EDUC 799: CATEGORICAL AND LIMITED DEPENDENT VARIABLE MODELING

Fall 2013

Location:
Lecture: 2224 School of Education
Lab: G442 School of Public Health
(109 S. Observatory)

Time:
Lecture: Tuesdays 9:00am to 12:00pm
Lab: Thursdays 9:00am to 11:00am

Primary Instructor:
Julie R. Posselt, Ph.D.
Assistant Professor, Center for the Study of
Higher and Postsecondary Education
University of Michigan
2108E School of Education
Office: (734) 764-8423
Email: jposselt@umich.edu
Office hours: Wed 2-4pm or by appt

Graduate Student Instructor:
Joe Howard
Doctoral Student, Higher Education
Office: 2018M School of Education
Email: joehow@umich.edu
Office hours: Wed Noon-2pm

COURSE DESCRIPTION:

This course focuses on multivariate regression methods in which the dependent variable is non-continuous, as well as the application of those methods to education-related research questions. These types of outcomes are ubiquitous in educational and social research (e.g., those that take the form of a Likert scale, whether a student graduates from high school, the type of college or major s/he enrolls in, which political party one votes for, etc). Yet employing the general linear model for these types of variables frequently violates regression assumptions and produces misleading results. Non-linear regression methods are now widely used as alternatives to OLS regression with non-continuous variables, and are not only an important set of skills for any quantitative researcher, but for qualitative researchers who want to be critical consumers of studies that employ such approaches as logistic, ordinal, probit, multinomial, and Poisson models.

As an education course, we will illustrate and practice these methods using education data, but students are welcome to use any data they wish for the final project. All assignments and lab activities will use Stata 12. The course is open to graduate students who are comfortable with multiple linear regression, which is the foundation and point of departure for the methods we cover. Like most statistics classes, the learning in this class is cumulative, with early content providing the foundation for later learning. Therefore, if you find yourself struggling with any of the course material, please don’t wait to approach us. We strive to create a learning culture of high expectations with high support.

LEARNING OBJECTIVES:

The course is designed with three learning objectives in mind:
(a) Understanding the need for and statistical properties of non-linear probability models,
(b) Estimating, interpreting, and optimizing multivariate models with categorical and limited
dependent variables as their outcome, and
(c) Applying these methods to answer research questions and questions of practice about education.

The curriculum, instruction, and assessments therefore emphasize technique, interpretation, and implications of results for decision making.

ORGANIZATION OF THE COURSE:

Our Tuesday meetings will emphasize (1) Theory underlying the methods we study, (2) Walking you through an example of the code and output in Stata, and (3) Discussion as well as time for questions/answers. Preparation for lecture primarily involves reading: I strongly encourage you to read the Long text before coming to class, even if you need to browse it again later.

Discussions will alternate each week between (1) Examples of peer-reviewed published research that utilizes the method we are studying and (2) A reading that examines one of the many contexts of quantitative research—historical, epistemological, visual, professional, etc.

Much of the class is devoted to development of technical skill in and theoretical understanding of a body of methods. These readings, and the discussions we will have about them in lecture, will foster awareness of and critical thinking about how scholars have conducted the type of work we are learning and the context in which the practice of statistics occurs. We are more than number crunchers or analysts; we are part of a long history and conversation about knowledge production.

Lab time on Thursdays will be devoted to (1) reinforcing material discussed in lecture, (2) exercises using Stata, and (3) consultation about assignments and final projects. The GSI is also available for office hours to provide assistance with assignments and final projects.

The course is organized as follows. Please note that the dates on which these topics will be discussed may be adjusted based on how quickly we progress through the materials, conflicts within the semester, and particular learning needs of the group.

<table>
<thead>
<tr>
<th>Week No.)</th>
<th>Topic Area</th>
<th>Meeting Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td>2)</td>
<td>Introductions /Linear Regression/ Stata Intro</td>
<td>September 3, 5</td>
</tr>
<tr>
<td>3)</td>
<td>Logistic Regression</td>
<td>September 10, 12</td>
</tr>
<tr>
<td>4)</td>
<td>Logistic Regression (cont’d)</td>
<td>September 17, 19</td>
</tr>
<tr>
<td>5)</td>
<td>Logistic/ Probit Regression</td>
<td>September 24, 26</td>
</tr>
<tr>
<td>6)</td>
<td>Hypothesis Testing and Goodness of Fit</td>
<td>October 1, 3</td>
</tr>
<tr>
<td>7)</td>
<td>Ordered Logit/Ordinal Outcomes</td>
<td>October 8, 10</td>
</tr>
<tr>
<td>8)</td>
<td>Out-of-Sample Prediction</td>
<td>October 17 (No lecture; Fall Study Break)</td>
</tr>
<tr>
<td>9)</td>
<td>Multinomial Outcomes</td>
<td>October 22, 24</td>
</tr>
<tr>
<td>10)</td>
<td>Multinomial Outcomes (cont’d)</td>
<td>October 29, 31</td>
</tr>
<tr>
<td>11)</td>
<td>Conditional Outcomes</td>
<td>November 5, 7</td>
</tr>
<tr>
<td>12)</td>
<td>Count Outcomes</td>
<td>November 12 (No lab, ASHE conference)</td>
</tr>
<tr>
<td>13)</td>
<td>Interaction Effects</td>
<td>November 19, 21</td>
</tr>
<tr>
<td>14)</td>
<td>Course Wrap-Up</td>
<td>December 3, 5 (Lab focus on final projects)</td>
</tr>
<tr>
<td>15)</td>
<td>In-Class Presentations and Turn in Papers</td>
<td>December 10, 12</td>
</tr>
</tbody>
</table>
A FEW WORDS ABOUT MATH:

The theory undergirding the methods we will discuss involves matrix algebra and calculus. Understanding these will enable the deepest mastery of the course content, especially as it is communicated in the Long textbook; however, I will not be evaluating you in ways that test your knowledge of matrix algebra or calculus. Rather, my primary goal is for every student to confidently and adeptly estimate the right type of regression for the research question and outcome; to accurately interpret the results and communicate them in both the precise language of statistics and the real-world meanings that policymakers, the media, and the public require (You will hear me often ask you what a finding means, substantively); and to execute such an analysis as part of a complete research project that focuses on education.

These tasks will require you to have a solid grasp on basic algebra, including work with exponents and logarithms. For the first four weeks (and beyond, if there is demand from students), I will devote the first ten minutes of class to reviewing these functions. If you don’t need the review or prefer to brush up on your own, you can come at Michigan time (9:10am). However, on the fourth week of class, I will give a low-stakes math quiz to the entire class during class time to check on whether everyone is solid on these basic functions as we enter the heart of the class.

COURSE REQUIREMENTS:

Grading will be on the A-F scale for three (3) semester credits. Incomplete grades are strongly discouraged, and must be arranged with the instructor before the end of the last regular class meeting (does not include the session(s) for final paper presentations). Your grade will be calculated based on the following requirements:

a. ATTENDANCE, ENGAGEMENT WITH READINGS, AND CLASS PARTICIPATION (25% of course grade)

Attendance is required, except in cases of contagiousness and conference travel. If you are sick, please stay home and notify the instructor for lecture or lab, as appropriate. If you have conference travel, please work this out well in advance with the primary instructor. More than one absence, except in unusual circumstances that have been discussed in advance, will negatively affect your course grade.

Participation means actively supporting our collective learning. I am committed to creating a focused, non-competitive atmosphere in which we come together to build mastery in methods and think critically about their use. To this end, I expect engagement through asking and answering questions (of the instructors and one another) and contributing to discussion in ways that demonstrate you have given thought to the weekly readings. It also means turning OFF your phone altogether (unless you have a child or other caregiving responsibilities) and not pulling up email or other social media during class.

Scholars are frequently called upon to conduct manuscript reviews for their colleagues—either formally for a journal, conference, publishing house, etc. or informally to guide revisions before the manuscript is submitted. Rarely do graduate students get practice in this type of review, however. Therefore, one element of your participation grade will be writing up a review and leading the class in discussion of an article that uses one of the methods we are studying. You will have the opportunity to sign up for this early in the class.
b. DELIVERABLES
   i. WEEKLY ASSIGNMENTS (35% of final grade)
      a. Sometimes, these assignments will involve conducting statistical analyses of education data, other times they will involve critiquing published research, and still other assignments may be designed for you to demonstrate your ability to write-up the results of statistical analysis. The assignments will give you practice in such important skills as critical reading, coding, interpretation, and communicating your ideas in writing.
      b. A hard copy of the assignment on the specified due dates, usually at the start of lecture or lab. Note that due dates follow lab to allow for time to discuss during lab any challenges you encountered in the design, coding, and/or interpretation of results.
      c. Please put the last four digits of your student ID at the top of the weekly exercises, NOT your name, so that we may blindly assess the work.
      d. For assignments that involve considerable coding, you will turn in your Stata log file in advance of the assignment’s due date.
      e. Graded through rubrics. For most assignments, you will receive separate grades on two dimensions: (1) The quality of technical work, and (2) The clarity and accuracy of your communication.
      f. Late assignments will negatively affect your grade on the assignments’ substantive dimension, and your grade in the course.

   ii. COURSE PROJECT (40% of final grade)
       The course project begins early in the semester, and will develop into an original, rigorous, conference-ready inquiry by the end of the semester, including a ~20 page paper (double spaced; 1 inch margins; excluding tables, figures, references, appendices) and 15 minute presentation. It should focus on a novel question about education, and should use one of the methods we cover in the course, using any data you wish.

       You are invited to collaborate with another member of the class on the course project, although this is not required. The project involves deadlines in several stages, and at each stage you will receive feedback from one or both of the instructors. Only the final paper and presentation will be graded. You will have the rubric for the final paper and presentation well in advance of their due date. Please use APA format for all assignments turned in toward the course project (unless you are from a discipline in which another publication style is normative; if this is the case, please notify me).

       i. Abstract/ Proposal: Sept 24
       ii. Literature review: Oct 17
       iii. Revised literature review and draft of proposed methodology. You will turn this in and briefly present it to the instructors and class for feedback.: Oct 31
       iv. Revised literature review and methodology, and draft of findings: Nov 26
       v. FINAL WEEK: Complete paper, including introduction and conclusion and Formal 15 minute presentation, including feedback from instructors and class: Dec 10, 12
          (This is a hard deadline, to which no exceptions will be made except in case of unlikely prior arrangement with the instructor.)
TEXTBOOKS:

Required:


The book is available at Ulrich's Bookstore, or Amazon and the following URLs for used copies:

- [http://www.fetchbook.info/compare.do?search=0803973748&startFrom=1](http://www.fetchbook.info/compare.do?search=0803973748&startFrom=1)


Make your life easier this semester and in the years to come and buy this companion to the textbook. It is widely regarded as one of the most useful books that a quantitatively oriented Ph.D. student in the School of Education can buy. Although its contents include user-written commands that are becoming outdated with the introduction of the `margins` and `marginsplot` commands in Stata 11 and beyond, it still contains a wealth of practical information on coding in Stata, model estimation, post-testing, and so forth. It is available in Ulrich’s, at Amazon, and used copies may be found at:


You will find additional weekly readings on CTools. Look under Resources/ Additional Readings.

Optional:


Another Stata companion. Most of this book is not about logistic or other categorical DV modeling, and I am making copies of the required chapters will available on CTools. However, it is such a great resource for using the margins and marginsplot to model and display findings and interaction terms across many types of regression that I am making it available in the bookstore.

The three books mentioned above are also available directly from Stata at: [http://www.stata.com/bookstore/cldv.html](http://www.stata.com/bookstore/cldv.html). In the past, the Stata website has had the lowest prices for new copies.

For additional background on the estimation technique that underlies the methods we will study, pick up Scott Eliason’s little green Sage book, Maximum Likelihood Estimation: Logic and Practice, Sage Publications ISBN 0803-941072. It is available from Ulrich’s and Amazon, as well as [http://www.fetchbook.info/compare.do?search=9780803941076](http://www.fetchbook.info/compare.do?search=9780803941076)

If you are new to Stata, you may want to purchase an introduction. A good one is: *A Gentle Introduction to Stata* by Alan C. Acock (see [http://www.stata-press.com/books/acock.html](http://www.stata-press.com/books/acock.html)). However, many websites (e.g., [http://www.ats.ucla.edu/STAT/stata/](http://www.ats.ucla.edu/STAT/stata/)) contain information about using Stata, including coding hints, sample programs, etc., so use your judgment about whether you need this additional book or not.
SOFTWARE:

For this course, you will need access to a recent copy of Stata, the statistical software with which we will conduct analyses. The Sites campus computing labs, including the one on the third floor of the School of Education, all have version 12.0 of Stata’s SE software installed. You can also access the software from any computer with a reliable internet connection using the library’s remote desktop system, virtual site: https://virtualsites.umich.edu/connect/setup/. However, you may find this a somewhat unreliable way to work, as the system will drop you if the internet connection is broken for any reason.

You may want to purchase your own copy of the software from Stata (at www.stata.com) to make conducting the analyses more convenient. Note that the most current version is Stata 13 (just released summer 2013!) and that different versions have different capabilities (see http://www.stata.com/products/whichstata.html) for details. The University of Michigan has a campus-wide GradPlan agreement with Stata allowing students to obtain a copy of Stata, either with an annual or perpetual license. These prices are discounted from the already-discounted academic price of the software. See current pricing and availability at http://stata.com/order/new/edu/gradplan.html. Stata’s phone number is: 800-82-8272 (Monday through Friday 8:00 to 5:00 Central Time), however, ordering online is easiest. If you order online, be sure to include your UMICH.EDU email address to ensure you receive the best price. Once your order is processed, you will be contacted by a StataCorp salesperson with campus pick-up times and location, usually 2 - 3 business days after the order is placed. You may pay using a credit card (Visa, MasterCard, American Express or Discover) or a faxed university purchase order.

ON DIVERSITY:

Although the topical focus of this class is statistics, diversity is nonetheless critical to building both our academic knowledge and an equitable learning community. Research by UM professor Scott Page finds that when a diverse group comes together around a shared problem, they will solve it more successfully and innovatively than homogeneous groups of experts, because diverse individuals contribute more angles on the problem and ideas toward its solutions (Page, 2008). However, this is only true when members of the group believe in the value of diversity (Ely & Thomas, 2001).

Please see the attached discussion guidelines, which we will review and add to (as you see fit) on the first day of class as part of our collective framing of this course. I expect that our discussions and work together will be respectful and encouraging of different perspectives, and inclusive of people regardless of background and identity. I also expect that some of us will have moments in the class where identities are triggered—positively and negatively. I am comfortable pausing our stats talk to address such issues as part of our learning during class time. If you have concerns about anything that comes up in your experience of the class or as a graduate student, generally, please know that my office is a safe, confidential space to raise them now and when the course is done.

In selecting sample articles for critique and additional readings, I have tried to reflect a diverse array of interests. I have also selected readings with an eye toward developing your capacity to use statistics thoughtfully with regard to race, gender, socioeconomic status, and other systems of inequality. You may be interested in institutional resources about diversity matters, non-discrimination, ADA, and more that the University of Michigan’s Office for Institutional Equity has posted online: http://www.umich.edu/~hraa/oie/
STUDENT RIGHTS AND RESPONSIBILITIES:

I urge you to understand what your rights and responsibilities are as a student. Information on these topics can be found at: http://www.umich.edu/~oscr/. You may also want to read the University’s General Catalogue, especially the section that details your rights as a student and the section that discusses the University’s expectations of you. (See http://www.rackham.umich.edu/policies/.)

ACCOMMODATIONS FOR SPECIAL NEEDS:

I am happy to confidentially discuss with you any accommodations for special needs that will help you succeed in this course. Please approach me early in the semester, & see http://www.umich.edu/~sswd/ for more information about the University’s services for students with disabilities.
WEEK 1 (September 3 & 5)

LECTURE: INTRODUCTIONS, EXPECTATIONS FOR THE COURSE & REVIEW OF LINEAR REGRESSION
We will discuss the syllabus, introduce the data sets we’ll be using during the class, and review linear regression.

LAB: INTRODUCTION TO STATA

REQUIRED READINGS:
Long, Chapters 1 & 2.

Long & Freese, Chapters 1 & 2

ASSIGNMENTS: Some of the Stata commands we will use are Long and Freese’s user-written commands not built into the current version of Stata. In their supplement to the text, Long and Freese tell you how to load these programs on to any computer you will use. Please familiarize yourself with this by next week. You may also want to check out Long’s web site that contains information about these “SPOST” commands (see http://www.indiana.edu/~jslsoc/spost.htm).

Also, please familiarize yourself with the CTOOLS site for this class (named “EDUC 799 001 F13”) and peruse the materials that are there as you will need to interface with this site on a regular basis. Focus on the “Stata Materials” folder and especially the “Introduction to Stata,” “Important Stata Commands,” “Stata Primer,” “Stata Review Materials,” and “Things to Know About Stata.”

DUE: Nothing.

WEEK 2 (September 10 & 12)

THE LINEAR PROBABILITY MODEL & LATENT VARIABLE MODEL FOR BINARY DEPENDENT VARIABLES

LECTURE 9:00-9:10: Reviewing basic exponential functions

REQUIRED READINGS:
Long, Chapter 3, pp. 34-50


DUE TUES SEPT 10:  1-2 page review of the McGrayne and Silver chapters. This will not be graded, but you will turn it in and it will be the basis for discussion in lecture.

WEEK 3 (September 17 & 19)

MAXIMUM LIKELIHOOD ESTIMATION & LOGISTIC REGRESSION
Nonlinear binary models are an appropriate alternative to the linear probability model (discussed in Session 2) when categorical outcomes are being analyzed. We discuss two of the most commonly used nonlinear binary regression techniques, logistic and probit regression.

LECTURE 9:00-9:10: Reviewing logarithmic functions

REQUIRED READINGS:
Long, Chapter 3, pp. 50-83;

Long & Freese, Chapter 4 pp. 129-154.

RECOMMENDED READING:
(Available in Library Reserves and at the Bookstores with other textbooks. Kind of a gem.)

DUE:
•  Tuesday, September 17 (by start of class) – Log for Assignment 1.
•  Thursday, September 19 (start of lab) – Assignment 1

WEEK 4 (September 24 & 26)
LOGISTIC REGRESSION (cont'd)

We continue our discussion of logistic regression modeling. This week, we’ll introduce predicted probabilities and marginal effects as ways to interpret your findings.

LECTURE will include a low-stakes math quiz emphasizing basic algebra, exponents, and logarithms.

REQUIRED READINGS:
Long & Freese, Chapter 4 pp. 154-181.


RECOMMENDED READINGS:
SUPPLEMENTAL READINGS:


DUE:
- Tuesday, September 24 (by start of class) – Log for Assignment 2
- Tuesday, September 24 (by start of class) – 2 paragraph proposal/abstract for final paper
- Thursday, September 26 (start of lab) – Assignment 2

WEEK 5 (October 1 & 3)

PROBIT REGRESSION

REQUIRED READINGS:
Reread Long Chapter 3;

Reread Long and Freese Chapter 4


DUE:
- Tuesday, October 1 (by start of class) – Log for Assignment 3
- Thursday, October 3 (start of lab) – Assignment 3

WEEK 6 (October 8 & 10)

HYPOTHESIS TESTING AND GOODNESS OF FIT STATISTICS

We review tests of hypotheses that can be used with maximum likelihood estimation techniques. Particularly important is a sound understanding of goodness of fit statistics such as the likelihood ratio test, and how to use this test to make inferences about the relative efficacy of one model vs. another.

REQUIRED READINGS:
Long, Chapter 4.

Long & Freese, Chapter 3 & Review Chapter 4.


DUE: Thursday, October 10 (start of lab) – Assignment 4

WEEK 7
(No lecture October 15 due to Fall Break; lab will meet on October 17)
OUT-OF-SAMPLE PREDICTION

We are now going to use logistic regression to predict out-of-sample. This strategy is used extensively by researchers who work in enrollment management within colleges and universities. Particularly important is a sound understanding of goodness of fit statistics such as the likelihood ratio test, and how to use this test to make inferences about the relative efficacy of one model vs. another.

REQUIRED READINGS:


DUE: Thursday, October 17 (start of lab) – Draft of literature review for final project

WEEK 8 (October 22 & 24)
ORDERED LOGIT & ORDINAL OUTCOMES

This week, we will focus on how to statistically model outcomes that can be ordered. It is not appropriate to model ordered dependent variables as though they are interval. Ordered logit and probit regression models allow us to estimate these models more appropriately than when using standard linear regression techniques.

REQUIRED READINGS:
Long, Chapter 5.

Long & Freese, Chapter 5.

SUPPLEMENTAL READINGS:

A good treatment of logit and probit models. STATA code provided.

An example of the use of ordered probit

DUE: Thursday, October 24 (start of lab) – Assignment 5
WEEKS 9 & 10 (October 29, November 1, 5, 7)
MULTINOMIAL LOGISTIC REGRESSION

In this section of the course, we move to discuss a statistical model specifically designed to estimate nominal outcomes, that is, dependent variables that are not ordered.

REQUIRED READINGS:
Long, Chapter 6, thru p.178.

Long & Freese, Chapter 6 thru p.235.

DUE:
• Tuesday, October 31 (start of class) – Revised literature review & draft of methodology
• Thursday, November 7 (start of lab) – Assignment 6

WEEK 10 (November 5 & 7)
MULTINOMIAL LOGISTIC REGRESSION, cont’d

REQUIRED READINGS: Read all three examples. Sign up to write a 1-2 page in-depth review/ critique.


WEEK 11 (November 12 – No lab on 11/14 due to ASHE)
CONDITIONAL LOGIT MODELS

In the multinomial logistic model each explanatory variable has a different effect on each outcome. The conditional logit model is a closely related technique in which the coefficients for the variable are the same for each outcome, but the values of the variables differ for each outcome.

REQUIRED READINGS:
Long, Chapter 6, pp. 178-186.

Long & Freese, Chapter 6, pp. 235-244.

SUPPLEMENTAL READINGS:


DUE: **Tuesday**, November 12 (start of class) – Assignment 7

WEEK 12 (November 19 & 21)
COUNT OUTCOMES: THE POISSON REGRESSION MODEL

Sometimes we study the frequency of an event occurring, such as the number of times a student is late or absent from school. Dependent variables that contain count information are often treated as continuous and linear regression techniques are applied. However, this particular application of the linear regression technique may result in inefficient, inconsistent, and biased estimates. The Poisson regression model redresses these problems.

REQUIRED READINGS:
Long, Chapter 8.


RECOMMENDED READING:

SUPPLEMENTAL READINGS:

ASSIGNMENT: Continue to work on your final papers.

WEEK 13 (November 26; No Lab—Happy Thanksgiving)
INTERACTION EFFECTS

REQUIRED READINGS: Mitchell, Chapter 18 (Nonlinear Models) from *Interpreting and Visualizing Regression Models Using Stata* (on ctools if you opted not to buy the entire book)

DUE: Tuesday, November 26 (start of class) - Literature review, revised methodology, draft findings
WEEK 14 (Dec 3 & 5)
PEOPLE’S CHOICE

LECTURE POSSIBILITIES:
Multilevel modeling options
Missing Data Management
Lab: Workshop time for final papers

WEEK 15 (Dec 10 & 12)
IN CLASS PRESENTATIONS DURING LECTURE AND LAB TIME