POLI 787 – Generalized Linear Models, Fall 2013
Thursday, 5:00-8:00 p.m., Greenlaw 301

Weekly Lab Wednesday, Noon-12:50 Hamilton Hall 452

Instructor

Dr. Thomas Carsey
E-mail: carsey@unc.edu
Office: 317 Hamilton Hall, 231-B in the Odum Institute
Phone: 962-1207 (but e-mail is a MUCH better way to contact me)
Office Hours: Tuesdays: 10-Noon, Thursdays, 3-5, and by appointment.

Teaching Assistant

Kristin Garrett
E-mail: kdgarret@live.unc.edu
Office: 451 Hamilton Hall
Office Hours: Mondays Noon-2:00, Wednesdays 10-Noon, and by appointment.

Course Description

This course builds on what students would have learned in a simple linear models course focused on Ordinary Least Squares estimation, such as POLI 784. This course generalized that knowledge to look at a class of models that at some level are linear, but are designed for circumstances when the simple linear model is not appropriate. Hence the name Generalized Linear Models, or GLMs. Much of what we will explore might also appear in courses with titles like "Categorical Dependent Variables" because the most common application of GLMs is to estimate models with such dependent variables. However, such courses often present those models as if they were a bunch of different models to be selected as if off a menu our out of a toolbox to deal with a bunch of different specific situations. That view misses the critical point that these models are all part of a single family of models, and in fact the simple linear model is one of the family members. Viewed through the lens of GLM, our goal is to see the common features of these models and recognize the power and flexibility of this approach to be adapted to any sort of liner modeling situation you might encounter. As always, our goal should be to fit our methods to our models and theories, not the reverse.

Students will also learn about a method of estimating the parameters of a statistical model called Maximum Likelihood Estimation or MLE. In fact, many departments teach this course as an MLE course and many researchers refer to the models we will explore and MLE models. That is not really correct. In POLI 784 or any similar course, you were exposed to Ordinary Least Squares or OLS. Again, many people refer to the simple linear model as an OLS model or an OLS regression. I've probably done so many times myself, but that is not quite right. OLS is a method of estimating parameters, and the simple linear model is a model. OLS is the most frequently used method of estimating the parameters of the simple linear model, but one can use OLS to estimate parameters of other models, and one can use something other than OLS to estimate the parameters of the simple linear model. Thus, OLS $\neq$ the simple linear regression model. N the same way, MLE $\neq$ GLM. We will explore all of this in class.

This is a really important class in your development as an empirical social scientist. If you have gotten this far, you have built a solid foundation in basic probability, hypothesis testing, simple statistics, the simple linear model, and OLS. You learned that there are a lot of places you can apply the tools you have, but you also know that the vast array of quantitative research being done in contemporary social science reaches beyond your base of knowledge. If we have been teaching you to crawl and walk before, this class will start teaching you how to run. You will not be an expert in GLMs or MLE at the end of this course, but you will have substantially expanded your knowledge of the tools available to you. There is always more to learn, but you will be well equipped to start doing independent quantitative research if

you successfully complete this class. Thus, I strongly encourage you to push yourself to get as much out of this course as you possibly can.

In order to engage the scholarly community in virtually every subfield of political science, one needs an understanding and working knowledge of statistical methods. Statistical analysis of data is certainly not the only, or even necessarily the best, approach to conducting research. However, every area in social science makes use of statistical methods. 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 class. 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 believe in learning by doing. Thus, we will have regular assignments. Ultimately, social scientists need to be able to formulate theories of social or political processes, translate theories into testable hypothesis, develop models that capture the theory and permit the testing of hypotheses, apply appropriate methods, interpret the results, and return to the theory in order to evaluate it. I think about this process as trying to move seamlessly back and forth between words, pictures/figures, and equations. This is one of the hardest parts of becoming a successful social scientist – just getting the statistics right is only one aspect of the process. The intuition I noted in the previous paragraph is a critical part of this larger process, as it provides a mechanism to facilitate translating our theories of social and political processes into statistical models that can be evaluated 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.

When doing research, a good rule of thumb is to think about trying to satisfy three types of reviewers: 1) the substantive expert, 2) the methods expert, and 3) a friend, college undergraduate, or elderly relative that is neither of #1 or #2. While this course is mostly about methods, learning methods tools in isolation of trying to satisfy the other two types of reviewers would be a mistake.

Course Requirements

Statistical analysis for assignments must be completed using R, and those assignments and your course paper must be written using LaTeX. I have more to say about each below.

There are two required texts for the course:

Gelman, Andrew and Jennifer Hill. 2006. *Data Analysis Using Regression and Multi-level/Hierarchical Models.* New York: Cambridge University Press.

King, Gary. 1998. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference.* Ann Arbor, MI: University of Michigan Press

Both books are in paperback, which should help with the price. We won't cover every single element of each book, but both are very well regarded and are good books to have on your shelf.

We'll also be reading parts of other books and some articles along the way. One is particular is a book written by Jeff Harden and myself called *Monte Carlo Simulation and Resampling Methods for Social Science*. Those readings will get posted on the course Sakai site as we proceed.

These books are all pretty readable as methods books go. 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.

Regular class attendance is expected and required, and 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 on time for every class meeting. We simply have too much work to do to proceed any other way.

Finally, there is a lab/workshop associated with this course that all students are **required** to attend. The lab sessions will give us time to address any issues with assignments, but will also be devoted to a developing a range of professional skills associated with doing scholarly research. Issues from translating theories into statistical models, writing academic articles, seeking grant support and many others will be discussed. These are issues that are critical to your professional success, but are not often dealt with systematically in a graduate program. These sessions will be tied to the main course and your assignments for that course, but my objective is for them to have a broader impact.

Assignments and Grading

We will have 1 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.

35%	Final Exam
20%	Course assignments
10%	Lab (participation, assignments, Poster, etc.)
35%	Final Paper

I reserve the right to make minor adjustments to final course grades based on overall performance in the class. However, as a rule of thumb, if you are scoring in the 90s, you are doing well in the course. If you are scoring in the 80s, you are making satisfactory progress but missing more than you should. If you are scoring in the 70s or below, you are not really doing satisfactory work. 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. 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.

Using R and LaTeX

Assignments will generally require you to perform some analysis using the statistical computing environment 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 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. It is best to use a text editor in conjunction with R. R has one built in, but there are others with many more useful features. I currently use RStudio, which runs on both Windows machines and Macs. 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 reading the documentation for R. I **STRONGLY** recommend that you begin with the manual called "An Introduction to R." This document 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 a few semesters ago and created a website for that, which is located at: http://socserv.mcmaster.ca/jfox/Courses/R-course/index.html.

The Odum Institute (http://www.odum.unc.edu/) is offering a workshop on R on ????. I **STRONGLY** encourage you to sign up if you have not worked with R before. The statistics consultants in Odum also support R, and we will devote the time necessary in class and in the lab to get you all up and running in R. Odum also has an online version of this short course available (http://www.odum.unc.edu/odum/contentSubpage.jsp?nodeid=665).

LaTeX is a document processing environment like R is a statistical computing environment. LaTeX is not really a point and click system, but it is a superior environment for producing publication quality documents, especially if they include tables, figures, and/or equations. In addition, there are some packages in R that will format the output of R functions with all the codes necessary to make the output look nice in LaTeX. All you need to do is copy and paste the output into your LaTeX document. LaTeX is actually used through a text editor, of which there are many. TeXnicCenter is a popular one for Windows users. TexStudio (the one I use now) and TexMaker work on Windows, Mac, and Unix platforms. They are all free an come with only help manuals. Odum is also offering a workshop on LaTeX, on ??? Odum has an online version of this course available as well (http://www.odum.unc.edu/odum/contentSubpage.jsp?nodeid=665).

While we will provide support and direction with R and LaTeX, you need to take the responsibility yourself to learn the tools you need to do your work. 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 exam and in your paper. 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.

The final exam will look like the GLM/MLE portion of the methods field prelim. It 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

The paper is designed to be a piece of original quantitative analysis conducted by you during the course of the semester. You will articulate a theory, derive testable hypotheses from that theory, test them using quantitative methods, present the results, and draw conclusions. You should think of this as writing a journal article like the best ones you see in your substantive classes. There is no formal page length, but for most of you I expect the paper will constitute 18-22 pages of text. Your paper must also include a detailed appendix documenting all of the methodological choices you made, models you estimated, tests you performed, and diagnostics you conducted. That appendix must also include the complete R code that would allow a scholar to load the raw data you started with, make any/all transformations in the data you made, and estimate the results you present. In other words, submit a file that permits a full replication of your analysis. Think of this file as your lab notebook.

If you plan on updating a previous paper, linking this assignment with another class's paper assignment, or using this as a foundation for your MA thesis or a chapter in your dissertation, we need to talk about that. I want you to work on a paper that is valuable to you, but I also think working on lots of papers while in graduate school is a good thing.

All students will read drafts of two other student papers in the course and provide written comments to them. Seeing the work of others 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.

Finally, the last lab session of the year will be devoted to a course poster session. Students will prepare a poster similar to what they would prepare for an academic conference. Faculty and graduate students from the entire department will be invited to attend (and maybe folks from other departments as well). We will work out the logistics of this when it gets closer.

I do not expect perfect papers ready for submission to APSR by the end of the semester. In fact, whether the paper is ever publishable or not is not the goal of this assignment. 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 can focus on how to push beyond that. Don't worry – it will be fun!

I have listed several due dates for aspects of the paper in the course schedule. I expect you to provide me with at least what is asked for 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 (the Poster gets graded as part of the Lab). 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 my expectations, your responsibilities, and the information covered in this course. I will send e-mails to the entire class. I maintain a Sakai 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. The best ways are to come to my office hours or send me an e-mail. I can get pretty busy sometimes, but I do all I can to be responsive to my students. It is important for you to stay in touch, particularly if any problems arise. 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 dishonesty 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 your future career – you should take responsibility for it.

Course Schedule

The schedule below 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 FOR MAIN CLASS

August 22	Introduction, Uncertainty and Inference Read: King chapters 1-2
August 29	APSA – R and LaTeX Review View Odum Online Courses. Read help manuals.
September 5	Probability Model of Uncertainty Read: King chapter 3; Carsey and Harden Chapters 1-3
September 12	Likelihood Theory Read: King chapter 4; Greene chapter 14; Pawitan chapter 2 **Paper Assignment:** 2-page proposal describing paper topic. Include research question, sketch of theory, and sketch of design, data, and analysis plan.
September 19	GLM Theory and the Exponential Family Form Read: McCullah and Nelder chapter 2; Carsey and Harden chapters 4-5
September 26	Binomial Models – Estimation and Interpretation Read: Gelman and Hill chapter 5; Carsey and Harden chapter 6 section 6.1-6.3
October 3	Binomial Models – Estimation and Interpretation Read: Carsey and Harden chapter 9
October 10	Count Models Read: Gelman and Hill chapter 6 section 6.2; Carsey and Harden chapter 6, section 6.4.1 **Paper Assignment:** Draft of theory, design, and methods section of paper, along with a status update/plan for data analysis.
October 17	**Fall Break – No Class**
October 24	Ordered and Unordered Models Read: Gelman and Hill chapter 6, section 6.5; Carsey and Harden chapter 6 section 6.3.2-6.3.3
October 31	Survival Models Read: Box-Steffensmeier and Jones chapters 1-4; Carsey and Harden chapter 6 section 6.4.2
November 7	Multilevel Models Read: Gelman and Hill chapters 11-12
November 14	Missing Data Read: Gelman and Hill chapter 25
November 21	Matching Read: Morgan and Winship chapters 1-2, 4; Iacus, King, and Porro (2011) **Paper Assignment:** Complete draft due to Instructor and Class Readers
November 28	**Thanksgiving – No Class**
Tuesday, December 10	Final Exam, 7:00 p.m.
????	**Final Paper Due**