



Journal of Statistical Software

April 2009, Volume 30, Book Review 6.

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

Reviewer: Anatol Sargin
University of Augsburg

Statistics and Data with R: An Applied Approach Through Examples

Yosef Cohen and Jeremiah Cohen

John Wiley & Sons, Chichester, United Kingdom, 2008.

ISBN 978-0-470-75805-2. 628 pp. EUR 60.00.

<http://turtle.gis.umn.edu/pmwiki/pmwiki.php?n=StatisticsandDatawithR.HomePage>

This book is an introductory course in statistics and R. It covers both basic and sophisticated statistical topics, and shows how to use these techniques in R. Due to its many real data examples this book can even be useful for experienced R users. Datasets and R code can be found on the book's homepage, giving users the possibility to run the code easily, and modify it if desired.

The difference between this and other R books is the focus on applications. It introduces the most important functions used to perform fundamental statistical analyses. This is done in depth at the expense of other R features. For example, programming in R is only briefly discussed.

The book consists of three major parts. The first part (97 pages) covers basic R concepts, different ways to handle data in R, and graphical methods. It is nice that **lattice** plots are mentioned here, though unfortunately they are not used much later. In the second part (144 pages) substantial concepts of probability theory and statistics are explained. This is not done in much detail, but this is not the purpose of the book. It is shown how these concepts can be implemented in R, and many examples are used to illustrate them. The last part (412 pages) discusses advanced statistics. However, only one chapter is on exploratory data analysis, while the others are on inference statistics. Here the book could have laid less stress on theoretical concepts (there are, for example, four chapters on hypothesis testing) and more on data analysis in R. The balance is somewhat redressed in the last chapter, which describes applications to two datasets. This interesting chapter could have been enhanced, too.

The datasets used in this book are interesting and new. For instance, in the last chapter data on the war in Iraq (from 2003-03-12 to 2007-10-10) and on the Second Intifada between Israel and militant Palestinian organizations (from 2000-09-27 to 2003-10-04) are analysed. It is nice to have new and recent datasets. Overall, the extensive use of empirical data is the major advantage of the book, complementing the statistical theory and its realization in R well.

The R code is mostly well-written, but at some points it is too complicated and unnecessarily verbose. Sometimes the authors are inconsistent in their advice, or do not stick to it themselves. But these problems are rare and do not have serious consequences. The only major drawback is the sloppy use of graphics, which is usually considered one of the most important features of R. For example, the authors always use the default point character, while a solid circle is usually much better. Also the use of lines instead of points would often improve the graphics.

To sum up, this book is a solid text for people getting started with R. The introduction is very detailed and many hints for working in R are provided. Readers should have some knowledge of statistics, because the theory is not covered in depth. People interested in the analysis of real datasets with R will also find plenty of useful material in this book.

Reviewer:

Anatol Sargin

Department of Computer Oriented Statistics and Data Analysis

University of Augsburg

D-86135 Augsburg, Germany

E-mail: anatol.sargin@math.uni-augsburg.de

URL: <http://stats.math.uni-augsburg.de/mitarbeiter/sargin/>