LPO 3460: Regression Analysis (Spring 2006)

Monday/Wednesday, 11:10-12:25 in Peabody Library Learning Commons (Room 304)

Instructor:
Tom Smith
Office 207D Payne Hall
Telephone: 322-5519
Email: tom.smith@vanderbilt.edu
Office hours: Monday, 3-5 pm or by appointment

Teaching Assistant:
Natasha Rumyantseva
Telephone: 668-7259
Email: nataliya.l.rumyantseva@vanderbilt.edu
Office hours: Friday, 10-12 pm in the Peabody Library Learning Commons or by appointment

Course Description

Statistical analysis can be a powerful tool for identifying social, educational, economic, psychological, and developmental relationships. As it is rare that we can collect data on every individual, classroom, teacher, and school of interest, statistical analysis allows us to examine data on a sample of individuals (or classrooms, schools, etc.) in order to infer patterns in a larger population. For example, we might want to examine data on achievement test scores and per-pupil spending for a sample of schools to determine whether there is an association between spending and achievement patterns in the population. Or we might want to examine the association between race/ethnicity and achievement. Moreover, if we find such an association, we might wish to ask additional questions, such as whether race/ethnic differences in achievement patterns can be accounted for by race/ethnic differences in family socioeconomic characteristics or in school or teacher quality.

In this course we will learn to answer such questions using regression analysis—a statistical tool that allows us 1) to describe average patterns of association among multiple variables observed in a sample and 2) to make inferences about the patterns of association among these variables in a population. Regression analysis is a powerful statistical method with many variations. Our goal in this course is to develop an understanding of the basic methods, including their limitations, and to develop skill in using regression analysis to answer educational research questions. Finally, because an important part of any analysis is communicating the results to an audience, we will also place considerable emphasis on learning to present (in writing, tables, and figures) the results of regression analyses.

By the end of the semester, students in this course should be sufficiently skilled in regression analyses that they can critically examine published research using regression, can carefully perform their own analyses, and write about what they have found.

This is not a math course, although the ability to describe relationships mathematically is an important part of quantitative research. As such, we will use math in the service of defining statistical models that correspond to the research questions we want to answer, and we will pay less attention to the mathematics by which computer programs estimate the parameters in our models. Our concern is with understanding how to use these methods to do good research and with learning to interpret the results they provide.

The course will put a lot of emphasis on conducting statistical analyses using sample data and on communicating the results of these analyses. The software program that we will use for statistical

analyses is called STATA (version 9), a general statistics software package similar in capacity to SPSS and SAS. The UCLA Academic Technology Services have excellent on-line resources for learning how to do things in STATA. See

http://www.ats.ucla.edu/stat/stata/

I have also ordered copies of *Statistics with Stata* by Lawrence Hamilton at the Vanderbilt Bookstore. This is an excellent resource for getting started using STATA. This text is not required, although you may find it useful to read and share between you. I will also put a copy of this text at the reserve desk in Peabody Library.

Personally, I find STATA easier to use than SAS and more powerful for the kinds of analyses that I typically want to do than SPSS. STATA is installed on all of the computers in the Learning Commons of the Peabody Library and should be accessible during regular library hours for homework exercises:

Sunday:	11 am – 2 am
Monday - Thursday:	7:30 am – 2 am
Friday:	7:30 am – 6 pm
Saturday:	10 am – 6 pm

If you prefer to have STATA installed on your own computer, you can purchase a 1 year license for $89 and a perpetual license for $145. You would order this directly from STATA within Vanderbilt's "Gradplan" program (you'll want Intercooled Stata 9) and subsequently pick up the installation CD and documentation from ITS (call 343-9999). See

http://www.stata.com/order/new/edu/gradplans/gp-campus.html

Prerequisites

This course is an introductory/intermediate level applied statistics course, designed for students who have completed an introductory statistics course, such as PSY 309 or SOC 310. To be prepared for taking this course you should be familiar with basic statistical concepts such as populations, sampling, means, variances, estimates, and statistical inference. A prior familiarity with simple (bivariate) regression is helpful, but we will cover that again in this course. The course relies heavily on algebra; no more sophisticated mathematics are necessary for our purposes.

Text

We will use *Introductory Econometrics: A Modern Approach* by Jeffrey M. Wooldridge (3rd Edition). ISBN 0324289782. Econometrics is just the development of statistical methods for estimating economic relationships and testing economic theories. As economists rarely have experimental data available to test their hypotheses, the tools they have developed are particularly useful for analysis of administrative data, survey data, or data collected with a quasi-experimental design. While the text includes some examples of analyses related to educational policy issues, we will supplement these with in-class examples.

Course Requirements and Grading

Course grades will be based on your performance on assigned problems and exercises (25%), an examination (25%), and a data analysis project and presentation (50%). Students are expected to attend all classes. If you miss a class you are responsible for making up the work. The class may be relatively large, but I strongly urge you to ask questions in class. If you have a question, it is likely that others do as well.

Problem sets and data analysis exercises

Problems will be assigned from the exercises at the end of each chapter in the text. Additional computer exercises may also be assigned as homework. Homework is due to Natasha (by email or in her box in Graduate Student Space in the basement of Payne Hall) by 5pm on Friday of the week that it is due. As a rule, I will not accept or review late assignments, but if you have extenuating circumstance I will consider them. No credit will be given for assignments received after we have gone over them in class (most often the Monday following when they were due). It is acceptable (and encouraged) that you work together on problem sets and data analysis exercises. Answers to the problem sets, however, should be written entirely on your own and in your own words. Do not submit assignments that duplicate word-for-word or paraphrase another student's assignment.

Exam

There will be an in-class examination on March 1st. It will cover Chapters 1-7 of the Wooldridge text.

Data Analysis Project and End of Term Presentation

Students will be expected to write and present a seminar paper in which they:

1. identify a research problem that can be studied using regression analysis,
2. review the relevant literature and discus what is known and not know about the problem,
3. select an existing dataset that can be used to study the problem,
4. construct appropriate variables to use in the analysis,
5. develop appropriate models to test hypotheses generated from the literature review,
6. test the models using STATA, and
7. describe and interpret the results.

The data analysis project should be of a quality that would warrant presentation and distribution at a professional conference (AERA, ASHE, ASA, AEFA, etc.). You are encouraged to explore more deeply one of the substantive topics that you have covered or are currently covering in another class, although your literature review should not duplicate an assignment that you will turn in for another professor. In addition to the paper that you will turn in, students will make a 10 minute conference-quality presentation of their project at the end of the semester. An overhead projector and/or PowerPoint projector will be available for this purpose.

Before undertaking data analysis for this project, students shall complete a literature review and write up and turn in for approval a project proposal that addresses the first five questions above. I would be happy to meet with you to discuss your project. I also have a number of large datasets available for you to use (National Educational Longitudinal Study of 1988; Third International Mathematics and Science Study, Schools and Staffing Survey, Early Childhood Longitudinal Study, as well as others). This "project proposal" will be due February 10th and will be worth 1/3 of your project grade. It should be 8-10 pages double spaced and should describe in detail the data you plan to use, as well the hypotheses you plan to test (including the equations that you plan to model). The final project paper is due Monday, May 1st. If you are not knee deep in your analyses by Spring Break, you will probably run out of time, do a "rush job", and accept a lower grade in the course. This paper should be 16-20 double-spaced pages, *plus* tables, figures, and references. You will also turn in the STATA Commands ("do file") that you used to restrict your sample, create analysis variables, run descriptive statistics, run regression models, etc.

Draft Schedule and Homework Assignments

Date	Topic	Reading	Homework (Due 5pm Friday of the week assigned)
01/11/06	Introduction	W Ch. 1	1.1, 1.3, C1.3, C1.4
01/16/06	Simple Regression	W Ch. 2, pp. 24-44	
01/18/06	Functional Form & OLS Assumptions	W Ch. 2, pp. 44-65	2.1,2.3,2.4,2.6,2.11,C2.2, C2.4,C2.6
01/23/06	Multiple Regression Analysis: Estimation	W Ch. 3, pp. 73-94	
01/25/06	Omitted Variable Bias and the Variance of OLS Estimators	W Ch. 3, pp. 94-111	3.1,3.2,3.3,3.4,3.5,3.7,3.8, 3.10,3.11,C3.1,C3.4,C3.5, C3.7
01/30/06	Multiple Regression Analysis: Inference	W Ch. 4, pp. 123-147	
02/01/06	Testing Hypotheses about Combinations of Parameters	W Ch. 4, pp. 147-167	4.1,4.2,4.4,4.5,4.6,4.9,C4.1, C4.2,C4.5,C4.6,C4.7
02/06/06	Multiple Regression Analysis: OLS Asymptotics (Consistency)	W Ch. 5, pp. 176-185	5.3,C5.1,C5.2
02/08/06	Conducting an Empirical Research Project	W Ch. 19, pp 678-705	**Term Project Proposal Due 2/10**
02/13/06	Multiple Regression Analysis: Data scaling, beta coefficients, functional form (Meet in Peabody Library Room 303)	W Ch. 6, pp. 192-206	
02/15/06	Goodness of Fit and Residual Analysis	W Ch. 6, pp. 206-222	6.1,6.4,6.5,6.6,6.8,C6.2, C6.3,C6.4,C6.7,6.9
02/20/06	Multiple Regression Analysis with Qualitative Information: Binary or Dummy Variables	W Ch. 7, pp. 230-249	
02/22/06	Testing for Differences in Regression Functions Across Groups	W Ch. 7, pp. 249-261	7.1,7.2,7.3,7.5,7.8,C7.1,C7.2, C7.3,C7.4,C7.6,C7.12
02/27/06	Data Coding & Exam Review	Handouts	
03/01/06	**EXAM**		
03/06/06	Spring Break		
03/08/06	Spring Break		
03/13/06	Heteroskedasticity & Robust Standard Errors	W Ch. 8, pp. 271-284	
03/15/06	Weighted Least Squares Estimation	W Ch. 8, pp. 284-298	8.1,8.2,8.3,8.4,C8.1, C8.2,C8.8
03/20/06	Limited Dependent Variable Models--Logit & Probit	W Ch. 17, pp 582-595 & Handouts	17.2,17.7,C17.1
03/22/06	Additional Topic TBA		Additional data analysis exercise
03/27/06	Specification and Data Problems	W Ch. 9, pp. 304-318	
03/29/06	Properties of OLS under Measurement Error	W Ch. 9, pp. 318-334	9.3,9.4,9.5,C9.2,C9.4, C9.6,C9.8
04/03/06	Introduction to Panel Data Methods	W Ch. 13, pp. 448-467	
04/05/06	Fixed effects models	W Ch. 13, pp. 467-476	13.1,13.2,13.3,13.6,13.7, C13.2,C13.5,C13.6,C13.7, C13.11
04/10/06	AERA: Optional Lab, NO CLASS		
04/12/06	Term Project Presentations		
04/17/06	Term Project Presentations		
04/19/06	Term Project Presentations		
04/24/06	Term Project Presentations		
05/01/06	**Term Project Paper Due**		

W=Wooldridge Text