Biometrics 62, 1269–1280
December 2006

BOOK REVIEWS

EDITOR:
THOMAS M. LOUGHIN

HACCOU, P., JAGERS, P., and VATUTIN, V. A. **Branching Processes: Variation, Growth and Extinction of Populations**. Cambridge University Press, Cambridge, U.K., 2005. 316 pp. US$95.00/£60.00, ISBN 0-521-83220-9 (hardback).

Branching processes form a classical area of applied probability with origins in the work of Galton and Watson (1875) on the extinction of families. Since then the topic has undergone substantial development and much of today's research literature, for example on superprocesses and coalescents, has a considerable mathematical sophistication.

Branching processes have been used in various areas, including nuclear splitting and cosmic rays, but the most important application area remains biology. The present book is an exposition of biological applications combined with a leisurely exposition of the mathematical background. The intended audience is mainly biologists, but the abundance and diversity of biological examples should make the text highly enjoyable also for probabilists and statisticians.

The book starts with illustrating stochastic and deterministic modeling, where a main difference is that stable stochastic population models will typically not exist: in all realistic models, the population will either explode or become extinct.

This leads naturally to introducing the benchmark process, the discrete Galton–Watson process, and demonstrating its most basic properties. The Galton–Watson process incorporates one sex only and assumes that individuals reproduce independently and according to the same law that remains fixed with respect to time, increasing population size, and other population attributes. The basic results are the form of the extinction probability as the solution to an equation involving the probability generating function, the exponential (Malthusian) growth when extinction does not occur, and the quasistationary behavior of an eventually extinct population.

Generalizations of the Galton–Watson process that are treated include processes with several types, sexual reproduction, sibling dependence, age-structured populations, environmental variation, and size dependence.

A topic of particular current interest is coalescents, as introduced by Kingman (1982), and their applications to ancestral inference. A coalescent can be seen as a backward description of the family tree in the classical Wright–Fisher model from population genetics (including mutations). Given the set of DNA sequences from the current population, coalescent models lead to probability statements on quantities such as the number of generations ago all individuals (or species) had a common ancestor.

Another application example is cell kinetics, where branching processes can help understand the structure of a population of cells, say in a tumor, according to age and composition according to the basic phases of a cell's life, G_1 (initial growth), S (DNA synthesis), G_2 (continued growth), and M (mitosis or cell division). A further main class of examples is the dynamics of population of plants or animals. However, as already mentioned, the book is rich in examples and the ones just outlined only form a sample.

In addition to the three main authors, altogether 15 well-known scientists have contributed with longer or shorter sections: M. Alexandersson, G. Alsmeyer, A. D. Barbour, M. Durinx, M. Gyllenberg, G. Högnäs, V. A. A. Jansen, M. Kimmel, F. Klebaner, T. G. Kurtz, J. A. J. Metz, P. Olofsson, S. Sagitov, N. Sollenwerk, and S. Tavare. In view of this, it is natural that the style and level of the book are not entirely uniform. However, efforts have been taken to make large parts of the text accessible to a biological audience with some mathematical background so the book will serve well for an advanced course in mathematical biology. It would also be possible to use it for a course for statistics and probability students at a rather early stage of their studies.

The classical texts on branching processes are T. E. Harris's *The Theory of Branching Processes*, published by Springer-Verlag (1963), which is more mathematical and stresses the applications to physical sciences more than the biological ones, and K. B. Athreya and P. E. Ney's *Branching Processes*, published by Springer-Verlag (1972), which also is more mathematical. Also P. Jagers's *Branching Processes with Biological Applications*, published by Wiley (1975), is probably only accessible to a rather limited number of mathematical biologists. A recent book of a rather similar spirit and level as the present one is M. Kimmel and D. E. Axelrod's *Branching Processes in Biology*, published by Springer-Verlag (2002), which in comparison is somewhat more narrow in its choice of examples and applications.

References

Galton, F. and Watson, H. W. (1875). On the probability of extinction of families. *Journal of the Anthropological Society of London* **4,** 138–144.

Kingman, J. F. C. (1982). On the genealogy of large populations. *Journal of Applied Probability* **19A,** 27–43.

Søren Asmussen
Department of Mathematical Sciences
Aarhus University
Aarhus, Denmark

GENTLEMAN, R., CAREY, V., HUBER, W., IRIZARRY, R., and DUDOIT, S. (eds) **Bioinformatics and Computational Biology Solutions Using R and Bioconductor**. Springer Science+Business Media, New York, 2005. xx + 473 pp. US$89.95/€74.85, ISBN 0-387-25146-4.

The statistical analysis of genomic data, namely microarray data, has attracted numerous statisticians, mathematicians, and computer scientists to the interdisciplinary fields of genetics, genomics, bioinformatics, and computational biology. This book is solid evidence of the influence that quantitative researchers can have on biological investigations. Organized into separate chapters of shared authorship, the book provides a valuable overview of the impact that the authors and their colleagues have had on the analysis of genomic data. The fluidity between chapters and subjects is sometimes lacking, but the main message of bringing statistical, computational, and graphical solutions to real biological problems rings clearly throughout the book.

Equally important to a potential reader of this text is knowing that it is centered on the Bioconductor project (`www.bioconductor.org`), a publicly funded effort that is quickly becoming known as the resource for the statistical analysis of microarray data, as well as other biological, bioinformatics, and computational biology investigations. One of the key objectives throughout the book is to showcase the broad capabilities of the Bioconductor project when considering the many suggested computational biology and bioinformatics approaches. Coupled with the software package R and the Bioconductor project, this book provides a powerful tool for the experienced statistics or bioinformatics reader, in that knowledge of statistics, computer programming, and R is required for successful implementation of the recommended bioinformatics and computational biology solutions. This text is not for the mathematically or quantitatively challenged, and is certainly not a "point and click" solution to bioinformatics problems. As discussed by the authors, this is a fully computable book that allows dynamic tables, figures, data, R code, and graphical output, as well as an interface with a companion web site (`www.bioconductor.org/mogr`) from which the reader can download, edit, and refine code to potentially suit their own application. As with all books that have a web interface, the reader may be left wondering how long the link will remain active once the text is either revised or out of print. While concentrated heavily on the statistical analysis of microarray data, the associated statistical issues, and the graphical presentation of results, there is a view to the future

of these analyses and results with chapters on sequencing, GO annotation, mass spectrometry protein data, and online resources.

Five relatively broad sections define the structure of this book. The 25 chapters and 1 appendix entail preprocessing data (108 pp.), meta-data (69 pp.), statistical analysis (145 pp.), graphs and networks (66 pp.), case studies (37 pp.), and details on selected resources (10 pp.). For quantitative researchers who are already comfortable with the biological concepts, technology, and statistical analyses and investigations of genomic data, this book is very useful in providing ideas about what "the next step" of the investigation might be. Typically, in practice, the results from rather sophisticated statistical analyses are given to the biologist in the form of significant results, lists of differentially expressed genes, changing protein levels, etc. The biologist might review the results, ask a few questions about the analyses, and then either continue on with more focused experiments or present their results in a scientific paper. The majority of quantitative researchers, until recently, have not been involved in exploring online resources for the purpose of improving their own analysis. The message from the authors of this text is that it is important for quantitative researchers to be able to connect their results with online genomic resources, such as gene or protein sequence information, databases of biological and molecular information, structure databases, and/or additional literary information, so that more complete analyses and biological understanding can be obtained. The readers are fortunate in that the Bioconductor project includes numerous R packages that provide avenues to resources that can be implemented in either a real-time or a static fashion via the World Wide Web or Web services. The authors expertly detail the capabilities allowed within the Bioconductor project, but are also careful to lay out obvious and nagging issues that so often cause problems, namely version tracking, incorrect annotation, differences in terminology, etc. Many of the chapters end with a short discussion of the importance of the main points and the anticipated need for future work, thus reminding the reader that the fields of bioinformatics and computational biology are highly active areas of research that will continue to provide many fruitful opportunities for quantitative researchers.

The concept of having a fully computable book that interfaces with a web link to provide dynamic resources was once a visionary idea, but as evidenced by the format of this monograph it is becoming the rule rather than the exception of the future. The fact that bioinformatics and computational biology themselves are such dynamic areas of research requires the rather dense, yet brief, treatment of bioinformatics and computational biology solutions that the authors have provided. As the Bioconductor project evolves to include more current bioinformatics challenges (e.g., molecular dissection of complex traits, expression quantitative trait loci [eQTL], epigenomics, etc.) the covers of this text will clearly expand beyond the current volume.

R. W. Doerge
Department of Statistics
Purdue University
West Lafayette, Indiana, U.S.A.

VITTINGHOFF, E., GLIDDEN, D. V., SHIBOSKI, S. C., and McCULLOCH, C. E. **Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models**. Springer, New York, 2005. xv + 340 pp. US$79.95/€74.85, ISBN 0-387-20275-7.

Degree programs to give physicians sufficient background to conduct independent research in epidemiology and clinical research are becoming increasingly popular. These programs typically have one to three courses that focus on biostatistics methods, and *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models* is designed to be the primary biostatistics textbook for the second or third of these courses. Linear, GLM, and Cox regression models are explained with a minimum of theoretical development; only first-year algebra is required to read this text. The focus is on understanding key statistical and analytical concepts—interpreting regression coefficients, understanding the impact of the failure of model assumptions, grasping how correlation in clustered sample designs affects analysis—rather than on mathematical derivations.

Chapter 1 introduces the text and the motivation behind it. Chapter 2 discusses descriptive statistical methods, including histograms, box-and-whisker plots, and Q–Q plots, as well as bivariate analysis of both continuous and categorical variables.

Chapter 3 introduces t-tests and contingency tables, linear regression, and Kaplan–Meier curves. Chapter 4 returns to linear regression as the prototype for all other regression models considered, discussing in detail confounding and interaction, and introducing the concepts of causal effects and counterfactuals to distinguish mediation from confounding. Transformations to improve normality and reduce heteroscedasticity, transformations, and categorization of continuous predictors, and outlying, high leverage, and high influence points are also discussed. Chapter 5 discusses causality in the context of model selection, and considers the issue of including predictors in both randomized and nonrandomized designs. The authors present the case for backwards selection when screening of covariates is required. Chapter 6 develops and interprets the standard logistic regression model and explains its use in case–control studies. Chapter 7 discusses survival analysis regression models, emphasizing the Cox proportional hazard model. Time-dependent covariates and interactions with time (failures of the proportional hazard model) are briefly discussed. Chapter 8 considers the analysis of longitudinal/hierarchical data, focusing on marginal and random effects models for linear and logistic regression. An excellent nontechnical discussion of the difference in the interpretation of marginal model and random effect model regression parameters is provided. Chapter 9 discusses the GLM, focusing on Poisson regression. Chapters 6–9 also reinforce basic concepts such as confounding and interaction, as well as specific details of model fit and checking, first developed for Gaussian linear regression in Chapter 4. Chapter 10 briefly discusses the analysis of complex sample design survey data, emphasizing the need to use either generalized-estimating-equation-type or replicate-type (jackknife, balance repeated replication) techniques for variance estimation. Chapter 11 briefly overviews data management and statistical analysis

plans. Problems at the end of each chapter focus on reviewing concepts; problems requiring students to conduct data analyses will have to be provided by the instructor.

Restricting mathematical development means that some concepts, such as the relationship among hazard, cumulative hazard, and survival functions, are hard to follow. Also, in attempting to touch on most core topics in biostatistics, some important material is given short shrift. Bootstrapping is described as close to a panacea for small-sample inference; while bootstrapping generally has better small-sample properties than other asymptotic alternatives, the target audience for this text may not understand that it too has limitations. Discussions of multiple comparison problems, measurement error, and alternative effect coding are very limited, and the description of matched case–control studies is incomplete. Power is not mentioned at all. However, the authors provide a number of standard references at the end of each chapter for readers who might desire more in-depth or mathematically formal explanations.

Two competitor texts are *Modern Epidemiology* (Rothman and Greenland, 1998) and *Epidemiology: Study Design and Analysis* (Woodward, 2005). These texts have broader goals than *Regression Methods in Biostatistics*, which is focused on regression modeling. *Modern Epidemiology* is a reference text collecting chapters from 15 authors; it covers many of the same topics as *Regression Methods in Biostatistics*, sometimes in greater depth, but not as concisely, and no problem sets are available. *Epidemiology: Study Design and Analysis* is a standard biostatistics textbook for epidemiologists. It also covers similar material in more detail, but requires a somewhat greater level of mathematical sophistication than *Regression Methods in Biostatistics*.

Regression Methods in Biostatistics was developed from a two-quarter course sequence that the authors taught to postgraduate physicians training for clinical research. Instructors working in a semester system will need to cut material for a single-semester course. The depth of methodological development is too shallow for statistics or biostatistics students, or epidemiology students with strong quantitative backgrounds. Nonetheless, the emphasis on concepts, rather than on the mathematical detail of methodological development, makes this book useful background reading for an instructor looking to reach students at both levels. The data analysis examples provided would be useful in applied biostatistics courses at any level, and are available at a web site provided by the authors. Instructors teaching biostatistics to clinicians, or to anyone without a background in calculus or linear algebra, should consider *Regression Methods in Biostatistics* as a primary or secondary text for their course.

References

Rothman, R. J. and Greenland, S. (1998). *Modern Epidemiology*, 2nd edition. Philadelphia: Lippincott-Raven.

Woodward, M. (2005). *Epidemiology: Study Design and Analysis*, 2nd edition. Boca Raton, Florida: Chapman and Hall/CRC.

Michael Elliott
Department of Biostatistics
University of Michigan
Ann Arbor, Michigan, U.S.A.

McKILLUP, S. **Statistics Explained: An Introductory Guide for Life Scientists**. Cambridge University Press, Cambridge, U.K., 2006. xii + 267 pp. US$75.00/£45.00 (hardback), US$29.99/£19.99 (paperback), ISBN 0-521-54316-9.

The book is intended to be used as a reader-friendly introduction to experimental design and statistical analysis, or as a refresher. The coverage is broader than that for a typical first (undergraduate) experience, including chapters on nonparametric tools, selected ANOVA (analysis of variance) topics beyond an introduction to one-way ANOVA, as well as sections on philosophy and ethics of doing science that often get left out of statistics texts. In sum, I think it reads reasonably well for the latter (refresher) purpose, but it appeals less for an introduction.

There are no homework exercises; indeed, the book is slanted much more toward discussion of conceptual issues rather than data-driven learning. My own feeling is that playing with data, learning to answer real and interesting science questions, is absolutely a great way to learn, and this book lacks appeal for that purpose. Case in point: as illustration of hypotheses and predictions, the author introduces a study he conducted to investigate whether Portuguese millipedes are attracted to light, interesting and important because these millipedes had become a serious (introduced species) pest in South Australia. Apart from a brief statement (p. 41) that the results were consistent with his hypothesis-driven predictions, my curiosity about the study (How dramatic were the results? What statistical tools did he use to analyze it?) was not satisfied.

The use of statistical language in the book is out of step with that currently in vogue in North America, which is still likely to make the book more attractive to an audience other than North Americans. For example, the pair (mensurative study, manipulative study) is rarely used on this side of the Atlantic for observational study (experiment). I prefer the pair (response, predictor) to McKillup's (dependent, independent) as descriptors of the variables in regression.

For primarily these reasons, I would not recommend the book for use as an introduction to statistics. It is, however, a very readable overview of a broad suite of statistical methods, and might be of interest to practicing biologists as review.

Ken Gerow
Department of Statistics
University of Wyoming
Laramie, Wyoming, U.S.A.

BADDELEY, A. and VEDEL JENSEN, E. B. **Stereology for Statisticians**. Chapman & Hall, Boca Raton, Florida, 2005. 395 pp. US$79.95/£44.99, ISBN 1-58488-405-3.

Stereology is the business of inferring information about three-dimensional structures from two-dimensional slices: imagine slicing a lump of spaghetti and trying to estimate the number of strands, their length, density, curvature, etc., from the resulting section. The theory of stereology has been developed for over a century. This book is the first comprehensive presentation from a statistical viewpoint. It is aimed at statisticians whose work involves collaborating with and being consulted by scientists using stereology. A stated purpose in

the book is to enable statisticians to better collaborate with such scientists.

Overall the level of the book is nicely balanced between explaining ideas with reference to the main issues in scientific use of stereology and giving clear mathematical derivations of the results. The diagrams are excellent, and help the reader to visualize how three-dimensional structures will appear in two-dimensional slices. Each chapter includes exercises. In earlier chapters some of these are practical, and include baking a cherry cake, slicing it, and estimating from these slices how many cherries were used.

After starting with a short introduction, Chapter 2 covers classical stereology, which is based on mathematical formulae, and does not properly address where the randomness lies (is it in the stochastic nature of the material or the position and orientation of the slicing?). In this classical approach, the variance of estimators is not fully addressed.

The next chapter (Chapter 3) provides an overview of a modern statistical approach to stereology. The main ideas and issues, which will be examined in more depth in later chapters, are introduced. The following three chapters cover the essential theoretical bases for stereology: geometrical identities (Chapter 4), geometric probability (Chapter 5), and the statistical formulation of stereology (Chapter 6). Chapter 4 presents results as mathematical identities, which are the basis of classical stereology. These relate quantities such as the integral of region area or boundary over two-dimensional slices to region volume or surface area, respectively, in the whole object. Chapter 5 revisits these ideas from a probabilistic viewpoint: important results now appear as hitting probabilities and expectations. In Chapter 6, statistics is introduced, and the basic properties of estimators of quantities of interest are presented.

The next few chapters present in detail the theory for stereological estimation in various situations. Chapter 7 considers the estimation of an absolute quantity such as the volume, surface area, or length of the feature of interest (think of the yolk in a hard-boiled egg). Chapter 8 describes sampling designs in which the two-dimensional section is not isotropically oriented, or is not uniformly positioned in experimental material that has a reference axis or point (think of a pear or banana). Chapter 9 discusses ratio estimation, in which a ratio of two population parameters is estimated from sample data by taking the ratio of the corresponding sample quantities.

Often material under study is not homogeneous, and we need to take a sample from it by first cutting it into pieces and then treating these as individual units in a sampling population. Chapter 10 considers how to design such an investigation when the properties of interest may not be uniform over the object. In Chapter 11 we suppose the interest to be in particles contained within the material of study (remembering the cherry cake we baked earlier). We may be interested in the density or size distribution of these particles, for example, and the issues of whether we have the whole object of interest available, or a sample from a much larger reference space, need to be addressed.

Chapter 12 departs from the mathematical treatment of previous chapters to consider some of the practical issues in designing stereological experiments. These include sampling issues: stratification, nesting, subsampling, and some practical issues in how to place slices where the design calls for them. Chapter 13 presents theoretical results about the variance of stereological estimators, and describes practical techniques for estimating the variance from sample data. This is considered for both the model-based approach (where the random variation is in the material) and the design-based approach (where it is in the placing of the sections). The important point is made that in biological studies, there is greater variance reduction from increasing the number of biological replicates than from increasing the number of sections per biological sample. The final chapter (Chapter 14) briefly presents some frontiers and open problems, and an appendix reviews the main results of standard (not specifically stereological) sampling theory.

In summary, this book succeeds in what it aims to do: to provide a comprehensive introduction and reference to stereology for statisticians. It will be a valuable addition to this reviewer's bookshelf.

Graham Horgan
Biomathematics & Statistics Scotland
Rowett Research Institute
Aberdeen, U.K.

LITTELL, R. C., MILLIKEN, G. A., STROUP, W. W., WOLFINGER, R. D., and SCHABENBERGER, O. **SAS® for Mixed Models, 2nd edition.** SAS Institute, Cary, North Carolina, 2006. xii + 813 pp. $89.95, ISBN 1-59047-500-3.

For the last 10 years, the first edition of *SAS for Mixed Models* (Littell et al., 1996) has been a staple in my work as a statistical consultant. Furthermore, as a former colleague of one of its authors, George Milliken, I became accustomed to my privileged access to his expertise in SAS analysis to help on problems not addressed in the book. My need for George's help did not decline as my experience with mixed models grew, however, because as soon as I mastered a set of tricks, SAS would upgrade its capabilities in such a way that some of my old problems were solved and I could begin to tackle new ones. As the MIXED procedure in SAS mushroomed and the GLIMMIX procedure became established, consulting *SAS for Mixed Models* became increasingly less fruitful, and bothering George (or the SAS consultants) essentially became my only recourse.

This is why I was excited to receive a copy of *SAS for Mixed Models*, 2nd edition to review. I would normally avoid the potential conflict of interest in reviewing a book coauthored by a former colleague, but this is a book whose existence means a great deal to me, professionally. I jumped at the opportunity to give the book a thorough inspection in the name of writing this review.

The second edition of *SAS for Mixed Models* updates the original to full compatibility with SAS 9 (with notes about compatibility with SAS 8). The authors have revised the SAS code from the original examples to make more complete use of the newer tools available for analyzing mixed models. The book also includes an augmented and up-to-date discussion of nonlinear and generalized linear mixed models. Lastly, three new chapters have been added, covering mixed model diagnostics, power calculations, and the potential for some limited (empirical) Bayes analysis of mixed models.

As in the first edition, the authors put the bulk of their emphasis on linear mixed model analysis. Estimates of means and their standard errors, means contrasts and comparisons, BLUPS, *F*-tests, and covariance matrix models are covered in detail in the context of an assortment of different types of problems. The book begins with a chapter that provides a refresher on statistical models and random effects, plus some flow charts to aid in the selection of SAS procedures for different data structures. Subsequent chapters deal with different problems in mixed modeling, including randomized block designs (both complete and incomplete), designs with multiple error terms, repeated measures, analysis of covariance, heterogeneous variances, spatial variability, nonlinearity of the mean function, and nonnormality of the response. The main body of the book concludes with a chapter of 11 case studies. This is followed by an appendix covering linear mixed model theory. All data sets used in the book are provided in print in a second appendix, and all programs used in the book are available both on an accompanying CD and on the book's website.

Each chapter starts with some mild theory and development of commonly used statistics for the topic. One or more examples are introduced and basic SAS statements are given for their analysis, exploring both the essential statements and a variety of useful options. Segments of output are presented, followed by explanations of parts that are immediately relevant to the example at hand. These explanations do not encompass every number in every table; readers have to accept that the remaining parts of output will be discussed later if they are needed. The level of detail in both the programs and the output explanations is sufficient to address many questions, or at least to steer a user in the right direction.

Because the subject of mixed models is not really "entry-level" material, the book is not really intended for statistical novices. Nonetheless, intrepid beginners might find parts of the book useful, provided they can properly identify their fixed and random effects in the first place. Really, the book is meant for more experienced data analysts, especially those with degrees in statistics, or at least past experience with linear models. For instance, the fact that much of the theory and development at the start of each chapter is in matrix notation precludes thorough digestion by novices.

Within its 800+ pages, the book covers far more than would be expected from a mere "manual" or a "handbook" of SAS analyses. The level of sophistication and detail is sufficient to genuinely instruct readers both on the uses of mixed models and on their analysis in SAS. It would be an excellent companion book to a design of experiments course or a "messy data" course, and it could legitimately serve on its own as a text for a course on applied mixed models, provided that SAS was to be the tool of choice for the course (there is a heavy emphasis on the specific capabilities of the MIXED, GLIMMIX, and NLMIXED procedures). The combination of statistical development with clear examples and practical advice makes it an excellent resource for self-study and a superb reference book on mixed models of all types. Finally, owners of the first edition who, like me, have become increasingly frustrated with its obvious age will likely find the new edition a worthwhile investment, as it makes as much use as possible of the current state of the art in SAS mixed modeling capabilities and general SAS programming.

Thomas M. Loughin
Department of Statistics and Actuarial Science
Simon Fraser University-Surrey
Surrey, British Columbia, Canada

DEMETS, D. L., FURBERG, C. D., and FRIEDMAN, L. M. (eds) **Data Monitoring in Clinical Trials. A Case Studies Approach.** Springer, New York, 2005. xxvi + 374 pp. $49.95/€39.54, ISBN-10: 0-387-20330-3, ISBN-13: 978-0387-20330-0 (paperback).

Monitoring of clinical trials is an essential process to ensure that trials do not continue longer than necessary to achieve their objectives. Thus, monitoring of a trial should address among other things recruitment of patients, quality of data, and early evidence of benefit or harm. For this purpose, investigators or trial sponsors often establish a group of independent experts to monitor a specific clinical trial. This group, sometimes referred to as a data safety monitoring board or data monitoring committee, is especially responsible for monitoring the trial with respect to safety and early benefit. The task is complex and conflicting. Situations can be thought of where decisions are not easy to make. This book helps to get insights into the complex decision processes and considerations necessary to ultimately recommend whether to continue or discontinue a trial.

The book is subdivided into four major sections where the single chapters are all written by experts with experience in planning, monitoring, or analyzing a clinical trial. What is special about the book is that all chapters discuss specific trials and the role of the data monitoring committees (DMCs), their decisions, and the consequences. The first three chapters (Section 1) may be regarded as an exception to this rule since they introduce the tasks of DMCs, the lessons learned from various practical examples, the history of DMCs, and the experience with such committees from the FDA's perspective, but even these overview or introductory chapters address numerous real trials. Section 2 (11 case studies) is devoted to clinical trials where an early benefit could be observed. Here, the DMCs had to face the conflict "to balance short-term results against the desire for long-term information." Section 3 presents nine clinical trials with just the opposite problem. These studies showed evidence of being harmful for the patients due to the trial intervention. Eight of them were stopped earlier than planned. Special issues are treated in Section 4 (9 case studies) which cannot be solved by just a single rule, such as data flow problems, changing endpoints, or relevance of the question posed.

This book is highly recommended for anyone who starts to serve on a DMC or is already a member of such a board, and also for people in academia, industry, and regulatory agencies as well as for students in biostatistics, epidemiology, clinical trials, and medical ethics. It would be helpful for the understanding of the issues treated in this book for the reader to

have a basic knowledge of statistical design and analysis of clinical trials.

Iris Pigeot
Bremen Institute for Prevention Research
and Social Medicine
Bremen, Germany

HOSHMAND, A. R. **Design of Experiments for Agriculture and the Natural Sciences, 2nd edition.** Chapman & Hall/CRC, Boca Raton, Florida, 2006. 437 pp. $79.95/£29.99 (hardcover), ISBN 1-58488-538-6.

This book is an introductory text that covers classical topics in the design and analysis of experiments, but with emphasis on their use in the agricultural and basic biological fields. The goal of this text is to present the "necessary guidelines and techniques for designing experiments" for students and researchers in agriculture and biology. Like most books of its type, this text concentrates much more on analyzing experiments than on designing experiments. The text is mainly well written and the statistical techniques are reasonably presented. The author is associate dean for graduate education and chair of business and management at Daniel Webster College, and received his Ph.D. in agriculture and resource economics.

Most chapters begin with a basic description and example of a design, and then proceed with a step-by-step analysis of data collected under that design. Chapters 1–3 present basic terminology and examples, assumptions of the classical analysis of variance, and concepts for reducing error (variability) via experimental design. These chapters are placed correctly at the beginning of the text, and prepare the reader with design foundations. Subsequent chapters present single-factor experiments (including randomized complete block, incomplete block, and Latin square designs), two-factor experiments (including split unit and strip unit designs and two-factor interaction), three or more factor experiments (including fractional replication and combinations of split and strip designs), classical comparisons of means (LSD and Duncan's multiple range tests), experiments over seasons and years, basic regression and correlation, and analysis of covariance. Appendices contain some design plans and tables of critical values for some commonly used test statistic distributions (e.g., Student's t test). Unfortunately, test results are presented only by comparing test statistic values with critical values. That is, no α- or p-values are presented. This is not consistent with the "modern" teaching of statistical hypothesis testing. Most statistical references are dated from the 1980s or earlier, whereas most application (e.g., examples, problems) references are dated from the 1990s and later.

One strength of this text is that there are many actual agricultural and biological examples and data analysis problems. Some problems alter the original data set to illustrate how analysis results change under "nonclassical" situations (e.g., missing data). This is a good connection to actual research situations. However, for a "design of experiments" text, there should be design problems to challenge the reader, and I could find none in this text.

Illustrations in an introductory text are important for conveying and enhancing the textual discussion of basic concepts. The text contains basic graphics that one would find in most any text of this type. These illustrations are not consistent, however, and many of them are of poor reproductive quality (e.g., some of the graphics look like they are low-resolution screen images).

Basic linear models are used, at least partially, to motivate the analysis method, typically analysis of variance. There is not a heavy reliance on using these models, just enough information to give the reader an idea of their use. However, there are many "hand calculation" equations, including intermediate computational steps. This approach is used in place of software code and output, of which there is none in the main text (an appendix gives a basic presentation of several analyses using MINITAB). Also, I could not find a web page (say, with downloadable data sets) connected with this text, and a data disk or CD-rom is not included.

Overall, *Design of Experiments for Agriculture and the Natural Sciences* presents classical design and analysis of experiments concepts and examples in a traditional approach. It excels with the agricultural and biological examples and problems, but lacks a "modern" approach to presentation that is being used currently by most design of experiments instructors. This text would be beneficial to those whose backgrounds are in agriculture and biology, those who would like to see basic computational details, and those who prefer the classical test statistic/critical value approach to hypothesis testing.

Jeffrey S. Pontius
Department of Statistics
Kansas State University
Manhattan, Kansas, U.S.A.

Editor's Note: Jeff Pontius died on August 28, 2006, after several years of battling cancer. He was 52 years old.

ALHO, J. M. and SPENCER, B. D. **Statistical Demography and Forecasting**. Springer, New York, 2005. xxvii + 410 pp. US$89.95, ISBN 0-387-22538-2.

This book represents a quite complete review of main statistical techniques in demography. It is very well structured as a course guide, both at undergraduate and postgraduate levels, but it is also useful as a reference text for consultants in the field. A compendium of standard techniques in statistical demography and forecasting is presented together with useful applications in economic and social prediction. Some of the most important aspects and concepts addressed in this book are Lexis diagrams, Markov chains, sampling theory, distribution theory for sampling, waiting times, survival probabilities and their estimation in forecasting, Poisson processes, regression models for counts and survival (as logistic, Poisson, and Cox regressions), multistate life tables and general linear growth models, demographic time series models, prediction errors, uncertainty in demographic and stochastic forecasts, predictive distributions, errors in census numbers, financial applications based on demographic forecasts, decision analysis, and small area estimates.

Main issues are explained with clarity and mathematical rigor, and its wide variety of practical examples, developed from empirical data obtained in different countries around the world, makes the exposition extremely attractive. Intergeneration of fiscal equity under population aging, sustainability of pension systems, and accounting for health-care benefits for future retirees are examples of policy problems in population forecasts and their uncertainty. Indeed, statistical demography is a necessary tool for key social policy issues such as pension and health-care systems in the developed countries, estimation of the size of elusive populations, unemployment, epidemiology, and other hot areas. It is relevant that some statistical theories and methodologies for estimating and forecasting of births, deaths, and migration are presented in a multistate setting.

The first chapter of the book introduces the role of statistical demography, statistical inference, time series, bias and sampling errors, and other topics, with some basic statistical notation and preliminaries, allowing the future design of short specific courses. For example, a self-contained introduction to statistical demography can be taken from Chapters 2, 4, 5, and 6, with Chapters 2 and 4 defining a potential shorter course on biometrics. An introduction to demographic data sources and their quality can be taken from Chapters 2, 3, 4, 10, and 12. And from Chapters 4, 6, 7, 8, 9, and 11 we can obtain an introduction to demographic forecasting, mainly for pensions and public finances.

The exposition meets standard requirements for mathematical statisticians, demographers, actuaries, and epidemiologists, assuming some knowledge of differential and integral calculus, matrix algebra, probability theory, and regression analysis. The book is supported by updated references, an author index, and a subject index. Many figures illustrate concepts and examples throughout the book.

M. Ruiz Espejo
Universidad Nacional de Educación a Distancia
Madrid, Spain
and
F. J. Montero
Universidad Complutense
Madrid, Spain

AMSTRUP, S. C., McDONALD, T. L., and MANLY, B. F. J. (eds) **Handbook of Capture–Recapture Analysis.** Princeton University Press, Princeton, New Jersey, 2005. xv + 313 pp. US$49.50/£32.50 (paperback), ISBN 0-691-08968-X.

Capture-mark-recapture (CMR) methods are a rapidly expanding area of biometrics where development of new statistical models, model selection procedures, and computer software has given biologists a valuable set of tools for tackling questions in population and community ecology. The essays in this edited book provide a useful overview to CMR methods by describing classical and new models with statistical notation accompanied by the analyses of example data sets. This book is a good introductory text for graduate students working at the interface of ecology and statistics. It will be a basic reference for biometricians with an interest in models for estimation of demographic parameters, but is less comprehensive than the recent textbook of Williams, Nichols, and Conroy (2002).

The book contains 10 chapters that can be subdivided into five main sections. First, Chapters 1 and 10 are bookend chapters that provide a general introduction and synthesis. Second, Chapters 2 and 4 consider classical and new developments in closed population models for live encounter data. Closed population models are often used for estimation of population size. Third, Chapter 3 introduces classical Cormack—Jolly–Seber (CJS) models for open populations and live encounter data, used for the estimation of apparent survival and encounter rates. Chapters 5 and 8 discuss newer extensions of CJS models, including temporal symmetry models for estimation of recruitment and population change, robust design models for estimation of temporary emigration, and multistate models for the estimation of state-specific demographic parameters. Fourth, Chapters 6 and 7 discuss models for analysis of dead recovery data, both alone and jointly with live encounter data. Lastly, Chapter 9 and the appendices provide an overview of the software tools that are available to implement mark-recapture analyses.

A major strength of the book is that the chapters are well written and the notation will be easy to follow for a reader without training in biometrics. However, the book could have benefited from better organization. If the book is read in sequence, the flow is disjointed because the chapters alternate between closed and open population models, and also between models for live encounter and dead recovery data. Editorial flaws are limited. The authors attempted to standardize notation across all chapters but a few minor discrepancies remain. For example, the formula for corrected quasi-AICc values is given in Chapter 5 but not in the introductory Chapter 1, and the number of estimated parameters is variously denoted as either P (p. 18) or K (p. 100).

For newcomers to CMR methods, one of the best ways to become familiar with the models is to reanalyze example data sets where the results are already known. The authors extend this opportunity to the reader by providing the encounter histories that accompany many of their illustrative analyses. Photographs of capture and marking techniques for wildlife populations provide additional color. Despite being an introductory text, this book will likely offer new ideas to readers familiar with CMR methods. Two new topics that caught my interest included continuous time models for closed populations and applications of multistate models in stratified sampling.

The book falls short in the treatment of two issues, one technical and one philosophical. One issue that is common to most CMR analyses is a need for initial goodness-of-fit (GOF) tests to assess the fit of the global starting model to the data in the encounter histories. If necessary, lack of fit can be addressed by calculation of a variation inflation factor (c-hat) and by adjustment of model selection criteria. Adjustments for lack of fit are a critical issue that warranted a separate chapter, but is not mentioned in Chapter 1, and is treated by most authors as a side issue related to model selection. At the time of writing, the main GOF procedures available were CJS model tests and a parametric bootstrap

method, but a new median c-hat procedure based on logistic regression is beginning to receive wider use.

The authors cite the major papers for each set of CMR models, but potential applications are not well developed. Examples from linguistics and social and computer sciences are only briefly mentioned in Chapter 4 (p. 72). In ecology, advanced models for open populations have led to a resurgence of interest in animal demography including the first quantitative estimates of breeding propensity, and even estimation of demographic parameters for unobservable life stages (Sandercock, 2006)! Closed population models have great utility, and have been used to estimate the number of song elements in the vocal repertoire of birds, the size of plant populations, and the species richness of communities. Overall, the book could have done more to promote the utility of CMR models by exploring a wider array of empirical examples.

References

Sandercock, B. K. (2006). Estimation of demographic parameters from live encounter data: A summary review. *Journal of Wildlife Management*, in press.

Williams, B. K., Nichols, J. D., and Conroy, M. J. (2002). *Analysis and Management of Animal Populations*. San Diego: Academic Press.

Brett K. Sandercock
Division of Biology
Kansas State University
Manhattan, Kansas, U.S.A.

CROWLEY, J. and ANKERST, D. P. (eds) **Handbook of Statistics in Clinical Oncology, 2nd edition.** Chapman & Hall/CRC, Boca Raton, Florida, 2006. 617 pp. US$89.95/£49.99, ISBN 0-8247-2339-2.

Clinical research in the field of oncology continues to thrive. Cancer research continues to be a primary focus of research funded by the U.S. federal government, and has increasingly become a major focus for the pharmaceutical industry as the demand (and profits) associated with cancer therapeutics has increased. This continued emphasis on cancer research has been accompanied by an explosion in technology-enhanced assessment of possible prognostic and predictive factors, novel endpoints (such as functional imaging-based scans), and other tumor markers. Thus, an overview of current topics in clinical oncology is welcome. The *Handbook of Statistics in Clinical Oncology*, edited by Drs Crowley and Ankerst, attempts to provide such an overview. This is an expanded and revised second edition of a text that was initially well received by the community.

The book is comprised of 33 chapters, grouped into 6 sections. The first section is devoted to phase I trials and includes five chapters that discuss in detail the choice of endpoints, theory and practice of the designs in phase I cancer clinical trials, particularly the continual reassessment methods with toxicity-based endpoints, and pharmacokinetic analyses. Recent work in the literature on dose escalation designs for bivariate outcomes and multiple agents could have been included for completeness. Section 2 addresses phase II study design. The four chapters in this section cover in comprehensive detail the standard frequentist approaches to phase II trials. The chapter on the use of time-to-event endpoints is particularly helpful, as with novel agents, there is a (at least perceived) need to move away from the usual phase II endpoint of tumor response. There is little discussion of newer Bayesian approaches, and no discussion of the increasing emphasis on laboratory endpoints that accompany phase II trials.

The section on phase III trials spans seven chapters. The first few chapters in this section provide an overview of the randomization procedures and their impact on analyses, factorial designs including a detailed discussion of power considerations in the presence of interactions, design issues in noninferiority trials, and implications on sample size and design characteristics when the proportional hazards assumption is violated. The latter set of chapters deals with monitoring of trials including an interesting Bayesian approach to early stopping for futility and success, and methods to analyze quality of life and economic endpoints. A comprehensive list of references makes up for a lack of detail and depth of discussion of some concepts in this section. As with phase I trials, recent work on newly emerging designs for biomarker studies incorporating genetic or other prognostic factors could have been included.

Section 4 on exploratory analysis and prognostic factors is a refreshing and well–thought-out addition. A detailed discussion of sample size calculations for prognostic factor studies is a unique and useful feature. The discussion of some of the commonly used statistical tools to identify predictive and prognostic factors (prognostic model), although mostly comprehensive, does not address issues such as prevalence of a factor, sample size considerations, and use of the c-statistic to assess prediction accuracy. A highlight of the section is Chapter 21, which discusses an intuitive analytical tool that combines clinical outcome with biologic data to monitor clinical trials.

Considerations for analysis of high-dimensional data sets such as microarray and proteomic data are discussed in the six chapters comprising the fifth section of the book. Various types of microarray and genetic data are discussed and basic concepts necessary for understanding proteomic data structure are given. The chapters provide a good initial overview of objectives commonly addressed in these data sets including normalization, class comparison, class discovery, class prediction, and multiple comparisons. An overview of the methods available to address these objectives and how they work are also included. While the chapters apply the methods to real data sets, the reader will need to do further investigation before being able to choose their analysis methods, as the pros and cons of the different analysis methods are not discussed. The chapter discussing software, where to find it, how to choose a tool, data resources, and how to contribute software is a refreshing, practical read and openly points out areas needing further work.

Finally, Section 6 provides an overview of a variety of topics that are important but not included in standard texts. Peter Thall contributes an interesting, though somewhat theoretical, proposal for exploring between-trial heterogeneity in single arm trials. The excellent nontechnical discussion of competing risks, which is very helpful for the practicing statistician, is noteworthy. The closing three chapters are devoted mainly to a discussion of proper practices that, although most

should be common sense, clearly need such a presentation based on the literature full of examples of improper practice. Topics covered in this section include sequential randomizations and dose-intensity analyses.

The text is extremely well referenced. The authorship represents the leaders in the field, and for a text that consists of a collection of chapters from different authors, there is little overlap or duplication of topics. A significant strength of this book is its example-driven approach to the "out-of-the-box" but practical issues faced by those conducting cancer clinical trials. The book provides a critical assessment of the state of the science through a thorough review of the methods, applications, and available resources. Suggestions for expanded emphasis in future editions include the topics of meta-analysis, surrogate endpoints, and pharmacogenomics. In summary, the *Handbook of Statistics in Clinical Oncology* provides a detailed and mostly comprehensive text on the topics of cancer clinical trial design, conduct, analysis, and interpretation.

Daniel Sargent, Sumithra Mandrekar, and Ann Oberg
Division of Biostatistics
Mayo Clinic
Rochester, Minnesota, U.S.A.

FARAWAY, J. J. **Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models**. Chapman & Hall/CRC, Boca Raton, Florida, 2006. x + 301 pp. US$79.95, ISBN 1-58488-424-X.

Extending the Linear Model with R uses the free software R to teach techniques of GLMs, mixed effects models, and nonparametric regression models. The book is aimed at people who already possess some knowledge of linear models and want to extend their knowledge and practice of statistical methodologies beyond linear models. As the Preface states, the book presents three extensions to the linear model framework: the scale of the response, the distribution of the random components, and the model for the predictors. Basic knowledge of R and a general understanding of statistical inference such as likelihood theory, hypothesis tests, and confidence intervals are necessary to appreciate the book. In fact, there is a quick introduction to likelihood theory and information about R for this book in the appendices. A reader is advised to consult some other books for a more detailed introduction to both statistical inference and R as suggested by the author. The book is very comprehensibly written and can therefore be recommended for beginners in GLMs, mixed effects models, and nonparametric regression models. Theory is captured in a compact manner and a more extensive bibliography is provided at the end of each chapter for further reading. The book contains 14 chapters that can be organized into 4 sections.

The first section reviews the main points of linear models in R, including descriptive statistics, estimation, inference, and diagnostics with examples.

The second section discusses GLMs in detail. Treatment for binomial data, count data, contingency tables, and multinomial data is addressed in different chapters followed by a general explanation of GLMs and model diagnostics as well as joint modeling of the mean and the dispersion of the response and quasi-GLMs. Besides fitting GLMs with R, the book presents methods to remedy problems that arise when a linear model is wrongly fit to the data. These topics cover about 50% of the book. Model diagnostics for GLMs in this book are very helpful in that they are rarely available in other books. This part of the book also has a discussion of model comparison and variable selection problems.

The third section focuses on estimation and inference for designs in random and mixed effects models, such as split plots, nested effects, crossed effects, and hierarchical (multilevel) models. It covers standard methodology for both normal and nonnormal responses.

The fourth section presents nonparametric regression methods. Topics covered include kernel estimators, splines, local polynomials, and wavelets for a single predictor, and additive models, GAM, classification and regression trees, and neural networks for multiple predictors. Comparison of methods and the curse of dimensionality problem in high dimensions are described briefly yet effectively.

This is a very pleasant book to read. It clearly demonstrates the different methods available and in which situations each one applies. It covers almost all of the standard topics beyond linear models that a graduate student in statistics should know. It also includes discussion of topics such as model diagnostics, rarely addressed in books of this type. The presentation incorporates an abundance of well-chosen examples that clarify both the use of each technique and the conclusions one can draw from the results. Throughout the book, the author explains very clearly about the scope of different methods and which method is preferable. Even though the book emphasizes computer software and its applications, theoretical parts are well explained. The book concentrates on the essential formulas and places more emphasis on the statistical reasoning beyond them. This provides the reader with an adequate theoretical background and helps to incorporate theory into a more general setting. Additional exercises are provided at the end of each chapter, which is very convenient for lecturers. Moreover, keywords are often highlighted and the layout of the book is clear. Most of the commands in the book can be downloaded from the book's web page. The R package *faraway* is necessary when submitting these commands in R, which includes additional functions and the data sets used in this book. There are a few typographical errors. Known errors as well as software modifications that occur due to the new development of R are listed in the errata that can be downloaded from the book's web page.

In summary, this book is highly recommendable as a textbook for computational courses on statistical methodologies beyond linear models. It is also an excellent supplementary book for courses on the theory of GLMs or statistical data mining, and for applied statisticians who want to explore methods for analyzing data with R for which ordinary linear model analysis is not appropriate.

Haiyan Wang
Department of Statistics
Kansas State University
Manhattan, Kansas, U.S.A.

BRIEF REPORTS BY THE EDITOR

DEAN, A. and LEWIS, S. (eds) **Screening: Methods for Experimentation in Industry, Drug Discovery, and Genetics.** Springer, New York, 2006. xv + 332 pp. $79.95/€71.64, ISBN 0-387-28013-8.

Design and analysis of factor-screening experiments is an area traditionally studied in the context of applications in the physical and industrial sciences and in engineering. With the growing use of microarrays and other large-scale data-collection methods, however, interest in screening experiments has crept into the life sciences. This book is an edited collection of articles written by an exceptional panel of experts in the design and analysis of screening experiments of varying kinds. Chapters 3–6 are likely to be of greatest interest to readers of *Biometrics*. These chapters cover experiments for blood screening (Jacqueline Hughes-Oliver), screening of chemical compounds in pharmaceutical drug discovery (David Jesse Cummins, as well as Hughes-Oliver), and gene expression through microarrays (Paola Sebastiani and coauthors; Jason Hsu and coauthors). Other chapters have a more traditional industrial flavor, written in the context of using experiments to select important factors from a large pool of experimental factors.

These four chapters, as well as others in the book, are generally overviews and introductions, rather than detailed research contributions. Furthermore, the book is structured so that the chapters mostly follow in a logical progression, but individual chapters are sufficiently self-contained that a person can read the chapters in any order or read only selected chapters. A cover-to-cover reading finds occasional redundancy, but often with the subject approached from different directions by authors with different perspectives.

Due to the nontechnical nature of most of the chapters, this book is a great place for interested readers to start learning about the subject of screening experiments. Each chapter is concluded with a listing of references, so that finding follow-up reading is easy. I can genuinely recommend this book to statisticians of all levels and to numerically minded practitioners. I would especially suggest the book to any graduate student embarking on a research program on some topic relating to screening experiments.

EVERITT, B. S. and HOTHORN, T. **A Handbook of Statistical Analyses Using R.** Chapman & Hall/CRC, Boca Raton, Florida, 2006. xv + 275 pp. $49.95/£29.99, ISBN 1-58488-539-4.

Another in the series of Handbook titles from Brian Everitt and coauthors, this book is intended to give "brief and straightforward descriptions of how to conduct a range of statistical analyses using R." The book is clearly meant to help a true beginner to get started with the R package. It begins appropriately with a chapter presenting a description of R and installation instructions, the help (simple help) and vignette (detailed help) commands, and other available documentation. This chapter also discusses basic data-handling techniques and methods for summarizing data.

The remainder of the book consists of 14 chapters, each of which describes a different type of analysis. Each chapter is organized in the same fashion. Example data sets and associated questions are presented first. A brief discussion of some of the fundamental methods for the nominal analysis comes next. The bulk of the chapter consists of basic R analysis, using a limited set of options, and interpretation of the subsequent output. Each chapter is then concluded by a few exercises. The chapters are generally well laid out and easy to understand.

The book covers ANOVA/MANOVA, several forms of regression, an assortment of multivariate analyses, and various other forms of statistical analysis. The selection of methods presented in each chapter is necessarily limited due to the small size and wide-ranging scope of the book, but presumably users can investigate the R options for their favorite analyses on their own. The book is not intended to teach the analysis methods; the descriptions of the analysis methods are too brief for anything other than to help an experienced analyst to orient her- or himself with the subsequent R commands and output. For the experienced analyst wanting to learn R, this book is a useful, compact introduction.

HO, R. **Handbook of Univariate and Multivariate Data Analyses and Interpretation with SPSS.** Chapman & Hall/CRC, Boca Raton, Florida, 2006. xiv + 406 pp. $89.95/£49.99, ISBN 1-58488-602-1.

Ho's book focuses on statistical tests. His intent is to help practitioners in selecting appropriate tests and then implementing them in SPSS. He deliberately uses both the menu-driven interface and syntax files, the latter included on the grounds that these files are better for repeating analyses, complex data manipulations, and student understanding. The book starts with a brief introduction to (or refresher on) the ideas behind hypothesis testing. It then turns to an introduction to SPSS, including data handling and writing a syntax file. An oddly placed chapter follows on the specialized topic of summarizing multiple-response data from survey questions where subjects may select more than one response from a list of categories. The meat of the book is Chapters 4–16, each of which covers a different set of statistical methods. Topics include elementary methods such as *t*-tests, ANOVA, and regression, as well as more specialized methods that are commonly used in the social sciences such as factor analysis, instrument reliability, and structural equation modeling. Various tests used with categorical data are strangely included in a concluding chapter on nonparametrics.

The approach to each chapter follows the same basic sequence. Each chapter begins with a very brief description of the goals of the nominal topic and a checklist of its basic assumptions or requirements. Next comes an example data set with useful instructions for preparing the data for SPSS. The example is then analyzed, click by click, using the menu-driven interface, and then in a separate section using the corresponding syntax statements. All resulting output tables are displayed, and the chapter is concluded with a discussion of

what the output means in the context of the example. The click-by-click instructions would clearly be useful for beginners to SPSS, but might be a little slow moving for more experienced users. Also, listing *all* of the output first and discussing it in a separate section several pages later is realistic in that it mimics the process of sifting through pages of output, but it is rather cumbersome to have to constantly flip back and forth between the output and its interpretation.

The examples and methods all have a strong social-science flavor, which is consistent with the aims of the book. The descriptions of the statistical methods are not intended to teach the methods, per Se, because they are far too tersely stated to be useful to a statistical novice. This book would therefore seem to be most appropriate for statisticians or practitioners in the social sciences who are already familiar with the statistical methods covered and want simply to learn to implement them in SPSS. Secondarily, the book can help more experienced SPSS users who want to learn to write syntax files. It should be noted that the version(s) of SPSS to which this book applies is not stated anywhere, so it is unclear which versions previous to those available at the time of publication may be compatible with this book.

NAGL, S. (ed) **Cancer Bioinformatics: From Therapy Design to Treatment.** John Wiley and Sons, Chichester, U.K., 2006. xii + 287 pp. $120.00/£65.00, ISBN 0-470-86304-8.

This is an edited volume of chapters discussing current challenges in cancer systems biology, with a special emphasis on the ways in which data can and will be used to aid cancer research. In an effort to make the mathematical and computational aspects of cancer bioinformatics "more transparent and accessible to all," the book is almost completely nonmathematical. Its 14 chapters are discussions and surface treatments of their respective topics. The chapters are arranged in five sections. The first, "Cancer Systems," contains three chapters that lay out an overview of cancer bioinformatics, describe some of the biology of cancer, and stress the importance of an integrated approach to cancer research that brings together information from diverse sources. The second section, "*In silico* Models," contains four chapters that describe some of the efforts to model cancer and cancer therapies. In particular, empirical and mechanistic models for tumor growth and vascularization are discussed, as are some of the challenges that must be overcome in cancer modeling. The role of computer simulation of tumor growth and response to therapy is considered, as well as the potential benefits of modeling the three-dimensional structure of cancer molecules and of structure-based drug design. A two-chapter section on "*In vivo* Models" follows, discussing some of the efforts made by a variety of research groups to develop databases for such things as mouse tumor biology, clinical and epidemiological records, and diverse model systems. The two chapters in the section called "Data" deal largely with issues involved in integrating databases of clinical, genomic, and other cancer data to facilitate mining and discovery. Finally, "Ethics" contains three chapters covering topics such as risk assessment in software design, use of electronic patient-identifiable data, and understanding gene-induced variation in response to drug therapies.

Because of the low mathematical level of the book, it is not really intended to be a research text on any of these subjects. Rather, it provides a readable overview of the subjects (provided that the reader has at least some knowledge of basic biology, cancer, genomics, and the notions of modeling and data collection). It would be good reading for anyone entering into some aspect of cancer research, whether it is biological, mathematical, or computational, if she or he wanted to get a view of the "big picture" behind informatics-based cancer research.