Objectives of the Course:
POL531 is the second course in our graduate sequence in methods and applied statistics. In POL530 we have dealt primarily with models in which the dependent variable is continuous and did not have much time to concentrate on categorical dependent variables (CDVs). Such models are based on maximum likelihood estimators and include probit, logit, ordered logit, and Poisson regression, among others. In this course we aim at providing you with a longer and more detailed treatment of the CDV models.

This course is designed in the format of a small lecture and computer training class; we hope that it will turn into almost a graduate tutorial course. Besides the lectures and computer lab sessions that we will lead, we also expect the graduate students to pick a paper topic of their interest on which they can work with the help of available data. They will be required to regularly make short presentations as to their data sources, theoretical background and variable selection as well as appropriate analyses. At the end of this semester the objective is to have not only mastered the basic methods of categorical data analysis but also to have a written paper with the methods learned that can eventually be submitted to a peer reviewed journal.

Background Assumed for the Course:
As we have done previously in POL530 this course is also designed differently from most introductory graduate courses in econometrics in that it dispenses with most theorems and proofs. However, students still need some basic mathematical tools to read, understand and equally importantly, to appreciate most methodological treatments in class discussions as well as the journal articles they will come across in their research. Accordingly, students are expected to polish their mathematical skills. However, we will also provide some outlines for mathematical prerequisites for this class.
Books:
The basic texts for this course are:

Required:

Recommended:

*Lecture and Lab Notes for POLS 531.* Copies of our lecture notes and lab materials will be made available electronically to all students.

Computer Applications:
Familiarity with SPSS is helpful. However, we aim in this class to introduce students STATA through applications of the CDV models and hope that they will switch to it for more sophisticated analyses.

Organisation of Classes:
Again similar to POLS 530 there will be two components of our classes. First, lectures, hand-outs and exercises will be geared toward building the necessary toolchest for a graduate level understanding and use of CDV analyses. Second, there will be regular one hour per week meetings to cover issues in computer and software use, applications of regression analysis and discussions of different exercises.
Term Project:
Students are also required to write an empirical research paper where they apply the basic methods covered in class in testing a series of hypotheses that they develop on their own. Students should focus on a topic of their choosing in consultation with me and find their own data from various web sources. For this purpose do have a look into the following website of the Association of Religion Data Archives: http://www.thearda.com/ among others.

There also exists a number of rich websites where data for already published papers can be obtained. We also have a data collection of our own that we could provide to interested students. Do consult us about these data.

A detailed discussion of how to bring together such a paper will be made in class. We expect the students to choose a topic with a well-developed set of hypotheses and a dataset by the end of the first month. A clearly written paper proposal should be written by July 20th and a short presentation should be made to the class. A handout summarizing basic principles of such an undertaking will be distributed. July 27th is the date for the first draft of your papers which should be finalized by the end of the Semester.

Student Evaluation:
We will evaluate your final grades based on computer assignments, your research progress and term paper, non-computer assignments, attendance and in class assignments.

The final evaluation will be determined by the following components:

- Computer assignments, problem sets 50%
- Term Paper 50%

Tentative Course Outline and Reading Assignments:
You will benefit the most if you read the materials before the class when they are discussed.

0. Course Introduction and Conceptual Overview (3 lectures)
Math review; guidelines for assignments.
Measurement; nonlinearity; linear and nonlinear models; right-hand-side and left-hand-side variables.
Review of linear regression; standardization.
interpreting the LRM; simple hypothesis testing.
Identification; degrees of freedom; MLE.

1. Binary Outcomes (2 lectures)

Expected values of 0/1 variables; binary outcomes; LPM.
Discrete and continuous variables; latent variable models for binary outcomes; Pr(y=1); scaling in regression.
Identification of Pr(y=1); nonlinear probability models; numerical methods, MLE and sample size; slope and intercept; using Excel.
Using Pr(y=1) to interpret the BRM; plots.
Discrete change; articles using Pr(y=1); example from the National Academy of Sciences.
Odds ratios; articles using odds ratios.
Catch-up; nonlinear, nonlinear models; group comparisons.


2. Hypothesis Testing and Goodness of Fit (1 lecture)

Principles of hypothesis testing; Wald tests.
LR tests; confidence intervals.
Residuals and outliers.
BIC and AIC; measures of fit.


3. Ordinal Outcomes (2 lectures)

Ordinal variables; a latent variable model.
Estimation; latent variable interpretations; Pr(y=k).
Odds ratios; parallel regression assumption and proportional odds; articles using ordinal models.


4. Nominal Outcomes (3 lectures)
Multinomial logit as a set of BLMs.
Tests for the MNLM.
Interpretation; discrete change; odds ratio plots.
Class examples of odds ratio plots; utility maximization; conditional logit; IIA.
Applications; multinomial probit.


5. Limited Outcomes and Count Data (3 lectures)
The Problem of Censoring
The Tobit Model for censored outcomes
Count models; adding heterogeneity; something for nothing. PRM.
Predicted probabilities in count models, NBRM, unobserved heterogeneity.

