

POLS6482: Advanced Multivariate Statistics

M, W: 10:00-11:30AM, Room: PGH 405

University of Houston

Department of Political Science

Fall 2017

Instructor

Ling Zhu

Assistant Professor

Department of Political Science

Email: lzhu4@central.uh.edu

Office: PGH 436

Phone: 713-743-2649

Office Hours: M, W: 1:00-4:00 pm, or by appointment.

Teaching Assistant

Ryan Jewell

Ph.D. Candidate

Department of Political Science

Email: rmjewell@uh.edu

Office: PGH 426

Office Hours: M, W: 1:00-3:00 pm, or by appointment.

Course Description

This is the third (full) course in quantitative methods in the University of Houston's political science Ph.D. Program. This course introduces students to a number of new and useful generalized linear models, with an emphasis on likelihood-based methods. We emphasize in this course that good social science research involves an appropriately developed theory and a correctly specified statistical model to map the underlying theory. Maximum likelihood (ML) offers a variety of models to evaluate uncertainty. Most of our attention will be given to models where the traditional assumptions of ordinary least square (OLS) regression are violated, because the dependent variable is non-continuous. Among the topics we cover are logit and probit models for both binary and ordinal dependent variables, discrete choice models for multiple alternatives (e.g. voting for multiple candidates or parties), event count models (e.g. international conflicts in a decade, presidential appointments during an administrative term, congressional hearings in a year, etc.), models for survival (time-to-event) data, and models for non-random selection (e.g. when you observe voters' preferences, but not non-voters').

Learning Objectives

This course will prepare students to read and critically evaluate quantitative political science research that employs maximum likelihood estimation. Furthermore, this course will prepare students to perform research of their own with attention to data management, data visualization, model specification, diagnostic analysis, post-estimation analysis, and the presentation of results. The goal of this course is to train students to produce independent, original research, and to provide a good foundation for students to undertake dissertation research.

Prerequisites

The background required for the course is a good introduction to probability and statistical inference, and at least one good regression course. Some familiarity with linear algebra and calculus is also assumed.

Textbooks and Additional Readings

Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage Publications. (An excellent introduction to the basic models covered in this course, very clearly presented.)

Faraway, Julian J. 2006. *Extending the Linear Model with R: Generalized Linear, Mixed Effects, and Nonparametric Regression Models*. Boca Raton, FL: Chapman&Hall/CRC. (An accessible introduction to various generalized linear models with R.)

Box-Steffensmeier, Janet M. and Bradford D. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*. New York, NY: Cambridge University Press. (An excellent text on event history analysis.)

(Recommended) King, Gary. 1989. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. University of Michigan Press. (The book that made maximum likelihood required study for political science. This is the Michigan Press reprint of King's original Cambridge University Press Book.)

(Recommended) Long, J. Scott and Jeremy Freese. 2014. *Regression Models for Categorical Dependent Variables Using Stata*. College Station, TX: Stata Press. (A well-stocked toolbox of how to implement MLE models using Stata.)

(Recommended) James E. Monogan III. 2015. *Political Analysis Using R*. New York, NY: Springer. (A nice overview of how R can be useful in the analysis of public administration, public policy, and political science data.)

Additional readings as necessary, all of which will be available on Blackboard Learn and/or through JSTOR.

Software, Statistical and Otherwise

You can use whatever statistical software you choose to complete the homework exercises, so long as the manner by which your results are generated and conclusions are transparent. However, due to the limits of lecture and laboratory time, my teaching assistant and I will only support two software packages, R and **Stata**. Both are available on the machines in the political science computing lab. All lectures and statistical labs will be instructed with an emphasize on R.

R

R is a statistical environment and high-level programming language for data analysis and visualization. It is the GNU version of the S language. R is a free and open-source software. The current (June 2017) version of R is 3.4.1. R is an object-oriented language, unlike **Stata** and most other statistical packages, it operates by assigning values to objects in the workspace. In the lecture notes, handouts, etc., R commands and outputs will be marked by “R Code” and “R Output”, respectively.

```
-----R Code-----  
Age<-cps2011$age  
summary(Age)  
-----R Output-----  
Min. 1st Qu.  Median    Mean 3rd Qu.  Max.   
0.00  15.00   34.00   34.99  52.00   85.00
```

The [Comprehensive R Archive Network \(CRAN\)](#) is the to-go spot for all things R-related. You may also consider using [RStudio](#), which is an integrated development environment for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management.

Stata

At the present time, **Stata** is probably the most widely-used statistical package in the social science. It is a powerful tool for data management, analysis, display, and boasts some of the best manuals and online help for any existing software package. **Stata** is commercial software package. The current version of **Stata** is 15.0, but previous versions (back to v.9, at least) can also be used for this class. In the class notes, handouts, etc., **Stata** command will be preceded by a period (“.”).

```
. logit Y X, vce(robust)
```

There are a number of useful **Stata** references on the web, including the [Stata](#) homepage, [Scott Long’s page](#) at IU, and an excellent **Stata** “[help page](#)” sponsored by UCLA.

Grading

Grades will be based on multiple assignments. First, homework problem sets will be set up to use either the R language or the **Stata** statistical package. I encourage collaborative work on problem

sets. The goal of homework problem sets is to help you to learn the materials and enable you to master various statistical models covered in this course. There is also a required term paper. The term paper must focus on a substantive question that is related to your main research area, and applies the statistical techniques covered in this class. You should format your term paper by following the [APSA Style Manual for Political Science \(Revised 2006\)](#). The term paper counts for 40% of the final grade and the exercises (weighted equally) together account for 40% of the final grade. At the end of the semester, we will hold a mini-conference to give you an opportunity to present your term paper and draw feedback from your peers. Each working paper will be assigned with a peer “reviewer” (i.e. a fellow student from this class). The peer-reviewer will provide a one-page review of the working paper. Your research presentation will account for 10% of the final grade. Your peer-review and regular class-participation together will account for 10% of the final grade.

The work you turn in (both exercises and term paper) should be professional quality. That means you **MUST** use proper mathematical notation. If you use **Microsoft Word**, I insist that you master their equation editors to make professional looking papers. Likewise, you must learn how to include graphics and edit tables properly within your word processor. You would be **MUCH** better off learning **L^AT_EX** and using that, but this is not a requirement. If at all possible, you should convert your Word document to Adobe Acrobat .pdf files before submitting them.

Five exercises: 40%

Term paper: 40%

Research presentation: 10%

Participation and peer review: 10%.

Final Grades:

A	=	100-95 (Excellent)
A-	=	94-90
B+	=	89-87 (Good)
B	=	86-84
B-	=	83-80 (Fair)
C+	=	79-77
C	=	76-74 (Poor)
C-	=	73-70
D+	=	69-67
D	=	66-64
D-	=	63-60
F	=	59-0 (Failing)

Attendance and Late Policy

1. Attendance. Absence from the class will only be accepted in extenuating circumstances with a university-accepted excuse. If you know in prior that you cannot attend the class, please inform the professor before the scheduled class-time. If there is an emergency, in which you need to leave early, you should avoid disturbing the lectures.

2. Late Policy. Late work (i.e. exercises and research paper) will be penalized by one letter grade for each day after the deadline. E.g. an A-paper turned in one day late will become a B-paper. Late work will only be accepted without grade-penalty if you have a university-accepted excuse.

Academic Integrity

As commonly defined, presenting the words or works of others' as your own is plagiarism. Plagiarism is one of the worst academic sins, for the plagiarist destroys the trust among colleagues, without which research cannot be safely communicated. Plagiarism is also a violation of the UH Academic Honesty Policy. If you are uncertain of what constitute academic dishonesty, you should contact me prior to submitting the assignment and/or check the UH Academic Honesty Policy from the university website: http://www.uh.edu/provost/policies/honesty/_documents-honesty/academic-honesty-policy.pdf. Students are expected to adhere to the UH Academic Honesty Policy. Cheating or plagiarism in course assignments, exams, and the final paper will lead to a grade of F.

Americans with Disability Act (ADA)

The Americans with Disabilities Act (ADA) is a federal anti-discrimination statute that provides comprehensive civil rights protection for persons with disabilities. Among other things, this legislation requires that all students with disabilities be guaranteed a learning environment that provides for reasonable accommodation of their disabilities. If you believe you need special accommodations and assistance due to a disability, please contact the Center for Students with DisABILITIES (CSD Building 568, Room 110) and the Learning Support Services (LSS, 321 Social Work Building), or call 713-743-5411 to make appropriate arrangements.

University of Houston CAPS Statement

Counseling and Psychological Services (CAPS) can help students who are having difficulties managing stress, adjusting to college, or feeling sad and hopeless. You can reach CAPS (www.uh.edu/caps) by calling 713-743-5454 during and after business hours for routine appointments or if you or someone you know is in crisis. No appointment is necessary for the "Let's Talk" program (www.uh.edu/caps/outreach/lets_talk.html), a drop-in consultation service at convenient locations and hours around campus.

Course Calendar

Week 1: Introduction

- August 21 Course Overview
Readings
 - Syllabus
- August 23. Lab 1: Getting Started with R
Readings
 - Lab 1 Handout: Getting Start with R.

- Faraway Chapter 4.
- (Recommended) Long&Freese, Chapters 1-2.

Week 2: Maximum Likelihood: Derivation and Properties

- August 28. Maximum Likelihood: Derivation and Properties
Readings
 - Long, Chapters 1-2
 - Aldrich, John. 1997. “R. A. Fisher and the Making of Maximum Likelihood 1912-1922.” *Statistical Science* 12(3): 162-176.
 - (Recommended) Fisher, Ronald A. 1925. “Theory of Statistical Estimation.” *Proceedings of the Cambridge Philosophical Society* 22: 700-725.
 - (Recommended) Neyman, Jerzy. 1934. “On the Two Different Aspects of the Representative Methods: The Method of Stratified Sampling and the Method of Purposive Selection.” *Journal of the Royal Statistical Society* 97(4): 558-625.
 - (Recommended) King 1989, Chapter 1-3 (skim), 4.1-4.3.
- August 30. Lab 2: Data Visualization with R (by Ryan Jewell)
 - Lab 2 Handout: Data Visualization Using R
 - Introduction to L^AT_EX

Week 3: Maximum Likelihood Estimation

- September 4. Labor Day Holiday, no class.
- September 6. MLE: Normal and Heteroskedastic General Linear Models
Readings
 - Long, Chapter 3.6.
 - (Recommended) King, Chapters 4.1-4.5.
 - (Recommended) Long&Freese, Chapter 3
 - Franklin, Charles H. 1991. “Eschewing Obfuscation? Campaigns and the Perception of Senate Incumbents.” *American Political Science Review* 85:1193-1214.
 - Lab 3: Estimating a Linear Regression Model Using MLE.

Week 4: Binary Response Models

- September 11. Binary Response Models: Fundamentals
Readings
 - Long, Chapters 3.
 - Esarey, Justin and Andrew Pierce. 2012. “Assessing Fit Quality and Testing for Misspecification in Binary-Dependent Variable Models.” *Political Analysis* 20(4): 480-500.
 - Herron, Michael C. 2000. “Postestimation Uncertainty in Limited Dependent Variable Models.” *Political Analysis* 8(1): 83-98.

- (Recommended) Berry, William D., Jacqueline H.R. Demeritt and Justin Esarey. 2010. “Testing for Interaction in Binary Logit and Probit Models: Is a Product Term Essential?” *American Journal of Political Science* 54(1): 248-266.
- (Recommended) Zhu, Ling and Christine Lipsmeyer. “Policy Feedback and Economic Risk: The Influence of Privatization on Social Policy Preferences.” *Journal of European Public Policy*, forthcoming.
- September 13. Lab 4: Estimating a Logit Model
Readings
 - Lab 4 Handout: Estimating Binary Response Models.
 - Faraway, Chapter 2.
 - (Recommended) Long&Freese. Chapters 5-6.

Week 5: Binary Response Models: Advanced Topics

- September 18. Binary Response Models: Advanced Topics
Readings
 - Long, Chapter 4.
 - (Recommended) King, Chapter 5.3.
 - Nagler, Jonathan. 1994. “Scobit: An Alternative Estimator to Logit and Probit.” *American Journal of Political Science* 38(1): 230-255.
 - Alvarez, R. Michael and John Brehm. 1995. “American Ambivalence Toward Abortion Policy: A Heteroskedastic Probit Method for Assessing Conflicting Values.” *American Journal of Political Science* 39(4): 1055-1082.
 - King, Gary and Langche Zeng. 2001. “Logistic Regression in Rare Events Data.” *Political Analysis* 9:137-163.
- September 20. Lab 5: Estimating Logistic Regression with Skewed Data.
readings
 - Lab 5 Handout: Binary Response Models: Advanced Topics.
 - King, Gary and Langche Zeng. 2001. “Explaining Rare Events in International Relations.” *International Organization* 53(3): 693-715.
 - (Recommended) King, Gary, Michael Tomz, and Jason Wittenberg. 2000. “Making the Most of Statistical Analyses: Improving Interpretation and Presentation.” *American Journal of Political Science* 44(2): 347-361.

Week 6: Ordered Response Models

- September 25. Ordinal Responses *Readings*
 - Long, Chapter 5.
 - (Recommended) King, Chapter 5.4.
 - Jones, Bradford S. and Michael E. Sobel. 2000. “Modeling Direction and Intensity in Semantically Balanced Ordinal Scales: An Assessment of Congressional Incumbent Approval.” *American Journal of Political Science* 44(1):174-185.

- Franklin, Charles H. and Liane C. Kosaki. 1989. “Republican Schoolmaster: The U.S. Supreme Court, Public Opinion, and Abortion.” *American Political Science Review* 83(3):751-771.
- Espenshade, Thomas J. and Haishan Fu. 1997. “An Analysis of English Language Proficiency among U.S. Immigrants.” *American Sociological Review* 62(2): 288-305.
- Sanders, Mitchell. 2001. “Uncertainty and Turnout.” *Political Analysis* 9(1):45-57.
- September 27. Lab 6: Models for Ordinal Outcomes
Readings
 - Lab 6 Handout: Ordinal Response Models.
 - Faraway, Chapter 5.3
 - (Recommended) Long&Freese, Chapter 7.

Week 7: Multinomial Choice Models

- October 2. Multiple Choice Models
Readings
 - Long, Chapter 6.
 - Born, Richard. 1990. “Surge and Decline, Negative Voting, and the Midterm Loss Phenomenon: A Simultaneous Choice Analysis.” *American Journal of Political Science* 34(3): 615-645.
 - Whitten Guy D. and Harvey D. Palmer. 1996. “Heightening Comparativists’ Concern for Model Choice: Voting Behavior in Great Britain and the Netherlands.” *American Journal of Political Science* 40(1):231-260.
 - Dow, Jay K. and James W. Endersby. 2004. “Multinomial Probit and Multinomial Logit: A Comparison of Choice Models for Voting Research.” *Electoral Studies* 23(1):107-122.
- October 4. Lab 7: Multinomial Logistic and Probit Regression
Readings
 - Lab 7 Handout: Multinomial Logit and Probit Models
 - Faraway, Chapter 5
 - (Recommended) Long&Freese, Chapter 8

Week 8: Censored and Truncated Variables

- October 9. Models for Censored and Truncated Variables
Readings
 - Long, Chapter 7.
 - Sigelman, Lee and Langche Zeng. 1999. “Analyzing Censored and Sample-Selected Data with Tobit and Heckit Models.” *Political Analysis* 8(2):167-182.
- October 11. Lab 8: Censored Data and Data with Sample Selection
Readings
 - Lab 8 Handout: Models for Censored and Sample Selected Datas

Week 9: Event Count Models

- October 16. Event Count Outcomes

Readings

- Long, Chapter 8.
- (Recommended) King, Chapters 5.5-5.10.
- King, Gary. 1988. “Statistical Models for Political Science Event Counts: Bias in Conventional Procedures and Evidence for Exponential Poisson Regression Model.” *American Journal of Political Science* 32(3):838-863.

- October 18. Lab 9: Poisson Regression Model

Readings

- Lab 9 Handout: Event Count Models
- Faraway, Chapter 3
- (Recommended) Long&Freese, Chapters 9.1-9.3

Week 10: Event Count Models: Advanced Topics

- October 23. Generalized Event Count Models

Readings

- (Recommended) King, Chapters 8.3, 9.4-9.5.
- King, Gary. 1989a. “Variance Specification in Event Count Models: From Restrictive Assumptions to a Generalized Estimator.” *American Journal of Political Science* 33(3): 762-784.
- King, Gary. 1989b. “A Seemingly Unrelated Poisson Regression Model.” *Sociological Methods and Research* 17(3): 235-255.
- King, Gary. 1989c. “Event Count Models for International Relations: Generalization and Applications.” *International Studies Quarterly* 33 (2):123-147.
- (Recommended) Winkelmann, Rainer. 2000. “Seemingly Unrelated Negative Binomial Regression.” *Oxford Bulletin of Economics and Statistics* 62(4): 553-560.

- October 25. Lab 10: Event Count Models: Advanced Topics

Readings

- Lab 9 Handout: Event Count Models: Advanced Topics
- (Recommended) Long&Freese, Chapters 9.4, 9.5-9.7

Week 11: Duration Analysis: Introduction

- October 30. Fundamentals of Duration Analysis

Readings

- Box-Steffensmeier, Janet M. and Bradford D. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*, Chapters 1-3.
- Allison, Paul D. 2014. *Event History and Survival Analysis*. Second Edition. Thousand Oaks, CA: Sage Publication, Chapters 1-2.

- Alt, James and Gary King. 1994. “Transfers of Governmental Power: The Meaning of Time Dependence.” *Comparative Political Studies* 27(2): 190-210.
- Bienen, Henry and Nicolas van de Walle. 1992. “A Proportional Hazard Model of Leadership Duration.” *The Journal of Politics* 54(3): 685-717.
- McCarty, Nolan and Rose Razaghian. 1999. “Advice and Consent: Senate Responses to Executive Branch Nominations.” *American Journal of Political Science* 43(4): 1122-1143.
- November 1. Lab 11: Cox’s Proportional Hazards Model
Readings
 - Lab 11 Handout: Survival Analysis: The Cox Regression Model
 - Box-Steffensmeier, Janet M. and Bradford D. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*, Chapters 4-5.
 - Teachman, Jay D. and Mark D. Hayward. 1993. “Interpreting Hazard Rate Models.” *Sociological Methods and Research* 21(3):340-371.

Week 12: Duration Analysis: Advanced Topics

- November 6. Duration Analysis: Diagnostics, Etc.
Readings
 - Box-Steffensmeier, Janet M. and Bradford D. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*, Chapter 8.
 - Box-Steffensmeier, Janet M. and Christopher Zorn. 2001. “Duration Models and Proportional Hazards in Political Science.” *American Journal of Political Science* 45(4): 972-988.
- November 8. Lab 12: Duration Models: Testing for Non-proportional Hazards and Duration Dependence
Readings
 - Lab 12 Handout: Survival Analysis: Diagnostics, Etc.
 - Keele, Luke. 2010. “Nonproportionally Difficult: Testing for Nonproportional Hazards in Cox Models.” *Political Analysis* 18(2):189-205.
 - Zorn, Christopher. 2000. “Modeling Duration Dependence.” *Political Analysis* 8(3):367-380.

Week 13: Survival Analysis: Extensions

- November 13. Discrete Time Alternatives and Frailty Models
Readings
 - Beck, Nathaniel, Jonathan N. Katz, and Richard Tucker. 1998. “Taking Time Seriously: Time-Serious-Cross-Section Analysis with a Binary Dependent Variable.” *American Journal of Political Science* 42(4):1260-1288.
 - Signorino, Curt and David Carter. 2010. “Back to the Future: Modeling Time Dependence in Binary Data.” *Political Analysis* 18(3): 271-292.
 - Box-Steffensmeier and Jones, Chapters 9, 11.

- Carpenter, Daniel. 2002. “Groups, the Media, Agency Waiting Costs and FDA Drug Approval.” *American Journal of Political Science* 46(3):490-505.
- November 15. Lab 13: Survival Analysis: Advanced Topics
Readings
 - Lab 13 Handout: Survival Analysis: Extensions.
 - Banerjee, Sudipto, Melanie M. Wall, and Bradley P. Carlin. 2003. “Frailty Modeling for Spatially Correlated Survival Data, with An Application to Infant Mortality in Minnesota.” *Biostatistics* 4(1):123-142.
 - Faraway, Chapters 8-9

Week 14: (Week of November 20) Thanksgiving Break (No Class)

- Full draft of term paper due on Blackboard Learn by Sunday (November 26), by the end of the day (5:00pm).

Week 15: Models with Nonrandom Selection

- November 27. Selection Bias: Fundamentals
Readings
 - Hackman, James J. 1979. “Sample Selection Bias as a Specification Error.” *Econometrica* 47(1):153-161.
 - Dubin, Jeffery A. and Douglas Rivers. 1989. “Selection Bias in Linear Regression, Logit and Probit Models.” *Sociological Methods and Research* 18(2-3): 360-390.
 - Timpone, Richard J. 1998. “Structure, Behavior and Voter Turnout in the United States.” *American Political Science Review* 92(1):145-158.
 - Lassen, David Dreyer. 2005. “The Effects of Information on Voter Turnout: Evidence from a Natural Experiment.” *American Journal of Political science* 49(1):103-118.
- November 29. Lab 14: Dealing with Selection Bias and Unobserved Heterogeneity
Readings
 - Lab 14 Handout
 - Alvarez, R. Michael and Garrett Glasgow. 1999. “Two Stage Estimation of Nonrecursive Choice Models.” *Political Analysis* 8(2): 147-165.

Week 16: Student Research Presentations

- December 4-6. Student Research Presentations
- December 11. Final paper is due on Blackboard Learn by the end of the day (5:00pm).

Caveat: The aforementioned weekly schedule and assignments in this course may be subject to change.