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: By appointment
Email: almquist@umn.edu
Telephone: 612-624-4300 (not recommended)

Teaching Assistant

Name: Yang Yang
Office: Ford Hall 496
Office Hours: TU 10:00am-11:00am
Email: yang3175@umn.edu

Course Objectives

SOC 8811 covers statistical methods for analyzing social data such as online behavioral
data, survey data or experimental 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 primar-
ily 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 com-
putational statistics (e.g., monte carlo tests). Students will receive both theoretical and
technical training in statistics through the use of the R statistical programing language,
and practical hands on data analysis experience with common social science data sets such
as the General Social Survey and US Census administrative data. 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 of campus computer labs.

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.

Labs & Homework

Homework and labs will be assigned every week for every Monday for the first 12 weeks and will be due one week later. Each assignment will be comprised of a lab component and homework component. The lab component will consist of computing exercises in R and \LaTeX, and the homework component will consist of writing up and interpreting the labs in the manner of a technical report or journal article. Homework and lab assignments are meant to achieve three results: (1) provide practice with the computational and statistical programing 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 R code from a given source should also be cited as you would a journal article.
Proposal

In the eighth 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, and (4) expected results.

Report

The research report should mimic a standard empirical article or technical report and will be due the last day of finals week. 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 an additional 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 of material presented; (2) slides, e.g., quality, detail, readability; (3) organization, and (4) general presentation skills.

Participation

Individuals are expected to attend every course, to have completed every reading and to participate with questions and discussions 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

<table>
<thead>
<tr>
<th>Component</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participation</td>
<td>10%</td>
</tr>
<tr>
<td>Homework</td>
<td>40%</td>
</tr>
<tr>
<td>Project Proposal</td>
<td>10%</td>
</tr>
<tr>
<td>Project</td>
<td>30%</td>
</tr>
<tr>
<td>Presentation</td>
<td>10%</td>
</tr>
</tbody>
</table>

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


Recommended Resources and Books for R and Statistics

Below is a list of resources to help you become more familiar with R. The first three are text-books with a good collection of examples (Fox and Sandy) and detailed statistical theory (Julian J. Faraway) which the you might find as useful supplement to the course material. Göran Broström provides a detailed text on how to perform event history analysis in R. I have listed three common internet-based resources for improving your ability to manipulate and program in R. The first is stackoverflow a crowsource website for all problems related to mathematics, computers science, statistics, etc. and two peer-reviewed publishing outlets for R statistical software, the Journal of Statistical Software and Springer Series UseR! series:

- [http://stackoverflow.com/](http://stackoverflow.com/)
- [Journal of Statistical Software](http://www.jstatsoft.org/)
- [Springer Series UseR!](http://www.springer.com/series/6991)
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
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
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
  - Fox & Weisberg (2011), Chapters 1.1

Week 2: Computational Social Science and R

- Simple Programing, Graphics and Manipulating Data
- Readings:
  - Matloff (2011), Chapters 7-10
- Case Studies
- Assignments:
  - Lab # 1
  - Homework # 1

Week 3: Computational Social Science and R

- Simple Programing, Graphics and Manipulating Data Continued
- Readings:
  - Matloff (2011), Chapters 11-13
- Recommended:
  - Matloff (2011), Chapters 14-16
  - Fox & Weisberg (2011), Chapter 2
- Case Studies
- Assignments:
  - Lab # 2
  - Homework # 2
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
- Case Studies
- Assignments:
  - Lab # 3
  - Homework # 3

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 # 4
  - Homework # 4

Week 6: Maximum Likelihood and Optimization Theory

- Maximum Likelihood and Optimization Theory
- Readings:
  - Almquist (2015), Notes on MLE and Optimization
  - Fox & Weisberg (2011), Chapters 5.10
- Assignments:
  - Lab # 5
  - Homework # 5
Week 7: Logistic Regression

- Introduction to Logistic Regression

- Readings:
  - Fox (2008), Chapters 14
  - Fox & Weisberg (2011), Chapters 5.1-5.3.1
  - Case Studies

- Assignments:
  - Lab # 6
  - Homework # 6
  - Project Proposal Due!

Week 8: Logistic Regression

- Introduction to Logistic Regression Continued

- Readings:
  - Fox (2008), Chapters 14
  - Fox & Weisberg (2011), Chapters 5.3.2-5.4
  - Case Studies

- Assignments:
  - Lab # 7
  - Homework # 7
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

- Assignments:
  - Lab # 8
  - Homework # 8

Week 10: Model Diagnostics

- Model Diagnostics

- Readings:
  - Fox (2008), Chapters 11-13
  - Fox & Weisberg (2011), Chapters 4.3-4.9
  - Case Studies

- Assignments:
  - Lab # 9
  - Homework # 9
Week 11: Time-Series Regression

- Time-Series Regression and Generalized Least Squares

- Readings:
  - Fox (2008), Chapter 16
  - Almquist (2015), Notes on Time-Series
  - Case Studies

- Assignments:
  - Lab # 10
  - Homework # 10

Week 12: Nonlinear Regression

- Introduction to Nonlinear Regression

- Almquist (2015), Notes on Nonlinear Regression

- Readings:
  - Fox (2008), Chapter 17

- Assignments:
  - Lab # 11
  - Homework # 11
Week 13: Missing Data

- Missing Data in Regression Models

- Readings:
  - Fox (2008), Chapter 21
  - Almquist (2015), Notes on Missing Data
  - Case Studies

- Assignments:
  - Lab # 12
  - Homework # 12

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

Week 15: Presentations

- Presentations Day 2.
- Presentations Day 3.

Week 16: Presentations

- Assignments:
  - Project Due: Last day of finals.