POLI 891 – Applied Data Analysis in Political Science

MW: 12:30-1:45
Hamilton Hall 351
(Updated 9-28-2009)

Instructor

Dr. Thomas Carsey
E-mail: carsey@unc.edu
Office: 317 Hamilton Hall
Phone: 962-1207
Office Hours: Monday and Wednesday, 2:00-3:30 and by appointment

Course Description

This course focuses on the application of statistical analysis to quantitative data in order to study theoretically and substantively interesting questions about politics. We will consider attributes of single variables, including their distributions and measures of central tendency and dispersion. We will then consider measures of association between two variables and introduce the concept and mechanics of hypothesis testing. We will conclude with the basic linear model as a way to introduce additional covariates. For each topic, we will examine methods for both categorical and continuous variables.

This course is designed for students in a terminal Masters degree program, particularly for those in the Trans Atlantic Masters program. The course may be appropriate for students in other MA programs some Ph.D. students in programs that do not require statistical/quantitative methods who are seeking an introduction to applied data analysis.

This course will have a general social science focus, but we will make heavy use of examples from Political Science. The class will cover some underlying mathematics and probability theory, but the emphasis will be on training students to properly use and interpret the results of quantitative methods. This goal will be supported by a broader consideration for general questions of research design and inference. Thus, while the weekly material will mostly consist of applied statistics, we will never lose sight of the fact that we are social scientists interested in substantive questions about politics.

Statistical methods are an essential component of modern social science training. It is not possible to engage the scholarly research in any area of social science, nor engage in serious applied policy research, without some basic understanding of statistical methods. It is also increasingly difficult to do your own research and analysis without these tools. That does not mean that statistical analysis is the only way to conduct research or gain insights into a political phenomenon, nor does it mean that it is always the best method to employ. It simply means that statistical methods are powerful tools that are widely used. Because of this, I strongly encourage you to push yourselves to get the most out of this class that you can.

This course will help you become users of statistical methods and critical consumers of research employing such methods. You will not be experts by the end of this course, but you will have built a strong foundation. We will learn mathematical/statistical principles, but what I'm really striving for is the development of your intuition about tackling empirical research problems. In other words, I want you to do more than learn the math – I want you to learn to think like a systematic empirical social scientist.

I am a big believer in learning by doing. Thus, we will have regular assignments in this class that require students to engage the topics we are covering. More will be said about the assignments below. Ultimately, social scientists need to be able to translate theories into testable hypothesis, develop models

that capture those hypotheses, apply appropriate methods, interpret the tests, and return to the theory in order to evaluate it. I often think about this process as being able to move seamlessly back and forth between words, pictures/figures, and equations. This is the hardest part of becoming a successful social scientist – just getting the statistics right is only a part of the process. The intuition I noted in the previous paragraph is critical in this larger process because it provides a mechanism to facilitate translating our theories of politics into testable statistical models of politics without loosing something in the translation. Of course, getting the stats right is a critical part for many scholars, and it is the central task of this course, but I want to make sure that students do not view learning about methods as something different and separate from learning and thinking theoretically about politics.

Course Requirements

<u>Grading</u>

Grading for the course will be based on a midterm exam, a final exam, and a series of assignments. How much each counts toward the final course grade is detailed below. The exact nature of the exams and assignments will be described in greater detail below and in class.

Assignments:	30%
Midterm Exam:	30%
Final Exam:	40%

<u>Required Text:</u>

Using R for Introductory Statistics by John Verzani

<u>Additional texts students may wish to review:</u>

Introduction to the Practice of Statistics by David S. Moore and George P. McCabe
Introduction Statistics with R by Peter Dalgaard
Designing Social Inquiry by King, Keohane, and Verba
Research Methods in the Social Sciences by Chava Frankfort-Nachmias and David Nachmias

<u>Additional Reading:</u>

The course may also include additional readings and other materials that will be available electronically on the course website.

Reading research methods textbooks is not like reading novels. You are expected to read the assigned material before coming to each class meeting, and you would certainly benefit from reading it again afterward. There is no substitute for just hammering away at this material, and I can tell you that the better you understand the material in this course, the better off you will be down the road in other courses (both substantive and methodological) and in writing your own papers. Class time will be devoted to nailing down the basics and making sure you know when and when not to use the methods we discuss. There will always be more material to cover than we have time for in class, so in that sense, you'll always be left wanting for more.

<u>Attendance:</u>

Attendance is required for every class meeting. Each class meeting will build on the material from previous sessions, so falling behind or missing some material is not an option. I expect students to be here every time and to come to class prepared. We simply have too much to cover to proceed any other way. Failure to attend will result in your overall course grade being lowered.

<u>Assignments:</u>

There will be an assignment handed out nearly every week. They will typically be handed out on Wednesday and due the following Monday. Some will require computational work by hand, but most will require use of a computer. All computer assignments should be conducted using the statistical software R. R is freely available. You can download the program online at: http://cran.r-project.org/ and you can learn more about R in general at the R-project homepage (http://www.r-project.org/). R is best thought of as a statistical computing environment rather than as software. R is not a point-and-click program like SPSS or even STATA. Learning R can be a bit more challenging than learning a point-and-click program, but it is much more powerful and flexible. R also has very good graphics capabilities, and R is increasingly the computing environment of choice for those doing statistical work across a wide range of disciplines including Political Science. I already mentioned that it is free – it is also compatible across platforms (Windows, Mac, etc.). These two features make R ideal for those going overseas or away from a university to do research. Also, because anyone can put R on any computer, any classroom becomes a de facto computer lab if students have laptops available.

There is no substitute for reading the documentation for R. I **<u>STRONGLY</u>** recommend that you begin with the manual available online called "An Introduction to R." This document is about 100 pages long, and it provides the R basics. You can find the manual by clicking the "Manuals" link on the CRAN homepage. The direct link to the .pdf file is here: http://cran.r-project.org/doc/manuals/R-intro.pdf This manual is also downloaded and stored on your computer when you install R.

There are also some very helpful short reference documents for R commands that you might want to print and keep handy, located at: http://www.rpad.org/Rpad/R-refcard.pdf and at: http://www.psych.upenn.edu/~baron/refcard.pdf . John Fox has a couple of very useful websites for materials on R. First, he wrote a book on Applied Regression, the web site for which is: http://socserv.mcmaster.ca/jfox/Books/Companion/index.html. Second, he taught a two-day workshop on R here at UNC recently and created a website for that, which is located at: http://socserv.mcmaster.ca/jfox/Courses/R-course/index.html. The statistics consultants in the Odum Institute at UNC also support R, and we will devote the time necessary in class to get students up and running in R.

While I will provide a lot of support and direction with R, you need to take the responsibility yourself to learn the tools you need to do your work. R has extensive help menus and a website supporting it. It is O.K. to ask each other questions when working on assignments, but you have to know how to do this stuff on your own. Your learning will be greatly enhanced by banging through the assignments, and that will no doubt be reflected on the exams. I would rather see you make your own mistakes on the assignments and learn from them as opposed to copying correct answers from others but not really understanding what you are doing.

Later in the semester, the assignments will get more involved. They will be longer and will make use of real-world data more so than just simulated data. These longer assignments will count more toward the final grade and you will be given more time to work on them.

<u>Exams:</u>

The exams will be a combination of conceptual questions, technical questions, and questions that ask you to interpret information provided to you. You will not need a computer for the exams, but you may want a calculator. Exams will take place in the classroom and will be closed-book and closed-note.

Communication

I make every effort to communicate to you my expectations, your responsibilities, and the substantive information covered in this course. I will send e-mails to the entire class. I maintain a Blackboard site for the class, and I will make announcements and issue some reminders in class. Note that I will only send e-mail out to your UNC e-mail accounts as listed on the course roster. I am also very easy to reach if you need to communicate with me. Come to my office hours, call my office, or even better, send me an e-mail. It is important for you to stay in touch, particularly if any problems arise. I don't like to change exam schedules. If a situation arises where I need to give a make-up exam, I reserve the right to give it during the final week of the semester. I reserve the right to give a make-up exam that differs substantially from the normal exam in order to protect the integrity of the exam process. I or any faculty member will be much more understanding if you just communicate with us early and up front.

A Note on Academic Honesty

Cheating, plagiarism, and all other forms of academic dishonestly are easy to spot and come with severe consequences. All students should familiarize themselves with the Academic Honor Code at UNC (http://honor.unc.edu/honor/code.html). Students caught violating the Honor Code may receive an F for the course and may be turned over for further disciplinary action by the University. By taking this class, you have committed to comply with all aspects of the Honor Code regarding all aspects of this course.

Students with Disabilities

Students with disabilities needing academic accommodation should; (1) contact the office of Learning Disabilities at UNC (http://www.unc.edu/depts/lds/index.html), (2) bring a letter to the instructor indicating the need for accommodation and what type. This should be done during the first week of class.

Responsibilities

The success of this course depends upon all of us meeting our responsibilities. I am responsible for being prepared each week to present and discuss course material, for challengingly you academically and stimulating your curiosity, and for being available for and responsive to your questions and inquiries. You are responsible for being prepared each week as well, for asking questions when you are confused and actively engaging the material, for doing your own work, for meeting the course requirements, and for pushing yourselves to get the most out of this course that you can. Ultimately, this is your education and you should take responsibility for it.

Course Schedule

The schedule provided serves as a guideline for the semester. As we proceed, we may discover that some topics take a bit longer than expected to cover while others take less time. We may also add or change a few of the topics along the way. Readings associated with each topic are listed on the schedule and should be read by you prior to coming to class. It may be the case that additional readings will be assigned during the semester. Those readings will be provided for you either in class or online. Announcements regarding such changes will be made in class and distributed to students via e-mail. However, the dates for the exams will NOT change, nor with the due date for the project.

DATE	TOPIC/ASSIGNMENT
Aug. 26	Course Introduction and Overview
Aug. 31	Experimental Data, Observational Data, Description, and Causation
Sept. 2	APSA – No Class
Sept. 7	Labor Day – No Class
Sept. 9	Introduction to R – Verzani, Chapter 1
Sept. 14	Introduction to R (continued)
Sept. 16	Single Variables – Verzani, Chapter 2
Sept. 21	Single Variables – Verzani, Chapter 2
Sept. 23	Two or More Categorical Variables – Verzani Chap. 3 (pp. 69-76), Chap. 4 (pp. 105-108)
Sept. 28	Two or More Continuous Variables – Verzani, Chap. 3(pp.77-104), Chap. 4 (pp.109-140)
Sept. 30	Two or More Continuous Variables – Verzani, Chap. 3(pp.77-104), Chap. 4 (pp.109-140)
Oct. 5	Describing Populations – Verzani, Chapter 5
Oct. 7	Describing Populations (continued)
Oct. 12	University Day – No Class
Oct. 14	Catch-Up and Review
Oct. 19	**MIDTERM EXAM**
Oct. 21	Fall Break – No Class
Oct. 26	Hypothesis Testing/Confidence Intervals – Verzani, Chapter 7
Oct. 28	Hypothesis Testing/Statistical Significance – Verzani, Chapter 8
Nov. 2	Hypothesis Testing (continued)
Nov. 4	Goodness of Fit – Verzani, Chapter 9
Nov. 9	OLS – Verzani, Chapter 10 (Review Chapter 3, pp. 90-102)
Nov. 11	OLS (continued)
Nov. 16	OLS (continued)
Nov. 18	OLS (continued) Introduction to Pooled Times Series
Nov. 23	OLS (continued) Introduction to Pooled Time Series
Nov. 25	Thanksgiving Break – No Class
Nov. 30	Logit and Probit – Verzani, Chapter 12, pp 343-351
Dec. 2	Logit and Probit (continued)
Dec. 7	Logit and Probit (continued)
Dec. 9	Catch-up and Review
TBA	**FINAL EXAM**