Intermediate statistical modeling for language science

Instructor: Amelia E. Kimball
Email: akimbal2@illinois.edu. Email is the best way to reach me. I respond to student emails within 24 hours on weekdays.
Office: FLB 4036
Office hours: Monday & Wednesday 2-3pm. I am available and willing to meet students outside scheduled office hours by appointment. Request an appointment by email.

Description: This course provides hands-on experience with the most common statistical analyses for quantitative data in the language sciences. This course will be most beneficial for advanced graduate students who have collected their own data and can analyse that data for the final project. The following statistical models (including their assumptions, when to use them, and how to implement them in the R statistical programming language) will be addressed: (1) ANOVA (2) Linear Regression (3) Generalized Linear Models (Poisson, logistic, multinomial and ordinal) (4)(Generalized) Linear Mixed-Effects Models. As time permits and student interest dictates, there may also be opportunities to investigate: Generalized Additive Models, Smoothing Spline ANOVA

Prerequisites: This course is designed for graduate students undertaking quantitative research in language science. A basic course in statistics, research methods, or computer programming is a suggested prerequisite. Students with no experience in statistics, research methods, or programming must get instructor permission to take this course. The course will be implemented using R programming software.

Required text: *Discovering Statistics Using R* by Field, Miles, and Field (available at the university bookstore). The software package "R" will also be required; it is free and available for all computing platforms.

Assessment:
Problem sets: 40% (4 x 10% each)
Problem set exercises will be based on data provided by the instructor. Students are allowed to work together on the homework, but each student must write up her/his own assignment.

Take-home Midterm exam:20%
The Midterm exam will be a take-home exam consisting of an analysis of a data set including descriptive statistics, visualizations of descriptive statistics, either ANOVA or linear regression, and visualizations of results. Students will be assigned a data set by the instructor and must choose the correct analysis for that data set. The exam will be passed in on the course website, in the form of programming code in a .R file. Students will be graded on their choice of

test, their presentation of the data, and to a lesser extent on using good programming technique.

Critical review of a published article: 10%
Each student will read an article that uses one of the statistical models presented in this class and critique the analysis within the article. Critiques should address: whether the analysis was appropriate, whether the assumptions of the test used were met, whether the information provided was sufficient for the reader to replicate the analysis, and what the student would suggest doing differently, if anything. A list of suggested articles is provided on the course website.

Final Project: Research proposal/ Analysis of research data 20%
For the final project students will analyse a data set of their choosing, either from their own research or with the instructor's help. Students will

Attendance and participation 10%
Active participation in class is encouraged and expected. Attendance will be taken by the instructor or via written notes from small group activities conducted in class. Participation includes coming prepared to class and contributing through speaking in class.

Ethical behavior is expected: Students are bound by the Student Code regarding plagiarism. No plagiarism will be tolerated, and students found plagiarizing will be given a minimum penalty of receiving zero credit for the plagiarized assignment. Additionally, all research must be approved by the Institutional Review Board, and all data provided for the final project must be in anonymized, shareable format with no individually identifiable information.

Course Schedule:
The course textbook by Field, Miles, and Field, is referred to as FMF below

Week	Topic	Readings
Week 1	Introduction	
Week 2	Descriptive statistics: Population parameters, sample statistics.	FMF chapter 2
Week 3	Introduction to R, Linear relationships	FMF chapter 3 chapter 6.3-6.7
Week 4	The linear model	FMF chapter 7.7, chapter 5.6-5.7,
Week 5	The linear model continued	chapter7.8-10
Week 6	Categorical data, regression coding	FMF chapter 7.12
Week 7	binary data and logistic regression	FMF chapter 7.12, Chapter 8.2-8.9

Week 8	Logistic regression II/ generalized linear models	FMF chapter 18.3-18.12
Week 9	Linear mixed effects models	FMF chapter 19.1-19.4
Week 10	Choosing random effects structure	Barr et al (2013) (available on course website)
Week 11	Generalized mixed effects models	
Week 12	Intro to GAMMs/ SSANOVA, time permitting.	