Although results of real-world studies are incorporated into several homework exercises, the chapters did not contain any case studies or larger investigations. Few homework exercises emphasize interpretation of results.

Finally, the Excel exercises at the end of the chapters that expose students to computer software are not very insightful, and some are poorly designed. For example, in Chapter 7, the following exercise in Excel was used to demonstrate an application of the central limit theorem (CLT): 200 samples of size 4 are randomly selected from a population of repeated sequences of numbers 20 through 29. The means for each sample are computed, and then the histogram of the sample means is constructed. The student is asked to repeat the exercise for samples of size six and eight. This is a poor illustration of the CLT for many reasons, not the least of which is that the sample sizes are extremely small! In addition, the author does not explain the purpose of the exercise.

Overall, *UBS* has some strengths, but falls short in the end. Perhaps instructors interested in an inexpensive alternative to higher-priced textbooks could use *UBS* if they were to supplement the textbook chapters with additional information, more insightful and challenging exercises, and motivating examples.

Jeffrey C. SKLAR
California Polytechnic State University

Handbook of Statistical Distributions with Applications.

K. KRISHNAMOORTHY. Boca Raton, FL: Chapman & Hall/CRC, 2006. ISBN 1-58488-635-8. 346 pp. $79.95 (H+CD).

Handbook of Statistical Distributions with Applications is a reference text for practitioners who are familiar with statistical methodology but who are not experts in statistics. This text is fairly successful for persons with such a background, which means that it may not be of high interest to many readers of *The American Statistician* (*TAS*).

This reference book discusses, as the title implies, many of the most common parametric distributions as well as some less common distributions and distributions of test statistics used in nonparametric inference. The book is divided into 36 chapters. The first chapter is a brief introduction to the theory of statistics. The next 28 chapters each cover one parametric distribution, and then five chapters cover topics in nonparametrics. Finally, the last two chapters cover tolerance factors for a multivariate normal population and the multiple correlation coefficient.

Chapter 1 attempts to give background for the rest of the text, which requires a primer on statistical methodology. It includes diverse topics such as point estimation, hypothesis testing, confidence intervals, Q-Q plots, the chi-square goodness of fit test, and random number generation, all in just 27 pages. It is, of course, not possible to comprehensively and accurately discuss anything but the basics in such a compressed space. In fact, most *TAS* readers could probably find some quibbles with this exposition. For example, the author notes that, "For a discrete distribution, median (sic) is not well defined, and it need not be unique" (p. 12). This is correct, but a novice statistical user could use a little more help on this topic. For testing, the author notes that, "A level α test rejects H_0 whenever the p value is less than or equal to α" (p. 22). You may recall debates on this topic within the pages of this journal (e.g., Donahue 2000; Hubbard and Bayarri 2003), even though applying this rule will not result in problems for the vast majority of situations encountered by the intended audience. These problems are minor, though, and should not greatly detract from the usefulness for the target audience.

The 28 chapters that cover various parametric distributions are generally fine. The lengths of the chapters range from two pages for the discrete uniform distribution to 36 pages for the normal distribution, length indicative of the relative importance of the material presented. Most chapters contain references to papers that have used the distributional assumptions discussed. The writing is brief, almost terse, but due to space limitations it would not be possible to discuss each example in great detail. Several distributions appear that I have never personally used (e.g., the Rayleigh distribution), but which should still be covered by a book of this scope. The one missing distribution ("missing" being defined as "I have used it and it is not discussed") is the Studentized range. The distribution of the maximum of several t variables is discussed in the chapter on the t distribution, but I could not find a discussion of the range.

The material on nonparametric distributions includes the run test, sign test, signed rank test and rank sum test. Even though I was initially surprised to see discussion of nonparametrics in what is ostensibly a handbook of parametric distributions, the author incorporates the ideas seamlessly and the inclusion of these topics enhances the usefulness of the book.

This reference book is promoted as being for "practitioners and researchers in disciplines other than statistics" (Preface). As such, the technical ability of the target audience is probably somewhat lower than what most readers of *TAS* would deem appropriate for their personal needs. The focus of the book is on being able to access statistical distributions, not on derivations or proofs. A corresponding book from the 1980s or earlier would probably include many pages of tables of critical values of various distributions. This book contains no such tables, instead providing a software package to generate the values of interest. The included software is an integral part of the book. Dubbed StatCalc, the software is extensively referenced in instructing the reader how to calculate percentiles of distributions, moments, p values, confidence intervals, and even sample sizes for planning an experiment. Without the software, the book will be of limited benefit to the target audience since most examples in the text (and there are many) require the software for the reader to confirm the result or solve similar problems. (Do not lose the disk that comes with the book.)

This book is more suited as a reference than a course text. A lending library that caters to users of applied statistical methodology will find it to be an asset. It is also worthy of consideration as a reference book in a departmental library that caters to graduate students in statistics, due in large part to extensive discussion of generating random variables and extensive computer codes that can be easily adapted into many standard software languages.

In summary, this book can be recommended to statistical practitioners who need a comprehensive yet brief reference on statistical distributions with applications. The worst I can say about it is that a statistical purist will occasionally be annoyed—even though it generally meets the needs of the target audience.

Brian WIENS
Gilead Sciences, Inc.

REFERENCES

Donahue, R. M.J. (1999), "A Note on Information Seldom Reported via the p Value," *The American Statistician*, 53, 303–306.

Hubbard, R., and Bayarri, M. J. (2003), "Confusion over Measures of Evidence (p's) and Versus Errors (α's) in Classical Statistical Testing," (with comment and rejoinder) *The American Statistician*, 57, 171–178.

Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining.

Glenn J. MYATT. Hoboken, NJ: Wiley, 2007, xii+280 pp., $74.95 (P), ISBN: 0-470-07471-X.

The preface of *Making Sense of Data* (MSD) indicates that it is aimed at various members of an entire interdisciplinary team working with data, and could also be used as an introductory textbook (undergraduate or graduate) for students in any discipline. Chapter 1 is entitled "Introduction," and gives an extremely brief overview of the subject and outline of the book. Chapter 2, "Definition," discusses objectives, deliverables, roles and responsibilities of team members, and the nature of a project plan, and closes with an outline of the elements of a case study. The third chapter, "Preparation," mentions data sources, types of data, scales of measurement, data preparation (cleaning, removing variables, transformation) and a few other odds and ends. The first three chapters comprise only 35 pages. Chapter 4 on tables and graphs takes just 14 pages to cover contingency tables, frequency polygrams, histograms, scatterplots, box plots, and multiple graphs, obviously with little detail, commentary, or interpretation.

The author endeavors to cover the basics of statistics in the 40+ pages of Chapter 5: descriptive statistics, sampling distributions, estimation, hypothesis tests (including chi-squared tests and one-way analysis of variance (ANOVA)), and descriptive correlation. I would not recommend the exposition to someone who has not previously seen this material. For example, in the lead-in to hypothesis testing, the author writes "We can, however, minimize the chance of an error by specifying a confidence level that reflects the chance of an error," and in the ANOVA subsection we are told that "Generally the hypothesis statement will look like H_0: the sample means are equal." The abbreviated and sometimes incorrect/incomplete descriptions of statistical methodology aside, I was surprised at the almost complete absence of real data. Too often there was not even a context supplied with artificial data, as in the introduction to correlation. There are only 11 end-of-chapter exercises, none containing real data.

The next two chapters, "Grouping" and "Prediction," each roughly 50 pages long, constitute the data mining part of the book. The major topics in the first of these chapters are clustering (hierarchical and k-means), associative rules, and decision trees. The exposition was easy to read and would be mildly informative for someone who knows nothing about these concepts or methods (but does such an individual really need 1.5 pages of text on the calculation of Euclidean distance?). The prediction chapter discusses traditional regression at a