Maximum Likelihood Methods
Political Science, 733
Spring Semester 2013
Location: LINK 2-070 Seminar 4
Time: Wednesdays 6:15-8:45 pm, January 16th, 2013 - April 17th, 2013
Labs: Friday 10:20-11:455 Perkins LINK 2-087 classroom 3
Final Homework Due electronically April 16th, 17:00 pm
Final Writeup is due on May 2nd, before 17:00 pm

Instructor: Michael D. Ward, email: michael.d.ward@duke.edu
Office Hours: Wednesdays, 9-12:00 and 13:00-17:00 by appointment; Perkins 408

Teaching Assistant: Simon Weschle, email: simon.weschle@duke.edu
Office Hours: Monday 13:00-15:00, and by appointment, Perkins 317

1 Goals of Course

This course introduces the theory and practice of likelihood inference for statistical models, as applied to the social sciences. It generally focuses on nonlinear models that are often called "generalized linear models," In practice this means that the putatively standard linear model is adapted to accommodate a wider range of outcome variables and a broader range of functional forms. Thus, we will examine distributions for data that expand considerably upon the Gaussian distribution. We'll examine binary, ordinal, nominal, and counted outcomes, as well as outcomes that are bounded in some way and/or interdependent.

Many social science problems do not easily conform to a statistical frame defined by *ordinary least squares*. The likelihood framework provides a powerful framework that expands the set of problems that can be statistically examined in a cogent fashion. The skills obtained in this course are focused on two areas: 1) formulation of likelihoods that are nuanced and specific to applied problems found in social science data and 2) implementation, estimation, and presentation of results obtained by empirically oriented social scientists using generalized linear models.

The knowledge obtained in this class should permit you to analyze and present a wide range of non-OLS statistical models. By the end of the class you should also be in a position to formulate your own likelihoods that may be appropriately contextualized for your specific research questions. Finally, the computational skills and theoretical foundations of the likelihood are intended to be useful in more advanced statistical courses.

2 How to Succeed in this Course

Tenacity is probably the most important determinant of success in this class. The Teaching Assistants for the class also stand ready to help you, and you should attend a weekly lab section

(TBA). I also suggest that you work in teams on your homework assignments. We will randomize the working groups in labs, so that everyone will have the opportunity to work with everyone else. Each lab (except the first) will have a project that each team will need to turn in. These will not be graded per se, but will influence my evaluations.

Reading needs to be done. I know everyone tells you to read before coming to class, but I want you to read after coming to class. Thus when you see reading assignments under Week 1, for example, I expect you to read them after my lecture that week, but before lab. We'll then spend the first section of the labs answering your questions about what was covered in lecture and in the readings. I'll devote the first section of each Week's lecture answering any additional questions from the previous week.

3   Required materials

3.1   Books

The main text for the class is one I have written. I will supply in in PDF format on the class website at the beginning of class. It will probably be revised throughout the semester, but it exists in is current form as:

- WARD, MICHAEL D., AND AHLQUIST, JOHN S. *Maximum Likelihood Strategies for Social Science*, version 0.8. pdf typescript, Duke University, 30 December 2012, via Sakai.

The other text that I suggest you acquire is:

- KING, GARY *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*, University of Michigan Press, 1998, 288 pages, ISBN 0472085549.

You may also find helpful sections (chapter 15 and 16) in the Fox book *Applied Regression Analysis, Linear Models, and Related Methods* (1997) as well as a few chapters (5, 7, 8, 18, 21 & 25) in the Gelman and Hill volume, *Data Analysis Using Regression and Multilevel/Hierarchical Models* (2007). These various chapters are listed as optionL readings.

3.2 $\mathcal{R}$ Software

The course will be organized around the $\mathcal{R}$ computer language, a specialized programming language. We will show you how to use this program, which is used in most advanced statistics courses at Duke University, and is widely used by political methodologists.[1] $\mathcal{R}$ is loaded on all of the political science lab computers, and is freely available for your own use. It runs on most operating systems, including Windows, lin/u/nix, and the Apple OS/X (and earlier) operating systems. You may obtain a copy of $\mathcal{R}$ at http://cran.r-project.org/. My own advice is that you download and install from the binaries directory. You should know how to use $\mathcal{R}$ before beginning this class. You should also have successfully completed an introductory regression course (e.g., Political Science 318).

---

[1]You may, of course, use any computer program you like to undertake the work in the course. However, we can only help you with $\mathcal{R}$ and will spend considerable resources during the course illustrate how to use it to do your weekly homework.

3.3 Editing

There are many ways to work with statistical files, but it is important to develop a habit of actually creating scripts, which you can document and archive, instead of working interactively until you get a) tired, or b) find an answer you are happy with given that you are a). Cutting and pasting out of Windows Notepad or any ascii editor works fine, but I recommend that you use a syntax checking editor such as *WinEdt* (www.winedt.com) or *emacs*. WinEdt is a super editor for both LaTeX and $\mathcal{R}$, since it knows about the syntax of each program. I also like it because it also has a column editing mode, so it is great for certain data related tasks, too. You can use it for free for about 100 times, then the popups get annoying, but it has a student license that is only $30. If you are a interested in a freeware option, you can get copies of the EMACS editor (www.xemacs.org), which links to $\mathcal{R}$ directly via the Emacs-Speaks-Statistics (ESS) package

3.3.1 Technical Word Processing

Most technical, scientific communications are processed in a dialect of the TeXlanguage invented by Donald Knuth and freely available via the TeX Users Group at http://www.tug.org/ or via the Comprehensive TeX Archive Network (CTAN), primary repository for TeX-related software on the Internet: `http://www.tug.org/ctan.html`. Most political scientists have settled on LaTeX. There is a good, free implementation of both for the Windows world via MikTex: `http://www.miktex.org/`. My advice is that if you plan to use technical material in your scientific communications, that you learn to use LaTeX(or TeX or AMS-TeX). There are N, where N is large, tutorials on LaTeX on the WWW, and you can find one to match your interests. The canonical list is at `http://www.tug.org.in/tutorials.html`, but in reality most find it useful to start with a file that works that someone else created, replace their stuff with your stuff, tweak it a bit, and declare victory. MikTex has several appropriate templates as well as a feature to create your own. Feel free to use whatever text processor you wish; if you are using Microsoft Office products for this task, you will need to learn about (and possibly purchase) MathType (`http://www.desssci.com/en/products/mathtype/`).

4 Major Tasks

1. Quasi-weekly Homework, due before lecture the week following its assignment

2. Replication and extension of an article of interest that focuses on non-OLS techniques

    a. Comment and feedback for colleague on their project

    b. Establish archive for sharing own work with others

    c. Presentation of a Poster on April 17, 17:00-19:00 detailing your project and your findings

    d. 10-15 page writeup of the "Methods" and "Results" sections of a paper on your project.

3. Midterm exam on the Maximum Likelihood principle, March 6.

5 Course Outline, in 14 moving parts

Week 1, January 16th, 2013

REVIEW OF OLS AND FIRST PRINCIPLES OF MAXIMUM LIKELIHOOD

Readings

- **required** Ward, *Everything You Wanted to Know about Regression but Were Afraid to Ask*, Sakai pdf

- **required** King, *Unifying . . .*, chapters 1-4

- **required** Michael D. Ward and John S. Ahlquist. Maximum Likelihood Strategies for Social Science, chapter 1

Week 2, January 23rd, 2013

SIGNIFICANCE TESTING
Extra assignment: You must turn in a one page summary of the work you have chosen to replicate, along with output presenting the summary statistics for the data that are to be used in this project. In addition you need to set up an archive that holds these data, ideally using `thedata.org`, details in lab. Finally,

Readings

- **required** King, *Unifying . . .*, chapters 1-4

- **required** Donald N. McCloskey. The loss function has been mislaid: The rhetoric of significance tests. *American Economic Review, Papers and Proceedings of the 97th Annual Meeting of the American Economic Association*, 75(2):201–205, May 1985.

- **required** Edward L. Leamer, *Let's Take the Con out of Econometrics*, American Economic Review, 73(1):3143, March 1983.

- **required** JEFF GILL The insignificance of null hypothesis significance testing. *Political Research Quarterly 52*, 3 (September 1999), 647–674.

Week 3, January 30th, 2013

CROSS VALIDATION, VISUALIZATION, PRESENTATION OF RESULTS

Readings

- **required** Michael D. Ward and John S. Ahlquist. Maximum Likelihood Strategies for Social Science, chapter 3 & 10

- **required** King, *Unifying . . .*, chapters 1-4

- **optional** Gelman & Hill, *Data Analysis Using ..., Chapter 7.*,

---

Week 4, February 6th, 2013

BINARY VARIABLES

Readings

- **required** Michael D. Ward and John S. Ahlquist. *Maximum Likelihood Strategies for Social Science*, chapter 2

- **required** King, *Unifying . . .*, chapter 5 & 6

- **optional** Fox, *Applied Regression . . .*, chapter 15

- **optional** Gelman & Hill, *Data Analysis Using ...*, chapter 5, sections 6.3 & 6.3

Week 5, February 13th, 2013

ORDINAL VARIABLES

Readings

- **required** Michael D. Ward and John S. Ahlquist. *Maximum Likelihood Strategies for Social Science*, chapter 4

- **required** King, *Unifying . . .*, chapter 5

Week 6, February 20th, 2013

NOMINAL VARIABLES

Readings

- **required** Michael D. Ward and John S. Ahlquist. *Maximum Likelihood Strategies for Social Science*, chapter 5

- **required** King, *Unifying . . .*, chapter 5 & 6

Week 7, February 27th, 2013

COUNT VARIABLES Extra assignment: you must present a draft of your final project to your peer reviewer (and the instructor) on this date. You must also provide access to the data and any code you may have developed. Your review is due back to the instructor and your colleague by the end of the week, before break.

Readings

- **required** Michael D. Ward and John S. Ahlquist. *Maximum Likelihood Strategies for Social Science*, chapter 6

- **required** King, *Unifying . . .*, chapter 6

- **optional** Gelman & Hill, *Data Analysis Using ...*, chapter

Week 8, March 6th, 2013

MIDTERM EXAMINATION, IN CLASS

Spring Break, March 13th, 2013: NO CLASS

Week 9, March 20th, 2013

SPATIAL DEPENDENCIES

Readings

- **required** Michael D. Ward and Kristian S. Gleditsch. *Spatial Regression Models*, Sage QASS Number 155, 2008, ISBN 978-1-4124-5415-0

Readings

- **required** Michael D. Ward and John S. Ahlquist. *Maximum Likelihood Strategies for Social Science*, chapters 7 & 11

- **required** King, *Unifying . . .*, chapter 8

- **optional** Gelman & Hill, *Data Analysis Using ...*, chapter

Week 10, March 27th, 2013

MORE ON SERIAL CORRELATION

Readings

- **required** Michael D. Ward and John S. Ahlquist. *Maximum Likelihood Strategies for Social Science*, chapters 8 & 9

Week 11 April 3, 2012 ISA MEETINGS IN SAN FRANCISCO; NO CLASS

Week 12, April 10th, 2013

EVENT HISTORY

Readings

- **required** Michael D. Ward and John S. Ahlquist. Maximum Likelihood Strategies for Social Science, chapter 9

Week 13, April 17th, 2013

POSTER SESSION
We will have a poster session for class projects, location to be determined. We'll also need an electronic copy of your poster. The project must be archived on thedata.org. I will invite all faculty and grad students to our poster session, and I will supply refreshments. There will be prizes for the best posters, selected by the Methods Faculty.

Grading

Grading will focus on

- Homework (34 %): There will be nine weekly homework assignments consisting of analytical problems, computer programming, data analysis, and the writeup of results for a social science audience. The homework needs to be word processed (see 3.3.1 for suggestions) with any displays integrated into the document. Each homework must also be accompanied by a file containing the code used to complete the analysis, if any. Grading will be assigned based on three categories (unacceptable, acceptable, exceptional). Each is worth four points, you may drop or not turn in one of the nine;

- Midterm Examination (20 %): this is an in class Midterm in which you will be asked to use a pencil and paper to analyze statistical output. You may bring any written, printed material you wish to help you;

- Peer Feedback (16 %); You will be asked to provide a mid-project review of a colleague's research. This will be in the form of a journal review, and will be shared with the colleague and the instructors. Its goal is to provide helpful guidance at a midpoint in the research project;

- Final Writeup & Poster (30 %): Your semester project is a replication and extension of some published work. This work must be chosen by the second week of the semester, and must be approved by the instructor at that time. More details in class.

I will not give grades of "incomplete" in this class, nor accept late assignments.

Reasonable Accommodation

The Duke University is committed to providing access, equal opportunity and reasonable accommodation in its services, programs, activities, education and employment for individuals with disabilities. For information or to request disability accommodation contact: Student Disability Access Office (SDAO): `http://www.access.duke.edu/students/index.php`, or via 402 Oregon Street, Suites 102 and 103, Box 90142, Durham, NC, TTY: 919-668-1329, FAX: 919-668-3977.