



Journal of Statistical Software

January 2007, Volume 17, Book Review 4.

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

Reviewer: Juana Sanchez
University of California, Los Angeles

Linear Models with R and Extending the Linear Model with R

Julian J. Faraway

Chapman & Hall/CRC, Boca Raton, Florida, 2004 and 2006.

ISBN 1-58488-425-8 and 1-58488-424-X. 229 and 297 pp. USD 69.95 and USD 79.95.

<http://www.maths.bath.ac.uk/~jjf23/>

The books “Linear Models in R” and “Extending the Linear Model with R” by Julian J. Faraway (hereby referred to as Volume 1 and 2 respectively) are a joy to read for anybody interested in applications of linear models and R. The author does an excellent job at promoting good data analysis practices. The applications are interesting and each is guided by a research question. The R code used throughout, perfectly integrated with the text, is smart and very well explained in the context of the examples.

The data sets are available in the books’ web sites. No matter how expert the reader is, there is an unexpected new insight or way of seeing the theory at every corner. Practice problems at the end of each chapter are very good. Faraway summarizes each topic in linear models so well and with such good writing style, that the books read effortlessly. The organization is excellent. These two books should be in every statistician’s library and should be recommended to any student of statistics and researchers.

Although the emphasis is on applications, and the books claim that the discussion can be followed by anybody with introductory knowledge of statistical inference and no knowledge of R, the joy among this audience may be mitigated by the effort needed to understand the concepts and R codes. This audience will find the books a great complement to a more theoretical book and will benefit from an instructor. More advanced data analysts, eager to apply the models in any discipline and with no time to dwell in theories will probably find them very helpful as well. The graduate student researcher that has already had a course in linear models, will find in these books a companion not to leave home without, the equivalent of what S-PLUS books were 10 years ago. Those expert in linear models but desiring to move to using R as the software in their applications, will not be able to live without the books, as no other books in the market are such a self-contained set of manuals for applied linear models. And, paradoxically, these books are a must read for experts in the theory of linear models, because the pedagogical and intuitive way in which some theoretical concepts are explained in these books can not be found in existing theory books.

The two volumes are very modern and are well organized manuals for using R to do multiple regression, its 2 special cases, ANCOVA and ANOVA, and extensions such as generalized linear models, mixed effects models and nonparametric regression models. All of these are explained with equal depth and quality. Both books start by explaining how to summarize data and promoting good data analysis. In each subsequent chapter, in both books, an interesting data set works as an example with a research question to illustrate the application of the models. Practical problems such as coding, redefining and transforming variables, are discussed if needed. The R programs, output and graphics for each of the examples of data discussed are given within the text, without breaking the continuity of the text, and many useful hints about what R does in the background are given, too. Occasionally, the author shows how to compute things from scratch (p. 28 and 29), for instance, how to compute the F test statistic in Chapter 2 of the first book, thus preparing the reader for whatever programming they might want to do that is not pre-packaged in R. In an effort to keep the book short, only the aspects of the theory of linear models that are useful for each practical regression analysis or its extensions are outlined, with depth just enough to be able to know what the numerical output and graphics given by R are; proofs are rare. There is very little jargon typically found in the theory of the linear model. But quite often, Faraway dwells into a nice geometric or intuitive interpretation of formulas that usually is lacking in many other books on regression analysis widely used, both theoretical and applied. This trademark of these books make them remarkable and a real pleasure for teachers of linear models and applied researchers as well.

Although both books conform to what we say above, the second book conforms more to it than the first. In the first volume, there are 16 chapters. Chapter 1 warns about data quality and issues such as missing data, so it recommends a great deal of preliminary univariate and bivariate data analysis to discover outliers, data entry errors and so on before engaging into multiple regression. More warning about unreasonable analysis are given throughout the book (p. 15, Section 2.5, where the author warns about choosing your output reasonably, too, Section 2.9, and many sections thereon). On p. 16, there is a typo: Gauss-Markov theorem requires equal variances of errors—the book says unequal. The reader can start getting the general flavor of the book on Chapter 2. That is, the reader will notice that the authors do not dwell much into interpreting the R output results in the context of the data set analyzed, hinting that the story behind the data is not the main point. There are no research questions posed about the data. The data sets are used to illustrate how to use R in different contexts and to point out the problems one needs to be aware of when doing multiple regression. The exercises can be described the same way. Thus, if this book was used as a textbook on introduction to regression, the instructor would have to spend quite some time with the interpretation of the regression coefficients. The chapter on inference, Chapter 3, starts with the likelihood ratio statistic, thus only students with some math stats background will be able to go through that. The flavor of all the remaining chapters is similar to Chapters 2 and 3. While the reader familiar with linear models will find this volume delightful reading and the researcher will make good use of it, the beginning student or applied researcher will have to work hard at the beginning and get some help. Volume 2 on the other hand, reads as if the author assumed that Volume 1 was already read, because the emphasis is less on technical aspects and more on the effective application, the interpretation and answering the research question. Perhaps the authors gained in expertise or responded to evaluations of the first volume, but certainly the second volume makes more mature writing and better pedagogy

than the first. It is also much better written.

I would use both volumes for a course sequence that introduces regression. The effort of a teacher in such a class would be only on explaining the concepts in more detail in lecture and guide the students. Everything else, the lab work, the code needed, the examples, the research questions, are done for the teacher. Most books require the instructor to do both tasks. And I certainly will keep these books as precious resources for my research.

Reviewer:

Juana Sanchez

University of California, Los Angeles

Department of Statistics

8125 Math Sciences Building, Box 951554

Los Angeles, CA 90095-1554, United States of America

E-mail: jsanchez@stat.ucla.edu

URL: <http://www.stat.ucla.edu/~jsanchez/>