Enhancing the documentation
by leaving useful traces

Kimmo Vehkalahti

University of Helsinki, Finland

Abstract

An appropriate documentation of the scientific work process 1) advocates the reproducibility of the research, 2) supports backtracking and following side tracks, 3) improves the data quality, and 4) helps to avoid reinventing the wheel.

Instead of formal documentation, we focus on *traces* that are left behind when working. By traces we refer to free-form notes and comments written down explicitly, but also to expressions, commands and work schemes created primarily to utilize various operations of the software. These pieces together reflect the *thinking process*, which needs to be retraced, say, when revising a paper after a possibly long review process. Leaving useful traces enhances the documentation and helps to get back on the track.

We demonstrate these ideas especially with the matrix operations of SURVO MM software. Its unique *editorial interface* promotes building and maintaining the work process so that each step is appropriately documented.

Our examples are taken from certain new developments in the area of multivariate statistical modeling with measurement errors.

Keywords Work process, Documentation, Reproducibility, Data quality, Matrix computations, Multivariate methods, Measurement error.

References:

Dasu, T. and T. Johnson (2003). *Exploratory Data Mining and Data Cleaning.* Hoboken, New Jersey: John Wiley & Sons.

Gentleman, R. and D. Temple Lang (2004). Statistical analyses and reproducible research. Bioconductor Project Working Papers, Working Paper 2 http://www.bepress.com/bioconductor/paper2.

Mustonen, S. (2001). *SURVO MM: Computing Environment for Creative Processing of Text and Numerical Data* http://www.survo.fi/mm/english.html. Freeware version *Survo Editor*: http://www.survo.fi/english/download.

Tarkkonen, L. and K. Vehkalahti (2005). Measurement errors in multivariate measurement scales. *J. Multivariate Anal.*, *96*, 172–189.

Vehkalahti, K. (2005). Leaving useful traces when working with matrices. *Research Letters in the Information and Mathematical Sciences*, *8*, 143–154. Proceedings of the 14th International Workshop on Matrices and Statistics. (Paul S.P. Cowpertwait, ed.) Massey University, Auckland, New Zealand, March 29 – April 1, 2005
http://iims.massey.ac.nz/research/letters/volume8/.

Vehkalahti, K., S. Puntanen and L. Tarkkonen (2006). Effects of measurement errors in predictor selection of linear regression model. *Reports on Mathematics, Preprint 439, Department of Mathematics and Statistics, University of Helsinki, Finland*
http://mathstat.helsinki.fi/reports/Preprint439.pdf.

Vehkalahti, K., S. Puntanen and L. Tarkkonen (2006). Estimation of reliability: a better alternative for Cronbach's alpha. *Reports on Mathematics, Preprint 430, Department of Mathematics and Statistics, University of Helsinki, Finland*
http://mathstat.helsinki.fi/reports/Preprint430.pdf.