Journal of Statistical Software

Reviewer: Pedro Valero-Mora
University of Valencia

Introductory Statistics with R (Second Edition)

Introductory Statistics with R by Peter Dalgaard is the second edition of a successful book targeting the acclaimed free statistical environment R. Its author has been one of the members of the core team, which is in charge of controlling the development of R, since 1997. This qualifies him as an expert in the inner workings of the R environment and its applications for statistical computing. He also has actual experience teaching the Ph.D. curriculum with R at the Faculty of Health Sciences of the University of Copenhagen, and he does statistical consultancy within the area of medical research.

The first edition of this book had chapters on the following topics: essentials of R, basic statistics and graphics, t tests, regression and correlation, ANOVA, tables, multiple regression, linear models, logistic regression, and survival analysis. The new edition of this book comes with additional chapters covering advanced data manipulation, Poisson regression and non-linear curve fitting. Besides, an R package named **ISwR** is provided that includes a number of data examples useful for practicing the lessons in the book. The version of R used in the book is 2.6.2, which was the current at the moment of its writing, and the material in it is not linked to any operating system in particular because R runs virtually identically on Unix, Windows and MacOS. The output of R is printed as is, so it is always possible to check if things are running as expected, and if the reader is following the examples correctly. It has 364 pages, and includes indexes, a large number of figures (produced in all cases by the R system), exercises with answers, and four annexes dealing with installation instructions, a compendium of the commands, and some other things.

As mentioned by the author in the preface, the data examples and the selection of statistical techniques are clearly biased towards biomedical applications. However, as the topics included in the book cover quite easily the spectrum of what an introductory course in any discipline would generally accommodate, and, as biomedical examples have the potential of appealing to wide ranges of readers (everybody cares about health, don't they?), I think this aspect should be of minimal concern in most of the cases. Anyway, with little extra effort, teachers not entirely satisfied with this orientation may supply their students with more specific databases for extra practice once the basics that have been learnt using the book.

The book presents the information in a very straightforward manner: Each chapter starts introducing a minimum of theoretical background on the topic discussed, and then proceeds to explain the commands in R that are related to the topic. As this is mainly a computer book, and not so much a statistics book, usage of R receives considerable more attention than the concepts underlying the techniques. An interesting additional aspect is that common mistakes often carried out by non-expert users when starting to use R are carefully explained, as well as ways of avoiding them. This should save beginners a lot of unnecessary sweat on their first steps into the system.

There are several books out there with the words "Introductory" and "R" in the title but not all of them address the same audience. A classification, there can be others, takes into account whether the reader has knowledge of introductory statistics, and if his degree of expertise using computers is already at the level needed to use R. A reader interested in learning introductory statistics has different needs of a reader with this knowledge already acquired. Thus, while the second will prefer a book focused on how to take advantage of software, the first will require a complete presentation of the statistical concepts, as well as sound examples to motivating him throughout the material.

The audience of *Introductory Statistics with R* is made up of students that already have a good grasp of introductory statistics and need to learn how to apply them in practice using R. This book is not so well-suited for an introductory course of statistics because the presentation of the concepts is so brief and intended as a refresh, not as a full explanation. Besides, some of the comments are set to a level clearly higher than basic and, consequently, readers without knowledge on statistics will have problems to follow them. In summary, even though the title of this book includes the word "Introductory", its target reader is to be set at an intermediate level. Moreover, familiarity with concepts of software programming would be useful to follow this book too, as a few of the concepts managed in it (objects, expressions, types of data, etc.) can probably be hard to comprehend otherwise (do not forget that R is not a statistical package, but a statistical programming language and the level of flexibility associated with that type of software involves more complexity). In summary, graduate students of any discipline looking for learning the currently most praised free system for statistics will find this book very appealing, but undergraduate students attending to their first course in introductory statistics will not.

Reviewer:

Pedro Valero-Mora
University of Valencia
Department of Methodology of Behavioural Sciences
Valencia 46010, Spain
E-mail: valerop@uv.es
URL: http://www.uv.es/valerop/