Statistical Methods
Bunche Institute, Summer 2006
Professor: Scott de Marchi

Office: 407 Old Chem
Office Phone: 660-4342
Home Phone: (919) 304-4907
Email: demarchi@duke.edu
Web: http://moria.poli.duke.edu/overallpage.htm

Course Description:

The purpose of this course is to expose advanced undergraduates to the type of reasoning and model building that is most prevalent in empirical work in the social sciences. By and large, every graduate program in the United States requires political science students to acquire solid empirical skills; this class is a first step in understanding this paradigm of research. The ability to take advanced coursework in research design and empirical work is very important, as often your research will push you to learn a very specific set of advanced statistical methods. Most students, however, need a great deal of help when they approach this material for the first time, due to

a. a lack of mathematical training at the college level (i.e., calculus and linear algebra);
b. poor preparation in the skills necessary to design and test a model.

Obviously, the length of the course forces us to select the *core* material of a graduate level statistical methods seminar. Our focus will be on developing the intuition, reasoning, and language skills (where the language is mathematics) necessary to begin independent research. By taking this course, students will begin a process of using data and statistical modeling to produce evidence that bears upon research questions of interest to the student. Further, this course will allow students to look at the research of other scholars critically and with a solid understanding of the statistical methods commonly employed by social scientists.

Topics:

Week 1:

Reading (Thursday):
http://moria.poli.duke.edu/mug/prelude.pdf (de Marchi book chapter).
Chapters 1-3 Moneyball (Michael Lewis).
Pages 1-26 Introduction to Econometrics (Christopher Dougherty).

1) Modeling topics
    a. A brief history of the social sciences

b. Causality and Inference (or, when does one thing cause another)
c. Constructing measures for phenomena of interest
d. Data and sampling (in particular, the difference between a sample and a population)
e. Testing relationships between variables with statistical models

2) Mathematical topics
   a. Random variables, expectation, and distributions.
   b. Functions, in particular linear functions
   c. Maxima / minima of functions (and how to find them)
   d. Basic ideas of probability theory
   e. Random vs. stochastic
   f. The idea of a statistical model

Week 2:

Reading (Tuesday):
Finish Moneyball (Michael Lewis).
Chapters 1 and 4 Freakonomics (Steven Levitt).
Pages 30-65 Introduction to Econometrics (Christopher Dougherty).

Reading (Thursday):
Reread pages 48-65 Introduction to Econometrics (Christopher Dougherty).

1) Modeling topics
   a. What's a good question?
   b. What are the conditions for a field to make progress? Is progress possible?
   c. Models, assumptions, and a bit of epistemology
   d. Correlation vs. causation and testing models
   e. Assumptions required for most statistical models

2) Mathematical topics
   a. Distributions (including how to visualize them with histograms)
   b. The Normal distribution
   c. OLS regression
   d. Error terms and why they matter

Week 3:

Reading (Tuesday):
Pages 65-79 Introduction to Econometrics (Christopher Dougherty).
Chapter 3 Freakonomics (Steven Levitt).

Reading (Thursday):

http://zia.hss.cmu.edu/miller/papers/tiebout.pdf (Kollman, Miller, Page).

1) Modeling topics
    a. What can go wrong with a statistical model?
    b. EITM style research

2) Mathematical topics
    a. Testing for significance
    b. Multivariate regression
    c. Overall measures of model fit

Week 4:

Devoted to an overview of all topics covered in the course. If time permits, sundry comments on multivariate regression and binary dependent variables will be made. Last, we will focus on the student projects!

Assignments and Grades:

Assignments fall into three main categories:

1) This course will have numerous quizzes throughout the semester; in general, expect a quiz each class. These quizzes are not graded, but they do provide me with an idea of your progress in the course, and I'll speak to you if you aren't performing at a decent level on them.
2) You will be asked to complete ***two*** models throughout the class. These assignments will require you to work with actual data and a statistics package (Stata). Teaching assistants for the course will be available in the computer lab to help you with these assignments. These assignments are graded, and each counts for 1/3 of your grade.
3) You will have a final exam. It will be take-home, and will involve some modeling work with data. It will be graded, and will count for 1/3 of your grade.

Given this is a course for aspiring graduate students, grading is used more to gauge progress than to evaluate the capabilities of students (especially since a wide range of backgrounds exist in any class). If you make an honest effort, your final grade won't be a problem. The key, however, is to see me if at any point you feel you are in over your head. The teaching assistants will also be available to help you over any rough spots. This material will likely be difficult, but at the end of the course you will have a new set of skills that are tremendously useful. To master these skills will require effort, and you may (wrongly) feel rather dim at times.

Assignment, Due Thursday, June 8 and Tuesday June 13.

Part 1 (June 8): First, define your research question in 1 page of text. Second, demonstrate that a data set exists that would allow you to tackle this question. Finally, define (list the actual questions) the dependent variable and 5 independent variables that you think are essential in addressing this question. Make certain to specify your expectation about the importance and signs of the independent variables in question.

Part 2 (June 13): Revise this statement to reflect the criticism you've received from your TA.