POLI 784 - Intermediate Statistics, Spring 2011
Tuesday and Thursday, 9:30-10:45, Hamilton Hall 351

Weekly Lab (POLI 891-2); Thursday 1:00-1:50, Hamilton Hall 420

Instructor

Dr. Thomas Carsey
E-mail: carsey@unc.edu
Office: 317 Hamilton Hall
Phone: 962-1207
Hours: Tuesday and Thursday 8:30-9:30, and by appointment.

Teaching Assistant

Elizabeth Menninga
E-mail: menninga@email.unc.edu
Office: 301Hamilton Hall
Hours: Friday 11:00-Noon, Monday Noon-2:00, and by appointment.

Course Description

This course is focused almost entirely on Ordinary Least Squares (OLS) regression. We will examine the properties of OLS, the assumptions underlying the model, the consequences of violating these assumptions, how you can detect such violations, and how you might begin to respond to them. This course builds on what students learned in POLI 780 (Scope and Methods) and POLI 783 (Statistics), and assumes that all students have completed these courses or those functionally equivalent. Students who have not taking POLI 783 in particular need the instructor's permission to take this course.

This is a great course to teach and to take! You already have a basic understanding of probability and probability distributions, hypothesis testing, and some simple statistics. You have also been introduced to questions of measurement, concepts like reliability and validity, and many aspects of research design. In short, you have many of the basic building blocks in place to begin doing real systematic quantitative social science research, but you have probably not really done it yet. In this class, you will. We will devote ourselves in this class to quantitative analysis, but the skills you refine in this course apply to systematic empirical research of all stripes. Thus, while this course appears to be a continuation of your statistical training alone, it has as much to do with theory and design as it does with crunching numbers. Statistical analysis is an important tool used by political scientists, but for most of us, the real goal is to learn something about politics, not just something about mathematics.

This course will help you become increasingly skilled users of statistical methods and critical consumers of research employing such methods. You will not be experts by the end of this course. In fact, you won't be close. However, you will have built a very strong foundation for future coursework, you will be better-positioned to read more of the literature, and you will be well-versed in using OLS regression to answer your own research questions. OLS and the linear model provide the basis for most published empirical political science research, and the jumping off point for most of the more sophisticated methods you will see and use. As a result, this course is extremely important in your Ph.D. training. I strongly encourage you to push yourself to get as much out of this course as you possibly can.

I believe that in order to engage the scholarly community in virtually every subfield of political science, one needs an understanding and working knowledge of statistical methods. That is not to say that statistical analysis of large data sets constitutes the only, or even necessarily the best, approach to

conducting research. However, you would be hard-pressed to find an area in political science where important work using statistical methods is absent. Furthermore, the general logic of the methods we will explore extends beyond large-N quantitative studies.

I have trouble thinking about a course in statistics that is not mathematical at some level, so of course we will be doing math in this course. However, I do not think the math will be a barrier to anyone (particularly anyone who took POLI 783). We will work through the math with the goal of providing a deeper understanding of the concepts under consideration, but it is that understanding, and not the math itself, that is the primary goal. I often use the word 'intuition' to describe the level of understanding that I want students to have regarding statistical methods. That intuition is not a substitute for the math, nor do I mean that you should be satisfied with some sort of general sense of what is going out without understanding the math. What I do mean is that understanding the logic of quantitative analysis runs deeper than just a set of mathematical rules and formulas. A constant theme in the course will be on why a practicing political scientist would want to know about the statistical topic at hand. If I fail to make it clear at any point in the semester why we are learning what we are learning, you should press me on it.

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 part of the process. The intuition I noted in the previous paragraph is critical in my view in this larger process, as 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

There is one required text for the course:

Basic Econometrics, 5^{h} Edition, by Damodar N. Gujarati and Dawn C. Porter (McGraw-Hill).

The text is no-doubt expensive, but I believe it is worth having. This is the sort of book you will place on your shelf and return to for years to come. This particular text is highly regarded and widely used for similar courses.

I have also listed one recommended book for this class:

An R Companion to Applied Regression, 2nd Edition, by John Fox and Sanford Weisberg, Sage.

This text is a companion to Fox's textbook *Applied Regression Analysis and Generalized Linear Models*, which itself is a very good book. The value of this companion text is that it will help you work through doing OLS analysis in R. I will have more to say about that below as well.

We will not read every chapter in Gujarati and Porter, but we will read many. It is pretty readable as statistics texts go, but it is still a statistics text. Thus, reading through this is not like reading a novel, nor really like anything you will read in your substantive seminars. You are strongly encouraged to read the assigned material before coming to each class, 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), in writing papers, in writing your dissertation, in publishing, in getting a job, and in getting tenure. 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.

I don't take attendance formally for graduate classes, but it will be obvious who is and is not here every class meeting. For a class like this, it is imperative that you keep up with the readings, assignments, and lectures. Thus, missing class is really not an option. I also expect you all here on time and ready to go every class meeting. We simply have too much work to do to proceed any other way.

Finally, there is a 1-credit hour lab/workshop associated with this course that all students are **required** to take (POLI 891-2, listed under my name). The lab sessions will be led either by myself or the TA. Some time will be devoted to mathematical issues, but much of the time will be devoted to practical issues associated with managing and analyzing data. This will include some support to help with the assignments for the course, and will also include some training in both R and STATA.

Assignments and Grading

We will have 2 exams in the class, a paper/project, and several assignments. You will also receive a grade for the lab, though that grade will be folded into your performance in the class overall and you will be given the same single grade for both the 3-credit class and the 1-credit lab.

20%	Midterm exam
30%	Final Exam
10%	Course assignments
10%	Lab (participation/assignments, etc.)
25%	Paper/project
5%	Paper comments

I reserve the right to make minor adjustments to final course grades based on overall performance in the class. I will not accept late papers or late assignments unless a compelling reason is provided to me in advance or a serious unanticipated problem arises (NOTE: computer or printer problems do not qualify). Anticipate having problems with the assignments and the paper and plan accordingly.

There will be an assignment handed out for the main course nearly every week. They will typically be handed out on Thursdays and due the following Tuesday. You can and should collaborate on the assignments, but you need to learn the material for yourself. I do NOT want to see identical assignments turned in by students. Use each other as a resource, but NOT as a crutch. Some of the assignments will require some computational work by hand, but most will require use of a computer. All assignments that require use of the computer will be conducted in 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. There are some Graphical User Interfaces (GUI's) available for R, but we won't be using them. Instead, you will be writing text files, called script files in R, that send R a series of commands to execute. Learning R can be a bit more challenging than learning a point-and-click program, but it is much more powerful, flexible, and is increasingly the computing environment of choice for those doing statistical work across a wide range of disciplines including Political Science. More importantly, our goal in this class is to learn about statistics, NOT about software. Using a programming environment like R is a far-superior way to teach you about statistics than is using a point-and-click program.

There is no substitute for simply reading the documentation for R. I **STRONGLY** recommend that you begin with the manual available online called "An Introduction to R." This is about a 100-page document that provides the core basics to understanding R as a statistical computing environment. 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.

Springer books (http://www.springer.com) has an entire series of books in their Use R series that are designed to be practical applications of R for users. Many of these can be accessed through the UNC library online for free. One in particular that is quite useful is *Data Manipulation with R* by Phil Spector. Another is *A Beginner's Guide to R* by Zuur et al.

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 last term and created a website for that, which is located at: http://socserv.mcmaster.ca/jfox/Courses/R-course/index.html. Finally, our own Jeff Harden has a web site that includes links to a number of helpful R documents and sites, located at: http://www.unc.edu/~jjharden/

Jeff will also be teaching a short workshop on R January 13th and 14th through the Odum Institute. I **STRONGLY** encourage you to sign up. The statistics consultants in Odum also support R, and we will devote the time necessary in class and in the workshop to get you all up and running in R. R is an extremely powerful and flexible computing environment, but it is not quite as easy to use as some other programs like STATA or SPSS. You may also have some assignments in STATA (which is a much better choice than SPSS). You may do your paper in whatever software you want, but the TA and I will be best able to help you with STATA and R.

While we will provide support and direction with R and STATA, you need to take the responsibility yourself to learn the tools you need to do your work. Both programs have extensive help menus and websites supporting them. Both are programmable, particularly R. It is O.K. to ask each other questions when working on assignments and such, but ultimately 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. That is in part why the assignments only count for 10% of your grade.

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.

The Paper/Project

The paper/project is designed to be a piece of original quantitative analysis conducted by you during the course of the semester. For this class, you should pursue a paper that is a replication and extension of an existing published paper. This will make it easier for you to present the literature review and theory sections for your paper since they will be closely tied to the paper you replicate. Your emphasis for the paper will be on properly analyzing data to test your hypotheses. You should plan to use an OLS model for your paper. If you have concerns about the appropriateness of OLS for your paper, or if you need assistance in developing a paper topic, you should consult with me early and often in the semester. There

is no formal page length, but for most of you I expect the paper will constitute 14-20 pages of text. You should model your paper after the quantitative papers you have seen in journals like APSR, AJPS or JOP, with the caveat that the front-end of your paper (everything up to the Data/Methods section) will be shorter than the typical journal article, and that you'll be asked to provide a bit more detail in the back half of your paper regarding the analyses, tests, etc. that you performed.

By "replication," I mean that your first task will be to reproduce the findings exactly as shown in the published report. This DOES NOT mean simply contacting the author or using a so-called "replication data set" in which all of the coding, modeling, and estimation decisions/commands are already done for you. Rather, it means going back to the primary (electronic) data source and proceeding from there. For example, suppose you are replicating a study by Bill Smith that uses survey data from the National Election Studies (NES) series. Rather than ask Bill Smith to send you any computer files that record all of his coding decisions that you simply have to run, I want you to download the original NES data, locate the proper variables, make any coding changes, and perform the analysis. In other words, your first task is an independent replication/verification of Bill Smith's reported analysis. If that is not possible, we can talk about contacting the original author and other strategies for proceeding.

By "extension," I mean that once you have replicated the results of an existing study, you will then build upon that analysis in some way. This might involve using a different coding of a variable, adding additional variables, considering different (maybe non-linear) model specifications, or adding additional data. Whatever extension you attempt, however, must be derived from a clear theoretical proposition and/or a clear methodological critique. In other words, don't toss in an interaction term, "just to see what happens." Remember, this is not just an exercise in number crunching – you are writing a paper with the goal of answering a theoretically motivated research question. However, by going through the process of trying to replicate another scholar's study, I hope you will learn the value of documenting every step of the research process.

Finally, all students will read drafts of two other student papers in the course and provide written comments to them. Seeing other people's work and seeing how others react to your work helps you to improve your skills as a researcher. You might be a bit nervous about sharing your work with others, let alone receiving their written comments and providing such comments yourself. It's O.K. to be nervous, but it is also time to start getting used to this. It is better to begin this among friends and colleagues before you have to deal with anonymous reviewers.

I do not expect perfect papers ready for submission to APSR by the end of the semester. However, I do expect your best professional effort. The only way that I and your classmates can help you to improve is if you do the best you can on your own with your first draft so our advice is focused on how to push beyond that. Don't worry – it will be fun!

Note there are several due dates for aspects of the paper listed in the course schedule. I expect you to provide me with at least what is listed for that date on that date. If you give me more, I will read more. The only part of the paper assignment that is graded is the final version of the paper you submit to me at the end of the course. So, being lax with these intermediate deadlines does not directly hurt your grade, but it does limit my ability to provide you with helpful feedback and leaves you behind schedule and scrambling to catch up. Hitting these deadlines also signals your effort on this project. So, I encourage you to follow the schedule as outlined. Of course, you are free to talk with me at any point along the way about your paper.

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 website 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 will not keep track of any other e-mail addresses that you might use. 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

In order for me to evaluate your work fairly, you have to do your own work. It is much easier to study, work hard, and complete your own assignments than it is to try and figure out some way to "beat the system" without getting caught. Cheating, plagiarism, and all other forms of academic dishonestly are pretty 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 cheating in any form in this course 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. Myself and the TA are 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 paper.

DAILY SCHEDULE

Jan. 11	Introduction and Overview: Start Reading "An Introduction to R" User manual
Jan. 13	Simple Regression: Read Gujarati Chap. 1-3
Jan. 18	Estimation of Simple Regression: Read Gujarati Chap. 4-5
Jan. 20	Multiple Regression: Read Gujarati Chapters 7-8
Jan. 25	Multiple Regression, Matrix Algebra: Read Gujarati Appendix B and C
Jan. 27	Multiple Regression **Paper Assignment Due:** 1-2 page statement of your paper topic, noting data source. Include copy of article you are replicating if you pursue that option
Feb. 1	Multicollinearity: Read Gujarati Chapter 10
Feb. 3	Multicollinearity, Factor Analysis, and Measurement
Feb. 8	Heteroscedasticity: Read Gujarati Chapter 11
Feb. 10	Autocorrelation: Read Gujarati Chapter 12
Feb. 15	Autocorrelation and Autogressive Models: Read Gujarati Chapter 17
Feb. 17	Model Specification: Read Gujarati Chapter 13 **Paper Assignment Due:** Outline of paper, including sketch of theory, hypotheses to be tested, and data to be used.
Feb. 22	Outliers, Normality, and Median Regression
Feb. 24	Cross-Validation
March 1	Midterm Exam – Day 1
March 3	Midterm Exam – Day 2
March 7-11	**Spring Break – No Class**
March 15	Dummy Variable Regression: Read Gujarati Chapter 9
March 17	Interaction Terms: Read Brambor et al "Understanding Interaction Models . . ." *Political Analysis* (2006) 14:63-82. **(Berry and Brambor paper)** **Paper Assignment Due:** First cut at analysis. Report hypotheses, statistical model, analysis, and results.
March 22	Interaction Terms
March 24	Pooled Time Series/Panel Data: Read Gujarati Chapter 16
March 29	Pooled and Clustered Data continued.
March 31	**No Class Meeting Due to Midwest Political Science Association Conference**
April 5	Non-linear models and other Transformation: Read Gujarati Chapter 6 pp.175-192; Chapter 7 section 7.10 **Paper Assignment Due:** **Send Via E-mail** First draft of paper – give me all you have.
April 7	Simultaneous Equations: Read Gujarati Chapter 18-20
April 12*	Simultaneous Equations: Read Gujarati Chapter 18-20
April 14	MLE
April 19	MLE **Paper Assignment Due:** Full Draft of Paper, copies for me and for your readers
April 21	Common Mistakes: Read King, Gary. How Not to Lie With Statistics: Avoiding Common Mistakes in Quantitative Political Science, *American Journal of Political Science*, 30(3):666-687.
April 26	Catch-up and Review: **Paper Assignment Due:** Provide written comments to authors and copies of them for me
Tuesday, May 3rd	**Final Exam: 8:00 a.m. Hamilton Hall 351 (allow up to 3 hours)**
TBA	**Final Paper Due, Noon, in my office**