Reviewer: Robert Gould
University of California, Los Angeles

A Modern Approach to Regression with **R**

Simon J. Sheather
Springer-Verlag, New York, 2009.
ISBN 978-0-387-09607-0. 393 pp. USD 89.95.
http://www.stat.tamu.edu/~sheather/book/

A Modern Approach to Regression with **R** is aimed, according to its preface, at first-year graduate students and senior undergraduates. This book fills an important niche in the regression textbook by providing a data-centered approach strong on graphics. *Modern Approach* is very much in the spirit of Cook and Weisberg (1999) and Weisberg (2005).

I have taught an undergraduate applied regression course for several years, and found few textbooks that are appropriate for modern students. To reveal my prejudices up front, what I want in an applied regression text are data, graphics, and computation. Data, because an applied regression course should give students an approach for understanding the world as it is revealed through data; the regression model is an end to a means but should not be treated as the end itself. Graphics, because not only are they important to modern regression, but they can help students develop a conceptual understanding of regression models. Computation, because data analysis via graphical and numerical regression is not possible without statistical software. The choice of software not only determines the sort of diagnostics and visualizations available to one's students, but also influences how students understand regression itself.

Sheather defines regression as "a method for investigating the functional relationship among variables". This definition puts the emphasis in exactly the right place and signals that the book is teaching a way of examining data, and is not a study of the linear model itself. An emphasis on real-world problems is carried throughout the book. A great strength of this book is the numerous case studies, which make for interesting reading and always include a clearly stated goal. Four primary case studies are introduced in great detail in the first chapter, and later chapters expand on these studies and introduce new problems, too. The examples are diverse: NFL kickers, wine critics, newspaper circulation, restaurant menus.

Modern Approach differs from some other books in that it provides a complete examination of simple regression (including weighted least squares) before beginning multiple regression. Chapters 2–4 cover all aspects of simple regression, and these chapters mirror the development of the multiple regression in the rest of the book. When teaching undergraduates, as I do, this means a substantial part of the ten-week term is spent on simple linear regression. This

is both good and bad. An advantage is that students learn the basics of regression in the most easily understood setting. Much of multiple regression then seems like "review". On the other hand, some topics, such as transformations, can appear confusingly elaborate when they are revisited, and some students are confused by what they see as a similar, but different, presentation of a topic they thought they had already learned. Another disadvantage is that one of the challenges of teaching multiple regression (at least to undergraduates) is teaching how to interpret a model. For example, what does it mean to say "controlling for other variables"? Students need to practice in a variety of contexts to begin to understand this, and this practice takes time. Interpretation in simple linear regression, on the other hand, is fairly straightforward. As a result, the first few chapters of this book emphasize the craft and technicalities more than interpretation. In latter chapters, interpretation is emphasized more, but, at least in a short academic term, there might not be enough time for this to sink in.

For a book that has ". . . with R" in the title, there is very little R code. The book itself is R-centric; output and graphics are clearly generated by R, and analysis procedures are consistent with functions provided by such packages as **MASS**, **car**, and **alr3**. Still, students will need to access the author's website to view the R code used to generate the text's figures and output. (Code is also provided for Stata and SAS.) This book will not teach students how to use R. If this is a student's first exposure to R (as is quite plausible for a regression course), then much will need to be explained. Students will likely need tutorials on some basic data management, the use of the formula syntax, and help understanding how packages are installed. Banishing R to the website is not necessarily a bad thing; I know of no regression books that successfully integrate the teaching of R, and the separation strategy is as good as any other I have seen.

I am particularly interested in teaching regression to undergraduates, and I used this book one term in an introduction to applied regression course. Previously, I used Weisberg (2005) and feel that Sheather improves on this book in many ways, at least as far as the undergraduate experience goes. The in-depth case studies are instructive and interesting, and the writing is very clear. The discussions of colinearity and model selection were particularly clear. undergraduates are fascinated with transforms, and will use them again and again to try to solve almost any problem, and Sheather provides just the right level of detail to help them understand how to use transforms without encouraging them to go transform-wild. As an introduction for undergraduates, though, the book falls short, but not irreparably so. Undergraduates need more examples of the consequences of failure to meet model conditions. They need a clear separation between tools for checking model validity and goodness of fit. (Students apparently arrive their first day of class believing that R^2 will measure everything.) They need more guidance in why the diagnostic plots take the shapes that they do. But these are aspects that the instructor can supplement.

In summary, this is a book I will use again, unless or until a more fully fleshed undergraduate version appears. Graduate students in particular will find the balance between applications and theory useful, and the minimal amount of formulae used means the book should be useful for students from a variety of disciplines. The well-motivated homework problems are interesting and sufficiently complex that students at all levels will be able to learn something from them.

References

Cook RD, Weisberg S (1999). *Applied Regression Including Computing and Graphics.* John Wiley & Sons, New York.

Weisberg S (2005). *Applied Linear Regression.* 3rd edition. John Wiley & Sons, New York.

Reviewer:

Robert Gould
Department of Statistics
University of California, Los Angeles
Los Angeles, CA 90095, United States of America
E-mail: rgould@stat.ucla.edu
URL: http://www.stat.ucla.edu/~rgould/