



Journal of Statistical Software

June 2004, Volume 11, Book Review 2/3.

<http://www.jstatsoft.org/>

Reviewer: Vanessa C. Beddo and Frauke Kreuter
University of California at Los Angeles

A Handbook of Statistical Analyses using SPSS

Sabine Landau and Brian Everitt
Chapman & Hall/CRC, Boca Raton, Florida, 2004.
ISBN 1-58488-369-3. vii + 354 pp. \$44.95 (P).

A Handbook of Statistical Analyses using Stata (Third Edition)

Sophia Rabe-Hesketh and Brian Everitt
Chapman & Hall/CRC, Boca Raton, Florida, 2004.
ISBN 1-58488-404-5. xiii + 308 pp. \$49.95 (P).

A Handbook of Statistical Analyses using SPSS by Landau & Everitt (2004) and *A Handbook of Statistical Analyses using Stata, Third Edition* by Rabe-Hesketh & Everitt (2004) join *A Handbook of Statistical Analyses using S-PLUS* and *SAS*. The handbooks, like their predecessors, are collections of applications of statistical techniques for their corresponding software packages. The handbooks are situation-specific learning tools with clearly explained step-by-step examples that can be followed by the reader using an overlapping set of data provided on the web.¹

Most chapters of the two handbooks reviewed here, revolve around a data analysis technique, span about 15 to 20 pages in length, and are easy to reference in most sections. This is what a reader would hope for when picking up a handbook. The chapters start with a research question, followed by a brief (usually between 1 and 5 page) introduction to the statistical technique used to answer the research question. However, each chapter's primarily goal is to present a worked example that shows how the analysis is carried out in SPSS or Stata. A short interpretation of analysis results and a battery of exercises follow the initial example. Solutions to the exercises can be found in the back of the book or on the web.

A Handbook of Statistical Analyses using SPSS begins with data entry fundamentals, explained clearly and concisely, followed by a brief explanation of elementary SPSS data management procedures, including variable label and missing value assignments. For the most part, this book assumes that data is ideally formatted for each analysis situation.

¹As of June 2004, <http://www.iop.kcl.ac.uk/iop/departments/BioComp/SPSSBook.shtml> and <http://www.stata.com/texts/stas3>

Simple statistical analyses appear in early chapters as descriptive and inferential techniques, usually involving univariate outcome variables, both continuous and categorical. Several examples are discussed in the context of using the SPSS user interface, a point-and-click GUI. The chapters are essentially a map of these menus, providing common analysis examples for the novice user. The statistical analysis methods within these sections are buried within the study discussions, which makes referencing analytic techniques somewhat time-consuming.

Subsequent chapters utilize various regression analysis techniques, such as multiple regression and ANOVA. A chapter dedicated to multiple regression analysis immediately follows the chapters describing simple analysis techniques for continuous and categorical data. The ANOVA examples span over three chapters; one-way ANOVA, factorial design ANOVA, and ANOVA for repeated measures analysis. ANCOVA is briefly discussed in the section on factorial design, while MANOVA is illustrated in the one-way ANOVA section. These chapters are clearly organized by analysis technique, with the exception of MANOVA and ANCOVA. This allows the text to genuinely resemble a handbook, since the sections are easier to reference.

Advanced analysis techniques, such as mixed effects models (in the second of two chapters dedicated to repeated measures), logistic regression, survival analysis (using Cox's regression) and various multivariate analysis techniques (such as factor, principal component, cluster and discriminant function analysis) are treated in a similar fashion. Each discussion of these analysis techniques is clearly described and demonstrated, again in the spirit of acting as a statistical handbook. Within a study on mixed effects models, the authors elaborates effectively on transposing a data matrix (changing variables to cases). This aspect of data management, as with the rest of the text, focuses on the GUI which makes the process painless for the user.

Although *A Handbook of Statistical Analysis using SPSS* initially appears to have two different organizational constructs, a reader should be able to navigate the text in search of a desired technique, so long as the procedure for which a reader searches has been used in one of the many example analysis sections. Once located, the examples will aid a user to effectively navigate the SPSS GUI in order to conduct the statistical analysis required.²

A Handbook of Statistical Analyses using Stata is in its 3rd edition. The text has been updated to Stata 8, enriched with three new chapters, includes new graphical features of Stata 8, and emphasizes diagnostics.

The handbook begins with a brief but broad introduction to Stata, from the use of basic syntax and data management to more advanced topics, such as looping through variables and observations, reshaping data, matrix commands, programming, and graphics. This introduction provides the reader with the necessary tools to work through the rest of the book. In addition throughout the following chapters various data management techniques and programming tricks are presented in the context of applications of the statistical analyses. This additional exposure to complex data management examples is a special feature of this book. The detailed index on techniques and Stata commands will aid the reader in finding them throughout the book.

Data description and simple inference form the earlier chapters in the series of data and techniques examples, followed by multiple regression, analysis of variance, logistic regression

²Editorial note: The header for Chapter 3 is "Data Description and Simple Inference for Continuous Data." It should read "Simple Inference for Categorical data."

and generalized linear models. After describing the research question and the statistical technique used to address it, each chapter starts with reading ASCII raw files, examining the data with the help of the new graphic features in *Stata* 8, followed by the analysis and diagnostic examples.

The discussion of longitudinal data analysis is expanded in the 3rd edition of this handbook. New is the chapter on random effect models and generalized estimation equations, along with a discussion of estimating the parameters of mixed models using **gllamm**. (Sophia Rabe-Hesketh is the principal developer of the *Stata* user written module **gllamm**.)

Known from former editions are the chapters illustrating different epidemiological designs, survival analysis, mixture models (in the chapter on maximum likelihood estimation), and principle component analysis. The last chapter on cluster analysis is also new in the 3rd edition introducing the *Stata* 8 extensions for cluster analyses.

A Handbook of Statistical Analysis using Stata will be especially useful for those readers who have some experience with the *Stata* language and basic programming and who appreciate examples as a means to work their way through statistical techniques. While the *Stata* GUI has been around since release 8.0, commands are used throughout the book, reinforcing the importance of replication. The advice given in the introductory chapter on how to build a file containing the commands necessary to carry out a particular data analysis, should be taken seriously. Not only because this is useful for a smooth replication of the analyses examples but as good practice for data analysis in general.

Though brief, this book contains a lot of useful and practical information on how to carry out statistical analyses in *Stata* 8. Readers should be able to follow the mix of challenging programming tools and advanced data analysis techniques after working through the introductory chapters.

Generally speaking the audience for both handbooks is quite similar, since they assume previous knowledge of statistics such as linear regression, correlation, significance tests and simple analysis of variance. On the other hand, the books differ in the amount of previous programming experience they expect from a reader. The *SPSS* book uses the GUI as the sole learning tool, and therefore neither demands programming knowledge from its reader nor introduces *SPSS* syntax statements, which make for reusable code. In contrast, the *Stata* handbook, with its more intensive and instructive concentration on code production, assumes a reader who is comfortable conducting statistical analysis via the command line.

Both handbooks consist of a healthy volume of actual research examples, many of them coming from biostatistics, public health, and epidemiology. This selection of interesting examples, such as the treatment of post-natal depression, schizophrenia or drug and alcohol dependencies, are well suited to motivate the reader to learn the software capabilities for various statistical analysis of intermediate and advanced level. However, given the wide range of topics covered in a rather few number of pages, understandably neither book provides a comprehensive introduction to either the statistical procedure or the statistical package. Nevertheless there are excellent hints, literature references, and software examples in either handbook to move forward on your own.

From a teacher's perspective, neither one of the books would be an obvious choice for the sole textbook for a particular class. However, both books have the potential to be used in lab sessions for graduate level courses. In a data analysis course, it would be recommended

that instructors elaborate further in places where the authors were understandably brief in their descriptions of statistical techniques. In a course whose focus is to teach either of these programming languages, it is recommended that an instructor supplement lessons with syntax statements (Landau/Everitt) or provides further explanations of the code in more detail, when necessary (Rabe-Hesketh/Everitt).

Reviewer:

Vanessa C. Beddo and Frauke Kreuter
University of California at Los Angeles
Department of Statistics
8130 Math Science Building
Los Angeles, CA 90095-1554, USA
E-mail: beddo@stat.ucla.edu, fkreuter@stat.ucla.edu