

Categorical and Limited Dependent Variables

Public Affairs 56:824:708:01

Spring Semester 2018, BSB 107, **Thursdays** 6-8:45pm

Paul A. Jargowsky, Ph.D.

856-225-2729; 321 Cooper St.; paul.jargowsky@rutgers.edu

Office Hours: Tuesdays 5-7pm and by appointment

Course Description

The estimation of empirical models is essential to public policy analysis and social science research. Ordinary Least Squares (OLS) regression analysis is the most frequently used empirical model, and is appropriate for analyzing *continuous* dependent variables with well-behaved distributions. This course examines several types of advanced regression models for troublesome dependent variables that violate one or more of the assumptions of the OLS regression model. For example, in social science research we frequently have to deal with dependent variables that are:

- *binary conditions*, such as being pregnant or not, employed or not, etc.;
- *categorical*, such as mode of transportation, level of education, etc.;
- *truncated* or *censored*, such as contributions to an individual retirement account that are limited by law to certain dollar amounts;
- *counts* that are always positive integers, like the number of children born to a given woman or the number of traffic accidents on a given day;
- *events* that may occur sooner or later for different subjects, like death, marriage, or recidivism.

The principal models examined in the course are binary logit and probit, multinomial logit, ordinal logit and probit, tobit, the family of Poisson regression models, and event history models. All these models are estimated using maximum likelihood estimation (MLE). The course focuses primarily on the application and interpretation of the models, rather than statistical theory. While there is of necessity quite a bit of math in this course, I will work hard to make the math accessible by discussing the intuition behind the math. By completing this course, you will be well equipped to analyze a broad array of variables frequently encountered in the social sciences.

Course Prerequisites

56:824:709 Quantitative Methods II or the equivalent is required. In other words, you should have a strong grounding in Ordinary Least Squares (OLS) regression at the level of Damodar Gujarati, *Basic Econometrics*, chapters 1-9, or Stock and Watson, *Introduction to Econometrics*, chapters 1-7.

Student Learning Objectives/Outcomes

- Students will learn the theory and practice of regression models for limited and categorical dependent variables, including logit, probit, ordinal logit, ordinal probit, multinomial logit, Poisson regression, Tobit and related models, and event history analysis.
- Students will learn how to interpret and critique these models by reviewing published papers drawn from social science literature.
- Students will develop proficiency in applying and interpreting these models using data provided by the instructor and/or data from their own research and employment.
- Students will demonstrate mastery of the material by writing an empirical paper using one or more of the models discussed in class and presenting their analysis and findings in class. The goal is to produce a publication-quality paper and submit it to an appropriate academic journal.

Textbooks and Materials

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

- This timeless classic is used throughout the course. It provides the underlying theory for most of the models we discuss.

Long, S. J. and Freese, J. 2015. *Regression Models for Categorical Dependent Variables Using Stata*, 3rd ed. College Station, TX: Stata Press.

- This book is more of a “how-to” guide for running the models in Stata. It has less theory but more specifics on executing the models and manipulating the output.

Mario Cleves, William W. Gould, Roberto G. Gutierrez, and Yulia Marchenko. 2016. *An Introduction to Survival Analysis Using Stata*, 3rd Edition. Stata Press.

- Used for the event history analysis part of the course only. It covers both theory and execution in Stata.

Course Software.

The software for the course is *Stata* version 15, although version 14 or any recent version will probably work just fine. There is no need to upgrade from version 14 unless you want to have the latest and greatest. (However, you should make sure you loaded all the patches and bug-fixes. Type “update query” at the command line while connected to the internet to check.)

You do not need to buy anything, because Stata software is available in the computing labs and online at <http://apps.rutgers.edu>. Information buying your own copy at a discount may be found here:

<https://www.stata.com/order/new/edu/gradplans/student-pricing/>

A 6 month license for the basic version of the program is \$45.

Requirements

1. Students are required to take two in-class tests. The first test covers material from the first half of the class: Logit and Probit and related models for binary, ordinal, and multinomial variables. The second test covers models for Tobit and related models, Poisson and related models, and Event History Analysis. The exams are open-note, open-book.
2. There are several short problem sets due at the beginning of class on the dates indicated on the schedule. The lowest problem set score will be dropped.
3. Students must complete an empirical paper on an approved topic using one or more of the techniques covered in this course. You are encouraged to think about topics and potential datasets early in the semester. A typical paper will be 15 to 20 double-spaced pages.
4. On the indicated dates, students will be asked to turn in a proposed paper topic and a first draft of the empirical paper. These items will not be graded.

Grading Policy

The grading in the course is based on the problem sets, examinations, and the empirical paper. The weights assigned to each are as follows:

6 Problem Sets (lowest dropped)	30%
Test I (October 27)	20%
Test II (December 8)	20%
Empirical Paper	30%

After computing the student semester average on a 100-point numeric scale and rounding to the nearest whole number, letter grades will be assigned as follows:

<u>Numeric Range</u>		<u>Letter</u>	<u>Grade</u>
<u>min</u>	<u>max</u>	<u>Grade</u>	<u>Points</u>
90	100	A	4.0
85	89	B+	3.5
80	84	B	3.0
75	79	C+	2.5
70	74	C	2.0
0	69	F	0.0

In other words, you have to average 89.5000..., which rounds up to a 90, or higher to get an A.

Course & Instructor Policies

Late Work. Problem sets will not be accepted late, because the answers are discussed in class on the day they are due. Due to a medical emergency or other valid reason, you may be excused from turning in a problem set. In such cases, the grade will be computed based on the remaining problem sets. Consult me *in advance of the due date*, if at all possible, if such a contingency should arise. Likewise, I cannot give early or late examinations. Arrange your schedule now to avoid potential conflicts.

Calculator. A calculator is a virtual necessity for this class. However, any basic scientific calculator will do. The following functions are necessary: square root, y^x , e^x , and $\ln(x)$. Such calculators can often be obtained for under \$10 and there are free smart phone apps. (The built in iPhone calculator has everything you need, but turn the phone sideways to access the scientific functions.) You will not need graphing capability or programmability.

Attendance. Attendance is entirely optional. However, be advised that you are responsible for any material covered in class, whether or not it was in the readings or lecture notes. You are also responsible for any announcements made in class. For most students, attendance is simply essential to learning the material. If you do need to miss a class, be sure to consult with a fellow student to learn what transpired.

Schedule, Readings and Assignments

Last modified January 17, 2018

1	Jan. 18	<p>Introduction to Course</p> <p><i>Optional background readings (review only if/as needed):</i></p> <p><u>1. Review of Probability and Statistics</u> (optional) Paul Jargowsky and Rebecca Yang, "Descriptive and Inferential Statistics" (optional) Gujarati, <i>Basic Econometrics</i>, Appendix A (optional) Stock and Watson, <i>Introduction to Econometrics</i>, Chapters 2 and 3</p> <p><u>2. Review of OLS</u> (optional) Gujarati, Chapters 2-9, especially 7-8 (optional) Stock and Watson, Chapters 4-7, especially 6-7</p> <p><u>3. Introduction to Stata</u> Stata Youtube Channel, Tour of the Stata 15 User Interface. Stata Learner, Getting Started with Stata Stata Learner, Keeping Track of your Stata Session Long and Freese, Chapter 2 (For now, just skim through this. It is a resource you can come back to when you have questions about Stata.)</p> <p>A Wee Bit of Calculus</p>
2	Jan. 25	<p>Binary Dependent Variables: Logit and Probit</p> <p>Long, Sections 3.1-3.4, 3.7-3.9 Long and Freese, 5.1-5.2, 5.4 Explore spreadsheet "Logit Function of XiB" and "Comparison of Probit and Logit" Williams and Nesiba (1997). "Racial, Economic, and Institutional Differences in Home Mortgage Loans," <i>Journal of Urban Affairs</i> 19: 73-103.</p> <p>Principles of Maximum Likelihood Estimation</p> <p>Myung (2003). "Tutorial on Maximum Likelihood Estimation," <i>Journal of Mathematical Psychology</i> 47:90-100, sections 1-2 (the rest is optional). (optional) Long, Section 2.6 (Highly mathematical)</p>
3	Feb. 1	<p>Interpretation and Hypothesis Testing in Logit and Probit Models</p> <p>Long, Sections 4.1, 4.3 Long and Freese, Chapter 5.3, 5.5, 6.1, 6.2, 6.6 Browne (1997). "Explaining the Black-White Gap in Labor Force Participation Among Women Heading Households," <i>American Sociological Review</i> 62: 236-252, especially pages 244-248. Weil (2001). "Assessing OSHA Performance: New Evidence from the Construction Industry," <i>Journal of Policy Analysis and Management</i> 20: 651-674.</p> <p><u>Problem Set 1 Due</u></p>

Schedule, Readings and Assignments

Last modified January 17, 2018

4	Feb. 8	<p>Ordinal Dependent Variables: Ordinal Logit and Ordinal Probit</p> <p>Long, Sections 5.1-5.4 Long and Freese, 7.1-7.4, 7.7-7.11 (optional) Long and Freese, 7.6 Winship and Mare (1984). "Regression Models with Ordinal Dependent Variables", <i>American Sociological Review</i> 49: 512-525. Hughes and Waite (2002). "Health in Household Context: Living Arrangements and Health in Late Middle Age," <i>Journal of Health and Social Behavior</i> 43:1-21. (Focus on Table 2 results.) (optional) Alvarez and Brehm (1998). "Speaking in Two Voices: American Equivocation About the Internal Revenue Service," <i>American Journal of Political Science</i> 42: 418-452. (Model is heteroskedastic ordered probit.)</p> <p><i>Problem Set 2 Due</i></p>
5	Feb. 15	<p>Nominal Dependent Variables: Multinomial Logit</p> <p>Long, 6.1-6.2, 6.4-6.6 Long and Freese, 8.1-8.8, 8.10-8.11 Stratton, O'Toole, and Wetzel (2008). "A Multinomial logit model of college stopout and dropout behavior," <i>Economics of Education Review</i> 27: 319-331.</p> <p><i>Problem Set 3 Due</i></p>
6	Feb. 22	<p>Nominal Dependent Variables: Conditional Logit</p> <p>Long, 6.7-6.10 Long and Freese, 8.12.2 Sections 1 and 2 of Alvarez and Nagler (1998). "When Politics and Models Collide: Estimating Models of Multiparty Elections," <i>American Journal of Political Science</i> 42: 55-71.</p>
7	Mar. 1	<p>Nominal Dependent Variables: Multinomial Probit</p> <p>Long and Freese, 8.12.3-8.12.4 Sections 3-5 of Alvarez and Nagler (1998). "When Politics and Models Collide: Estimating Models of Multiparty Elections," <i>American Journal of Political Science</i> 42: 71-96. Carole J. Wilson (2008). "Consideration Sets and Political Choices," <i>Political Behavior</i> 30: 161-183. (optional) Hausman and Wise (1978). "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing interdependence and Heterogenous Preferences," <i>Econometrica</i> 46: 403-426.</p> <p><i>Problem Set 4 Due</i></p>
8	Mar. 8	<p style="text-align: center;">Test I <i>Open book, open note</i> <i>Bring a calculator</i></p>

Schedule, Readings and Assignments <i>Last modified January 17, 2018</i>		
9	Mar. 22	<p>Censored and Truncated Dependent Variables: Tobit Long, Chapter 7 Beron (1990). "Child Support Payment Behavior: An Econometric Decomposition," <i>Southern Economic Journal</i> 56: 650-663. (optional) McDonald and Moffitt (1980). "The Uses of Tobit Analysis," <i>The Review of Economics and Statistics</i> 62: 318-321. <u>Proposed Empirical Paper Topic Due</u></p>
10	Mar. 29	<p>Censored and Truncated Dependent Variables: Extensions Jargowsky, "Using Stata's ML Utility." Gould (1992). "At Home Consumption of Cheese: A Purchase-Infrequency Model," <i>American Agricultural Economics Association</i> 74: 453-459 Winship and Mare (1992). "Models for Selection Bias," <i>American Sociological Review</i> 18: 327-350 (Focus on examples of selection bias and the Heckman estimator) (optional) James J. Heckman (1979). "Sample Selection Bias as a Specification Error," <i>Econometrica</i> 47: 153-161 <u>Problem Set 5 Due</u></p>
11	Apr. 5	<p>Count Dependent Variables: Poisson Regression and Related Models Long, Chapter 8 Long and Freese, 9.1-9.4, 9.6-9.7 Hughes & Waite (2002). "Health in Household Context: Living Arrangements and Health in Late Middle Age," <i>Journal of Health and Social Behavior</i> 43: 1-21 (Yes, it's the same article as before. This time look at the Poisson results.) (optional) Minkoff (1997). "The Sequencing of Social Movements," <i>American Sociological Review</i> 62: 779-799 <u>First Draft of Empirical Paper Due</u></p>
12	Apr. 12	<p>Event History Analysis: Theory Cleves et al., Chapters 1-4 <u>Problem Set 6 Due</u></p>
13	Apr. 19	<p>Event History Analysis: Interpretation and Implementation Cleves, Chapters 5-8, 9.1, 12.1, 13.1-13.3 Finocchiaro and Lin (2000). "The Hazards of Incumbency," unpublished paper. (optional) Cox, D. R. (1972). "Regression Models and Life Tables," <i>Journal of the Royal Statistical Society, Series B</i> 34: 187-220. (Cited over 9000 times!)</p>
14	Apr. 26	<p>Test II <i>Open book, open note</i> <i>Bring a calculator</i></p>

Schedule, Readings and Assignments

Last modified January 17, 2018

Monday,
Apr. 30

Empirical Paper Due