General information

This course is a more advanced course on statistical modeling with an emphasis on various kinds of regression modeling; it presupposes a good understanding of the second edition (2013) of my textbook *Statistics for Linguistics with R: a practical introduction*. We begin with a first recap of linear and generalized linear regression modeling. We then discuss the use of contrasts and general linear hypothesis tests for linear and generalized linear regression models, followed by some ideas on how to explore curvature in data (regressions with breakpoints, polynomial regressions, and generalized additive models). This is followed by a larger chunk on linear and generalized linear mixed-effects (or multilevel) modeling, where we reanalyze published data and discuss numerical and visual exploration of regression results. The last two sessions are then devoted to influential data points and validation approaches and classification and regression trees / random forests. Obviously, 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 and/or assignments so that you can discuss them and/or ask about things you have not understood;
iii. my current idea is to have you do a comprehensive (!) analysis of a data set (ideally, one of yours), using as much of what you have learned in this class.

The assignment is due as an executable R script called `<204_assignment_lastname.r>`; together in one email with the data file that you analyzed as a tab-separated `.csv` file called `<204_assignment_lastname.csv>`; assignments that do not conform to this will be considered as not submitted! The final grade will depend on your number of points. You can get 100 points by

i. active participation in class (20 points);
ii. submitting the assignment in good quality and in a timely fashion (90 points); the assignment can be submitted early once to get feedback before the final submission as an executable R script called `<204_assignment_lastname_draft.r>`.

Contact

Office hours: Wed 14:30-15:30 in SH 3506 and upon appointment
Web: <http://tinyurl.com/stgries>
Email: <stgries@linguistics.ucsb.edu>
Course plan

(1) 04/02: Recap: (fixed-effects linear) and binary logistic regression models
    Read as follow-up: SFLWR 5.1-5.4
    Read for next time: Field, Miles, & Field (2012: Sections 10.4.1-10.4.4, 10.6.7, 12.5, 14.6); Crawley (2013:426-436)

(2) 04/09: extending/tweaking regression models 1: contrasts, GLHTs
    Read as follow-up: Gelman & Hill (2008: Ch. 3)
    Read for next time: Crawley (2013:209-211, 262-264, 466-471)

(3) 04/16: extending/tweaking regression models 2: breakpoints, polynomials, GAMs
    Read for next time: Gries (2015), Zuur et al. (2009: Ch. 5)

(4) 04/23: Mixed-effects modeling, part 1
    Read for next time: Gries (2013: Section 5.5), Pine et al. (2013), Gries & Adelman (2014)

(5) 04/30: Mixed-effects modeling, part 2
    Read for next time: Gries & Adelman (2014)

(6) 05/07: Mixed-effects modeling, part 3
    Read for next time: Zuur, Ieno, & Elphick (2010)

(7) 05/14: no class (teaching workshop in Michigan)
    Read for next time: Wilkinson & the Task Force on Statistical Inference (1999)

(8) 05/21: Influential data points and validation
    Read as follow-up: PennState Stat 501: lesson 11 on outliers and influential points (<https://onlinecourses.science.psu.edu/stat501/node/336>),

(9) 05/28: no class (Memorial Day)
    Read for next time: Crawley (2013:768-784)

(10) 06/04: Classification trees and random forests
References / Bibliography

Books, articles, websites

Statistics for linguists (with R)

General statistics and/or general R and/or applications
Cohen, Jacob. 1994. The earth is round ($p < 0.05$). American Psychologist 49.12:997-1003.
Loftus, Geoffrey R. 1996. Psychology will be a much better science when we change the way we analyze data. Current Directions in Psychological Science 5.6:161-71.

© Stefan Th. Gries
(Regression/linear) modeling