Quantitative Methods for Linguistic Research

LSA Linguistic Institute 2017, University of Kentucky

Professor: Natasha Warner

Class meets:
Tuesday/Friday 1:30-3:20, JSB 108

Email: nwarner@email.arizona.edu

Books/materials:
Some optional but possibly helpful readings are posted on the Canvas site.

General organization of the course:
This course is a statistics and experimental design course using examples from all areas of linguistics. The focus is on the concepts behind various statistical analyses, including enough math to understand where the statistical results come from, and what type of analysis to use in order to answer which questions. Topics include: hypothesis testing, power, effect size, one-factor designs, interaction, two-factor designs, understanding and interpreting a 3-way or higher level interaction, between-subjects vs. repeated measures factors, and random vs. fixed factors. (We may have to cut down on a few of those though, depending on time.) Each topic is addressed both from the perspective of ANOVA and from that of LME, but practice in how to do analyses will use ANOVA because of the short duration of the course. Many topics are easier to understand first as related to ANOVA, and then can be transferred directly to LME later. The course uses the R statistics language, but it is not a course in how to use R. No previous statistics experience is necessary.

While a large part of the class will be spent on basics of how to analyze quantitative data, another goal of the class is to examine the statistical analyses which appear in actual published linguistic literature. We can use this class to get a sample of what statistical methods are being used in which subfields of linguistics or in various journals currently.

There will be some in-class lab activities. These will be graded only for participation, and if you have to miss a class you can turn in a written version to count as participation. There will be two homework assignments focusing on concepts and interpretation of statistical results rather than on conducting the analysis yourself. An additional part of the homework assignments will give you optional ungraded opportunities for practice conducting the analyses. Each student will also create a bullet-point style report of approximately 1-2 pages summarizing current usage of statistical tests in a sub-field of linguistics or in one or more journals. These will be shared with the whole class.

Prerequisites:
The only math you will need will be addition, subtraction, multiplication, and division (just a whole lot of them!). No previous training in statistics is required. The only assumption about background is that you are currently or plan to eventually be involved with quantitative data in some linguistically related field. At this point, you may only be reading literature which uses statistical analyses, or you may already be collecting your own data and need to analyze it. However, if you do have past statistics experience, you are likely to still find new things to learn.

Requirements/grading:

In-class activity participation:	40% (divided among however many we do)
Homework assignments:	40% (20% each) (turn in as hard copy)

Handout on published stat. methods: 20%

Late policy: 10% reduction of the possible grade per day late, except for very serious excuses.

Tentative schedule:

Date	Topic	Outside of class requirements
Friday July 7	Introduction, 1 factor ANOVA with 2 levels, with 3+ levels, pairwise comparisons, power	
Tuesday July 11	2 factor ANOVA, interaction	
Fri. July 14	interaction continued, follow-up tests	Homework 1 due
Tues. July 18	3 factor ANOVA, higher designs	
Fri. July 21	within-subjects designs, interaction, follow-up tests	
Tues. July 25	mixed designs	Handout on statistics in a sub-field due
Fri. July 28	random factors, counterbalancing	
Tues. Aug. 1	by subjects and by items tests, comparison to LME	Homework 2 due

What to turn in for the handout on published statistical methods:
Please choose:

- one subfield of linguistics/language research (e.g. phonetics, psycholinguistics, discourse analysis, second language pedagogy, neurolinguistics), OR
- one or two methods (e.g. eye-tracking or analysis of a text corpus or EEG; or a comparison of two methods), OR
- a few journals in one subfield (e.g. Journal of the Acoustical Society of America, Journal of Phonetics, and Journal of the International Phonetic Association).

Only choose one of those options. For whichever set/area of research you choose, go find a good collection of articles in that area. You might, for example, take one year's publications on phonetics in JASA and one year's publications in J Phonetics. Or you might take a collection of say 15-20 recent papers on eye-tracking, distributed across several journals. The articles you choose should be relatively recent (published in the last 5-8 years), unless you specifically want to compare publications at earlier and later times. The total number of articles you look at should probably be around 15-20. State in your write-up what set of research you're looking at, and why you selected it (e.g. subfield, to compare two journals you use often).

You do not need to read all the articles! That would take way too long for this course. Just skim for what statistical tests they use. One article may use more than one. At the least, you can classify the statistical tests by overall type (e.g. t-tests, ANOVA, LME, GAMs, correlation, traditional multiple regression, chi-squared, other non-parametric tests, discriminant analysis, other interesting things like hierarchical clustering or multi-dimensional scaling). Count how often you see which type of test in which journal or year or subfield or whatever. You may have to google to find out what some tests are because we can't cover everything in class. Summarize number of articles using what, don't list every one.

You may be able to notice some interesting patterns. In phonetics, is LME more common when used by linguists or for more obviously linguistic questions like intonation than it is when used by speech and hearing people, or for more biological questions about glottal function or for clinical questions? Does the switch from ANOVA to LME as the main method depend on which journal or which year you're looking at? Are people using a particular type of method to answer a particular type of question? Is that because of the nature of the data (e.g. time series analysis for eye-tracking) or because of conventions in the field, as far as you can tell?

If you have time and room left in your write-up, you could also look into how people report a specific type of statistical test (in tables or prose? how long is the write-up of a particular model or statistical test? is there a standard convention across papers and authors, or is it completely variable? is there a separate section explaining the statistical methods?). You could look at what people use figures vs. tables for. You could look at how large/small sample size is for various types of research, and whether it has increased over time. Basically, we'll be interested in whatever you find out about how statistics are used and reported within the area you choose to look at.

The final product should be about 1-2 pages (single-spaced) of bullet-point style/handout style material. It should not be a prose term paper. You should post it on the Canvas site for everyone to learn from.

Various statements:

Students with Disabilities:

If you have a disability that affects how you will need to learn in this course, please discuss this with me.

Academic misconduct:

Please see the LSA's Ethics statement (https://lsa2017.as.uky.edu/sites/default/files/Ethics_Statement.pdf) and Sexual Harassment statement (https://lsa2017.as.uky.edu/sites/default/files/lsa-stmt-sexual-harassment.pdf).

For this course, study groups (especially for homework) are encouraged, but you are NOT allowed to write up your homework together or share files in the writing process. If you work with a study group, you should make a note on the top of your homework stating who you studied with.

Conduct in class:

Please turn cell phones to silent when in class. Please do not use electronic devices in class other than for class-related purposes or emergencies. Everyone is required to treat others in class with respect. Disruptive behavior is prohibited.