

Probability and statistics with R

María Dolores Ugarte, Ana F. Militino and Alan T Arnholt

Chapman & Hall / CRC

ISBN: 978-1-58488-891-9

On the one hand, this book offers an extensive and well structured introduction to statistics. On the other hand, in my opinion, the best aspect of this book is that it opts for R as the statistical package for the entire learning process and for performing the statistical analyses. The book is intended for a wide audience, but I think that teachers and students will especially find it a valuable resource for introductory courses on statistics. The chapters cover the content of a classic one-year course (or two semesters courses) on descriptive statistics, probability, basic inference and regression methods, giving special attention to practical issues through interesting examples that use mostly real data sets, most of them provided by the authors. Illustrated examples as well as proposed exercises are also useful pedagogical tools and reflect the authors' experience in university teaching during the last years. This book is also the evolution of a previous book in Spanish , but the current book is much more complete both from the theoretical and practical point of view so that the previous one has to be considered as the initial seed.

Firstly, a brief introduction to S is provided, mainly regarding data objects and their manipulation, but also some concepts about programming and producing graphs are introduced. Here, one important point that needs to be clarified is the difference between the S-PLUS and R languages. Even though the title of the book only refers to R, the authors use S to refer to both R and S-PLUS, which can be a bit confusing, specially for readers that have recently started to work with R, and who may not have ever heard about S or S-PLUS. The point is that most code in S is also valid in R and vice versa since S was the initial language but R is the open source version. However, nowadays R has evolved by itself and has become the "lingua franca" of computational statistics, as the authors remark and S-PLUS is a commercial evolution of S. After this preliminary introduction to R, chapters 2 to 5 offer a well structured introduction to descriptive statistics and basic probability, including univariate and multivariate probability distributions. Chapter 6 covers issues regarding sampling and sampling distributions and the next three chapters (7 to 9) provide methods for point estimation and classic inference, including confidence intervals and hypothesis testing. Non-parametric methods for performing statistical inference are presented in chapter 10 and

ANOVA models and other methods of experimental design are covered in chapter 11. Finally, chapter 12 provides an extensive introduction to linear models, with special attention to the mathematical aspects.

The authors of the book have also developed a R library, named PASWR, that contains the datasets and some functions used through the book, which makes the content really interactive and provides facilities for using this book for teaching. Scripts from the commands used in the chapter can be found at the following web page: <http://www1.appstate.edu/~arnholta/PASWR/>. A generous number of exercises are proposed at the end of each chapter, most of which have to be solved using R. A manual with the complete solutions to the problems has also been edited, which is available only for teachers.

In conclusion, I strongly recommend this book, especially for students and teachers of statistics and also many other fields, since they will find a large amount of material for learning statistics and R. From my own experience of teaching statistics, I can state that R is one of the best options for supporting practical sessions with students. Although R can be difficult to handle at the initial learning stage, after a reasonable period it can become one of the most effective platforms for performing statistical analyses. In addition to this, I think that this book can become a reference book for learning statistics with R, jointly with the books of Venables and Ripley and Dalgaard. The authors, in the preface of the book, finally thank “the geniuses of this age who first conceived of the idea of an excellent open source software for statistics and those who reared the idea to adulthood, our gratitude is immeasurable. May the lighthouse of your brilliance guide travelers on the ocean of statistics for decades for come. Thank you, R Core Team “. In this sense, I think that the authors of this book have to be thanked for their important contribution that will help statistical practitioners, students, teachers and scientists to definitely adopt R as the gold-standard package for statistics.

References

- Ugarte, M. D. and Militino, A. F. (2001). *Estadística Aplicada con S-PLUS*. Pamplona, Universidad Pública de Navarra.
- Venables, W.N. and Ripley. B. D. (2002). *Modern Applied Statistics with S*. New York, Springer.
- Dalgaard, P. (2008). *Introductory Statistics with R*. New York, Springer.

Joan Valls Marsal
joan.valls@iconcologia.net
Catalan Cancer Registry, Catalan Institute of Oncology, Barcelona
Department of Mathematics, Autonomous University of Barcelona, Barcelona