

Book review

KA Aho, *Foundational and applied statistics for biologists using R*. CRC Press: Boca Raton, FL, 2014; 596 pp. ISBN: 978-1-4398-7338-0 (hbk)

Reviewed by: Nick R Parsons, Warwick Medical School, University of Warwick, Coventry, UK

Do we really need another book describing the use of R for a range of widely used statistical analysis methods? There do seem to be a lot around – so does this book offer anything new or different from other publications?

The target readership for this book, as the title suggests, is biologists. The book contains a host of examples to illustrate various methods of analysis and statistical concepts; the majority of the example datasets used are from ecology, which is clearly where the author's interests lie. The book is divided into two main sections, *foundations* and *applications*. The former takes the reader on a journey from the very basis of scientific thinking, through probability theory, density functions, parameters and statistics, interval estimation, hypothesis testing and experimental design, with some nice historical background along the way. The statistical concepts described are illustrated with some terrific interactive GUIs and sliders; code to implement these is provided in the R package *asbio* which accompanies the text. More than a plaything for the distracted statistician, these are a great resource for teaching students and conveying statistical ideas to the non-statistically trained. The second section of the book focuses on applying statistics and implementation in R, covering correlation, regression, ANOVA and tabular analysis. As for all such general texts, the selection of the material is rather idiosyncratic. For instance, there is no discussion of time-to-event analysis or multivariate methods, and only cursory descriptions of mixed-effects models. These are topics that occur often in biological settings, with vast associated statistical literatures, that a reader will be surprised not to see discussed in such a text. On the other hand, the selected topics are covered in some depth; my preference would have been for increased breadth at the expense of some of the detail, which I suspect will be hard going to all but a few of the most mathematically inclined biologists. The distinction between the two parts (*foundations* and *applications*) seems fairly arbitrary at times, but overall the latter part of the book contains more detailed examples and R code that take one through, for instance, the basics of how to fit a linear regression, interpret and draw inferences on parameter estimates and undertake diagnostic testing of important assumptions.

There is no basic introduction to R syntax in this book, which is surprising in such a long and otherwise comprehensive text. One might argue that there are many such texts and resources available elsewhere, particularly online; however, a concise summary of basic syntax and methods for importing and exporting data would have been useful for an R novice using this as an introductory text. Clearly, one is never going to be able to cover all statistical topics and R implementation in sufficient depth to satisfy all readers, so it is perhaps churlish to list the omissions and inadequacies of what is overall a very readable and well-structured book.

My main observation after reading this book is exactly who the target readership really is; applied statisticians or biologists? The former group, where I would place myself, will find plenty to interest them in this text, as a source book for useful R code and as a teaching aid, where thanks to the

interactive tools, this excels. The latter group is more heterogeneous. The author clearly encounters many individuals with high-level mathematical skills, who are able to grasp matrix algebra and integral calculus; however, the majority of biologists will no doubt not be familiar with such methods, or at best have forgotten them. For this group, there are certainly more friendly (less mathematical) texts on the market that they will find preferable. Returning to where I started, does this book offer anything new, not available elsewhere in the crowded market of statistics books using R? Yes it does. The approach, focussing on statistical foundations first, building upwards from basic philosophical concepts, before progressing to implementation and real world applications, is certainly novel. I would, without a doubt, recommend this book to those statisticians working with biologists, and if they can persuade those biologists to read the book and start writing R code themselves, then I think this book will have succeeded in its aims.