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). 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 15. 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 outcomes, 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 Friday 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 at office hours to provide assistance with assignments and final projects. If you need to meet with the GSI, please sign up for an appointment on her calendar. If you do not sign-up on her calendar 12 hours in advance, the GSI cannot guarantee that she will be at office hours. If these appointment slots do not work for you, please email her.

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.) Topic Area</th>
<th>Meeting Dates</th>
<th>Assignments Due this Week</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Introductions /Linear Regression/ Stata Intro</td>
<td>Sept. 4, 7</td>
<td>Assign. 1</td>
</tr>
<tr>
<td>2) Linear Probability and Latent Variable Models</td>
<td>Sept. 11, 14</td>
<td>Assign. 2 &amp; paper proposal</td>
</tr>
<tr>
<td>3) Logistic Regression</td>
<td>Sept. 18, 21</td>
<td>Assign. 3</td>
</tr>
<tr>
<td>4) Logistic Regression (cont’d)</td>
<td>Sept. 25, 28</td>
<td>Assign. 4 &amp; lit. review</td>
</tr>
<tr>
<td>5) Probit Regression</td>
<td>Oct. 2, 5</td>
<td>_assign. 5, lit review and methods</td>
</tr>
<tr>
<td>6) Hypothesis Testing and Goodness of Fit</td>
<td>Oct. 9, 12</td>
<td>Review/critique of article</td>
</tr>
<tr>
<td>7) Out-of-Sample Prediction (No lecture – Fall Break)</td>
<td>Oct. 19</td>
<td>Assign. 6</td>
</tr>
<tr>
<td>8) Ordinal Outcomes</td>
<td>Oct. 23, 26</td>
<td>Assign. 7</td>
</tr>
<tr>
<td>9) Multinomial Outcomes</td>
<td>Oct. 30, Nov. 2</td>
<td>Lit review, methods, findings</td>
</tr>
<tr>
<td>10) Multinomial Outcomes (cont’d)</td>
<td>Nov. 6, 9</td>
<td></td>
</tr>
<tr>
<td>11) Conditional Logit</td>
<td>Nov. 13, 16</td>
<td></td>
</tr>
<tr>
<td>12) Count Outcomes (No lab - Thanksgiving)</td>
<td>Nov. 20</td>
<td></td>
</tr>
<tr>
<td>13) Interaction Effects</td>
<td>Nov. 27, 30</td>
<td></td>
</tr>
<tr>
<td>14) An Introduction to Causal Inference</td>
<td>Dec. 4, 7</td>
<td></td>
</tr>
<tr>
<td>15) In-Class presentations and turn in papers</td>
<td>Dec. 11</td>
<td></td>
</tr>
</tbody>
</table>
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, the weekly 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. An electronic copy of the assignment on the specified due dates must be submitted, usually at the start of lecture or lab. Note that due dates include one lab session in between to allow for discussion of any challenges you encountered in the design, coding, and/or interpretation of results.

c. When naming your assignments, please use the following naming convention:

[LastName]_Assignment[#].[fileextension]

Example: Oster_Assignment1.doc, Oster_Assignment2.pdf.

For the final project assignments, please use the naming conventions below. If you are working in a group, please list the last names alphabetically.

[LastName]_Step1.[fileextension]
[LastName]_Step2.[fileextension]
[LastName]_Step3.[fileextension]
[LastName]_Step4.[fileextension]
[LastName]_FinalPresentation.[fileextension]
[LastName]_FinalPaper.[fileextension]

d. Weekly assignments are 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.

e. 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 a 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 and the GSI).

i. Abstract/Proposal: Sept 25
ii. Literature review: Oct 19
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: October 30
iv. Revised literature review and methodology, and draft of findings: Nov 27
v. FINAL WEEK: Complete paper, including introduction and conclusion and a formal 15 minute presentation, including feedback from instructors and class: Dec 11. (This is a hard deadline, to which no exceptions will be made except in the case of prior arrangement with the instructor or a documented emergency.)

TEXTBOOKS:

Required:

The book is available at Amazon (link below). Additionally, many students have taken the course. It might be possible to get a used copy from them.

https://www.amazon.com/Regression-Categorical-Dependent-Quantitative-Techniques/dp/0803973748/ref=sr_1_1?ie=UTF8&qid=1533585307&sr=8-1&keywords=Regression+Models+for+Categorical+and+Limited+Dependent+Variables


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. It contains a wealth of practical information on coding in Stata, model estimation, post-testing, and so forth. It is available at Amazon (link below).


Additional Required Readings:
You will find additional weekly readings on Canvas. Look under Resources/Additional Readings tab.
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 available on Canvas. However, it is such a great resource for using 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. 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 at Amazon:

https://www.amazon.com/Maximum-Likelihood-Estimation-Quantitative-Applications/dp/0803941072/ref=sr_1_1?s=books&ie=UTF8&qid=1533585821&sr=1-1&keywords=Maximum+Likelihood+Estimation%3A+Logic+and+Practice

For an additional primer on categorical and limited dependent variable modeling, please see the cited handbook chapter below. It is on Canvas in the Resources/Additional Reading tab.


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.com/bookstore/gentle-introduction-to-stata/). However, many websites (e.g., https://stats.idre.ucla.edu) contain information about using Stata, including coding hints, sample programs, etc., so use your judgment about whether you need this additional book or not.

Also, Stata Press has a YouTube channel that offers helpful videos and short tutorials on how to use the software. The channel is available at: https://www.youtube.com/user/statacorp/featured

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 second floor of the School of Education, all have version 15.0 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: http://its.umich.edu/computing/computers-software/campus-computing-sites/virtual-sites. 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 to make conducting the analyses more convenient. Note that the most current version is Stata 15 and that different versions have different capabilities (see https://www.stata.com/products/which-stata-is-right-for-me/ 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: 1-800-782-8272 (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.

**APA STYLE:**
For all assignments (including lab assignments and the final project), use APA style unless you come from a discipline that uses a different citation style. If that is the case, please notify Steve and the GSI early in the semester. If you need help with APA style, please see the links below:

http://www.apastyle.org/manual/

https://owl.purdue.edu/owl/research_and_citation/apa_style/apa_formatting_and_style_guide/general_format.html

**ON DIVERSITY:**
I intend to present materials and activities that are respectful of our diverse world and multiple opinions and points of view. I encourage you to provide suggestions on how we can incorporate new materials to improve the course for all students. The University of Michigan has a number of affirmative action policies and these policies can be found at: 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).

**GSI EMAIL POLICY:**
In general, the GSI will strive to answer emails as quickly as possible during both the week and weekend. During the week, the GSI will return emails within 24 hours. During the weekend, the GSI will continue to strive for a 24-hour response time; however, it is possible that her response may be delayed to the following Monday. If multiple students have similar questions, an announcement (via CANVAS) will be sent to the entire class.

**ACCOMMODATIONS FOR SPECIAL NEEDS:**
I am happy to discuss with you any accommodations for special needs that will help you succeed in this course. Please approach me early in the semester and see http://www.umich.edu/~sswd/ for more information about the University’s services for students with disabilities.
Course Schedule

WEEK 1 (September 4 & 7)
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.

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/spost13.htm).

Also, please familiarize yourself with the Canvas site for this class (named “EDUC 799 001 F18”) and peruse the materials that are there as you will need to interface with this site on a regular basis. Please fill out the pre-class survey sent out by the GSI if you haven’t already done so.

If you are new to Stata or need a refresher, focus on the files in the Stata Materials folder—especially “Introduction to Stata with 50 Basic Commands,” “Getting Started with Stata 15,” “Stata Command Reference,” or this URL https://www.stata.com/features/overview/quick-starts/

DUE: Nothing.

WEEK 2 (September 11 & 14)
THE LINEAR PROBABILITY MODEL & LATENT VARIABLE MODEL FOR BINARY DEPENDENT VARIABLES

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


RECOMMENDED READING:

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

REQUIRED READINGS:
Long, Chapter 3, pp. 50-83.
Long & Freese, Chapter 5

RECOMMENDED READING:

DUE: Tuesday, September 18 (by start of lecture) - Assignment 1.

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

REQUIRED READINGS:
Long & Freese, Chapter 6

RECOMMENDED READINGS:

SUPPLEMENTAL READINGS:

DUE:
- Tuesday, September 25 (by start of lecture) - Assignment 2
- Tuesday, September 25 (by start of lecture) - 2 paragraph proposal/abstract for final paper

REQUIRED READINGS:
Reread Long Chapter 3.
Reread Long & Freese Chapter 5


DUE: Nothing. Work on literature review.
WEEK 6 (October 9 & 12)
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 & 5.


DUE: Tuesday, October 9 (by start of lecture) - Assignment 3

WEEK 7
(No lecture October 16 due to Fall Break; lab will meet on October 19)
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:
• Friday, October 19 (by start of lab) - Draft of literature review for final project
• Friday, October 19 (by start of lab) - Assignment 4
WEEK 8 (October 23 & 26)
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 7.


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


An example of the use of ordered probit

DUE: Nothing. Work on literature review and methodology.

WEEK 9 (October 30 & November 2)
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 8 through p. 444.

DUE:

• Tuesday, October 30 (by start of lecture) - Revised literature review & draft of methodology
• Tuesday, October 30 (by start of lecture) - Assignment 5
WEEK 10 (November 6 & 9)  
MULTINOMIAL LOGISTIC REGRESSION, cont’d

REQUIRED READINGS: Read all three examples. Write a 1-2 page in-depth review/critique (unless you have an interest in a substantive area, readings will be randomly assigned).


DUE: Review/critique of one of the above articles

WEEK 11 (November 13 & 16)  
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 8, pp. 454-464


SUPPLEMENTAL READINGS:

DUE: Tuesday, November 13 (by start of lecture) - Assignment 6
WEEK 12 (November 20. No lab on Nov. 23 because of Thanksgiving)
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: Tuesday, November 20 (by start of lecture) - Assignment 7

WEEK 13 (November 27 & 30)
INTERACTION EFFECTS

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

DUE: Tuesday, November 27 (by start of lecture) - Literature review, revised methodology, draft findings.
When analyzing observational data, many social scientists have used categorical dependent variable regression models to make inferences about cause and effect. For instance, scholars have estimated a logistic model to analyze whether receiving various types of financial aid causes students to persist to the second year of college. Doing so, however, can be problematic when the independent variable of interest is not randomly allocated (as is the case with financial aid). This week’s readings provide an overview of causal inference and an introduction to quasi-experimental methods that employ non-linear models as a first step in a two-stage process for producing estimates that are not tainted by self-selection bias.

REQUIRED READINGS:

Pay particularly close attention to the sections on the counterfactual framework (p. 195-196) and propensity score matching (p. 200-202).


RECOMMENDED READING:


ASSIGNMENT: Continue to work on your final papers.

Lab: Workshop time for final papers

WEEK 15 (December 11 & Possibly December 14th)
IN CLASS PRESENTATIONS DURING LECTURE & POSSIBLY A MAKE-UP Lab