

Preface to the 3rd Edition

The third edition of this book on Applied Multivariate Statistical Analysis offers the following new features.

1. A new Chapter 8 on Regression Models has been added.
2. Almost all numerical examples have been reproduced in MATLAB or R.

The chapter on regression models focuses on a core business of multivariate statistical analysis. This contribution has not been subject of a prominent discussion in earlier editions of this book. We now take the opportunity to cover classical themes of ANOVA and ANCOVA analysis. Categorical responses are presented Section 8.2. The spectrum of log linear models for contingency tables is presented in Section 8.2.2 and applications to count data e.g. in the economic and medical science is presented there. Logit models are discussed in great detail and the numerical implementation in terms of matrix manipulations are presented.

The majority of pictures and numerical examples has been now calculated in the (almost) standard language R & MATLAB. The code for each picture is indicated with a small  sign near the picture, e.g.  MVAdenbank denotes the corresponding quantlet for reproduction of Figure 1.9, where we display the densities of the diagonal of genuine and counterfeit bank notes. We believe that these publicly available Quantlets (see also www.quantlet.com) create a valuable contribution to distribution of knowledge in the statistical science. The symbols and notations have also been standardized. In the preparation of the 3rd edition, we received valuable input from Song Song, Weining Wang and Mengmeng Guo. We would like to thank them.

Berlin
Louvain la Neuve
June 2011

Wolfgang Karl Härdle
Léopold Simar

Preface to the 2nd Edition

The second edition of this book widens the scope of the methods and applications of Applied Multivariate Statistical Analysis. We have introduced more up to date data sets in our examples. These give the text a higher degree of timeliness and add an even more applied flavour. Since multivariate statistical methods are heavily used in quantitative finance and risk management we have put more weight on the presentation of distributions and their densities.

We discuss in detail different families of heavy tailed distributions (Laplace, Generalized Hyperbolic). We also devoted a section on copulae, a new concept of dependency used in the financial risk management and credit scoring. In the chapter on computer intensive methods we have added support vector machines, a new classification technique from statistical learning theory. We apply this method to bankruptcy and rating analysis of firms. The very important CART (Classification and Regression Tree) technique is also now inserted into this chapter. We give an application to rating of companies.

The probably most important step towards readability and user friendliness of this book is that we have translated all Quantlets into the R and Matlab language. The algorithms can be downloaded from the authors' web sites. In the preparation of this 2nd edition, we received helpful output from Anton Andriyashin, Ying Chen, Song Song and Uwe Ziegenhagen. We would like to thank them.

Berlin
Louvain la Neuve
June 2007

Wolfgang Karl Härdle
Léopold Simar

Preface to the 1st Edition

Most of the observable phenomena in the empirical sciences are of a multivariate nature. In financial studies, assets in stock markets are observed simultaneously and their joint development is analyzed to better understand general tendencies and to track indices. In medicine recorded observations of subjects in different locations are the basis of reliable diagnoses and medication. In quantitative marketing consumer preferences are collected in order to construct models of consumer behavior. The underlying theoretical structure of these and many other quantitative studies of applied sciences is multivariate. This book on Applied Multivariate Statistical Analysis presents the tools and concepts of multivariate data analysis with a strong focus on applications.

The aim of the book is to present multivariate data analysis in a way that is understandable for non-mathematicians and practitioners who are confronted by statistical data analysis. This is achieved by focusing on the practical relevance and through the e-book character of this text. All practical examples may be recalculated and modified by the reader using a standard web browser and without reference or application of any specific software.

The book is divided into three main parts. The first part is devoted to graphical techniques describing the distributions of the variables involved. The second part deals with multivariate random variables and presents from a theoretical point of view distributions, estimators and tests for various practical situations. The last part is on multivariate techniques and introduces the reader to the wide selection of tools available for multivariate data analysis. All data sets are given in the appendix and are downloadable from www.md-stat.com. The text contains a wide variety of exercises the solutions of which are given in a separate textbook. In addition a full set of transparencies on www.md-stat.com is provided making it easier for an instructor to present the materials in this book. All transparencies contain hyper links to the statistical web service so that students and instructors alike may recompute all examples via a standard web browser.

The first section on descriptive techniques is on the construction of the boxplot. Here the standard data sets on genuine and counterfeit bank notes and on the Boston housing data are introduced. Flurry faces are shown in Section 1.5, followed by the

presentation of Andrews curves and parallel coordinate plots. Histograms, kernel densities and scatterplots complete the first part of the book. The reader is introduced to the concept of skewness and correlation from a graphical point of view.

At the beginning of the second part of the book the reader goes on a short excursion into matrix algebra. Covariances, correlation and the linear model are introduced. This section is followed by the presentation of the ANOVA technique and its application to the multiple linear model. In Chapter 4 the multivariate distributions are introduced and thereafter specialized to the multinormal. The theory of estimation and testing ends the discussion on multivariate random variables.

The third and last part of this book starts with a geometric decomposition of data matrices. It is influenced by the French school of *analyse de données*. This geometric point of view is linked to principal components analysis in Chapter 10. An important discussion on factor analysis follows with a variety of examples from psychology and economics. The section on cluster analysis deals with the various cluster techniques and leads naturally to the problem of discrimination analysis. The next chapter deals with the detection of correspondence between factors. The joint structure of data sets is presented in the chapter on canonical correlation analysis and a practical study on prices and safety features of automobiles is given. Next the important topic of multidimensional scaling is introduced, followed by the tool of conjoint measurement analysis. The conjoint measurement analysis is often used in psychology and marketing in order to measure preference orderings for certain goods. The applications in finance (Chapter 18) are numerous. We present here the CAPM model and discuss efficient portfolio allocations. The book closes with a presentation on highly interactive, computationally intensive techniques.

This book is designed for the advanced bachelor and first year graduate student as well as for the inexperienced data analyst who would like a tour of the various statistical tools in a multivariate data analysis workshop. The experienced reader with a bright knowledge of algebra will certainly skip some sections of the multivariate random variables part but will hopefully enjoy the various mathematical roots of the multivariate techniques. A graduate student might think that the first part on description techniques is well known to him from his training in introductory statistics. The mathematical and the applied parts of the book (II, III) will certainly introduce him into the rich realm of multivariate statistical data analysis modules.

The inexperienced computer user of this e-book is slowly introduced to an interdisciplinary way of statistical thinking and will certainly enjoy the various practical examples. This e-book is designed as an interactive document with various links to other features. The complete e-book may be downloaded from www.xplore-stat.de using the license key given on the last page of this book. Our e-book design offers a complete PDF and HTML file with links to MD*Tech computing servers.

The reader of this book may therefore use all the presented methods and data via the local XploRe Quantlet Server (XQS) without downloading or buying additional software. Such XQ Servers may also be installed in a department or addressed freely on the web (see www.i-xplore.de for more information).

A book of this kind would not have been possible without the help of many friends, colleagues and students. For the technical production of the e-book we

would like to thank Jörg Feuerhake, Zdeněk Hlávka, Torsten Kleinow, Sigbert Klinke, Heiko Lehmann, Marlene Müller. The book has been carefully read by Christian Hafner, Mia Huber, Stefan Sperlich, Axel Werwatz. We would also like to thank Pavel Čížek, Isabelle De Macq, Holger Gerhardt, Alena Myšíčková and Manh Cuong Vu for the solutions to various statistical problems and exercises. We thank Clemens Heine from Springer Verlag for continuous support and valuable suggestions on the style of writing and on the contents covered.

Berlin
Louvain-la-Neuve
August 2003

W. Härdle
L. Simar

In this new edition we have responded, also, to the reality that previous editions have been enthusiastically embraced by the community of practicing circuit designers, even though *The Art of Electronics* (now 35 years in print) originated as a course textbook. So we've continued the "how we do it" approach to circuit design; and we've expanded the depth of treatment, while (we hope) retaining the easy access and explanation of basics. At the same time we have split off some of the specifically course-related teaching and lab material into a separate *Learning the Art of Electronics* volume, as shown in the PREFACE TO THE 3rd EDITION. The 3rd edition of this Handbook is augmented with the most important results of investigations carried out in recent years. Some of the sections in the book have been refined and changed. The Handbook has been composed on the basis of processing, systematization, and classification of the results of a great number of investigations published at different times. The essential part of the book is the outcome of investigations carried out by the author. In giving the Third Edition of this work to the public, I have little else to do than to express my acknowledgments to those to whom I am under obligations, for enabling me thus far to bring it to a successful issue. To Mr. Murray, of Albemarle Street, London; Mr. Vaux, of the British Museum; and Messrs. Black and Messrs. Chambers, Edinburgh, I am specially indebted for permission to copy woodcuts belonging to them.