

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](#). 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 105. We use the open source [programming language R](#) and, as an IDE, [RStudio](#).

Course requirements and grading

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

Attendance is not required and will not be monitored. All four assignments must be sent to Michael Fiddler (see below) as R reports (i.e. as self-contained HTML files generated with RStudio) and must have the following file name structure: `<104_lastname_assignment0#.html>` (as in `<104_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. 20 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 `<104_smith_assignment02-draft.html>`.

Contact (STG)	Office hours: Zoom, upon appointment Web: http://www.stgries.info Email: stgries@linguistics.ucsb.edu
Contact (MF)	Office hours: Monday, 10:00-12:00, South Hall 4431 Email: mfiddler@ucsb.edu

Course plan

(1) 01/04: Fundamentals of statistical methods

Read as follow-up: <104_01_intro-stat.pdf> and SFLWR³ 1.1-1.5
Read for next time: SFLWR³ 2.1-2.5

(2) 01/11: R: functions, arguments, data structures

Read as follow-up: SFLWR³ 2
Oblig. assignment (graded): <104_04_assignment1.r>; deadline: 18 Jan at 09:30 PST
Read for next time: SFLWR³ 3.1, 3.4

(3) 01/18: univariate stats and visualization

Read for next time: SFLWR³ 3.2-3.3; 3.5.3

(4) 01/25: bivariate stats and correlation

Oblig. assignment (graded): <104_04_assignment2.r>; deadline: 01 Feb at 09:30 PST
Read for next time: SFLWR³ 4.1

(5) 02/01: distributions & frequencies (goodness of fit and independence/difference)

Read for next time: SFLWR³ 4.2-4.3.1

(6) 02/08: dispersions and means (goodness of fit)

Read for next time: SFLWR³ 4.3.2-4.4

(7) 02/15: means (independence/difference) and correlation/regression

Oblig. assignment (graded): <104_07_assignment3.r>; deadline: 22 Feb at 09:30 PST
Read for next time: SFLWR³ 3.5.1-3.5.3

(8) 02/22: plotting, functions, & recommendations for code

Read as follow-up: SFLWR³ 3.5.4

(9) 03/01: exercise/practice session

Read for next time: SFLWR³ 1-4
Work on UCB admissions data set and bring code with you to class

(10) 03/08: a brief exploration of multifactorial statistics

Oblig. assignment (graded): <104_11_final.r>; deadline: 17 Mar at 09:30 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>>).