.

Additional Readings

In addition to the required text, there are some articles that you will probably be asked to read, depending on how our schedule evolves. They are listed below, and scheduled in the course outline later in this syllabus.

Schrodt, Philip A. 2010. "Seven Deadly Sins of Contemporary Quantitative Political Analysis"

Collier, David, Henry E. Brady and Jason Seawright. 2011. "Outdated Views of Qualitative Methods: Time to Move On". *Political Analysis*. 18:506-513.

There are some others that I'll add as well, but not much.

Software

We will learn the basics of two statistical packages this fall: Stata and R.

R is a free, open-source statistical computing package available online at: http://cran.r-project.org. Learning R is challenging, and sometimes a bit of a headache. But it is incredibly powerful, affordable (free!), and will allow you to easily estimate your own models and create custom graphs. Think of it as the workshop or laboratory equipment.

Stata is a commercial, but relatively inexpensive software package. It is probably the most-used software package by political scientists today. It includes "canned" routines for most estimation procedures, and can be programmed to estimate any custom function. It is available in the labs, but I recommend that you purchase it under your academic discount. It comes with a getting started manual. There are extensive online resources, most notably at UCLA's statistical consulting website:
"http://www.ats.ucla.edu/stat/stata/sk/default.htm". In addition, online courses are available from Stata's website, "www.stata.com"; consider spending $150 or so to take one or two next summer.

Stata's strengths include ease of use, a wide range of estimation procedures and some nice graphics. It has some limitations in advanced data management, is missing some advanced statistics, and Stata has the annoying habit of making subtle changes in syntax and file formats across versions. Think of Stata as the minivan of packages. It may not be the best for certain applications, but it will do almost anything. But don't try an engine swap.

Stata now comes in four versions: Small (mostly for small datasets used in undergraduate courses), IC (Intercooled, the base model), SE (Special Edition, handles bigger datasets), and MP (big datasets and parallel processing). If you KNOW you will be working with big datasets right away, consider Stata/SE or Stata/MP. If you aren't sure yet, get Stata/IC and use a lab computer (or your colleagues' computers) if you exceed IC's capabilities.

Why learn two packages? To be honest, two won't be enough. Although some will tell you otherwise, there is no perfect statistical package - all have their strengths and weaknesses. Learning both packages will give you more power and flexibility to accomplish basic statistical computing. Having familiarity with three or four packages will assure you give you the flexibility to work across platforms, co-authors, and computing problems.

Avoid the Excell trap at all costs. Some of the assignments in this course *can* be completed in excell, though you will probably do most of them wrong. In addition, if you avoid learning a "real" package you will regret it for the rest of your career.

There are two other widely used statistical software programs (SAS and SPSS) and many other specialized programs (Limdep, BUGS, Gauss, Xlisp-stat, and so on). We can discuss their utility during the course.

The Syllabus

The syllabus and course outline is intended to provide an overview of the course. You cannot claim any rights from it. In particular, scheduling and dates may change. Although the syllabus should be a fairly reliable guide for the course, official announcements are always those made in class.

New This Quarter / Caveats

This year for the first time I am adopting a twice a week format, with two, two-hour class meetings. I am hoping that this will reduce student fatigue associated with three-hour classes, and that it will allow us to cover more

topics. The other proposed change (I'm open to your thoughts) is that rather than have the last three weeks of class include many student presentations, perhaps we'll only have two sets of student presentations, and cover more topics. Two immediate impacts of this change - we probably won't take any breaks, and we may have to adjust the schedule as we move forward.

As time allows, I will work in some new topics this quarter, including heteroscedasticity, time-series, missing data imputation, matching, regression discontinuity, small and large-sample nonparametric, and an overview of Bayesian methods. That's ALOT of material, but if I can give you an intuitive understanding of some of these topics you'll be in great shape after just a quarter.

We'll also at least partly use the university's online course resources, all at ted.ucsd.edu. If you are not enrolled in the course, I'll need to get you access - please remind me. The site is most useful for massive undergraduate courses, but two things might be especially useful for us: the online discussion forums and the online gradebook. We'll give it a shot; please let me know if you have any suggestions.

Assigments and Lectures

All reading should be completed before class. I may call on individual students to participate in discussions. I reserve the right to adjust the lecture and reading schedule as needed. Weekly topics and readings are listed below.

Week	Topic	Reading
1	Introduction, Visualizing Data, Lab: Basics of Stata	Fox 1,3 Schrodt Article.
2	Basic Probability Lab: Basics of R	Fox D1-D4
3	Hypothesis Tests Competing Approaches to Inference Lab: Data management, tests, simulations	DeGroot 8 Goertz, Beck, & Collier+
4	Introduction to Regression Lab: Regression in R & Stata	Fox 2, 5.1
5	Multiple Regression and Matrix Algebra Lab: Linear algebra in R, Multiple Regression	Fox B.1, 5.2
6	Inference and Regression Gauss-Markov Lab: Advanced data management	Fox 6
7	Extensions: Dummy Variables and Interactions Analysis of Variance Lab: Predicted Values and Presentation Tools	Fox 7, 8
8	Diagnostic Tools and Fixes Student Presentations Lab: Stata examples	Fox 11,13
9	Violating Regression Assumptions + Imputation of Missing Data Student Presentations	Fox 12, 4
10	Maximum Likelihood, Logit/Probit, Matching Student Presentations	Fox 14