

A Review of *An R and S-Plus Companion to Applied Regression*

Malcolm J. Ree

Our Lady of the Lake University

Here is a book that John Fox produced as a companion for a text on regression. I acquired the book from the publisher with the idea of using the R and S-Plus programs for simulation studies of imputation of missing data. It is impossible to review this book without also reviewing R and S-Plus computer languages, but first the book.

The *Companion* is available in both hard and soft cover and is very reasonably priced. Because it is tied to another text by Fox (1997) and to a programming environment, it first introduces the reader to R and S-Plus. It does so through simple examples based on data that are provided on-line. The reader is lead through a basic introduction to methods of reading data, exploring and transforming data, and then fitting linear models. Past this are chapters on more advanced topics such as the generalized linear model, regression diagnostics, graphics, and writing your own programs in R and S-Plus. It is best to read this book with your computer turned on and with R or S-Plus on the monitor. A high-speed Internet connection is desirable.

A very complete website is provided for the book with revisions on a timely basis. The website has numerous links and useful information including a link to an 8.5 by 11 inch "R Reference Card" by Jonathan Baron. Thanks Jonathan! Additionally the book directs the reader to other R and S-Plus websites including CRAN, the Comprehensive R Archive Network that has numerous statistical functions that can be downloaded and executed immediately.

The book is written in a lively style with an occasional witticism interspersed. The writing is scholarly without being stilted. For example: "Examination—particularly graphical examination—of the data is an important prelude to statistical modeling, and a step that is skipped at the peril of the data analyst." (p. 85)

Of special note was the chapter on data transformation (Chapter 3). Numerous transformations and their appropriate use are presented. The graphic capability of R and S-Plus add to the explanation of the utility of the transformation. Scatter plots presented in figures 3.7, 3.8, and 3.9 are especially instructive. The chapter (Chapter 5) on the generalized linear model was particularly well written and very informative. I particularly like the handling of ANOVA as a specific case of linear

John Fox. *An R and S-Plus companion to applied regression*. Thousand Oaks, CA: Sage, 2002. 312 pages.

models. I would have handled the coding of “dummy” variables a little differently with more of an emphasis on potential linear dependency and adjusting degrees of freedom. Further, I always preferred to call them the more descriptive term—“categorical.” These are small issues and may very well be handled in the regression text. The target audience for the *Companion* should be the same as for the regression text. The *Companion* seems easy enough for an advanced undergraduate class or for a first multivariate graduate course.

The R and S-Plus programming environment deserve specific comment. Both are powerful high-level programming languages. S was originally developed by Bell Labs and continues to be developed. S was originally available from Bell Labs but now is only available commercially from Insightful Corporation. The commercial version of S-Plus has the advantage of an easy to use GUI. In many ways, it is similar to programs like SPSS but has the capability to use numerous programs built from script that conduct analyses either not programmed or not easily programmed in SPSS. A demo disk is available from Insightful Corporation.

R is a free source version of the language that can be downloaded via the Internet. It differs in some ways from S-Plus and does not come with a GUI. However, it runs in a Windows interface and does not require you to operate in DOS mode. MAC and UNIX versions are also available. Much of what follows is applicable to both R and S-Plus. In the *Companion*, Fox points out the differences in text boxes.

Both R and S-Plus are translator languages meaning that they do not compile an executable object but rather are executed line by line in real time. This is both bad and good. It is bad in that it is less efficient and good that changes can be made easily and quickly without time out for compiling. The language reminds me of a 1970s language called APL. It too was a translator language and had powerful operators. A typical R or S-Plus line might look like:

```
“Sample.20<-sort(sample(nrow(women1f), 20) #20 random observation”
```

This code (p. 66) selects and sorts a random sample of twenty observations from the data set “women1f”) and assigns them to the variable Sample.20. The R and S-Plus languages have numerous built in functions (Think of a “function” as a program in R or S-Plus) to compute things such as means and other descriptive statistics as well as familiar regression techniques. As stated before, CRAN and other sources have literally hundreds of other functions that are free and ready to use. It is in this environment that the *Companion* works.

Baker (2002) has said it nicely addressing the combination of the *Companion* and the Fox (1997) statistical text. He observed that Fox has “. . . made a fantastic contribution to the world of quantitative social science methodology.” (page 6). My assessment exactly.

References

- Baker, R. (2002). [Review of the book *Applied regression analysis, linear models, and related methods and R and S-Plus companion guide from the graduate student perspective*], *The Political Methodologist*, 11, 5–6.

A Review of An R and S-Plus Companion to Applied Reg

Fox, J. (1997). *Applied Regression Analysis, Linear Models, and Related Methods*. Thousand Oaks, CA: Sage.

Author

MALCOLM J. REE is Associate Professor, Center For Leadership Studies, Our Lady of the Lake University, 411 S. W. 24th St., San Antonio, TX 78207; reemal@lake.ollu.edu