

Statistical Computing in R: Strategies for Turning Ideas into Software

Achim Zeileis

WU Wirtschaftsuniversität Wien

Abstract

Computing and computational methods play an important role in quantitative (social science) research. Therefore, both researchers and practitioners typically also become authors of code/software (for their own applications or simulations), but more often than not this software is not made available (or not in such a way that makes it easily reusable by others). Hence, we discuss two questions: *Why* should researchers write and publish software? *How* should such software be written in order to be efficient and still easy to use? The answer to the first question is that publication of software is in the interest of both, the authors (to promote their methods/results) and the readers/users (to assure reproducibility of research). The *how* is less straightforward because its details depend on the particular problem at hand. However, it is possible to discuss general strategies for turning ideas into software. The R system for statistical computing assists these strategies by providing various language features that facilitate the development of effective software, e.g., object orientation, functions as first-class objects, and lexical scope among others. We illustrate how these features have been utilized in the design of the package *sandwich* for HC and HAC covariance matrix estimation. Its application in practice is exemplified by regression modeling of count data.

Keywords: R, statistical computing, open-source software, reproducible research, object orientation, lexical scope, covariance matrix estimation, count data regression.

Affiliation:

Achim Zeileis

Department of Statistics and Mathematics

WU Wirtschaftsuniversität Wien

Augasse 2–6

1090 Wien, Austria

E-mail: Achim.Zeileis@R-project.org

URL: <http://statmath.wu.ac.at/~zeileis/>