

# Quantitative Analysis in Political Science II

## Generalized Linear Models

Political Science 582, Friday 1-3, Location: Seigle 205.

### Course Description

This course extends what you did in previous methods courses by focusing on nonlinear model forms for the outcome variable. These are typically called “generalized linear models,” although for historical reasons people in political science call them “maximum likelihood models.” The principle we will care about is how to adapt the standard linear model that you know so that a broader class of outcome variables can be accommodated. These include: counts, dichotomous outcomes, bounded variables, and more. There is a strong theoretical basis for the models that we will use. Also, the bulk of the learning in the course will take place outside of the classroom by reading, practicing using statistical software, replicating the work of others, and doing problem sets. Keep in mind that the skills attained in this course are those that the discipline of political science expects of any self-declared data-oriented researcher.

The second aspect of the course is focused on the statistical package R which is completely free for downloading for Mac, Unix, Linux and that other platform at CRAN, the Comprehensive R Archive Network. R is an implementation of the S language, which is the default computational tool for research statisticians. Quite simply R is the most powerful, extensively featured, and capable statistical computing tool that has ever existed on this planet. And as mentioned, its free. We will not use Stata; don’t ask.

### Prerequisite Details

The only official prerequisite for this course is QPA I. However, each student should be familiar with: basic probability theory, statistical inference, hypothesis testing, and least squares estimation. The course will also assume a working knowledge of calculus and linear algebra at the level of Essential Mathematics for Political and Social Research. Jeff Gill, 2006, Cambridge University Press. Knowledge of R is assumed.

## Course Grade

The final grade will be based on three components: problem sets (40%), a replication assignment (30%), and an exam (30%) on MLE theory and basic models. The exam covers material from the first 7 weeks of the course plus the assigned readings (Faraway and articles). Consequently, we will discuss the readings in as much detail as the class desires. The problem sets will be a combination of analytical and computational assignments and given in each meeting. See Alicia Uribe's tips on success with the problem sets. For the replication assignment, find a published work in your field of interest, obtain the data, and exactly replicate the author's model results. It is usually easier to find an article that uses the readily available datasets in the discipline (COW, ANES, GSS, etc.), but some authors are forthcoming about distributing their data if asked. The relevant model should be one of the nonlinear forms studied in this course. Gary King has some useful tips and links to his PS paper on the subject here, and a recent success story (publication) is described by two of his students here. All submitted work must be from LaTeX source.

## Office Hours

Friday 9-10.

## Incompletes

None given.

## Teaching Assistant

Jonathan Homola, [homola@wustl.edu](mailto:homola@wustl.edu). Office hours: Monday 3-5, Wednesday 3-5, in Seigle 277.

## Homework

Assigned each day and due the following week at classtime. No late homework accepted. All homework must be LaTeX'd.

## Required Text

Title: Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models.

Author: Faraway.

Publisher: Chapman & Hall/CRC.  
ISBN: 158488424.

Also: “Practical Regression and Anova using R” by Faraway (from PS581) , and  
“All of Statistics: A Concise Course in statistical Inference” by Larry Wasserman. Springer,  
2004, ISBN: 978-0387402727.

## Optional Texts

(These are for background; see me before making any purchases.)

Title: A Guide to Econometrics.  
Author: Kennedy.  
Publisher: MIT Press, 2003.  
Edition: Fifth or Sixth.  
ISBN: 0-262-61183-X.

Title: Generalized Linear Models: A Unified Approach.  
Author: Gill.  
Publisher: Sage, 2001.  
Edition: First.  
ISBN: 0761920552.

Title: Modern Applied Statistics with S.  
Author: Venables and Ripley  
Publisher: Springer-Verlag, 2003.  
Edition: Fourth.  
ISBN: 0387954570.

Title: Linear Models with R.  
Author: Faraway.  
Chapman & Hall/CRC  
Edition: First.  
ISBN: 1-58488-425-8.

## List of Topics/Dates

### 1. September 2. No Class (APSA meeting).

Reading:

- Leamer, Let's Take the Con Out of Econometrics.
- How Not to Lie With Statistics, by Gary King and Ellie Powell.
- TPM (The Political Methodologist) Volume 11, No. 2, articles: (1) Jackman, (2) Anderson, et al., (3) Gill (pages 20-26). Available at: The Society for Political Methodology.
- Wasserman Chapter 7.
- R code from the lecture.

Homework:

- Problem Set #1.
- Wasserman exercises 7.3 and 7.10.

### 2. September 9. Uncertainty, Inference, and Hypothesis Testing.

Reading:

- Faraway, Chapter 1.
- R Tutorial.
- Gill, The Insignificance of Null Hypothesis Significance Testing.
- McCloskey, The Loss Function Has Been Mislaid: The Rhetoric of Significance Tests.
- Tressoldi et al. High Impact = High Statistical Standards? Not necessarily so.
- Wetzels, et al. Statistical Evidence in Experimental Psychology An Empirical Comparison Using 855 t-Tests.

Homework:

- Problem Set #2.

### 3. September 16. The Likelihood Model of Inference.

Reading:

- Faraway, Appendix A.

- Wasserman page 122-134.
- Binomial PMF likelihood grid search Model syntax summary.
- R code from the lecture.
- slides from the lecture.

Homework:

- Problem Set #3.
- Wasserman exercise 9.3.

#### 4. September 23. Models for Dichotomous Outcomes.

Reading:

- Faraway, Chapter 2.
- Wasserman pages 223-226.
- Altman, The cost of dichotomising continuous variables.
- R Code for Chapter 2, draft chapter.

Homework:

- Faraway, Chapter 2, Exercises 1-7. For Exercise 2.2, download the `wbca.txt` data from <http://www.maths.bath.ac.uk/~jjf23/ELM/>. Also for Exercise 2.2, do not use the step function in part (b), use your own intuition).
- Wasserman exercise 13.9.
- Find a dataset with a dichotomous outcome that you are interested in. Run an appropriate glm model submit the output with a paragraph defending the model fit.

#### 5. September 30. Models for Count Outcomes.

Reading:

- Faraway, Chapter 3.
- Poisson Example.
- Negative Binomial Example.
- R code for Chapter 3.
- Statistical Models for Political Science Event Counts: Bias in Conventional Procedures and Evidence for the Exponential Poisson Regression Model, Gary King.

Homework:

- Faraway Chapter 3, Exercises 1-7.

6. October 7. Models for Contingency Tables.

Reading:

- Faraway, Chapter 4.
- R code for Chapter 4.
- contrasts.

Homework:

- Faraway, Chapter 4, Exercises 1-7.

7. October 14. Models For Ordered Categorical Data.

Reading:

- Faraway, Chapter 5.
- Multinomial Probit and Logit: A Comparison of Choice Models for Voting Research, by Jay K. Dow and James W. Endersby.
- R code for Chapter 5.

Homework:

- Faraway Chapter 5, Exercises 1-6.
- Consider a proportional odds model using the logit link function with only one explanatory variable in addition to the constant. Express the odds ratio (i.e. not-logged) for a one-unit change in the explanatory variable. What does this simplify to?

8. October 21. Models for Unordered Categorical Data.

9. October 28. How to Handle Missing Data in Models. The EM Algorithm and Multiple Imputation.

Reading:

- mice: Multivariate Imputation by Chained Equations by Stef van Buuren and Karin Groothuis-Oudshoorn.
- Multiple Imputation in R.

- R code from the lecture.

Homework:

- Problem Set #4.

10. November 4. Exam On Fundamentals

Homework:

- Turn in an electronic copy of your replication data and one regression model using these data (necessary to sit for the midterm).

11. November 11. The GLM Theory and the Exponential Family Form.

Reading:

- Faraway, Chapter 6.
- R code from the lecture.
- Bloodpressure data.
- GLM Chapter (Sage).
- The Epic Story of Maximum Likelihood, by Stephen M. Stigler.

Homework:

- Faraway Chapter 6, Exercises 1-5.

12. November 18. Nonparametric Regression, Additive Models.

Reading:

- Faraway, Chapters 11-12.
- R code. Dust data. GAM test data.
- Some lecture slides.
- Generalized Additive Models, by Trevor Hastie and Robert Tibshirani.

Homework:

- Faraway Chapter 11, Exercises 1-5; Faraway Chapter 12, Exercises 1-5.

13. November 25. Thanksgiving Holiday.

14. December 2. Other GLMs, Quasi-Likelihood Estimation.

Reading:

- Faraway, Chapter 7.
- R code from the lecture. Scottish voting data. Ship data.
- Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method, by R.W.M. Wedderburn.

Homework:

- Faraway Chapter 7, Exercises 1-7.

15. December 9. **Submission and Presentation of Replications.**