



Journal of Statistical Software

December 2005, Volume 14, Book Review 8.

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

Reviewer: John Fox
McMaster University

The Basics of S-PLUS, Fourth Edition

Andreas Krause and Melvin Olson
Springer-Verlag, New York, 2005.
ISBN 0-387-26109-5. xxii+442 pp. USD 59.95 (P).
<http://www.elmo.ch/doc/splus-book/>

Written to coincide with Version 7.0 of S-PLUS, the fourth edition of *The Basics of S-PLUS* furnishes a gently paced, but detailed, introduction to the S-PLUS statistical computing environment. The book assumes a basic understanding of statistical data analysis but no prior exposure to S.

Following a brief introductory chapter that provides some background information, the authors devote successive chapters to the graphical user interface (GUI) to S-PLUS; to an initial command-line session introducing interaction with the S-PLUS interpreter, vectors, arithmetic operations, and basic indexing; and to a “second session” discussing additional data structures (matrices, arrays, lists), simple user-defined functions, and the representation of missing values. Chapters 5 and 6 take up, respectively, traditional S graphics and **Trellis** graphics. It is easy to get lost in the intricacies of **Trellis** displays, and Krause and Olson’s easy-to-follow coverage of the topic is one of the high points of the book. All of the material in the first several chapters—and in much of the rest of the book as well—deals with the practical details of formulating S-PLUS commands. A nice feature of the book is the exercises, which appear at the end of each chapter, and for which detailed answers are provided.

Chapter 7 describes numerical summaries and graphical methods for exploring data. Chapter 8 introduces statistical modeling in S-PLUS, describing primarily the `lm()` and `aov()` commands for fitting linear models, although there is also a brief discussion of fitting logistic-regression models with `glm()`, and of estimating Kaplan-Meier survival curves with the `survfit()` function. There is a description of the general S formula notation for specifying linear (and related) statistical models, but the treatment of the subject is quite superficial.

Chapters 9 and 10 introduce programming in S-PLUS. The first of these chapters deals with basic programming techniques, such as operations on lists, function definition, control structures, and debugging. Chapter 10 describes the “S4” object-oriented programming system described by [Chambers \(1998\)](#). Despite its brevity, this chapter is another strong point of the book, clearly negotiating a potentially difficult subject. Curiously, there is no discussion of the older, easier to understand, and more extensively used “S3” object-oriented programming

system. Although S4 objects are arguably to be preferred for new projects, so much existing S code uses S3 objects that it is hard to understand why one would ignore the topic.

The last several chapters deal with odds-and-ends: data input and output (Chapter 11); miscellaneous “tips and tricks” (Chapter 12); interactions between S-PLUS and the operating system, and scoping issues (Chapter 13); sources of information about S-PLUS (Chapter 14); and differences between S-PLUS—the proprietary implementation of the S programming language and statistical-computing environment, and the focus of this book—and R, the free, open-source implementation of S (chapter 15).

A book that makes it to a fourth edition is obviously doing something right, and there is indeed a lot to like about this latest incarnation of *The Basics of S-PLUS*: The novice user who works carefully through the book will acquire substantial facility with S-PLUS.

Nevertheless, there are several aspects of the book that are unfortunate:

- The writing is often awkward, and occasionally obscure. Consider, for example, the following passage (on page 105): “If not the entire data frame consists of numeric variables, the result is hard to predict because the data frame will be converted before running `apply` on it.” It is, frankly, difficult to fault the authors here (who, I believe are not writing in their first language), but the publisher might have invested more in copy-editing the fourth edition of a clearly successful book.
- There are occasional slips. For example, in their discussion of **Trellis** graphics, the authors refer in passing (on page 163)—and without attribution—to Cleveland’s idea of adjusting the aspect ratio of a graph to “bank” the data to 45 degrees (Cleveland 1993), a reference that will surely prove cryptic to those unfamiliar with this idea.
- The statistical examples in the text are developed only in a cursory manner. Given the authors’ focus on the mechanics of using S-PLUS, however, this is not a serious fault, and after working their way through Krause and Olsen’s text, readers will be ready to tackle other sources on statistical data analysis and programming in S.
- As I have mentioned, the authors ignore the ubiquitous S3 object system.
- Finally, I would have preferred to see some discussion of the differences between S-PLUS and R—and of the unique features of R—throughout the text, rather than in a very brief chapter on R tacked on at the end. It is true, as the authors mention, that most of the code in the book will run under R as well as under S-PLUS, but where there are differences, the focus is clearly on the latter. This is not unfair given the title of the book, which refers only to S-PLUS, but it does restrict the book’s utility at a time when, by all indications, use of R is expanding rapidly.

In sum, and despite these reservations, the *Basics of S-PLUS* generally succeeds at providing a gentle, extended introduction to S-PLUS.

References

Chambers JM (1998). *Programming with Data: A Guide to the S Language*. Springer-Verlag, New York.

Cleveland WS (1993). *Visualizing Data*. Hobart Press, Summit NJ.

Reviewer:

John Fox
McMaster University
Department of Sociology
1280 Main Street West
Hamilton, Ontario, Canada L8S 4M4
E-mail: jfox@mcmaster.ca
URL: <http://socserv.socsci.mcmaster.ca/jfox/>