Commercial meets Open Source - Tuning STATISTICA with R

Christian H. Weiß*

March 10, 2008

Abstract

R is an extremely powerful environment for statistical computing: It provides packages designed for different areas such as data mining, econometrics, epidemiology, biostatistics, it offers methods from different statistical disciplines like time series analysis, statistical process control, bootstrapping, cluster analysis, and others. Besides its mere extent, R differs from competing statistics environments also in the fact that it reflects the state-of-the-art in statistical sciences. And not to forget: R is freely available.

On the other hand, R is not particularly user-friendly: It does not offer a graphical user-interface, where the repertoire of methods is fully integrated and available also for users, who have not learnt the R language. It does not offer a powerful spreadsheet environment, which enables an intuitive way of data manipulation. Therefore, (potential) users from applied sciences and industry often do not have the heart to work with R.

In this talk, I propose to combine the power of R with the comfort of a commercial package like STATISTICA. STATISTICA can be used as an easily operated interface with a respectable basic equipment of statistical procedures, see Weiß (2006). But if required, one can easily integrate specialised statistical procedures and sophisticated techniques offered by R into the user interface of STATISTICA. Besides the base version of STATISTICA with its Visual Basic development environment, and besides R together with the required packages, the user only needs to install the R DCOM Server of Baier & Neuwirth (2007).

The necessary procedure and essential commands to access R from STATISTICA are explained, also refer to StatSoft (2003). A number of examples highlight situations, where R can be used to extend the functionality of STATISTICA. Among others, we explain how an ARL calculator for computing average run lengths of EWMA and CUSUM control charts can be programmed, using the spc package of Knoth (2007). The ARL calculator supports the design of these control charts, which are themselves available through STATISTICA.

---

*Institute of Mathematics, Department of Statistics, University of Würzburg, Germany.
Email: `christian.weiss@mathematik.uni-wuerzburg.de`

References

BAIER, T., NEUWIRTH, E.: *R/Scilab (D)COM Server V 2.50*. March, 2007.
http://cran.r-project.org/contrib/extra/dcom/

KNOTH, S.: *The spc Package (Statistical Process Control), Version 0.21*. October, 2007.
http://cran.r-project.org/src/contrib/Descriptions/spc.html

STATSOFT: *STATISTICA Data Miner: Integrating R Programs into the Data Miner Environment*.
StatSoft Business White Paper, June, 2003.
http://www.statsoft.com/support/whitepapers/pdf/STATISTICA_Integrating_R.pdf

WEISS, C.H.: *Datenanalyse und Modellierung mit STATISTICA*. Oldenbourg Wissenschaftsverlag, München, 2006.