POLSCI-935: Seminar in Advanced Political Science Methodology

7-9:40PM, Thursdays (9/3/09-12/17/09), Bolton 293

Instructor: Dave Armstrong
Office: 624 Bolton Hall
Hours: 3-5PM Mondays or by appointment
Phone: (414) 229-4239
E-mail: armstrod@uwm.edu
Web: www.quantoid.net

This course is designed to make you better consumers and produces of quantitative social science research. It will extend the linear model and address a new class of models - generalized linear models. The bulk of the course will be spent discussing models for categorical dependent variables, while the remainder of the course will consider extensions to the linear model (e.g., hierarchical models and models for time-series cross-sectional data). The course will be very applied meaning that we will spend a considerable amount of time on the computer implementing the techniques discussed in class. With any luck, at the end of the course students will be able to critique, run, interpret and present models using categorical dependent variables.

I expect that everyone in the class has completed the 701-702 sequence, or its equivalent. We will spend some time reviewing the linear model - specifically, we'll spend a lecture on OLS and its assumptions along with some discussion of interpretation and presentation that should be review, largely, but my experience is that these are always useful things to learn multiple times. Further, as you are graduate students, I expect you to take responsibility for learning the material just as I will take responsibility for presenting the material in a clear and accessible way. Your responsibilities, then, include reading and really trying to engage the material outside of class.

We will be using STATA as the primary statistical package for this course. STATA is a relatively easy-to-use program and is available for use outside the classroom in Bolton 617 (3 computers), the computer lab in Bolton 225, as well as other labs on campus. In addition, students can obtain discounted versions of STATA via the grad plan offered by STATA Corporation (http://www.stata.com/order/new/edu/gradplan.html). The current version of Stata on offer is 11 and is available in a one-year license for $98 and in a perpetual license for $179. If there is any interest, I can also point you to resources engaging this material in **R**, an open-source (i.e., free) and *very* powerful statistical programming language.

Textbooks

Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage Publications.

Long, J. Scott and Jeremy Freese. 2006. *Regression Models for Categorical Variables Using Stata*, 2nd ed. College Station, TX: Stata Press

Both of these books are currently available in the UWM Bookstore. I will also provide some additional reading materials that will provide you with examples of these techniques being used well in print. I will provide links to these resources when appropriate. All supplementary readings will be available in electronic format on the course's D2L site under the "content" link.

You final grade in the course will depend on the following:

Homework	50%
Research Paper	50%

Work Schedule

Homeworks	1	handout after covering topic 1
	2	handout after covering topic 4
	3	handout after covering topic 6
	4	handout after covering topic 8
Research Paper	Topic	9/17
	Research Question	9/24
	Brief lit review/hypotheses	10/7 ($\leq$ 2 pages)
	Codebook and Descriptives	11/4
	Final paper	12/16 (Noon, CST)

Homeworks

The homework exercises will be almost entirely applied; that is, they will ask you to use the computer to generate and interpret results with actual data. While I encourage you to think of you colleagues as a resource and to talk about common problems that you might have and about *how* to do things. That said, I want each of you to *do* your own work. To be perfectly clear, you should each hand in an original effort at answering the homework problems. This includes writing your own code to generate the results and using your own words to interpret the results. You should not ask your colleagues for, nor should your colleagues provide if asked, a "template" or skeleton of either the code or the interpretation. I truly believe that the only way that you will learn how to do these things is by doing them yourselves.

Research Paper

The research paper I want you to write is one that has a *very* brief literature review proceeding from a research question and leading to a set of hypotheses. This should be original in that it should not be a simple replication of someone else's work. The

literature review should be simply enough to provide a bit of theoretical backing for the hypotheses of interest. This should be no more than two (2) pages. You will have to gather some data that will help you test the hypotheses of interest. In the remainder of the paper, you will describe the data you have collected, run a model (or models) that allow you to test the hypotheses of interest, evaluate the models and their appropriateness to the task of testing the hypotheses of interest and interpret the findings both statistically and substantively. Null findings are alright here - I want to see whether you understand the logic of the process and are able to produce and interpret a set of results that bear some resemblance to the hypotheses they are meant to test. As you can see from the schedule of work, the final preliminary piece of the project is due on 11/4 with the final draft due on the date of the final (12/16). I will read and comment on any preliminary draft I receive before 11/25 so it is to your benefit to be getting things done early.

You may obtain information on UWM policies concerning academic issues and course conduct here: `http://www.uwm.edu/Dept/SecU/SyllabusLinks.pdf`. While predictions of widespread outbreaks of H1N1 are (hopefully) of the worst-case-scenario variety, there is some possibility that an outbreak of H1N1 could disrupt University operations. In the event of disruption of normal classroom activities due to an H1N1 swine flu outbreak, the format for this course may be modified to enable completion of the course. In that event, you will be provided an addendum to this syllabus that will supersede this version.

Outline

1. Linear Model Refresher

   - Properties of OLS estimators
   - Presentation and Interpretation
   - Interactions
   - Non-linearity

   Readings: Brambor, Clark and Golder 2006; Long and Freese 2006, Chapter 9: § 9.1, 9.2, 9.3

2. Maximum Likelihood Estimation

   - What is MLE?
   - Why do we need it?
   - How does it work?

   Readings: Long 1997, Chapter 2; Long and Freese 2006, Chapter 3

3. Models for Binary Dependent Variables: Probit and Logit

   - The Likelihood Function
   - Model Estimation
   - Interpretation of coefficients
   - Marginal Effects
   - Substantive Effects
   - Graphical Presentation

   Readings: Long 1997, Chapter 3; Long and Freese 2006, Chapter 3; Hanmer and Kalkan 2009; Fox 1987

4. Model Fit and Evaluation

   - How well does the model fit these data?
   - How well does the model fit other data?
   - Nested model comparison
   - Non-nested model comparison

   Readings: Long 1997, Chapter 4; Herron 1999; Burnham and Anderson 2004

5. Models for Ordinal Dependent Variables

   - The Likelihood Function
   - Model Estimation
   - Interpretation of Coefficients
   - Marginal Effects
   - Substantive Effects
   - Graphical Presentation

   Readings: Long 1997, Chapter 4; Long and Freese 2006, Chapter 5; Fox and Andersen 2006;

6. Models for Unordered Dependent Variables

   - The Likelihood Function
   - Model Estimation
   - Interpretation of Coefficients
   - Marginal Effects
   - Substantive Effects
   - Graphical Presentation
   - Testing Ordinality

   Readings: Long 1997, Chapter 6; Long and Freese 2006, Chapters 6-7; Dow and Endersby 2004

7. Models for Counts: Poisson and Negative Binomial

   - The Likelihood Function
   - Model Estimation
   - Interpretation of Coefficients
   - Marginal Effects
   - Substantive Effects
   - Graphical Presentation

   Readings: Long 1997, Chapter 8; Long and Freese 2006, Chapter 8

8. Models for Duration Data

   - The Likelihood Function
   - Model Estimation
   - Interpretation of Coefficients
   - Marginal Effects
   - Substantive Effects
   - Graphical Presentation
   - Exponential vs. Weibull
   - Time-varying covariates

   Readings: Box-Steffensmeier and Jones 2004, pages 1-68

9. Missing Data and Multiple Imputation

   - Why do we have missing data?
   - Types of missing data.
   - Consequences of missing data?
   - The wrong ways of dealing with missing data: listwise deletion, mean imputation, etc...
   - Using Multiple Imputation to handle missing data.

   Readings: Mcknight et al. 2007, pages 1-64 and 196-212

10. Models for TSCS Data

    - (Re)Fresher on time-series issues
    - Unit- and time-effects in the linear model
    - Models for TSCS binary, ordinal and multinomial data.

    Readings: Wooldridge 2002, Chapter 10; Beck and Katz 2009; Beck et al. 2002; Box-Steffensmeier and Jones 2004, Chapter 5

References

Beck, Neathaniel and Jonathan N. Katz. 2009. "Modeling Dynamics in Time-Series Cross-Section Political Economy Data." California Institute of Technology Social Science Working Paper # 1304.

Beck, Nethaniel, David Epstein, Simon Jackman and Sharyn O'Halloran. 2002. *Alternative Models of Dynamics in Binary Time-Series Cross-Section Models: The Examples of State Failure*. Prepared for the 2001 Annual Meeting for the Society for Political Methodology - Atlanta, GA.

Box-Steffensmeier, Janet M. and Bradford S. Jones. 2004. *Event history modeling : a guide for social scientists*. Cambridge; New York: Cambridge University Press.

Brambor, Thomas, William Clark and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14(1):63–82.

Burnham, Kenneth P. and David R. Anderson. 2004. "Multimodel Inference: Understanding AIC and BIC in Model Selection." *Sociological Methods and Research* 33(2):261–304.

Dow, Jay K. and James W. Endersby. 2004. "Multinomial Probit and Multinomial Logit: A Comparison of Choice Models for Voting Research." *Electoral Studies* 23:107–122.

Fox, John. 1987. "Effect Displays for Generalized Linear Models." *Sociological Methodology* 17:347–361.

Fox, John and Robert Andersen. 2006. "Effect Displays for Multinomial and Proportional-Odds Logit Models." *Sociological Methodology* 36:225–255.

Hanmer, Michael J. and K. Ozan Kalkan. 2009. "Behind the Curve: Clarifying the Best Approach to Calculating Predicted Probabilities and Marginal Effects from Limited Dependent Variable Models." Unpublished Manuscript.

Herron, Michael C. 1999. "Postestimation Uncertainty in Limited Dependent Variable Models." *Political Analysis* 8(1):83–98.

Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks: Sage Publications.

Long, J. Scott and Jeremy Freese. 2006. *Regression Models for Categorical Dependent Variables Using Stata,* 2nd *ed.* College Station: Stata Press.

Mcknight, Patrick E., Katherine M. McKnight, Souraya Sidani and Aurelio Jose Figueredo. 2007. *Missing data a gentle introduction*. New York: Guilford Press.

Wooldridge, Jeffrey. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.