

Advanced Social Statistics (Sociology 8811)

Zack W. Almquist
Spring Semester, 2015

Class Schedule

Lecture: M/W 1:00 – 2:15pm Social Science 1114

Lab: M 03:35 – 05:30pm Bleg Hall 440

URL: <http://moodle.umn.edu>

Note: Requires UMN login and registration in class to access.

Professor

Name: Zack W. Almquist
Office: 960 Social Science Building
Office Hours: 5:00-6:00 TH
Email: almquist@umn.edu
Telephone: 612-624-4300 (not recommended)

Teaching Assistant

Name:
Office:
Office Hours:
Email:

Course Objectives

SOC 8811 covers statistical methods for analyzing social data. This course is designed for Sociology graduate students and assumes a background equivalent to Soc 5811 Intermediate Social Statistics, i.e., linear models, ANOVA and descriptive statistics. The class will be comprised primarily of introduction to modern statistical techniques such as categorical data analysis (e.g., logistic regression), time series analysis (e.g., event history analysis), and modern computational statistics (e.g., monte carlo tests). Labs are organized to help

students with the data analysis required to complete the exercises, develop the term paper, and to further training in statistical software used by social science researchers.

Prerequisites

SOC 5811 or equivalent; students are assumed to have an elementary understanding of the basic concepts of probability and statistics. English language proficiency appropriate to a graduate university class is assumed.

Course Requirements

Computers

It is not required that students bring their computers/laptops to lecture and lab (if one is owned), but it is *highly* recommended since both lecture and lab will make extensive use of the computer software R. Computer labs are available on campus, please consult with the TA if you have trouble finding the various locations that computer labs reside on campus.

Readings

Weekly readings are assigned on the course syllabus. All readings are assumed to be completed before each lecture/seminar. You are expected to read over the class notes each week and make sure you are familiar with the material as the course progress – questions are encouraged.

Homework

Homework assignments will be administered on a semi regular basis for the first 13 weeks of the course. Homework assignments are meant to achieve three results: (1) provide practice with the computational and statistical programming language R and (2) provide practice with the statistical concepts discussed in class and (3) provide a chance to demonstrate your mastery of material and highlight areas where more work is needed. You may work in a group, but all write-ups must be done independently. All collaborators should be appropriately cited in your write up and any detailed R code should also be cited.

Proposal

In the ninth week of class you are expected to submit a research proposal of no more than 2 pages. This proposal should consist of the following elements: (1) introduction to the problem, (2) hypothesis to be tested, (3) methods to use, (4) expected results.

Report

The research report should mimic a standard empirical article or technical report. This should be no more than 20 pages (excluding references and figures). Note that **maximum** is **not a minimum**, well written 10 page papers will receive higher marks than a poorly written 20 page paper. This should follow the same basic structure as the proposal, but fully flushed out and with careful analysis and conclusions. Your report must come with an Appendix which includes all R code used in the analysis of this project.

Presentations

Presentations will be 15 minutes in length with 5 minutes of questions for a total of 20 minutes. Presentations should be in standard conference format (e.g., American Sociological Association). Points will be deducted if the presentation goes more than 17 minutes. Presenters will be graded on the following standards: (1) composition, (2) slides, (3) organization, and (4) general quality and presentation skills.

Participation

Individuals are expected to attend every course, to have completed every reading and to participate with questions and discussion on each topic as presented. If you plan on missing any class period you are responsible for all material and for contacting the instructor in a timely manner.

Grading

Participation:	10%
Homework:	40%
Project Proposal:	10%
Project:	30%
Presentation:	10%

Lectures, readings, labs, and review sessions are provided for each student's benefit. It is the responsibility of the student to take advantage of these opportunities to acquire and demonstrate mastery of course material, so as to achieve his or her desired grade.

Letter grade assignment

A	93%+
A-	90-92.99%
B+	87-89.99%
B	83-86.99%
B-	80-82.99%
C+	77-79.99%
C	73-76.99%
C-	70-72.99%
D	60-69.99%
F	<59.99%

Required Texts

- John Fox. 2008. *Applied Regression Analysis and Generalized Linear Models*. 2nd ed. Sage.
- John Fox and Sanford Weisberg. 2011. *An R Companion to Applied Regression* 2nd ed. Sage.
- Peter Dalgaard. 2008. *Introductory Statistics with R*. 2nd ed. Springer. (*Downloadable from the library.*)
- Norman Matloff. 2011. *The Art of R Programming: A Tour of Statistical Software Design*. No Starch Press.
- Zack Almquist. 2015. *Notes for Soc 8811*.

Recommended Resources and Books for R and Statistics

- Julian J. Faraway. 2006. *Extending the Linear Model with R*. Chapman & Hall.
- Göran Broström. 2012. *Event History Analysis with R*. Chapman & Hall.
- <http://stackoverflow.com/>
- Journal of Statistical Software
- Springer Series UseR!

Readings

Be prepared to discuss all readings assigned at anytime in lecture/seminar.

Required Software

We will be using the R statistical programming language. R can be downloaded at <http://www.r-project.org/>.

Recommended Software

RStudio IDE (Integrated Development Environment) is a software application which facilitates interaction with the R statistical programming language. It is often preferred to the GUI (Graphic User Interface) made available through CRAN. You can download it at <http://www.rstudio.com/>.

Course Policies

Missing Class, etc.

It is expected that each member of the class will attend every lecture/discussion. If there is an appropriate reason to miss class it is expected that the individual will email or discuss in person with the instructor at least one week in advance. For any medical issues please see the UMN website for university policies.

Cheating, etc.

All work is assumed to be your own and all individuals are expected to follow the university policy on cheating and misconduct. If you have any questions please consult the UMN website for university policies.

Readings and Assignments

Week 1: Computational Social Science and R

- Introduction to R
- Introduction to Computer Science Concepts and R Data Structures
- *Readings:*
 - Matloff (2011), Chapters 1-6 (*Recommendation:* Attempt all the examples)
 - Fox & Weisberg (2011), Chapters 1.1
 - *Case Studies*
 - * Matthew J. Salganik, Peter S. Dodds, and Duncan J. Watts (2006). “Experimental study of inequality and unpredictability in an artificial cultural market.” *Science*, 311:854–856.
- *Assignments:*
- *Lab:*
 - Lab # 1

Week 2: Computational Social Science and R

- Simple Programming, Graphics and Manipulating Data
- *Readings:*
 - Matloff (2011), Chapters 7-10 (*Recommendation:* Attempt all the examples)
- *Assignments:*
 - Handout Homework # 1
- *Lab:*
 - Lab # 2

Week 3: Computational Social Science and R

- Simple Programming, Graphics and Manipulating Data Continued
- *Readings:*
 - Matloff (2011), Chapters 11-13 (*Recommendation:* Attempt all the examples)
- *Recommended:*
 - Matloff (2011), Chapters 14-16
 - Fox & Weisberg (2011), Chapter 2
- *Assignments:*
- *Lab:*
 - Lab # 3

Week 4: Review

- Review of Univariate Statistics and Linear Regression in R
- *Readings:*
 - Fox (2008), Chapters 1-6
 - Dalgaard (2008), Chapters 4-6
 - Fox & Weisberg (2011), Chapters 1.2-1.5, 3
- *Assignments:*
 - Homework # 1 Due
 - Handout Homework # 2
- *Lab:*
 - Lab # 4

Week 5: Statistical Theory for Linear Models

- Statistical Theory for Linear Models
- *Readings:*
 - Fox (2008), Chapters 9-10
 - Fox & Weisberg (2011), Chapters 4.1-4.2
- *Assignments:*
- *Lab:*
 - Lab # 5

Week 6: Model Diagnostics

- Model Diagnostics
- *Readings:*
 - Fox (2008), Chapters 11-13
 - Fox & Weisberg (2011), Chapters 4.3-4.9
 - *Case Studies*
 - * Adrian Raftery (1995). “Bayesian model selection in social research” (with Discussion). *Sociological Methodology*, 25:111–196.
- *Assignments:*
 - Homework # 2 Due
 - Handout Homework # 3
- *Lab:*
 - Lab # 6

Week 7: Maximum Likelihood and Optimization Theory

- Maximum Likelihood and Optimization Theory
- *Readings:*
 - Almqvist (2015), Notes on MLE and Optimization
 - Fox & Weisberg (2011), Chapters 5.10
- *Assignments:*
- *Lab:*
 - Lab # 7

Week 8: Logistic Regression

- Introduction to Logistic Regression
- *Readings:*
 - Fox (2008), Chapters 14
 - Fox & Weisberg (2011), Chapters 5.1-5.3.1
 - *Case Studies*
 - * Mullen, Ann L., Kimberly A. Goyette, and Joseph A. Soares (2003). "Who goes to graduate school? Social and academic correlates of educational continuation after college." *Sociology of Education*, 76: 143–169.
- *Assignments:*
 - Project Proposal Due
 - Homework # 3 Due
 - Handout Homework # 4
- *Lab:*
 - Lab # 8

Week 9: Logistic Regression

- Introduction to Logistic Regression Continued
- *Readings:*
 - Fox (2008), Chapters 14
 - Fox & Weisberg (2011), Chapters 5.3.2-5.4
 - *Case Studies*
 - * Kalleberg, Arne L., Barbara F. Reskin, and Ken Hudson (2000). "Bad jobs in America: Standard and nonstandard employment relations and job quality in the United States." *American Sociological Review*, 65: 256–278.
- *Assignments:*
- *Lab:*
 - Lab # 9

Week 9: Generalized Linear Model

- GLM General Forms and Extensions
- *Readings:*
 - Fox (2008), Chapter 15
 - Fox & Weisberg (2011), Chapters 5.5-5.9
 - *Case Studies*
 - * Haynie, Dana L (2001). “Delinquent Peers Revisited: Does Network Structure Matter?” *American Journal of Sociology*, 106, 4:1013–1057.
- *Assignments:*
 - Homework # 4 Due
 - Handout Homework # 5
- *Lab:*
 - Lab # 9

Week 10: Time-Series Regression

- Time-Series Regression and Generalized Least Squares
- *Readings:*
 - Fox (2008), Chapter 16
 - Almqvist (2015), Notes on Time-Series
 - *Case Studies*
 - * Isaac, Larry W., and Larry J. Griffin (1989). “Ahistoricism in time-series analyses of historical process: Critique, redirection, and illustrations from US labor history.” *American Sociological Review*: 873-890.
- *Assignments:*
- *Lab:*
 - Lab # 10

Week 11: Nonlinear Regression

- Introduction to Nonlinear Regression
- Almquist (2015), Notes on Nonlinear Regression
- *Readings:*
 - Fox (2008), Chapter 17
- *Assignments:*
 - Homework # 5 Due
 - Homework # 6
- *Lab:*
 - Lab # 11

Week 12: Missing Data

- Missing Data in Regression Models
- *Readings:*
 - Fox (2008), Chapter 21
 - Almquist (2015), Notes on Missing Data
 - *Case Studies*
 - * Little, Roderick JA, and Donald B. Rubin (1989). “The analysis of social science data with missing values.” *Sociological Methods & Research* 18: 292–326.
- *Assignments:*
- *Lab:*
 - Lab # 12

Week 13: Model Selection

- Model Selection, Averaging and Validation
- *Readings:*
 - Fox (2008), Chapter 22
 - Fox & Weisberg (2011), Chapters 4.5-4.9, 6

- *Case Studies*

- * Adrian Raftery (1995). “Bayesian model selection in social research” (with Discussion). *Sociological Methodology*, 25:111–196.

- *Assignments:*

- Homework # 6 Due
- Handout Homework # 7

- *Lab:*

- Lab # 13

Week 14: Advanced Topics and Presentations

- Spatial Models, Panel Data Models, Bayesian Data Analysis, etc.

- Presentations Day 1.

- *Readings:*

- Almquist (2015), Notes on Advanced Topics

- *Case Studies*

- Almquist, Zack W. and Carter T. Butts (2013). “Dynamic Network Logistic Regression: A Logistic Choice Analysis of Inter- and Intra-group Blog Citation Dynamics in the 2004 US Presidential Election.” *Political Analysis*, 21(4), 430-448.

- *Assignments:*

- *Lab:*

Week 15: Presentations

- Presentations Day 2.

- Presentations Day 3.

- *Assignments:*

- Homework # 7 Due

- *Lab:*

Week 16

- *Assignments:*
 - Project Due