

# Regression modeling and multimodel inference.

Andreas Baumann  
Department of English and American Studies, University of Vienna  
andreas.baumann@univie.ac.at

Many subfields of linguistic science rely on quantitative data collection and analysis. Thus, for testing hypotheses based on collected data, statistical techniques have become increasingly relevant, regression modeling belonging to the most popular statistical tools. The benefits of regression modeling are evident as the technique allows to simultaneously analyze the respective effects of multiple factors on an outcome variable one is interested in.

Sometimes, however, the choice of independent factors entering a regression model is relatively arbitrary, e.g. if factors are included by the researcher just to be on the safe side and to prevent any criticisms of having neglected this or that potential covariate. This can have severe consequences for the quality of the model, since every additional factor consumes data. In some cases, dropping a certain factor increases the model's predictive power. The question now is: based on what criterion should a factor be included in a model?

In the first session of this workshop, we quickly review linear regression modeling and afterwards discuss a number of model-selection criteria (R-squared; AIC; AICc; BIC) which help to tackle the latter question. Most often, these criteria are used to identify a single optimal model among a set of possibly model candidates. However, this is not their only purpose. In the second session, we discuss information-theoretic techniques which allow to combine and draw inferences from multiple models at the same time (model averaging; multimodel inference; relative variable importance; cf. Burnham & Anderson 2002; Burnham et al. 2011). This approach is powerful, as it helps to minimize loss of information contained in the data. Computations will be done in R with a particular focus on the MuMIn package (Barton 2016).

## References

- Barton, K., 2016. MuMIn: Multi-Model Inference. R package version 1.15.6. <https://CRAN.R-project.org/package=MuMIn> [1]
- Burnham, K.P., Anderson, D.R., 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York.
- Burnham, K.P., Anderson, D.R., Huyvaert, K., 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behav Ecol Sociobiol* 65, 23-35.