Journal of Official Statistics
Vol. 6. No. 1, 1990. pp. 101–111
© Statistics Sweden

Book Reviews

Books for review are to be sent to the Book Review Editor Gösta Forsman, Department of Mathematics, University of Linköping, S-581 83 Linköping, Sweden.

Chaudhuri, A. and Vos, J.W.E., Unified Theory and Strategies of Survey Sampling. North-Holland, Amsterdam, 1988. ISBN 0-444-70357-8. xvii + 414 pp., $92.75.

Outline of contents:

Part A – Unified Theory of Sampling	No. of pages
1. Introduction	2
2. Principal notation	3
3. Design-based estimation	56
4. Admissibility and other optimality properties of sampling designs	4
5. Sufficiency and related concepts in survey sampling	8
6. Sample Survey and General Statistical Model	18
7. Further details on likelihood and Bayesian approach	11
8. Predictive approach	19
9. Robustness	5
10. Concluding remarks	2

Part B – Strategies of Sampling	No. of pages
1. Introduction and sampling	2
2. The Horvitz-Thompson estimator (HTE)	4
3. Subdivision of summary schemes	2
4. Sampling schemes for use of the HTE: $n = 2$	30
5. Sampling schemes for use of the HTE: general n	77
6. Sampling with other strategies	28
7. Alternative estimators	6
8. Ratio, product and regression estimators	11
9. Variance estimation	8
10. Special designs	15
11. Comparison of strategies	15

Part A of this book is essentially a literature review on inference in finite population sampling. It sets out deliberately to complement Cassel, Särndal, and Wretman (1977) and is therefore not comprehensive on all topics. Part B is mainly a literature review, largely overlapping with and updating Brewer and Hanif (1983). I will refer to these two comparison books as CSW and BH, and to the book under review as CV.

The most noticeable difference between CSW and part A of CV is in the greater use of references in the latter. There is also a considerable difference in the arrangement of material.

CSW is characterized by its logical presentation. After the introductory first chapter it deals with inference under the fixed population model, then inference under super-population models, and finally robustness as an issue relevant to both. Excluding the references and index, it contains 178 pages compared with Part A's 141. It lists 169 references, almost three quarters of which are also cited by CV. Citations of papers in the text, while highly clustered, average only one or two per page. CV's Part A (as its title suggests) leans heavily on the unified theory of survey sampling as developed by Godambe and his co-workers, notably Joshi and Thompson. Godambe's name, in particular, is mentioned almost 100 times and can be found on more than one third of the pages. Total references to published articles average about eight to the page. Inevitably this heavy emphasis on references makes Part A less readable and its presentation less logical than that of CSW.

One example of this shortcoming in presentation is the rather scattered treatment of admissibility and related issues such as hyperadmissibility and necessary bestness. These are first considered in Section 3 on Design-based estimation. Admissibility is defined on page 17. It dominates the discussion to page 36 but then gives way to other topics (labels, permutations of orderings, and minimax estimates). Necessary bestness and hyperadmissibility have a subsection of their own from pages 48 to 53. The final part of Section 3 deals with the use of superpopulation models. Section 4 on "Admissibility and other optimality properties of sampling designs" is on pages 61 to 65. The difference between the discussion of admissibility in Sections 3 and 4 is that the former deals with the admissibility of estimators for a given design and the latter with admissibility of designs for a given estimator (the HTE). (Since the HTE is a design-based estimator it is not at all clear why Section 4 should not be a subsection of Section 3. The same could be said of Section 5, which deals with sufficiency and related concepts.)

A second example of problems with logical presentation is CV's treatment of super-population models. There are basically three ways of using these models, distinguished by the fashion in which they interact (or fail to interact) with randomisation and design-based inference.

a. The comparative use.
Before 1970 super-population models were used almost exclusively to compare the accuracies of design-based estimators.
b. The predictive use.
During the early 1970s Royall and his co-workers started to advocate a purely predictive approach to sample survey inference, based essentially on a simple super-population model. Robustness against more general models was achieved by 'balancing' the sample; that is, by choosing it in such a way that the sample moments of the auxiliary variable were as close as possible to its corresponding population moments. This approach proved highly contentious, with most survey statisticians remaining sceptical as to whether, even with balancing, it was sufficiently robust against model breakdown.
c. The integrated use.
By the later 1970s there was an increasing awareness that super-population models could safely be used for more than comparisons of the accuracies of design-based estimators. Godambe (1976) was probably the first to suggest explicitly that samples should be chosen such that their design-based estimators would also have good super-population model-based properties. However the same notion also

seems to be implicit in the paper by Cassel, Särndal, and Wretman (1976) on generalised difference estimation.

(For a fuller description and an evaluation of these three uses see my review of this book in the *Australian Journal of Statistics* 31 No. 2, August 1989).

At first sight, CV's treatment of super-population models does seem to correspond to this threefold classification. Subsection 3.5 deals with the comparative use, Section 8 with the predictive use and Section 9, "Robustness," with the integrated use. But the concern of the authors for completeness of treatment seems to have driven them into putting part of the work on the integrated use of super-population models into Subsection 3.5 and most of the rest into Section 8, leaving hardly anything to write about in Section 9 when it comes to the robustness issue.

In summary, Part A is long on thoroughness but shorter on the logical arrangement of issues. This deficiency is exacerbated by the lack of any index, either for authors or for topics, making it rather difficult to find one's way around. It offers the already informed reader a wealth of information, but those who are not already familiar with the issues should read CSW first.

The coverage of Part B (Strategies and Sampling) and that of BH are very similar. Excluding bibliography and indexes, BH contains 146 pages compared with Part B's 208. It lists 229 references, nearly two-thirds of which are also referenced by CV. After some preliminaries, both books get down to describing procedures for sampling with unequal probabilities without replacement. BH lists 50 of these, CV adds at least another ten. Both sets of authors also compare and evaluate unequal probability sampling strategies and deal with multi-stage issues. It is only towards the end of each treatise that the differences in emphasis emerge clearly. The final chapter of BH is on an optimal sample strategy based on the integrated use of population models. Part B, on the other hand, deals with ratio, product and regression estimators, double sampling, sampling on successive occasions, and controlled sampling.

Four years ago, Hanif and I started (but had to abandon) work on a revision of BH. What we were then aiming for would have looked very much like Part B. While I cannot speak for my co-author at some 15 000 km distance, I must myself admit to a rather envious appreciation of these authors' achievement.

There is, however, one important issue which would have been included in such a revision and which is absent from CV. This is the search for a "perfect" selection scheme for sampling with unequal probabilities without replacement. Such a scheme would have the following requirements:

i. The number of units in the sample would be predetermined.
ii. It could be any number n subject only to $n \leqslant Z/Z_{\max}$ where $Z_{\max}$ is the measure of size of the largest unit and Z is the total measure of size for the population.
iii. The first order inclusion probabilities π_i would be exactly proportional to the measures of size Z_i (yielding an unbiased HTE).
iv. The process of sample selection would be simple.
v. The values ϕ_{ij}, defined by $\phi_{ij} = \pi_{ij}/(\pi_i\pi_j)$ where the π_{ij} are the second order inclusion probabilities, would be $\geqslant (n-1)/n$ – ensuring that the HTE was always at least as efficient as the Hansen-Hurwitz estimator (HHE) in the corresponding PPSWR strategy – but $\leqslant 1$, ensuring non-negativity of the Sen-Yates-Grundy estimator of variance (SYGEV). These limits for the ϕ_{ij} would also automatically ensure the variance of the SYGEV to be close to its minimum possible value.
vi. The values π_{ij} (and hence the SYGEV) would be easily calculable.
vii. Additional units could easily be added to the sample at any time, subject only to property (ii) above, making the scheme suitable for use in rotating samples.

What has changed since the early 1980s is that there is now a greater acceptance of model-based variance estimation. A model-based estimator to replace the SYGEV was suggested by BH (page 68). An equivalent to this estimator (using a double rather than single summation) was investigated by Kumar, Gupta and Agarwal (1985). They found it to have fairly small bias and to be somewhat more stable than the SYGEV, both properties holding over an extreme range of assumptions about heteroskedasticity.

If this variance estimator is now acceptable, requirement (vi) and most of requirement (v) cease to be necessary and the random systematic selection scheme becomes almost perfect. It falls short of perfection only in that for some atypical populations, the HTE is less efficient than the HHE with PPSWR.

In all CV has 525 references to 345 in the union of CSW and BH. Naturally many of the additional references are from the period 1982 onwards, but the majority are not. It seems that no one has yet brought about a fully comprehensive bibliography, either on the foundations of sampling theory or on sampling strategies; at least not since Murthy's (1967) textbook.

In summary, CV's claim to have brought out a book which in Part A complements CSW and in Part B provides an alternative and more up-to-date reference to BH is fully justified. Neither part is easy to read (nor is BH!) but this is largely because BH and CV are essentially books of reference, not self-contained courses of instruction like CSW. From that point of view I regard CV as valuable, but find the lack of subject and author indexes a distinct obstacle to its easy use.

References

Brewer, K.R.W. and Hanif, M. (1983). Sampling With Unequal Probabilities. Lecture Notes in Statistics, 15, Springer-Verlag, New York.

Cassel, C.M., Särndal, C.E., and Wretman, J.H. (1976). Some Results on Generalized Difference Estimation and Generalized Regression Estimation for Finite Populations. Biometrika, 63, 615–620.

Cassel, C.M., Särndal, C.E., and Wretman, J.H. (1977). Foundations of Inference in Survey Sampling. Wiley, New York.

Godambe, V.P. (1976). A Historical Perspective of the Recent Developments in the Theory of Sampling from Actual Populations. Technical Report, Indian Society of Agricultural Statistics, 1–12. (Referenced by CV but not sighted by reviewer.)

Kumar, P., Gupta, V.K., and Agarwal, S.K. (1985). On Variance Estimation in Unequal Probability Sampling. Australian Journal of Statistics, 27, 195–201.

Murthy, M.N. (1967). Sampling Theory and Methods. Calcutta: Statistical Publishing Society.

Kenneth R.W. Brewer
Australian Bureau of Agricultural and Resource Economics
Lyneham, A.C.T.
Australia

Chhikara, R.S. and Folks, J.L., The Inverse Gaussian Distribution: Theory, Methodology, and Applications. Marcel Dekker, Inc., New York, 1989. ISBN 0-8247-7997-5. viii + 213 pp., $79.95 (U.S. and Canada), 95.50 (all other countries).

The inverse Gaussian distribution is the first-passage-time distribution to a fixed barrier in a Wiener diffusion process (i.e., Brownian motion with drift). The distribution family has many attractive theoretical properties and has been applied widely, primarily in settings that require a model for positive right-skewed data and especially for duration data. Applications have included, for example, equipment failure times, strike durations, and lengths of hospital stay. The distribution has also been used as a mixture distribution for positivevalued parameters of common probability distributions, such as the Poisson distribution. The Poisson-inverse Gaussian mixture distribution, for instance, has been used to describe word frequency patterns, accident proneness of individuals, and species abundance. Although the derivation of the inverse Gaussian distribution dates from the early part of this century (Schrödinger 1915; Smoluchowski 1915), the first extensive investigation of its statistical properties began with the pioneering work of Tweedie (1941, 1957a, b). Inference characteristics of the family are generally tractable and frequently simple, often involving familiar chi-squared, t and F statistics.

The book aims to provide a survey of published research on the subject, up to the date of publication, for readers having interests in either research or applications. This book reminds one of Aitchison and Brown's invaluable reference book on the lognormal distribution (Aitchison and Brown 1957). The authors' contributions to the inverse Gaussian literature have been extensive and their intimate knowledge is reflected in their command of the subject in the book. The book, in an expanded format, updates their earlier survey paper (Folks and Chhikara 1978). Although broadly aimed, the book will be most useful to statisticians with a solid background in mathematical statistics. The topics covered in the book's eleven chapters are reflected in the chapter titles: Introduction, Properties of the Inverse Gaussian Distribution, Genesis, Certain Useful Transformations and Characterizations, Sampling and Estimation of Parameters, Significance Tests, Bayesian Inference, Regression Analysis, Life Testing and Reliability, Applications, and Additional Topics. The theoretical results are presented in a reasonably readable and organized form with mathematical derivations of key results, demonstrations of inference methods using several real data sets, and references to original literature. The book's coverage of the literature is quite complete to 1985 but a number of more recent publications are omitted, as may be seen from recent editions of the *Current Index to Statistics* compiled by the American Statistical Association and Institute of Mathematical Statistics. Even though one chapter is devoted to applications and real data sets are selectively employed throughout the book, the applied statistician, scientist, or engineer is not likely to find the motivation, explanation or discussion in the book adequate to give a thorough appreciation of potential applications of the distribution or of its limitations as a model for real-world phenomena. It is regrettable that the book contains numerous errors, mostly typographical, that will force a user to check formulas and results carefully for accuracy. The index is skimpy, making it difficult to locate key results. The book also lacks an authors' index. It is rare for a survey book on any subject to cover all of the favorite topics of a reviewer or to present the topics that are covered in a manner that is completely pleasing, and this book is no exception. For instance, one topic that is not covered in sufficient depth from this reviewer's point of view is the inverse Gaussian distribution as an exponential reproductive family [see Barndorff-Nielsen and Blæsild (1983)]. In contrast, the authors seem especially attracted to the topic of minimum variance unbiased estimation (MVUE) and give it a great deal of attention.

In the final analysis, however, the book will be a valuable reference for investigators who are concerned with finding elegant and sensible mathematical models for real-world phenomena and data. Official statisticians have good reason to look closely at the inverse Gaussian distribution family, and hence at this book. As an alternative to the normal distribution model for data analysis, the inverse Gaussian has much to offer, both in terms of methodology (for example, ratio estimation and regression methods) and as a realistic data model for vital, economic, and environmental statistics.

References

Aitchison, J. and Brown, J.A.C. (1957). The Lognormal Distribution: With Special Reference to its Uses in Economics. Cambridge: University Press.

Barndorff-Nielsen, O. and Blæsild, P. (1983). Reproductive Exponential Families. Annals of Statistics, 11, 770–782.

Folks, J.L. and Chhikara, R.S. (1978). The Inverse Gaussian Distribution and Its Statistical Application – A Review. Journal of the Royal Statistical Society, ser. B, 40, 263–289 (with discussion).

Schrödinger, E. von (1915). Zur Theorie Der Fall- und Steigversuche An Teilchen Mit Brownscher Bewegung. Physikalische Zeitschrift, 16, 289–295.

Smoluchowski, M.V. (1915). Notiz Über Die Berechnung Der Brownschen Molekularbewegung Bei Der Ehrenhaft-millikanschen Versuchsanordnung. Physikalische Zeitschrift, 16, 318–321.

Tweedie, M.C.K. (1941). A Mathematical Investigation of Some Electrophoretic Measurements on Colloids. M.Sc. Thesis, University of Reading, England.

Tweedie, M.C.K. (1957a). Statistical Properties of Inverse Gaussian Distributions I. Annals of Mathematical Statistics, 28, 362–377.

Tweedie, M.C.K. (1957b). Statistical Properties of Inverse Gaussian Distributions II. Annals of Mathematical Statistics, 28, 696–705.

G. Alex Whitmore
McGill University
Montreal
Canada

Cleveland, W.S. and McGill, M.E. (Eds.), Dynamic Graphics for Statistics. Wadsworth & Brooks/Cole, Belmont, California, 1988. ISBN 0-534-09144-X. xii + 424 pp., $44.95.

The book gives an overview over the history and the state-of-the-art in dynamic computer graphics for statistical applications. The essential characteristic of a dynamic graphical method is the direct manipulation of elements of a graph on a computer screen, where in high-performance implementations the screen can be changed nearly instantaneously by an input device like a mouse. Typical methods are 3-dimensional (3D)-rotation and brushing in scatterplot-matrices. In the future, probably most graphical data analysis will be carried out in a dynamic graphics environment, since the addition of dynamic capabilities to the methodology of traditional static data display provides an enormous increase in the power of graphical methods to convey information about data, and the price and availability of powerful statistical-computing environments are rapidly evolving in a direction that will permit the use of dynamic graphics.

Statisticians use roughly two kinds of methods for the analysis of data. If one has a great deal of prior knowledge about the data, and has confidence in what to expect in them, then statistical inference may be appropriate. But, if one is given data about which one has very little prior knowledge, the goal is primarily to look at the data to discover interesting features. In such *Exploratory Data Analysis* (Tukey 1977) dynamic computer graphics, emphasizing interaction and real-time-reaction, should be useful.

The book contains a collection of papers about dynamic graphics. Some papers are new, others are technical reports from the past and republications of journal articles. The writings cover a time period of about two decades, ranging from the late 1960s to the present. In particular, the book covers the historical development of what is called the PRIM-family of dynamic graphic software, which concentrated on the implementation of 3D-rotation. The first of these systems, called PRIM-9, was developed in 1972 (Fisherkeller, et al. (3)). Through a combination of Projection and Rotation PRIM-9 could display an arbitrary 3D-subspace of up to 9D data. Isolation and Masking were used to divide the data into subsets. One successor to PRIM-9 was PRIM-H (Harvard) in 1979/80 (Donoho, Huber, et al. (4)). A significant advantage of PRIM-H is its embedding into a flexible statistical package called ISP. ORION I (1980/1) went beyond 3D-scatterplots. It first implemented the idea of two linked views, where linking was realized by painting corresponding objects in both views with the same colour depending upon the object's distance from one actively changeable object in one of the views (Friedman, et al. (5); McDonald (7)). Moreover, ORION I was the first system interesting to non-specialists. The total cost for the computer needed was about $40,000 in 1982, whereas its predecessors had price tags of several hundreds of thousands of dollars. However, in the meantime there are even implementations of the PRIM family on very small personal computers. For example, a new version of ISP (1988), not described in the book, is available on IBMs, and MacSpin (Donoho, et al. (13)) on Macintoshs. Both kinds of computers are affordable to the whole statistical community. MacSpin offers a very easy to handle user interface, but no command language like ISP. On the other hand, the state-of-the-art PRIM-successor Plot-windows (Stuetzle (9)) is again only available on more exotic and expensive Symbolics Lisp-machines. – Since it is one of the forefathers of the PRIM-systems, Fowlkes q-q-plot-system is described in the book (Andrews, et al. (2)). Moreover, ideas for PRIM-system-design are described as well (Tukey (6)).

In the other mainstream of dynamic graphic techniques, mainly associated with the software system *S*, various versions of brushing in scatterplot-matrices were developed, a very elegant method of linking corresponding objects in different scatterplots (Becker, Cleveland, Wilks (1); Becker, Cleveland (8)). Meanwhile, there is effort to combine the two research directions "3D-rotation" and "linked scatterplots" in that both kinds of methods are implemented in newer systems (IRIS (Becker, Cleveland, Weil (10)), ISP (1988), Plot-windows (Stuetzle (9)), S-PLUS (1988)). Other very recent dynamic graphics systems are described in Carr, Nicholson (12) (4D exploring), Papathomas, Julesz (14) (portraying parallel planes), and Wang, Gugel (15) (General Motors system).

Last but not least, there are two articles which provide the link to more general systems including dynamic graphics just as a building block. Buja, et al. (11) discuss preconditions, merits, and drawbacks of the various dynamic graphics methods and propose a reasonable ordering for practical applications (viewing pipeline). Also, Young, et al. (16) stress the need of a combination of dimension reducing (algebraic) multivariate techniques (e.g., principal components analysis, canonical discriminant analysis) with the dynamic graphics techniques. Moreover, in both articles new graphical methods and software systems are proposed as well.

One of the merits of the book is the first "real" appearance of some "gray" literature up to now only available as technical reports. This way, there is the chance to fairly well trace the historical development. Unfortunately, the ordering of the papers appears to be a bit confusing. Instead of insisting upon "full historical" ordering, an ordering by topics might have been more instructive, e.g., first all the PRIM articles, then the articles related to brushing, then those articles where both kinds of methods appear together, then others. This appears to be particularly true since the historical ordering is in a way broken as it seems that some of the historical articles have been updated by the addition of newer references.

Moreover, since in the preface there is neither any motivation for the selection of the papers nor hints for further reading, the book is not more than a source of important articles, presenting them in one collection for the first time. Indeed, some of the topics seem to be over-represented, such as ORION I in two somewhat similar articles (Friedman et al. (5); McDonald (7)) and brushing, where the article of Becker, Cleveland (8) is mainly included in the "overview-article" of Becker, Cleveland, Wilks (1). Moreover, the article of Papathomas, Julesz (14) about meteological simulation and portraying parallel planes has little relation to statistics and might have been discarded. Also, the inclusion of the article of Tukey (6) about ideas for the construction of the apparently never realized system

PRIM81 appears to be questionable, since it is too technical and very hard to read. On the other hand, too little emphasis was layed upon evaluation of methods. Only two exceptions have to be mentioned: Becker, Cleveland, Weil (10), comparing brushing and rotation, and Buja, et al. (11), presenting the viewing pipeline.

Unfortunately, most of the book is not easily read. But this is in most cases not the fault of the authors. Indeed, in most of the articles there are illustrative examples stemming from real data, and many impressive snapshots of computer-screens and even stereograms together with a stereopticon viewer are included. Nevertheless, it remains difficult to describe graphical methods, and interactive computer controls by words.

The best motivation for reading such a book is by no doubt the planning of the analysis of a high-dimensional data set stemming from a statistical experiment with little a priori information about the relation between the involved variables. Unfortunately, the book is not thin, and little help is offered for a decision to read just some articles. From the point of view of an applied statistician, the following articles may be the most interesting: the overview by Becker, et al. (1) with discussion; the comparison of brushing and rotation by Becker, Cleveland, Weil (10); the viewing pipeline of Buja, et al. (11); the Macintosh-implementation MacSpin of Donoho, et al. (13); the SAS-related-implementations by Wang, Gugel (15), and Young, et al. (16). Moreover, some material not included in the book might be interesting also, e.g., ISP (1988) and S-PLUS (1988).

References

The articles in the book are referenced by the name(s) of the author(s) together with the paper number in the book.

ISP (1988). ISP-SGS/DGS User's Guide. Carlisle, MA: Artemis Systems.

S-PLUS (1988). S-PLUS User's Manual. Seattle, WA: Statistical Sciences.

Tukey, J.W. (1977). Exploratory Data Analysis. Reading, MA: Addison Wesley.

Claus Weihs
CIBA-GEIGY Ltd.
Basel
Switzerland

Råde, L. and Westergren, B., BETA Mathematics Handbook. Studentlitteratur and Chartwell-Bratt, Lund, 1988. ISBN 91-44-25051-7 and 0-86238-140-1. 412 pp.

This book is a synopsis of basic facts of college mathematics, including probability and statistics. The book is intended for students and professionals, and it mainly consists of detailed listings of useful formulas, definitions and theorems in a wide range of mathematical fields. Blackboard-style clarifying illustrations are also provided. There are no numerical tables on roots, logarithms, etc., which are nowadays available through calculators and computers. Differential and integral calculus, in one and several variables, are of course covered. So are topics like linear algebra, geometry, transforms, complex analysis, optimization, numerical analysis, and statistics.

Near the end of the book there is one chapter on Probability Theory and one on Statistics, in total comprising 89 pages. They have an emphasis on probability distributions, parameter estimation and tests of significance, but applications such as reliability of systems are also treated. There is even a one-page subsection on sampling. Of course, a book like this must largely be limited to basics, and one cannot expect the selection of topics to satisfy more specific needs. For instance some basic stochastic processes are dealt with, but statistical analysis of time series is not. Unlike other parts of the book, these two chapters are to a substantial part devoted to numerical tables, which supplement the presentation of important distributions and tests.

It must be stressed that this book is not a guide for the novice, but rather it is a convenient overview and reference-work for those who do have some mathematical maturity. It should however be useful to readers at various levels of mathematical background. The presentation is organized so as to largely follow a usual college curriculum, so having taken some mathematics courses, the reader may feel at home in the book by concentrating on the corresponding chapters. The chapter and section titles should on the whole give adequate orientation. Furthermore, the alphabetical index is rather detailed, comprising nearly 15 pages, and separate indexes of symbols and of functions are also provided. There are no references to other literature for further reading.

Apparently the book has been prepared with great care, and it contains a fairly great amount of detail. The notation and terminology mostly adhere to common standard, and explanations of the notation are usually easily found. Perhaps, though, the authors could have given somewhat more attention to alternative conventions occurring in other literature, which some readers may be used to. The natural logarithm is denoted by ln as a real function and by log as a complex function, seemingly a tacit compromise between two traditions.

Definitions and enunciations are generally stated in a terse style, saving words and enhancing the overview. Usually they will be intelligible to the intended readers, but they are apparently not well suited as a "gold standard" for word-by-word quotation. The general definition of a confidence interval is exceptional, being stated in a much too loose and ambiguous nonmathematical fashion, although it is followed by a more formal treatment of confidence intervals in terms of pivot variables.

Statisticians and statistics students with a basic college-level mathematical training should find much of this book relevant. The statistician may occasionally encounter various mathematical problems and may need to look up one or another fact, and in such situations the book may be quite handy.

Martin Ribe
Statistics Sweden
Stockholm
Sweden

Robertson, T., Wright, F.T., and Dykstra, R.L., Order Restricted Statistical Inference. John Wiley & Sons, Chichester, 1988. ISBN 0-471-91787-7. xix + 521 pp., £50.00.

Seventeen years ago, there appeared an extensive and very good book on isotonic regression entitled *Statistical Inference Under Order Restriction* by R.E. Barlow, D.J. Bartholomew, J.M. Bremner, and H.D. Brunk (1972). But it is now time for a new comprehensive book which covers new results. The present book by T. Robertson, F.T. Wright, and R.L. Dykstra follows closely the outline and basic ideas from Barlow et al. and the authors have not tried to avoid similarities. This is not a disadvantage but an advantage; as in sports, a winning team should not be changed.

The field of order restricted statistical inference is important and vivid, and the book by Robertson et al. covers the field very well, both with respect to theory and application.

The contents are as follows. After a general introduction to isotonic regression, there are chapters on tests of ordered hypotheses for the normal means case and approximations of related distributions. Tests of ordered hypotheses in other models are also treated including distribution-free methods. There is one chapter devoted to multinomial parameters. Those chapters present the theory and applications in an intermingled fashion. Then a chapter on duality, which requires some advanced mathematics, follows. Furthermore, there are two chapters on distributions that are subject to shape restrictions and conditional expectation given a σ-lattice. The discussions in these chapters are essentially based on Barlow et al. and finally there is a chapter with more recent contributions.

To read this type of reference book, the reader must have a very broad background in statistical theory and mathematics. On the other hand, the elementary parts are illustrated by examples and can be read by anyone working with statistical application. Nevertheless, most of the book requires a good knowledge of statistical theory and mathematics.

Parts of the book can be used for courses on order restricted inference, but its main use is as a modern reference book for researched, graduate students, and professional statisticians.

References

Barlow, R.E., Bartholomew, D.J., Bremner, J.M., and Brunk, H.D. (1972). Statistical Inference Under Order Restriction. London: John Wiley.

Sture Holm
University of Göteborg
Göteborg
Sweden

Rawlings, J.O., Applied Regression Analysis: A Research Tool. Wadsworth & Brooks/Cole, Pacific Grove, California, 1988. ISBN 0-534-09246-2. xvii + 553 pp., $49.95.

The preface of this book points out that it "is the outgrowth of several years' teaching an applied regression course to graduate students in the sciences . . . Its basic purpose . . . is to develop an understanding of least squares and related statistical methods without becoming excessively mathematical."

The text achieves these ends very well. The writing is concise, clear and the notation and terminology are consistent. Paragraphs have been kept short and the layout is attractive with wide pages and paragraph headings in the margins which facilitate finding passages of interest. Many chapters have brief introductions and summaries which should please students, and many short exercises which should please instructors. Answers to selected exercises are also supplied. Coverage is broad so that most topics of interest are referred to, but the size of the text is kept reasonable as the explanations are brief.

Coverage is rather traditional as shown by the chapter headings which are: Review of simple regression; Introduction to matrices; Multiple regression in matrix notation; Analysis of variance and quadratic forms; Case study: Five independent variables; Geometric interpretation of least squares; Model development: Selection of variables; Class variables in regression; Problem areas in least squares; Regression diagnostics; Transformation of variables; Collinearity; Case study: Collinearity problems; Response curve modelling; Case study: Response surface modelling; Analysis of unbalanced data; and Case study: Analysis of unbalanced data.

A good balance has been struck between explanations and examples, which are interspersed throughout. Students should find the text well organised and clear to follow. The exercises at the end of chapters do require answers rather than an encouragement to ask questions. This reviewer would plead for more considerations of the latter approach by supplying data sets. This was attempted to a modest degree in Brook and Arnold (1985). Instructors using this text would certainly appreciate other interesting and practical examples but there is the usual trade-off which must be made between realistic, if untidy, large data sets, on the one hand, and small sets, or aggregated data on the other, which are more amenable to short questions.

A practitioner will appreciate the clear explanations but will not find many helpful tips on how to carry out a statistical investigation of data. Data sets are given and analysed without checking to see if basic assumptions appear to be satisfied. The biomass data supplied by R.A. Linthurst are referred to repeatedly throughout the text yet the method of collection of these data is referred to only briefly. What were the aims of the biomass inquiry and were these best fulfilled by the collection of five samples at nine points? This implied structure of a 3×3 cross factor design, or, perhaps, a split-plot design, should suggest a particular approach to analysis. It is not until page 356 that the structure is considered but, then, only ridge regression and principal components are used to understand and reduce collinearity.

Some nodding acquaintance of Exploratory Data Analysis, EDA, would be appropriate in these days of computer ubiquitousness. The biomass data referred to above follow a positively skewed distribution so that a transformation may be appropriate before least squares analyses are carried out. The data also has a few typographical errors in the column labelled "Zn" so that the correlation coefficients are not as quoted. Such errors, as well as possible interesting trends, may be revealed by EDA. Indeed, it would appear to this reviewer that an "applied" statistician should counsel clients to think carefully about data before commencing an analysis.

Least squares regression is so prevalent and computer packages are so readily available that there is often a tendency to use this method almost without thinking. This text may not discourage this approach, for it is not until halfway through the text that nonnormality is discussed and then only briefly. Methods other than least squares receive scant attention. One of the reasons for the emphasis on least squares may be that the examples tend to be biological. Economic and time series are limited so that common econometric topics are briefly mentioned or ignored. Prediction is not discussed in detail. Perhaps, a book such as Hoaglin, Mosteller, and Tukey (1983) should be read in conjunction with Rawlings to remind the reader that much interest and research is continuing into EDA and robust methods.

Graphical displays from (micro)computers are available to most users of statistics that one may have expected to encounter more than appear in this book. Gabriel's biplot is used often but half-normal plots are not. Chambers, Cleveland, Kleiner, and Tukey (1983) do suggest a number of other graphical methods which could be used prior to, and after, traditional analyses.

The coverage of techniques such as analyses of residuals, influence statistics (four are examined) and collinearity diagnostics (three are examined) is good but few hints are given as to the situations where one may be preferred above another or which peculiarities in the data may lead to one being more effective than another.

Overall, this book is well written, succinct and virtually free from typographical errors. Instructors and students will appreciate the wealth of topics covered. Practitioners may wish that it was more "applied" by giving more advice on how to investigate data but they should find the clarity of explanations helpful with the notation and mathematics kept to a minimum.

References

Brook, R.J. and Arnold, G.C. (1985). Applied Regression and Experimental Design. New York: Marcel Dekker.

Chambers, J.M., Cleveland, W.S., Kleiner, R., and Tukey, P.A. (1983). Graphical Methods for Data Analysis. Belmont: Wadsworth.

Hoaglin, D.A., Mosteller, F., and Tukey, J.W. (eds.) (1983). Understanding Robust and Exploratory Data Analysis. New York: John Wiley.

Richard J. Brook
Massey University
Palmerston North
New Zealand

Printed in May 1990.