POLSCI 733: MAXIMUM LIKELIHOOD ESTIMATION
Spring Semester 2016

Lecture times: **Tuesdays/Thursdays, 3:05-4:20**
Lecture room: **Bunche Lab**
Instructor: **Chris Johnston**
Contact: **cdj19@duke.edu, (919)660-4345**
Office: **294J Gross Hall**
Office hours: **By appointment, or just stop by!**
Lab times: **Fridays, 10:05-11:20**
Lab room: **Bunche Lab**
TAs: **Anh Le (anh.le@duke.edu), Jan Vogler (jan.vogler@duke.edu)**

Books with required readings

- Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.
- Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage.
- Eliason, Scott R. 1993. *Maximum Likelihood Estimation: Logic and Practice*. Sage.

Grades

- Eight homework assignments: 5% each
- Three "short" paper assignments: 10% each
- One final research paper: 30%

I will use the following scale to assign final grades:

A	93-100	A-	90-93		
B+	87-90	B	83-87	B-	80-83
C+	77-80	C	73-77	C-	70-73
D+	67-70	D	60-67	F	<60

Policy

Late paper assignments will be reduced by one full letter grade for each 24-hour period beyond the deadline. The policy for late homework assignments is specified below. I will follow the University's policy in any event of plagiarism and academic dishonesty.

Software

I will use and teach R in this class, and you will need to do the same. If this presents a problem for you, come talk to me immediately and we can discuss. I will often post example R-code on Sakai (with the relevant data) which you may find helpful for your assignments or in your work outside of class. Use these files as you wish.

Homework

Each homework assignment will deal with the concepts and methods discussed in the previous week's lectures. Homework assignments will be posted after Friday's lab and should be turned in via the course website by the beginning of lab the following week. For each homework, students are required to cross-grade another student's homework. The deadline for the submission of the cross-graded homework is the beginning of the Tuesday lecture. For the cross-grading procedure, we will use the following website (some of you will be familiar with this system from *POLSCI 630: Introduction to Empirical Approaches*): http://ps630-f15.herokuapp.com/

Students are required to submit their homework as a PDF document created by knitr as well as the source code for knitr. The submitted PDF file is expected to contain (1) written answers, (2) R code, and (3) R output. In order to ensure that the cross-grading procedure takes place anonymously, students are asked not to make any reference to their identity in the submitted documents. A maximum of 16 points can be earned for each homework. For the calculation of the final course grade, the average of all homework grades will be rescaled to a 100-point scale. For your TAs' sake, please try to keep your code and output concise and organized!

As cross-grader, each student has an important responsibility in this class. Students are expected to take this responsibility seriously and carefully evaluate the homework they have received from another student. If the TAs recognize that a student does not grade another person's homework carefully and in accordance with the answer key, they may reduce the grader's homework score.

Students can discuss their answer strategies but are expected to write their answers and code individually and not to copy from other students or external sources (such as websites or online lectures). If it is recognized that answers are copied from students or external sources, the grade of the involved students will be reduced, and further action may be taken in line with the University's policies on academic dishonesty. These assignments are not busywork; they are designed to help you better understand the course material. If you are having a lot of trouble, come see me!

If the homework is late, 3 points will be subtracted from the grade of the homework for each 24-hour period it is late. If the submission of the cross-grading is late, 1.5 points will be subtracted from the grade of the grader's homework for each 24-hour period it is late.

"Short" Paper Assignments

Treat each paper assignment as practice for writing the methods and results sections of a research article to be submitted to a peer-reviewed journal.

Paper 1: Binary Dependent Variables

Find a dataset with a binary dependent variable of interest to you. First, propose a hypothesis regarding the relationship of this variable to at least one other variable. Test that hypothesis with appropriate controls. You should report all regression estimates with representations of uncertainty, as well as the ePCP, the proportionate reduction in error, and the area under the ROC curve for the model. Provide an appropriate representation of your primary hypothesis test, and interpret your results in words. Second, propose an interactive hypothesis. You can use the same or a different set of variables, but you must include at least one (theoretically reasonable) interaction in your model. Plot your results in an appropriate fashion (with some representation of uncertainty) and interpret them. For each of these two tasks, you should briefly introduce your hypothesis, data, operationalizations of variables, and model specifications before moving ahead to results. You can do this in one section or two (it is up to you). Be sure to interpret your results sufficiently and clearly!

Paper 2: Ordinal, Nominal, and Count DVs

First, find a dataset with an ordered dependent variable and several relevant independent variables and estimate an ordered probit or logit model; report all regression estimates, and provide an appropriate graphical representation of the relationship of at least three of your model variables with the dependent variable, including representations of uncertainty. Then relax the parallel lines/proportional odds assumption for one focal independent variable, re-estimate the model, and report and interpret the new results for this variable. Second, find a dataset with a nominal dependent variable and several relevant independent variables, and estimate a multinomial logit model; provide an appropriate graphical representation of the relationship of at least three of your model variables with the dependent variable, including representations of uncertainty. Third, find a dataset with a count dependent variable and several relevant independent variables. Estimate a poisson model, report all regression estimates, and graphically report and interpret your results for at least three model variables including representations of uncertainty. Then, re-estimate using a quasi-poisson or a negative binomial model and compare to the results from the poisson model. You may divide this paper into three separate "mini-papers," or you may write combined methods and results sections. It is up to you.

Paper 3: Multilevel Modeling

Find a dataset with a multilevel structure. First, propose and test two hypotheses using a "random intercept" model. The first hypothesis should entail a relationship between the DV and a "first-level" variable, and the second should entail a relationship with a "second-level" variable. Second, propose and test a cross-level interaction using a "random slope" model. The dependent variable can be anything you wish. Describe and justify all data and modeling choices, and provide sufficient and appropriate representation and discussion of all results.

Final Paper

You have two options for your final paper. Begin looking into this early on in the semester. You must clear your project with me by March 1st at the latest. **The final paper is due on May 6th by 12pm!**

Option 1:

Go to one of the major journal websites and find a paper on a topic that interests you. Using the provided replication data, or other relevant publically available data, you should improve upon or extend the analysis reported by the authors of this paper using the methods learned in this course. In other words, your job is to find an existing analysis and make it better. That might mean using a more appropriate methodology than the authors of the paper to check the robustness of the results, it might mean using a more appropriate model specification, or it might mean extending the results of the paper by proposing and testing a new hypothesis that the authors did not check. There are a large number of options here. The only requirements are that you use one or more of the methods learned in this class (no OLS), and that your contribution is meaningful enough to warrant writing the paper! With respect to the latter use the following rule of thumb: would my paper be desk-rejected at a peer-reviewed journal? If the answer to this is "probably yes," then find a different topic.

You will write-up your new results in a research paper. This means: (1) introduce the topic and the previous research on this topic (obviously including the paper that inspired your own paper), (2) explain the purpose of your own paper, (3) describe the data and variable operationalizations you will use, (4) describe your modeling approach, (5) present and interpret your findings, and (6) conclude. Again, the paper should (with only relatively minor edits) be appropriate as a submission to an actual journal. We will grade your paper as-if we were reviewing your article for a journal.

Option 2:

Write a full research article on a topic of your choosing that could (with at most *minor* edits) be submitted to a peer-reviewed journal. This means: (1) introduce the topic and the previous research on this topic, (2) explain the purpose of your own paper, (3) describe the data and variable operationalizations you will use, (4) describe your modeling approach, (5) present and interpret your findings, and (6) conclude. The only requirement is that you use one or more of the methods learned in this class (no OLS). We will grade your paper as-if we were reviewing an article for a journal.

Whichever option you choose, use this paper as an opportunity to advance your research agenda. Do not waste your time!

Course Schedule

- **Required**
  - **Recommended**
    - **Lab-related**

Week 5 (February 8th-12th): Binary Dependent Variables II

- G&H: Sections 6.6, 7.3, & 7.4
- Hanmer and Kalkan. 2013. "Behind the Curve."
- Berry et al. 2010. "Testing for Interaction in Binary Logit and Probit."
- ➢ Lab, and homework #3 is due to your TAs, on Friday

Week 6 (February 15th-19th): Binary Dependent Variables III

- Review readings and lectures; practice!!!
  - Gelman et al. 2008. "A Weakly Informative Default Prior Distribution for Logistic…"
  - King and Zeng. 2001. "Logistic Regression in Rare Events Data."
- ➢ Lab, and take-home assessment is due to your TAs, on Friday

Week 7 (February 22nd-26th): Ordered Dependent Variables

- Long: Chapter 5
- G&H: Section 6.5
  - Gabel. 1998. "Economic Integration and Mass Politics."
  - King and Wand. 2007. "Comparing Incomparable Survey Responses."
- ➢ Lab, and homework #4 is due to your TAs, on Friday

Week 8 (February 29th-4th): Nominal Dependent Variables

- Long: Chapter 6
- G&H: Section 6.5
  - Dow and Endersby. 2004. "Multinomial Probit and Multinomial Logit."
  - Cheng and Long. 2007. "Testing for IIA in the Multinomial Logit Model."
- FIRST PAPER DUE AT BEGINNING OF LAB ON FRIDAY!
- ➢ Lab on Friday, but no homework this week

Week 9 (March 7th-11th): Count Dependent Variables

- Long: Chapter 8
- G&H: Section 6.2 & 6.3
- Zeileis. 2007. "Regression Models for Count Data in R."
  - King. 1988. "Statistical Models for Political Science Event Counts."
  - King. 1989. "Variance Specification in Event Count Models."
- ➢ Lab, and homework #5 is due to your TAs, on Friday

Week 10 (March 14th-18th): NO CLASS FOR SPRING BREAK

Week 11 (March 21st-24th): Limited Dependent Variables

- Long: Chapter 7
- Sartori. 2003. "An Estimator for Some Binary-Outcome Selection Models."
- ➢ Lab, and homework #6 is due to your TAs, on Friday

Week 12 (March 28th-April 1st): Duration Models

- Box-Steffensmeier and Jones. 1997. "Time Is of the Essence."
- Beck et al. 1998. "Taking Time Seriously."
- Carter and Signorino. 2010. "Back to the Future."
- SECOND PAPER DUE AT BEGINNING OF LAB!
- ➢ Lab on Friday, but no homework this week

Week 13 (April 4th-8th): Basics of Multilevel Modeling

- G&H: Chapters 1, 11 & 12
- o G&H: Chapters 21 and 22
- o Steenbergen and Jones. 2002. "Modeling Multilevel Data Structures."
- o Jackman. 2004. "Bayesian Analysis for Political Research."
- o Kruschke. 2015. *Doing Bayesian Data Analysis*, Chapter 19. (on Sakai)
- ➢ Lab, and homework #7 is due to your TAs, on Friday

Week 14 (April 11th-15th): Modeling Regression Coefficients and Other Complexities

- G&H: Chapters 13
- o Aguinis et al. 2013. "Best-Practice Recommendations for Estimating Cross-Level…"
- o Bafumi and Gelman. 2006. "Fitting Multilevel Models When Predictors and Group…"
- o Stegmueller. 2013. "How Many Countries for Multilevel Modeling?"
- ➢ Lab, and homework #8 is due to your TAs, on Friday

Week 15 (April 19th): Multilevel GLMs, Applications, and Extensions

- G&H: Chapters 14 & 15
- o Lax and Phillips. 2009. "How Should We Estimate Public Opinion in the States?"
- o Bafumi et al. 2005. "Practical Issues in Implementing and Understanding Bayesian Ideal Point Estimation."
- o Park et al. 2004. "Bayesian Multilevel Estimation with Poststratification."
- o Lebo and Weber. 2015. "An Effective Approach to the Repeated Cross-Section Design."
- o Martin and Quinn. 2002. "Dynamic Ideal Point Estimation via MCMC…"
- o Barberá. 2015. "Birds of a Feather Tweet Together."
- o Grimmer. 2010. "A Bayesian Hierarchical Topic Model…"
- ➢ Lab on Friday, but no homework this week
- ➢ THIRD PAPER DUE TO YOUR TAs BY FRIDAY AT 12PM!