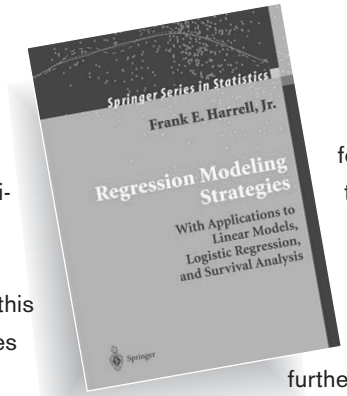
BOOK REVIEW

Frank E. Harrell, 2001:
Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis.
New York: Springer-Verlag. xxiii + 568 pp; illus. $79.95 (HC).
ISBN 0 387 95232 2.



There are many books that are excellent sources of knowledge about individual statistical tools, but the art of data analysis is about choosing and using multiple tools. Instead of presenting isolated techniques, this book emphasises problem solving strategies that address the many issues arising when developing multivariable models using real data and not standard textbook examples. A standard regression book would be expected to include linear and nonlinear regression, but the author goes much deeper including imputation methods for dealing with missing data effectively, methods for dealing with nonlinear relationships and for making the estimation of transformations a formal part of the modelling process, methods for dealing with 'too many variables to analyse and not enough observations', and model validation techniques based on the bootstrap. Most of the methods discussed in the 20 Chapters apply to all regression models, but the author gives special emphasis to some of the most popular ones: multiple regression using least squares, the binary logistic model, two logistics models for ordinal responses, parametric survival regression models, and the Cox semiparametric survival model. There is even a chapter on nonparametric transform-both-sides regression. The author's aim was to provide a 'hands-on', ready to use book for Masters or PhD graduate students who have had a general introductory probability and statistics course and who are well versed in ordinary multiple regression and intermediate algebra. This book can also serve as a reference for data analysts and statistical methodologists as it contains an up-to-date survey and bibliography of modern statistical regression modelling techniques. The author provides many very useful 'rules of thumb' and 'strategies' based on his own impressive experience. These allow the reader to develop a strategy that is tailored to a particular problem. They are presented depending on whether the goal of the problem is prediction, estimation or hypothesis testing. Emphasis is given to detailed and very useful case studies (mainly in 7 Chapters) for these methods as well as for data reduction, imputation, model simplification, and other tasks. These case studies use freely available high-level S-PLUS functions that make the multiple imputation, model building, validation, and interpretation tasks described in the book relatively easy to do; see below for further information. The majority of examples are from biomedical research. However, the methods presented have a broad application to other areas. The book uses extensively two excellent libraries of available S-PLUS functions ('Design' and 'Hmisc') written by the author and which can be downloaded freely from the book's homepage at http://hesweb1.med.virginia.edu/biostat/rms/ The latter holds most of the data sets mentioned in the book for downloading. Links to additional related resources are also found on this site, as is information for instructors, extra problems, solutions to these and to many of the book's problems, and possible syllabuses for courses using parts of the book. Finally, note that the S-PLUS libraries were recently adapted to the open-source statistical software package R (http://www.r-project.org). The author has a very motivating style and includes opinions, remarks, and summary and 'Further Reading' sections that lead to further thoughts and enable the reader to deepen her/his knowledge. The logical path chosen on how to present the material is excellent. In summary, this book realistically deals with model uncertainty and its effects on inference to achieve what the author calls 'safe data mining'. Overall speaking, considering the fun I had reading the book, I think that the author's aims are met and I highly recommend everybody to have a look at the book. Moreover, I recommend any library purchasing the book.

Dr. Diego Kuonen
Statoo Consulting
PO Box 107 1015 Lausanne
Switzerland
kuonen@statoo.com