

Analyzing Receiver Operating Characteristic Curves With SAS.

Mithat GÖNEN. Cary, NC: SAS Institute, Inc., 2007, x+134 pp., \$31.95 (P), ISBN: 978-1-59994-298-8.

Mithat Gönen's new book describes statistical methods involving receiver operating characteristic (ROC) curves and their implementation in SAS. In this brief volume, Dr. Gönen gives a comprehensive treatment of the basic ideas and methodology of ROC curves. Theoretical aspects of the subject are neglected in favor of practical matters. The book requires no familiarity with ROC curves, medical diagnostics, or the assessment of predictive accuracy, but instead introduces these topics from an elementary level. Very little statistical sophistication is required of the reader, although the more advanced methods that Dr. Gönen covers do involve such topics as probit regression, nonlinear mixed models, generalized estimating equations, and the Cox regression model, so some familiarity with these techniques is helpful.

After introducing basic ideas briefly in Chapters 1 and 2, ROC curves for a single continuous predictor are discussed in Chapter 3. Covariate adjustment and comparison of multiple ROC curves in paired designs is the focus of Chapter 4, while ROC curves for ordinal predictors are discussed in Chapter 5. Having discussed the binomial model for ROC curves in earlier chapters, the author then introduces an alternative family of ROC curves in Chapter 6 which he terms the Lehmann family. His description of methods for Lehmann ROC curves and their implementation using Cox regression is based on as yet unpublished work (Gönen and Heller 2007). ROC methods for censored data are discussed in Chapter 7, while Chapters 8 and 9 discuss the use of ROC curves to assess the predictive accuracy of multivariable models such as logistic multiple regression models (Chap. 8) and more modern models available in SAS Enterprise Miner such as decision trees and neural networks (Chap. 9).

Throughout the text, Dr. Gönen illustrates the methods with examples drawn not only from the medical diagnostic field where ROC curves are most well known, but also from fields such as weather forecasting and credit scoring. Of course, implementation of the methods in SAS is central to the book, and SAS code and output is presented and available on the publisher's website. A wide variety of SAS procedures are used to implement the methods, but the author also makes available several macros to perform computations not possible with standard PROCs.

The book is well-written and easy to follow, and being unfamiliar with ROC methods, I learned much from it. For SAS users who have an interest in statistical methods related to ROC curves and their implementation, this is a handy and useful book which seems well worth its modest price.

Daniel B. HALL
University of Georgia

REFERENCES

Gönen, M., and Heller, G. (2007), "Lehmann Family of ROC Curves," submitted. Available online at <http://www.bepress.com/mskccbiostat/paper11/>.

Introduction to Design of Experiments (3rd ed.).

Jacques GOUPY and Lee CREIGHTON. Cary, NC: SAS Institute, Inc., 2007, xiv+448 pp., \$69.95 (P), ISBN: 978-1-59994-422-7.

This is the first English edition of the book. The previous two editions authored by Goupy were published only in French. Goupy and Creighton wrote the third edition of the book in French, translated it, and added new material.

The book covers response surface designs, mixture designs, D-optimal designs, and designs for discrete variables, in addition to the material on full and fractional designs covered in the first two editions. It caters to professionals who routinely perform experiments, providing them with basic tools to conduct well-organized experiments that can lead to useful insights.

The book is well written and very easy to read. It includes 20 examples, most of which are presented as small case studies with an introduction that describes the particular study, one or more sections on preparing, designing and carrying out the experiment, and finally, one or more sections that interpret the results and state the conclusions of the study. The examples are done in JMP (SAS Institute Inc.). The data and tables are available at the SAS Press companion website of the book. The book does not intend to serve as a JMP manual to experimental designers—rather, it is an introductory book on experimental design

with many examples—and so, readers that have never used JMP before would need to get accustomed with the software before working with the examples.

Raquel PRADO
University of California, Santa Cruz

Medical Statistics: A Textbook for the Health Sciences (4th ed.).

Michael J. CAMPBELL, David MACHIN, and Stephen J. WALTERS. Chichester, West Sussex, U.K.: Wiley, 2007, xii+331 pp., \$37.50 (H), ISBN: 978-0-470-02519-2.

This venerable text is now in its fourth edition in 17 years. A new co-author, Stephen J. Walters, has been added and the book has seen significant changes since the third edition was published in 1999. The book still aims to explain "medical statistics with as little technical detail, so as to make the textbook as accessible to a wide audience," and the authors certainly achieve this. It is intended for both *consumers* of statistics—those whose primary need of statistical knowledge is for critical evaluation of published research, and for *doers* of statistics—those who need to know how to conduct basic statistical analyses. The first half of the text is targeted at the consumers, while the remainder is aimed at the doers.

The new edition is much longer (more than 100 pages) than the previous edition. Several new chapters have been added: "Survival analysis," "Reliability and method comparison studies," and "Sample size issues." Each chapter is now laid out so that more mathematical details are given their own section ("Technical details") rather than being relegated to an appendix. Most of the examples are motivated by journal articles published in the last five years, and some generic computer output is included. These two improvements lend the book a fresher, more modern feel. The chapter of multiple choice exercises found in the previous edition has been replaced with a variety of in-chapter exercises with detailed solutions provided at the end of the text. Finally, the new edition has a more logical layout of chapter topics that students will unknowingly appreciate.

Given these substantial changes, I agree with the authors' that this fourth edition is "essentially a new textbook." Outside of the fact that the text is now quite a bit longer, I think that all of the changes are for the better. This an excellent text for teaching an introductory course in biostatistics and for those interested in self-study.

J. Wade DAVIS
University of Missouri

SAS Programming for Enterprise Guide Users.

Neil CONSTABLE. Cary, NC: SAS Institute Inc., 2007, xiv+282 pp., \$49.95 (P), ISBN: 978-1-59994-540-8.

When I agreed to review this book, I was not aware that the author assumes a "good working knowledge of SAS Enterprise guide" and that the author recommends *The Little SAS Book for Enterprise Guide 4.1* by Slaughter and Delwiche as a prerequisite. Although I am a user of Enterprise Guide 4.1, it became clear that in order to fully appreciate the usefulness of this book the user needs to have not only a strong background in Enterprise Guide but also a good knowledge of SAS Proc SQL.

The author uses the "<double click to insert code>" option frequently as a means to add additional SAS commands to the code generated by Enterprise Guide. However, when I gave a talk and used that technique a number of the SAS people suggested that it is easier to set up the WYSIWYG, run it, and then open Last Code Submitted and modify that.

With the proper background, this book will be extremely useful in SQL, formats, reports, graphics and the Output Delivery System (ODS).

F. Michael SPEED
Texas A&M University

Statistics (11th ed.).

James T. MCCLAVE and Terry SINCICH. Upper Saddle River, NJ: Pearson Education, Inc, 2009, 835 pp., \$129.33 (H), ISBN: 0-13-206951-2.

The sixth edition of this book was given a positive review in *TAS* by Katherine Halvorsen back in 1995, and earlier editions have also been praised by reviewers in *Technometrics* and *The Statistician*. By the time a book reaches its 11th edition (advanced middle age?), changes from the previous edition can be expected to be very modest, and that is certainly true here. The authors have

revised/updated about 30% of the 1200+ exercises, many of which are based on real and interesting data. There are some new case studies, and a CD that accompanies the text contains demonstration applets. Chapter 14 on nonparametrics is no longer in the printed text, but instead appears on CD and online. If the authors continue to delete one chapter per new edition, I guess the 24th edition will be distributed in pill form.

Here are a few miscellaneous observations: It is nice to see plausibility checks for normality assumptions included in examples. The definition of a p -value is gratifyingly correct and complete, unlike sloppy statements I have seen recently in other introductory books. I do wish, however, that a more complete t table had been included to allow for statements like p -value ≈ 0.078 rather than just $0.05 < p$ -value < 0.10 . The authors still use the standard normal table with areas from 0 to z rather than a cumulative table. I didn't find "bootstrap" in the index, but then noticed a biographical sketch of Brad Efron which mentioned it in passing. But Efron is not listed in the index, although Fisher, Gossett, Neyman, and the two Pearsons, among others, also get bios and index citations. Some of the references are over the top for the intended audience (e.g., Scheffé's ANOVA book), and some have not been updated to reflect new editions.

This book has stood the test of time and captured a sizable audience. It is a very solid choice for a general introductory course with no calculus prerequisite. The inclusion of much real data and actual problem scenarios should provide students with convincing evidence for the relevance of the subject to their professional and general interests.

Jay DEVORE
California Polytechnic State University

Unlocking the Census with GIS.

Alan PETERS and Heather McDONALD. Redlands, CA: ESRI Press, 2004, xviii+309 pp., \$44.95 (P), ISBN: 1-58948-113-5.

Unlocking the Census with GIS is a clear, basic guide to the U.S. census and related data sources, with an emphasis on georeferenced data and visualization. It is a nice reference text for any scientist working with census data, and it is particularly useful for graduate students or researchers using census data for the first time. Limited prior knowledge of, or capability with, GIS is required, extending the book's utility as a reference text for those mystified about how to obtain geospatial data and use them for mapping.

The first chapter is an introduction to the census, including a summary of questions asked, a reproduction of the short-form questionnaire, a discussion of the census geography hierarchy and FIPS codes, and a comprehensive summary of census data, tabulations, and reports. Chapter 2 provides several step-by-step procedures for data acquisition, with an emphasis on obtaining digital boundary files for map display of geographic units. A complete and well-written description of the TIGER/Line files and topology is provided. Each of the next four chapters focuses on a particular dataset or collection of variables. In Chapter 3, the focus is on demographic and social variables (e.g., age, race, education) and includes a description of the Public Use Microdata Sample that can be used to enhance census tabulations. Chapter 4 focuses on the analysis of economic conditions and the major economic variables from the Census of Population and Housing, including income, poverty status, and employment and labor force status. The latter part of Chapter 4 focuses on labor force data, including data from the Current Population Survey and North American Industry Classification System. It also gives an overview of data available from the Bureau of Labor Statistics, including the National Compensation Survey and the Quarterly Census of Employment and Wages. I found this chapter to be extremely comprehensive, and I believe it will be useful to readers who are unfamiliar with these data sources. Chapter 5 covers housing and community development data and includes descriptions of housing units, as well as housing costs and financing. It also provides a description of other sources of housing data (e.g., data available from the Department of Housing and Urban Development). Chapter 6 is a unique and interesting chapter on the analysis of transportation issues, based on such variables as place of work, usual means of transportation, and time of departure for work. This chapter provides a fairly thorough description of the Census Transportation Planning Package specifically designed for transportation analysis. The final chapter, Chapter 7, is a short chapter containing ideas on how to distribute maps in a cost-effective and friendly way. This chapter also includes practical tips on Web-based mapping.

The real strength of the book is its comprehensive treatment of data from the U.S. Census Bureau and related data from other federal agencies that can be useful in answering questions related to population, transportation, and labor economics. A discussion of the Area Resource File produced by the U.S. Department of Health and Human Services would be a nice addition that would allow the authors to illustrate data on, for example, health facilities, health professions, and health status, as well as spatial linkage to other data discussed in

the text. No discussion is included of ZIP Code Tabulation Areas, relatively new geographic areas developed to overcome the difficulties in precisely defining the land area covered by each ZIP Code. The main use of GIS in the text is simply visualization/mapping; other basic functions such as spatial query and buffering are not illustrated, yet these functions are part of the foundation of any GIS. The authors note in Chapter 7 that "designing clear, understandable, and attractive maps is an art," and they defer discussion of these skills to other texts. However, map construction and design also have a scientific foundation, and I believe any book on mapping or GIS should start with a basic overview of the principles of map visualization and display.

In summary, *Unlocking the Census with GIS* provides a comprehensive description of census data. The authors believe that understanding the precise meaning of the variables is key to using the data effectively, and I agree completely. This book will help anyone make more effective use of all the data the U.S. Census Bureau provides and increase their awareness of other data sources that can be useful in any geographic analysis.

Carol A. GOTWAY CRAWFORD
Centers for Disease Control and Prevention

Workshop Statistics: Discovery With Data (3rd ed.).

Allan J. ROSSMAN and Beth L. CHANCE. Emeryville, CA: Key College Publishing, 2008, xxxv+660 pp., \$69.95 (P+CD), ISBN: 978-1-59757-113-5.

In the Preface, the authors provide the following quote from George Cobb: "Shorn of all subtlety and led naked out of the protective fold of educational research literature, there comes a sheepish little fact: lectures don't work nearly as well as many of us would like to think." With this teaching philosophy in mind, this edition retains the hallmark features of previous editions of *Workshop Statistics*, providing a resource for teaching an introductory statistics course that supports active learning and conceptual understanding through the analysis of real data and the use of technology. New features in this edition include increased student support with more expository material, examples, definitions, and hints; new and updated datasets and new activities; authentic studies addressing specific, relevant, and engaging research questions; a new sequence of topics emphasizing statistical thinking, data collection, and scope of conclusions; and links between activities and interactive Java applets. Supplements include print and online instructor resources and an AP guide for secondary instructors.

The book is organized into seven units, each consisting of several topics. Each topic begins with an Overview section, which presents a motivating example and introduces the topic objectives. A Preliminaries section engages students with study contexts and statistical issues in the topic. In-Class Activities guide students through a series of directed questions within the context of an authentic study and using real data. Space for written activity answers is provided within the book's perforated, three-hole punched pages. Self-Check Examples cover the key ideas of the activity and includes detailed solutions. A Wrap-Up section provides a review of major ideas studied in the topic. Homework Activities test students' conceptual understanding and ability to apply what they have learned about the topic.

The units cover collecting data and drawing conclusions (variables, distributions, studies, random sampling, and experiments); summarizing data (two-way tables, measures of center and spread, and displaying distributions); randomness in data (probability, normal distributions, and sampling distributions); inferences for one mean and one proportion; inferences for paired data and for comparing two means as well as two proportions; inferences with categorical data; and regression and correlation.

Included are more than 275 activities based upon media examples that cover a diverse range of topics (e.g., entertainment, sports, current events, health, and medicine) and more than 20 that use student-generated data. Roughly one-half of the activities require the use of a software package or a graphing calculator. The book comes with a Student CD, which contains applets and data files in various formats for all of the activities. The data files and applets can also be accessed from a Web-based Student Resource Center (which contains selected solutions to In-Class Activities).

This book is an ideal fit for an introductory statistical literacy course. The varied nature of the topics covered by the activities should provide virtually every student with real-life applications of statistics that they will find relevant and interesting.

Mary RICHARDSON
Grand Valley State University