

Short Book Reviews

Editor: Simo Puntanen

Festschrift in Honor of Jorma Rissanen on the Occasion of his 75th Birthday

Peter Grünwald, Petri Myllymäki, Ioan Tabus, Marcelo Weinberger, Bin Yu (Editors)
Tampere International Center for Signal Processing, 2008, viii + 320 pages, softcover
ISBN: 978-952-15-1962-8

Table of contents

I. Introduction

1. Pentti Huuhtanen, Erkki P. Liski, and Simo Puntanen: A conversation with Jorma Rissanen

II. Research Articles

2. Andrew R. Barron, Cong Huang, Jonathan Q. Li, and Xi Luo: The MDL principle, penalized likelihoods, and statistical risk
3. Guilherme V. Rocha and Bin Yu: Greedy and relaxed approximations to model selection: a simulation study
4. Vijau Balasubramaniana, Klaus Larjo, and Ravi Sheth: Experimental design and model selection: the example of exoplanet detection
5. Álvaro Martín, Gadiel Seroussi, and Marcelo Weinberger: Enumerative coding for tree sources
6. Antonio Galves and Eva Löcherbach: Stochastic chains with memory of variable length
7. Ioannis Kontoyiannis: Some information-theoretic computations related to the distribution of prime numbers
8. Erkki P. Liski and Antti Liski: MDL model averaging for linear regression
9. Jerzy Konorski and Wojciech Szpankowski: What is information?
10. Ioan Tabus and Gergely Korodi: Genome compression using normalized maximum likelihood models for constrained Markov sources
11. Petri Myllymäki, Teemu Roos, Tom Silander, Petri Kontkanen, and Henry Tirri: Factorized NML models
12. Janne Ojanen, Jukka Heikkonen, and Kimmo Kaski: Towards the multicomponent MDL denoising
13. Søren Forchhammer: Context adaptive coding of bi-level images
14. Ciprian Doru Giurcaneanu: Estimation of sinusoida regression models by stochastic complexity
15. K. Vela Velupillai: A stochastic complexity perspective of induction in economics and inference in dynamics
16. Boris Ryabko: Compression-based methods for nonparametric on-line prediction, regression, classification and density estimation of time series

III. Personal Notes

17. Peter Grünwald: That simple device already used by Gauss
18. Paul Vitányi: A great mind
19. Terry Speed: My encounters with MDL
20. Tjalling Tjalkens and Frans Willems: Jorma's unintentional contributions to the source coding research in Eindhoven

Readership: Statisticians, both theoretical and applied, and information theorists, primarily those interested in the Minimum Description Length (MDL) principle.

This volume celebrates the 75th birthday of the outstanding information theorist and statistician Jorma Rissanen. It contains 20 articles, the majority providing new contributions to the theory and practice of the Minimum Description Length (MDL) principle, which was introduced by Rissanen and developed mainly by him to an attractive general approach to statistical inference.

Item 1, “A conversation with Jorma Rissanen”, presents his biography, main scientific achievements, research philosophy, and also some personal details as his passion for soccer.

Of the next 15 research articles, items 2 and 16 address theoretical issues connected with MDL, and so do also 5 and 6, specifically relative to tree models introduced by Rissanen. Items 3, 4, 8, 10, 11, 12, 13, 14 are more of applied nature, featuring simulations and/or real-world problems, though they also address the theoretical background. Item 7 points out a remarkable connection of information theory to number theory, and item 9 addresses a philosophical problem, offering more than just philosophy. Item 15 is mainly philosophical.

Finally, there are 4 “Personal Notes”, some hinting to their authors’ scientific benefits from contacts with Jorma Rissanen. Item 19 is an exception, its author, a leading statistician, says that he appreciates MDL but in his daily work only occasionally encounters problems where he can use it. This reminder that MDL, while highly commendable under the right conditions, is no panacea adds to the merits of the volume.

Imre Csiszár: *csiszar@renyi.hu*
 Alfred Renyi Institute of Mathematics
 Hungarian Academy of Sciences
 POB 127, H-1364 Budapest, Hungary

Interactive and Dynamic Graphics for Data Analysis – With R and GGobi

Dianne Cook, Deborah F. Swayne

Springer, 2007, xviii + 190 pages, US\$ 59.95 / £ 36.00 / € 46.95, softcover

ISBN: 978-0-387-71761-6

Table of contents

- | | |
|------------------------------|-------------------------|
| 1. Introduction | 5. Cluster analysis |
| 2. The toolbox | 6. Miscellaneous topics |
| 3. Missing values | 7. Datasets |
| 4. Supervised classification | |

Readership: Later year undergraduates, beginning graduate students, and researchers and graduate students in any discipline needing to explore and analyze multivariate data.

This is one of a number of recent books that are witness to the widespread (and increasing) popularity of the R system for data analysis and graphics. Much of R’s power comes from its many packages. This includes interfaces, implemented as R packages, that give access to systems and software that are independent of R.

In this book, R commands are used to exploit the power of the GGobi system for interactive and dynamic graphics. An alternative, not described in this book, is to use GGobi’s own graphical user interface. As with R, GGobi is free and can be freely downloaded from the internet.

Chapter 1 starts with a brief history of data visualization. An example follows that illustrates the use of data visualization in practical data analysis. The final two sections discuss process, commenting on analysis approaches and offering caveats. Later chapters follow the same style of example-based exposition.

Chapter 2 describes “tools” that are available in R or R packages, in GGobi, and (rather briefly) from other sources. It discusses plot types that range all the way from static plots through to interactive dynamics plots and “tours”, and enhancements that include brushing, linking and scaling (zoom in or out).

Chapter 3 uses graphical methods to diagnose the distribution and seriousness of missing values, and to assess the results of imputation.

Chapter 4 describes supervised classification, known also as discriminant analysis. Examples, with extensive use of graphs, give insight into the data and into results. Methods that are described are: linear discriminant analysis, trees, random forests, neural networks and Support Vector Machines. The discussion of random forests does not explain how the two-dimensional representation in Figure 4.11 was obtained. This will leave some readers puzzled.

Cluster analysis methods described in Chapter 5 include hierarchical, projection pursuit, model-based methods, and self-organizing maps. Again, graphical presentations are used to good effect. Two distance measures are explicitly mentioned – Euclidean distance with equal variable weights, and a correlation based measure of difference in profile. The effect and common importance of transformation is not discussed.

Chapter 6 takes up on several further topics – inference, network data and multi-dimensional scaling.

For detail, including discussion of some of the traps (e.g., selection effects with very high-dimensional data), readers will need to look elsewhere. This book is however a very useful brief overview of the insight that a powerful modern suite of graphics tools may offer.

Data sets, and other supplementary materials, are available from the web site <http://www.ggobi.org/book/>.

John H. Maindonald: john.maindonald@anu.edu.au
 Centre for Mathematics & Its Applications
 Australian National University
 Canberra ACT 0200, Australia

The Frailty Model

Luc Duchateau, Paul Janssen

Springer, 2008, xviii + 318 pages, € 64.95 / £ 50.00 / US\$ 79.95, hardcover

ISBN: 978-0-387-72834-6

Table of contents

- | | |
|-----------------------------------------------------------------|----------------------------------------|
| 1. Introduction | 4. Frailty distributions |
| 2. Parametric proportional hazards models
with gamma frailty | 5. The semiparametric frailty model |
| 3. Alternatives for the frailty model | 6. Multi-frailty and multilevel models |
| | 7. Extensions of the frailty model |

Readership: Students in statistics and mathematics; researchers and students in survival analysis.

The book by Duchateau and Jansen is generally easy to follow. The book starts with introduction to the most popular parametric and semiparametric survival models. It covers several approaches to model clustered survival data. In addition to frailty models assuming different frailty distributions, also e.g. fixed effects models with their limitations are covered. Methods are compared using both real and simulated data sets. Simple maximum likelihood methods are generally not sufficient in analyzing frailty models, thus the EM algorithm, the penalized partial likelihood approach, and Bayesian methods are applied in several accessible examples, which contain also intermediate phases of derivations. Therefore this book can be recommended

also for undergraduate students in statistics. On the other hand, the book contains several further extensions of frailty models such as multifrility and multilevel models with references. Therefore it is valuable also for researchers in survival analysis.

Tommi Härkänen: tommi.harkanen@ktl.fi
National Public Health Institute
Mannerheimintie 166
FI-00300 Helsinki, Finland

A Matrix Handbook for Statisticians

George A. F. Seber

Wiley, 2008, xix + 559 pages, € 78.00 / £ 61.95, hardcover

ISBN: 978-0-471-74869-4

Table of contents

- | | |
|-----------------------------------------------------------------------|---------------------------------------------------------|
| 1. Notation | 12. Inequalities |
| 2. Vectors, vector spaces, and convexity | 13. Linear equations |
| 3. Rank | 14. Partitioned matrices |
| 4. Matrix functions: Inverse, transpose, trace, determinant, and norm | 15. Patterned matrices |
| 5. Complex, Hermitian, and related matrices | 16. Factorization of matrices |
| 6. Eigenvalues, eigenvectors, and singular values | 17. Differentiation and finite differences |
| 7. Generalized inverses | 18. Jacobians |
| 8. Some special matrices | 19. Matrix limits, sequences and series |
| 9. Non-negative vectors and matrices | 20. Random vectors |
| 10. Positive definite and non-negative definite matrices | 21. Random matrices |
| 11. Special products and operators | 22. Inequalities for probabilities and random variables |
| | 23. Majorization |
| | 24. Optimization and matrix approximation |

Readership: Researchers interested in matrix calculus, postgraduate students and researchers of statistics.

This matrix handbook for statisticians is organized by topic rather than mathematical developments. Each chapter has a uniform structure where a definition is followed by a list of results and a short list of references to related topics in the book. The book does not have proofs, but it indicates where proofs can be found. It allows the author to put an enormous number of mathematical results into his 500 pages. There are no numerical examples, no exercises, but far more formulas and many specialized chapters. The book is not intended as a textbook for a graduate course. It contains many applications and specializations inspired by statistics, but the largest number of pages discuss general matrix calculus. There are already quite a few books on matrix algebra for statisticians, but this book maintains its uniqueness among the competition through its extensive referencing to proofs and comprehensive coverage of topics not found in any other one book.

Erkki P. Liski: erkki.liski@uta.fi
Department of Mathematics and Statistics
FI-33014 University of Tampere, Finland

Lattice: Multivariate Data Visualization with R

Deepayan Sarkar

Springer, 2008, xviii + 268 pages, € 44.95 / £ 34.50 / US\$ 54.95, softcover

ISBN: 978-0-387-75968-5

Table of contents

- | | |
|--------------------------------------------|------------------------------------------|
| 1. Introduction | 8. Plot coordinates and axis annotation |
| Part I Basics | 9. Labels and legends |
| 2. A technical overview of lattice | 10. Data manipulation and related topics |
| 3. Visualizing univariate distributions | 11. Manipulating the “trellis” object |
| 4. Displaying multiway tables | 12. Interacting with Trellis displays |
| 5. Scatter plots and extensions | Part III Extending Trellis Displays |
| 6. Trivariate displays | 13. Advanced panel functions |
| Part II Finer Control | 14. New Trellis displays |
| 7. Graphical parameters and other settings | |

Readership: For R users at any level interested in producing high quality graphics. And especially for users of the R package lattice.

D. Sarkar is a member of the R core team and author of the award winning R package lattice. The package implements the Trellis graphics system in R. Although the package comes with full documentation getting started using lattice or grasping all possibilities it offers seems a bit difficult. This book can be seen as a valuable source for lattice users at all levels. The book consists of three parts where the first part is mainly for new users. Nevertheless also more experienced users will find probably still some interesting features. The second part describes in detail how to fine tune lattice graphics and how to interact with them. The last part is for advanced users and describes how to extend the existing possibilities of the lattice package by for example writing own panel functions or creating new displays. The book nicely shows that making good graphics is a process and the reader is guided by the author in a wealth of examples through the various steps needed to reach the final result. A nice feature of the book is that it has an accompanying homepage that contains all the R code and figures of the book. The book can be read from the beginning to the end or just consulted as a reference when trying to make a certain graphic.

Klaus Nordhausen: klaus.nordhausen@uta.fi
 Tampere School of Public Health
 FI-33014 University of Tampere, Finland

Applied Multiway Data Analysis

Pieter M. Kroonenberg

Wiley, 2008, xxii + 579 pages, £ 60.95 / € 76.70 / US\$ 115.00, hardcover

ISBN: 978-0-470-16497-6

Table of contents

Part I. Data, Models, and Algorithms	11. Graphical displays for components
1. Overture	12. Residuals, outliers, and robustness
2. Overview	Part III. Multiway Data and Their Analysis
3. Three-way and multiway data	13. Modeling multiway profile data
4. Component models for fully-crossed designs	14. Modeling multiway rating scale data
5. Algorithms for multiway models	15. Exploratory multivariate longitudinal analysis
Part II. Data Handling, Model Selection, and Interpretation	16. Three-mode clustering
6. Preprocessing	17. Multiway contingency tables
7. Missing data in multiway analysis	18. Three-way binary data
8. Model and dimensionality selection	19. From three-way to four-way data and beyond
9. Interpreting component models	Appendix A: Standard notation for multiway analysis
10. Improving interpretation through rotations	Appendix B: Biplots and their interpretation

Readership: Beginning graduate students who have a basic understanding of linear algebra, statistics, and multivariate methods (such as factor analysis, principal components, and correspondence analysis). Also professionals working with multiway data in universities or elsewhere will find the book a valuable source of reference.

This book is all about counting “one, two, multi” – instead of “one, two, three, four, . . .”. This is stressed both in the beginning and in the end of the book. The arguments given in between are very convincing.

There are lots of situations in several fields of study, where the data appear to be more than two-way. With a large and a rich collection of examples from a range of fields the book demonstrates that the usual two-mode methods are not sufficient, because they do not respect the multiway design of the data. It seems that it is not essential whether the “multi” means three or four (or more), as the serious step is from two to three-way.

The book comprises three parts. Part I gives a nice motivation and orientation by introducing the key concepts of the multiway world, such as the Tucker and Parafac (parallel factor) models and their algorithms.

Part II goes through the phases needed to cope with multiway data: preprocessing, handling missing data, thinking about the implications of the dimensionality in model selection, as well as interpreting the results, plotting, rotating, and analysing residuals and outliers. Most of these correspond to the usual stages of data analysis, but the multiway nature of the data brings new views and challenges to each one of them.

Part III is extremely interesting, as it shows how the multiway models relate, for example, to structural equation models (SEM), repeated measures ANOVA, latent growth curves, and multivariate autoregressive processes. Many fascinating applications are worked through, and they are easy to follow, as they are structured in five stages: 1) objectives, 2) data description and design, 3) model and dimensionality selection, 4) results and their interpretation, and 5) validation. The majority of these applications come from the social and behavioral sciences, but data sets from environmental studies, growth studies, chemistry and agriculture are analyzed as well.

The book is written in clear style and mostly in conceptual rather than mathematical level. It emphasizes the author's over thirty years' personal experience and practical side of performing multiway data analyses. It is easy to recommend this book, as it really opens new views of the world.

Kimmo Vehkalahti: *kimmo.vehkalahti@helsinki.fi*
 Department of Mathematics and Statistics
 FI-00014 University of Helsinki, Finland

Handbook of Data Visualization

Chun-houh Chen, Wolfgang Härdle, Antony Unwin (Editors)

Springer, 2008, xiv + 936 pages, £ 197.00 / € 249.00 / US\$ 319.00, hardcover

ISBN: 978-3-540-33036-3

Table of contents

- | | |
|--------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------|
| I. Data Visualization | III.8. Structural adaptive smoothing by
propagation–separation methods |
| I.1. Introduction | III.9. Smoothing techniques for visualisation |
| II. Principles | III.10. Data visualization via kernel machines |
| II.1. A brief history of data visualization | III.11. Visualizing cluster analysis and finite mixture
models |
| II.2. Good graphics? | III.12. Visualizing contingency tables |
| II.3. Static graphics | III.13. Mosaic plots and their variants |
| II.4. Data visualization through their graph
representations | III.14. Parallel coordinates: Visualization, exploration
and classification of high-dimensional data |
| II.5. Graph-theoretic graphics | III.15. Matrix visualization |
| II.6. High-dimensional data visualization | III.16. Visualization in Bayesian data analysis |
| II.7. Multivariate data glyphs: Principles and
practice | III.17. Programming statistical data visualization in
the Java language |
| II.8. Linked views for visual exploration | III.18. Web-based statistical graphics using XML
technologies |
| II.8. Linked views for visual exploration | IV. Selected Applications |
| II.9. Linked data views | IV.1. Visualization for genetic network reconstruction |
| II.10. Visualizing trees and forests | IV.2. Reconstruction, visualization and analysis of
medical images |
| III. Methodologies | IV.3. Exploratory graphics of a financial dataset |
| III.1. Interactive linked micromap plots for
the display of geographically referenced
statistical data | IV.4. Graphical data representation in bankruptcy
analysis |
| III.2. Grand tours, projection pursuit guided tours,
and manual controls | IV.5. Visualizing functional data with an application
to eBay's online auctions |
| III.3. Multidimensional scaling | IV.6. Visualization tools for insurance risk processes |
| III.4. Huge multidimensional data visualization:
Back to the virtue of principal coordinates and
dendrograms in the new computer age | |
| III.5. Multivariate visualization by density estimation | |
| III.6. Structured sets of graphs | |
| III.7. Regression by parts: Fitting visually
interpretable models with GUIDE | |

Readership: Researchers and practitioners of almost any field.

This handbook shows hundreds of ways to visualize data by using modern, high-quality statistical graphics. The articles of 66 authors reveal basics and backgrounds as well as details and dynamics

of this fascinating area, which should be an essential part of any data analysis or statistical modelling.

The book includes over 500 examples of different graphs, such as parallel coordinate plots, grand tours, mosaic plots, matrix diagrams, micromap plots, and linked views. The interplay between multivariate statistical methods and various graphics is evident in several articles. In visualizing huge data sets efficiently, the advances in computer hardware and software have made totally new possibilities available. It is most enjoyable to see such a large number of specialists sharing their insights of these methods within one volume.

This book really feeds the imagination of the reader. High-dimensionally recommended!

Kimmo Vehkalahti: kimmo.vehkalahti@helsinki.fi
Department of Mathematics and Statistics
FI-00014 University of Helsinki, Finland

Statistical Design

George Casella

Springer, 2008, xxiii + 307 pages, £ 55.99 / € 69.95 / US\$ 84.95, hardcover

ISBN: 978-0-387-75964-7

Table of contents

- | | |
|-------------------------------------------------|--------------------------|
| 1. Basics | 5. Split plot designs |
| 2. Completely randomized designs | 6. Confounding in blocks |
| 3. Complete block designs | A. Designs illustrated |
| 4. Interlude: Assessing the effects of blocking | |

Readership: Researchers and graduate students.

In the Preface of his new book, George Casella states the following: “There are plenty of ‘Design of Experiments’ books available, many of which do fine job of describing not only how to design experiments, but also how to analyse them. So why bother with another book?” A good question into which Casella gives a good answer. He gives two main reasons to write one more book, ending up into a summary saying that “the goal is to describe the principles that drive good design, which are also the principles that drive good statistics”. I have difficulties to resist picking up also the following sentence from the Preface: “My best advice is that if you really want to understand statistical design, read (or even better, reread) Fisher”.

Casella succeeds exceptionally well to reach his goals. I greatly enjoyed browsing through this book. The author’s experience and writing skills together make this an excellent course book. Concepts are presented in a very reader-friendly and instructive way. I like the style to emphasize important facts with highlighted phrases like “A most important identity”. Such explanations help the reader to understand that now it’s time to take a breath and try to make a serious attempt to understand what is really going on. The layout of the book, huge amount of examples, and very clear writing make this a book highly recommended for anyone interested in statistical design.

Simo Puntanen: simo.puntanen@uta.fi
Department of Mathematics and Statistics
FI-33014 University of Tampere, Finland

Survival Analysis for Epidemiologic and Medical Research

Steve Selvin

Cambridge University Press, 2008, xiv + 282 pages, £ 50.00 / US\$ 95.00, hardcover (also available in softcover)

ISBN: 978-0-521-89519-4

Table of contents

- | | |
|-------------------------------------------------------|-----------------------------------------|
| 1. Rates and their properties | 7. Analysis of two-sample survival data |
| 2. Life tables | 8. General hazards model: parametric |
| 3. Two especially useful estimation tools | 9. General hazards model: nonparametric |
| 4. Product-limit estimation | Examples of R |
| 5. Exponential survival time probability distribution | Data |
| 6. Weibull survival time probability distribution | Problem set |

Readership: Users of Statistics, particularly in the biological sciences.

The author's stated aim is to introduce Survival Analysis 'using only elementary mathematics, with an emphasis on examples and intuitive explanations'. This does give a fair idea of the flavour of the book. However, it is all too easy to slip from this ideal unless great care is taken. For example, the first ten pages of Chapter 1 give a very gentle introduction to survival functions without taxing one's algebraic skill. But then, on page 11, after three short sentences, it says 'Therefore, ...' and the algebraic formula for the binomial probability distribution is given. The reader might well wonder why it is 'therefore', when no derivation has been given. Moreover, the formula contains the familiar (to mathematicians) bracket notation for a binomial coefficient without comment. Perhaps American biosciences students are better grounded in mathematics than English ones. One has to acknowledge the difficulty of presenting a mathematical subject to non-mathematicians. (But, saying that 'only one straight line touches a [continuous] curve at a specific point' (page 25) is a bit careless.)

On the positive side, the book does do what it sets out to do. That is, to take the reader through many of the routine computations that form the body of basic applied statistics: significance tests, confidence intervals, likelihood functions (skillfully introduced in Chapter 3), etc. A major strength of the book is that each technique is illustrated numerically, going through the arithmetic in detail. Moreover, the data referred to is 'real', taken from actual health studies; four data sets are listed at the end. There is also an introduction to the use of the R-language for Survival Analysis and some problems for the student to practise on.

In summary, while the mathematical statistician might grumble a little (see first paragraph above) I think that the book can be recommended for the target readership. The topics covered are the basic methods used in the field and the student will be well-equipped to be recruited into one of the associated professions.

Martin Crowder: m.crowder@imperial.ac.uk
 Mathematics Department, Imperial College
 London SW7 2AZ, UK

Linear Models in Statistics, Second Edition

Alvin C. Rencher, G. Bruce Schaalje

Wiley, 2008, xvi + 672 pages, £ 75.50 / € 91.10 / US\$ 127.50, hardcover

ISBN: 978-0-471-75498-5

Table of contents

- | | |
|-------------------------------------------------------------------------|-----------------------------------------------------------------------|
| 1. Introduction | 10. Multiple regression: random x 's |
| 2. Matrix algebra | 11. Multiple regression: Bayesian Inference |
| 3. Random vectors and matrices | 12. Analysis-of-variance models |
| 4. Multivariate normal distribution | 13. One-way analysis-of-variance: balanced case |
| 5. Distribution of quadratic forms in y | 14. Two-way analysis-of variance: balanced case |
| 6. Simple linear regression | 15. Analysis-of-variance: the cell means model
for unbalanced data |
| 7. Multiple regression: estimation | 16. Analysis-of-covariance |
| 8. Multiple regression: tests of hypotheses and
confidence intervals | 17. Linear mixed models |
| 9. Multiple regression: model validation and
diagnostics | 18. Additional models |

Readership: Statistics students, researchers willing to learn more about regression.

This book will definitely have readers whose nerves are in good mood having not been frustrated with difficult-to-read material. Namely this text is written in a particular clarity that deserves congratulations. There are numerous situations where the authors have been exceptionally patient in explaining and writing up the explanations in fuller details than usually. This means that not heavy mathematical background is necessary for the reader. Many topics needed are briefly reviewed before putting into use. According to the authors, one or two statistical methods courses would though be helpful.

In the Preface the authors state that “One overriding objective in preparation of this book has been the clarity of exposition”, where the authors type ‘clarity of exposition’ in italics – the only words in italics in the Preface! They continue that they hope that students, instructors, researchers, and practitioners will find this book more comfortable than most. After browsing the book, I’m inclined to believe that they have succeeded worth applause. As for notation used, I was pleased to observe that it is pretty close to that one I’m personally used to. This indeed clearly written book will do great service for advanced undergraduate and also for PhD students.

Simo Puntanen: simo.puntanen@uta.fi
 Department of Mathematics and Statistics
 FI-33014 University of Tampere, Finland

A Primer on Linear Models

John F. Monahan

Chapman & Hall/CRC, 2008, xiv + 287 pages, £ 25.99 / € 40.20 / US\$ 49.95, softcover

ISBN: 978-1-4200-6201-4

Table of contents

- | | |
|----------------------------------------------|-----------------------------------------|
| 1. Examples of the general linear model | 7. Further topics in testing |
| 2. The linear least squares problem | 8. Variance components and mixed models |
| 3. Estimability and least squares estimators | 9. The multivariate linear model |
| 4. The Gauss-Markov theorem | Appendix A: Review of linear algebra |
| 5. Distributional theory | Appendix B: Lagrange multipliers |
| 6. Statistical inference | |

Readership: Graduate students in courses of linear models, researchers interested in foundations of basic linear models.

For a few weeks, I have been having this book on my desk, not to be reviewed but to be utilized in my teaching: I'm giving right now a course in linear models. Anyways, I was in browsing this book to find some particular properties related to experimental design and I'm pleased to tell that I found the book very helpful. It is naturally much more fun to write a review about a good book than about not so good one – this opinion may surely vary from a person to person though. So I decided to write a short book review about Monahan's book.

As the author puts it, he has two features in particular in mind: the use of non-full-rank design matrices to seamlessly incorporate regression, analysis of variance and mixed models, and attention to the exact, finite sample theory supporting common statistical methods. The book has 287 pages which is less than half of many books in linear models. Simply the author does not plan to cover everything, he has chosen topics following his experiences and I find the result is very nice, very readable and in particular I like the idea of avoiding leaps in the development and proofs, or referring to other sources for the details of the proofs. This is a useful well-written instructive book.

Simo Puntanen: simo.puntanen@uta.fi
 Department of Mathematics and Statistics
 FI-33014 University of Tampere, Finland

Probability and Statistics with R

Maria Dolores Ugarte, Ana F. Militino, Alan T. Arnholt

Chapman & Hall/CRC, 2008, xxvi + 700 pages, £ 46.99 / US\$ 89.95, hardcover

ISBN: 978-1-58488-891-8

Table of contents

- | | |
|---------------------------------------------|----------------------------------------------------------------|
| 1. A brief introduction to S | 9. Hypothesis testing |
| 2. Exploring data | 10. Nonparametric methods |
| 3. General probability and random variables | 11. Experimental design |
| 4. Univariate probability distributions | 12. Regression |
| 5. Multivariate probability distributions | Appendix A: S commands |
| 6. Sampling and sampling distributions | Appendix B: Quadratic forms and random vectors
and matrices |
| 7. Point estimation | |
| 8. Confidence intervals | |

Readership: Students taking introductory statistics.

I approached this book with great enthusiasm. It appeared to be exactly the text that I sought for the introductory statistics courses that I teach to engineering students. I have for years been dissatisfied with traditional introductory texts that did not take advantage of statistical software, especially freely-available, high-quality software like R. (For the record, I am a member of the core development team for the R Project.) Typical introductory texts may include a few figures produced with a computer package or output from, say, fitting a linear model or perhaps even a generalized linear model, but all too frequently the presentation of topics has changed little from the 1960's when I took such a course. In those pre-calculator days, probabilities from all but the simplest distributions could only be evaluated from tables, graphics were limited to what you could draw by hand on a piece of graph paper, we didn't bother considering the concept of a p-value because they couldn't be evaluated and fitting even a simple linear regression model was agonizing. Needless to say, computing has changed a bit since then.

I had hoped that this text would present the statistical theory and practice for an introductory course using R as the computational and graphical engine that can allow us to drop the tedious, convoluted and often decidedly suboptimal calculation formulas and explanations that plague the typical text. I'm sorry to say that my hopes were not completely realized. In the initial chapters I more-or-less agreed with the choice of topics and the presentation of the probability and statistics content. I wasn't enthusiastic about some of the R code that is shown. I would have done many things differently but one of the advantages of R, and the S language in general, is that you can do things in the way that makes sense to you. However, I don't think I would be alone in saying that the R code shown here is unnecessarily complicated and much less effective than it could be.

Nevertheless, I would say that through chapter 6 I agreed with the general principles even if I would quibble about some of the details, especially the number of complex formulas presented. In later chapters, especially chapters 11 and 12, I felt that many of the advantages of using R were not being realized. These chapters are very formula-heavy in a way that distracts from the practical results that can be obtained from the software. Why devote several pages of dense formulas deriving expressions for least-squares estimates of the coefficients in a simple linear regression when the estimates aren't even calculated according to those formulas? Our students, especially engineering students, will spend the working lives in front of a computer so they don't

need such formulas to be able to evaluate the estimates and I don't see how exposure to such formulas will help them to understand the simple linear regression model better.

In summary, I like the idea of such a book but I was not enthusiastic about the execution. It does not use R nearly as effectively as I would like to see.

Douglas M. Bates: bates@stat.wisc.edu

Department of Statistics, University of Wisconsin
1300 University Avenue, Madison, WI 53706–1532, USA

Applied Data Analytic Techniques for Turning Points Research

Patricia Cohen (Editor)

Routledge, 2008, viii + 243 pages, £ 21.99 / US\$ 39.95, softcover (also available in hardcover)

ISBN: 978-0-8058-5452-7

Table of contents

1. The origins of this book
 2. The interrelationship of temporally distinct risk markers and the transition from childhood physical aggression to adolescent violent delinquency
 3. Estimating time-varying causes and outcomes, with applications to incarceration and crime
 4. Turning points in family contact during emerging adulthood
 5. Testing turning points using latest growth curve models: Competing models of substance abuse and desistance in young adulthood
 6. Modeling age-based turning points in longitudinal life-span growth curves of cognition
 7. Bereavement as a potential turning point: Modeling between-person variability in adjustment to conjugal loss
 8. Application of change point theory to modelling state related activity in fMRI
 9. Using an econometric model of change points to locate turning points in individual time series
 10. Developmental structural change in the maturity of role assumption
- Appendix 1: Computer programs for change-point analysis
- Appendix 2: Equivalent forms of the Level 1 equation for the bereavement model in Chapter 7
- Appendix 3A: The variance of the EWMA statistic assuming ARMA (1,1) noise
- Appendix 3B: Maximum likelihood estimation of change point
- Appendix 3C: EM algorithm for Gaussian mixture model
- Appendix 3D: Estimating variance components using restricted maximum likelihood

Readership: Advanced students and researchers interested in detecting changes in data series.

Turning points, change points, or sometimes break points are changes in values of variables in a stream of observations. Tools for detecting such changes are becoming more important, driven by the growth of interest in streaming data. This book contrasts 'potentially many change points over time in a single measured unit' with 'turning points' which represent 'a relatively lasting "new stage" of function, development, or problem.' It consists of a collection of nine examples, plus an introductory chapter. The volume also includes several appendices giving software examples and mathematical details, although I think the software examples are too limited to be of much value – in a second edition this appendix could usefully be expanded. The book is aimed at social and behavioural scientists, with chapters describing applications such as adolescent violent delinquency, incarceration and crime, substance abuse, growth curves of cognition, conjugal loss, and so on.

Much work in this area is narrowly focused on methods for detecting turning points in data series isolated from any context. This book is rather unusual in that all of the chapters are driven by specific applications. That, of course, is where the power and potential of new developments will come from – now that the basic ideas of turning point detection have been (well-)established, significant progress will come from identifying important additional aspects of the particular application and taking advantage of that in the modelling.

This book will certainly be valuable for social or behavioural scientists concerned with detecting turning points. It will also be valuable for statisticians seeking a broader, and applied perspective on change-point detection techniques themselves – it would make good supplementary reading for a course on that topic.

David J. Hand: *d.j.hand@imperial.ac.uk*
Mathematics Department, Imperial College
London SW7 2AZ, UK

Probability and Information: An Integrated Approach, Second Edition

David Applebaum

Cambridge University Press, 2008, xvi + 273 pages, £ 65.00 / US\$ 130.00, hardcover (also available in softcover)

ISBN: 978-0-521-89904-8

Table of contents

- | | |
|--------------------------------------------------------|-------------------------------------------------------------------------------------|
| 1. Introduction | 9. Random vectors |
| 2. Combinatorics | 10. Markov chains and their entropy |
| 3. Sets and measures | Exploring further |
| 4. Probability | Appendix 1. Proof by mathematical induction |
| 5. Discrete random variables | Appendix 2. Lagrange multipliers |
| 6. Information and entropy | Appendix 3. Integration of $\exp(-1/2x^2)$ |
| 7. Communication | Appendix 4. Table of probabilities associated with the standard normal distribution |
| 8. Random variables with probability density functions | Appendix 5. A rapid review of matrix algebra |

Readership: Beginning students in mathematics, statistics, computer science, or economics.

This is an elementary introduction to probability, at a level suitable for a university first year introduction to the subject. It requires no mathematics beyond basic calculus. It successfully introduces measure concepts, sidestepping the full (and non-introductory) sigma-algebra development by driving the discussion using Boolean algebras. At first glance, it competes with numerous other introductions to probability. Where this volume differs from many of the alternatives is that it places considerable emphasis on information theory, communication, and entropy. Given the intimate connections between the domains of probability and information theory, not to mention the more advanced relationships (beyond the scope of this book) between statistical topics such as Bayesian inference and minimum message and minimum description length inference, this is a valuable and useful strategy. I recommend it for first year undergraduate students, who often fail to grasp the links between different mathematical domains.

This updated edition has minor amendments and an additional chapter on Markov chains and entropy. Solutions to most exercises are given on the web.

David J. Hand: *d.j.hand@imperial.ac.uk*
Mathematics Department, Imperial College
London SW7 2AZ, UK

Statistical Decision Theory: Estimation, Testing, and Selection

Friedrich Liese, Klaus-J. Miescke

Springer, 2008, xvii + 677 pages, £ 55.99 / € 69.95 / US\$ 89.95, hardcover

ISBN: 978-0-387-73193-3

Table of contents

- | | |
|---------------------------------------------------|--------------------------------------------------------|
| 1. Statistical models | 6. Large sample approximations of models and decisions |
| 2. Tests in models with monotonicity properties | 7. Estimation |
| 3. Statistical decision theory | 8. Testing |
| 4. Comparison of models, reduction by sufficiency | 9. Selection |
| 5. Invariant statistical decision models | |

Readership: Advanced Master's students, PhD students, researchers in mathematical statistics and decision theory.

The central theme of this book is the nature of optimal decisions, in both general and specific problems, and how to reach them. The first part of the book treats estimators, tests, and selection in parallel. Chapter 1 describes the technical and mathematical background and tools, and then Chapter 2 introduces the Neyman-Pearson lemma and its extensions, before progressing, in Chapter 3, to the general framework of decision theory. Chapters 4, 5, and 6 are concerned, respectively, with comparison of models, invariant decision models, and large sample approximations. The last three chapters treat the topics of estimating, testing, and selection separately.

The authors say that they were motivated to write the book by a deficiency of coverage of selection rules in books on decision theory, and contrast the ad hoc nature of most of the prior literature with the more unified presentation in this volume. The book describes both frequentist and Bayesian approaches.

This is a deep theoretical and mathematical presentation of decision theory, which I imagine will become a leading reference work for researchers in the area.

David J. Hand: *d.j.hand@imperial.ac.uk*
Mathematics Department, Imperial College
London SW7 2AZ, UK

Proportional Hazards Regression

John O'Quigley

Springer, 2008, xviii + 542 pages, € 62.95 / £ 49.99 / US\$ 79.95, hardcover

ISBN: 978-0-387-25148-6

Table of contents

- | | |
|------------------------------------------------|------------------------------------------------|
| 1. Introduction | 9. Inference: Likelihood |
| 2. Background: Probability | 10. Inference: Stochastic integrals |
| 3. Background: General inference | 11. Inference: Small samples |
| 4. Background: Survival analysis | 12. Inference: Change point models |
| 5. Marginal survival | 13. Explained variation |
| 6. Regression models and subject heterogeneity | 14. Explained randomness |
| 7. Inference: Estimating equations | 15. Survival given covariates |
| 8. Inference: Functions of Brownian motion | 16. Proofs of theorems, lemmas and corollaries |

Readership: Researchers in Statistics, students and academics.

In his Preface the author refers to the widely-adopted modern treatment of the theory of Survival Analysis in terms of counting processes, based on the rather esoteric concepts of martingales and stochastic integrals. His aim is to provide an alternative treatment, mostly in terms of more classical probability results such as central limit theorems for sums of random variables and the functional version leading to the Wiener process limit. Although, as the author says, the approach should make the theory more accessible, a reasonable level of mathematics, say the first year of an English degree course, is still required of the reader.

The book is clearly intended to be student-friendly. Each chapter begins with a section called Summary and a following one called Motivation; each chapter ends with some exercises and class projects. Theorems, etc., are stated within the chapters but proofs are relegated to the end, in Chapter 16. The huge Bibliography contains 710 references and the four-page Index looks fairly comprehensive.

My personal impressions of the book are as follows. It is very carefully written, with detailed explanation and discussion everywhere. Practical issues are highlighted though no data sets are numerically processed. Many topics are covered: to pick a few chapter headings, there is prerequisite treatment of probability, statistical inference and survival analysis, followed by regression models and subject heterogeneity, functions of Brownian Motion, stochastic integrals, change point models, and explained variation and randomness. I believe that the book can be thoroughly recommended to the student starting his research in the field and to the practitioner who needs to understand some of the theory.

Martin Crowder: m.crowder@imperial.ac.uk
 Mathematics Department, Imperial College
 London SW7 2AZ, UK

Global Sensitivity Analysis: The Primer

Andrea Saltelli, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana, Stefano Tarantola
 Wiley, 2008, xii + 292 pages, £ 55.00 / € 74.30 / US\$ 110.00, hardcover
 ISBN: 978-0-470-05997-5

Table of contents

- | | |
|-----------------------------------------|--------------------------------------------------------|
| 1. Introduction to sensitivity analysis | 4. Variance-based methods |
| 2. Experimental designs | 5. Factor mapping and metamodelling (with Peter Young) |
| 3. Elementary effects method | 6. Sensitivity analysis: From theory to practice |

Readership: Postgraduate students and practitioners in statistics, mathematics, engineering, physics, chemistry, environmental sciences, biology, toxicology, actuarial sciences, and econometrics; engineers working on risk analysis and financial analysts concerned with pricing and hedging.

This book is written by leading sensitivity analysis practitioners with mainly computer science and mathematical backgrounds. It aims to provide an easy-to-read and self-contained treatment of the subject, covering global methods, sensitivity measures and how they may help find more robust or parsimonious models. It takes a practical approach with motivation for sensitivity analysis, reviews required statistical concepts, and provides a guide to potential applications. Chapter 4 and each section of chapters 1, 2, 3 and 5 present lists of questions in the very beginning to draw the reader's attention, and then answer the questions to introduce the ideas and methods. It contains suitable examples and exercises to help understand the material and illustrate the applications. Each of the first five chapters has solutions to the exercises and chapter six presents a collection of examples taken from various areas.

This book focuses on global methods. It does not mention existing research on local methods of statistical diagnostics, say the work in the book *Sensitivity Analysis in Linear Regression*, by S. Chatterjee and A.S. Hadi (1988, Wiley), nor the work based on Jacobians in *Matrix Differential Calculus with Applications in Statistics and Econometrics*, second ed. by J.R. Magnus and H. Neudecker (1999, Wiley). However, it is a welcome addition to its sister volume *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models* by A. Saltelli, S. Tarantola, F. Campolongo and M. Ratto (2004, Wiley).

Shuangzhe Liu: shuangzhe.liu@canberra.edu.au
 University of Canberra
 Canberra ACT 2061, Australia

Statistical Practice in Business and Industry

Shirley Coleman, Tony Greenfield, Dave Stewardson, Douglas C. Montgomery (Editors)
 Wiley, 2008, xii + 435 pages, £ 55.00 / € 74.30 / US\$ 110.00, hardcover
 ISBN: 978-0-470-01497-4

Table of contents

1. Introduction (*Shirley Coleman, Tony Fouweather and Dave Stewardson*)
2. A history of industrial statistics (*Jeroen de Mast*)
3. Statistical consultancy
 - 3.I. A statistician in industry (*Ronald J. M. M. Does and Albert Trip*)
 - 3.II. Black belt types (*Roland Caulcutt*)
 - 3.III. Statistical consultancy units at universities (*Ronald J. M. M. Does and András Zempléni*)
 - 3.IV. Consultancy? . . . What's in it for me? (*Roland Caulcutt*)
4. The statistical efficiency conjecture (*Ron S. Kenett, Anne De Frenne, Xavier Tort-Martorell and Chris McCollin*)
5. Management statistics (*Irena Ograjenšek and Ron S. Kenett*)
6. Service quality (*Irena Ograjenšek*)
7. Design and analysis of industrial experiments (*Timothy J. Robinson*)
8. Data mining (*Paola Cerchiello, Silvia Figini and Paolo Giudici*)
9. Using statistical process control for continual improvement (*Donald J. Wheeler and Øystein Evandt*)
10. Advanced statistical process control (*Murat Kulahci and Connie Borror*)
11. Measurement system analysis (*Giulio Barbato, Grazia Vicario and Raffaello Levi*)
12. Safety and reliability
 - 12.I. Reliability engineering: The state of the art (*Chris McCollin*)
 - 12.II. Stochastics for the quality movement: An integrated approach to reliability and safety (*M. F. Ramalhoto*)
13. Multivariate and multiscale data analysis (*Marco P. Seabra dos Reis and Pedro M. Saraiva*)
14. Simulation in industrial statistics (*David Ríos Insua, Jorge Muruzabal, Jesús Palomo, Fabrizio Ruggeri, Julio Holgado and Raúl Moreno*)
15. Communication (*Tony Greenfield and John Logsdon*)

Readership: Business and Industrial people with statistical interests.

According to the preface, “This book was originally conceived as a handbook arising from the European Commission funded Fifth Framework project entitled European Network for Promotion of Business and Industrial Statistics (pro-ENBIS).” It is essentially a compendium of summaries of statistical work in quite general terms, the sort of book that readers can refer to in order to get started on a topic. For that purpose it is excellent but, naturally enough, most of the topics do not take the reader very far and the personal orientations of the authors limit the specific directions suggested for further reading. Two general non-technical chapters that are especially good are numbers 3 and 15, respectively Statistical Consulting and Communication.

Norman R. Draper: draper@stat.wisc.edu
 Department of Statistics, University of Wisconsin – Madison
 1300 University Avenue, Madison, WI 53706-1532, USA

Quantitative *Eco-nomics*: How sustainable are our economies?

Peter Bartelmus

Springer, 2008, xxx + 330 pages, € 94.95 / £ 75.00 / US\$ 149.00, hardcover

ISBN: 978-1-4020-6965-9

Table of contents

Part I: Questions, Questions, Questions

1. What on Earth is wrong?
2. What's economics got to do with it?
3. Sustainable development – blueprint of fig leaf?

Part II: Assessing the Physical Base of the Economy

4. Statistics and indicators
5. Aggregation: from indicators to indices
6. Energy and material flow accounting

Part III: Greening the Economic Accounts

7. Linking the physical and monetary accounts
8. SEEA – the system for integrated environmental and economic accounting
9. Corporate accounting: accounting for accountability

Part IV: Analysis – Modelling Sustainability

10. Diagnosis: *has* the economy behaved sustainably?
11. Prediction: *will* economic growth be sustainable?
12. Policy analysis: *can* we make growth sustainable?

Part V: Strategic Outlook

13. Strategies: tackling the limits to growth
14. Globalisation and global governance
15. Questions, questions, questions – and some answers

Annexes

- I. Market failure and environmental cost internalisation – a primer
- II. Economic rent and natural resource depletion
- III. SEEA Germany – a pilot case study

Readership: Government officials, accountants, corporations, researchers, scholars, statisticians, and students of environment, economics and sustainable development.

Three very important terms in everyday discussions about the future of the world: economy, ecology, and sustainable development, are the basis of investigations presented in the book. Starting with some historical comments and going through measures and indicators (not only statistical and macroeconomic) used to quantify the degree of sustainability of development, the author suggests (almost constructs) an operational conception of how to analyse economic growth and economic policies which take into account ecology and the environment. There are many figures and diagrams explaining the problems discussed, a bit of mathematics in models considered in general terms, and many references to other publications addressing those problems. Each chapter ends with review questions provoking one to think over the presented material. It is a really valuable contribution, even though it gives only some answers to the many inspiring questions posed in the book. It could be a guide not only for students, but also for researchers and people who aspire to be advisers to politicians and government officials.

Jan Hauke: jhauke@amu.edu.pl
Adam Mickiewicz University
Poznań, Poland

Probably Not: Future Prediction Using Probability and Statistical Inference

Lawrence N. Dworsky

Wiley, 2008, xvi + 310 pages, £ 35.50 / € 42.90 / US\$ 59.95, softcover

ISBN: 978-0-470-18401-1

Table of contents

- | | |
|-------------------------------------------------------|-----------------------------------------------------------------|
| 1. An introduction to probability | 10. Combined and conditional probabilities |
| 2. Probability distribution functions and some basics | 11. Scheduling and waiting |
| 3. Building a bell | 12. Stock market portfolios |
| 4. Random walks | 13. Benford, Parrondo, and Simpson |
| 5. Life insurance and social security | 14. Networks, infectious disease propagation, and chain letters |
| 6. Binomial probabilities | 15. Bird counting |
| 7. Pseudorandom numbers and Monte Carlo simulations | 16. Statistical mechanics and heat |
| 8. Some gambling games in detail | 17. Introduction to statistical analysis |
| 9. Traffic lights and traffic | 18. Chaos and quanta |

Readership: Practitioners and students who would like to learn more about the role of probability and statistics in everyday decision making.

The book addresses important problems – usually analysed with the use of advanced methods of probability theory and statistics – by discussing clearly presented basic issues. It is worth noting that this is done without complex formulas. To understand them, it suffices to have only a basic knowledge of mathematics and basic mathematical skills, but combined with logical thinking. For such readers the book is a very good source of improving and extending their knowledge. The real-world examples from simple games through biological problems and stock market investments to mechanics are described using the previously introduced and discussed issues of probability theory in an easily understandable way. The graphs included in the book also contribute to the elucidation of the problems discussed. So, the book is a very good tool for students and specialists in different areas who do not like advanced statistics, but are obliged to understand, at least partly, the process of decision making.

Jan Hauke: jhauke@amu.edu.pl
Adam Mickiewicz University
Poznań, Poland

Analysis of Multivariate Social Science Data, Second Edition

David J. Bartholomew, Fiona Steele, Irini Moustaki, Jane Galbraith

Chapman & Hall/CRC, 2008, xii + 371 pages, £ 29.99 / US\$ 69.95, softcover

ISBN: 978-1-58488-960-1

Table of contents

- | | |
|----------------------------------|-----------------------------------------------------------------|
| 1. Setting the scene | 7. Factor analysis |
| 2. Cluster analysis | 8. Factor analysis for binary data |
| 3. Multidimensional scaling | 9. Factor analysis for ordered categorical variables |
| 4. Correspondence analysis | 10. Latent class analysis for binary data |
| 5. Principal components analysis | 11. Confirmatory factor analysis and structural equation models |
| 6. Regression analysis | 12. Multilevel modelling |

Readership: Graduate social science students and researchers.

This is the second edition of *The Analysis and Interpretation of Multivariate Data for Social Scientists*. The major difference is the extension from methods in which all the variables had the same status, to include asymmetric situations such as regression (with predictor and response variables), confirmatory factor analysis, structural equation models, and multilevel models. Given that Bartholomew is one of the leading thinkers in the area of latent variable models, it will come as no surprise to learn that this book emphasises them – in particular noting the links between latent variable models which are normally treated quite separately.

I am pleased that the authors emphasise that the book is in no sense a cookbook. Attempts at cookbooks in statistics fail to convey the integrated whole of the discipline, and how superficially distinct methods may in fact be closely related. However, one consequence of this is that the book is not really to be dipped into (it is based on a course, taught and developed over 20 years) but is best read in order.

It assumes a knowledge of basic statistical modelling and inference up to and including the analysis of contingency tables, analysis of variance, and regression. However, the presentation is well-matched to its intended audience, relying on only the minimal necessary mathematics and driving the development with examples, figures, and verbal descriptions.

There is an associated website giving the data sets and software. Each chapter concludes with a section of ‘additional examples and further work’, along with references through which the reader can pursue deeper studies.

This is the sort of book from which I would have liked to have learnt multivariate statistics.

David J. Hand: d.j.hand@imperial.ac.uk
 Mathematics Department, Imperial College
 London SW7 2AZ, UK

Linking and Aligning Scores and Scales

Neil J. Dorans, Mary Pommerich, Paul W. Holland (Editors)

Springer, 2007, xx + 396 pages, £ 43.99 / € 54.95 / US\$ 69.95, hardcover

ISBN: 978-0-387-49770-9

Table of contents

- | | |
|------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|
| 1. Overview | 11. Concordance: the good, the bad, and the ugly
(Mary Pommerich) |
| Part 1: Foundations | 12. Some further thoughts on concordance
(Richard Sawyer) |
| 2. A framework and history for score linking
(Paul W. Holland) | Part 5: Vertical Scaling |
| 3. Data collection designs and linking procedures
(Michael J. Kolen) | 13. Practical issues in vertical scaling
(Deborah J. Harris) |
| Part 2: Equating | 14. Methods and models for vertical scaling
(Richard J. Patz and Lihua Yao) |
| 4. Equating: best practices and challenges to best
practices (Nancy S. Petersen) | 15. Vertical scaling and no child left behind
(Wendy M. Yen) |
| 5. Practical problems in equating test scores: a
practitioner's perspective (Linda L. Cook) | Part 6: Linking Group Assessments to Individual
Assessments |
| 6. Potential solutions to practical equating issues
(Alina A. von Davier) | 16. Linking assessments based on aggregate
reporting: background and issues
(David Thissen) |
| Part 3: Tests in transition | 17. An enhanced method for mapping state
standards onto the NAEP scale
(Henry I. Braun and Jiahe Qian) |
| 7. Score linking issues related to test content
changes (Jinghua Liu and Michael E. Walker) | 18. Using aggregate-level linkages for estimation
and validation: comments on Thissen and
Braun & Qian (Daniel Koretz) |
| 8. Linking scores derived under different modes
of test administration (Daniel R. Eignor) | 19. Postscript |
| 9. Tests in transition: discussion and synthesis
(Robert L. Brennan) | |
| Part 4: Concordance | |
| 10. Sizing up linkages
(Neil J. Dorans and Michael E. Walker) | |

Readership: Educationalists and other researchers who need to match different forms of a test or score.

Science progresses, and in the process, improved measurement methods and technologies take over from older ones. This poses the problem of continuity, something particularly apparent in areas such as time series of economic and official data, where changing definitions mean that one may have suspicions about how well they are matched, and also in education, where different kinds of test procedures seem to appear with regularity.

Scale linking is the construction of transformations relating the results of measurements made by one method to those of another. Such linking may be done in various ways, with various ultimate objectives. This book divides these objectives into predicting (predict an examinee's score on a test from other information about them), aligning (establishing a link so that two distinct tests are comparable), and equating (to link the scores to such an extent that they can be used interchangeably). Although the book focuses on educational tests, its relevance is clearly much wider.

This is an edited book, but the editors have made a convincing effort to produce a unified whole. It is not a collection of disparate research papers, but a very useful compilation of articles

covering the area. One might say that the editors have done a first class job of linking the separate contributions.

David J. Hand: *d.j.hand@imperial.ac.uk*
Mathematics Department, Imperial College
London SW7 2AZ, UK

Statistical Analysis and Modelling of Spatial Point Patterns

Janine Illian, Antti Penttinen, Helga Stoyan, Dietrich Stoyan

Wiley, 2008, xx + 536 pages, £ 70.00 / € 94.50 / US\$ 140.00, hardcover

ISBN: 978-0-470-01491-2

Table of contents

- 1. Introduction
- 2. The homogeneous Poisson point process
- 3. Finite point processes
- 4. Stationary point processes
- 5. Stationary marked point processes
- 6. Modelling and simulation of stationary point processes
- 7. Fitting and testing point process models
- Appendix A. Fundamentals of statistics
- Appendix B. Geometrical characteristics of sets
- Appendix C. Fundamentals of geostatistics

Readership: Researchers in a wide variety of application areas including environmental statistics, ecology, physics, materials science, geostatistics, biology, astronomy, and medicine.

Spatial point processes describe the configurations of objects which are randomly distributed in some space. There are many visually obvious examples of such situations – the distribution of trees in a forest, stars in the sky, people in a crowd, pores on the skin, and so on – but there are also countless more abstract examples. One perspective on the subject is that it is the study of short-range interactions between points, so leading to an understanding of the configurations and distributions of the points.

Spatial point processes represent one of the intersections of probability theory and statistics, and the authors reflect this in their description of the book as dealing with both the *analysis* of point distributions and the *synthesis* of point process models. The abstract flavour this brings to the subject means that methods may have very wide applicability over different application domains. This applicability, in turn, is reflected by the large number of interesting examples described in the book.

The book provides a comprehensive overview of the area.

David J. Hand: *d.j.hand@imperial.ac.uk*
Mathematics Department, Imperial College
London SW7 2AZ, UK

Meta Analysis: A Guide to Calibrating and Combining Statistical Evidence

Elena Kulinskaya, Stephan Morgenthaler, Robert G. Staudte

Wiley, 2008, xiv + 260 pages, £ 37.50 / € 50.70 / US\$ 75.00, softcover

ISBN: 978-0-470-02864-3

Table of contents

Part I. The Methods

1. What can the reader expect from this book?
2. Independent measurements with known precision
3. Independent measurements with unknown precision
4. Comparing treatment to control
5. Comparing K treatments
6. Evaluating risks
7. Comparing risks
8. Evaluating Poisson rates
9. Comparing Poisson rates
10. Goodness-of-fit testing
11. Evidence for heterogeneity of effects and transformed effects
12. Combining evidence: fixed standardized effects model
13. Combining evidence: random standardized effects model

14. Meta-regression

15. Accounting for publication bias

Part II. The theory

16. Calibrating evidence in a test
17. The basics of variance stabilizing transformations
18. One-sample binomial tests
19. Two-sample binomial tests
20. Defining evidence in t -statistics
21. Two-sample comparisons
22. Evidence in the chi-squared statistic
23. Evidence in F -tests
24. Evidence in Cochran's Q for heterogeneity of effects
25. Combining evidence from K studies
26. Correcting for publication bias
27. Large-sample properties of variance stabilizing transformations

Readership: Statistical consultants and researchers in medicine, the social sciences, forensic statistics and other areas.

Meta analysis describes the integration of results from diverse studies, to shed light on a common question. The area has not been uncontroversial in the past, though the problems arose more from a lack of rigour in the implementation than fundamental problems with the idea. Nowadays the approach is recognised and generally respected.

This book presents a novel approach, based on transforming the results of the different studies to a common scale, permitting a more ready combination and interpretation. This common scale is intended to represent the 'evidence' in different studies. While I endorse this aim, I have to say that I have always found Royall's notion of evidence (*Statistical Evidence*, Chapman and Hall, 1997) very convincing. He argues that, although p -values are often regarded as measures of evidence, they have a number of inconsistencies, and instead argues in favour of using likelihood ratio as a measure of evidence.

The book is divided into two parts, the first dealing with the practical methods and the second the underlying theory and justification. This order is rather unusual, in that it is more usual to present the underlying theory first, and then move on to applications and examples, but it certainly works here. There is a link to a website containing appropriate R software. The book is illustrated with copious examples.

David J. Hand: d.j.hand@imperial.ac.uk
 Mathematics Department, Imperial College
 London SW7 2AZ, UK

The Concise Encyclopedia of Statistics

Yadolah Dodge

Springer, 2008, x + 616 pages, € 249.00 / £ 191.50 / US\$ 299.00, hardcover

ISBN: 978-0-387-31742-7

Readership: Students and practitioners using statistics and visiting libraries.

A “concise encyclopedia” is an exceptional challenge, since, according to Merriam-Webster Online Dictionary at <http://www.merriam-webster.com/>, the first word “suggests the removal of all that is superfluous or elaborative”, while the second word refers to “a work that contains information on all branches of knowledge or treats comprehensively a particular branch of knowledge”.

This book claims to concentrate “on the most important topics” (of Statistics) and explain those “as deeply as space has allowed”. In addition, it is said in the preface that the book “leaves no central topics out”. It is not an easy task, of course, as no “official” list of the most important or central topics exists. Hence, the inclusions (and exclusions) will definitely reflect a personal taste and expertise of the author.

The process leading to the current version of the book has taken 25 years. The earlier versions were written in French, and only quite recently the book was translated in English. Some consequences may be seen in the inclusion, exclusion, and weighting of the topics. For example, the French School of data analysis is rather heavily represented. In that light, the most important method of data analysis would be Correspondence Analysis (CA). Its description is indeed one the longest in the whole book, occupying the pages 119–126.

According to p.150, “data analysis can be considered to be the essence of statistics to which all other aspects of the subject are linked”. This is a nice view, but unfortunately the given classification of data analysis does not seem very comprehensive. In addition, the essence could have been better stressed under the term Statistics (p.518).

I have nothing against CA, but I would certainly have included (or at least mentioned) certain multivariate methods like Factor Analysis and Principal Components as well, since I would easily consider those as central topics of Statistics. Of course, that reflects my personal taste and expertise, but still I find it quite odd that the term “factor” refers only to the design of experiments in a book with such a general title.

There are also topics I would definitely exclude, such as Inertia Matrix (which merely allocates two more pages for CA). Also Lagrange Multiplier, Linear Programming, and Operations Research are not very central topics of Statistics, I think. And while Regression Analysis is indeed central, I would say that Partial Least Absolute Deviation Regression is far from central.

The book also includes biographies of some more or less famous statisticians, in alphabetical order with the terms and concepts. I would find it useful to have an alphabetical list of the biographies, and actually another list of the terms in general. Now, the book begins from Acceptance Region and ends to Yule, George Udny. There is no way to browse the contents in a concise form. Such a list could have been provided automatically, and perhaps it could still be provided on web, for example. Even better (although demanding a bit of hard work) would be a grouped list of the terms, which would reveal the structure and the weighting of the book. Based on the current version, it is not easy to see what has been considered “central”.

Many formulas could be more concise, for example those on pages 167, 187, 306, 314, and 471. However, in general, the book is quite easy to read, and the cross-references are useful. In the topic of Distance Table (which I would have called Distance Matrix) some editorial work for the next version would be needed, since the example of five students and their grades is exactly the same that appears in the previous topic, Distance. Another topic is Statistical

Software, which for some reason includes two different versions of some explanations, for example, about BMDP, SAS, SPSS etc. The texts here seem to give quite an old-fashioned picture of these issues. Perhaps it would be more dynamic to refer to some web content, such as the list of software providers in Stata web site http://www.stata.com/links/stat_software.html. These are rare exceptions, however, and I did not find much typos in the book. In all, it is a useful reference that should be found in many academic and corporate libraries.

Kimmo Vehkalahti: kimmo.vehkalahti@helsinki.fi
 Department of Mathematics and Statistics
 FI-00014 University of Helsinki, Finland

Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis

Uffe B. Kjaerulff, Anders L. Madsen

Springer, 2008, xviii + 318 pages, £ 51.99 / € 64.95 / US\$ 79.95, hardcover

ISBN: 978-0-387-74100-0

Table of contents

Part I. Fundamentals

1. Introduction
2. Networks
3. Probabilities
4. Probabilistic networks
5. Solving probabilistic networks

Part II. Model Construction

6. Eliciting the model
7. Modeling techniques
8. Data-driven modeling

Part III. Model Analysis

9. Conflict analysis
10. Sensitivity analysis
11. Value of information analysis

Readership: Practitioners of applied artificial intelligence.

The basic theory of Bayesian Networks evolved in the nineties leading to the very well known monographs by Pearl and Lauritzen. Lauritzen's book was based on the work of a distinguished group consisting of Dawid, Spiegelhater, Lauritzen etc. The object was to solve structured but real life decision and diagnostic problems through a directed graphical model representing causal relations, decision nodes and utility nodes, complex enough to represent reality, simple enough to aid quick, but expert decision making.

The book under review is by two well known contributors to this general area. The first part of the book, consisting of five chapters, provides a very clear, insightful introduction to this theory. There is a lot of material but, rightly, the stress is on the DAG's (Directed Acyclic Graphs), which are a very natural model for an inference/decision engine based on causal analysis and conditional probabilities. Of course there is much else, but even with all this, the DAG's are much simpler than the chain graphs, which have a lovely theory and are more general but more difficult to understand and use.

After developing the basic theory, Kjaerulff and Madsen discuss in the remaining two parts of the book how to construct the graph for particular problems and related issues like sensitivity and conflicts between some part of the model and some of the evidence or between different evidential elements. These two parts are also well written and show novelty and imagination, but they are probably not yet in a final form. Especially the issue of conflict between Bayesian modeling and

data raise interesting theoretical or philosophical issues that need further discussion. Suppose I get data which is very rare under my model. When do we treat it as unusual but acceptable? When would we prefer to discard the model and start all over again?

This is a good book for learning or teaching an important, practically useful topic. There are many good examples and example-based discussions.

Jayanta K. Ghosh: *ghosh@stat.purdue.edu*
 Department of Statistics, Purdue University
 West Lafayette, IN 47909, USA

Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning

Alan Julian Izenman

Springer, 2008, xxvi + 734 pages, £ 47.99 / € 59.95 / US\$ 89.95, hardcover

ISBN: 978-0-387-78188-4

Table of contents

- | | |
|----------------------------------------------------------|--------------------------------------------------------------|
| 1. Introduction and preview | 10. Artificial neural networks |
| 2. Data and databases | 11. Support vector machines |
| 3. Random vector and matrices | 12. Cluster analysis |
| 4. Nonparametric density estimation | 13. Multidimensional scaling and distance geometry |
| 5. Model assessment and selection in multiple regression | 14. Committee machines |
| 6. Multivariate regression | 15. Latent variable models and blind source separation |
| 7. Linear dimensionality reduction | 16. Nonlinear dimensionality reduction and manifold learning |
| 8. Linear discriminant analysis | 17. Correspondence analysis |
| 9. Recursive partitioning and tree-based methods | |

Readership: Advanced undergraduate students, graduate students, and researchers in statistics, computer science, artificial intelligence, psychology, cognitive sciences, business, medicine, bioinformatics, and engineering.

This massive opus is not a textbook to go through in one course – unless the course is a pretty long one! Izenman has an ambitious task to cover multivariate techniques from traditional multivariate methods, such as multiple regression, principal components, canonical variates, linear discriminant analysis, factor analysis, clustering, multidimensional scaling, and correspondence analysis, to the newer methods of density estimation, projection pursuit, neural networks, multivariate reduced-rank regression, nonlinear manifold learning, bagging, boosting, random forests, independent component analysis, support vector machines, and classification and regression trees – this long list is all from the back cover of the book. Being not a master in all these areas I do not wish to give final judgement how well Izenman succeeds in all these matters but that much as I have been browsing the book, I think he has done excellent job.

Izenman states in the Preface that it is hoped that this book will be enjoyed by those who wish to understand the current state of multivariate statistical analysis in an age of high-speed computation and large data sets. I can easily believe that persons interested in learning new trends of multivariate methods would find Izenman's book very helpful. Yet the classical methods are

handled in the book as well, but often derived as special cases of a common theoretical framework. Being a matrix-man, I paid attention to his matrix representation which I found very pleasant. The full-color graphics is quite impressive – well done! There are numerous real-data examples from many scientific disciplines so that not only statisticians may find this book useful and interesting.

Simo Puntanen: *simo.puntanen@uta.fi*
Department of Mathematics and Statistics
FI-33014 University of Tampere, Finland