Ling 313: Quantitative Methods in Linguistic Research

Course Syllabus
Spring 2012

Class Meetings  Mon/Wed 3:10pm - 4:30pm, Vollum 302

Instructor       Jennifer Nycz
                 jnycz@reed.edu
                 Eliot 100A
                 Office hours: Mon 1-3pm & Tues 2:30-3:30pm

Prerequisites    Linguistics 211 (or equivalent), or instructor consent

Texts            There is no required textbook for this course. Readings will be drawn from a variety of sources, and will be distributed on Moodle.

Software        All students will be required to use R, a statistical software environment that can be downloaded from http://www.r-project.org/

Course Overview. Modern linguistic research increasingly involves quantitative analysis: we count or measure various aspects of linguistic behavior, and want to use this data to answer questions about the way language works. This course will guide students through the description, visualization, and basic statistical analysis of both discrete and continuous linguistic data using R, an open-source statistical software environment.

Students will develop skills throughout this course that will enable them to:
  • formulate clear and answerable research questions
  • make appropriate methodological choices when planning future research projects
  • establish good habits in the management of data, analytical work, and analysis results
  • summarize, visualize, and otherwise explore data using a variety of methods
  • understand the conceptual underpinnings of some common statistical tests, and apply them appropriately
  • critically evaluate quantitative analyses in a range of linguistic subfields
  • wield R for fun and profit

Course Requirements and Assessment

Participation (30%): You are expected to attend every class meeting and participate in classroom activities such as discussions, group problem-solving, and occasional quizzes. You may also be asked to participate in Moodle forums or other online activities.

Problem sets (50%): Five problem sets will be given. These will be handed out in class approximately every other week, and due by class time one week later; please submit hard copies of these assignments.

Assignments must be turned in on time to receive full credit and comments. Extensions will be granted only in cases of illness, family emergency, etc. Unexcused late assignments will be penalized 10% of total possible points.
for each day they are late. Late assignments will not be accepted for credit if turned in after the problem set has been returned to students, or discussed in class, whichever comes first.

You are encouraged to work on problem sets together, but your answers must be written up separately and in your own words.

**Final project (20%)**: Students will propose and complete a final project that demonstrates their grasp of the course material. Each student will give a short (~15 minute) presentation of the progress of their final project in the last week of classes. More details on this project will be distributed by Spring break; one-page project proposals will be due on **March 28th**, and the final write-up due on **May 7th**.

---

**Course Schedule**

**Week 1** (January 23 & 25): **Introduction to the course & Research design**.

Why quantitative methods? What are we trying to accomplish in this course? Fundamentals of research design: asking answerable questions.


**Week 2** (January 30 & February 1): **Getting started with R & Graphical exploration of data**.

Installing packages, loading data, manipulating variables and dataframes. Visualizing data: histograms, density plots, scatterplots, and other ways to see what’s going on in your data set.

- Fruehwald, Josef. Introduction & R Basics.

**Week 3** (February 6 & 8): **Descriptive statistics & Probability distributions**.


**Week 4** (February 13 & 15): **Drawing conclusions about the population from a sample: Counts**.


Week 5 (February 20 & 22): **Drawing conclusions about the population from a sample: Measures.**

Inferential statistics for (1 or 2) measures. t tests (all flavors). parametric v. non-parametric tests. Reporting results.


---

Week 6 (February 27 & 29): **Correlation and simple linear regression.**

Exploring the relationship between two measurement variables.


---

Week 7 (March 5 & 7): **Power & Plots.**

Making sure you collect enough data so you can answer the question you’re asking. Best practices in the graphical presentation of results.


Recommended:

---

**Spring Break!**

---

Week 8 (March 19 & 21): **Analysis of variance & the Kruskal-Wallis test.**

Comparing measures across more than two groups. Post hoc tests.


---

Week 9 (March 26 & 28): **Multiple regression.**

When you’ve got more than one predictor.

Week 10 (April 2 & 4): Logistic regression.

When you’ve got more than one predictor, and your dependent variable is nominal.


What do sociolinguistic data sets tend to look like? How do the particular characteristics of these data affect the analytical choices we make?


Week 12 (April 16 & 18): Issues in phonetic research.

Same questions as last week, different field: What do phonetic data sets tend to look like? What assumptions do they violate, and what are the implications for analysis?


Week 13 (April 23 & 25): Project presentations, final reflections & course wrap-up.

Final projects due Monday, May 7

Additional sources to check out:


Websites:

Handbook of Biological Statistics: http://udel.edu/~mcdonald/statintro.html
Statistics Hell: http://statisticshell.com
Joe Fruehwald’s R Study Group page: http://www.ling.upenn.edu/~joseff/rstudy/index.html
The Little Handbook of Statistical Practice: http://www.tufts.edu/~gdallal/LHSP.HTM