

## **Linguistics 330: Spring 2006**

### **Inductive Statistics**

Instructor:	Matt Goldrick	Janet Pierrehumbert
Email:	goldrick@ling.northwestern.edu	jbp@northwestern.edu
Phone:	(847) 467-7092	(847)491-8050
Office Hours:	T 12:30-1:30 and by appointment	by appointment
Office:	2016 Sheridan Rd. Room 22	2016 Sheridan Rd. Room 30

Class time: TTh 11-12:20

Class location: Annenberg 101 / Library Mac/PC lab

#### What's this course all about?

Learners—everyday people as well as scientists—make generalizations on the basis of incomplete data. Children learn to speak an infinite range of utterances based on a finite sample of sentences. Scientists make generalizations regarding people based on detailed observations of small groups participating in experiments. How can we specify these abilities in a precise way (using the language of mathematics)? We'll pursue this goal by answering three questions:

1. How does one determine the structure of a given set of behavioral data?
2. How can one determine the behavioral reflexes of some underlying physical (cognitive) process?
3. Do the predicted behavioral reflexes of some process accord with observed data?

#### What am I going to do in this class?

We'll work towards these goals by:

1. Using the R statistical package to analyze data. This is a powerful tool for visualizing and analyzing data. It's also an open-source, platform independent program, free to all users.
2. Reviewing (using R, readings and lectures) the theory and application of statistical methods that allow us to generate the behavioral reflexes of some underlying processes and assess how well these predictions accord with observed data.
3. Using statistical methods to explore real data sets. We'll go through a series of 5 laboratory analyses, drawing on real data sets from language structure at various level (from sound to meaning), music, and clinical populations.
4. Using these methods to examine data sets outside the course. Final project groups will critically examine, in light of the conceptual and methodological background of the course, new data sets and/or primary literature that analyze data. They will present their findings to the class and write up the results.

Prerequisites: One course at the 200 level in Linguistics, CSD, or Music Cognition, or the equivalent, or permission.

Webpage: Blackboard. You should be able to access it by going to [courses.northwestern.edu](http://courses.northwestern.edu). Let us know if you have trouble.

### Assessments: What You Need to Do

<u>Assessment Type</u>	<u>% of Final Grade</u>
Participation in and preparation for class meetings	30%
Labs (5)	60%
Final paper	
Presentation	2%
Paper	8%

1. Participation in and preparation for class meetings.
  - a. Participation will be assessed through attendance and ability to meaningfully participate in discussion. This doesn't mean you have to know everything before you come to class. Come prepared with clarification questions!
  - b. Preparation will be assessed in completion of readings and assignments from the textbooks.
2. Labs. These labs will be assigned throughout the quarter. You will have 1 week to complete each assignment. Due dates for the 5 labs are listed below:
  - Lab 1: Assigned 3/30, due 4/6
  - Lab 2: Assigned 4/11, due 4/18
  - Lab 3: Assigned 4/20, due 4/27
  - Lab 4: Assigned 5/2, due 5/9
  - Lab 5: Assigned 5/11, due 5/18
3. Final project. In a few weeks, you'll be given a list of broad topics to choose from. You can work on these in groups or individually. You will present your research on this topic in one of the last two class sessions and write up the work for a final paper.
  - Final presentations: 5/23, 5/25
  - Final papers due: 6/5

### Course Policies

1. You have to do your work on time. Due dates are firm; attendance in class is not optional. If this is a problem, see #2. The highest possible letter grade for work handed in during the first 24 hours after class will be a C. We will give comments and suggestions on work turned in later than this, but you will receive no credit for the assignment.
2. If you have a problem, give us sufficient time to help you! If you've tried to do an assignment, but can't seem to complete it, come to us well before it's due. If you are unable to come to a class, try to let us know beforehand, or as soon as possible after class. Your reasons for class absences need to be verifiable.
3. You should work together, but the finished product must be your own. Working together is a big part of our in-class work; we hope this will extend outside of the classroom. However, for most assignments, your written work must be your own.

### Tentative Course Plan

Textbooks: (downloadable from Blackboard)

**B:** Baayen, R. H. Practical Data Analysis for the Language Sciences with R

**V:** Vasishth, S. The foundations of statistics: A simulation-based approach

Date	Topic	Readings	Assignments
3/28	Basic R syntax; functions	B 1.1	
3/30			Lab 1 assigned: Distributions in speech and music
4/4	Visualization in R	B 1.2, 2.1	
4/6			Lab 1 due
4/11	Randomness + probability; Behavioral reflexes of simple mechanisms	V 2,3	Lab 2 assigned: Monte Carlo simulation of lexical acquisition
4/13	Reflexes of more complex mechanisms: Monte Carlo methods		
4/18	Assessing simple mechanisms	V 4	Lab 2 due
4/20			Lab 3 assigned: Generating behavior using parametric and non-parametric methods
4/25	Assessing complex mechanisms; efficiency vs. robustness	B 2.2	
4/27			Lab 3 due
5/2	Discrimination vs. significance	V 8	Lab 4 assigned: Disambiguation
5/4			
5/9	Clustering and classification	B 3	Lab 4 due
5/11			Lab 5 assigned: Semantic clustering
5/16			
5/18			Lab 5 due
5/23	Final project presentations		
5/25			
6/5	Final papers due, 5pm		