Journal of Statistical Software

November 2012, Volume 51, Book Review 1. http://www.jstatsoft.org/

Reviewer: Irina Kukuyeva
University of California, Los Angeles

R for Statistics

Pierre-André Cornillon, Arnaud Guyader, François Husson, Nicolas Jégou, Julie Josse, Maela Kloareg, Eric Matzner-Løber, Laurent Rouvière
Chapman & Hall/CRC, Boca Raton, FL, 2012.
ISBN 978-1-4398-8145-3. 306 pp. USD 59.95 (P).
http://www.crcpress.com/product/isbn/9781439881453

R for Statistics is a translation of *Statistiques avec R* that came out in 2008. The authors implicitly assume one introductory course in statistics and no prior knowledge of R. As a result, the book is accessible for statisticians of all levels and areas of expertise as well as for novice and advanced R users. It is broken up into two parts, introducing the reader to the field and the software.

Part I of the book focuses on providing an introduction to R for novices as well as a reference for advanced users. It covers topics such as software installation, object types, subsetting objects and output, importing and exporting data files, data processing (such as formatting variables and handling missing data), graphics (2D, 3D and interactive) and function creation (such as loops and vectorized functions in the `apply` family). Each chapter closes with a set of exercises and hints to review the material covered; the answers to the exercises can be found in the appendix.

Part II of the book is an introduction or review of statistical methods. It begins with a review of key concepts from Part I for those that want to skip ahead. The following six chapters discuss hypothesis testing, regression, analysis of variance, classification methods (via linear discriminant analysis, logistic regression and decision trees), dimension reduction techniques (principal component analysis, correspondence analysis and multiple correspondence analysis), and the most common clustering approaches (ascending hierarchical clustering and k-means). Each method is presented in its own section and includes an overview of the method, a worked out example with an interpretation of findings as well as a references for extensions of the methodology aimed at relaxing the data assumptions.

Based on the presentation and scope of the book, I recommend it for anyone who wants to learn about the why and how of the most commonly employed statistical methods and their extensions.

The only drawbacks to the books are as follows. There were minor formatting issues, such as too much or lack of whitespace. The last two chapters of the book (on dimension reduction

techniques and clustering) analyze the examples in the context of the **FactoMineR** package, created by some of the co-authors of the book, even though most of the analysis can just as easily be called from base R and with arguments in English (in contrast to `choix` and `habillage` from within the `plot` function of their package). While the authors stress that all of the examples are reproducible, as all of the datasets can be downloaded from http://www.agrocampus-ouest.fr/math/RforStat, unfortunately at the time of the writing of this review, the websites for datasets for both the American and French versions of the book were under construction (and not all of the datasets introduced in the book are available in the **FactoMineR** package).

Reviewer:

Irina Kukuyeva
Department of Statistics
University of California, Los Angeles
8125 Math Sciences Building, Box 951554
Los Angeles, CA 90095-1554, United States of America
E-mail: ikukuyeva@stat.ucla.edu
URL: https://sites.google.com/site/ikukuyeva/

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume 51, Book Review 1 *Published:* 2012-11-05
November 2012