BIOMETRICS 59, 735–740
September 2003

BOOK REVIEWS

EDITOR:
M. S. RIDOUT

YAN, W. and KANG, M. S. **GGE Biplot Analysis: A Graphical Tool for Breeders, Geneticists, and Agronomists.** CRC Press, Boca Raton, Florida, 2002. ix + 271 pp. $119.95/£79.99. ISBN 0-8493-1338-4.

This book focuses on the analysis of genotype by environment interaction (GEI), particularly as observed in multi-environment trial (MET) data from agricultural experimentation. Importantly, the authors interpret GEI in concert with the genotype main effect (G), as production level cannot be ignored. An integrated graphical analysis is achieved using the genotype and the genotype by environment (GGE) biplot methodology available in Windows application software developed by Yan (2001).

The objective is to demonstrate the power of this graphical analysis and enable other scientists to effectively use the methodology and software to gain more insight into their own data. I believe that Yan and Kang have achieved this objective and that someone who has not used this ordination approach before could do so with this book as their guide.

The contents are organized into three sections: Section I, GEI and stability analysis (Chapters 1 and 2); Section II, GEI biplot and MET data analysis (Chapters 3 to 5); and Section III, GGE biplot software and applications in analyzing other types of two-way data (Chapters 6 through 11). While the first section gives a comprehensive introduction to GEI, the second section is the heart of the book. The underlying theory of the biplot, originally introduced by Gabriel (1971), is discussed and it is shown how the visual presentation of the data in the form of a biplot can be interpreted to answer numerous questions. The choices that have to be made to produce a biplot are clearly explained and recommendations on what should be used to answer particular questions are put forward in a concise manner. These are demonstrated by using the biplot technique to visualize MET data to (1) understand the target environment, (2) evaluate cultivars, (3) evaluate test locations, and (4) understand the underlying causes of GEI. The first chapter of the third section explains the GGE biplot software, including terminology, data preparation, structure, analysis options, and output options (with a wide variety of publication-quality graphics easily produced). Each of the remaining chapters illustrates the analysis of data from different plant-breeding situations.

There are a few minor typographical errors, particularly in relation to the software. It appears that a different version was sometimes being referred to in the text from the version being illustrated in the figures. Fortunately, this does not detract from the general clarity and well laid out nature of the book. It is quite easy to read, with the summaries at the beginning of each chapter being most helpful.

A major advantage of the biplot methodology is that it can be applied to any two-way data with multiple entries and multiple attributes in order to obtain a better understanding of the relationships among entries, relationships among attributes, and the interaction between entries and attributes. Hence the material in this book will be of benefit to any scientist working with such data. I believe no other book covers this methodology and software in equivalent depth or breadth. Note that the software does not come with the book. While I use biplots extensively, I must admit that I have not yet used the GGE biplot software.

I have no hesitation in recommending this book to breeders, geneticists, agronomists, and ecologists, and to biometricians working with them. The biplot methodology greatly enhances the ability to understand and interpret two-way data (of the type described above) and this graphical analysis can be readily implemented with the GGE biplot software.

References

Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika* **58,** 453–467.

Yan, W. (2001). GGEBiplot—a Windows application for graphical analysis of multi-environment trial data and other types of two-way data. *Agronomy Journal* **93,** 1111–1118.

K. E. Basford
School of Land and Food Sciences
University of Queensland, Australia

CHOW, S.-C. and SHAO, J. **Statistics in Drug Research: Methodologies and Recent Developments.** Marcel Dekker, New York, 2002. xiv + 367 pp. $165.00/ £114.00. ISBN 0-8247-0763-X.

Statistics in Drug Research discusses a variety of statistical topics relevant to the drug development process. Beyond clinical statistics (represented by topics like bioavailability and bioequivalence, therapeutic equivalence and noninferiority, analysis of incomplete data, meta-analyses, quality of life, and medical imaging), it covers nonclinical statistics (pharmaceutical validation, dissolution, and stability testing) as well as a general introduction briefly describing the drug development process.

The aim and scope of the book is to "provide a well-balanced summarization of current and emerging practical issues and the corresponding statistical methodologies." When measured against this goal, the book fails to achieve what the authors promise. The selection of topics appears to be arbitrary, and it seems to reflect the authors' research interests more than anything else. The selection of nonclinical and preclinical topics is very limited; important areas such as toxicology and carcinogenicity studies are left out entirely. In the clinical area, some recent hot topics are discussed in a very superficial way, like the choice of control group in clinical trials, or are missing entirely, for example, QTc studies. There is an entire section on clinical imaging, but other disease areas with interesting statistical aspects are left out. A methodological aspect that may have deserved more attention is adaptive design and sample size reestimation.

A similar comment as for the table of contents applies to some of the chapters of the book: the content often appears to be a reflection of the authors' research interests, rather than a balanced overview of the topic. For example in the chapters on stability analysis and on bioequivalence and bioavailability, the majority of references are to papers by S.-C. Chow and J. Shao themselves. Having said this, I also have to say that I found the chapter on bioequivalence and bioavailability particularly interesting and good. It provides a clear overview of average, population, and individual bioequivalence, and can serve as a useful starting point for anyone who wants to learn more about this area. Equally, the chapter on stability analysis is well written and provides good ideas for someone interested in this area.

Some chapters in the book seem to lack focus. As an example, take the chapter on therapeutic equivalence and noninferiority. I think that there is a lot that can be said about the choice of control groups, and the choice of delta. But the choice of control group is mentioned at various points of the chapter, without a real systematic discussion. The choice of delta is hardly discussed at all, and I find that the technical discussions on hypotheses and designs do not add too much to the topic. Similar comments apply to the chapter on randomization and blinding, which appears to merely scratch the surface.

In summary, *Statistics in Drug Research* is a selection of topics relevant to a statistician in the pharmaceutical industry. Some of the chapters are well written and provide good ideas to colleagues interested in these areas. Other chapters are less well written and appear superficial. The book discusses some topics of current interest—however, other interesting recent developments are not covered in the book.

G. Heimann
Clinical Statistics
Pfizer Global Research and Development
Sandwich, U.K.

RAMSAY, J. O. and SILVERMAN, B. W. **Applied Functional Data Analysis: Methods and Case Studies.** Springer, New York, 2002. x + 190 pp. $39.95/£42.00. ISBN 0-387-95414-7 (pbk).

In their simplest form, *functional data* extend the notion of a sequence of independent numerical observations to a sequence of independent functional observations. That is, for each of a number of independent units, we observe a single function. Such functions are (frequently, but not always) continuous, and often sampled at discrete points. The goals of functional data analysis (FDA) are to extend the usual aims of statistics to this new setting while retaining the functional identity of the data.

Applied Functional Data Analysis is a companion volume to Ramsay and Silverman's 1997 publication *Functional Data Analysis.* While the earlier book provides a thematic introduction to the ideas, mathematics, and statistics of functional data, the focus of this new volume is applications. The authors address an impressive range of problems with a stated goal "to demonstrate how large is the potential scope of functional data." More importantly, these fascinating case studies encourage readers to develop insight into "functional thinking" about both statistical methods and substantive problems.

An introduction and eleven case studies are presented in 12 chapters. Each case study is described by a short case background, includes sections on statistical issues and analysis, and is summarized by a valuable "What have we seen?" section. Mathematical descriptions are included in the text, but many technical details are postponed to chapter's end. An exceptionally useful feature is a companion website (`http://www.maths.bris.ac.uk/~bernard/fdacasebook/`) linking to computational details for each chapter. Indeed, most chapters have links to the original data, a detailed description of the data analysis, and S-PLUS or MATLAB code for performing the analysis.

The style of presentation encourages accessibility of the material to quantitative subject-matter specialists, while also serving trained statisticians. This well-written book is targeted toward both groups. Students coming to functional data analysis for the first time will find this text very helpful. However, the new volume is best viewed as a supplement to the authors' 1997 publication. Although the authors claim that neither book is a prerequisite for the other, only statistically sophisticated readers will be able to proceed directly from the applications book to a full-scale analysis of a new problem.

Because the book relies on case studies to carry the theme, a summary of the problems and associated statistical methods is warranted.

Chapter 1: Introduction—A brief introduction to the ideas of FDA, and a summary of the case studies discussed in the book.

Chapter 2: Life Course Data in Criminology—The data track the number of arrests for each of 400 individuals between the ages of 11 and 35. FDA methods include forming functions from discrete data, regularization, functional principal components analysis (PCA), and calculations involving basis expansions.

Chapter 3: The Nondurable Goods Index—This study tracks seasonal variation in the economic index for the U.S. between 1920 and 2000. The authors introduce and interpret phase-plane plots (a plot of index acceleration vs. velocity) as an analysis tool to describe seasonal dynamics. They also demonstrate smoothing to estimate first and second derivatives from noisy data.

Chapter 4: Bone Shapes from a Paleopathology Study—The data are femur bone shapes characterized by landmarks. The analysis uses functional PCA (with varimax rotation) to identify modes of variability, and graphical display of results.

Chapter 5: Modeling Reaction Time Distributions—This study compares reaction time distributions for children diagnosed with attention deficit hyperactive disorder vs. matched controls. Here the functional data are the estimated probability densities, and variation in these densities is explored via PCA.

Chapter 6: Zooming in on Human Growth—Childrens' height records and monotonic smoothing are used to characterize growth. Curve registration is introduced to accommodate phase differences in growth.

Chapter 7: Time Warping Handwriting and Weather Records—Examples of handwriting and weather records are used to further develop ideas of curve registration.

Chapter 8: How Do Bone Shapes Indicate Arthritis?—The bone shapes from Chapter 4 are revisited using contour information rather than landmark locations. PCA of vector functions and functional discriminant analysis are used to distinguish arthritic and nonarthritic bone shapes.

Chapter 9: Functional Models for Test Items—FDA methods are used to model a latent trait in the performance of 5000 candidates to 60 items on a test of mathematics achievement. PCA applied to log odds-ratio functions is used to identify test items that discriminate between candidates of different ability, and to assess test fairness.

Chapter 10: Predicting Lip Acceleration from Electromyography—Functional extensions of correlation and linear models are used to examine the mechanism relating measurements of neural activity and lip movement.

Chapter 11: The Dynamics of Handwriting Printed Characters—A dynamic model is developed for the handwriting data of Chapter 7. This study demonstrates a novel method for estimating a differential equation model from noisy data. Further, it suggests how these models might be used to identify an individual from a handwriting sample.

Chapter 12: A Differential Equation for Juggling—A coupled, three-dimensional differential equation is used to model fingertip velocity during juggling.

Formal inference procedures are notably absent from this new text. Indeed, most inference is treated informally (e.g., by interpretation of graphical summaries). Practitioners would benefit from examples of more quantitative inference procedures such as functional extensions of t- and F-tests (introduced in the authors' earlier text), or permutation-based methods. Certainly, inference methods for FDA are in their infancy, and this immaturity likely contributes to their omission. However, a discussion of currently available approaches, and their pitfalls, would be beneficial.

Overall, this is a very interesting, readable book showing the potential scope of FDA. It will be useful to anyone wishing to develop insight in this emerging field.

REFERENCE

Ramsay, J. O. and Silverman, B. W. (1997). *Functional Data Analysis.* Springer, New York.

D. BILLHEIMER
Division of Biostatistics
Vanderbilt University
Nashville, Tennessee, U.S.A.

PAN, J. X. and FANG, K. T. **Growth Curve Models and Statistical Diagnostics.** Springer, New York, 2002. xvii + 387 pp. $79.95/£55.00. ISBN 0-387-95053-2.

This book provides a comprehensive and nicely structured introduction to some recent developments in multivariate diagnostic techniques, with emphasis on the theory and applications to growth curve models (GCMs). It is intended for scientists or researchers who have a basic knowledge of multivariate analysis and analysis of variance. This book gives not only theoretical discussions of modeling data, but also shows extensively how to detect outliers and identify influential observations.

The basic diagnostic principles are discussed within the classical likelihood-based and Bayesian frameworks. The likelihood displacement approach proposed by Cook and Weisberg (1982) plays an important role in the former framework, whereas Bayesian Kullback-Leibler divergence in information theory, as suggested by Box and Tiao (1968), serves as a basis for the Bayesian framework. Within the two frameworks, the authors show the ways to apply multivariate diagnostic techniques to the fitted model. Also, several computer programs that implement various calculations described in the book have been written in S-PLUS and GENSTAT and can be found on the authors' web pages.

The book comprises seven chapters. Chapter 1 gives a short introduction to existing models and the related model selection criteria with respect to possible covariance structures. Chapter 2 introduces the fundamental concepts of GCMs and illustrates these with several practical examples. These examples recur in later chapters and are reanalyzed with the diagnostic techniques as they are introduced. Chapter 3 is concerned with deriving the maximum likelihood (ML) estimates and restricted maximum likelihood (REML) estimates of parameters in the context of a growth curve model with Rao's simple covariance structure (SCS) and unstructured covariance (UC). Expectations, variance-covariance matrices, and the distributions of ML estimates are also discussed. Chapters 4 and 5, within the likelihood framework, are devoted to exploring outliers or influential observations in the GCMs with SCS and UC, respectively. Chapter 4 uses the case deletion (or global influence) approach to build up multiple outlier detection criteria based on the concept of likelihood displacement. Chapter 5 focuses on the use of local influence to diagnose the adequacy of GCMs. This approach can tackle *masking* and *swamping* phenomena as well. Finally, parallel developments of diagnostic tools within a Bayesian framework are covered in Chapters 6 and 7. As addressed by Pan, Fang, and von Rosen (1999), Bayesian diagnostics in general reveal more information on influential observations than the likelihood-based approach. Under a noninformative prior distribution, the fundamental idea is to replace the likelihood displacement with the Kullback-Leibler divergence (or Bayesian entropy).

All chapters provide the details of theoretical proofs, bibliographic notes, and further references on the topic. Moreover, all examples are explained thoroughly and the use of well-designed graphics really facilitates understanding for readers. I definitely recommend the book to researchers and graduate students as a guide for implementing statistical diagnostic tools in growth data.

References

Box, G. E. P. and Tiao, G. C. (1986). A Bayesian approach to some outlier problems. *Biometrika*, **55,** 119–129.

Cook, R. D. and Weisberg, S. (1982). *Residual and influence in regression.* Chapman and Hall, New York.

Pan, J. X., Fang, K. T., and von Rosen, D. (1999). Bayesian local influence for growth curve model with unstructured covariance. *Biometrical Journal* **41,** 641–658.

J. C. Lee
National Chaio Tung University
Taiwan

QUINN, G. P. and KEOUGH, M. J. **Experimental Design and Data Analysis for Biologists.** Cambridge University Press, Cambridge, 2002. xvii + 537 pp. $45.00/£29.95. ISBN 0-521-00976–6 (pbk).

The title of this textbook suggests that it might be of general interest and value to a wide range of biologists requiring support and advice in the design and analysis of their experimental work. A quick glance, however, at the background of the authors (both have research interests in ecology and environmental sciences), and at most of the examples used throughout the book, gives the feeling that the book has been mis-titled. A specific reference to "environmental scientists" in the opening chapter, and the approach taken in introducing many of the statistical methods included, reinforces the view that it would really only be of value to those working in ecology and environmental sciences. This is not a criticism of the content of the book, just a feeling that the title is a little misleading.

The early chapters provide an introduction to the scientific method and the role of statistics and probability, and a revision of ideas of estimation (Chapter 2) and hypothesis testing (Chapter 3). Although these chapters are illustrated with a number of examples (an excellent feature of the whole book), both areas are covered quite quickly, and the authors state an assumption "that readers are familiar with basic statistical concepts." This basic knowledge is probably an absolute requirement, and I suspect that many biologists would struggle with both these and the subsequent chapters even having first attended an undergraduate service statistics course. Chapter 4 covers graphical exploration of data, including testing the assumptions of linear models. This chapter seems out of place, since the models themselves have not yet been covered, and the material might be better in later chapters, immediately following the introduction of linear models.

The remainder of the book, excluding the last chapter on the presentation of results, covers four broad statistical areas—two chapters on correlation and regression, six on experimental design and analysis of (co)variance, two on generalized linear models (specifically logistic regression and log-linear models for frequency data), and four on multivariate analysis. In the chapters on regression there is remarkably little on non-linear regression, but there are sections on smoothing, regression trees, and path analysis. Experimental design is approached from an observational, sampling approach rather than from the more traditional viewpoint of the planned field experiment, leading to an occasional unusual use of standard terminology—for example the randomized complete block design is referred to as an "unreplicated two-factor design." The authors also refer to "repeated measures" designs in these chapters, though from the analysis approach described I am not sure that these produce what I would consider to be conventional repeated measures data. The general introduction to generalized linear models is short, with the reader referred to a number of more specialist texts, concentrating on the two most commonly used GLMs, logistic regression for binary data and log-linear models for frequency data. The extent of the sections on multivariate analysis was a surprise for a nonspecialist text, and again emphasizes the focus of the text on ecological and environmental sciences where the use of such approaches is much more common than in many areas of biology.

The book is generally well laid out—the examples are easy to identify, being printed in boxes with a grey background. One criticism is that there are a number of points where particular methods or tests are used in examples prior to being introduced in the text, leaving the reader unsure as to whether he or she has missed something in the preceding pages. One nice feature is the inclusion at the end of every chapter of a summary of "General issues" raised in the chapter and "Hints for analysis." There are a number of typographical errors, notably associated with the labeling of figures or subsections.

In this book the authors aim to provide biologists with a better understanding of the statistical approaches used to analyze their data, both to aid them in the selection of the best approach for a particular data set (when working alone) and to aid the communication between biologist and consultant statistician (by improving the biologist's knowledge of the statistical language). The book covers a wide range of statistical approaches, and the use of examples (although primarily ecological or environmental) will certainly assist the understanding of these methods, but I fear that many biologists will be put off by the amount of algebraic notation included. The book will be of most value to statistically inclined ecologists or environmental scientists, but only if they already have a good basic grounding in statistics. I guess that this is only a subset of its intended audience. I suspect that it will not be of particular appeal to most readers of *Biometrics*.

A. Mead
Horticulture Research International
Wellesbourne, U.K.

AERTS, M., GEYS, H., MOLENBERGHS, G. and RYAN, L. M. (editors). **Topics in Modelling Clustered Data.** Chapman and Hall, Boca Raton, Florida, 2002. xv + 308 pp. \$69.95/£46.99. ISBN 1-58488-185-2.

This edited volume has arisen primarily as a result of a collaboration between biometricians at Limburgs Universitair Centrum and Harvard School of Public Health, and the majority of the twelve contributors, including all four of the editors, are from one or the other of these institutions. Although the title is general, the book is primarily concerned with modeling clustered *binary* data, particularly with application to developmental toxicity data. Many papers published in *Biometrics* are concerned with clustered binary data, and the book is therefore of considerable potential interest.

Edited collections of this sort can often be of variable quality, with unnecessary repetition and inconsistent notation. The editors of this collection are very aware of this problem, and have tried hard to achieve something that is closer to a standard monograph, for example by providing good cross-referencing. I think they have been successful in this.

After a brief introductory chapter, Chapter 2 describes a series of example data sets that are used throughout the book. The next two chapters cover general issues in modeling clustered data and families of models (primarily) for binary data. The latter covers conditionally specified models, marginal models, and cluster-specific models. These chapters cover a lot of ground in a fairly short space, and are therefore not ideally suited to beginners; but there is a wealth of material here for readers with some experience of the analysis of binary data.

The next three chapters of the book are concerned with inference. Although maximum likelihood methods provide an obvious approach to inference, one might seek alternatives either to improve robustness or because the likelihood is difficult to work with. Many models for clustered binary data fall into the latter category because of the presence of an awkward normalizing factor. This book discusses, in successive chapters, generalized estimating equation methods as a robust alternative and pseudolikelihood as a method for dealing with awkward likelihoods. A third chapter discusses analogues of the Wald, score, and likelihood ratio statistics for pseudolikelihood. These are excellent chapters, providing a succinct account of the general theory, details of application to binary data, worked examples, and simulations to look at questions of efficiency.

The next chapter looks at improving the flexibility of modeling covariates either parametrically, through the use of fractional polynomial models (i.e., extending standard polynomial models by allowing noninteger powers of a covariate, such as $x^{-1/2}$ or $x^{3/2}$), or nonparametrically, using local polynomial models. Chapter 9 then discusses a variety of methods for assessing the fit of a model. The next chapter is concerned with estimating safe levels of exposure based on dose-response modeling. This is a difficult area, partly because of the inherent problems of extrapolation. This chapter discusses many other issues, such as the choice between individual-based risk and litter-based risk. Estimation of safe doses arises again in the next chapter, which considers model misspecification. The authors discuss several interesting techniques for providing more robust inference when the underlying model may be misspecified in some way.

A short chapter on exact dose-response inference provides a method of testing for the presence of a dose response based on clustered binary data that is useful when data are sparse. It does not provide a general account of exact inference methods for binary data. A basic feature of clustered data regression problems is that covariates may exist at the cluster level, at the individual level, or both. This is the subject of Chapter 13. Chapter 14 considers multivariate data that comprise a mixture of continuous and discrete variables. In the context of developmental toxicology, for example, one variable might be binary, indicating the presence or absence of malformations, another might be the length or weight of the foetus.

Throughout the book, the authors have stressed that the methodology they describe is applicable to areas other than developmental toxicology and this is emphasized in the last chapter, which is concerned with modeling data from a complex hierarchical survey, the Belgian Health Interview Survey. The book closes with an appendix on the rather complicated constraints that arise when trying to develop models based on the Bahadur representation.

The book thus provides a nice mix of chapters on general inference issues and more specialized topics such as safe dose estimation. It is well written, and will be essential reading for anyone working on statistical problems in developmental toxicology, but beyond this I would recommend it to anyone with an interest in the analysis of binary data.

M. S. Ridout
University of Kent
Canterbury, U.K.

BRIEF REPORTS BY THE EDITOR

COLLETT, D. **Modelling Binary Data, 2nd edition.** Chapman and Hall, Boca Raton, Florida, 2003. xii + 387 pp. $59.95/£29.99. ISBN 1-58488-324–3 (pbk).

The first edition of this useful book was reviewed favorably in *Biometrics* **50,** pp. 587–588. The main change in this new edition is the introduction of two new chapters. The first is on mixed models for binary data. The second is on exact methods, and concentrates mainly on Fisher's exact test and on exact logistic regression with a single covariate. Other chapters have also been updated and revised, and in particular the examples no longer use the GLIM package. Discussion of software is left until the final chapter, which contains a useful section on implementing nonstandard models within standard software packages.

GELMAN, A. and NOLAN, D. **Teaching Statistics: A Bag of Tricks.** Oxford University Press, New York, 2002. xv + 299 pp. $40.00/£25.00. ISBN 0-19-857224-7 (pbk).

The idea of this book is that the reader should browse through the bag of tricks, many of which require active student participation, and select particular ideas that suit his or her own teaching style. There are a great many ideas here, and while some of them are well known or "obvious," I found many of them to be original and interesting. Much of the material is targeted at "basic" probability and statistics courses, with topics such as descriptive statistics, simple data collection, linear regression and correlation, etc. This is the subject of Part I of the book. Part II has a useful "How to do it" chapter, and a second chapter giving a detailed outline for an introductory statistics course. Part III has ideas for more advanced courses, such as decision theory and Bayesian methods, survey sampling, and stochastic processes. There are also ideas for student projects, and indeed the very last of these, exploring response surface methods through trying to optimize the design of a paper helicopter, is one that I've used successfully myself this year for a final-year undergraduate project, having come across it in another book by one of the authors (Nolan and Speed (2000). reviewed in *Biometrics* **57,** pp. 646–647). The authors write with great enthusiasm, and it is difficult to come away from the book without a resolve to attempt to enliven one's own teaching.

Reference

Nolan, D. and Speed, T. (2000). *Stat Labs: Mathematical Statistics Through Applications.* Springer, New York.

EDITOR'S NOTE

In the March 2003 issue, I included a brief report (p. 208) on the second edition of Burnham and Anderson's *Model Selection and Multimodel Inference,* the first edition of which was reviewed more fully in *Biometrics* **57,** pp. 320–321. In this report I included the parenthetical statement "assuming that the true model lies within the set of models considered." The authors have asked me to point out that they do not in fact make this assumption, and, on the contrary, assert in various places in the book that they do *not* assume that the set of models considered includes the true model. Indeed, they argue that "the concept of a 'true model' in biology seems of little utility." I apologize to the authors for this error.

On this apologetic note, I complete my term as book review editor for the journal. I should like to thank the many people who have provided book reviews during my period as editor, and wish my successor well.