Intermediate Quantitative Analysis

SOCI 498/898
Spring (Jan 11, 2016–May 6, 2016)
Tuesday, Thursday 3:30 PM – 4:45 PM
Oldfather Hall 707
University of Nebraska-Lincoln

Professor: Dr. Jeffrey A. Smith
Office: 706 Oldfather Hall
Email: jsmith77@unl.edu
Office Hours: Tuesday, Thursday 11:00 – 12:00 AM or by appointment

Course Description
This is a graduate level course covering statistical models for limited dependent variables and their application to substantive questions in the social sciences. We will cover outcomes that are binary (smoke; not smoke), ordinal (Disagree Strongly; Disagree; Neutral; Agree; Strongly Agree), nominal (Married; Divorced; Never Married; Widowed) and count (number of children). Many of the variables that sociologists face will have one of these forms (or something similar). Unfortunately, when we have limited dependent variables, our standard OLS model is inappropriate, and can lead to incorrect inference and interpretation. Worse yet, we run the risk of embarrassment when presenting at conferences or submitting to journals if we apply the wrong model to our data. This class will introduce a set of methods that are appropriate for modeling outcomes that are not continuous. Note that this class is designed to be practical. The mathematical treatment will be on the light side and we will focus on proper specification, interpretation and presentation.

Overall, there are three main objectives:
a) gain experience with methods appropriate for limited dependent variables
b) learn how to analyze and interpret limited dependent variables using Stata
c) apply models for limited dependent variables to a substantive case and write-up an original research project based on the results.

Prerequisites
I am assuming a working of knowledge of OLS but we will do a short review (just in case we forget something…). The class will make heavy use of Stata. I am assuming that you have had some experience with *a* statistical programming language (e.g., SAS), although I make no assumption that you have any experience with Stata.

Text and Software

Textbook: Long, Scott and Jeremy Freese. *Regression Models for Categorical Dependent Variables Using Stata*, Third Edition; ISBN-13: 978-1-59718-111-2

Software: This class will use Stata. We will walk through Long and Freese, which offers detailed examples using Stata. Stata is available on the sociology cluster. You may use an

alternative program (SAS, R, etc.) but the course will be conduced in Stata and you are on your own in terms of completing the homework.

Format of the class
The class will be a mix of lectures, discussion and labs. It is important that you read the material prior to class. You should come prepared with questions about things you are confused about. The hope is that you will understand *some* of the material beforehand, and then we can work through the more difficult parts together. The more you ask questions, the more likely you will be successful in this class. You are also expected to be an active participant during the labs, where we will learn how to apply the ideas presented in class to actual data. The labs will be held in our normal classroom, unless folks do not have laptops, in which case we will have to venture over to the computer lab. Note that the lectures are likely to bleed into the labs, as this is a very hands-on course, and we will learn by doing things in Stata and interpreting the results. For each major topic we will cover the following: a) the basic model and its rationale (When should we use the model? Why is it appropriate? What are the assumptions of the model?); b) how to interpret the results of the model (What do the parameters mean? How can we graphically depict the results? How can we get predicted values from the model?); c) hypothesis testing; d) diagnostic procedures (How can test if the assumptions of the model are met? How do we know if there are model specification problems?).

Finally, some of our time in class will be spent discussing your projects. One day in February will be spent describing your project to the rest of the class. You should be prepared to describe the topic, data, and models you are tentatively expecting to use. We will also spend the end of the class doing presentations. Each student is expected to give a 15 minute presentation describing their project.

Readings
The main readings for each week will be a chapter in the Long and Freese book. I will also occasionally include additional readings to supplement the book. We will also go over substantive papers throughout the semester. The papers will be selected by the discussion leader for that day (see details below).

Grading
Participation: 15%
Participation is a key requirement of this class. You need to come to class prepared, interested, and ready to discuss the material at-hand. You will be graded on general participation (did you come to class, engage in lab, ask questions, etc.) as well more specific days where you must take a more active role in class. This takes two forms:

Discussion Leader
For each major topic in the class (e.g., logistic regression, ordinal regression, etc.) we will discuss empirical papers that employ the model in question. This will give us an opportunity to see how these models were used in published papers, examining exemplars of analysis and interpretation. We will discuss two or three empirical papers per major topic. Each student in the class will act as a discussion leader for one empirical

paper. The discussion leader(s) will choose for the class the paper they would like to discuss. The paper **must** employ the model/variable type that we are discussing that day. The discussion leader will choose the paper of interest and send it to everyone to read for that day (at least a day before class). Students will have their day assigned next class and this cannot be changed later on in the semester unless there are extenuating circumstances. Each discussion leader should prepare for a 10 minute discussion about the paper, where the discussion leader should:

a) summarize the main problem tackled by the readings
b) summarize the main argument about the problem
c) describe the data employed
d) describe the model employed
e) examine the interpretation and presentation of the results (did they do a good job?; would you have done anything different?)
f) examine the assumptions made by the author
g) facilitate discussion about *how* the author went from an initial research question to an actual test to interpretation of results.

Class Presentation
Each student will give a final presentation of their project at the end of class. Each student is expected to give a formal 15 minute presentation (plus answering questions). This will give an opportunity to present an early version of your work in a forgiving environment. Think of this as good practice for the real thing. Students will also report on their progress in class on March 1. You should expect to talk informally for 5 minutes.

Homework: 25%
We will have 8 homeworks assigned during the class. The homework is designed to get you practice at working through Stata to answer substantive questions. You will have one week to complete each assignment. You will be graded based on completeness and general accuracy, but I will not grade every homework closely. The point is to get you *practice* at using Stata and interpreting results. Nonetheless, homeworks that are incomplete and/or show a lack of understanding will get lower marks.

Project: 50% for actual research paper; 10% for initial submission of proposal and pieces of project

The main assignment in the course is a research paper, in which you perform an empirical analysis on real data and write up the results. You must employ (at least) one of the methods from this course in your paper. To be clear, it is insufficient to rehash and update a paper from another course that utilizes OLS to analyze a continuous variable. Your outcome of interest must be binary, nominal, count, ordinal, etc. The research paper must include: an introduction with a substantive/theoretical justification for the project; a short literature review; a description of the data and models; tables/figures describing the results; a summary of results; and a conclusion. The hope is that this will be a start to a publishable paper. The paper is due during the scheduled final for the class.

I leave it you to decide what question you want to answer, what data you would like to use, etc., but you must get my approval for the project before you begin. I can also help you find data and a research question if you are struggling on your own. We will discuss the researcher paper in more detail during class. One strong constraint you will face during this class is time. It is not a good idea to leave the project to the very last minute, hoping to get it all done in a mad rush. You should be working on the project throughout the entire semester. This means that you cannot pick a dataset that will only be available late in the semester (or even worse, after the semester is over). You need to plan ahead and make sure the data is available early enough to make the project feasible. With this in mind, you should have a usable dataset by March 1. You need to pick a different project if that is not possible.

You will be required to turn in initial results and pieces of the project as the semester moves along. This is designed to ensure that you are making sufficient progress. For example, an initial proposal is due on April 5. This is a short (5-7 pages) description of your proposed project; it will include a summary of the question, an outline of the theoretical argument, and a brief description of the data.

Key Dates for Project

Feb 9: Deadline for getting my approval for your proposed research project. This includes the core research question and the data you will use. After this date, I will take 1% off your final grade everyday until you discuss your proposed project with me.

Mar 1: In-Class Progress Report. Be prepared to talk for 5 minutes about your project, including basic research question, theoretical motivation, data sources and variables of interest.

March 15: Descriptive statistics on key variables due

April 5: Initial 5-7 page proposal due. This should include a summary of the question, basic theoretical arguments and initial literature review, and description of the data.

April 12: Tables and figures showing results (must use a model discussed in this class)

April 21, 26 or 28: In-Class Presentation

May 2 at 10:00 p.m.: Final Paper Due

Course policies

Academic Misconduct (or Don't Cheat):

"The maintenance of academic honesty and integrity is a vital concern of the University community. Any student found guilty of academic dishonesty shall be subject to both academic and disciplinary sanctions. Academic dishonesty includes, but is not limited to, the following: Cheating; Fabrication or Falsification; Plagiarism; Abuse of Academic

Materials; Complicity in Academic Dishonesty; Falsifying Grade Reports; Misrepresentation to Avoid Academic Work."
Quoted from the UNL Student Code of Conduct

Disabilities
"It is the policy of the University of Nebraska-Lincoln to provide flexible and individualized accommodation to students with documented disabilities that may affect their ability to fully participate in course activities or to meet course requirements. To receive accommodation services, students must be registered with the Services for Students with Disabilities (SSD) office, 132 Canfield Administration, 472-3787 voice or TTY."

If you need accommodations it is your responsibility to discuss this with me early on in the semester.

Paper Policy
The final paper must be turned in by the end of the official exam period for the class, stipulated by the university (May 2th at 10:00 pm). There will be no papers accepted after the fact unless there are extraordinary circumstances and the student has received permission from me to turn in the exam late.

Course Schedule (subject to change)

Day 1: Introduction to the Class (Jan 12)

Day 2: OLS Review (Jan 14)
Required Readings
1. Chapter 7 in OpenIntro Textbook: Introduction to Linear Regression and Multiple Regression
2. Part of Chapter 8 in OpenIntro Textbook: Multiple Regression

Day 3: Stata Introduction (Jan 19)
Required Readings
1. Chapter 1-2 in Long and Freese

Day 4: Stata Introduction (Jan 21)
Required Readings
1. Chapter 1-2 in Long and Freese

Day 5: OLS in Stata (Jan 26)
*Assignment: Discussion Leader #1 for OLS models
Required Readings

1. Selected Article

Day 6: Venturing in: Introduction to MLE and GLM (Jan 28)
*Assignment: HW 1 due
Required Readings
1. Chapter 3 in Long and Freese
2. Myung, In Jae. "Tutorial on maximum likelihood estimation." *Journal of mathematical Psychology* 47.1 (2003): 90-100.

Day 7: Binary Outcomes (Feb 2)
Required Readings
1. Chapter 5 in Long and Freese
2. Sperandei, Sandro. "Understanding logistic regression analysis." *Biochemia medica* 24.1 (2014): 12-18.

Day 8: Binary Outcomes Continued (Feb 4)
*Assignment: HW 2 due
Required Readings
1. Chapter 5 in Long and Freese
2. Morgan, S. Philip, and Jay D. Teachman. "Logistic regression: Description, examples, and comparisons." *Journal of Marriage and the Family* (1988): 929-936.

Day 9: Binary Outcomes Continued (Feb 9)
*Assignment: Discussion Leader #1 for binary outcomes
Required Readings
1. Chapter 4 in Long and Freese
2. Selected Article

Day 10: Binary Outcomes Continued (Feb 11)
*Assignment: HW 3 due and Discussion Leader #2 for binary outcomes
Required Readings
1. Chapter 6 in Long and Freese
2. Selected Article

Day 11: Binary Outcomes: Prediction and Presentation (Feb 16)
*Assignment: Discussion Leader #3 for binary outcomes
Required Readings
1. Chapter 6 in Long and Freese
2. Selected Article

Day 12: Binary Outcomes: Prediction and Presentation (Feb 18)
*Assignment: HW 4 due and Discussion Leader #4 for binary outcomes
Required Readings
1. Selected Article

Day 13: Binary Outcomes: Advanced Topics (Feb 23)
Required Readings

1. Allison, Paul D. "Comparing logit and probit coefficients across groups." *Sociological Methods & Research* 28.2 (1999): 186-208.
2. Mood, Carina. "Logistic regression: Why we cannot do what we think we can do, and what we can do about it." *European Sociological Review* 26.1 (2010): 67-82.

Day 14: Binary Outcomes: Advanced Topics (Feb 25)
*Assignment: HW 5 due
Required Readings

1. Williams, Richard. "Fitting heterogeneous choice models with oglm." *Stata Journal* 10.4 (2010): 540.
2. Williams, Richard. "Using heterogeneous choice models to compare logit and probit coefficients across groups." *Sociological Methods & Research* 37.4 (2009): 531-559.

Day 15: In Class Progress Report (Mar 1)

Day 16: Ordinal Outcomes (Mar 3)
Required Readings

1. Chapter 7 in Long and Freese

Day 17: Ordinal Outcomes (Mar 8)
Required Readings

1. Chapter 7 in Long and Freese

Day 18: Ordinal Outcomes (Mar 10)
*Assignment: Discussion Leader #1 for ordinal outcomes
Required Readings

1. Selected Article

Day 19: Ordinal Outcomes (Mar 15)
*Assignment: Discussion Leader #2 for ordinal outcomes
Required Readings

1. Selected Article

Day 20: Nominal Outcomes (Mar 17)
*Assignment: HW 6 due
Required Readings

1. Chapter 8 in Long and Freese

Day 21: Spring Break (Mar 22)

Day 22: Spring Break (Mar 24)

Day 23: Nominal Outcomes (Mar 29)
Required Readings
1. Chapter 8 in Long and Freese

Day 24: Nominal Outcomes (Mar 31)
*Assignment: Discussion Leader #1 for nominal outcomes
Required Readings
1. Selected Article

Day 25: Nominal Outcomes (Apr 5)
*Assignment: Discussion Leader #2 for nominal outcomes
Required Readings
1. Selected Article

Day 26: Catch-up On Research Projects (Apr 7)

Day 27: Count Models (Apr 12)
*Assignment: HW 7 due
Required Readings
1. Chapter 9 in Long and Freese

Day 28: Count Models (Apr 14)
*Assignment: Discussion Leader #1 for count outcomes
Required Readings
1. Selected Article

Day 29: Count Models (Apr 19)
*Assignment: HW 8 due and Discussion Leader #2 for count outcomes
Required Readings
1. Selected Article

Day 30: In Class Presentations (Apr 21)

Day 31: In Class Presentations (Apr 26)

Day 32: In Class Presentations (Apr 28)

May 2: **Final Paper due**