Journal of Statistical Software

September 2016, Volume 72, Book Review 2. doi: 10.18637/jss.v072.b02

Reviewer: Abdolvahab Khademi
University of Massachusetts

Applied Univariate, Bivariate, and Multivariate Statistics

Daniel J. Denis
John Wiley & Sons, Hoboken, 2016.
ISBN 9781118632338. xxxiv+726 pp. USD 130 (H).
http://www.wiley.com/WileyCDA/WileyTitle/productCd-1118632338.html

---

Most general statistics books for social and behavioral sciences are characterized by extensive exposition of concepts for the sake of simplifying the mathematical and technical aspects of statistics. In scope, they are very similar in presenting techniques from the very basic to more intermediate. This depth and scope of materials satisfy most learning and teaching goals for students in undergraduate programs, such as psychology, education, nursing, and economics, among others. More advanced topics, such as factor analysis, principal component analysis, and structural equation modeling are usually presented in more specialized books, and usually in graduate programs or workshops. *Applied Univariate, Bivariate, and Multivariate Statistics* attempts to be a book that starts from intermediate statistics up to more specialized techniques commonly found in graduate programs. As the title of the book suggests, the author has intended an applied treatment of the subject matter but with the distinguishing aspect of being more technical even in the topics commonly covered in the majority of textbooks in behavioral sciences. The programming language R and the statistical package SPSS (script) are used to illustrate how the techniques are implemented. The author assumes that the readers already know how to use the R and SPSS statistical packages.

The book is organized in sixteen chapters, with conceptual questions and exercises along with a summary and list of highlights. An appendix briefly presents an overview of matrix algebra with R code. There are two companion websites, one of which is hosted by the publisher (which includes instructional details, data, and selected solutions) and the other is the author's personal domain that includes a continuously updated guide to the book and an errata file. In a rather detailed preface, the author sets out his intention of writing the book (balance between practice and theory), the audience (advanced undergraduates and graduate students), structure and coverage of the book, rationale for knowing about the history of a statistical technique (the book contains many historical sketches about statistical papers and techniques), conventions and symbols used, and some notes on hypothesis testing and missing data (which I think is too early to be included in the preface).

Chapter 1, *Preliminary Considerations*, serves as an introduction to statistical modeling and the empirical research world, where the author talks about epistemology, representation of

reality (in social and hard sciences), and some comments about mathematical understanding needed to achieve a working knowledge of applied statistics in social and behavioral sciences. The author reminds the reader that taking a course in statistics involves logic and a perspective of reality in addition to mathematics and computation. Chapter 2, *Mathematics and Probability Theory*, presents the essential concepts in probability, such as set theory, basics of probability, Bayes' theorem, and statistical inference in a general manner. The mathematics section reviews limits, differentiation, integration, and some matrix algebra.

Chapter 3, *Introductory Statistics*, sets the beginning of the statistical core of the book. The chapter starts with theoretical distributions (normal, binomial, and chi-square) and along the way touches on z-score and chi-square goodness-of-fit test. Other topics include scales of measurement, moments and expectations, estimation, degrees of freedom, sampling distributions, bootstrapping, likelihood ratio test, covariance, the t distribution and the family of t-tests, power, blocking, and some effect size measures.

In Chapter 4, *Analysis of Variance: Fixed Effects Models*, the author provides a thorough treatment of the one-way ANOVA model, assumptions (checking), hypothesis testing, effect size measures, contrasts and posthoc tests, and power. Chapter 5, *Factorial Analysis of Variance: Modeling Interactions*, continues the treatment of ANOVA with two or more independent variables, their interaction and main effects. *Introduction to Random Effects and Mixed Models* in Chapter 6 continues the discussion of ANOVA models with random and fixed effects, their definition and their estimation methods (maximum likelihood and restricted ML). In addition to ANOVA, the author explains variance components analysis. Finally, mixed models (with random and fixed effects) are introduced with their relation to multilevel modeling. Chapter 7, *Randomized Blocks and Repeated Measures*, discusses the last member of ANOVA family models. Both additive and non-additive models are discussed.

Linear regression is thoroughly discussed, both in theory and in application, in three chapters. Chapter 8, *Linear Regression*, begins the discussion with a brief account of Galton's study and use of regression. Linear model, least squares (parameter) estimation, model assumptions, hypothesis testing, matrix formulation of regression, ANOVA table in regression, model fit criteria, model diagnostics, and power analysis are presented. In Chapter 9, *Multiple Linear Regression* (MR), the author introduces multiple regression starting out with the concept of partial and semipartial correlation (which I think unnecessarily overcomplicates the introduction of MR), MR in matrix form, polynomial regression, collinearity, variable selection with fit criteria (e.g. AIC and BIC), and mediation. Interaction in MR predictors is treated separately in Chapter 10, *Interactions in Multiple Linear Regression: Dichotomous, Polytomous, and Continuous Moderators*. The author explains interaction in MR drawing an analogy to factorial ANOVA. Both continuous and categorical variables are considered in the discussion (presented as mediator, hinting at the audience with psychology and sociology backgrounds). A brief note on ANCOVA is also presented in this chapter. Logistic regression and generalized linear models (GLM) are discussed in Chapter 11, *Logistic Regression and the Generalized Linear Model*. Because regression has received a very good amount of attention in the past three chapters of the book, the author appropriately allocates much detail about the logic of logistic regression and the GLM family of methods. Different link functions are presented briefly. Other topics in this chapter include the logit, interpretation of a logistic regression analysis, predictions, and sample size considerations.

Chapter 12, *Multivariate Analysis of Variance* (MANOVA), discusses the important topic of analyzing high dimensional data. The author begins with a caveat that MANOVA is not just

about combining the dependent variables, and then presents an example where composite variables are appropriate. The rest of the chapter is concerned about the theory and formulation of MANOVA, assumptions (testing and alternative tests when assumptions are not met), multivariate contrasts, univariate analysis in the presence of MANOVA, and sample size considerations.

Chapter 13, *Discriminant Analysis*, presents linear discriminant analysis for classification of responses into two or more categories. Analogy is made to logistic regression (for classification) and MANOVA (for combination of predictors). Canonical correlation is also discussed in detail at the end of the chapter.

Data reduction methods are presented in the later chapters of the book. *Principal Component Analysis* is presented in Chapter 14, with a brief history and theory. Adequate presentation is also provided on the pivotal concept of eigenvalues. Chapter 15, *Factor Analysis*, discusses another common data reduction method in psychology and social sciences. A history of the common-factor model is given, together with distinction between exploratory and confirmatory factor analysis methods. Scale indeterminacy and non-unique solutions, estimation methods, factor rotation, and implementation of factor analysis in R and SPSS are presented. The author chooses to present cluster analysis at the end of the chapter (though, a separate chapter would suit better to distinguish it from latent variable modeling). *Path Analysis and Structural Equation Modeling* (SEM) are presented in Chapter 16, the last chapter of the book. In this chapter, the author presents a history of path analysis and then proceeds to the concept of causal modeling. The SEM part starts with a discussion of CFA (the measurement component) and then proceeds to the structural part, with discussions of model fit criteria, model identification, and implementation in R.

All the foregoing chapters and contents are characterized by several points that make this book stand out in the well-represented behavioral statistics realm. The detailed exposition of the topics in every chapter and the majority of the exercises support the author's intention as a readable theoretical book on applied statistics. Most chapters are rich with theoretical and logical explanations of familiar topics in statistics. Readers with experience in intermediate statistics may find a wealth of advanced information about familiar methods and new information about advanced statistics. Therefore, even those who have already applied most methods in the book may still find insights about the logic and theory of those methods, making the book very useful for an advanced audience. Therefore, the book will be very helpful for advanced students and researchers who intend to dig deeper into the theory, logic, and (to some extent) the math beneath the methods they are already familiar with.

Another outstanding feature of the book is the wealth of references it provides for those advanced readers who may need more specialized resources. The references are excellent in terms of the selection of authoritative or original scholars almost in every chapter topic. In addition, providing historical sketches about a method may appeal to those who like to be well-grounded in a quantitative field. Nevertheless, I found the in-text historical sketches (some running for several pages) and references a bit distracting. For future editions, I would recommend to place the historical notes in side bars or demarcated out of the main passage, and to move the references to the end of the chapter.

Another great feature of the book is the extensive set of conceptual questions and exercises. If worked out thoroughly, these questions can firmly instill the theory and concepts of several statistical methods commonly used in the behavioral sciences. On the other hand, they may

seem inadequate for the readers who plan to get hands-on experience with the methods and software because there are too few exercises involving data analysis.

As for statistical software, the book does not bear a distinct emphasis on and dedication to computing. Rather, the use of software, mainly R and SPSS scripts, is for the sake of implementing the methods. This is in line with the author's intention for writing a book notable for its emphasis on theory and logic. Throughout the book and where relevant, examples are given and solved using R code and SPSS script. Knowledge of these software packages is taken for granted and there is no introduction or guidelines for installation or running the commands. It might have served some readers if an appendix or a few introductory pages were dedicated to the basics of the software used in the book.

Overall, the book would be valuable for graduate students in social and behavioral sciences and researchers who need to know the more advanced topics. As stated above, because the book emphasizes theory over practice, those coming for hands-on experience and data analysis opportunities with software may find the book lacking in this area. But for those with practical background in statistical methods, especially students from behavioral sciences, this book makes a very valuable theoretical resource that adequately fills the gap for an in-depth treatment of advanced statistics for behavioral sciences.

Reviewer:

Abdolvahab Khademi
University of Massachusetts
Department of Mathematics and Statistics &
College of Education
Amherst MA 01002, United States of America
E-mail: khademi@math.umass.edu

Journal of Statistical Software http://www.jstatsoft.org/
published by the Foundation for Open Access Statistics http://www.foastat.org/
September 2016, Volume 72, Book Review 2 *Published:* 2016-09-02
doi:10.18637/jss.v072.b02