

LING 792 Quantitative Methods in Linguistics, Winter 2017

Instructor: Jonathan Brennan

- **Time** 2:00 to 5:00 pm, Mondays
- **Location** 473 Lorch Hall
- **Website** Via Canvas
- **Office Hours** T 9:30am-11:30am, W 1:30pm-3:30pm @ 414 Lorch Hall
(click to schedule office hours visit)
- **Email** jobrenn@umich.edu

Course Description

This course offers a practical introduction to the quantitative analysis of linguistic data. We will discuss best practice in four domains: (1) building statistical models appropriate for categorical and continuous linguistic data, including (generalized) linear mixed models, (2) designing simulations to test models and check assumptions, (3) visualizing data for exploration and for publication, (4) and constructing analyses optimized for reproducibility and sharing. Class time will be a mix of mini-lectures and hands-on activities. Course assignments will include problem sets, lab reports, and two data analysis reports based on data from outside the class. You will learn to use the free and popular R statistical package along with several extensions, including lme4, GGplot, R Markdown, Shiny, and “the tidyverse”. Prior experience with statistics and/or with R is beneficial, but is not a requirement.

Course Goals

To develop expertise and best practice habits in quantitative data management, analysis, and reporting for linguistics.

Recommended Prerequisites

This course **does not assume** any statistical background. However, we will be moving quickly in the first few sessions to establish foundations with basic

statistics, and with some particular tools for statistical analysis. Here's what you can do to be prepared:

- Install R and RStudio on your computer. (These will also be available on desktop computers in the classroom).
- If you have not used R before, then **I highly recommend that you complete the following on-line tutorials:** Sessions 1 (“Intro to Basics”) and 2 (“Vectors”) of the free Introduction to R produced by DataCamp.

Readings

Readings will be a mix of book excerpts, journal articles, and web resources. They will be made available on Canvas. In addition, there are several good books and course notes that cover similar (but not the same) material as will be covered in the course. Below, I've highlighted **two optional books** and **one useful website**. Other books and resources will be highlighted throughout the term.

Optional: Gelman, A & Hill, J (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge (ISBN: 978-0521686891) [Link](#)

Optional: Vashishth, Shravan (n.d.) Course Notes: Introduction to Statistical Analysis

Optional: Johnson, Keith (2008) *Quantitative Methods in Linguistics*. Wiley-Blackwell (ISBN: 978-1-4051-4424-7) [Link](#)

##Assignments & Grading

Participation & Discussions (10%)

Much of the intellectual work in the class hinges on working through problems individually, in small groups, or as a full class. Attendance and engagement are crucial. One class meeting can be missed without penalty.

In addition, students are invited to post questions or comments to the Canvas discussion forums *at least one hour prior to class*. These questions will help to shape in-class discussions and activities.

Labs (30%)

Labs will be weekly in-class assignments that you submit via Canvas. These will develop quantitative analysis, visualization, and reporting skills. You will have time in-class to work through these exercises, which will generally be woven together with group activities and lecture. You will submit your lab via Canvas by 12 pm on the Wednesday following the class meeting.

When evaluating these, the most important question is *Did you make a thoughtful attempt at each component of the in-class assignment?* Showing your work is an key part of this. Getting the “right” answer is less important. I do not expect

that you will need time outside of class to complete your work. The lowest two scores, including incompletes, will be dropped.

Problem Sets (30%)

Problem sets will be bi-weekly take-home assignments that invite you to explore concepts introduced in class and to push your analytic skills further. These will be due via Canvas on Fridays at the end of the day on the weeks indicated in the syllabus. The lowest score, including incompletes, will be dropped.

Data Analysis Reports (30%)

You will complete two comprehensive data analysis reports. One will be due about half-way through the term, and one will be due at the end of the term.

These will be long-form assignments based on data that you bring in from outside the class. The data may be collected by you or may be taken from another source, such as an open data repository (with appropriate permissions). **Your report will be presented in R Markdown.**

Each report will include:

1. A research question or hypothesis along with citations to selected background literature (no extensive literature review or discussion)
2. A succinct description of the relevant measures and how the data were collected
3. A prose statistical analysis plan
4. Descriptive statistics of the data
5. Exploratory visualizations (graphs)
6. Hypothesis-based statistical analysis of the effects of interests
7. Visualizations for the effects of interest
8. Simulation(s) to check for statistical reliability or to evaluate the adequacy of the statistical tool(s) used
9. Prose description of the primary statistical results
10. Brief discussion of how the results bear on the research question

The first report should address continuous or binomial data that are separated into at most two variables (e.g. comparing two conditions, or comparing two conditions in two groups).

The second report should address continuous or binomial data that require the modeling of continuous linear or non-linear relationships and/or covariates. **Optional** Take a look at the tools for building so-called Shiny apps in RStudio and consider building a simple one to allow the reader to explore some key aspects of the data and statistical analysis.

Late Policy

Labs, problem sets, and *first* lab report:

- >0-2 days: 25% penalty
- 2-7 days: 50% penalty
- >7 days: not accepted for credit.

Second lab report: late work will not be accepted for credit

Grading Scale

	A	B	C	D	F
+	97-99	87-89	77-79	67-69	
	93-96	83-86	73-76	63-67	<60
-	90-92	80-82	70-72	60-62	

Schedule

The schedule below sketches where the course will go. It is almost sure to change during the term as the interests of the class become more focused. Be sure to check Canvas for updates!

Readings are posted on Canvas or via links that are embedded below.

Unit 1: Simulations and Data Management

- **Jan 09: Week 1**
 - Introduction; Binomial events; Simulations; R Basics
 - No reading
- **Jan 16: No Class, MLK Day**
 - Problem Set 1 due Friday Jan 20
- **Jan 23: Week 2**
 - Continuous data; Normal distribution; T-tests, ANOVA in R; Basic plotting
 - Read Vasishth Ch. 2
 - Problem set 2 due Friday Jan 27
- **Jan 30: Week 3**
 - Correlations and Linear regression; Data transformations; Tidy data; GGplot
 - Read Gelman & Hill Ch. 3 (Optional: Ch. 4)
- **Feb 6: Week 4**
 - Regression continued
 - Problem set 3 due Friday Feb 10
- **Feb 13: Week 5**
 - Logistic regression; Generalized linear models
 - Gelman & Hill Ch. 5

Unit 2: Linear mixed modeling

- **Feb 20: Week 6**
 - Group analysis and repeated measures; Linear mixed modelling
 - Read Baayen et al., 2008
 - *First data analysis report due Friday Feb 25*
- **Feb 27: Spring Break**
- **Mar 6: Week 7**
 - Advanced linear and logistic mixed models
 - Read Jaeger et al., 2008
 - Problem set 4 due Friday Mar 10
- **Mar 13: Week 8**
 - Structuring complex models; checking model quality with simulations
 - Read Barr et al., 2013

Unit 3: Advanced topics

- **Mar 20: Week 9**
 - Hypothesis testing, contrasts, and p-values
 - Read Gelman & Stern 2006
 - Review slides from Gillespie & Jaeger 2013
 - Problem set 5 due Friday Mar 25
- **Mar 27: Week 10**
 - Exploratory and confirmatory analyses; Forking paths
 - Read Gelman & Loken 2014 and Gelman & Loken 2017
 - Optionally read Simmons et al., 2011
- **Apr 03: Week 11**
 - Data management revisited
 - Read Fruewald’s Data Management Notes and This data analysis example from Wei, Brennan & Boland (under review)
- **Apr 10: Week 12**
 - Power analysis revisited; Inferences based on the null; Frequentist vs. Bayesian inference;
 - Read Sprouse & Almeida in press and Gelman 2014 (3 pp)
 - Optionally read Gelman & Garlin 2013
 - Problem set 6 due Friday Apr 15
- **Apr 17: Week 13**
 - Wrap-up discussion
 - *Second analysis report due Monday April 17*

Possible topics not included in the schedule above

(This list will be updated during the term)

- More on generalized linear models: Poisson, bias-correction etc.
- Mixed effects modeling with Stan
- Time-series and GA(M)Ms

Email Policy

The best place to ask questions is via the Canvas Discussion forum where I as well as fellow students may post a reply. This open forum allows for easy follow-up discussion and makes your contribution available to other students who may have similar questions. I will monitor the forum to ensure answers are clear and accurate. E-mail should only be used for personal concerns. I will respond by email within 24 hours on weekdays.

LSA Community Standards of Academic Integrity

The LSA undergraduate academic community, like all communities, functions best when its members treat one another with honesty, fairness, respect, and trust. The College holds all members of its community to high standards of scholarship and integrity. To accomplish its mission of providing an optimal educational environment and developing leaders of society, the College promotes the assumption of personal responsibility and integrity and prohibits all forms of academic dishonesty and misconduct. Academic dishonesty may be understood as any action or attempted action that may result in creating an unfair academic advantage for oneself or an unfair academic advantage or disadvantage for any other member or members of the academic community. Conduct, without regard to motive, that violates the academic integrity and ethical standards of the College community cannot be tolerated. The College seeks vigorously to achieve compliance with its community standards of academic integrity. Violations of the standards will not be tolerated and will result in serious consequences and disciplinary action. (<http://www.lsa.umich.edu/academicintegrity/>, Dec 21, 2012)

Accommodations for Students with Disabilities

If you think you need an accommodation for a disability, please let me know at your earliest convenience. Some aspects of this course, the assignments, the in-class activities, and the way the course is usually taught may be modified to facilitate your participation and progress. As soon as you make me aware of your needs, we can work with the Office of Services for Students with Disabilities (SSD) to help us determine appropriate academic accommodations. SSD (734-763-3000; <http://ssd.umich.edu>) typically recommends accommodations through a Verified Individualized Services and Accommodations (VISA) form. Any information you provide is private and confidential and will be treated as such.

Syllabus Updated: March 8, 2017