The influence of model selection on selected measures of fit

Märt Möls

University of Tartu, Estonia, martm@ut.ee

Keywords: model selection, mixed models

Selecting models based on AIC, covariate statistical significance or by some other model selection criterion can introduce bias into statistics usually used to describe the model fit (and can also lead to biased estimates of effect sizes). A typical remedy of the problem is to use a separate validation dataset to get unbiased estimates of model fit characteristics. But sometimes it is not practical to divide the original dataset into model selection and validation datasets. Therefore some theoretical alternatives to decrease the bias introduced by the model selection process may still be of interest.

How to use available but biased (due to model selection) statistics to derive still meaningful (bias-reduced) estimates of the quantities of interest? A mixed model for truncated response variable is introduced and used to estimate some quantities of interest to reduce the model selection bias. The proposed mixed model behaviour in some practical applications will be evaluated.