CRIM 6480: Multivariate Statistics

Professor: Dr. Marisa Omori (she/her/hers)
Email: marisa.omori@umsl.edu

Class: Tuesdays 5:30pm-8:10pm in Social Science Business 134

Course Description
This course is designed to introduce students to multiple regression techniques that can be used with questions around criminology and criminal justice. We will take an applied approach to these techniques, learning about the assumptions of the models, and some techniques to diagnose and minimize the possibility of improper usage of the models. We will focus mostly on continuous outcomes using ordinary least squares (OLS), although some time will be taken to learning extensions of these regression models for non-continuous outcomes, including dichotomous, count, and truncated outcomes. We will use Stata software and learn about using coding and syntax. Because understanding statistics occurs in our social context, I will also discuss criminology, criminal justice applications, and other social scientific applications throughout the course, and how we can use statistics to work towards racial and social justice.

Learning Objectives
- Critically read and evaluate multiple regression models with continuous outcomes
- Apply multiple regression techniques in social science research and specify models, including diagnosing common problems that may come up with basic statistical research
- Conduct the basics of data management and applied analysis in Stata
- Build a solid foundation for more advanced statistical techniques
- Understand criminology, criminal justice, and social science applications for statistics

Course Logistics
- Although we are meeting in person, we may need to make adjustments as the semester goes on due to COVID-19.
- We will have a lab session for half of class, starting on week 2 (7 lab sessions total). This will give you a hands-on opportunity to learn to use Stata and apply concepts learned in class to real-world examples.
- I will post the syllabus, slides, lab assignments, and lab dataset on Canvas, so make sure you check it regularly.
- Please let me know if you need any disability accommodations. I will happily make accommodations through the office of disability services (314.516.6554).
- Academic dishonesty/integrity: I encourage you to work together on the homeworks and labs, and give each other help and feedback on your final papers, although each of you must use your own words for interpretations (this is grad school!), and turn in your own homework and lab. I do take academic dishonesty seriously, including cheating, collusion, and plagiarism. Here are the UMSL guidelines about academic dishonesty.

Texts
Required

Wooldridge, J.M. (2013). *Introductory Econometrics: A Modern Approach*. (5th Edition). Mason, OH: Cengage Learning
-note: other versions of the book contain similar material!

Optional: basic introductory books, data management, and intro to Stata and R

Hamilton, L.C. (2013). *Statistics with Stata (Updated for version 13)*. Cengage Learning (for those new to Stata, this is a helpful resource)

Acock, A.C. (2012). *A Gentle Introduction to Stata*. (Revised Third Edition). College Station, TX: Stata Press (for those new to Stata, this is a helpful resource)

Hess, F. (2020). Using R for Introductory Econometrics. This is a companion to our textbook for those who want to use and learn R. It's also free online.

Muenchen, R. & Hilbe, J. (2010). *R for Stata Users*.

Kennedy, P. (2008). *A Guide to Econometrics*. (Sixth Edition), Malden, Mass.: Blackwell Publishing.

Greene, W. H. (2008). *Econometric analysis* (6th ed.). New Jersey: Pearson Prentice-Hall.

Fox, J. (1991). *Regression Diagnostics*. Newbury Park, CA: Sage Publications

Aiken, L. and West, S. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage Publications.

Long, S. (2014). *The Workflow of Data Analysis Using Stata*. Stata Press.

Mustillo, S. A., Lizardo, O. A., & McVeigh, R. M. (2018). Editors' comment: A few guidelines for quantitative submissions.

Further reading: advanced and special topics

Long, J.S. (1997). *Regression Models for Categorical and Limited Dependent Variables.*

Long, J.S., & Freese, J. (2006). *Regression Models for Categorical Dependent Variables Using Stata.*

Winship C, Mare RD. Regression Models with Ordinal Variables. *American Sociological Review*. 1984.

Long, J. S., & Mustillo, S. A. (2018). Using predictions and marginal effects to compare groups in regression models for binary outcomes. *Sociological Methods & Research*, 0049124118799374.

Long, J. S. (2009). Group comparisons in logit and probit using predicted probabilities. *Department of Sociology, University of Indiana.*

Berk, R., & MacDonald, J. M. (2008). Overdispersion and Poisson regression. *Journal of Quantitative Criminology*, *24*(3), 269-284.

Wooldridge, Jeffrey. *Econometric Analysis of Cross-Sectional and Panel Data.*

Raudenbush, S.W. & Byrk, A. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd ed.

Snijders, T. & Bosker, R. (2011). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*.

Rabe-Hesketh, R. & Skrondal, A. (2012). *Multilevel and Longitudinal Modeling using Stata.*

Bollen, K. (1989). *Structural Equations with Latent Variables*.

Kline, R. (2016). *Principles and Practice of Structural Equation Modeling*.

Cleves, M., Gould, W., & Gutierrez, R. (2004). *An Introduction to Survival Analysis Using Stata, Revised Edition*.
Bloosfeld, H., Glosch, K., & Rohwer, G. (2007). *Event History Analysis with Stata*.
Morgan, S. & Winship, C. (2015). *Counterfactuals and Causal Inference.*
Winship, C. & Mare, R. (1992). Models for Sample Selection Bias. *Annual Review of Sociology*. 18:327-350.
Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, 153-161.
Bushway, S., Johnson, B. D., & Slocum, L. A. (2007). Is the magic still there? The use of the Heckman two-step correction for selection bias in criminology. *Journal of Quantitative Criminology*, *23*(2), 151-178.
Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International economic review*, 693-709
Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, 436-455.
Wasserman and Faust (1994) *Social network analysis: methods and applications*
Anselin, L. (2013). *Spatial econometrics: methods and models* (Vol. 4). Springer Science & Business Media.
O'Brien, R. (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality & quantity*, *41*(5), 673-690.

Software

We will use Stata code/syntax throughout the course, which is a popular software for social scientists. You can "point and click" in Stata, but I will emphasize the use of writing code, which is a must for conducting statistical research. We have access to Stata in the classroom, although if you want to learn the program long-term, I also encourage you to access Stata on your own computers at home. For the purposes of the class, Stata BE (used to be called Stata IC) should be sufficient. There are two options for getting Stata at home:

- You can buy a student copy of Stata for your computer ($125 for a 1 year license).
- It is also available for free through TritonApps

In addition to the Hamilton and Acock books listed above, here are good websites for Stata:

- UCLA website (lots of great Stata examples). Both the data analysis examples and annotated output are super helpful.
- UNC website (has some good Stata data management examples)

Assignments and Grading

1. **Lab assignments (6 x 10% each=60% of total grade):** these will consist of conceptual questions as well as Stata exercises with interpretations. No late work will be accepted unless for an excused absence.
    - The labs should include two components: 1) The main body of the lab, including relevant cut-and pasted output from Stata (use Courier New, size 8 or 9 to format correctly in Word), as well as your interpretations/responses to questions, and 2) Your .do file, including the commands you used for that particular lab.
    - Where appropriate, PLEASE SHOW YOUR WORK (and box your answers), or you will not receive credit for the problem.

2. **Final paper (30% of total grade):** you will write an empirical paper applying the regression techniques used in the class, where you can use your own dataset (pending my approval), or one provided in the class. The paper should be modeled after a short quantitative journal article (aim for between 4,000-7,000 words, including bibliography), with much greater attention paid to the methods and results, and less attention paid to the introduction/lit review/discussion compared to a normal journal article.
    - I encourage you to select a continuous dependent variable (or a variable that approximates a continuous outcome). While you are certainly welcome to use dichotomous, count, or other types of outcomes, we will not get to these other techniques until the very end of the course.
    - Your methods section should include a discussion about running diagnostics and making appropriate corrections (joint hypothesis test, multicollinearity, heteroskedasticity, etc.) This would likely be given minimal discussion in a real journal article, but I want you to include it here.
    - At least 3 models, including a "baseline" model, a "full" model, and a model including an interactions, polynomial, or other specification. You should conduct nested model tests to help assess fit.
    - At least 2 figures plotted in Stata
    - Make sure you interpret your coefficients and figure in your results section.
    - Include journal-article worthy tables AND figures including predicted values for your key variable(s)

3. **Final paper progress assignments (10% of total grade):** There will be 3 assignments demonstrating progress on the final paper throughout the course.
    1. Abstract (250-400 words) of your project, including the research question, background literature, a description of your dataset, and the potential variables you are interested in using, particularly your dependent variable
    2. Descriptive table of your dependent and independent variables (all of your variables must be cleaned/recoded) and a draft of the methods section of your paper. The methods section includes a) a description of your data and sample; b) the variables section with information on the dependent and independent variables, how they are measured and coded, and an interpretation of the descriptive statistics in your table; and c) the analysis plan of the model you are running, the assumptions of that model, and any diagnostics.
    3. Preliminary models and figures that you have run/created—they don't need to be finalized, or written up, but I want to see that you've run some.

Grade categories

93.5 +	A	73.5-76.4	C
89.5-93.4	A-	69.5-73.4	C-
86.5-89.4	B+	66.5-69.4	D+
83.5-86.4	B	63.5-66.4	D
79.5-83.4	B-	59.5-63.4	D-
76.5-79.4	C+	Below 59.5	F

Schedule—we may make adjustments as we go along!

Week	Reading	Assignment (what you should work on this week)
Week 1: Aug 23 Introduction and statistics review Simple regression	 Ch. 1 Ch. 2	
Week 2: Aug 30 Simple regression Lab 1: Stata basics and simple regression model	 Ch. 2	Lab 1 due Sep. 6
Week 3: Sep 6 Multiple regression: estimation	 Ch. 3	
Week 4: Sep 13 Binary predictors Lab 2: Multiple regression estimation	 Ch. 7.1-7.3	Lab 2 due Sep. 20
Week 5: Sep 20 Inference	 Ch. 4	
Week 6: Sep 27 OLS asymptotics Lab 3: Inference & binary predictors	 Ch. 5	Lab 3 due Oct. 4
Week 7: Oct 4 Binary predictors & interactions	Ch. 7.4, 7.6	Short description of dataset, research question, and variables due Oct. 11
Week 8: Oct 11 Interactions, functional forms, data scaling, and further issues Lab 4: Data management	 Ch. 6	Lab 4 due Oct. 18
Week 9: Oct 18 Interactions, functional forms, data scaling, and further issues Heteroskedasticity	 Ch. 6 Ch. 8	
Week 10: Oct 25 Heteroskedasticity Lab 5: Interactions, polynomials, and further issues	 Ch. 8	Lab 5 due Nov. 1
Week 11: Nov 1 More specification, missing data, other data issues	 Ch. 9	Descriptive table of variables and methods section due Nov. 8
Week 12: Nov 8 Limited dependent variables I: Binary response Lab 6: Heteroskedasticity, specification, and data issues	 Ch. 7.5, 7.7, 17	Lab 6 due Nov. 15
Week 13: Nov 15 No class: work on preliminary models for final projects		Preliminary models and figures for final paper due Nov 29
Week 14: Nov 29 Limited dependent variable II: Categorical and count Nested and multilevel data, advanced topics Lab 7: Limited dependent variable models		Lab 7 due Dec. 6
Week 15: Dec 6 Wrap up		Final paper due Dec 13