Reviewer: Jan de Leeuw
University of California at Los Angeles

An R and S-PLUS Companion to Multivariate Analysis

If book publication is a reliable indicator of the state of statistical software, then R and to a lesser degree S-PLUS have definitely arrived and are starting to ascend to a considerable degree of dominance in the statistics market. We are rapidly approaching the stage in which computationally oriented books in statistics will by default come with code in R, either in the text or on a website.

This development has an important corollary. There are quite a few books appearing now for which there would not be a market if they had not included code in R. This is because the books are otherwise unremarkable, their main redeeming feature is that they include the computer code in some form or another. The book by Everitt under review here is one in the growing series "Introduction to X with R". It is a more or less standard introductory treatment of multivariate analysis techniques. In fact, because of the extraordinary productivity of the author, we can almost say it is a more or less standard Everitt introduction to multivariate analysis. But the book also has a ten-page appendix teaching the basics of R and S-PLUS, and throughout the book there are snippets of code that actually perform the calculations and make the graphs.

One interpretation of the book, and of other simalar books, is that this is one of a new generation of statistics cookbooks. The typical reader now has R in her kitchen, and consequently there is no reason any more to talk about hand-held calculators or mainframes. I should emphasize that this, however, does not just produce old wine in a new bottle. It actually is a large step ahead compared with previous cookbooks, because the R skills that people pick up are transferable, and because interactive computing means being in much closer contact with both techniques and data.

In this particular book, if you strip out the R code and the examples, you probably have 50 pages of written introduction to the standard body of multivariate analysis techniques. This, in itself, does not take you very far beyond what is already included in Venables and Ripley (2002). Clearly, one cannot expect much depth. There is a useful list of references if the reader wants to know more. There are many examples, data are often included in the text,

and analysis are done using mostly the R base packages. I would not hesitate to recommend this book for an advanced undergraduate course in multivariate analysis, and even for a graduate course in the social or health sciences. The benefits that come with forcing students to install and use R and to actually do the computations substantial, although skeptics may argue that this are not very different from the old practice of forcing students to compute correlation coefficients by hand.

From the production and organization point of view, the book could be improved a great deal. Given the fact that the book has a website, there seems to be no need to actually print the data sets, which can be quite large. The practice of just inserting R output and input in a computer font directly in the text makes for a pretty ugly book – not unlike the obnoxious color panels in previous generations of cookbooks. Default R plots are not very nice either when copied into a book. In fact, one could argue that having the code and the data on a website makes it easy to produce the plots and print output on demand. And, of course, this ultimately points to the fact that books like these should be interactive, so that people can run their code, input and edit their own data, and produce their graphs by just clicking on the appropriate places in the text.

The code in the book studiously avoids the use of functions (let alone methods and objects). R is used as a calculator, almost exclusively from the command line. This may give people the wrong impression, because it very much underrepresents the power of the R system. Although using objects is taking things too far, I think presenting the code in fuctional form, even if many of the functions are trivial, would have been better.

All in all, I am bit ambiguous about this book. Its treatment of multivariate analysis is minimal, and neither exciting nor offending. It only uses a tiny percentage of the power of R. If your hobby is to own and peruse beautiful books, then there is no reason to add this one to your collection. But on the other hand it is, without a doubt, useful for some levels of routine teaching, including self-teaching. Consequently books written according to this new recipe have the same ulitarian value as the never-ending stream of introductory statistics textbooks. They have a somewhat mechanical after-taste, and ultimately they reinforce the view of statistics as a boring bunch of recipes, but unfortunately that is still what the market wants. At least, after using this book, the good students will have R on their machines.

References

Venables WN, Ripley BD (2002). *Modern Applied Statistics with S. Fourth Edition.* Springer. ISBN 0-387-95457-0, URL http://www.stats.ox.ac.uk/pub/MASS4/.

Reviewer:

Jan de Leeuw
University of California at Los Angeles
Department of Statistics
Los Angeles, CA 90095-1554
United States of America
E-mail: deleeuw@stat.ucla.edu
URL: http://gifi.stat.ucla.edu/