Goal of the course

The goal of this course is to familiarize students with a range of statistical techniques that are available for the analysis of one response variable (e.g., reaction time, or pupil dilation, pitch, accuracy) that is to be modeled as a function of one or more predictors. These predictors can be factors (e.g., native versus non-native speakers), numerical covariates (e.g., frequency of occurrence), or combinations of factors and covariates. Modeling techniques will be introduced conceptually, and emphasis will be on worked examples of their application. Hands on sessions provide training in applying statistical models to real data sets from linguistics and psychology. For this course to be profitable to them, participants should be familiar with basic concepts from statistics (random variables, quantiles, mean, variance, normal distribution, t-distribution, t-test, hypothesis testing, confidence intervals). As the course will make use of the R statistical programming environment, participants should bring their laptops with R installed, and should know how to install packages and how to load data into R.

The course has four blocks.

1. **The (generalized) linear mixed model.** This model is widely used for data with response variables collected from multiple subjects and multiple items. It allows the analyst to take into account how uncertainty about model estimates varies with subjects and items. The vexed question of how complex a model should be to be minimally adequate will be discussed in detail.

2. **The generalized additive model: basic concepts.** The generalized additive model (GAM) relaxes the assumption that the functional relation between the response and one or more predictors is linear. It is ideal for modeling wiggly curves and wiggly (hyper)surfaces. Model criticism and tools for dealing with model residuals that are not identically and independently distributed will be introduced.

3. **Extending the generalized additive model.** Factors may interact with numeric predictors, resulting in wiggly curves and wiggly surfaces that may have a different shape depending on the levels of that factor. Two ways of assessing such interactions are discussed, including the modeling of a difference curve or a difference surface. Furthermore, examples are provided of how ordinal data can be modeled with GAMs.

4. **Quantile regression and survival analysis.** The generalized linear model and the generalized additive model predict the expected value (the mean) of the response. However, it is often of interest to know whether the effect of predictors is different depending on which quantile
(other than the median) is modeled. Quantile regression with GAMs on the one hand, and
dynamic survival analysis on the other, provide very different and complementary approaches
for coming to grips with the full distribution of the response.

Selected references (including references to studies applying regression methods):

**linear regression**: Baayen (2008), Harrell (2015), Anscombe (1973), Friedman and Wall (2005),
Wurm and Fisicaro (2014), Baayen (2013), Baayen (2010); **generalized linear model**: Donnelly
**linear mixed model**, Pinheiro and Bates (2000), Baayen et al. (2008), Galecki and Burzykowski
(2013), Janda et al. (2010), Barr et al. (2013), González et al. (2014), Bates et al. (2015), Lele et al.
(2012), Matuschek et al. (2017), Johnson (2009); **random forests**: Breiman et al. (1984), Breiman
(2001), Strobl et al. (2009), Tagliamonte and Baayen (2012); **generalized additive model**: Wood
(2006), Baayen et al. (2017), Baayen et al. (2016), Wieling et al. (2011), Nixon et al. (2016), Hendrix
et al. (2016), Koesling et al. (2012), Wieling et al. (2016), Wieling et al. (2014), Tremblay and Newman
(2014), Tomaschek et al. (2017); **ordinal regression**: Baayen and Divjak (2017), Kapatsinski
et al. (2017); **quantile regression**: Koenker (2005), Fasiolo et al. (2017), Baayen (2017) (chapter
7); **survival analysis**: Scheike and Martinussen (2007), Scheike and Zhang (2011), Schmidtke
et al. (2017), Baayen (2017) (chapter 8).

References


version 0.1.3.


157.


- completely revised second edition. Manuscript in preparation for Cambridge University Press,
Cambridge, U.K.


Makarova, A., Dickey, S. M., and Divjak, D. S., editors, *Thoughts on Language: Studies in Cog-
nitive Linguistics in Honor of Laura A. Janda*, page in press. Slavica Publishers, Bloomington,
IN.


