General information

This course is a selective introduction to predictive modeling applications in linguistics. We start with a one-session intro of predictive modeling with an emphasis on regression modeling, which will survey model formulation, model selection, multifactoriality, and validation. Then, we work our way through a variety of regression modeling applications: linear regression, binary logistic regression, multinomial, and ordinal regression models. Then, one session will be concerned with model diagnostics and, perhaps, model validation. Finally, there are two sessions on tree-based approaches: classification and regression trees as well as random forests. Like its prerequisite course Ling 104, this course is based on the third edition of my textbook Statistics for linguistics with R: a practical introduction (2021) and uses the open source programming language R and, as an IDE, RStudio.

Course requirements and grading

i. a mid-term-like take-home assignment;
ii. a course-final take-home assignment.

Attendance is not required and will not be monitored. Both 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: <105_lastname_assignment0#.html> (as in <105_**smith**_assignment0**2**.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 both assignments in good quality and in a timely fashion; each small assignment is worth max. 50 points. Each assignment can be submitted early once to get feedback before the final submission; this, too, would be an R report called <105_**smith**_assignment0**1**-draft.html>.

Contact (STG)	Office hours:	Zoom, upon appointment
	Web:	<https://www.stgries.info>
	Email:	<stgries@linguistics.ucsb.edu>
Contact (MF)	Office hours:	Fri, 10:00-12:00, South Hall 4431A or via Zoom
	Email:	<mfiddler@ucsb.edu>

Course plan

(1) 03/30: Overview of modeling fundamentals
Read as follow-up: go over the slides and the HTML report
Read for next time: SFLWR[3] 5.2.1-5.2.3 (without 5.2.3.3)

(2) 04/06: Linear modeling 1
Read for next time: SFLWR[3] 5.1, 5.2.4-5.2.7

(3) 04/13: Linear modeling 2
Read for next time: SFLWR[3] 5.3.1-5.3.3

(4) 04/20: Binary logistic regression modeling 1
Read for next time: SFLWR[3] 5.3.4-5.3.5

(5) 04/27: Binary logistic regression modeling 2
Read for next time: SFLWR[3] 5.4.1

(6) 05/04: Multinomial regression
Read for next time: SFLWR[3] 4.3.2-4.4

(7) 05/11: Ordinal regression
Read for next time: SFLWR[3] 5.6-5.7

(8) 05/18: Model assumption & diagnostics
Read for next time: SFLWR[3] 7.1

(9) 05/25: Trees
Read for next time: SFLWR[3] 7.2

(10) 06/01: Random forests

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>).