MURDOCH RESEARCH REPOSITORY

This is the author's final version of the work, as accepted for publication following peer review but without the publisher's layout or pagination. The definitive version is available at

Brouwers, N.C. (2012) Biostatistical design and analysis using R: A practical guide (book review). Austral Ecology, 37 (3). e19-e19.

http://researchrepository.murdoch.edu.au/12292/

Book Review

Biostatistical Design and Analysis Using R: A Practical Guide.

Murray Logan . Wiley-Blackwell, Chichester, 2010. xxviii + 546 pp. Price AUD $79.95 (paperback). ISBN 978 1 4051 9008 4 (Hardback: AUD $200.00, ISBN 978 1 4443 3524 8).

The power of R is increasingly being recognized in the scientific community, more and more becoming the preferred tool for statistical analysis and graphical data presentation (and it is free!). This fact has inspired me to start performing all my analyses in R, walking the path of 'learning by doing', using my own ecological datasets. I had been looking for a useful reference book with real worked examples using R for some time, and along came this practical guide.

This book (or guide) builds on existing biostatistical textbooks and tries to cater for biological researchers and students that do not desire to spend much time programming in R to perform critical analyses and graphical presentation of their datasets.

The focus of this book is: 'to highlight the major concepts, principles and issues in contemporary biostatistics as well as demonstrate how to use R (as a research design, analysis and presentation tool) to complete examples from major biostatistics textbooks' (p. xvi).

Within this framework, the objectives of this book are stated to be threefold:

- 'To provide very brief refresher summaries of the main concepts, issues and options involved in a range of contemporary biostatistical analyses

- To provide key guides that steps users through the procedures and options of a range of contemporary biostatistical analyses

- To provide detailed R scripts and documentation that enable the users to perform a range of real worked examples from statistics texts that are popular among biological and environmental scientists' (p. xvii)

It meets these objectives very well!

The book is very complete in its approach, so it caters for the first-time user of R as well as people that are already up to speed with the basics of the R software and approach. It starts with talking you through the installation of the software, the basics on how the syntax works (i.e. what commands do you type in and what do you get out), and how to get your datasets into the right format to get going in R. For me these are vital points to address to get off to a good start with a minimum amount of frustration when learning a new software package.

The book then introduces briefly the principles of statistics and the importance of sampling design, which paves the way to the more detailed chapters that deal with the range of

statistical analyses most frequently used in the field of biostatistics. First, however, the book introduces the capabilities of R in visualizing data. This part is mainly focused on graphical data presentation using the most commonly used plotting functions in R (e.g. x/y plots, histograms and boxplots), again informing the more sophisticated examples found in the following chapters.

The rest of the book chapters are structured by introducing the various statistical approaches that can be used followed by detailed worked examples using real datasets. These datasets are downloadable from an accompanying website (http://www.wiley.com/go/logan/r), and the book uses these to guide you step-by-step through the analyses. Most importantly, it does not only explain and show you what commands to use, and the output it generates, it also explains how to interpret these outputs (although in some instances this could be more complete and detailed). It then often gives further examples of how to present the data that were analysed in graphical format to support the statistical results.

I started using this book from the point that I wanted to analyse my datasets, knowing what statistics to use, but not knowing how to do it in R. Skipping through most of the first introductory chapters (I do not recommend this though(!), see example below), it was easy to find what I was looking for and how to do it. Core features of this book are namely a handy statistics key to point you towards the right chapters in the book, a table with useful R commands also pointing you to the right locations in the main text, and two comprehensive R and statistics reference indexes in the back of the book.

When using the steps of the analyses as presented in the book using my own large dataset (instead of the ones used in the example available from the website), I naturally stumbled across problems in getting the different command syntax to work. However, the author (M. Logan) was ready to respond to an email with my request for help. We found the issues were mainly being caused by the format of my input dataset, and how R recognized the individual variables. The importance of this and other related issues, however, are addressed in Chapter 2 of this book.

Overall, I can highly recommend this book either for getting started in R, inspiration on statistics and/or as a practical guide to analyse your biological datasets in a robust way. Enjoy!

NIELS C. BROUWERS
State Centre of Excellence for Climate Change,
Woodland and Forest Health
School of Environmental Science, Murdoch University
Murdoch,Western Australia, Australia
Email: n.brouwers@murdoch.edu.au