Reviews of Books and Teaching Materials


To link to this article: http://dx.doi.org/10.1080/00031305.2015.1115663

Published online: 29 Dec 2015.

Submit your article to this journal

Article views: 217

View related articles

View Crossmark data
Adaptive Design Theory and Implementation Using SAS and R (2nd ed.).

The revision of the 2008 edition of this text is substantially expanded and reorganized, adding approximately 246 pages of material. The second edition has 29 chapters, after adding 12 new chapters and removing one. The book is largely reorganized in many chapters, especially in Chapters 8, 11, 14, 19, 20 and 22, where new materials such as analytical methods, sequential parallel designs with rerandomization, the skeleton approach, etc. are added. The book has corrected some issues in the first edition, for instance, the indistinguishable citations of papers with the same authors and publication year, unnecessary repetition of equations and some typographical mistakes. As in the first edition, the book provides rich and invaluable examples (adding about 22 new ones) within each chapter. Throughout the book, about 20 new SAS macros and R functions for the design and simulation of adaptive trials are added. The examples and computer programs are great aids for the reader to understand the discussed theory and to implement them in practice. Each chapter ends with a set of exercises but the practice sections remain somewhat sparse with only about three to four problems in each chapter.

Huixia WANG
George Washington University

Applied Predictive Modeling.

With the growing emphasis on data mining and now data analytics, there has been a proliferation both of statistical modeling techniques and software to implement them. The book Applied Predictive Modeling, is geared toward an audience interested in learning about the many predictive methods available, and aims to provide this audience the tools needed to implement these methods in R. In the preface, the authors introduce their work as a “practitioner’s guide to the predictive modeling process and a place where one can come to learn about the approach.” Their motivation is to “describe the techniques in a way that helps develop intuition for its strengths and weaknesses instead of its mathematical genesis and underpinnings. For the most part we avoid complex equations.” The authors do exactly as advertised, providing a text that will appeal to a broad audience regardless of academic preparation. The lack of equations and technical details, while welcome to those keen on the background, general application, and benefits and cautions, may be too mysterious for the technically curious. To the authors’ credit, they point to other texts, including Hastie, Tibshirani, and Friedman (2009), for a more rigorous treatment.

This book covers a multitude of techniques appropriate for various data types. It provides a thorough treatment of each method, how it compares to its competitors, and how it is applied in R, while embedding other examples and case studies between major topic headings. For the practitioner, this book is not just a welcome addition to the library, but more like a handbook that enables those only vaguely familiar with some of these techniques to roll up their sleeves and dig in and try out each method in short order. For those not familiar with R, the authors provide an R introduction in the appendix.

The many predictive modeling approaches are cataloged and described across 20 chapters. The first part (Chapters 1–4) actually prepares the practitioner for data analysis by covering the modeling process, data pre-processing (e.g. data transformations, removing or adding predictors), the importance of training and test subsets, and how generally to tune the parameters of modeling strategies by resampling or data splitting. Part II chapters (5–10) deal with models for continuous responses, starting with linear regression and “its cousins” (partial least squares, and penalized or ridge regression), followed by the many nonlinear regression models including neural nets, multivariate adaptive regression splines (MARS), support vector machines, and k-nearest neighbors. Other nonlinear regression models discussed include regression tree variants (bagging, random forests, boosting), and rule-based methods, including cubist, and are all covered in a separate chapter. Part III (Chapters 11–17) addresses classification models for categorical responses. Among the models discussed are: 1) linear - logistic regression, discriminant analysis, and penalized models, 2) nonlinear - nonlinear discriminant, neural nets, support vector machines, k-nearest neighbor and naive Bayes, and 3) tree methods. Chapter 14 returns to the classification varieties of the tree methods described in Part II. The authors offer an extensive collection of modeling techniques that span the top current alternatives well. The only method possibly missing from that list is Gaussian stochastic process, or kriging models. The final set of chapters (Part IV, Chapters 18–20) cover some important modeling issues, some of which I would prefer to have received more emphasis and to have appeared earlier in the book. Although the authors clearly state in the preface and elsewhere that the objective is to present models that predict well, in many instances we are at least as interested in interpreting the fitted model. This topic is addressed in two of the final three chapters, which describe measuring predictor importance, feature (or variable) selection, and finally, some general modeling advice, such as looking out for Type III errors and when to trust in your predictions.

The bulk of each of the methods chapters is devoted to describing the methods and contrasting them with their closest competitors, with a concluding section giving directions on computing. Example datasets from R packages are presented, along with the necessary R code and selected output. This organization and the flow of the book is perfect for those who want to gain experience by implementing these methods right away. As a casual R user, I found this aspect of the book quite useful and was happy to gain valuable skills without much toil.

The authors provide case study chapters at the end of each of the book’s parts. These help to reinforce the reader’s understanding of the models and illustrate the relative strengths of each. Chapter exercises are also provided and there is a website containing links to related information.

I expect this book will be a welcome addition for anyone interested in learning about the many and varied methods for predictive modeling. The authors do an admirable job not only in presenting each model in a digestible fashion, but also of framing the techniques with chapters devoted to advice on good statistical practice. These chapters will coach novices and fairly seasoned practitioners in terms of the work required up front, the need to try multiple modeling approaches, and the advantages and disadvantages of model complexity. As an example, the authors advise the analyst to solicit subject matter expertise as a part of the modeling process, by stating that “the best, most predictive models are fundamentally influenced by a modeler with expert knowledge and context of the problem. This expert knowledge should first be applied in obtaining relevant data for the desired research objectives” (pg. 5). This book, possibly with a more technical companion text, should suit the needs of the aspiring predictive modeler well.

Jim SIMPSON
JK Analytics

REFERENCE
Hastie, T., Tibshirani, R., and Friedman, J. (2009), The Elements of Statistical Learning (2nd ed.), New York: Springer.

Basic Statistics and Pharmaceutical Statistical Applications (3rd ed.).

The third edition of Basic Statistics and Pharmaceutical Statistical Applications presents an excellent introduction to statistics for undergraduate and
graduate students in pharmacy, individuals desiring a brief introduction to the field of statistics, and those in need of a quick reference for statistical problem solving. It is a handbook, a guide, and a reference for non-statisticians and researchers in need of understanding basic statistical concepts and conducting simple statistical analysis. It does not deal with theoretical details; rather, it serves as a quick and practical tool for the understanding of the most commonly employed statistical concepts and methods. Similar to the first two editions, the book is divided into three major sections: (1) the underpinnings required to understand inferential statistical tests; (2) inferential statistics to help in problem solving; and (3) supportive materials in the form of flow charts and tables. Each chapter in this edition has been reviewed and edited to clarify or expand on information previously discussed. Virtually every chapter has some new information either in the form of additional paragraphs or entirely new sections. A major addition is the inclusion on information on Excel® and Minitab® for performing statistical analysis.

Chapters 1 through 8 primarily focus on basic statistical and probabilistic concepts, including sampling approaches, communication of data analysis, visual display of graphs and summary statistics, various probability distributions, sampling distributions and confidence intervals, and hypothesis testing and types of errors. Chapters 9–12 are concerned with settings with discrete independent variables and continuous outcomes. Chapters 13, 14 and 20 discuss tests where both the dependent and independent variables are continuous. Chapter 15 through 19 continue the presentation of tests concerned with discrete independent variables, but in these chapters the dependent variables are also discrete. Chapter 21 provides nonparametric or distribution-free statistics for evaluating data that does not meet the criteria required for many of the tests presented in chapters 9 through 14. Statistical tests for equivalence are described in chapter 22. Outliers are often a concern in data analysis and chapter 23 discusses various ways to identify them. Some frequently encountered statistical errors in the literature are detailed in chapter 24. In addition, the book contains four appendices: flow charts for selection of appropriate inferential tests in Appendix A, some commonly used statistical tables in Appendix B, a summary of commands for Excel® and Minitab® in Appendix C, and answers to example problems from each chapter in Appendix D.

The book covers an enormous number of topics in statistics, most at an introductory level. It does not deal with theoretical derivation of most of the formulas presented in the book. This is consistent with the author’s intended audience of non-statisticians and researchers in need of understanding basic statistical concepts and conducting simple statistical analysis. Overall, this book covers statistical topics most relevant to those in the pharmaceutical industry and pharmacy practice. It focuses on the fundamentals required to understand descriptive and inferential statistics for problem solving. The inclusion of how to use Minitab’s and Excel’s “Data Analysis” add-in feature to perform statistical tests is very helpful. This book would serve as an excellent introduction to statistics for undergraduate and graduate students whose major are not statistics or as a reference book for nonstatisticians in biomedical research.

Weili He
Merck Sharp & Dohme Corporation

Data Analysis with MPlus.

In Data Analysis with MPlus, Geiser presents a comprehensive introduction to statistical modeling using the MPlus statistical package. Given the rising popularity of MPlus, this provides a timely contribution that will aid both the beginning researcher and the seasoned investigator. The utility of this book is in no small part due to the analyses and techniques that Geiser chose to focus on that are in increasing demand in social science research. The six chapters in this book take the reader from the very basics of data management through longitudinal analyses and basic multi-level and mixture modeling.

This book has a distinctly practical focus with the first two chapters dedicated to the issues of preparation and management of data for use in MPlus. Chapter one focuses on preparing data using SPSS for importing into the MPlus framework, with attention given to the appropriate management of missing values. Chapter two describes the process of importing data into MPlus and setting up the basic syntax. Here Geiser demonstrates how MPlus can use summary data including covariance and correlation matrices with means and standard deviations.

In chapter three the reader is introduced to structural equation modeling (SEM). This chapter is designed with the beginner in mind, starting with simple linear regression models using manifest variables before moving onto the more complex topics and examples. Once the reader is made familiar with simple linear regression models, with manifest variables, treated as a special case of more complex SEM’s, Geiser introduces modeling latent regression models in MPlus, including a discussion of model specification as well as various model fit indices. An overview of confirmatory factor analyses (CFA) follows, including a step by step example of a first order and a second order CFA. Formal path modeling and mediation analyses are covered in the final section of chapter three. Here again, Geiser begins with path modeling in MPlus using manifest variables before moving onto latent path models. The reader is introduced to the MPlus functions for modeling indirect effects allowing for tests of mediation. Overall, chapter three covers the foundational material allowing the reader to engage in basic SEM modeling. It is structured in such a way that the beginning reader with little experience or theoretical knowledge will be able to follow and understand the material, and will be able to work through the detailed examples Geiser provides to get practical experience in SEM modeling.

Chapter four of the book moves beyond the typical scope of most introductory SEM books and describes approaches to modeling longitudinal data. The focus here is on latent state (LS) and latent state trait (LST) models. Perhaps the most interesting aspect of this chapter is the practically focused discussion of measurement invariance in the context of longitudinal research. This chapter also includes a tutorial on testing for the different levels of measurement invariance in MPlus. With the increasing interest and popularity of longitudinal research and longitudinal modeling approaches, the issue of measurement invariance is becoming a central concern. As such, it is very encouraging to see an accessible discussion of these issues in a beginners SEM modeling book. Another highlight of this chapter is the discussion of latent growth curve (LGC) modeling. Geiser introduces the reader to both first order and second order LGC models and describes the steps of associated graphical analyses offered in Mplus.

Multilevel modeling and the handling of data with a nested structure is the topic of chapter five. Here Geiser briefly discusses the theoretical properties of multilevel models and the necessity of appropriately modeling clustered data structures by referring to accessible practical examples. As earlier in the book, the chapter introduces the reader to multilevel modeling in MPlus through examples of simple, intercept only models, before finally moving onto the more complex, random intercept and slope models. The example syntax and outputs provide the reader with clear demonstrations of how various multilevel models are specified using the MPlus commands, including the modeling of effects at different levels of a two-level model.

Finally, chapter six deals with mixture modeling, specifically focusing on latent class analysis (LCA). The reader is taken through an example of an LCA with three latent classes. The author includes a discussion of model fit, an issue of particular importance and some dispute in the literature. An overview of the most important criteria for evaluating LCA model fit is provided in a very clear and methodical manner.

In its final pages the book offers useful information in the appendices, including a glossary of the MPlus commands used throughout the book. Also included is, a list of common mistakes encountered MPlus programming and a steps for troubleshooting the problems that may arise.

Overall, Data Analysis in MPlus showcases the possibilities and the accessibility of MPlus as versatile statistical modeling software. The book provides a good introduction to SEM modeling with MPlus that is accessible to both beginner and advanced investigators. One of the greatest advantages of the book is the extremely thorough presentation of the practical steps in running various analyses in MPlus, with detailed discussion of syntax and output. The high level of detail provides an excellent complement to the many syntax files already available for MPlus. The analyses Geiser chose to focus on are an interesting and a refreshing selection for a text like this. The first three chapters provide a necessary introduction to the practical issues of working with MPlus, as well as to the basic SEM modeling, while the remaining chapters provide a working overview of much more advanced analyses, all of which are very topical and growing in popularity. The consistent tips and practical advice Geiser provides
Throughout round off this book as an impressive package that will undoubtedly prove very useful to the beginner as well as more advanced researchers.

Petar MILOEV 
Massey University
Chris G. SIBLEY
University of Auckland

Financial and Actuarial Statistics: An Introduction (2nd ed.).

This is an excellent book written for students in financial and actuarial statistics; it would fit nicely into a one-semester course. A somewhat rapid presentation allows the authors to cover numerous topics. A number of changes have been made in the second edition. For example, lots of new exercises have been added along with excel solutions and built-in excel functions. Chapter 1 of the first edition, an introduction to probability and statistics, is now expanded and divided into two chapters: new examples, discussion of nonnegative variables, and a section on model diagnostics have all been added. The topics of option pricing and stochastic surplus models have been expanded (now in Chapter 2) and new discussion on the probability of ruin for standard surplus models is included. The new Chapter 7 includes multiple future lifetime and multiple decrement models. Markov chains and their applications to actuarial sciences appear in a new chapter (Chapter 8). Chapter 9, on scenario testing, is reorganized into deterministic and stochastic cases. Simulation techniques have been extended and nonparametric prediction intervals have been added to the new edition.

Chapter 1 focuses on introductory topics in probability and statistics including random variables, expectations, moment generating functions, and some well known distributions. Chapter 2 continues with more topics in probability and statistics presenting the readers with a variety of topics such as sampling distributions, confidence intervals and confidence sets, CLT, aggregate variables, distributions, and regression. Although there is no need for a prior exposure to these topics, having some background would be helpful to understand a rapid presentation of these topics.

The main focus of the book begins with Chapter 3 where financial and actuarial models are presented in this and subsequent chapters. In Chapter 3, the reader is introduced to topics such as the fixed financial rate model, fixed rate annuities, and stochastic rate models. Chapter 4 focuses on deterministic status models and presents loss and risk models. Aggregate status models in which the status may contain stochastic components appear next. Chapter 5 begins with the continuous and discrete future lifetime models, which look at survival and the force of mortality from financial and actuarial perspectives. Other topics include survivorship groups, life model and life tables, and life table confidence sets and predication intervals.

Chapter 6 is the longest chapter in the book and returns to the stochastic rate model presented in Chapter 3. Future economic decisions are important in financial and actuarial sciences and several related topics are presented in this chapter, including stochastic present value functions, risk evolution and calculations (discrete, continuous, and mixed), life insurance (continuous and discrete), loss functions, and reserves (continuous and discrete). The (annual) time period of discrete actuarial status models, first introduced in Chapter 3, is extended here to allow for sub-intervals such as monthly time periods. The chapter ends with a discussion on expense models and computations where execution and administration of financial contracts are presented.

Chapter 7 considers further aspects of the stochastic status model, including time decrement such as pension and retirements. Multiple future lifetimes including joint life status, last survival, and status are presented. Continuous and discrete multiple decrement models and single and multiple decrement life tables are also introduced.

Contributions are concerned with the applications of discrete Markov chains. Life table data are revisited using both time-homogeneous and time-heterogeneous Markov chains. Single and multiple decrement chains are also considered. Topics covered in Chapter 5 are revisited using discrete Markov chains. Membership groups covered in Chapter 5 are divided into subgroups (strata). The authors nicely apply the discrete Markov chains to financial and actuarial fields.

Chapter 9 begins with scenario testing for deterministic and stochastic models. A clear understanding of the previous chapters is essential in order to follow the topics covered here. Demonstration of financial viability under various economic scenarios due to factors such as variations in interest rates is discussed. Bootstrap and simulation sampling techniques are presented and used to estimate sampling distribution for financial and actuarial statistics. Models presented throughout the book are considered using simulation techniques in Chapter 9. Bootstrap and simulation sampling are deployed to generate samples for testing different financial and actuarial models. The book presents interesting approaches in modeling financial and actuarial topics via simulation.

Overall, the book is clearly written and provides a well organized and effective structure. Each chapter starts with a description of its overall objective, topics that will be covered, and what the reader can expect to learn. After introducing the topics, the author summarizes what the reader should have learned and accomplished. The book also contains substantial examples and exercises, giving the reader an opportunity to practice modeling and graphing using R. This is presented in such a way as to be useful for those with little experience with R or computer programming.

Chapter 1 describes growth curve analysis using R, a free software environment for statistical computing and graphics. Methodologically it uses multilevel regression models, which are popular in longitudinal data analysis because they allow researchers to explore individual differences over time. The book does not devote much space to statistical theory; instead it briefly introduces what the reader needs to know, while keeping notation to a level that is easily understandable. A such, it is particularly appropriate for applied researchers and data analysts in the fields of behavioral sciences such as cognitive science, neuroscience and psychology.

Overall, this book would make a great choice for a course in finance and actuarial statistics and is also a good reference for researchers in the field. The strength of the book is the presentation of the topics via example and the problems and solutions included at the end of each chapter. A background at the level of calculus should be sufficient for the readers to follow the topics covered in the book. Readers with no prior exposure to probability and statistics should also be able to handle the book.

Growth Curve Analysis and Visualization Using R.

This book is designed to be a practical guide for implementing growth curve analysis methods using R, a free software environment for statistical computing and graphics. Methodologically it uses multilevel regression models, which are popular in longitudinal data analysis because they allow researchers to explore individual differences over time. The book does not devote much space to statistical theory; instead it briefly introduces what the reader needs to know, while keeping notation to a level that is easily understandable. A such, it is particularly appropriate for applied researchers and data analysts in the fields of behavioral sciences such as cognitive science, neuroscience and psychology.

Overall, this book would make a great choice for a course in finance and actuarial statistics and is also a good reference for researchers in the field. The strength of the book is the presentation of the topics via example and the problems and solutions included at the end of each chapter. A background at the level of calculus should be sufficient for the readers to follow the topics covered in the book. Readers with no prior exposure to probability and statistics should also be able to handle the book.

The American Statistician, November 2015, Vol. 69, No. 4 427
tical predictors that have more than two levels. Chapter 6 considers the use of logistic growth curve analysis for binary outcome variables and also introduces the generalized form of the multilevel regression model in R using the \texttt{glmer} function in the \texttt{lme4} package. This chapter, however, does not describe the problems that can arise with ordinal or count outcome variables.

Chapter 7 discusses two methods for quantifying individual differences: treating individual measures as fixed effects in the multilevel model, and using random effects to model individual effects. Chapter 8 presents seven complete examples with detailed R scripts, demonstrating, for example, how to plot the observed data, fit the appropriate growth curve models, plot the model fits, and report the results. Since the examples stand alone, this chapter can serve as a handy reference for R programming. Although the book covers a wide range of issues related to growth curve analyses, there is no discussion of missing data, which frequently occurs in repeated measurement studies.

In summary, this book provides a practical introduction to the visualization and modeling of repeatedly measured data as it arises in the behavioral sciences. It illustrates growth curve analysis with various examples using R. It will serve as a good reference for applied researchers who are interested in learning how to apply and implement the presented statistical methodologies using R. Snijders and Bosker (2012) is a suitable alternative for readers that are interested in the theoretical underpinnings of multilevel regression.

Hyunkeun Cho
Western Michigan University

REFERENCE


Handbook of Spatial Point-Pattern Analysis in Ecology.
Thorst\textsc{e}n Wie\textsc{g}and and Kirk A. Moloney. CRC Press, 2013, xxvii+510 pp., \$83.95(H), ISBN: 978-1-42-008254-8.

The objective of the book is to elucidate the framework of statistical point-pattern analysis for practicing ecologists. It is refreshing to read as its presentation differs from theory-based, standard statistical textbooks. It carefully describes the cycle of statistical analysis beginning with forming a question, assembling point-pattern data, determining data type, selecting summary statistics, and deciding on appropriate null models to assess ecological hypotheses. The authors refer to statistical tools by names more often than with specific notation. Subtle issues of applying the tools are addressed in a practical, ecological context as well as in both theoretical and statistical realms. I am excited about the potential of the book since many scientists, including statisticians, who conduct point-pattern analysis could learn not only the framework of statistical analysis but also the details of a variety of statistical tools.

The book is comprised of five chapters, with the first and last being very brief. Chapter 1 opens with the motivation for point-pattern analysis in testing ecological theories. Chapter 2 lays the groundwork of statistical analysis and categorizes point-pattern data by the type of information provided and by the number of characteristics represented per location (or sampling unit). The authors inform readers that when one can correctly classify a point-pattern dataset, then one is able to set an ecologically relevant baseline theory and adopt a statistically viable null model to determine whether the observed phenomenon deviates far from the baseline theory or not.

Chapter 3 introduces several statistical measures for each data type, as titled “Estimators and Toolbox”. Its preliminary focus is on the interpretation of mathematical and statistical concepts. Since a single measure cannot describe a complex point pattern, the book suggests using several measures to compare patterns. Their properties are detailed along with graphical illustrations, and often a parallel is drawn to a related measure in a number of worked-out analyses.

Chapter 4 is the highlight of the book, as it explores pattern-generating mechanisms that align well with certain scenarios of observed point-pattern and test a simple hypothesis. I would have titled Chapter 4 “Testing models of a point process” instead of its printed title of “Examples”. It presents a Poisson process, Thomas process, Cox process, and Gibbs process, and each of them with some twists to incorporate inhomogeneity. It emphasizes assessing a dataset by considering the spatial scale of the data organization, verifying underlying assumptions, and providing summary measures to evaluate various aspects of a complex dataset.

Chapter 5 provides the summary of the book and suggests using it as a graduate-level ecology textbook over a 60-hour time period; an outline of a syllabus is provided for a 6-hour/day, 10 day-long program (pg. 444). The book includes a compilation of mathematical symbols with section references and a fully descriptive glossary.

I note two shortcomings the authors could easily remedy. The first involves simulation envelopes, that is, the expected range of a statistical measure under a baseline scenario. The idea is reiterated several times, but little guidance is provided on how to generate a simulation envelope for statistical measures. This may not be straightforward because it requires specifying a null scenario and data-generating process model then simulating from it. This sequence may be less familiar to those who are new to the framework of statistical inference. I wished the book added a few examples or pseudo-code of generating a simulation envelope, which would parallel the detailed descriptions for the use of statistical measures.

Second, the book is missing a low-level guide for analyzing point-pattern data. It may have presumed that the computing requires separate attention because software for spatial point-pattern analysis is still in development. Wiegand introduced the \texttt{Programita} software in the preface, but there is no mention of using it in the rest of the text. Users can request to download the software free of charge via http://www.programita.org (last updated in February of 2014) and learn more about it here: http://www.thorst\textsc{e}n-wiegand.de/howe/programita.html. Note that the software is currently compatible only with a Windows operating system. I did not find the interface to be straightforward since I could not immediately display a dataset upon loading and because users must recognize notational symbols to apply statistical tools. If you are familiar with R, the alternative is to explore \texttt{apopstat} (Baddeley and Turner 2005), \texttt{PtProcess} (Harte 2010), or \texttt{stpp} (Gabriel, Rowlingson, and Diggle 2013) package.

The book should be helpful not only to graduate students but also to working professionals in ecology and spatial statistics. It is not intended to be browsed through from cover to cover; it largely reads as a compendium of statistical methods. The authors repeatedly mention the importance of data quality and the method of data collection since describing the effective spatial scale of a phenomenon is crucial. It includes ample references in ecology but mainly focuses on how to test for basic patterns according to a context-driven hypothesis and provides a variety of statistical measures to analyze the data. It uses mathematical notation as necessary to formalize methods and add clarity. It is rare to encounter such a statistically thorough and user-friendly graduate level text. I highly praise the meticulous presentation of the materials and the examples. If you are more interested in a mathematical presentation of point processes, you should refer to Lieshout (2000), Møller and Waagepetersen (2004), Diggle (2013), and Chiu et al. (2013).

Ecologists and field biologists often ask the question of where a phenomenon of interest is taking place, in what formation, and why. With an increased effort to restore local flora and fauna, understanding the density of its population over a region, learning its dispersal pattern, and discovering the operative spatial distance of interaction within a species is important to intervene and adjust the spatial density of the observed species. For other field scientists dealing with point-pattern data, their questions should boil down to understanding the features of a point pattern essentially in the same fashion as ecologists. Hence, the book can also serve researchers in astronomy, neuroscience, social science, epidemiology, and geography wherever point-pattern data analysis is necessary.

Exploring new topics is easy through this book. This book provides an extensive, comprehensible, and accessible account of when to use a specific statistical tool. The authors devote considerable attention to the interpretation and implication of each tool and the assessment of the model fit. Various data types and the properties of statistical tools are illustrated in graphs, and the comparisons between related statistical methods are demonstrated via simulation and accompanied by text. To enhance insights, I would love to see some exercises added at the end of each chapter. This would enhance the learning experience of the readers and could be used to shed light on the limiting conditions of statistical tools. This book lays out the basic framework of statistical analysis of spatial point-pattern data, and there are plenty of statistical and ecological insights to glean.

Eunice J. Kim
Amherst College
REFERENCES


Introduction to Probability Models (11th ed.).

This is the 11th edition of Ross’s popular Introduction to Probability Models. This text is organized so that the first two chapters give a whirlwind overview of basic probability theory. Chapter 3 focuses on conditional expectations and probability, while Chapter 4 provides an introduction to Markov chains. Chapter 5 provides more detailed coverage of the exponential distribution and introduces Poisson processes. Markov chains are revisited in Chapter 6, which covers Markov chains in continuous time, with an emphasis on birth and death models. The topics of Chapters 7 and 8 are renewal theory and queuing theory, respectively. Reliability theory is the subject of Chapter 9. The book is rounded out with chapters on Brownian motion (Chapter 10) and simulation (Chapter 11).

As in the previous editions, the strengths of this text are the clear writing style and the vast number of examples of varying difficulty (all with detailed solutions). In addition to numerous new examples and exercises throughout the text, the new material in this edition includes a revised section on long run proportions and limiting probabilities of Markov chains (Section 4.4), new material on Hawkes processes (Section 5.5), a new section on reversed chains in continuous time Markov chains (Section 6.7), and a new section on the maximum of Brownian motion with drift (Section 10.5).

In the preface, the author indicates that the approach taken in this text is to help the student develop “an intuitive feel for the subject [probability theory] that enables him or her to ‘think probabilistically’ ” (p. xi). While the clear exposition facilitates this approach, I think the text slightly misses the mark due to the lack of graphical illustrations to supplement the written explanations—in the first two chapters, the only basic probability concepts for which there are graphical illustrations are the CDF of a discrete random variable and the normal density curve. Thus, in my opinion, this text is best suited for individuals who have had some previous exposure to basic probability theory. Indeed, a quick Google search suggests that this text is a common choice for introductory courses in stochastic processes or as an extension to a first course in basic probability theory.

Jessica L. CHAPMAN
St. Lawrence University

Introduction to Robust Estimation & Hypothesis Testing (3rd ed.).

In his third edition, to Introduction to Robust Estimation & Hypothesis Testing, the author presents a multitude of robust methods for standard inference problems including: one- and two-sample location problems; multivariate problems; one-way and higher order ANOVA type problems; correlation and tests of independence; and simple and multiple regression situations. For most of these problems, the author discusses several robust methods, comparing them with each other and their traditional least squares (LS) counterparts. The comparisons are usually based on results for various examples and small simulation studies. Besides estimation, inference is stressed throughout the book. These inference procedures use estimates of asymptotic variances of estimators and tests and, also, bootstrap procedures (percentile type and others). Bootstrap-based inference is discussed throughout the book. The author has made available to the reader a library of R functions to compute most of the estimation and inference procedures that he discusses, including the bootstrap procedures. I applaud the author’s scope on methods and his use of the bootstrap.

This third edition of the book has the same chapter titles as the second edition. Many of the procedures that are discussed in the second edition have been retained in the third edition, while several new procedures have been added in the third edition. In the second edition, S-PLUS and R were discussed but in the third edition only R code is discussed.

I do have a few reservations with the text; in general, the author does not clearly state the necessary assumptions for consistency of the procedures under discussion. Throughout the text, the author worries about heteroscedasticity of scale, but the text contains little on robust procedures for fitting heteroscedasticity. Such procedures (testing and estimation) exist and, further, unlike the traditional $F$- and $\chi^2$ procedures, they possess robustness of validity over a wide range of distributions. One such procedure is the Fligner-Killeen procedure. Hettmansperger and McKean (2011), Section 2.10, discussed the correct version of the Fligner-Killeen procedure for testing and estimation of scale effects. Their simulation studies confirm the validity of the procedure. Software for this procedure can be downloaded at CRAN; see Sections 3.3 and 5.7 of Kloke and McKean (2014).

Another form of estimator that is used throughout the author’s book is to first remove the “outliers” and then compute a selected procedure on the remaining data. This procedure, though, generally does not possess robustness of validity. Hopefully the author considers these reservations in his next edition.

Joseph W. MCKEAN
Western Michigan University

REFERENCES


Multistate Analysis of Life Histories with R.

Willekens offers a detailed, concise, and well-written description of the Biograph package in Multistate Analysis of Life Histories with R. He opens with a brief introduction to multistate models, providing an overview of life history data and the conceptualization of them as observations of stochastic processes. Specific topics covered in this overview are driven by the methods implemented in Biograph. The package estimates transition rates using two different approaches: the Nelson-Aalen estimator, and a semi-parametric alternative commonly used in demography where the transition rates are assumed to be piecewise constant by age (i.e., occurrence-exposure rates). Willekens outlines these approaches and briefly describes the nonparametric estimation of the cumulative hazard function and parametric approaches involving maximum likelihood estimation of transition rates under exponential and piecewise exponential models, and estimation of transition probabilities and state occupation times. He then describes the process of generating synthetic life histories, or expected life histories, from the transition rates estimated based on the sample. Bootstrapping and longitudinal microsimulation are combined to obtain the expected life trajectory based on the estimates and to integrate the uncertainty in these estimates.
Chapter 3 is devoted to a description of a Biograph object and instructions on how to arrange data to ensure it is consistent with the package formatting requirements. The next two chapters are dedicated to the descriptive statistics that can be produced in Biograph and some of the data visualizations that can be created using the R packages lattice, ggplot2 and Epi. Chapter 6 consists of a survey of some of the other R packages that can be used to conduct more advanced modelling and estimation (i.e., survival, eha, mstate, mva, em and msm). A variety of semi-parametric and parametric models are fit with and without covariates to highlight the capabilities of these packages and to provide examples of output. Biograph is used to convert data to formats that are required by these other packages. Chapter 7 shows how multistate life tables can be computed to produce useful statistics (e.g., transition probabilities, state probabilities, expected state occupancy and sojourn times) using Biograph and related packages and Chapter 8 describes a detailed application illustrating use of the packages and functions presented in the book.

A major strength of the book is the rich set of applications based on real data that were used to illustrate the use of R functions and their resulting output throughout. Appealing colour graphics facilitate interpretation of the results. The main dataset, the Netherlands Family and Fertility Survey of 1998, is analyzed in Chapter 8 to demonstrate a multistate data analyses using Biograph and related packages. Other datasets used include a subsample from the German Life History Survey, the National Family Health Survey of India, the Survey of Health, Ageing and Retirement in Europe and the European Registry of Blood and Marrow Transplantation. Several of these datasets are available directly within R packages, while others are available for download online, so readers are able to recreate the output and experiment with the code as they follow along. Another strength is that Willekens emphasizes the importance of exploratory data analyses and visualizations to summarize and explore the data. He devotes two of his nine chapters to this.

In his Preface, Willekens acknowledges that the purpose of the book is to motivate more use of survival and event history analyses in social sciences. I applaud this. So many applications can be expressed as transitions between states. The challenge is packaging the methods in such a way that they are accessible to those who do not already have a related foundation, while at the same time, preventing misinterpretation and inappropriate use. He also identifies this book as useful for a graduate-level course in event history analysis, but assumes that readers already have some basic knowledge about survival and multistate analyses.

If you use or are thinking about using the Biograph package to analyze life history data, this book is definitely worth considering as a resource for yourself and/or your graduate students. The detailed documentation about the features of Biograph, as well as the worked examples, would be quite useful to someone learning the Biograph package. However, if you are not using Biograph, this book may be of limited value to you. It does not provide a comprehensive treatment of how to conduct multistate analyses in R using other packages. Only one of the nine chapters is devoted to a survey of the analyses that can be conducted using the packages other than Biograph.

Multistate Analysis of Life Histories with R is not sufficient as an introductory book in multistate data analyses. Students who are new to multistate analyses will need additional resources to complement this book. Novices will require more instruction on the fundamentals to understand the methods and models. Specifically, they would benefit from more instruction on the choice of appropriate summaries and models, as well as the underlying assumptions and process of model selection in a variety of event history analysis situations.

Bethany J.G. White
University of Western Ontario

Noninferiority Testing in Clinical Trials: Issues and Challenges.

The book Noninferiority Testing in Clinical Trials: Issues and Challenges provides a comprehensive overview and discussions on many important aspects of noninferiority (NI) testing. It is written in a way that is easily understood and at a level suitable for both statisticians and scientists in drug and biologics development. The author of this book is an expert on this topic and has developed many concepts, statistical methods, and considerations for using NI trials. The book provides some historical overview with respect to the development of ideas on setting NI margin, estimating the effect size of active control from historical trials, and considerations on a few critical components for NI trials including assay sensitivity, constancy assumption, discounting, preservation, analysis population, simultaneous superiority testing, and three-arm trials.

The book has 12 chapters. Chapter 1 introduces concepts for the NI hypothesis, including some interesting historical development about the formulation of NI and choice of margin. Chapter 2 to Chapter 4 discuss the choice of NI margin, a main issue for conducting NI trials, for different endpoints from a continuous mean difference, mean ratio, hazard ratio for survival endpoints, to proportion difference, odds ratio, or risk ratio for a binary endpoint. Simple numerical examples and graphical illustrations are used to convey ideas.

Chapter 5 introduces the fixed margin and synthesis methods. Terminologies for establishing the NI margin such as preservation, discounting, double-discounting, and constancy assumption are explained with simple examples. Chapter 6 provides an interesting discussion about simultaneously testing NI and superiority. Some controversial view points on potential logic flaw and increase of false discovery rate are discussed. Chapter 7 covers how to use meta-analysis to estimate treatment effect and determine NI margin when there are multiple historical trials. Chapter 8 introduces the gold-standard 3-arm trial for NI, which includes test treatment, active control, and placebo in the study to allow assessment of assay sensitivity and within-trial validation of the choice for NI margin.

A brief summary of regulatory guidance documents for NI trials is presented in Chapter 9, which covers the guidance issued by the European Medicines Agency (EMA) and the US FDA draft guidance on NI trials. Chapter 10 discusses the populations used for analysis of clinical trials for both superiority and NI testing, including: intention-to-treat (ITT), per-protocol (PP), and as-treated (AT). Some regulatory perspectives and recommendations are presented based on guidance documents. A useful summary about handling missing data based on recent published guidance documents are provided in the appendix. Chapter 11 provides a real example of constructing the NI margin for testing thrombolytic therapies in treatment of acute myocardial infarction. The last chapter summarizes the issues discussed throughout the book, as well as some controversial topics and challenges that may require future research.

In summary, the book provides concise discussions about several critical issues in conducting NI trials. It not only covers basic concepts such as assay sensitivity, constancy assumption, discounting, and preservation in constructing NI margin, but also shares many historical developments of the ideas presented with citations from key publications. Each chapter contains reference articles for readers to get more in-depth information as needed. In comparison to another book by Rothmann, Wiens, and Chan (2012), this book is more focused on the issues and regulatory considerations for NI testing without getting into too much of the statistical and technical details (e.g., analysis methods or sample size calculations for designing NI trials). This would be a very good introductory or reference book for anyone, statistician or non-statistician, who is involved in NI trials to capture the issues and regulatory considerations.

G. Frank Liu
Merck & Co. Inc.

REFERENCE

Risk Assessment and Decision Analysis with Bayesian Networks.

Risk Assessment and Decision Analysis with Bayesian Networks serves as an introduction to probabilistic models based on directed acyclic graphs, with an emphasis on applications in modern decision-making. While there exist numerous specialized books in Bayesian networks written for scholars and graduate students with high levels of training in mathematics, statistics, and
computing, there is little written on Bayesian networks which is accessible to
the less proficient reader. This book expertly fills this gap in the literature, even
going so far as to carefully describe and motivate the necessary mathematical
and probabilistic underpinnings.

The authors aptly motivate the text with a series of important cases of both
individual and organizational-decision-making including cases in finance and
government. Both require high levels of decision making and risk assessment in
setting where decisions are often made with a very minimal amount of historical
data, and rely heavily on expert (subjective) information.

Chapter one begins with the financial crisis of 2008, pointing out that armies
of analysts and statisticians failed to predict either the event or its scale. The
authors examines how the methods employed in this book could have helped
predict the crisis, by improving analyst’s ability to empirically leverage expert
opinions. The authors then discuss a series of case studies from important policy
issues in medicine, law, and other areas. This chapter also reviews famous
statistical misconceptions such as Simpson’s paradox and introduces several
themes which recur in later chapters.

Chapter two covers classical limitations of traditional statistical approaches
and demonstrates how one can employ causal or explanatory models for risk
assessment. Again this chapter focuses on motivating examples such as in
the section “Are you More Likely to Die in an Automobile Crash When the
Weather Is Good Compared to Bad?” Here, the authors discuss misconceptions
of correlation and causation in detail.

Chapters three and four review classical (frequentist) and Bayesian ap-
proaches to statistics, with a careful introduction to basic probability theory.
The authors assume very little prior training in formal mathematics or statistics
and thus review everything from sigma notation to the chain rule in probability
theory. Chapter five covers Bayes’ Theorem and conditional probability, and
begins with the assertion that all probabilities are conditional (e.g. in a specific
context).

Chapter six is where the meat of the book is found as the preliminaries
are (mostly) finished. It provides the bridge from Bayes’ Theorem to Bayesian
Networks, and aggressively uses visualization techniques based on a software
package developed by the authors. (My only major complaint about the book
pertains to the software: I would have preferred to see it made available in an
open source environment.) The authors provide a formal definition of Bayesian
Networks, keeping the text light with a series of simple examples centered
around whether “Norman” is late given that there is a train strike. Chapter seven
extends Chapter six, offering guidance in building the graphical component
of a Bayesian Network model. Chapter eight provides the necessary information
for fully specify the Bayesian Network model. Chapter nine takes what has thus
far been discrete nodes and discrete probability distributions for building the
Bayesian Networks, and extends them to the continuous case. It is motivated
by a highly engaging problem of predicting automobile costs in the context of
hybrid cars.

Chapter 10 is a bit incongruous in that it attempts to engage the reader in the
use of Bayesian Networks for classical hypothesis testing – which the authors
rall against repeatedly. In a generally well-paced textbook, this chapter largely
feels out of place.

Chapter 11 returns to the task at hand, modeling risk. Specifically, it tar-
gets operational risk which covers holistic risk assessment for organizations
looking to quantify both internal and external unpredictability to their opera-
tions. Chapter 12 extends this idea to modeling system reliability in a Bayesian
Network framework. Chapter 13 concludes the book with an example of us-
ing Bayesian methods in legal case studies, a topic first raised in chapter one.

A strength of this text is its holistic approach, allowing a reader or instruc-
tor to explain a complex topic to an individual with only basic mathematical
training. This is especially useful for popularizing these important ideas and
concepts through the larger social science, financial, and policy fields. Ideally,
this text provides a good background for practitioners of decision-making and
risk assessment. I believe this will be great resource for both self-study and
undergraduate instruction, with the only major weakness being the reliance on
specialized, non-open source software.

Zack W. ALMQQUIST
University of Minnesota

Stated Preference Methods Using R.
Hideo AIZAKI, Tomoaki NAKATANI, and Kazuo SATO. Boca Raton, FL: CRC

The stated preference (SP) methods discussed in this book are based on the
random utility theory (Thurstone, 1927; McFadden, 1974) that is used to derive
families of statistical models to represent people’s preferences for and tradeoffs
among attributes (features) of choice alternatives. Typically, the choice alter-
 natives are arrayed into choice sets using some type of statistical experimental
design procedure, and participants in the experiment choose one or more of
the alternatives in each choice set, depending on the type of choice task. A
long-lived and appealing underlying theoretical foundation and considerable
flexibility has made these models popular both with academics and practition-
ers in many disciplines, including transportation research, applied economics
(e.g., agricultural, environmental, food, health, etc.), finance, political science
and marketing, to name just a few.

Stated Preference Methods Using R is targeted at beginners, and the ob-
jective is to demonstrate how to design and implement SP methods using the
open source software R. Simultaneously integrating SP theory and experimental
design with practical implementation in R leads to significant benefits for ki-
naesthetic learning. The authors mostly accomplish their objective by avoiding
too much theory and detail in order not to overburden beginners who want to
learn how to apply the methods. A focus on R is particularly attractive and useful
for this target group because costs are negligible (except, of course, learning)
and the open source code enables users to understand theory and calculation pro-
cesses better, providing an in-depth learning experience. Indeed, the widespread
availability of R and the many resources associated with it make it likely to be
an important source for statistical applications for many years into the future.

The text is divided into five chapters and two appendices. The first chapter
briefly introduces the simplest variants of the three methods covered in this book,
namely (i) single-bounded and double-bounded dichotomous choice Contingent
Valuation, (ii) Discrete Choice Experiments based on orthogonal main-effect
designs, and (iii) object-case Best-Worst-Scaling, based on orthogonal main-
effects designs or balanced incomplete block designs. The book also briefly
summarizes random utility theory, discrete choice models and willingness-to-
pay (WTP) calculations for analysing data resulting from the three methods.
The summary is short and moves quickly, so beginners interested in applying
and understanding the three types of SP methods are likely to need one or more
other textbooks. The authors partly compensate for this by referring readers to
more detailed literature, but important gaps remain such as preference or scale
heterogeneity (e.g. Louviere and Eagle 2006) and estimating models in WTP
space.

Chapters 2, 3, and 4 discuss implementation of Contingent Valuation, Dis-
crete Choice Experiments and Best-Worst Scaling in R using packages partly
developed by the authors. The chapters walk readers nicely through different
steps in the SP methods, including the experimental design of surveys and the
mathematical foundations of preference elicitation. The authors explain in de-
tail how to implement these steps in R and illustrate the preference evaluation
for each method using at least two datasets (some synthetic and some real). A
common weakness in these chapters relates to experimental design construc-
tion (generation); the discussion often is somewhat vague and/or fails to offer
readers guidelines (or at least references) about how to proceed in nonstandard
situations (e.g., when a balanced-incomplete block design does not exist for
their problem). Chapter 5 concludes the book by discussing basic operations in
R. This is a nice feature because it enables readers to start using R immediately,
obviating a need to do extensive research into the use of R beforehand.

Two appendices complement the book and provide a guide for readers who
wish to go further in their understanding and application: the first appendix
briefly discusses other packages in R that can be used for SP evaluation, and
the second appendix offers several examples of what the authors call “contrivances
in empirical research”.

Discrete choice models are described by a number of stand-alone books. These
include an overview of Stated Choice Methods (aka “Discrete Choice
Experiments” or DCEs; e.g., Louviere, Hensher and Swait 2000), an overview
of the Design of Experiments for DCEs (e.g., Street and Burgess 2007), dis-
susions of theory and associated econometric methods for choice modeling (e.g.,
Hensher, Rose and Green 2005), a detailed overview of Contingent Valuation

The American Statistician, November 2015, Vol. 69, No. 4 431

Downloaded by [North Carolina State University] at 08:29 05 February 2016
(CV) theory and methods (e.g., Carson, 2012), along with an in press introduction to theory and applications of Best-Worst Scaling (e.g., Louviere, Marley and Flynn, 2015, forthcoming). Apart from these books, there are hundreds, if not thousands of academic papers and book chapters dealing with various aspects of this rich and expanding field.

No existing book, chapter, or paper, however, provides a comprehensive introduction to the major aspects of SP modeling utilizing widely available, noncommercial software in the manner provided by this book. A text that can be used in university classes and shortcourses without requiring students to buy or use commercial software is a welcome addition. Having taught many such courses over the years, a frustrating aspect of teaching is deciding which of several commercial software packages to use. This decision requires tradeoffs between the need to learn how to use the software in a short period of time, its availability and cost after students leave the course, its flexibility to accommodate various course objectives, and its ability to handle future applications. To date these tradeoffs have not been satisfactorily addressed with commercial software, although some packages like Stata come close (e.g., Hole 2013).

This is potentially an important book because it puts almost everything a beginner needs in one place and on one statistical programming platform. One can overlook some clunky aspects of the text, such as the approach used to produce experimental designs and associated discrete choice surveys. Today most researchers in the field have access to web-based survey software, and this typically allows one to produce fairly elaborate surveys that include converting experimental designs directly into choice sets, blocking the surveys, embedding multi-media information in choice sets or surveys, and many other features. Future updates of the book might take such “best practice” technology into account.

In conclusion, this text provides an overview at an introductory level of three types of SP methods (CV, DCEs and BW) and how to design and implement them in R. It provides in-depth examples from several real and synthetic agricultural and environmental economics datasets, and provides several free R packages written by the authors. The book is not designed to be used as a textbook for a SP course, but could be used as a text (possibly supplemental) for courses where researchers wish to apply these methods without delving too deeply into the underlying theory. The book is an excellent resource for beginners in both SP methods and R.

Jordan J. LOUVIERE
University of South Australia
Christine ECKERT
University of Technology, Sydney
Spring SAMSON
University of Queensland

REFERENCES


Statistical Methods in Biology: Design and Analysis of Experiments and Regression.


In writing Statistical Methods in Biology: Design and Analysis of Experiments and Regression, Welham and colleagues have brought a new child into the already crowded family of books describing (or instructing) the use of statistical techniques in biology. The abundance of such works is due in part to the breadth of both fields. What type of statistics? What type of biological question?

Statistical Methods in Biology deals primarily with the use of ANOVA in experimental studies and with regression. The first two chapters describe some foundational concepts in linear models and in statistics more generally. The authors describe these chapters as either introductory or a review but the pace of the writing is probably more appropriate for someone with some (perhaps rusty) background in such concepts rather than for the complete novice. The third chapter covers common forms of experimental design. Here, the pace at which new concepts are introduced slows, seemingly an indication that the authors feel they have completed the review of what they expect readers to understand a priori and are entering the heart of their topic. The next three chapters describe the conceptual basis and use of one factor ANOVA, methods of checking the requisite assumptions, and the use of transformed data, respectively. They complete the groundwork for understanding the discussion of the use and limitations of linear models in common biological contexts which is to follow.

Chapters 7 through 11 deal with specific types of experimental design and the appropriate ANOVA models for each. Topics included are complete randomized block designs (Chapter 7); factorial and multi-stratum (nested) designs (Chapter 8), Latin square, split-plot and balanced incomplete block designs (Chapter 9), power analysis (Chapter 10), and fractional and unbalanced factorial designs (Chapter 11).

The remainder of the book includes seven chapters about linear regression and associated extensions and a final chapter with advice about designing experiments (Chapter 19). Specifically, Chapters 12 and 13 cover simple linear regression and its assumptions. They are followed by chapters about multiple regression (Chapter 14), ANCOVA (Chapter 15), mixed models (Chapter 16), polynomial and nonlinear regression (Chapter 17), and GLMs (Chapter 18). Of course, the last three of these topics are very broad in scope so an individual chapter devoted to each really serves as an introduction rather than a full treatment.

Thus, the answer to the question of what type of statistics includes a suite of very common techniques based mostly on linear models. As for the biology, Welham et al. include example datasets throughout, mostly drawn from the fields of agriculture and entomology, but readers learning from this text should have little difficulty in applying their new knowledge to other sub-fields of biology. Overall, the content this book greatly overlaps with many others on the market already. Nevertheless, I see value in this retelling of the story using linear models in biology. The reason why a sporting goods store offers many types of running shoe is that individual runners have diverse needs—their shoes must feel comfortable on their feet and must be appropriate for their intended distances, anticipated conditions, etc. So also, a wide variety of biology professionals, students and academics will read about statistical techniques for diverse reasons and with diverse goals in mind.

A reader’s comfort with books that teach statistics to non-statisticians will be heavily influenced by level of mathematical sophistication demanded of the reader. A given reader will easily be lost and discouraged by a text that demands too much mathematical background but she or he can also be frustrated by long winded explanations of concepts that could be represented and understood by a simple equation. The preface states that the book is intended for students in last undergraduate year through PhD or for investigators who wish to analyze their own data or collaborate effectively with analysts. Within this audience, no single book will satisfy all but Welham et al. seem to have the level of math calibrated
correctly for many. For example, the book does a thorough job explaining most
equations it presents and uses calculus in equations very rarely. Most of the
readers who understand Chapter 2 will have explained to them much of the
notational conventions used later in the book.

Just as runners require different shoes for sprinting, long distance track
running, and cross country, so the diversity of statistical methods books is
motivated by the type of skills and understanding that they offer. At one end of
the spectrum “cookbooks” describe how to accomplish a particular task using
a particular software (e.g., Teteer 2011) without allowing readers to gain any
understanding of the underlying conceptual basis (or demanding that they do).
At the other end of the spectrum concepts can be presented with little emphasis
on implementation (e.g., Sokal and Rolf 2011). Here, Welham et al. are much
closer to the latter approach. The real strength of their book is its ability to
incorporate a relatively deep or nuanced treatment of the material without
resorting to highly technical descriptions. For example, they introduce and
explain basic components of statistical models very early in Chapter 1 which
allows them to introduce the concept of a one factor ANOVA by describing
the model equation first. They make use of the model equation throughout the
chapter and finish with a description of alternative forms of the same model that
readers may encounter from other sources. This is in refreshing contrast to some
presentations that first introduce the idea of statistical models only after explain-
ing some other key concepts (e.g., Quinn and Keough 2002). I also found their
treatment of some individual topics such as Type I, II and III sums of squares
(Chapter 11) clearer and more precise than similar discussions in many other
texts.

Welham et al. do make some efforts to help readers with implementing the
techniques they cover. For example, in textbook style, each chapter contains
worked examples within the text and a set of approximately five to ten exercises
at the end. Nevertheless, readers who have data in hand and wish to learn how
to analyze it will likely be disappointed. The preface accompanies a companion
website (www.stats4biol.info) and promises all of the example datasets will be
posted as well as programs in GenStat, R and SAS for each text example and
each end-chapter exercise. As of the time of this writing, neither the data files
or the GenStat and R code for the text examples were available. (SAS code
was only available up to Chapter 7.) The code and solutions for the end-chapter
exercises was only available up to Chapter 6. The code that was there was
relatively sparsely commented and would be entirely insufficient to someone
who did not already have a basic working knowledge of the corresponding
programming language.

Overall, I believe that the book will be successful in the sense that a successful
running shoe is comfortable for many but not all runners. In the preface the
authors write that it is their hope that the book will provide readers with an
appreciation of the importance of statistics and experimental design, a sound
understanding of the techniques covered, knowledge of how to use one or more
statistical packages, and the ability to correctly interpret results. Based on my
experience, readers who have even a limited background in statistics or higher
math should find the book helpful in all these respects, except perhaps in learning
how to use statistical software. Compared to other books on statistics in biology
that I have read, this book most reminds me of Quinn and Keough’s Experimental
Design and Analysis for Biologists (2002) in terms of its coverage of topics and
level of sophistication in language, math, and presentation. Relative to Quinn
and Keough, Welham and colleagues’ new work is less heavily referenced, more
attractive in layout, and more clearly written in some places. In particular, I am
hopeful that some readers will find that the equations are presented with more
supporting text and are thus more understandable.

Ben Rubin
University of Western Ontario

REFERENCES

Quinn, G. P. and Keough, M. J. (2002), Experimental Design and Data Analysis
for Biologists, Cambridge, UK: Cambridge University Press.

Statistical Studies of Income, Poverty and Inequality in Europe:
Computing and Graphics in R Using EU-SILC.
Nicholas T. LONGFORD. Boca Raton, FL: CRC Press, 2014, xiiii + 354 pp.,

This book is a tour de force giving a thorough and readable description of what
is and what can be done with the European Union Statistics on Income and Living
Conditions (EU-SILC). To anyone who enjoys exploring the complete picture,
it will be a pleasure to read. For those who wish to gain a deeper understanding
of parts of the analysis, the book will also be very satisfying. It introduces
the concept of poverty and explains that its definition revolves around measures
of household income and household size. Nothing is as simple as you might
think and we are immediately engaged in considering what should be the level
of relative resources required for each additional adult member of a household
and for each child (defined as being aged 14 or under). The usual levels are
0.5 for each additional adult and 0.3 for each child. Naturally, changing these
levels will lead to changes in the resulting equivalized household size (eHS) and
we encounter our first equation that is eHS = 1 + c_a(H_a - 1) + c_c(H_c - 1) where
c_a and c_c are the levels for adult and child and H_a and H_c are the numbers of
adults and children in the household.

Dividing the equivalized household size into the household income gives the
 equivalized income (eHI), which is the fundamental variable analyzed in the
book. Thereafter the book deals with every conceivable analysis that can be

carried out with yearly eHI values for 29 EU countries. Section 1.2 tells us that
“An individual or a household is classified as poor if its value of eHI is below 60%+
of the median national eHI.” From this definition all sorts of analyses arise
and most include varying the threshold around 60% and presenting threshold-
linked poverty rates as a function of the threshold.

All terms in the subsequent chapters are thoughtfully appraised and so the
book takes the novice through a process of revelation that is intellectually
stimulating and rewarding. No doubt the experienced user can skip through
the early parts of the book but nevertheless a book that deals step by step with
the fundamentals is very reassuring even for those to whom some or all of the
subject matter is familiar.

The book is well written and virtually error free so that I was pleased with
myself that I found a typo on page 69 where infiltration is misspelt as “in-
filtration”! There are lots of diagrams, everything is justified and explained;
everything you would hope from a superb teacher. One problem is that because
of the number of diagrams they occasionally get out of sync with the text.

Compared with other thorough, learned books, this has a superbly flowing
style and takes the reader forward in methodical logical steps. There are many
tables and graphics and at the end of each chapter all the R code used to carry
out the analysis and to prepare the graphics is given in an appendix and the
reader is invited to replicate everything for themselves. The R code is clear and
all practicalities are addressed in the text, for example, on page 106, it says that
the deviations (in Figure 4.5, p. 109) “are generated as a random sample from
a centered normal distribution with a small standard deviation (set to 0.006 in the
diagram).”

One shortcoming is that the author chooses to focus on helping the reader to
understand the analytical challenges and statistical methods rather than actually
analyzing the latest EU-SILC data. This is reasonable but because the book
is so well written the reader would like to see the methods applied in their
entirety to all the data so that the experience of each country can be compared
and the data can be completely pulled apart and scrutinized. This is clearly not
the aim of the book. Although some of the analysis is applied to data from all
the EU countries, often it is just applied to a subset. For clarity and to compare
alternative methods, the author often focuses on data from Austria between 2004
and 2010. The reader may find themselves wanting to see the results for their
own country in 2015 but they will need to do this for themselves. In this way, the
reader is invited to apply the techniques themselves to make a more complete
investigation. This is one of the reasons that comprehensive R code is included.

The author reports extensively on the results of the analyses and their graph-
ical representation. He also comments on the implications of some of the analy-
yses, for example, chapter 8 deals with social transfers in which we compare
the real (R) world with an alternative (A) world in which social transfers such as
child support and disability entitlements are disregarded as a source of income.
The chapter proceeds to identify individuals who are poor or rich in either or
both of these R and A worlds and then to study the conditional distribution of

The American Statistician, November 2015, Vol. 69, No. 4 433

Downloaded by [North Carolina State University] at 08:29 05 February 2016
the social transfers given the value of eHI in the A world scaled by the median eHI. The final section makes observations about the perils of indices and says somewhat oxymoronically "So long as the deviation of the index from the target is inexplicable, the index can be used with confidence." However, the main thrust of this section is to note the problem of people being tempted to fine tune indices to give a more "desirable" result.

The book is fairly complex and although all the concepts, terms, and methods are explained, the reader will benefit from a knowledge of statistics at least at an undergraduate level. Although parts of the book could be omitted, to be able to understand the whole book, the reader should be comfortable with descriptive statistics, distributions, transformations, summation, some linear algebra, and some calculus. The book deals with a niche area and is probably too statistically complex for a purely qualitative course in poverty studies. For an undergraduate or graduate statistics module, however, the book would make an excellent textbook from which to teach; although there are no end of chapter questions as such, there are often suggestions for further reading and the R code in each chapter’s appendix is extensive often filling nearly as many pages as the chapter itself.

The author makes some profound comments, for example, on page 226 in his chapter on multivariate mixture longitudinal models of log(eHI) for panel data, Longford prefers to keep options open on a variety of models and says that to do otherwise is a "denial of uncertainty." Another of the author’s nice descriptions refers to mixtures of improper and proper distributions. He notes that some of the implausible features of his longitudinal models “become less prominent in the proper components when an improper component is included in the model. This suggests that the improper component fulfills its role of recruiting outliers or, more generally, households with exceptional profiles.”

The final chapter 9 studies cause and effect in observational data. As an example, the author considers the effect on eHI of postsecondary education as a cause. The subtleties are thoroughly dealt with leading to some complicated sentences. For example, within the setting of "good advice" being given to individuals about the benefits of education, “the comparison of the outcomes of two groups would be meaningless, because the realized outcomes of those who did not study are a poor replacement for the would-be outcomes of those who did study, had they not done so.” The estimated effect of postsecondary education is found to be positive for all countries in all years. It would be interesting to have the author’s views on the implications of the size of the effect. For example, in 2011, UK, Austria, Germany and Estonia have smaller values and Poland, Hungary, France, and Spain have higher values. Finally, unexplored questions are formerly gathered into an Epilog that invites the reader to consider further analysis such as question 16 "Is the current economic crisis reflected in EU-SILC data?"

The book is definitely worth buying if you have an interest in or need to study statistical methods in this subject area; besides being a good read, it acts as a step by step textbook for learning and using a wide range of statistical techniques. The area is specific but nevertheless fundamental and the skills learnt are transferable to other areas of analysis. The book is unusual in that it is academic but in a nonconfrontational way; the author clearly knows his field inside out and relishes all its minutiae; his enthusiasm is contagious. It is amazing what depth and breadth of analysis can be done with a set of household sizes and incomes for a number of years from a number of countries!

Shirley Y. Coleman
Newcastle University