

book.org), where instructions for downloading packages with the datasets for R and Python users are given, together with an essential part of the book, and additional materials. Other resources on modeling in R and Python and on meaningful regressions can be found in the given references.

References

- Lipovetsky S. (2020), *Introduction to Data Science: Data Analysis and Prediction Algorithms with R*, by Rafael A. Irizarry, *Technometrics*, 62, 280–282.
- (2020), *Statistical Inference via Data Science: A Modern Dive into R and the Tidyverse*, by Chester Ismay and Albert Y. Kim, *Technometrics*, 62, 283.
- (2021), *Advanced Statistics with Applications in R*, by Eugene Demidenko, *Technometrics*, 63, 273–275.
- (2021), *Linear Models with Python*, by Julian J. Faraway, *Technometrics*, 63, 426–427.
- (2021), “Game Theory in Regression Modeling: A Brief Review on Shapley Value Regression,” *Model Assisted Statistics and Applications*, 16, 165–168.
- (2021), “Modified Ridge and Other Regularization Criteria: A Brief Review on Meaningful Regression Models,” *Model Assisted Statistics and Applications*, 16, 225–227.

Stan Lipovetsky
Minneapolis, MN



Semiparametric Regression with R, by Jaroslav Harezlak, David Ruppert, and Matt P. Wand. Springer, 2018, ISBN: 978-1-4939-8853-2, xi + 331 pp., eBook, \$87.20.

This book covers the fundamentals and recent developments of semiparametric regression, and related approaches. Classical, mixed model, and Bayesian approaches are given for most of the methods. Since it is part of the Use R! series, the focus is on R and, for Bayesian applications, STAN (via the rstan package). The number of packages in R is bewildering, and this book provides a valuable and welcome roadmap to those that implement semiparametric methods. It is applications-focused but has plenty of solid mathematical background. More theoretical aspects of the material are given by a companion book by the same authors (Ruppert, Wand, and Carroll 2003).

The coverage of this book is very thorough. Here is a brief overview. After some introductory material, penalized splines are described, focusing on scatterplot smoothing. Since penalized splines are at the heart of many semiparametric applications, sufficient time is dedicated to providing a good basis of understanding. Of the many methods for implementing penalized splines, the book uses O’Sullivan spline basis functions. A mixed model representation is then presented, followed by a Bayesian approach. Semiparametric additive models are then introduced, where parametric model terms for factors, continuous variables, and interactions are added to the semiparametric terms. These approaches are extended to generalized additive models. Bivariate function estimation is introduced (via tensor product and thin plate splines). Covariance function estimation is described, along with the use of principal components dimension reduction for high-dimensional data. The book ends

with miscellaneous topics: robust/resistant smoothing, including quantile regression, kernel machines (including support vector machines), and approaches to missing data and measurement errors via acyclic directed graphs to construct elegant and intricate Bayesian models.

Unlike more reputedly stable software systems, such as SAS and STATA, texts that are based on R may require more updating. The companion website, <http://semiparametric-regression-with-r.net>, contains supporting information about the book’s companion HRW package, software updates, and errata. The book uses many different R packages, and all are subject to changes that could break the code as displayed in the book. The *software updates* document (current as of April 2021) describes the changes needed. Notably, when R Version 4 was introduced the fundamental properties of factors were changed, causing havoc with older code. So, there were many places in the book’s examples that required revision. The Bayesian applications use STAN which is under active development and is a source of future code fragility; this is pointed out by the authors. Readers wanting to run all the examples in the book are warned that there may be some “bumps in the road” at times. Nevertheless, the authors appear committed to maintaining the book and the HRW package.

The book is very well written and produced. So is the code used in the examples. However, most of the examples show the code with the console prompt (“>”) and line continuation characters (“+”), making it tedious to copy and paste into R. Suppressing these characters would make copy/paste much easier. However, all the code in the examples is contained in files stored in the HRW package folder, and the book gives pointers for finding these files.

Overall, this is a very impressive work. It could easily serve as a text for graduate courses in statistics and data science programs. To that end, each chapter concludes with exercises, some of which could be challenging. The datasets used in the examples are not “toy datasets” that are used in many texts. They are substantial real-life datasets obtained from selected packages (HRW, mlbench, Ecdat, refund, and fda.usc), and cover applications across a wide variety of fields.

Reference

- Ruppert D., Wand, M.P., and Carroll, R.J. (2003), *Semiparametric Regression*. Cambridge, New York: Cambridge University Press; pp. 386. [145]

Charles E. Heckler
University of Rochester - Retired



Statistical and Machine-Learning Data Mining: Techniques for Better Predictive Modeling and Analysis of Big Data, 3rd Edition, by Bruce Ratner. Chapman and Hall/CRC, Taylor & Francis Group, Boca Raton, FL, 2020; ISBN 978-0-367-57360-7; xxxiii + 655 pp., \$54.95 (pbk).

The book’s 3rd edition has been significantly extended to 44 chapters from 31 chapters of the 2nd edition in 2011 (the