

## **BOOK REVIEWS**

*An R and S-PLUS Companion to Applied Regression.* By John Fox. Thousand Oaks, CA: Sage, 2002. 312 pp. \$44.95.

DOI: 10.1177/0049124105277200

John Fox's (1997) previous book, *Applied Regression Analysis, Linear Models, and Related Methods*, is an introduction to linear models that begins with least squares fits, covers a range of diagnostic methods, and ends with chapters on logistic regression and various extensions to the linear model, including generalized least squares, robust estimation, and nonparametric approaches. The book is thorough in its treatment of the underlying theory but also strongly oriented toward applied researchers, in that it emphasized good data visualization and diagnostic practices in model building. What it lacks, however, are examples developed in a computing environment that would allow students of the book to follow along for themselves. *An R and S-PLUS Companion to Applied Regression* fills this gap.

S-PLUS and R are, respectively, commercial and free implementations of the S statistical computing language. R is a very successful open-source software project that has shown rapid and sustained development over the past few years. For instructors, it has three main benefits. First, it is freely available for all the operating systems students are likely to use, including Windows, Macintosh, and Linux. (It can be downloaded at <http://cran.r-project.org>.) Second, it self-consciously implements a "best practices" approach to data analysis in a way that discourages users from thinking about statistics as a set of canned solutions to standard problems. And third, it has powerful graphics capabilities that allow for very useful visual representations of data and models. Given R's free availability, Fox treats it

as the default application in the book. Where S-PLUS diverges in its implementation (in the way contrasts are coded within factors, for example), text boxes are used to explain the differences.

The graphical user interface of S-PLUS is not discussed. The focus instead is on reproducible data analysis done interactively via the command line or from scripts.

The first two chapters of the book introduce R and S-PLUS and explain how to read and manipulate data within the applications. An extended illustration in Chapter 1 (using Duncan's occupational prestige data) allows the novice reader to get a feel for what using R is like and might be of use to a reader wanting to compare S to a more familiar application. S has a slightly steep learning curve (especially for users more familiar with point-and-click applications), but Fox's exposition is clear and well paced. Chapters 3 through 6 go through the practical work of exploring data, fitting models, and diagnosing problems. Ordinary least squares (OLS), analysis of variance (ANOVA), and general linear model (GLM) fits are explored. Throughout, Fox makes good use of S's graphics capabilities, some of which he supplies himself through his add-on package, CAR (Companion to Applied Regression). Chapter 7 deals with the construction of graphs from scratch, and Chapter 8 introduces the basics of programming in S (rather than simply writing scripts or functions). The book's Web site (<http://socserv.socsci.mcmaster.ca/jfox/Books/Companion/>) provides the code for all the examples used in the book. More impressively, a substantial amount of more advanced material is available in an online appendix. The additional topics introduced there include nonlinear and nonparametric regression, mixed-effects and structural equation models, bootstrap and robust regression techniques, and survival analysis. The online supplements are presented with the same high standard as the text, and together they amount to another half a book's worth of content.

The approach throughout is hands-on, with Fox taking the reader step-by-step through the analyses. The book uses freely available social science data sets throughout, and all the plots and results should be reproducible by the reader. At the same time, Fox provides enough discussion of the general literature to

make sure the reader is not encouraged to forget about the relevant theory.

When *An R and S-PLUS Companion* appeared in 2002, R had matured enough to be a stable platform for statistical computing. It has continued to evolve since then: the core application has improved in numerous ways, and many add-on packages have been contributed. Most of these changes will be invisible or not immediately relevant to users who learn R with this book. A significant addition to R's capabilities comes through the lattice graphics package. It offers a powerful set of methods for graphically presenting data and model estimates, particularly data with repeated observations stratified by some unit such as the experimental subject, school district, state, or country. These methods mostly postdate the book, and so at the time of writing, Fox was able to mention them only very briefly. (They are used more extensively in some of the online material, however.) I should emphasize that the newer packages do not make the techniques presented by Fox obsolete, by any means. It is just unfortunate that the lattice approach could not be treated in more depth.

The book should be of use to a variety of audiences. For instructors teaching a second- or third-semester graduate statistics course, it fits very well (as one would expect) with Fox's own textbook. It should also slot easily into any course of this kind that uses R or S-PLUS as a computing environment, regardless of the main textbook. Students or social scientists already familiar with other statistics packages will find the book makes it easy to learn R and provides additional context (both theoretical and applied) to the informative but often terse documentation supplied with the software. When it appeared, *An R and S-PLUS Companion* was one of very few introductions of its kind. The range of choices has expanded in the past two years, and although it is still a useful companion for self-starters, readers wanting to learn R for themselves might also want to look at Dalgaard (2002) (for an elementary introduction) or Maindonald and Braun (2003) (for more advanced topics). Nevertheless, for classroom use by social scientists at its intended level—especially in conjunction with Fox (1997)—it is still the best book available.

## REFERENCES

- Dalgaard, Peter. 2002. *Introductory Statistics with R*. New York: Springer.
- Fox, John. 1997. *Applied Regression Analysis, Linear Models, and Related Methods*. Thousand Oaks, CA: Sage.
- Maindonald, John and John Braun. 2003. *Data Analysis and Graphics Using R: An Example-Based Approach*. New York: Cambridge University Press.

—KIERAN HEALY  
University of Arizona

---

*Handbook of Data Analysis*. Edited by Melissa Hardy and Alan Bryman. Thousand Oaks, CA: Sage, 2004. xviii + 704 pp. \$130.

DOI: 10.1177/0049124105277555

Melissa Hardy and Alan Bryman have assembled a collection of 30 essays dealing with various aspects of quantitative and qualitative data analysis, written by some of the leading data analysts of our time. Following an introduction by the editors pointing out some common themes and issues that arise in various methods for analyzing data, a section on “foundations” contains chapters dealing with methods for constructing variables, summarizing distributions, drawing inferences, and analyzing incomplete data sets and historical data. A chapter on feminism reviews discussions and debates about the implications of feminist thought for research. The issues raised in this chapter go beyond data analysis and deal with issues that transcend gender.

A section on the general linear model and its extensions covers multiple regression, categorical variables, contingent effects, categorical outcomes, and log-linear analysis. The section on longitudinal models contains chapters on modeling change, fixed and random effects in panel data, trajectory models, event history analysis, and sequence analysis. Another section deals with such “new methods” as sample selection bias models, structural equation modeling, multi-level modeling, causal inference, and social network analysis. The final section takes up qualitative methods, including content analysis,