Statistics for Language Studies
Syllabus
Ling 270

May 8, 2018

This course is an introduction to statistics for language studies that develops and illustrates fundamental ideas in statistics with examples and applications from language. Through the lectures, readings, and assignments you will get experience in applying concepts and using statistical methods to collect, analyze, and interpret data from a diverse set of linguistic sources. We will emphasize those statistical concepts that facilitate linguistic explanation through the analysis of quantitative data.

1 Goals

The primary goal of the course is to acquaint students with the basic ideas of probability theory and to show them how those ideas can help us understand linguistic data and explain facts about variation in language, how language changes, and how humans process language.

This course is preparation for a linguistics major. It is also a Foundations GE course in the area of Natural Sciences and Quantitative Reasoning. Quantitative reasoning refers to a range of academic capacities that includes learning from data, communicating quantitatively, analyzing evidence and assertions, and employing quantitative intuition.

Goal One in this area is to apply appropriate computational skills and use basic mathematical concepts to analyze the natural world and solve problems. As your computational tool, you will use R, a statistics package, to
input, organize, and count your data, as well as to create meaningful statistical analyses. Goal Two in this area is to use methods of quantitative reasoning to solve and communicate answers to real-world problems. In this course you will use statistical models to describe dialect variation, explain the frequency distributions of words, and to predict speaker judgments of sentence grammaticality.

2 Practice

The course will use two required texts,


There will also be two online manuals. The second is required for only one assignment:


There will be weekly assignments and several inclass quizzes, a midterm, and a final exam.

3 Learning Outcomes

1. Use R to provide counts from linguistics data, construct contingency tables, compute basic descriptive statistics, plot probability distributions, and explore data graphically in other ways.

2. Conduct basic significance tests and explain the meaning of the results;
3. Construct a linear regression model for dependent variable and predictors, and describe what the model shows about each predictor;

4. Discriminate between different types of probability distributions, and test goodness of fit of data to distributions;

5. Distinguish between fixed effects and random effects, specifically in language data using multiple linguistic productions from multiple speakers with multiple linguistic forms.

4 Prerequisites

Course in intermediate algebra.

5 Grading

Grading will be based on assignments, quizzes, and a final exam.

- Assignments: 40%
- Quizzes: 20%
- Midterm 20%
- Final exam 20%

6 Late Assignments

The general structure of the course is not well-suited to late assignments or missed quizzes. Assignment solutions will be discussed in detail on the day they are turned in, and thus students who turn assignments in late will be at an advantage. Quizzes are designed to test understanding of foundations needed for further work, and without those foundations, progress will be slowed. However, to allow for some flexibility, late assignments will receive partial credit. Here is the lateness policy:

1. Up to one week late: 50% credit for assignment. Late assignments must include all problems for which solutions have not been posted in order to receive any credit at all.
2. More than one week late: not accepted.

7 Attendance

Attendance is not a formal part of your grade. However, be aware that your performance on assignments is a formal part of your grade, and extensive amounts of class time will be devoted to working through the steps required to complete assignments. Similarly, hints on how to complete parts of the assignments and answer questions the midterms are handed out liberally in class. These hints will not be posted on the web pages.

8 Accessibility

Students who need accommodation of their disabilities should contact me privately to discuss specific accommodations for which they have received authorization. If you have a disability, but have not contacted Student Disability Services at 619-594-6473 (Calpulli Center, Suite 3101), please do so before making an appointment to see me.
9 Course Outline

<table>
<thead>
<tr>
<th>Weeks</th>
<th>Topics</th>
<th>Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2</td>
<td>Probability and counting. Binomial distribution. Introduction to R</td>
<td>Data manipulation in R, counting, contingency tables, coin-tossing simulations.</td>
</tr>
<tr>
<td>3</td>
<td>Conditional probability, Bayes’ Theorem</td>
<td>Pencil and paper assignment.</td>
</tr>
<tr>
<td>4</td>
<td>Mean, Variance, Normal distributions</td>
<td>Vowel formant measurements assignment, measuring individual and community variation.</td>
</tr>
<tr>
<td>5</td>
<td>Other types of probability distributions</td>
<td>Quantile/quantile plots. Testing for Poisson distribution of “the”.</td>
</tr>
<tr>
<td>6, 7</td>
<td>Statistics and hypothesis testing</td>
<td>Bresnan et al. Dative alternation data (animacy and syntactic role in explaining grammaticality judgments). Written/spoken language assignment: Sentence length.</td>
</tr>
<tr>
<td>9</td>
<td>Word distributions, burstiness of language</td>
<td>Poisson distribution assignment redux: Testing for Poisson distributions of various words.</td>
</tr>
<tr>
<td>12, 13</td>
<td>Correlation and Regression</td>
<td>Vowel formant data assignment (variation between communities and vowel fronting).</td>
</tr>
<tr>
<td>14,15</td>
<td>Mixed models</td>
<td>Bresnan et al. data assignment redux: Controlling for lexical effects.</td>
</tr>
</tbody>
</table>