

General information

This course is a hands-on introduction to fundamentals of quantitative/statistical methodology in linguistics. It is based on the third edition (2021) of my textbook *Statistics for Linguistics with R: a practical introduction* (also used for Ling 202 and 204). 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 and 204. We use the open source software tool R.

Course requirements and grading

- i. three small take-home assignments;
- ii. a course-final take-home assignment.

Attendance is not required and will not be monitored. All three assignments are due as R reports, i.e. as self-contained HTML files generated with RStudio (from .R or .Rmd files) 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 submitting all assignments in good quality and in a timely fashion (each small assignment is worth max. 30 points; the final take-home assignment is worth max. 40 points). Each assignment can be submitted early once to get feedback before the final submission; this, too, would be an R report called `<201_smith_assignment02-draft.html>`.

Contact

Office hours: Zoom, upon appointment
Web: <https://www.stgries.info>
Email: stgries@linguistics.ucsb.edu

Course plan

- (1) 01/09: Fundamentals of statistical methods**
Read as follow-up: <201_01_intro-stat.pdf> and *SFLWR*³ 1.1-1.5
Read for next time: *SFLWR*³ 2.1-2.5
- (2) 01/16: no class (MLK)**
Read for next time: *SFLWR*³ 2.1-2.5
- (3) 01/23: R: functions, arguments, data structures**
Read for next time: *SFLWR*³ 3.1, 3.5.3
- (4) 01/30: descriptive stats: frequency distributions, means, dispersions**
Oblig. assignment (graded): fundamentals/descriptives; deadline: 05 Feb at 15:00 PST
Read for next time: *SFLWR*³ 3.2-3.4
- (5) 02/06: descriptive stats: standardization, confidence intervals, bivariate statistics**
Read for next time: *SFLWR*³ 4.1
- (6) 02/13: distributions (goodness of fit and independence/difference)**
Read for next time: *SFLWR*³ 4.2-4.3.1
- (7) 02/20: no class (President's Day)**
Oblig. assignment (graded): statistical tests 1; deadline: 26 Feb at 15:00 PST
- (8) 02/27: dispersions and means (goodness of fit)**
Read for next time: *SFLWR*³ 4.3.2-4.4
- (9) 03/06: means (independence/difference) and correlation/regression**
Read for next time: *SFLWR*³ 1-4
Work on UCB admissions data set and bring code with you to class
- (10) 03/13: a brief exploration of multifactorial statistics plus review / evaluation**
Oblig. assignment (graded): statistical tests 2; deadline: 24 Mar at 15:00 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>>).