

**LING 660: Research Seminar in Sociolinguistics
Spring 2017**

COURSE TIME	Mon. 1-3 pm
LOCATION	3401-C Walnut, Room 326
WEBSITE	https://canvas.upenn.edu/courses/1345696
INSTRUCTOR	Dr. Meredith Tamminga
CONTACT	tamminga@ling.upenn.edu
OFFICE	3401-C Walnut #316
OFFICE HOURS	By appointment

OVERVIEW

The topic of this research seminar for Spring 2017 is **tools for quantitative data analysis**. The seminar has two goals: 1) to equip participants with a comprehensive set of skills to implement common quantitative analyses of linguistic data, and 2) to introduce debates around best practices in statistical analysis for linguistics. In-class activities will be a mix of paper discussion and practical tutorials.

COURSE REQUIREMENTS

This seminar has no official statistics prerequisites. If you have not taken either STAT 500 or STAT 501 (or equivalent) you may need to do extra background reading to catch up and keep up. If, on the other hand, you already have more extensive statistics background, you may sometimes need to be patient and actively support your classmates' learning.

There will be several small assignments during the course of the semester, which all participants should plan to do. Participants taking the course for credit will additionally submit a term project, the nature of which will be decided on a case-by-case basis. Please consult with me early in the semester to get a project plan approved.

SCHEDULE OF TOPICS AND READINGS

Jan. 11 **Manipulating quantitative data in R**
 ****Monday classes on a Wednesday****

Jan. 16 **No class - MLK Jr. Day**

UNIT ONE: Issues in the responsible analysis of quantitative data

Jan. 23 **Misconceptions and misuses of p-values**

- Nuzzo. 2014. Statistical errors. *Nature* 506: 150-152.
- Goodman. 2008. A dirty dozen: twelve p-value misconceptions. *Seminars in Hematology* 45: 135-140.
- Wasserstein & Lazar. 2016. The ASA's statement on p-values: context, process, and purpose. *The American Statistician* 70(2): 129-133
 - Each student to pick one supplemental commentary to read and report on to the class
- Gelman & Stern. 2006. The difference between "significant" and "not significant" is not itself statistically significant. *The American Statistician* 60(4): 328-331.

Jan. 30 **Researcher degrees of freedom**

- Simmons, Nelson & Simonsohn. 2011. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science* 22(11): 1359-1366.
- Head, Holman, Lanfear, Kahn & Jennions. 2015. The extent and consequences of p-hacking in science. *PLoS Biology* 13(3): e1002106. doi:10.1371/journal/pbio.1002106
- Nosek, Spies, & Motyl. Scientific utopia II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science* 7(6): 615-631.

Feb. 6 **Statistical power**

- Hallahan & Rosenthal. 1996. Statistical power: concepts, procedures, and applications. *Behavior Research and Therapy* 34(5/6): 489-499.
- Green & MacLeod. 2016. SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7: 493-498.
- Lakens. 2014. Performing high-powered studies efficiently with sequential analyses: Sequential analyses. *European Journal of Social Psychology*, 44(7), 701-710.

Feb. 13 Trimming and transforming data

- Cox. 2005. Transformations: an introduction. <http://fmwww.bc.edu/repec/bocode/t/transint.html>
- Ratcliff. 1993. Methods for dealing with reaction time outliers. *Psychological Bulletin* 114(3): 510-532.
- Baayen & Milin. 2010. Analyzing reaction times. *International Journal of Psychological Research* 3(2): 12-28.
- Gelman & Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Sections 4.1-4.5. Cambridge University Press.

UNIT TWO: Practical considerations in regression modeling

Before the beginning of this unit, each student needs to choose a “pet” data set for use throughout the unit. The data should be linguistic in nature, may have a continuous or binary dependent variable, and should include at least two categorical and two continuous predictors. You may use a data set of your own or find one somewhere else.

Feb. 20 Practical intro to linear and logistic regression

- Gelman & Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Chapters 3 & 5. Cambridge University Press.

Feb. 27 Contrast specification and interpreting interactions

- Gillespie. 2013. Coding categorical predictors. Lecture for LI539 at LSA 2013 Summer Institute.
- Gelman & Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press. Revisit Chs.3-5 with special attention to sect. 3.3, sect. 4.4, and p. 93-95.

Mar. 6 No class - spring break**Mar. 13 Mixed effects modeling: the basics**

- Gelman & Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press. Sections 1.1-1.3 and Chapters 11-12.

Mar. 20 Mixed effects modeling: current debates

- Barr, Levy, Scheepers & Tily. 2013. Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language* 68: 255-278.
- Bates, Kliegel, Vasishth & Baayen. 2015. Parsimonious mixed models. arXiv: 1506.04967.

Mar. 27 Model comparison and multicollinearity

- Wurm & FisiCaro. 2014. What residualizing predictors in regression analysis does (and what it does *not* do). *Journal of Memory and Language* 72: 37-48.
- TBA?

UNIT THREE: Data visualization**Apr. 3 Principles of data visualization**

- Tufte. 1983. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press. Ch. 2 (Graphical Integrity), Ch. 4 (Data-Ink), and Ch. 5 (Chartjunk)
- Wickham. 2010. A layered grammar of graphics. *Journal of Computational and Graphical Statistics* 19(1): 3-28.

Apr. 10 Practical intro to ggplot2

- Fruehwald. 2012. Workshop on ggplot2. Advances in Visual Methods for Linguistics, University of York, Sept. 26. <https://jofrhwld.github.io/avml2012/>

UNIT FOUR: A peek outside of frequentist statistics**Apr. 17 Bayes' Rule**

- Krushke. 2015. *Doing Bayesian Data Analysis*. London: Academic Press. Ch. 2 (Introduction) and Ch. 5 (Bayes' Rule)

Apr. 24 Proving the null

- Gallistel. 2009. The importance of proving the null. *Psychological Review* 116(2): 439-453.