

Book Review of Julian J. Faraway: Linear Models with R

Chapman & Hall/CRC, Boca Raton FL, USA, 2005
229 pages, ISBN 1-58488-425-8
<http://www.stat.lsa.umich.edu/~faraway/LMR>

This book is useful to serve for the practical aspects of an advanced undergraduate linear regression course. Regarding the potential readership, the Author writes in the preface that the book is “not introductory” and that it “presumes some knowledge of basic statistical theory and practice. Readers are expected to know the essentials of statistical inference such as estimation, hypothesis testing and confidence intervals. A basic knowledge of data analysis is presumed. Some linear algebra and calculus are also required.” Thus this book is most suitable for undergraduate statistics majors at least half way through their degrees. The book needs to be accompanied by a theoretical book, such as [Seber and Lee \(2003\)](#). A somewhat similar competitor to the book is [Fox \(2002\)](#).

With a large number (16) of chapters in this smallish book, most of them are short and mention ideas briefly. Of course, some chapters are on core practical topics such as residual diagnostics, transformations, weighted and generalized least squares, ANOVA (factorial designs, block designs), ANCOVA, variable selection techniques, and inference. A good feature of the book is that more ‘optional’ topics are covered, include missing values, regression splines, robust regression (M -estimation and least trimmed squares), permutation tests, shrinkage methods such as partial least squares, measurement error models including SIMEX, latin squares and iteratively reweighted least squares. Most instructors are unlikely to be familiar with all these topics, and the book does a good job in giving a very gentle introduction to these topics. The reader is usually referred to references at the back for further reading.

The book has two small appendices. The first describes R installation, functions and data, while the second is a quick introduction to R. The examples in the book are based on R 1.9.0. In my copy there were a few font problems, e.g., p.189 has $(\phi)_{ij}$ instead of $(\alpha\beta)_{ij}$.

The tone of the book is that of an informal tutorial with some necessary theory interdispersed throughout. Explanations are good overall, and the mathematical notation is quite standard. The Author’s website serves as the repository for the book’s package, called *faraway*, and includes errata, R commands and data sets (the package is also on CRAN.) The Author writes “Data analysis cannot be learned without actually doing it” and this is facilitated by practical exercises (which are usually short and

sharp) at the end of each chapter. Solutions appear unavailable.

The book centers on `lm()` and assumes familiarity with the S language, e.g., I found it referred to “.” in a formula without reminding or telling the reader that it means all variables in the data frame other than the response. The book is rather unsuitable as a reference book, for example, summation constraints $\sum_{j=1}^J \alpha_j = 0$ is a popular parameterization for a factor with J levels. The function `contr.sum` is alluded to but not mentioned directly in the book, nor how to switch between different parameterizations. Ideally, a handy summary of topics such as the Wilkinson-Rogers operators (e.g., `*`, `/`, `:`) and all the different contrast options available should be put into an appendix for fast and easy reference. This would make [Chambers and Hastie \(1991\)](#) less necessary for students to buy or refer to. The book makes good and frequent use of basic graphics with a liberal sprinkling of plots everywhere, and many examples.

Altogether, there are 27 data sets used in the book; the large majority of these are small. Even though the figures are in black and white, the reader is not encouraged to use colors in the plotting—something very useful on a computer screen. It would be good also that the reader be reminded how to create PDF files of figures (especially if the user uses Linux), which is useful if the user writes reports.

The book has many good points but I picked up a few bad points. For example, the use of extractor functions is not always taken where possible, e.g., `coef(fit)` should be used instead of `fit$coef`. Also, numerous examples use cut-and-paste from previous output rather than extracting quantities from objects, for example, p.20, `60.975` is used rather than `mdl1$sigma`. I would have liked the paragraph on additive models expanded to make the reader aware of how to check the linearity assumptions using `lm()` and regression splines.

I thought there were a number of omissions. The prediction problem should be mentioned (in S-PLUS data-dependent terms such as `scale(x)`, `poly(x, 2)` and `bs(x)` give problems, whereas in R, a data-dependent function called inside another function are problematic, e.g., `poly(scale(x), 2)`). Other omissions include not mentioning computational details such as the QR-algorithm and not using more advanced graphics such as those found in the `lattice` package.

Although I agree that the book is “not introductory”, neither is it advanced. For example, a broken stick regression model is fitted in

Section 7.2.1. Improvements to this example include using `coef()` instead of `$coef`, and using the `predict` generic function, i.e., `predict(gb, data.frame(pop15=seq(20,48,by=1)))`. The two (dotted) lines described in the text actually do not show up on the plot. Furthermore, although it is instructional to create two basis functions `lhs()` and `rhs()`, the reader is not informed that `bs(pop15, degree=1, df=1, knot=35)` would be equivalent.

In conclusion, the book is quite suitable to serve for the practical component of an advanced undergraduate course in linear models to reasonably prepared students. A sequel on generalized linear models and extensions would be a natural next step!

R Foundation News

by Bettina Grün

Donations and new members

Donations

Adelchi Azzalini (Italy)
 BC Cancer Agency, Vancouver (Canada)
 David W. Crawford (USA)
 Peter L. Flom (USA)
 Google Inc., Mountain View, California (USA)
 Faculty of Economics, University of Groningen (Netherlands)
 Shigeru Mase (Japan)
 Merck and Co., Inc. (USA)
 Network Theory Ltd, Bristol (United Kingdom)
 Stanford University, California (USA)
 Douglas Wilson (Canada)
 Ivo Welch (USA)

New benefactors

Burns Statistics Ltd., London, U.K.
 Loyalty Matrix Inc., California, USA
 Statisticon AB, Uppsala, Sweden
 Merck and Co., Inc., USA

Changes in R

by the R Core Team

User-visible changes

- box plots by `boxplot()` or `bxp()` now have the median line three times the normal line width

Bibliography

- J. Fox. *An R and S-Plus Companion to Applied Regression*, 2002. Thousand Oaks, CA: Sage Publications. [56](#)
- G. A. F. Seber and A. J. Lee. *Linear Regression Analysis*, 2nd Edition, 2003. New York: Wiley. [56](#)
- Chambers, J. M. and Hastie, T. J. *Statistical Models in S*, 1991. Pacific Grove, Calif.: Wadsworth & Brooks/Cole. [56](#)

Thomas Yee
 University of Auckland, New Zealand
yee@stat.auckland.ac.nz

New supporting institutions

Baxter Healthcare Corp., California, USA
 Department of Mathematics and Statistics, Utah State University, USA
 Department of Statistics, Iowa State University, USA
 Dipartimento di Statistica, Università Ca' Foscari di Venezia, Italy
 School of Economics and Finance, Victoria University of Wellington, New Zealand

New supporting members

Claudio Agostinelli (Italy)
 Roger J. Bos (USA)
 Dianne Cook (USA)
 Gregor Gorjanc (Slovenia)
 Ivan Kojadinovic (France)
 Iago Mosqueira (Spain)
 Jonas Ranstam (Sweden)
 Christian Schulz (Germany)
 Mario Walter (Germany)

Bettina Grün
 Technische Universität Wien, Austria
Bettina.Gruen@ci.tuwien.ac.at

in order to distinguish it from the quartile ones.

- Unix-alike versions of R can now be used in UTF-8 locales on suitably equipped OSes. See the internationalization section below.