

# Statistical Modeling for Linguists

Dr Bodo Winter

June 1-5, 2020. UiT

The language sciences are undergoing a quantitative revolution. There is ever more data, an ever-growing toolkit of statistical methods, and a dedicated push towards incorporating more empirical research into linguistic theorizing. This workshop provides an in-depth introduction into what is arguably the most useful statistical method for linguists, the linear model and its various extensions, including generalized linear models (Poisson and logistic regression) and linear mixed effects models. The linear model framework is an alternative to standard ‘cookbook’ significance testing approaches, and it encourages users to express their research questions precisely in the form of an equation. The workshop will focus on issues that researchers often struggle with, such as the interpretation of models with interactions, the correct interpretation of model coefficients of logistic regression models, how to make choices about random effects structures, convergence issues, and so on. The workshop will also teach a lot of practical R skills, including an introduction to the tidyverse. No mathematical background is required. Also, no statistical background is required: while we will deal with fairly advanced approaches towards the end of the workshop, the idea of this workshop is that we start with the foundations but move at a fast pace, so that everything is presented in an integrated fashion, and so that learners from different levels of statistical expertise stand to benefit from this. Prior to the workshop, some exercises and preparation materials will be distributed to participants to save ourselves time with the basics of R during the workshop.

The course will consist of 6 lectures over the course of three days, one in the morning and one in the afternoon. The lectures will be a combination of lectures and detailed discussion of data sets. Lunch will be provided on all three days. There will be one course dinner for all the participants at the

end of the three days.

**Day 1:** 2 x 2 hours

introduction to data processing with the tidyverse and reproducible research work flows, introduction to linear models

**Day 2:** 2 x 2 hours

linear models with categorical predictors, interactions, generalized linear models: logistic regression

**Day 3:** 2 x 2 hours

generalized linear models: Poisson regression, (generalized) linear mixed effects models (and if there's extra time can also do intro to Bayesian modelling)

To get credit for the course (5 ECTS)

- The student must do the preparation and readings assigned in advance.
- Attend all lectures
- Do an analysis of a dataset of their choice with at least one linear model, at least one figure. In addition it should include (i) a Markdown and knitted Markdown file, and (ii) it should be uploaded on GitHub or OSF. This should be handed in in the form of a paper, after the end of the course.