

Oregon State University
School of Public Policy
Spring 2013

Sociology 516: Conducting Social Research
4 Credits

Lecture: MW 10:00-11:50, MLM 033

Lab: F 9:00-9:50 (Johnston) **OR** F 1:00-1:50 (Burkhardt), WITH 205

Dr. Brett Burkhardt
Fairbanks Hall 309B

E-mail: brett.burkhardt@oregonstate.edu
Office Hours: TBD

Dr. Alison Johnston
Gilkey Hall 300b

E-mail: Alison.Johnston@oregonstate.edu
Office Hours: TBD

This course is designed as an introduction to the use of quantitative analysis in social science research. The intent is to cover topics that graduate students will need in working on their theses and dissertations. Students are required to have their own data for this course (finding a dataset will be one of the assignments). In this course, we will take a hands-on approach. As noted by Studenmund's *Using Econometrics: A Practical Guide*, "Econometric education is a lot like learning to fly a plane; you learn more from actually doing it than you learn from reading about it." The course will consist of two parts: lectures and labs. In the lectures, we will discuss the mathematics of multivariate statistical modelling, how they have been applied to academic papers, and work through relevant exercises. In the computer labs, using STATA, we will work with various datasets, exploring the methods that we learn in lecture. This course will focus on: linear regressions, logistic regressions (binary dependent variables), ordinal logistic regressions (dependent variables with an ordinal scale), multinomial regressions (dependent variables that are unrankable categories), Poisson-type regressions (count-based dependent variables), and experimental and quasi-experimental research designs.

Student Learning Outcomes:

By the end of the course students will:

1. Evaluate and apply methods of multivariate quantitative analysis.
2. Develop a theoretically-informed research question, testable hypotheses and robust regression models to test the theory
3. Judge empirical papers that utilize different regression techniques
4. Conduct statistical modelling within the regression software program STATA and interpret the output.

Prerequisites

Successful completion of ECON 524 and SOC 515.

Core Texts

Required:

Studenmund, A.H. (2011) *Using Econometrics: A Practical Guide*. Pearson: Addison and Wesley. 6th Edition.

Long, J.S. and Freese, J. (2006) *Regression Models for Categorical Dependent Variables Using STATA*. STATA Press. 2nd Edition.

Johnston, A. (2012) *Methods/STATA Manual for the School of Public Policy*

Supplementary readings:

In addition to the core text, there will be supplementary articles assigned for each week where relevant methods are used. We **STRONGLY** recommend you read at least one article.

While not required, the following book is an excellent resource for best practices using Stata. It is highly recommended for those intending to use Stata in the future: Long, J. Scott. 2009. *The Workflow of Data Analysis Using Stata*. Stata Press. It is available at: <http://www.stata.com/bookstore/>.

Assessment of Outcomes

- Participation (30%)
 - STATA manual (10%)
 - Lab home works (total of 4, worth 5% each)
- Midterm Exam (Take home exam: Due in lab Week 5) (15%)
- Peer Review Assignment (Due Wednesday Week 6) (15%)
- Final Project (25%)
 - Selection of dataset (For feedback, due **by week 3 in lab**)
 - Presentation of testable hypothesis (5% - Due by **week 6 in lab**)
 - Final report (20% - Due in **week 10 in lab**)
- Final Exam (in final lab) (15%)

The *STATA manual* is a log of all commands learned within the labs. We require that you keep a manual of all commands used, their purpose, and the output they

produce. This manual is for your benefit, because you will be able to use it for reference in future courses and for your dissertation/thesis.

Peer review assignment (due in Week 6): The editor of an academic journal has assigned you the task of reviewing an academic paper (i.e. comment on its research design and methodology, and offer your opinion on whether it should be published). In a 1,200 word (maximum) essay, you must provide the following: 1.) a brief summary of the paper and its contribution to the field; 2.) a summary of the methods used; 3.) a critique of the methods used (i.e. what are the shortcomings, advantages for using the particular methods the authors did), and; 4.) based on these three criteria, your assessment of whether you believe the article is worthy of publication, and what revisions need to be made before it is published.

Final Project (due in Week 10): Your final project will be a 10 page paper (maximum – not including output tables and references) that takes the form of a journal article. In your project, you must include: a research question and testable hypothesis (or hypotheses), a brief review (2-4 recent articles) of the literature relevant to your hypothesis, a description of the data and the regression model you developed to test your hypothesis including the results, and concluding remarks that discuss the implications of your research. This paper should have at least 7 academic references, and must rely upon the use of one econometric method, along with relevant tests associated with the method's assumptions, discussed in class. The dataset you will use for your paper is **due in Week 3's lab**. The hypothesis you aim to test is **due in Week 6's lab**. You will be able to test the methods used in labs on your dataset after completion of lab assignments.

The **midterm exam** is a take home exam (due at the beginning of lab in week 5) which will require the application to concepts learned in class to an assigned dataset. We require STATA command and output be attached to your solutions, which will be a discussion of your results.

The **final exam** is open book, open notes and open internet, and will be a lab based application of all modelling concepts learned in class.

Grading Scale

- A:	94-100%	- C:	73-76%
- A-:	90-93%	- C-:	70-72%
- B+:	87-89%	- D+:	67-69%
- B:	83-86%	- D:	63-66%
- B-:	80-82%	- D-:	60-62%
- C+:	77-79%	- F:	<60%

Other Expectations

Plagiarism: Academic work must be your own. It is plagiarism to claim work (such as writing, exams or projects) done by anyone other than the author(s) named. Plagiarism also includes cutting and pasting information from websites without

attribution of **AND** paraphrasing someone else's ideas or writing. It is not sufficient to re-arrange or re-state someone else's writing or ideas. A zero tolerance policy will be applied towards plagiarism and any work which is plagiarized will automatically result in a **COURSE GRADE OF F**. For more information on how the university handles academic misconduct, go to <http://oregonstate.edu/admin/stucon/achon.htm>.

Disrespectful behavior: Disrespectful behavior towards students on grounds of race, gender, economic background, age, sexual orientation, religion, or any other factors which individuals have no choice or are irrelevant to the class will not be tolerated. **Disrespectful behavior can result in course expulsion.** For more information on the university's policy regarding academic conduct go to <http://oregonstate.edu/admin/stucon/achon.htm>.

Disabilities: Students with accommodations approved through the Services for Students with Disabilities (SSD) are responsible for contacting the faculty member in charge of the course prior to or during the first week of the term to discuss accommodations. Students who believe they are eligible for accommodations but who have not yet obtained approval through SSD should contact SSD immediately at 737-4098.

Course Schedule:

Week 1: Review of Multivariate Linear Regression (Johnston)

Review of hypothesis testing, multivariate linear regression (ordinary least squares - OLS), interpreting beta coefficients, R-squared values, and F-statistics and the classic assumptions of OLS, learned in ECON 524.

Core Reading:

- Studenmund (2011): Chapters 2 and 4

Supplementary Reading:

- Kennedy, B., Kawachi, I., Prothrow-Stith, D., Lochner, K., and Gupta, V. (1998). "Social Capital, Income Inequality, and Firearm Violent Crime", *Social Science and Medicine*. Vol. 47, Issue, 1: Pg. 7-17.
- Sher, K., Bartholow, B., and Nanda, S. (2001) "Short- and Long-Term Effects of Fraternity and Sorority Membership on Heavy Drinking: A Social Norms Perspective" *Psychology of Addictive Behaviors*. Vol 15, No. 1.
- Fisman, R. and Miguel, E. (2007) "Corruption, Norms, and Legal Enforcement: Evidence from Diplomatic Parking Tickets", *Journal of Political Economy*. 115(6): 1020-1048. ****I highly recommend you read this article as its data will feature in your second homework. Don't worry too much about the model design for now, just the main argument of the paper.****

Lab Exercises:

- **First week lab assessment.** This assessment will be a review of multivariate regression analysis and hypothesis testing. You are allowed to use notes and the STATA manual for reference purposes. The assignment will test your application of the concepts learned in Econ 524 on a new dataset. It will NOT count towards your grade.

Week 2: Non-linear OLS variables, dummy variables and interaction terms (Burkhardt)

Applications of non-linear variables (inverse terms, quadratic terms, double and semi-log form), dummy and categorical independent variables and interaction terms between two independent variables within the classic OLS model.

Core Reading:

- Studenmund (2011): Chapter 7
- Lab Reading: Lesson 8 in the STATA Manual

Supplementary Reading:

- Tienda, M. and Lii, D.T. (1987) “Minority Concentration and Earnings Inequality: Blacks, Hispanics, and Asians Compared”, *American Journal of Sociology*. Vol. 93, No. 1, pg. 141-165.
- Svensson, R. and Pauwels, L. (2010) “Is a Risky Lifestyle Always Risky? The Interaction Between Individual Propensity and Lifestyle Risk in Adolescent Offending: A Test in Two Urban Samples”. *Crime and Delinquency*, Vol. 56, pg. 608-626

Lab Exercises:

- Generation of non-linear variables in STATA and interpreting their coefficients
- Creation of dummy/categorical variables and their inclusion within OLS
- Generation of interaction terms and interpreting their coefficients.

Week 3: The classic linear model specification (Burkhardt)

Discussions of the impact of omitted (independent) variables, the inclusion of irrelevant variables, influential cases (outliers) and restricted sampling, the misuse of specification criteria on the OLS linear model and alternative goodness of fit approaches.

Core Reading:

- Studenmund (2011): Chapter 6
- Lab Reading: Lesson 9 in the STATA Manual

Supplementary Reading:

- Torras, M. and Biyçe, J. (1998) “Income, inequality, and pollution: a reassessment of the environmental Kuznets Curve” *Ecological Economics*, Vol. 25, pg. 147-160
- Stern, D. and Common, M. (2001) “Is there an environmental Kuznets Curve for Sulfur?”, *Journal of Environmental Economics and Management*. Vol. 41, pg. 162-178.

Lab Exercises (*First homework due at the beginning of Friday’s lab*):

- Examining the omitted variable and irrelevant variable bias on regression outcomes
- Ramsey RESET specification tests
- dFBetas and dFFits tests for influential cases
- Jack-knife analyses

Week 4: The classic linear model (violated assumptions) (Johnston)

Discussions about the impact of including correlated independent variables (multicollinearity) within OLS regression models, and having an error term without a constant variance (heteroskedasticity).

Core Reading:

- Studenmund (2011): Chapter 8 and 10
- Lab Reading: Lesson 10 in the STATA Manual

Supplementary Reading:

- Binder, S., and Neumayer, E. (2005). “Environmental pressure group strength and air pollution: An empirical analysis”, *Ecological Economics*. Vol. 55, Issue 4: Pg: 527-538 (Note: *focus particularly on OLS model, NOT panel data RE estimator*).
- Djankov, D., Ganser, T., McLiesh, C., Ramalho, R., Shleifer A. (2010) "The Effect of Corporate Taxes on Investment and Entrepreneurship," *American Economic Journal: Macroeconomics*. Vol. 2, No. 3: 31-64

Lab Exercises:

- Pair-wise correlation testing of significance between independent variables
- Variance Inflation Factor and Tolerance (TOL) tests for multivariate multicollinearity
- Breusch-Pagan and White tests for heteroskedasticity.

- Means to correct for multicollinearity and heteroskedasticity within the linear model (robust standard errors, exclusion of terms)

Week 5: Count Models (Burkhardt)

Analyzing count outcomes with Poisson and Negative Binomial Regression models. Using predicted probabilities to interpret results.

Core Reading:

- Long & Freese (2006): Chapter 8 (8.1, 8.2, and 8.3)
- Coxe, Stefany, Stephen G West, and Leona S Aiken. 2009. “The analysis of count data: a gentle introduction to poisson regression and its alternatives.” *Journal of personality assessment* 91(2):121–36.

Supplementary Readings:

- Jacobs, David, and Jason T. Carmichael. 2002. “Subordination and violence against state control agents: Testing political explanations for lethal assaults against the police.” *Social Forces* 80(4):1223–1251.
- Drury, a. J., and M. DeLisi. 2008. “Gangkill: An Exploratory Empirical Assessment of Gang Membership, Homicide Offending, and Prison Misconduct.” *Crime & Delinquency* 57(1):130–146.

Lab Exercises (***Second homework due at the beginning of Friday’s lab***):

- Analyzing count-based dependent variables with Poisson regression.
- Testing for overdispersion in the dependent variable.
- Analyzing count-based dependent variables with Negative Binomial Regression in cases of overdispersion.
- Generating predicted probabilities with `prcounts` (after installing `SPost`).

Week 6: Logistic Regressions (Burkhardt)

Regression models for binary dependent variables: maximum likelihood models (i.e. logistic regressions). Discussion of marginal effects, odds ratios and predicted probabilities.

Core Reading:

- Studenmund (2011): Chapter 13
- Long and Freese (2006): Chapter 4 (up to section 4.6.6 ONLY, pages 131-168)
- Lab Reading: Lesson 11 in STATA Manual

Supplementary Reading:

- Piquero, N., Schoepfer, A. and Langton, L. (2010) “Completely Out of Control or the Desire to Be in Complete Control? How Low Self-control and the Desire for Control Relate to Corporate Offending”, *Crime and Delinquency*. Vol. 56, pg. 627-647
- Croco, S. (2011) “The Decider’s Dilemma: Leader Culpability, War Outcomes, and Domestic Punishment”, *American Political Science Review*. Vol. 105, No. 3: Pg. 457-477. (Note: *focus particularly on logistic model, NOT ordinal logit outcomes*)

Lab Exercises:

- Revisiting contingency tables/cross-tabs with binary dependent variables and categorical independent variables.
- Applications of logistical modelling to datasets and the interpretation of their results.
- Calculating and graphing fitted probabilities.

Week 7: Logistic Regressions (violation of assumptions and specification issues) (Johnston)

Applications of multicollinearity, heteroskedasticity, and misapplications of interaction terms to logistic regression models

Core Reading:

- Norton, E., Wang, H., and Ai, C. (2004) “Computing interaction effects and standard errors in logit and probit models”, *The STATA Journal*. Vol. 4, No. 2, pg. 154-167
- Lab Reading: Lesson 12 in STATA Manual

Supplementary Reading:

- Bobak, M. (2000). “Outdoor Air Pollution, Low Birth Weight, and Prematurity”. *Environmental Health Perspectives*. Vol. 108, No. 2. Pg. 173-176.
- Hamann, K., Johnston, A, and Kelly, J. (Forthcoming, 2013) "Striking Concessions from Governments: Explaining the Success of General Strikes in Western Europe, 1980-2009." *Comparative Politics*.
- Evans, L. (2011) “Expertise and Scale of Conflict: Governments as Advocates in American Indian Politics”, *American Political Science Review*. Vol. 105, No. 4: Pg. 663-682.

Lab Exercises (***Third homework due at the beginning of Friday’s lab***):

- Testing for and controlling for multicollinearity and heteroskedasticity within logit models

- Problems with interaction terms in logit models. Calculating fitted probabilities which account for interactions between variables.

Week 8: Ordinal Logistic Regressions (Johnston)

Regression modelling for dependent variables which are ordinal (i.e. survey outcomes with Likert Scales). Assumptions of the ordinal logistic model (parallel regression/proportional odds assumption). Calculation of fitted probabilities regarding the influence of independent variables on ordinal dependent variables.

Core Reading:

- Long, S. and Freese, J. (2006), Chapter 5
- Lab Reading: Lesson 13 in STATA Manual

Supplementary Readings:

- Hood, M.V. and Morris, I.L. (1998) “Give us your tired, your poor,... but make sure they have a Green Card: The effects of documented and undocumented migrant context on Anglo opinion towards immigration”, *Political Behavior*. Vol. 20, No. 1: Pg. 1-15

Lab Exercises:

- Applications of ordinal logistic modelling with survey data
- Testing the parallel regression assumption
- Stereotype logistic model as an alternative to ordinal logistic models
- Calculating fitted probability tables and applying these to graphics

Week 9: Multinomial Logistic Regressions (Johnston)

Regression modelling for dependent variables which are nominal (i.e. categorical data whose values have no numerical meaning). Assumptions of the multinomial logistic model (independence of irrelevant alternatives assumption). Interpretation of beta coefficients for multinomial logit given a baseline category. Calculation of fitted probabilities regarding the influence of independent variables on non-ranked categorical dependent variables.

Core Reading:

- Long, S. and Freese, J. (2006), Chapter 6

Supplementary Readings:

- Alvarez, M., Nagler, J., and Bowler, S. (2000) “Issues, Economics and the Dynamics of Multiparty Elections: The British 1987 General

Election” *American Political Science Review*. Vol. 94, No. 1: 131-149.

- Dow, J. and Endersby, J. (2004) “Multinomial probit and multinomial logit: a comparison of choice models for voting research” *Electoral Studies*. Vol 23: 107-122.

Lab Exercises: (*Fourth homework due at the beginning of Friday’s lab*):

- Applications of multinomial logistic modelling with survey data
- Testing the independence of irrelevant alternatives assumption via a Hausman test
- Calculating fitted probability tables and applying these to graphics

Week 10: Best practice and review (Both)

Discussion of best practices for writing empirical papers and a review of non-linear models.

Core Reading:

- Studenmund Ch. 11
- Gerner Hariri, J. (2012) “The Autocratic Legacy of Early Statehood”, *American Political Science Review*. 106(3): 471-494.

Lab Exercises

- Lab used for final exam review