

From the Editor

Reporting modern statistical analyses: reproducible and transparent

Mehtätalo L. (2019). Reporting modern statistical analyses: reproducible and transparent. *Silva Fennica* vol. 53 no. 3 article id 10257. 2 p. <https://doi.org/10.14214/sf.10257>

Large majority of the articles published in *Silva Fennica* include statistical analyses of empirical quantitative data. In reporting of the materials and methods in such an article, an important requirement is reproducibility: the reader should be able to repeat the data collection and analyses based on the description in the article. A closely related requirement is transparency: the authors should give the necessary information for readers to evaluate whether the method is justified and has been implemented correctly. In *Silva Fennica*, we want to promote reproducibility and transparency of scientific publishing also in the future. To illustrate what it means in practice, I will discuss analyses based on linear models in more detail. In this context, analysis of variance and (t-) tests of sample means are seen as special cases of the classical linear model.

The use of classical linear model has a long history, dating back to the works of Karl Pearson, R. A. Fisher and G. U. Yule in the early 20th century. For example, Fisher's classical textbook "Statistical methods for research workers" already presented the t-tests, analysis of variance, and linear regression as basic tools to analyze empirical data. These basic methods are fully optimal whenever the implicit assumptions about the independence and constant variance of the data are met. The use of those rather simple and widely known methods is straightforward and their use in a research report can be described, by writing e.g., "the logarithmic data were analyzed by one-way ANOVA and Tukey's post-hoc tests; the applied transformation showed constant error variance according to standard diagnostic plots". There is usually no need for a formal presentation of the implicitly assumed classical linear model. Of course, the empirical data and data collection procedures should be described transparently so that the reader can critically evaluate whether the selected method is justified.

Nowadays methods are widely available to take into account such properties of the data that could not be taken into account in the classical linear model. For example, there are good methods to analyze dependent data with heteroscedastic errors. However, a data set can be independent and homoscedastic in only one way, but dependent and heteroscedastic in infinitely many ways. Therefore, secondary sub-models are needed for dependence and heteroscedasticity. Furthermore, there may be several alternative methods for parameter estimation and inference. For example, in grouped data sets generalized linear mixed-effects models can be used to model non-normal grouped

data. Formulating a linear mixed-effects model involves choices about the levels of grouping and random-effects structure for each level of grouping, in addition to the variance-covariance structure of the residual errors. For non-normal data, additional choices are needed about the link function, parameter estimation methods, applied methods for inference, and the models about zero-inflation and overdispersion. All these choices can have a major effect on the results about the factors of main interest and should therefore be reported.

Transparent publication of today's statistical analyses requires reporting and justification of all non-trivial choices made in the model selection. Reproducible and transparent reporting of such an analysis is seldom possible without formal presentation of the model. Also tables showing the estimates of all model parameters are often useful, as well as carefully selected diagnostic graphs about model fit. The space limitations of the papers are no more a problem for a sufficiently detailed reporting of the methods and models, because they can be included as an electronic supplementary file.

Lauri Mehtätalo

Associate Editor for Biometry and Methods