

Data Analysis for the Social Sciences

Quantitative Methods in the Social Sciences Program

G4015, Fall 2012

Thursday, 9:00am-12:00pm

252 Engineering Terrace

Instructor: Nick Beauchamp

Email: nb2389@columbia.edu

Office: 712, International Affairs Building

Office Hours: MW, 10am-11am, and by appt.

TA: Lara Pheatt

Email: Pheatt@exchange.tc.columbia.edu

This course is meant to introduce students to the core concepts behind statistical analysis in the social sciences. The chief goal is to help students analyze and interpret quantitative data in useful and provocative ways. Only basic mathematical skills are assumed; this course will develop students' facility with numbers and their relationships, from basic concepts like averages, proportions, ratios, etc, to more complex approaches like regression. These skills will serve them very well in all areas of their lives, especially (but not only!) their future professional lives.

Another important goal of the course is to teach students how to manipulate and analyze data themselves using statistical software. We will focus mainly on the program STATA, but other software will be touched upon, from Excel to the Survey Documentation and Analysis (SDA) archive at <http://sda.berkeley.edu/archive.htm>. The last hour of each class will be devoted to using these software tools to practice analytic techniques and to develop a paper using the General Social Survey (GSS) or the American National Election Study (ANES) data at SDA.

Grades. The course grade comprises: Two Exams, 50%. Independent Project and STATA Labs, 30%. Attendance, Participation, and HW, 20%.

Attendance and Class Participation. Your attendance and participation are necessary at every meeting.

Exams. There will be two take-home exams. They will include short answer, longer answer and multiple choice questions.

Homework. Homework problems will be assigned as the semester progresses. It is expected that you will do your homework. It will be graded.

Textbook. *Statistical Methods for the Social Sciences*, 4th Edition, by Alan Agresti and Barbara Finlay. This book is also on reserve: Two copies are shelved in Lehman Library Reserves and one copy is shelved in Social Work Library Reserves (all under call number QA276.12 .A34 2009).

Other Readings. In some weeks, there will be additional readings from other sources, mostly

from *Introductory Econometrics: A Modern Approach*, 4th Edition, by Jeffrey Wooldridge. These readings will be available on Library E-reserves through Courseworks. Additional readings will also be suggested as the course progresses, to illustrate the concepts and methods in action in actual research articles and books.

Software. The last hour of class will be focused on learning to use STATA. This software is loaded on most of the Columbia computers around campus. If you do not mind using campus computers, you can just use the computer in the lab, and do not need to purchase your own personal copy of STATA. But if you would prefer the flexibility of having STATA at home or on your laptop, then Vanessa may have some STATA licenses to give to you; contact her. Otherwise, Columbia has a deal with STATA, giving you three choices: If you really think you will never use this software again, then the cheapest option is to order Stata/IC with a 1-year license, which gives you 12 months of STATA use for \$95. The next cheapest option is to get Stata/IC with a perpetual license. This will allow you always to have STATA on your computer, for \$155. Finally, if you expect to do a lot quantitative work while at Columbia, you may want to splurge on Stata/SE (Special Edition) with a perpetual license, for \$335; this is the most powerful version available. Further information is available at: <http://www.columbia.edu/acis/software/licenses/stata/index.html>

Independent Data Analysis and Interpretation. In addition to developing a general facility with quantitative analysis using STATA, there will also be a semester-long independent project based on the SDA Archive (<http://sda.berkeley.edu/archive.htm>). This final exercise will require students to integrate many of the skills and lessons learned throughout the semester into a final research report, and more information will be given about this assignment as the semester progresses. Students will have weekly STATA lab assignments throughout the semester that will practice skills and move their projects along. Students will be graded in terms of their ability to operate STATA, select the most appropriate statistics for each type of analysis, interpret the statistics generated, and write brief summaries about what they have learned. In short, you will develop your own social theory using the the GSS or ANES data at SDA. Depending on time, students may present their final results to the class.

Plagiarism and Academic Dishonesty. Students must do all their work within the boundaries of acceptable academic norms. See the Academic Honesty page of the CU website regarding college policy on plagiarism and other forms of academic dishonesty, http://www.columbia.edu/cu/history/ugrad/main/handbook/academic_honesty.html. Students found guilty of plagiarism or academic dishonesty will be subject to appropriate disciplinary action, which may include reduction of grade, a failure in the course, suspension or expulsion. This includes lab reports: if they are copied from another student, severe penalties may be applied.

Late policy. Students will lose points for handing in late assignments, at the discretion of the instructor and teaching assistant.

Note: Changes to this schedule will also be posted on Courseworks along with other announcements.

Schedule of Lectures

- Sept 6 Introduction to Statistics: Ch. 1.
- Sept 13 Sampling, Basic Descriptives & Probability: Chs. 2-4.
- Sept 20 Hypothesis Testing: Chapters 5-7.
- Sept 27 Regression 1: Ch. 9, through p. 283; Ch. 10.
- Oct 4 Regression 2: Ch. 11 through p. 340.
- Oct 11 Regression 3 (The Assumptions): p. 283-289; p. 345-355; Ch. 14 through p. 462.
Log Transformation, *Wooldridge*, p. 43-46.
- Oct 18 Regression 4 (Interactions): p. 340-345; p. 384-389.
Quadratics, p. 463-468.
- Oct 25 Linear Probability Model: p. 484; *Wooldridge* p. 246-251.
Scales, p. 532-537.
- Nov 1 Logistic Regression 1: Ch. 15, p. 483-493 and 496-502.
Take-Home Midterm Handed Out
- Nov 8 Logistic Regression 2 (Assumptions, Diagnostics, & Postestimations): Ch. 15, p. 493-496 and p. 507-512.
- Nov 15 First Differenced Analysis; Fixed Effects. *Wooldridge* p. 455-465 and Ch.14.
Lagged Dependent Variables, *Wooldridge* p. 310-312.
- Nov 29 Categorical Data Analysis & ANOVA, p. 369-376.
- Dec 6 Review and Presentations.
Take-Home Final Handed Out