

**SOCIOL 400: INTRODUCTION TO QUANTITATIVE DATA ANALYSIS
FALL 2008**

Class: Tuesday, 2-5pm, Parkes 222
Lab: Monday, 2-3pm, Library B182

Professor: Jeremy Freese
1810 Chicago Avenue, Rm 222
e-mail: jfreese@northwestern.edu
Office Hours: T, 10-12 (and by appt)

TA: Ben Ruiz
1812 Chicago Avenue, Rm 108
e-mail: b-ruiz@northwestern.edu
Office hours: T 12-2

OVERVIEW

This course provides an introduction to the principles of statistical reasoning and data analytic practice that are presumed by a graduate-level introduction to regression analysis. The goal is to start developing your abilities to draw substantially meaningful and accurate inferences from quantitative social data and to assess claims made by others. Any practical treatment of quantitative methods in social science involves computing; this course includes an introduction to the use of general-purpose statistical software.

PREREQUISITES

We presume familiarity with linear algebra, arithmetic and logical reasoning, and the animating goals of social science. All statistical computing in the course will be conducted using Stata.

READINGS

The required book for the course is:

Agresti, Alan and Barbara Finlay. 2009 [sic]. *Statistical Methods for the Social Sciences, Fourth Edition*. Upper Saddle River, NJ: Pearson Prentice Hall.

The book should be available from the NU bookstore and also online.

COURSE REQUIREMENTS

Final grades for the course will be based on the combination of weekly exercises, Stata exercises, and one in-class examination. The rough heuristic as I type this syllabus is ~40% for the weekly exercises, ~35% for the Stata exercises, and ~25% for the examination, *except* for the requirement of a certain level of performance on the final exam as discussed below. Do not expect the same exactitude regarding calculation of your final grade (e.g., specific cut-offs and policies regarding rounding scores) that characterize undergraduate courses.

Weekly exercises. Exercises will be due at 9pm each Monday night from 10/6 through 11/24. These will be numbered exercises from the Agresti and Finlay textbook. You are encouraged to discuss your work with your fellow students and to learn from them, but you must complete your work on your own. Most certainly, what you write must manifest your own understanding and words.

Exercises should be e-mailed to me (jfreese@northwestern.edu), with a subject line that makes their contents clear. I aspire to go through all the exercises prior to each Tuesday class so that we can reconsider any matters on which there were problems. Given the imperative for timely turnaround and the possibility of discussing problems in class, I will not provide extensive marking on the exercises and will be assigning them a holistically-derived assessment of up to 10 points. No promises about the fate of any exercises turned in after the deadline.

While you can e-mail a comment, question, or suggestion regarding the course at any time, the e-mail you send to turn in your exercise may be a particularly propitious time to do so. I sometimes require students to send a weekly “check-in” of some kind to make sure we are in dialogue, but it seems superfluous to have that as a separate requirement when the structure of the course already has you providing something weekly.

Stata exercises. The course will include several exercises that require you to perform operations in the statistical package Stata. These will be given to students in class and also posted on Blackboard.

There will always be a lab session between when the assignment is presented and its due date. Exercises should be e-mailed both to me and to the TA for the course. Statements above about the importance of doing your own work and turning it in on time apply.

Stata is available for use in several locations on campus, including virtually through the Social Sciences Computing Cluster. Stata is also available for purchase at a substantial discount to graduate students. The subsequent courses in the statistics sequence in sociology also use Stata. Stata has a modest learning curve, especially if one is unfamiliar with command-line interfaces, but once you grasp its architecture it is sublime.

For data, many examples in the course will draw upon the General Social Survey, a survey of adults in the United States that has been conducted annually or biennially since 1972. An abridged copy of GSS data is available on Blackboard. The General Social Survey presently has perhaps the most confusing web presence of any major social science data resource; presently, two less useful websites are easier to locate using standard search engines than their “real” website. The URL for their real website is:
<http://www.norc.org/GSS+website/>

Examination. This course partly serves a certification function, in the sense that a suitable grade in the course is supposed to mean that one has adequate background to continue with the regression sequence. We will have a closed-book exam at the conclusion of the course to assess this adequacy. *You must attain at least a B on this exam in order to attain*

at least a B overall in the course. The examination will not require the memorization of specific equations or any especially demanding calculations.

ADDITIONAL GUIDELINES FOR SUBMITTING OUTPUT

Including output. For assignments involving the estimation and interpretation of data in Stata, as well as your final paper, you will turn in Stata output along with your assignment. Details of this will be discussed with the assignments. But:

- (a) As part of *.do* files you use for generating results for assignments, you must use comments (i.e., using lines preceded by * or //) that indicate what part of the output corresponds to what.
- (b) You must highlight numbers in your output that correspond to numbers in your assignment. You can use, for example, boldface type to highlight.
- (c) You will preserve output using Stata's log commands for doing so (rather than, e.g., copying and pasting).
- (d) You must use a fixed-width font (like Courier or Andale Mono) for the output and your lines must not wrap.
- (e) As with anything else you turn into others in This Academic World, we advise that you look over your work with an eye for how it may look through the eyes of another person. In general, anything sloppy will irk.

COMMUNICATION

Students are presumed to check e-mail regularly. We will use e-mail to send announcements to the class as needed. *I strongly prefer e-mail to the telephone as a means of contact regarding the course.*

Blackboard will be used as a repository for this syllabus and other course materials. Please notify me by e-mail of any technical or other problems with materials provided via Blackboard.

SCHEDULE OF TOPICS AND READINGS

Wk	Date	Topic	Reading
1	9/23	Orientation	Syllabus
2	9/30	Sampling and Measurement	Agresti and Finlay, Chapter 1-2
3	10/7	Descriptive Statistics and Sampling Distributions	Agresti and Finlay, Chapter 3-4
4	10/14	Confidence Intervals and Significance Testing	Agresti and Finlay, Chapter 5-6
5	10/21	Comparing Groups and Cross-tabulation	Agresti and Finlay, Chapter 7-8
6	10/28	Bivariate Regression and Correlation	Agresti and Finlay, Chapter 9
7	11/4	Multivariate relationships and multiple regression	Agresti and Finlay, Chapter 10-11
8	11/11	Analysis of Variance	Agresti and Finlay, Chapter 12-13
9	11/18	Statistical models	Agresti and Finlay, Chapter 14
10	11/25	TBD	
11	12/2	Examination	