Journal of Statistical Software

November 2015, Volume 68, Book Review 1. doi: 10.18637/jss.v068.b01

Reviewer: Joseph M. Hilbe
Arizona State University

The New Statistics with R: An Introduction for Biologists

Andy Hector
Oxford University Press, New York, 2015.
ISBN 978-0-19-872906-8. 208 pp. USD 49.95 (P).
http://ukcatalogue.oup.com/product/9780198729068.do

The New Statistics with R: An Introduction for Biologists was written for researchers in the biological sciences, and ecology in particular, to help provide them with the statistical background needed for their research studies. The text, about 200 pages in length, provides instructions on how to construct, interpret, and evaluate foremost statistical methods to biological data. R is used throughout the text for examples, which are all based on real data.

Following a six-page introductory chapter, the statistical subjects examined in the book include, Chapter 2: *Comparing Groups: Analysis of Variance*; Chapter 3: *Comparing Groups: Student's t-test*; Chapter 4: *Linear Regression*; Chapter 5: *Comparisons Using Estimates and Intervals*; Chapter 6: *Interactions*; Chapter 7: *Analysis of Covariance: ANCOVA*; Chapter 8: *Maximum Likelihood and Generalized Linear Models*; Chapter 9: *GLMs for Data with Non-Normal Distributions*; Chapter 10: *Mixed Effects Models*; Chapter 11: *Generalized Linear Mixed-Effects Models*; and Chapter 12: *Final Thoughts*. An appendix is provided that reviews the major points of R.

The level of discusson is at the graduate school and professional researcher/analyst level. Accomplished upper division undergraduates in ecology or the environmental sciences who have successively passed a first level statistics course with a regression component should be able to understand and use the book as well. Readers do not need to have previous knowledge of R, although it would help.

The book is thoroughly within the frequentist framework. Bayesian methods are not mentioned. Moreover, no reference section is given in the book. Citations are provided in the context of the discussion. For example, in a sidebox on page 108, the author states at the end, "... beyond the scope of the book, but see Hector, A., von Felton, S., and Schmid, B. (2010). Analysis of variance with unbalanced data: an update for ecology and evolution, *Journal of Animal Ecology* **79**: 308–316)." This is a rather unusual manner of citation, and makes it difficult to track references given in the book. On the other hand, in the *Final Thoughts* chapter, the author provides a "Further reading" box with suggestions of books, articles, and websites which the author recommends to the reader for additional information.

The various statistial methods that are discussed in the book are explained well. Code and explanation for designing graphs for assessing model fit is well done. I also like the fact that the author discusses current issues such as why it is advisable when modeling to focus on estimates and confidence intervals and not so much on p-values. In fact the author recommends not using p-values at all when reporting modeling results in a research report or journal article. There is also discussion about information criteria statistics and goodness-of-fit tests that go well beyond what is normally discussed in similar books of this size. I believe that these types of discussions are valuable. I would not make such an all-encompassing restriction on the use of reporting p-values, but the underlying advice of making sure of what p-values actually mean in a given context is very important to understand when modeling.

There are also places in text where there are mistakes. For instance, in Chapter 9 the author discusses R's `glm` function. He asserts that the quasibinomial option is a GLM family, which it is not. It is simply a post-estimation adjustment to model standard errors formed by multipling them by the square root of the Pearson χ^2 dispersion statistic. Standard errors adjusted in this fashion are commonly referred to as scaled standard errors. He also writes that the Poisson assumption of equidispersion is met when the ratio of the residual deviance and residual degrees of freedom approximates 1.0. However, simulation studies demonstrated some time ago that this ratio is based on the Pearson dispersion, not deviance disperson. This is consistent with how quasi-Poisson models are constructed. See Cameron and Trivedi (1998), Hilbe (2011), and Hilbe (2014) for details.

Except for the few errors – which are easy to fix – such as the ones mentioned above, the book is accurate and well written. However, I wish it had a reference section, as well as an errata document posted to the book's website. No such errata appears to have been prepared thus far, but it is advised. It would also have been helpful to readers to have a place on the book's website to download the R scripts used througout the book for examples. But given the otherwise excellence of the presentation of topics discussed, I recommend the book to analysts and researchers in ecology and the environmental sciences.

References

Cameron AC, Trivedi PK (1998). *Regression Analysis of Count Data.* Cambridge University Press, Cambridge.

Hilbe JM (2011). *Negative Binomial Regression.* Cambridge University Press, Cambridge.

Hilbe JM (2014). *Modeling Count Data.* Cambridge University Press, Cambridge.

Reviewer:

Joseph M. Hilbe
Arizona State University

T. Denny Sanford School of Social and Family Dynamics
Tempe, Arizona, United States of America
E-mail: hilbe@asu.edu
URL: http://works.bepress.com/joseph_hilbe/

Journal of Statistical Software http://www.jstatsoft.org/
published by the Foundation for Open Access Statistics http://www.foastat.org/
November 2015, Volume 68, Book Review 1 *Published:* 2015-11-18
doi:10.18637/jss.v068.b01