

**USP/SOC 982, Fall 2019**  
**Advanced Quantitative Analysis**  
**Wednesdays, 4:30 – 7:20 p.m., Room 289, Bolton Hall**

Professor Marcus Britton  
E-mail: britton@uwm.edu

Office Hours: Tues., 1-2 pm, Bolton 728  
Phone: 414-229-5308

**Course Description and Goals:** The focus of this course is on the practical application of statistical modeling to theory-based analyses of quantitative data. Students will choose a quantitative data set and work throughout the semester on a research paper based upon their analysis of this data. The final product will be an original research paper suitable for presentation at a profession conference and publication in a scholarly journal. Toward this end, the course will briefly review key concepts from basic statistics, including sampling distributions, confidence intervals, and hypothesis testing, as well as several elementary statistics and techniques for analyzing univariate, bivariate and multivariate data (including cross tabulation, correlation coefficients, ordinary least squares regression, and logistic regression). The bulk of the course, however, will be devoted to developing the practical skills needed to use data-management and statistics software (Stata) to carry out the sort of quantitative analyses that facilitate critical evaluation of social scientific theories. The course will introduce students to a variety of advanced methods of quantitative analysis, including several statistical models appropriate for analyzing categorical and other non-continuous outcome variables, such as binary probit, ordered probit and logit, generalized ordered logit, multinomial logit and count models. The course will also briefly introduce additional, advanced topics in quantitative analysis, including selection models and propensity score models. The presentation of advanced statistical models in the course will emphasize how to select models that are appropriate for specific research questions and specific forms of data, while providing only a relatively superficial treatment of the mathematics used to derive the models and implement model estimation. The course will also emphasize how to interpret the results of quantitative analysis techniques that are widely used in sociology, urban studies and other social scientific fields. As such, this course is intended to enable students to read and critically evaluate social scientific publications and to complete the steps in the process of quantitative empirical research, moving from a research idea through data analysis to presentation of findings.

**Prerequisites:** All students must have taken at least one graduate statistics course that included OLS regression and logistic regression and must have achieved 85% or higher on the diagnostic exam (or the final exam for Sociology 760). The course work assumes that students are familiar with both basic principles of statistical inference, including sampling distributions, confidence intervals and hypothesis testing, and basic methods of quantitative analysis, including cross tabulation, multiple linear regression, and logistic regression. As noted above, we will review this material, but this review will presume prior familiarity and a basic conceptual grasp of these principles and techniques.

**Workload Expectations:** It is expected that students will spend 9–15 hours per week completing tasks related to this course, in addition to the three hours allocated for lecture and lab. While the workload will vary from week to week, students should generally expect to spend about 3–5 hours reading assigned course materials, about 3–5 hours completing homework assignments and lab exercises, and about 3–5 hours working on the final paper project and/or preparing for the final exam.

**Required Texts:**

Aneshensel, Carol. 2013. *Theory-Based Data Analysis for the Social Sciences*, 2nd edition. Thousand Oaks, CA: Pine Forge Press.

Long, J. Scott and Jeremy Freese. 2014. *Regression Models for Categorical Dependent Variables Using Stata*, 3rd Edition. College Station, TX: Stata Press.

**Recommended Texts:**

Heeringa, Steven G., Brady T. West, and Patricia A. Berglund. 2010. *Applied Survey Data Analysis*. Boca Raton, FL: Chapman & Hall/CRC Press.

Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage Publications.

Long, J. Scott. 2009. *The Workflow of Data Analysis Using Stata*. College Station, TX: Stata Press.

NOTE: Several chapters from Long (1997) are required reading (see below). However, the required chapters are available for download as PDF files on the course D2L site, as are all other assigned readings not included in the required texts. I nevertheless encourage you to purchase your own copy.

**Software:**

This course will make extensive use of Stata, which is among the most widely used and versatile data-management and statistical software packages available to social scientists.

Stata 16 is the most current version of the software. While not required for the course, I encourage you to purchase your own copy (e.g., so that you can work on labs or other course assignments on your laptop computer). Details about pricing for different versions of the software are available here (but see my recommendations below):

<http://www.stata.com/order/new/edu/gradplans/student-pricing/>

*Earlier versions of Stata and using Stata in campus computer labs.* While Stata 16 is the most recent version of the software, the examples in both lecture and the main text (i.e., Long and Freese 2014) will usually run on version 11 or later. Accordingly, you are welcome to use an earlier version of the software, so long as it is version 11 or later. The lab computers in Bolton 293 have version 15, and the computers in the statistics lab in the Sociology Department will have a recent version of the software, as well (currently, version 14).

*Purchasing recommendations.* If you do not already own your own copy of Stata, I recommend purchasing Intercooled Stata 16 (also known as “Stata/IC”), which is available with the following licensing options:

- Intercooled Stata 16, 6-month license: \$48
- Intercooled Stata 16, annual license: \$94
- Intercooled Stata 16, perpetual license: \$225

The six-month license is sufficient for the course, but you may wish to purchase an annual or perpetual license if you plan on incorporating quantitative analysis into future graduate work (e.g., a master’s thesis or dissertation).

For most students in this course, it will not be necessary or useful to purchase Stata/SE or Stata/MC.

NOTE: I only have Stata 15 in my office. Accordingly, if you are using version 16, you will need to use the `saveold` command in order to provide copies of your data, as required in many of the assignments for the course.

### **Grading:**

**1) Homework assignments (30% of your grade).** There will be five homework assignments. Each homework assignment will give you the opportunity to complete a key step in the process of completing your final paper (see below). Late assignments will be downgraded by one letter grade, plus one additional letter grade for each full day (24 hours) that they are late. Note: Homework assignments 1 through 4 will each be worth 5% of your final grade, while homework assignment 5 will be worth 10%.

**2) Lab exercises (20% of your grade).** There will be five required lab exercises, in which you will use Stata to apply some of the concepts and techniques learned in lecture and the assigned readings. I strongly encourage collaboration among students with respect to mastering the material and debugging statistical software runs. However, you are required to write up your own answers to the lab exercises, independent of other students. You are also required to produce and upload your own do files and log files whenever required by the exercise. You will typically be able to complete most of each exercise during lab, but some additional outside work may be required. Lab assignments that are not completed during lab should be uploaded to D2L by the following Friday. Your lowest lab exercise grade will be dropped, but you must complete all five

exercises in order to pass the course.

**3) Take-home Exam (10% of your grade).** A cumulative take-home exam will be posted to the course D2L site after Thanksgiving break (i.e., at least by December 2nd). The focus of the exam will be on correct interpretation of model results, comparable to those you might encounter in a published social-scientific research article. You will not need to work with data files, and you will not be required to use Stata to complete the exam, though you may use the program to make calculations if you wish (any standard scientific calculator will serve just as well). This will be an open-book, open-note exam, but you are required to work individually. In order to receive credit for your exam, you will need to upload your completed answers to D2L by Monday, December 16<sup>th</sup>.

**4) Empirical research paper (25% of your grade).** The final project will require you to select a dataset of interest to you, draft one to four hypotheses you wish to test using these data, motivate the research with a literature review, complete the statistical analysis, and write up the results. The final paper is expected to be of the quality one would find in a peer-reviewed social science journal. **Work on the final paper should begin the first week of class!** You will be required to provide me with an electronic copy of both the original version of your data and the cleaned version that you used for the analysis, along with do files that you prepared for data cleaning, preliminary analysis, and model estimation. I should be able to reproduce the results you report in your paper using the data and do files you submit. The paper and accompanying documentation are due (i.e., must be uploaded to D2L) by Saturday, December 21st, the last day of exam week.

**5) Empirical research paper: presentation (10% of your grade).** In addition to writing a final paper, you will give a professional quality presentation to the class (20-30 minutes) of your findings. This presentation will be similar to a presentation you would give at a professional conference. All students will grade and critique the presentations.

**6) Discussion leader and general participation (5% of your grade).** Twice during the semester, you will serve as the facilitator of the discussion of an assigned article that illustrates the use of a specific type of statistical model. In addition, students will be expected to contribute regularly to class, both by asking questions about lectures and related materials and by actively participating in class discussions.

**Overview of Assignment Due Dates:**

<i>Assignment</i>	<i>Due Date</i>
Homework 1 (Data & Hypothesis)	Wednesday, Sept. 11
Homework 2 (Preliminary Analysis)	Wednesday, Sept. 18
Lab Exercise 1 (OLS)	Friday, Sept. 20
Lab Exercise 2 (OLS Interactions)	Friday, Sept. 27
Homework 3 (Multiple Imputation)	Wednesday, Oct. 13
Lab Exercise 3 (Binary Logit)	Friday, Oct. 18

Lab Exercise 4 (Ordered/Multinomial Logit)	Friday, Nov. 8
Lab Exercise 5 (Count Models)	Friday, Nov. 15
Homework 5, rough draft	Wednesday, Nov. 20
Thanksgiving Break	Nov. 27 – Dec.1
Homework 4 (peer review of H5 rough draft)	Wednesday, Dec. 4
Homework 5, final	Wednesday, Dec. 11

Final Presentation: Wednesday, December 11th (last day of class)

Take-Home Final Exam: due Monday, December 16<sup>th</sup>

Final Paper: due Saturday, December 21<sup>st</sup>

### **Class Website:**

Lecture slides, exercises, and all readings not included in the required texts will be posted on the course website (D2L).

### **Lecture and Lab Format:**

Class will take place in Bolton 293 from 4:30 to 7:20 p.m. on Wednesdays; this time will be loosely divided between formal lecturing, going over examples in Stata, and discussing assigned articles, depending upon the nature of the material to be covered. We may occasionally run over the allotted time (i.e., class may end as late as 7:30 p.m.). Note that **attendance is required.**

You are expected to have read the assigned material prior to the lecture. The lectures will only make sense when you have prepared adequately for class. Since this is a small, Ph.D.-level course, all students will be expected to contribute to the discussion and explication of material during the class. Office hours are intended to be used for clarification of material and assistance with assignments, readings, and the final project and not to reiterate material that was covered in lecture.

## **Weekly Course Schedule**

*Week 1: Introduction to the Course & Review of Descriptive Statistics (9/4)*

Topics: Course Overview and Logistics, Theory-Based Data Analysis, Levels of Measurement, Univariate & Bivariate Distributions, Measures of Central Tendency and Variability, Measures of Bivariate Association

Required Readings:

Aneshensel (2013), Chapters 1–4

Recommended Readings:

Long (2009), *Workflow*, Chapters 1 & 2

Lab: Introduction to Stata & work on Homework Assignment 1

*Week 2: Descriptive Analysis & Statistical Inference from Complex Samples (9/11)*

HOMEWORK ASSIGNMENT #1 DUE (DATA & HYPOTHESIS)

Topics: Sampling Distributions, Confidence Intervals, Hypothesis Testing, Introduction to Sampling Weights and Statistical Inference with Complex Samples

Required Readings:

Long and Freese (2014), “Part I: General Information,” pp. 7–81, 99–102

Johnson, David R., and Lisa A. Elliot. 1998. “Sampling Design Effects: Do They Affect the Analyses of Data from the National Survey of Families and Households?” *Journal of Marriage and Family* 60:993–1001.

Discussion Article:

Leahey, Erin. 2005. “Alphas and Asterisks: The Development of Statistical Significance Testing Standards in Sociology.” *Social Forces* 84(1):1–24.

Guidance on Using Weights and Complex Variance Estimation in Stata:

Read “Introduction to survey commands” in the Stata Survey Data Reference Manual (type “help survey” in Stata’s command window and then click on the blue “[SVY] survey” link to access the PDF file)

Recommended Readings:

Heeringa, West and Berglund (2010), Chapters 2, 4–6

Long (2009), Chapters 3 & 5

Lab: Using Stata for Preliminary Analysis (Homework 2)

*Week 3: The Linear Regression Model, Part I (9/18)*

HOMEWORK ASSIGNMENT #2 DUE (PRELIMINARY ANALYSIS)

Topics: Review of Ordinary Least Squares (OLS) Regression; Estimating Linear Regression Models from Complex Samples

Required Readings:

Aneshensel (2013), Chapters 5–8

Long and Freese (2014), pp. 83–93, 102–115

Fox, John. 2008. “What is Regression Analysis?” Pp. 13–17 in *Applied Regression Analysis and Generalized Linear Models*, 2nd ed. Thousand Oaks, CA: Sage Publications.

Long, J. Scott. 1997. “Continuous Outcomes: The Linear Regression Model.” Pp. 11–33 in *Regression Models for Categorical and Limited Dependent Variables*. Advanced Quantitative Techniques in the Social Sciences, vol. 7. Thousand Oaks, CA: Sage Publications.

Mustillo, Sarah A., Omar A. Lizardo, and Rory M. McVeigh. 2018. “Editors’ Comment: A Few Guidelines for Quantitative Submissions.” *American Sociological Review* 83(6):1281–83.

Winship, Christopher and Larry Radbill. 1994. “Sampling Weights and Regression Analysis.” *Sociological Methods & Research* 23(2):230–57.

Discussion Article:

Klebanov, Pamela K., Jeanne Brooks-Gunn, Greg J. Duncan. 1994. “Does Neighborhood and Family Poverty Affect Mothers' Parenting, Mental Health, and Social Support?” *Journal of Marriage and the Family* 56:441–55.

Recommended Readings:

Heeringa, West and Berglund (2010), Chapter 7

Lab Exercise 1: Interpreting Linear Regression Results

*Week 4: The Linear Regression Model, Part II (9/25)*

Topics: Interactions, Mediation vs. Moderation, Postestimation Regression Diagnostics

Required Readings:

Aneshensel (2013), Chapters 9 & 11

Baron, Reuben M. and David A. Kenny. 1986. “The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic and Statistical Considerations.” *Journal of Personality and Social Psychology* 51(6):1173–82.

Fox, John. 2008. Selections from Chapters 13, “Collinearity and Its Purported Remedies.” Pp. 307–313, 323–329 in *Applied Regression Analysis and Generalized Linear Models*, 2nd ed. Thousand Oaks, CA: Sage Publications.

Tabachnick, Barbara G., and Linda S. Fidell. 2007. “4.1.4 Outliers,” “4.1.5 Normality, Linearity and Homoscedasticity,” and “5.3.2 Practical Issues.” Pp. 72–92, 123–128 in *Using Multivariate Statistics*, 5th ed. Boston, MA: Pearson/Allyn & Bacon.

Discussion Article:

Musick, Kelly and Ann Meier. 2014. “Variation in Associations Between Family Dinners and Adolescent Well-Being.” *Journal of Marriage and Family* 76:13–23.

Lab Exercise 2: Graphing Interactions from OLS Models

*Week 5: Data Screening and Preparation (10/2)*

Topics: Data Management, Dealing with Missing Values, Multiple Imputation

Required Readings:

Long and Freese (2014), pp. 93–99

Firebaugh, Glenn. 2008. “Internal Reality Checks.” Pp. 65–68 in *Seven Rules for Social Research*. Princeton, NJ: Princeton University Press.

Johnson, David R. and Rebekah Young. 2011. “Toward Best Practices in Analyzing Datasets with Missing Data: Comparisons and Recommendations.” *Journal of Marriage and Family* 73:926–45.

Schafer, Joseph, and John W. Graham. 2002. “Missing Data: Our View of the State of the Art.” *Psychological Methods* 7: 147–77.

Guidance on Implementing Multiple Imputation in Stata:

Social Science Computing Cooperative, University of Wisconsin. “Multiple Imputation in Stata.” [https://www.ssc.wisc.edu/sscc/pubs/stata\\_mi\\_intro.htm](https://www.ssc.wisc.edu/sscc/pubs/stata_mi_intro.htm) (READ ALL 8 SECTIONS!)

Read “Introduction to multiple-imputation analysis” and “Introduction to mi” in the Stata PDF documentation. (Type `help intro substantive` and/or `help mi` if you have trouble finding the documentation.)

Recommended Reading:

Heeringa, West and Berglund (2010), Chapter 11

Long (2009), Chapter 6

Lab: Compare OLS Results with and without Multiple Imputation (Homework 3)



*Week 6: Binary Logit and Probit Models, Part I (10/9)*

HOMEWORK ASSIGNMENT #3 DUE AT THE BEGINNING OF CLASS (OLS)

Topics: Review of Logistic Regression, Introduction to Probit Models, Comparison of Logit and Probit Models, Model Fit in Non-Linear Models, Maximum Likelihood Estimation

Required Readings:

Aneshensel (2013), Chapter 12

Long and Freese (2014), pp. 115–131, Chapters 5

Long, J. Scott. 1997. “Binary Outcomes: The Linear Probability, Probit and Logit Models.” Pp. 34–84 in *Regression Models for Categorical and Limited Dependent Variables*. Advanced Quantitative Techniques in the Social Sciences, vol. 7. Thousand Oaks, CA: Sage Publications.

Discussion Article:

Britton, Marcus L. 2013. “Race/Ethnicity, Attitudes, and Living With Parents During Young Adulthood.” *Journal of Marriage and Family* 75:995–1013.

Lab Exercise 3: The Elaboration Model with Binary Logit (begin working)

*Week 7: Binary Logit and Probit Models, Part II (10/16)*

Topics: Approaches to Interpreting Binary Logit and Binary, Interactions in Binary Logit and Probit Models

Required Readings:

Long and Freese (2014), Chapters 4 and 6

Long, Scott J. and Sarah A. Mustillo. 2018. “Using Predictions and Marginal Effects to Compare Groups in Regression Models for Binary Outcomes.” *Sociological Methods & Research*. Advanced online publication. [doi.org/10.1177/0049124118799374](https://doi.org/10.1177/0049124118799374)

Discussion Article:

Egan, Mark, Michael Daly and Liam Delaney. 2016. “Adolescent Psychological Distress, Unemployment, and the Great Recession: Evidence from the National Longitudinal Survey of Youth 1997.” *Social Science & Medicine* 156:98–105.

Lab Exercise 3: The Elaboration Model with Binary Logit (complete)

### *Week 8: Ordered Logit and Ordered Probit (10/23)*

Topics: Measures of Association for Ordinal Data, Modeling Ordinal Outcome Variables

Required Readings:

Long and Freese, (2014), Chapter 7

Long, J. Scott. 1997. "Ordinal Outcomes: Ordered Logit and Ordered Probit Analysis." Pp. 114–147 in *Regression Models for Categorical and Limited Dependent Variables*. Advanced Quantitative Techniques in the Social Sciences, vol. 7. Thousand Oaks, CA: Sage Publications.

Discussion Articles:

Henig, Jeffrey R., and Jason A. MacDonald. 2002. "Location Decisions of Charter Schools: Probing the Market Metaphor." *Social Science Quarterly* 83: 962–80.

Recommended Readings:

Heeringa, West and Berglund (2010), Chapter 9, pp. 277–286

### *Week 9: Generalized Ordered Logit (10/30)*

Topics: Modeling Ordinal Outcome Variables, continued

Required Readings:

Long and Freese, (2014), "The generalized ordered logit model," pp. 371–74

Fullerton, Andrew S. and Jun Xu. 2016. Pp. 55–107 in "Chapter 3: Partial Models" and "Chapter 4: Nonparallel Models." *Ordered Regression Models: Parallel, Partial, and Non-Parallel Alternatives.* Boca Raton, FL: CRC Press.

### *Week 10: Multinomial Logit (11/6)*

Topics: Modeling Categorical/Nominal Outcome Variables

Required Readings:

Long and Freese (2014), Chapter 8 (skim pp. 444–79)

Long, J. Scott. 1997. "Nominal Outcomes: Multinomial Logit and Related Models." Pp. 148–186 in *Regression Models for Categorical and Limited Dependent Variables*. Advanced Quantitative Techniques in the Social Sciences, vol. 7. Thousand Oaks, CA: Sage Publications.

Discussion Article:

Krysan, Maria. 2008. "Does Race Matter in the Search for Housing? An Exploratory Study of Search Strategies, Experiences, and Locations." *Social Science Research* 37: 581–603.

Lab Exercise 4: Ordered & Multinomial Logit

*Week 11: Count Models (11/13)*

Topics: Conceptualizing Count Outcomes, Poisson Regression, Negative Binomial Regression, Zero-Inflated Count Models, Selecting the Right Model for Count Outcomes

Required Readings:

Long and Freese (2014), Chapter 9

Long, J. Scott. 1997. "Count Outcomes: Regression Models for Counts." Pp. 217–250 in *Regression Models for Categorical and Limited Dependent Variables*. Advanced Quantitative Techniques in the Social Sciences, vol. 7. Thousand Oaks, CA: Sage Publications.

Discussion Article:

McCord, Eric S., and Jerry H. Ratcliffe. 2007. "A Microspatial Analysis of the Demographic and Criminogenic Environment of Drug-Markets in Philadelphia." *Australian and New Zealand Journal of Criminology* 40(1):43–63.

Recommended Readings:

Heeringa, West and Berglund (2010), Chapter 9, pp. 286–298

Lab Exercise 5: Count Models

*Week 12: Selection Models (11/20)*

ROUGH DRAFT OF ASSIGNMENT #5 DUE AT THE BEGINNING OF CLASS

Topics: Sample Selection Bias, Heckman Models

Required Readings:

Berk, Richard. 1983. "An Introduction to Sample Selection Bias in Sociological Data." *American Sociological Review* 48: 386–398.

Guo, Shenyang and Mark W. Fraser. 2015. "Chapter 4: Sample Selection and Related Models." Pp. 95–128 in *Propensity Score Analysis: Statistical Methods and Applications*. Advanced Quantitative Techniques in the Social Sciences Series. Vol. 11. Thousand Oaks, CA: Sage.

Discussion Article:

Rotolo, Thomas and John Wilson. 2007. "Sex Segregation in Volunteer Work." *Sociological Quarterly* 48:559–85.

**Thanksgiving Break: 11/27 – 12/1**

*Week 13: Causal Inference and Propensity Score Analysis (12/4)*

HOMEWORK ASSIGNMENT #4 (PEER REVIEW) DUE: PLEASE E-MAIL YOUR RESPONSE TO YOUR PARTNER'S DRAFT VERSION OF ASSIGNMENT #5 TO HIM OR HER BEFORE CLASS

Topics: Causal Inference in Regression, Propensity Score Estimation, Propensity Score Weighting

Required Readings:

Gelman, Andrew and Jennifer Hill. 2007. "Chapter 9: Causal Inference Using Regression on the Treatment Variable" and "Chapter 10: Causal Inference Using More Advanced Models." Pp. 167–233 in *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.

Guo, Shenyang and Mark W. Fraser. 2015. "Chapter 7: Propensity Score Weighting." Pp. 239–54 in *Propensity Score Analysis: Statistical Methods and Applications*. Advanced Quantitative Techniques in the Social Sciences Series. Vol. 11. Thousand Oaks, CA: Sage.

Discussion Article:

MacDonald, John, Robert J. Stokes, Greg Ridgeway and K. Jack Riley. 2007. "Race, Neighbourhood Context and Perceptions of Injustice by the Police in Cincinnati." *Urban Studies* 44(13):2567–85.

*Week 14: Student Research Presentations (12/11)*

HOMEWORK ASSIGNMENT #5 DUE AT THE BEGINNING OF CLASS