

**MULTIVARIATE ANALYSIS II**  
**SOCIOLOGY 312**  
**VANDERBILT UNIVERSITY, SPRING 2006**

**Professor:** Michael E. Ezell  
**Office:** 312 Garland Hall  
**Office Hours:** M 1-2, W 11-12, & by appointment  
**Office Phone #:** 322-7536  
**Email:** mike.ezell@vanderbilt.edu

**TA:** Emily Tanner-Smith  
**Office:** 321F Garland Hall  
**Office Hours:** By appt.  
**Office Phone #:** 322-0187 (email is better to set up a time to meet)  
**Email:** emily.e.tanner-smith@vanderbilt.edu

**COURSE OVERVIEW**

This is a graduate course continuing an examination of the statistical methods commonly employed in the social sciences. This class, coupled with (its first-semester sibling class) Sociology 311, focuses on the application of multivariate quantitative methods, with a particular emphasis on alternative specifications of the regression model.

Sociology 312 has two main sections. The first section will continue the systematic exposition of the linear regression model involving a single, continuous dependent variable (i.e., the ordinary least squares regression model) that we began in Sociology 311. The second section of Sociology 312 covers the *generalized linear model* (e.g., logistic regression, multinomial logistic regression, Poisson regression) that serves as the backbone of most contemporary multivariate research in the social sciences today.

Like Sociology 311, this course will involve both a practical, applied focus and a theoretical focus. Thus, there will be emphasis on both theory and substance, and we will put particular emphasis on presenting the estimates from the generalized linear model such that the results are substantively interpretable.

**PREREQUISITES**

- Enrollment in Sociology 312 **requires** successful completion of Sociology 311 (grade B or higher) or a comparable graduate-level introductory statistics course.

**TEXTBOOKS AND REQUIRED READINGS**

**Required Textbooks**

Moore and McCabe. (2006). *Introduction to the Practice of Statistics, 5<sup>th</sup> Edition*. W. H. Freeman and Company: New York. [You should already have this from Sociology 311]

Hoffmann (2004). *Generalized Linear Models: An Applied Approach*. Allyn & Bacon: Boston.

## **Recommended Textbooks:**

I highly recommend:

Long and Freese. (2003). *Regression Models for Categorical Dependent Variables Using Stata, Revised Edition*. Stata Press: College Station, TX.

For new participants that were not in Sociology 311 last semester, I recommend these to get you up and running with Stata:

Lawrence Hamilton (2004). *Statistics with Stata 8*. Duxbury: Pacific Grove, CA. [Stata 8]

Lawrence Hamilton (2006). *Statistics with Stata 9*. Duxbury: Pacific Grove, CA. [Stata 9]

Alan Acock (2006). *A Gentle Introduction to Stata*. Stata Press: College Station, TX.

## **Required Readings**

On the class schedule below, “Required Readings” denote the readings that you are expected to have read *prior* to the beginning of class for that day.

## **Substantive Readings**

On relevant days, I will also be providing citations for substantive (not methodological) articles that employ the use of the methods discussed in class on that day. These substantive articles are highly recommended and will help you in the transition from the learning the method to employing the method in substantive application.

## **Recommended Readings**

In addition to the assigned readings, I have also listed other “Recommended Readings” that further discuss relevant topics. These are not required readings, but are expositions that are relevant for more in-depth coverage of the material should you be interested in further information.

## **COURSE REQUIREMENTS**

Your course grade will be determined on a grand total of 1,000 points.

	<u>%</u>	<u>Points</u>
Problem Sets (5 @ 9%):	45%	450
Brief Final Exam:	15%	150
Final Paper:	40%	400

### **Problem Sets (45%)**

Participants will complete five problem sets during the semester. The problem sets will involve questions focusing on both your substantive understanding of the topics and your ability to complete specified data analysis tasks.

You will need to turn in the both the “.do” Stata syntax file and the corresponding “.log” output file with each problem set.

**Late problem sets will not be accepted** unless you have received permission from the professor in advance of the assignment due date. The datasets for the problem sets will be available via the web as Stata is “web aware” and can directly load the data sets into Stata over the web using the command:

use [http://people.vanderbilt.edu/~mike.ezell/courses/soc312/data/filename](http://people.vanderbilt.edu/~mike.ezell/courses/soc312/data/<u>filename</u>)

### Final Paper (40%)

Each student in this class will complete a paper employing the use of one of the methods covered in this class (OLS or a variant of the generalized linear model) on an empirical topic of your choice (hopefully one in which your research agenda is situated). The paper will be graded based on the quality, sophistication, and rigorousness in applying the statistical models and techniques taught in this course. The paper is expected to be 15-30 pages in length, and will be written in an academic journal style and the format should model articles in the top sociological journals (e.g., *American Sociological Review*, *Journal of American Sociology*). There are several stages in the process of developing this research paper:

- Feb. 3: Brief proposal for the research paper is due.
- Feb 25: Full proposal for the research paper is due.
- April 21: Paper is due

### Final Exam (15%)

There will a brief “final exam” in this course that will make sure you understand the main concepts from the course. This will be a short (30 minute or less) exam.

## COURSE POLICIES

1. Assignments will **not** be **accepted via e-mail**. If you are unable to print your document due to printer-related problems, you may email me your assignment so that I know it was completed on time (e.g., before class started). I will not download your paper and print it out for you, and you must provide me with a hard copy of the assignment that same day. You must print out your assignment immediately after class and turn it in then. That version must be identical to the one you submitted via email.

## HONOR SYSTEM ISSUES

The Vanderbilt Honor System states “all work submitted as part of course requirements is produced by the student, without help from any other source unless acknowledgement is given in a manner prescribed by the instructor. Cheating, plagiarizing, or otherwise falsifying results of study are specifically prohibited.”

In this class, I encourage you to work with each other to figure out both the logic of statistical reasoning and the various statistical methods covered in this class, as well as helping each other figure out how to operate Stata. This is how much professional social science research is accomplished--with the help and aid of other researchers. However, you must **not** simply be copying other people’s work or statistical programs. Rather you should be colleagues that actively contribute in the process of learning. Working together to figure something out is vastly different from one student copying another student’s work. **Do not copy** each other’s work or programs. A group solution to the assignment sets is not acceptable, and submission of identical software code is not acceptable as well. In other words, you must do your own work, but you may collaborate and help each other (not free load off one person) with the various problem sets in this class. Please let me know if we need to discuss this further.

## STATISTICAL SOFTWARE

The statistical software package “Stata” (Version 8 or 9) is the official software package of this course. This software package will be used frequently in class throughout this course. The professor and the teaching assistant in this course will support the use of this package only. Although we have experience with other packages (SPSS, R, S-PLUS, SAS, LIMDEP), we will not be able to effectively support the use of all other

packages. If you have a deep commitment to one of these other programs (and are very knowledgeable regarding the use of another package) and wish to use them in this course, please discuss this issue with the professor.

### COURSE OUTLINE

Below is a tentative outline of the topics to be discussed in class each day and the dates that each of the problem sets will be handed out. This schedule is more or less firm, but if I have inaccurately projected the time required for each topic of discussion, I reserve the right to make modifications to the topics to be discussed and/or the dates of assignments/exam. The adjustments will be made on an as needed basis.

Date	Topic
<b>Week 1</b>	
January 15 <sup>th</sup>	<b>Getting Back Up To Speed: Matrix Algebra &amp; The Multiple Regression Model</b>  <b>Reading:</b> Appendix B of Kleinbaum <i>et al.</i>  Chapter 1 (Pp. 1-17) of Hoffman  Chapter 12 of Moore and McCabe
<b>Week 2</b>	
January 20 <sup>th</sup>	<b>Confounding &amp; Interaction Effects</b>  <b>Readings:</b> Chapter 1 (Pp. 17-20) of Hoffman  Chapter 11 of Kleinbaum <i>et al.</i>  Marsden, Peter V. (1981). "Conditional Effects In Regression Models." Pp. 97-116 in <i>Linear Models in Social Research</i> . Peter V. Marsden (ed). Sage.  <b>Recommended Reading:</b>  Allison, Paul D. (1978). "Testing For Interaction Effects In Multiple Regression." <i>American Journal of Sociology</i> 83: 144-153. Sharpe, N. Radke, and R. A. Roberts. (1997). "The Relationship Among Sums of Squares, Correlation Coefficients, and Suppression." <i>The American Statistician</i> 51: 46-48.  <b>Substantive Reading:</b> Baron, Stephen W. 2004. "General Strain, Street Youth, and Crime: A Test of Agnew's Revised Theory." <i>Criminology</i> 42: 457 - 483. [Read First] LaGrange, Teresa C., and Robert A. Silverman. 1999. "Low Self-Control and Opportunity: Testing The General Theory of Crime as an Explanation for Gender Differences in Delinquency." <i>Criminology</i> 37: 41- 72.

### Week 3

January 27<sup>th</sup>

---

#### **Nonlinearities & Curvilinear/Polynomial Regression; Multicollinearity**

❖ Assignment 1 Due

#### **Readings:**

Chapter 12 (Pp. 212-228; 237-248) of Kleinbaum *et al.*

Chapter 13 of Kleinbaum *et al.*

#### **Recommended Readings:**

Stinson, James A., Edward G. Carmines, and Richard A. Zeller. (1981).

“Interpreting Polynomial Regression.” Pp. 88-96 in *Linear Models in Social Research*. Peter V. Marsden (ed.). Sage.

Stine, Robert A. (1995). “Graphical Interpretation Of Variance Inflation Factors.” *The American Statistician* 49: 53-56.

#### **Substantive Reading:**

DeMarie et al. 2004. “College Students’ Memory for Vocabulary in their Majors: Evidence for a Nonlinear Relation Between Knowledge and Memory.” *Canadian Journal of Experimental Psychology* 58: 181-195.

### Week 4

February 3<sup>rd</sup>

---

#### **Heteroskedasticity; Robust (Sandwich) Standard Errors**

#### **Readings:**

Beck, Nathaniel. (1995). “Reporting Heteroskedasticity Consistent Standard Errors.” *The Political Methodologist* 7(2): 4-6

Long, J. Scott, and Laurie H. Ervin. (2000). “Using Heteroskedasticity Consistent Standard Errors In The Linear Regression Model.” *The American Statistician* 54: 217-224. (Available on ProQuest).

#### **Recommended Reading:**

McKinnon, J. G., and H. White. (1985). “Some Heteroskedasticity Consistent Covariance Matrix Estimators With Improved Finite Sample Properties.” *Journal of Econometrics* 29: 305-325.

## Week 5

February 10<sup>th</sup>

---

### Bootstrapped Standard Errors; Outliers

❖ Assignment 2 Due

#### Readings:

Chapter 12 (Pp. 228-237) of Kleinbaum *et al.*

Chapter 14 of Moore, Pages 14.1 - 14.42 (on the CD that came with the book)

Bollen, Kenneth, and Robert Jackman. (1990). "Regression Diagnostics: An Expository Treatment Of Outliers And Influential Cases." Pp.257-291 in *Modern Methods of Data Analysis*. John Fox and J. Scott Long (eds.). Sage.

#### Recommended Readings:

Mooney, Christopher Z. (1996). "Bootstrapping Statistical Inference: Examples And Evaluations For Political Science." *American Journal of Political Science* 40: 570-602.

Mooney, Christopher Z., and Robert D. Duvall. (1993). *Bootstrapping: A Nonparametric Approach to Statistical Inference*. Sage.

Stine, Robert. (1990). "An Introduction To Bootstrap Methods: Examples and Ideas." Pp.325-373 in *Modern Methods of Data Analysis*." John Fox and J. Scott Long (eds.). Sage.

## Week 6

February 17<sup>th</sup>

---

### Introduction to the Generalized Linear Model; MLE

#### Readings:

Chapter 2 of Hoffman

#### Recommended Readings:

Chapter 22 of Kleinbaum *et al.*

## Week 7

February 24<sup>th</sup>

---

### Logistic Regression

❖ Assignment 3 Due

#### Readings:

Chapter 3 of Hoffman

Chapter 16 of Moore and McCabe, 16.1-16.14 (On CD)

## Week 8

March 3<sup>rd</sup>

---

### Logistic Regression/Probit Regression

#### Substantive Readings:

Felson et al. 2005. "Police Intervention and the Repeat of Domestic Assault." *Criminology* 43: 563-588.

Steffensmeier et al. 1998. "The Interaction of Race, Gender, and Age in Criminal Sentencing: The Punishment Cost of Being Young, Black, and Male." *Criminology* 36: 763-798. [OLS and Logistic; Nonlinearities]

Warr, Mark. 1998. "Life-Course Transitions and Desistance from Crime." *Criminology* 36: 183-215. [OLS and Logistic]

## Week 9

March 17<sup>th</sup>

---

### Ordinal Logit Regression (Time Pending)

❖ Assignment 4 Due

#### Reading:

Chapter 4 of Hoffman

#### Substantive Readings:

Emerson, Michael O. and George Yancey. 2001. "Does Race Matter in Residential Segregation?" *ASR* 66: 922-935.

Horney, Julie and Cassia Spohn. 1996. "The Influence of Blame and Believability Factors on the Processing of Simple Versus Aggravated Rape Cases." *Criminology* 34: 135-162. [Proportional Odds Assumption Violated]

## Week 10

March 24<sup>th</sup>

---

### Multinomial Logistic Regression

#### Readings:

Chapter 5 of Hoffman

Chapter 16 of Moore and McCabe, 16.1-16-15 (On CD)

## Week 11

March 31<sup>st</sup>

---

### Multinomial Logistic Regression

#### Substantive Readings:

Browning, Christopher, and Edward O. Laumann. 1997. "Sexual Contact between Children and Adults." *ASR* 1997 62: 540-560. [OLS, Logistic, Mult. Logis]

Klorinek et al. 2005. "Through Thick and Thin: Layer of Social Ties and Urban Settlement among Thai Migrants." *ASR* 70: 779-800.

Ireland et al. 2002. "Developmental Issues in the Impact of Child Maltreatment on Later Delinquency and Drug Use." *Criminology* 40: 359- 396.

## Week 12

April 7<sup>th</sup>

---

### Poisson/Negative Binomial Regression

❖ Assignment 5 Due

#### Readings:

Chapter 6 of Hoffman

Ezell and Cohen (2005). Appendix A, (just read Pp 217-278) of *Desisting From Crime: Continuity and Change in the Long-term Crime Patterns of Serious Chronic Offenders*. Oxford University Press: Oxford, UK.

#### Recommended Readings:

Chapter 24 of Kleinbaum *et al.*

#### Substantive Readings:

- Pogarsky et al. (2003). "The Delinquency of Children Born to Young Mothers: Results from the Rochester Youth Development Study." *Criminology*, 41: 1249–1286. [Negative Binomial]
- Bernberg, Jon Gunnar, and Krohn, Marvin D. (2003). "Labeling, Life Chances, and Adult Crime: The Direct and Indirect Effects of Official Intervention in Adolescence on Crime in Early Adulthood." *Criminology*, 41: 1287–1318. [Modified Poisson]

## Week 13

April 14<sup>th</sup>

---

### Where To Go From Here: Introduction to Advanced Methods

Brief In-Class Final Exam today

#### Readings:

Chapter 8 of Hoffman

## Week 14

April 21<sup>st</sup>

---

### Paper Presentations

#### Readings:

None ☺