

General information

This course is a hands-on introduction to fundamentals of quantitative/statistical methodology in linguistics. It is based on the second edition (2013) of my textbook *Statistics for Linguistics with R: a practical introduction*. We begin by looking at a few basic notions such as variables and hypotheses. We then discuss the logic of quantitative studies using the null-hypothesis falsification approach and familiarized ourselves with how to set up factorial experiments and how data from experiments and corpora should be set up for subsequent statistical evaluation. Then, we are concerned with a variety of descriptive graphs and statistics for frequency data, averages, dispersions, and correlations. The largest part is concerned with a variety of statistical tests: distribution fitting tests, tests for independence, and tests for differences for frequencies, means, dispersions, and correlations. We end with a small primer for the kind of multifactorial methods that are the subject of Ling 202. We use the open source software tool R.

Course requirements and grading

- i. regular attendance in class;
- ii. preparation for, and active participation in, class. That is, I expect you to
 - do the readings/assignments so that you can discuss them and/or ask about things you have not understood;
 - work on and submit three small assignments (by email, see below);
- iii. one office hour visit in the quarter during which we talk about how statistical methods may be applied to something you are working on or something you find interesting;
- iv. a course-final take-home assignment, which covers the contents of the complete class and for which you can use all materials, notes etc.

All four assignments are due as **R reports**, i.e. as self-contained HTML files generated with **RStudio** and must have the following name (structure): `<201_lastname_assignment0#.html>`; as in `<201_smith_assignment02.html>` assignments that do not conform to these requirements will be considered as not submitted! The final grade will depend on your number of points. You can get 100 points by

- i. submitting all assignments in good quality and in a timely fashion (each small assignment is worth max. 20 points; the final take-home assignment is worth max. 30 points). Each assignment can be submitted early once to get feedback before the final submission; this, too, would be an **R report** called `<201_lastname_assignment0#-draft.html>`.
- ii. coming to the office appointment with questions regarding course contents or how course contents could maybe be applied in your own research (worth 10 points).

Excellent participation and/or homework assignments can result in up to 10 bonus points. Note that the fact that you will collaborate in class does *not* mean that the assignments are supposed to be solved collaboratively – I expect those to be the results of your individual work!

Contact

Office hours: Wed 14:30-15:30 in my office and upon appointment
Web: <http://tinyurl.com/stgries>
Email: stgries@linguistics.ucsb.edu

Course plan

- (1) 10/04: Fundamentals of statistical methods**
Read as follow-up: <201_01_intro-stat.pdf> and SFLWR 1.1-1.3
Read for next time: SFLWR 2.1-2.5, <<http://github.com/rstudio/cheatsheets/raw/master/base-r.pdf>> (bring to classes from now on)
- (2) 10/11: R: functions, arguments, data structures**
Read as follow-up: SFLWR 1.4-1.5, SFLWR 2.1-2.5
Oblig. assignment (graded): <201_02_assignment1.r>; deadline: 24 Oct at 09:00 PST
Read for next time: SFLWR 3.1.1-3.1.3
- (3) 10/18: descriptive stats: frequency distributions, means, dispersions**
Read for next time: SFLWR 3.1.4-3.2.3
- (4) 10/25: descriptive stats: standardization, confidence intervals, bivariate statistics**
Oblig. assignment (graded): <201_04_assignment2.r>; deadline: 01 Nov at 09:00 PST
Read for next time: SFLWR 4.1
- (5) 11/01: distributions (goodness of fit and independence/difference)**
Read for next time: SFLWR 4.2-4.3.1
- (6) 11/08: dispersions and means (goodness of fit)**
Read for next time: SFLWR 4.3.2-4.4
- (7) 11/15: means (independence/difference) and correlation/regression**
Oblig. assignment (graded): <201_07_assignment3.r>; deadline: 22 Nov at 09:00 PST
- (8) 11/22: your own little functions: tailoring things to your needs**
Read for next time: SFLWR 1-4
- (9) 11/29: no class, plz do your own exercise/practice session, I'll provide an answerkey**
Read for next time: SFLWR 1-4
Work on UCB admissions data set and bring code with you to class
- (10) 12/06: a brief exploration of multifactorial statistics plus review / evaluation**
Oblig. assignment (graded): <201_11_final.r>; deadline: 15 Dec at 20:00 PST

Preparation: you should make sure you have the following software installed (in this order):

- R (<<https://cran.r-project.org/>>);
- RStudio (<<https://www.rstudio.com/products/rstudio/download/#download>>).

Then start R (ideally with administrator access) and run the following line:

```
install.packages(c("amap", "car", "combinat", "effects", "formatR",  
                  "ggplot2", "gvlma", "party", "qcc", "rms", "rpart", "vcd"),  
                dependencies=TRUE)
```