



Journal of Statistical Software

January 2011, Volume 37, Book Review 3.

<http://www.jstatsoft.org/>

Reviewer: Jeroen Ooms
University of California at Los Angeles

SAS and R

Ken Kleinman and Nicholas J. Horton
Chapman & Hall/CRC, Boca Raton, FL, 2010.
ISBN 978-1-4200-7057-6. 323 pp. USD 71.95.
<http://SAS-and-R.blogspot.com/>

The book SAS and R arose from the popular blog with the same name and is the first in a series of currently three books by Kleinman and Horton about statistical computing in SAS and R. This book features an extensive list of techniques and worked examples in data management, statistical analysis, and graphics, illustrated in both R and SAS. In addition it has two appendices with brief introductions to both systems. The book has not been written to be read cover to cover; it rather is a convenient reference text to quickly learn by example how to perform common tasks in both software packages. To navigate through the examples, the book has a comprehensive table of contents and three indices: a detailed subject index in English, a SAS index organized by SAS syntax, and an R index organized by R syntax.

The authors affirm that *“the book functions in the same way that an English-French dictionary informs users of both the equivalent nouns and verbs in the two languages as well as differences and grammar.”* Therefore, it is probably not the best introductory text to statistical computing or either software package. A basic understanding of the general concepts in statistics and programming seems to be assumed and is required to understand the examples. However, for the reader that meets these requirements, the book provides a powerful starting point to a wide variety of statistical techniques available in SAS and R. The multiple indices effectively locate the appropriate sections, especially if one already has some experience with either SAS or R. Also it is clear and pleasant that the authors are extremely proficient and experienced with both languages. Although they *“do not claim to provide the most elegant solution”*, the quality of the code is actually one of the strong points of the book. The examples are clear and understandable and the code is efficient and readable.

The main chapters in the book cover respectively data management, statistical analysis, and graphics, however the scope of these chapters is somewhat selective and traditional with a preference for methods relevant to biostatistics applications. For example, the first chapter discusses the usual variable manipulation and reading/writing to several data formats e.g. CSV and XML, but it does not treat databases at all other than mentioning that both systems have SQL interfaces. The major part of the statistical analysis chapters is focused on regression related methods, e.g. ANOVA, GLM, GAM, time series, survival analysis and

mixed models. The book does not cover more exotic multivariate techniques, like PCA, SEM, or network analysis to name a few. The chapter on graphics is very detailed, but most of the examples are again the usual suspects: histograms, scatterplots, and smooth lines.

In conclusion, this book does exactly what it promises: it facilitates a translation between SAS and R, without getting overly detailed or technical. It is mainly useful as a starting point for those who already know either R or SAS, and want to learn the other language, without going over extensive manuals or introductory texts.

Reviewer:

Jeroen Ooms

University of California at Los Angeles

Department of Statistics

Los Angeles, CA 90024, United States of America

E-mail: jeroen.ooms@stat.ucla.edu

URL: <http://www.stat.ucla.edu/~jeroen/>