Journal of Statistical Software

July 2011, Volume 43, Book Review 1. *http://www.jstatsoft.org/*

Reviewer: Yue Shentu
Merck Research Laboratory

Clinical Trial Data Analysis Using R

Din Chen, Karl E. Peace
Chapman & Hall/CRC Press, Boca Raton, FL, 2010.
ISBN 9781439840207. 387 pp. USD 89.95 (P).
http://www.crcpress.com/product/isbn/9781439840207

Clinical Trial Data Analysis Using R is the latest member in the family of "Using R" books from CRC biostatistics series. The goal of this book, as stated by the authors, is to fill the knowledge gap that exists between developed statistical methods and the applications of these methods. Overall, this book achieves the goal successfully, and does a nice job covering most, if not all major aspects of clinical trial statistics. For those who are well versed in R, this book can serve as a good reference to the established clinical biostatistics methodology; for veteran biostatisticians, this book provides a gentle introduction to the use of R in clinical trial analysis.

The level of technical detail in this book is comparable to that of Venables and Ripley (2002), which is sufficient as a general review to the subject matter, but clearly not intended to be all inclusive. Practitioners who look for in-depth treatment of specialized topics should not expect to find all the answers here, but references provided at the end of each chapter do provide a good starting point for further research. On the other hand, since readers are expected to gradually pick up R with the help of examples and other resources referenced in the book, first-time users of R will need to actively seek out more details on R commands and functionalities on their own in order to get the most out of this book.

The first chapter of the book is a short introduction to R. The authors did not attempt to write yet another tutorial to the S language in this chapter, instead they used a simulated clinical trial to demonstrate the use of R. From this simulation exercise, the readers should get a taste of three major components of data analysis in R: data manipulations used to simulate the data, graphical utilities used to explore the data, and the statistical tools used to analyze the data. This style of presentation, which illustrates important concepts and methods through examples, is adopted throughout the book and works well for its purpose. The regulatory perspective on the use of R in clinical trials is also briefly discussed in this chapter. A bit more discussion on this topic, however, seems to be wanted to motivate newcomers to R to learn a new language specifically for clinical trial analysis.

Chapter 2 gives an excellent overview of clinical trials for those who look to start their careers in clinical biostatistics. The responsibility of a clinical biostatistician includes not only the

analysis and interpretation of data, but also the design of clinical studies from statistical perspective. Understanding how statistics plays its part in the bigger picture of clinical development is important. This chapter does a great job summarizing the major phases of clinical trials, the overall clinical development plan, and major aspects of clinical trials that require statistical input. Readers who wish to learn more details about the biostatistics considerations for clinical trials may also find a book-length treatment of this topic in Peace and Chen (2011).

Starting from Chapter 3, the book presents a series of diverse topics in clinical trial statistics. Most classic subjects can be found here, including the treatment comparisons using ANOVA and ANCOVA, analysis of longitudinal and time-to-event data, sample size and power calculation, bioequivalence trials, and analysis of adverse events. There are also chapters dedicated to Bayesian analysis in clinical trials, meta-analysis, and microarray analysis. In each chapter, background information and objectives for the clinical case study are introduced first, together with detailed descriptions of data collected from clinical trials. This is followed by the discussion of statistical methodologies, which in general gives as much detail as one would expect from the first graduate course on respective topics. The third part of a chapter then presents step-by-step implementations of the statistical analyses in R. The code snippets used in the book are very nicely written, accompanied by sufficient annotations and explanations for their intended purposes. Functions from popular R extension packages are frequently used to showcase the versatility of R. In addition, simple yet effective graphics are used throughout the book to visualize data and present the results of statistical analysis.

A very important topic of clinical biostatistics that is missing from this book is, indeed, the problem of missing data. Except the censored data in survival analysis, most datasets presented in the book are free of subject attrition and missing data points. The issue of missing data is largely hidden from the readers. A general discussion of the missing data problem, and relevant R utilities for missing data handling would be a welcomed addition for future editions of the book.

Another criticism of the book is the absence of any mention of SAS. It is understandable that the authors may want to avoid the perennial debate between R and SAS. Such comparisons, however, are probably unavoidable in an industry where SAS has been so widely used. It is this reviewer's opinion that potential reader of this book will benefit more if he or she truly understands the comparative strengths of these two powerful statistical software suites, and appreciates how the flexibility and extendability of R could empower a statistician beyond the modular statistical analyses of SAS.

Finally, there does not appear to be a straightforward way to obtain the datasets used in the book. Most programs in the book point to files on the local `C:` drive. For a book that emphasizes step-by-step illustrations, this is an unfortunate oversight. Making these datasets available online or in the format of an R package will definitely benefit those readers who are eager to try out the examples in the book.

Despite these minor criticisms, the book by Chen and Peace is a great introductory book for clinical biostatistics with an emphasis on R implementations. I would highly recommend it to two audiences: those who know R but wish to learn more about clinical biostatistics, and those who are experienced in clinical biostatistics but would like to explore the wonderful world of R. The example-based approach is easy to follow and makes the book a very helpful desktop reference for many biostatistics methods.

References

Peace KE, Chen D (2011). *Clinical Trial Methodology*. Chapman & Hall/CRC Press, Boca Raton.

Venables WN, Ripley BD (2002). *Modern Applied Statistics with S*. 4th edition. Springer-Verlag, New York.

Reviewer:

Yue Shentu
Merck Research Laboratory
126 E. Lincoln Avenue, PO Box 2000, RY34-A304
Rahway, NJ, 07065, United States of America
E-mail: yue_shentu@merck.com

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume 43, Book Review 1 *Published:* 2011-07-18
July 2011