Journal of Official Statistics
Vol. 5. No. 2, 1989, pp. 191–203

Book Reviews

Books for review are to be sent to the Book Review Editor Jan Wretman, Statistical Research Unit, Statistics Sweden. S–115 81 Stockholm, Sweden.

Redfern, P., A Study of the Future of the Census of Population: Alternative Approaches. Office for Official Publications of the European Communities, Luxembourg, 1987, ISBN 92-825-7429-6, 269 pp., U.S.$14.50.

It is no mean feat to write a book about population censuses in the Western world that is at once a highly useful reference volume about current census-taking practices and at the same time stimulates the reader to think long and searchingly about whether and how such massive operations will be carried out in the future. Yet Philip Redfern has succeeded admirably in serving both descriptive and analytical functions in his review of current and alternative approaches to census taking in 15 European and North American nations.

Redfern's study was commissioned by the Statistical Office of the European Communities (Eurostat). He published an interim report in 1983 based on visits to six countries – Denmark, France, the Federal Republic of Germany, the Netherlands, Sweden, and the United Kingdom. A second interim report, published in 1985, included material based on visits to another eight countries – Belgium, Canada, Ireland, Italy, Luxembourg, Portugal, Spain, and the United States. The current 1987 volume contains updated information on all fourteen of these countries plus information obtained on a visit to Greece.

The 1987 report comprises a main body of text followed by 15 "annexes" that provide information on the population censuses of the countries included in the study. The material in the annexes covers not only census procedures but problems and issues that surround the census in each nation.

The annexes should prove to be an invaluable reference tool for census data users, statistical agency officials, and others. Averaging 11 pages of text each, they follow a similar organization. The entry for Belgium, for example, includes the following subheadings:

Summary;

Organization (i.e., the agency responsible for the census);
The objectives of the census: its frequency;
The census fieldwork: the role of the communes;
The census questionnaires: the population base (e.g., whether residence is defined de facto or de jure);
Census processing and dissemination;
Census coverage and quality;
The cost of the census;
The public's response to the census;
Registers of the resident population;
Intercensal population estimates;
Sample surveys; and
The future of the census of population.

Oddly, the entry for the United States omits discussion of intercensal population estimates which represent an important application of administrative records, notably vital statistics and tax data.

The entry for West Germany provides an extensive discussion of the problems that caused the postponement until 1987 of the census originally planned for 1981 and scheduled for 1983. Similarly, the entry for the Netherlands covers the problems that led to the cancellation of the 1981 census.

The entry for Denmark includes quite a different set of subheadings from the set used for most other countries:

Summary;
The development of population registers in Denmark;
The last conventional census of 1970;
Further steps toward a register system;
A Data Surveillance Authority;
The register-based census of population and housing of 1981;
Data tapes;
The role of sample surveys; and
The future of the register-based system.

These subheadings reflect the fact that Denmark, alone among Western nations, has made a transition from the traditional population census to a "census" based entirely on administrative records.

Obviously, the annexes do not provide an exhaustive characterization of the population census in each country. (Regrettably, they do not include lists of references that would have greatly increased their utility.) Nonetheless, they highlight major features in a format that encourages government officials and data users to deepen their understanding of their own census in the context of how it compares and contrasts with censuses in other countries.

As a user of U.S. census data, I was struck by the fact that the issue of population undercount, which is very "hot" in the United States politically, appears to be of relatively little concern, outside of the technical offices, in other countries. Conversely, where, in the U.S., the emphasis in questions about race and ethnicity has been to add more detailed categories in order to "recognize" newly-important minority groups, many other countries have never included a race question because of the perceived sensitivity of the topic.

The main body of the report presents Redfern's own cross-national comparative analysis based on his visits to the 15 countries and the conclusions that he draws regarding the future of the modern census. The text is organized into seven chapters:

Introduction;
The Conventional Census of Population: Merits and Demerits;
The Conventional Census of Population: Comparisons Among Countries and with the Register-Based Census;
The Role of Sample Surveys;
Register-Based Censuses;
The Administrative and Policy Reasons for Improving Registers; and
Summing Up.

Redfern takes as the starting point for his discussion two basic premises. The first (p. 6) is that "there is a growing demand for the statistics that the census of population provides: basic figures of the numbers of people and of their main demographic, social and economic characteristics, particularly for small geographic areas and small groups of the population." His second premise is that the role of the conventional census is being increasingly called into question for three major reasons: (1) high costs, (2) the growing perception that the

census is a burden on the public, an invasion of privacy, and a threat to confidentiality, and (3) the fact that alternative statistical methods have emerged during the past two or three decades for providing at least some of the information that has traditionally come from the census. These methods include sampling and the use of data from administrative records or registers.

In a brief review of the uses of sampling, Redfern notes the many advantages that sample surveys have over censuses. They are less costly overall (although not on a per person basis), less apt to attract public attention and hostility, and, consequently, can be conducted on a more frequent basis than censuses. They can also include a longer and more varied list of questions and obtain more accurate responses through use of skilled interviewers and greater attention to editing and coding. On the debit side, sample surveys invariably obtain less complete population coverage than censuses.

In considering sampling to replace the traditional census, however, all of these comparisons of pros and cons fall by the wayside, given that sample surveys by definition are not designed to obtain responses from all persons. Hence, Redfern concludes, as did the Panel on Decennial Census Methodology of the U.S. Committee on National Statistics (Citro and Cohen (1985)), that sample surveys cannot substitute for complete censuses because they cannot provide needed statistics for small geographic areas and population subgroups nationwide. Sample surveys can support censuses in many ways, as Redfern briefly notes in his section on "Some roles for samples."

Redfern gives his greatest attention to the possible use of registers or administrative records to substitute in whole or in part for traditional enumerative censuses. Redfern identifies the following advantages of register-based censuses: they can provide more frequent, perhaps annual, data; they are well suited to longitudinal studies, for example, of migration or marital and childbearing histories; they place little additional burden on the public, because most of the data have already been collected for other purposes; and they are less costly in that no fieldwork is involved, but only the marginal costs of gaining access to and linking the administrative records and extracting the statistics.

However, as Redfern drily notes (p. 64), "the fact that few countries outside Scandinavia have shown a positive interest in moving to a register-based census demonstrates that the method is seen to have serious disadvantages and difficulties that must be set against the advantages." These disadvantages include technical difficulties related to data content and quality that are inherent in register-based censuses even in countries with well-coordinated administrative records systems. Thus, registers will not be a reliable source of data on topics such as household composition, language, occupation, and other items for which there is no administrative need for the information. Adding such questions to the forms is not likely to be popular with either the respondent or the administrative agency and hence is likely to result in poor quality data. Conversely, items may be added or deleted and definitions change in ways that suit administrative needs but play havoc with the continuity of statistical series. Moreover, the time period to which administrative data refer may not be the same for all persons or all items.

In Scandinavia, where administrative records are coordinated through an effective central population register and personal reference numbers, the advantages of a register-based census are felt to outweigh the disadvantages. Countries like Denmark and Sweden are moving vigorously to implement and improve techniques for obtaining "census" information via administrative records.

The question is whether register-based censuses are technically feasible in countries that lack population registration and personal reference numbers. Redfern's answer in his 1983 report was a guardedly optimistic "yes," provided that the available administrative files, taken together, provide good coverage of the population. He felt that information on address, name, and

birthdate could be extracted and used to link records for the same individual.

In the 1987 report, Redfern concludes that his earlier optimism was misplaced. He supports the objective of moving toward a register-based census, but concludes that data items such as name and address are too unreliable to serve as cost-effective means of generating high-quality linkages. Hence, he states (p. 69) that, "instead of relying on the increasing power of computers to extract meaningful statistical information from an uncoordinated record system (that is, a system with a high level of 'noise'), it would be better to put right the weaknesses of the record system: in other words, adopt the coordinated record system of the Scandinavian kind." Redfern goes further to argue that there are public policy reasons for improving the reliability of information in administrative records systems, specifically on addresses, and for assigning personal reference numbers that apply across these systems. With these improvements, he argues, administrative functions would be less costly, more effective, and more fair. He argues that the vast majority of law-abiding citizens would support such improvements which would ensure that they obtain the benefits to which they are entitled and bear no more than their fair share of the tax burden.

I find myself in complete sympathy with Redfern's plea to improve administrative records systems for their own sakes. However I remain less convinced about the wisdom of working towards the goal in all Western countries of replacing enumerative methods with an administrative records census, as the concept is referred to in the United States (see Alvey and Scheuren (1982)). It may prove very difficult, as a practical matter, to integrate administrative records and improve them to the extent that is necessary to support their use in a census context. In this regard, Jabine and Scheuren (1987) paint a less than encouraging picture of the process in the United States in achieving six goals that were articulated early in the decade for enhancing statistical uses of administrative records. Even the most basic goal of standardizing reporting units and data items across systems has had rough sledding.

Moreover, there is the issue of the desirability of obtaining census data entirely from administrative records. In the United States, the traditional census has deep historical roots. Kruskal (1984) has aptly referred to the U.S. census as a "national ceremony." Hence, statistics compiled from administrative records, for use in such basic political functions as reapportionment of the U.S. Congress, would appear to carry less weight than the numbers obtained through current methods.

This is not to say that the traditional census cannot and should not be improved in many ways and made more cost-effective. The judicious use of sampling in some phases of census operations may very likely effect improvements as may the use of statistical techniques to adjust the counts for coverage error. Administrative records can also play a useful role, by facilitating many phases of census operations, supplementing the data content of the census on selected topics, and contributing to the evaluation of the coverage and quality of census data.

Although I do not always find myself agreeing with Redfern's arguments and conclusions about the future of the population census, I remain grateful to him for producing a most valuable and thought-provoking report.

References

Alvey, W. and Scheuren, F. (1982): Background for an Administrative Records Census. American Statistical Association, Proceedings of the Social Statistics Section.

Citro, C.F. and Cohen, M.L. (eds.) (1985): The Bicentennial Census: New Directions for Methodology in 1990. National Academy Press, Washington, D.C.

Jabine, T.B. and Scheuren, F. (1987): Statistical Uses of Administrative Records in the United States: Where Are We and Where Are We Going? Paper presented to the International Symposium on Statistical Uses of Administrative Data, Ottawa, Canada.

Kruskal, W. (1984): Remarks. In Federal Statistics and National Needs, Prepared by the Congressional Research Service for the U.S. Senate Subcommittee on Energy, Nuclear Proliferation, and Governmental Affairs. U.S. Government Printing Office, Washington, D.C.

Constance F. Citro
National Research Council
Washington, DC
U.S.A.

Rao, C.R. and Kleffe, J., Estimation of Variance Components and Applications. North-Holland, Amsterdam, 1988, ISBN 0-444-70023-4, xiii + 370 pp., Dfl.160.00.

This book is the third in the North-Holland series in Statistics and Probability edited by P.R. Krishnaiah, C.R. Rao, M. Rosenblatt, and Y.A. Rozanov. Although not belonging to a series of monographs, this book does bill itself (page vii) as a monograph, and it certainly is. Despite its broad-sounding title, its viewpoint is very limited, namely just that of estimating variance components by the methodology known as minimum norm estimation. This is a method that for many sets of real-life data yields estimates that depend on "prior values α_i of the parameter θ_i" (page 90), θ_i being one of the variance components to be estimated. Thus different users of the same minimum norm quadratic estimation (MINQE) method will get different estimates from the same data set, depending on their particular prior values α_i. Not only is the book limited to this methodology, but the ". . . and Applications" of its title is a total misnomer. Real data are referred to only twice (pages 226–7 and 316), on each occasion of very limited quantity, with no actual data shown, and with no more than a line or two of cursory description.

The MINQE methodology has (page 93) many variants MINQE(·), where the parentheses are variously (U), (I), (U,I), (U,NND) and (I,NND) for unbiased, invariant and non-negative. There is also (PU) for partially unbiased (page 106) and (AU) for almost unbiased (page 125). Furthermore, several of these have extensions to O-MINQE (page 98), C-MINQE (page 149) and I-MINQE (Chapter 9), the latter being iterated MINQE, which is related to maximum likelihood estimation (MLE) and to marginal (or restricted) maximum likelihood (MLE or REML), although the authors never use the latter acronym, common though it is in the U.S.A. But this profusion of variations in the MINQE theme is presented with almost no motivation for the practitioner and with little or no discussion of when one variation is to be preferred over others. The book is therefore of little practical use or interest to the applied statistician who is concerned primarily with analyzing and interpreting data.

On the other hand, both authors are well known for their work in mathematical statistics including, particularly, their developments of the minimum norm method of estimating variance components. It is in this context that the book shines—as a reference. It begins with a chapter on matrix algebra, followed by chapters on quadratic forms, models, identifiability and estimability, estimation, pooling data, optimality, computation, iteration, asymptotics, minimum variance estimation, and selection. All of these topics are dealt with with great mathematical thoroughness, couched in full mathematical generality of projection theory, conic spans, semi-inner products and so on. Such thoroughness is blinding to the applied statistician, but is fine for the reader who is strongly oriented towards the mathematics of statistics. That reader will find this book to be a gold mine of research ideas and results.

The mathematical perspective of this book is demonstrated by the example of having just two observations, modeled as $Y_1 = \beta + \varepsilon_1$ and $Y_2 = \beta + \varepsilon_2$, and wanting to estimate β, σ_1^2 and σ_2^2 from them. This is used on pages 230–1 to illustrate the possibility of being able to use MINQE but not MLE. This may be so, but the practical naivete of the example detracts from its instructiveness. No one would ever contem-

plate using just two observations to estimate two variances and a mean! Other minor irritants of the essentially mathematical style of writing include several instances of forward referencing (e.g., on page 92 requiring the proof on page 162), and (despite the first chapter on matrix algebra) several occasions of quoting matrix results of some complexity without indication of their derivation or reference thereto; e.g., for $M = I - XX^+$, below (9.2.4), we find $R = (MVM)^+$
$= V^{-1} - V^{-1}X(X'V^{-1}X)^-X'V^{-1}$
and for $BX = 0$, below (9.3.4), there is $R = B(B'VB)^{-1}B'$. This naked display of such results does not make for easy reading. Finally, although the book is indeed well written and seems delightfully free of misprints in its technical content, errors abound in the reference list: Harville (1976) of page 6 appears as Harville (1976b), Patterson and Thompson (1971) has the wrong title, and numerous authors have wrongly-spelled names: Lamotte, Russel, Brandly, Mallela, Searl, and Hendrson are some of them.

To summarize: the shortcomings of this book are such as to detract from its appealing to applied statisticians. Its attractions are its thorough mathematical treatment of, primarily, one method of estimating variance components (that depends on "using prior values" for those variance components) – and its high mathematical standard and generality in developing this treatment. Although the preface gives no specific indication of the intended readership, it is clearly expected to be those of strong mathematical persuasion; for them it will be a valued monograph – but not a text for students.

Shayle R. Searle
Cornell University
Ithaca, NY
U.S.A.

Wolter, K.M., Introduction to Variance Estimation. Springer Series in Statistics, Springer-Verlag, New York, 1985, ISBN 0-387-96119-4, xi + 427 pp., DM160.00.

Wolter's book centers around six main subjects with a separate chapter devoted to each, namely

The Method of Random Groups (6 pp.),
The Jackknife Method (43 pp.),
Generalized Variance Functions (17 pp.),
Taylor Series Methods (22 pp.),
Variance Estimation for Systematic Sampling (56 pp.).

Variance is here understood as sampling variance (design–variance). Model-based inference is confined to a special case: systematic sampling.

Characteristic of modern large-scale, multiple-purpose surveys, what Wolter refers to as "modern complex surveys," is a combination of several selection and estimation techniques and large masses of data. Typically, in a survey of this kind, we would have multi-stage sampling with stratification, and systematic selection with varying probabilities within strata. And difficulties with variance estimation compounded by the ubiquitous problem of nonresponse and measurement errors! In some instances, it is conceivable to construct an unbiased estimator of the design variance, whereas in other cases there does not even exist an unbiased estimate of the variance, e.g., for systematic sampling schemes. A pedestrian way out of this predicament is to let formulas derived for simple random sampling approximate the variance in other more complex situations, hopefully with conservative estimates of variance as a result. This is, of course, not at all satisfactory. There is obviously a need for more rigorous methods, based on sound theory and empirical experience, as argued by Wolter in the introduction.

Criteria for choosing good variance estimators are:

small mean square error, short confidence intervals, or high interval coverage probability;
cost effectiveness, simplicity and flexibility;
administrative convenience and availability of software.

These are important and, unfortunately, often conflicting considerations.

The first three methodologies treated are based on the concept of repeated sampling. Another resampling technique – the bootstrap – although of potential use, is not yet fully developed for application in survey research. Wolter leaves it for future consideration. As examples of applications, Wolter uses well-known surveys such as the Health Examination Survey conducted by the U.S. National Center for Health Statistics, and the Census Bureau's Retail Trade Survey. Wolter places the reader right in the middle of survey sampling practice! Incidentally, statisticians employed by these two U.S. agencies have contributed much to the development of variance estimation over the last 30 years.

Chapter 4 shifts attention to the jackknife method, introduced by Quenouille in the 1950's. Interestingly, some techniques probably not very familiar to survey statisticians are taken up here: e.g., the jackknife method for unequal probability sampling without replacement. Chapter 5, the shortest in the book, introduces the idea of modeling the variance as a simple function of the expectation. In large-scale surveys with many items and subgroups of interest – perhaps thousands of basic estimates – the computation and publication of separate standard deviations for each estimate tend to become unmanageable. *Generalized Variance Functions* (GVF's), as Wolter calls them, are of great help here. To obtain estimates of variability, we simply evaluate the model at the survey estimates with the aid of charts or tables. For concreteness, consider the estimate of the total number of people in a population having a certain attribute: $\hat{X} = N\hat{P}$, say, with relative variance (random sampling)

$$\text{Relvar}(\hat{X}) = N/nX - 1/n = \alpha + \beta/X,$$

with $\beta > 0$. This simple GVF has been used by the Census Bureau since 1947 for its Current Population Survey. Other similar models, with possibly less theoretical justification than the one above, are discussed by Wolter.

Although estimates in survey sampling often have a very simple form, they are sometimes by necessity non-linear in the observations. The simplest example here is the popular ratio estimate. In such cases, simple unbiased estimators of the design-variance are not usually available. We then have to resort to Taylor series expansions to obtain a linearization. Chapter 6 draws on work by Fuller, and contains more theory than the previous chapters. An important application discussed in the chapter is composite estimation, where simple estimators are combined in order to utilize a correlation structure to obtain increased precision. The Census Bureau's Retail Trade Survey, where total monthly sales are estimated for different kinds of business, illustrates the technique. For Statistics Sweden, the labor force surveys – AKU – present similar estimation problems.

Chapter 7 reviews systematic sampling with equal and varying probabilities, an area in which Wolter himself has contributed to the theory of variance estimation. In an interesting section, eight estimators are presented and discussed in the light of a superpopulation set-up. As is done throughout the book, the theoretical discussion is supplemented with an empirical investigation using real data. Chapter 8, finally, contains a summary where the author attempts to answer the "ultimate question": Which of the various methods can be recommended and under what circumstances? Interval coverage probability is given primacy here and results from the literature are reported together with some Monte Carlo studies. The choice of a variance estimation method involves a complex trade-off between factors such as accuracy, cost, and flexibility. Clearly, no one method can be expected to be best over all. The book closes with six brief appendices, a useful one presenting software for variance estimation.

Perhaps I am allowed a few critical remarks to conclude this review. The purpose of the book, as stated by Wolter, is to describe a number of techniques for variance estimation suggested in recent years and to illustrate their use in the context of modern complex surveys. Wolter says (p. 5) that most textbook discussions stop short of dealing with important features of such surveys. I am not sure I agree here! Cochran's

classic text (1977) discusses, e.g., shortcut computation of variance in stratified sampling (6.13), takes up collapsed strata (5A.12), and introduces the jackknife (6.17). Cochran mentions complex surveys (11.17–18), and, finally, devotes the entire chapter 13 to problems of non-response and measurement errors. Obviously, in an introductory text emphasis must be on standard procedures and simple methods, while Wolter can concentrate on non-standard methodologies. In all justice to Cochran, however, it must be recognized that he does touch upon several of the problems Wolter finds characteristic of modern complex surveys.

Some further details: When discussing the random group method (independent groups), Wolter does not mention the possibility of forming an estimate based solely on the distinct units of the total sample. Estimation based on the distinct units of the sample is deferred to a subsequent section. Similarly, it is tempting to consider other estimators of the population mean than the simple weighted average (p. 255) when a systematic sample is supplemented by a simple random sample. In the chapter on variance functions, finally, I expected a treatment of the kind of variance functions discussed by Cochran in chapter 9.5, i.e., models relating the within-cluster variance to the size of the cluster. But I suppose there is no established terminology here.

What I have said here does not detract from the value of Wolter's book. Prior to his text, we had no systematic treatment of this important methodology, only descriptions widely scattered through the literature. It is a pleasure to recommend this book to colleagues engaged in teaching, research, and survey practice.

Reference

Cochran, W.G. (1977): Sampling Techniques (Third edition). John Wiley & Sons, New York.

Sven Berg
University of Lund
Lund
Sweden

Oakes, M., Statistical Inference: A Commentary for the Social and Behavioural Sciences. Wiley, Chichester, 1986, ISBN 0-471-10443-4, ix + 185 pp., £20.50.

Michael Oakes is a psychologist who teaches statistics to psychologists at the University of Sussex, U.K. He intends his book for teachers and students of the social and behavioural sciences, and more generally, for research workers in fields where statistical methods form an integral part of the methodology. To me his book is both interesting and irritating. I am a statistician who has taught statistics to social science students.

The first two chapters of the book are on significance tests of a simple null hypothesis concerning the parameter of interest. The standard example is a two-sided t-test of the difference between two means. These chapters are not easy to read because Oakes contaminates his account of significance tests with an acid polemic against a number of psychologists – and one unfortunate statistician – who have written about the logic of significance tests. The propositions that Oakes quotes are, indeed, horrendous. The worst one perhaps being that you should have more confidence in a small than in a large sample. I do not believe that such "wisdoms" deserve repetition or refutation.

In the constructive part of his treatment Oakes defines the significance level of a test as the preset p-value for which you look up the critical t-value, and the associated probability as the p-value you look up for an observed t-value. My only addendum is that the latter concept is often referred to as the observed significance level.

Oakes emphasizes the importance of knowing the power of one's test. He points out that ". . . in current psychological practice the sense of balance between the two types of error is lost." He has the nerve to put part of the blame for this on Fisher. Anyway, he cleverly lists some negative consequences of the neglect of power considerations. The most convincingly argued of these is that the use of low power tests leads to poor replicability of research find-

ings. Another, more controversial one, is based on the "publication bias." If only significant results are published, and if they are all based on a set significance level, say 0.05, and if the power of the tests is low, then the percentage of false positives among the published results is much higher than 5%.

Oakes's critique of significance tests has several aspects. The first being that the Neyman–Pearson tradition of rejecting or accepting the null hypothesis when the test statistic is larger or smaller than a preset critical value is a decision rule, and not an inferential procedure. The second is that tests whether with fixed or observed significance levels treat the null and alternative hypothesis in an asymmetric way which not always corresponds to the scientifically relevant question. On this score I wish Oakes had contemplated Cox's (1977) careful analysis of the many different purposes the null hypothesis can serve.

A more original critique of Oakes's is that the asymmetric, but easily used machinery of significance tests has kept much theorizing in the social and behavioural sciences on an unnecessary primitive level. Instead of imaginative modelling of data, research workers content themselves with simplistic 'is there such an effect' questions. Oakes provides three explanations for the retention of significance tests in the face of much argument against them. The reasons are: (i) bad habits engraved at an early statistics course, (ii) the scaring subjectivism and even some fanatism of the Bayesian alternative, (iii) the superficial philosophical respectability of significance tests bestowed on them by logical positivism and Popper's fallibilism.

In his third chapter Oakes attacks significance tests in a shrewd empirical way. He has collected data from experienced academic psychologists about their conceptions of significance tests. On the whole this is a well designed and nicely reported study, and Oakes succeeds in getting across the message that the confusion about significance tests is alarmingly widespread. However, I want to defend the psychologists on one point of some general bearing. He made his subjects contemplate a situation where the difference in means between the control and experimental groups gives rise to a t-value of 2.7, on 18 degrees of freedom, with a p-value of 0.01. The subjects were then asked to mark six different statements as either true or false. I shall reproduce only two of these with the percentage of "true" answers attached:

1. You have found the probability of the null hypothesis being true; 36%.
2. You know, if you decided to reject the null hypothesis, the probability that you are making the wrong decision; 86%.

The first statement is clearly wrong. Oakes claims that the second statement is, on close analysis, identical to the first, and therefore concludes that the overwhelming majority of psychologists interpret p-values in a Bayesian way. I believe that the second statement is all right, and would like to refer to Cox's comment on a recent paper by Berger and Delampady (1987) for support. Berger and Delampady, on the other hand, do not think that a valid frequentist interpretation of p-values can be given. At least, we can conclude that the questions Oakes asked his colleagues are not simple, and the logic of statistical inference far from agreed upon.

The second part of the book is called "Schools of Statistical Inference", Oakes starts by introducing the different interpretations of the probability concept. He treats the classical, the frequency, the fiducial, the logical, and the personal probabilities in a concise and informative way. I have only one minor objection: von Mises's idea of probability as a limiting value is hardly the nucleus of the present day frequency probability. The nucleus is that a fixed but unknown probability is axiomatized to exist for phenomena which have statistical stability. The difference is like defining a point in geometry, either as the limiting value of a sequence of diminishing chalk points, or by a set of axioms.

The sixth chapter is the shortest in the book and, by far, the most informative. Oakes contrasts the Neyman-Pearson, the Fisherian, the Bayesian, and the likelihood

schools of statistical inference. He summarizes his comparison in a table which assesses the different schools with regard to the following characteristics: their probability concept, their emphasis on decision versus inference, their emphasis on testing versus estimation, their use of prior information, their use of initial versus final precision, the location of their inference on the sample domain versus on the parameter domain, and finally, their sensitivity to the sampling procedure. For me the contrast between the Neyman-Pearson school as locating its inference on the sample domain, and the other schools locating the inference on the parameter domain was especially enlightening.

Then follows a chapter which is an elaboration of the preceding one. A simple "difference between two means" example is analyzed by the approach of each of the four schools. Oakes declares his "own mild preference" for likelihood inference. He also concludes that none of the approaches is entirely satisfactory. This is all fair and well, but I cannot find that Oakes has cause enough for the following formidable claim: "It is not unreasonable to conclude, as Popper did with induction, that statistical *inference* is a myth."

The last chapter covers a number of problems seldom treated in statistics courses but of importance in social science. Is there philosophical justification for statistical tests in observational as contrasted to experimental research? Is meta-analysis a new *bête noire*? Suppose that in a given experimental treatment 60% of subjects exhibit behaviour X. Does this mean that each subject has a 0.6 probability of producing X? Or does it mean that 60% of subjects deterministically exhibit X and 40% deterministically fail to do so? Can we have explanation in science without having prediction, and can we have prediction without explanation? Is it a scientifically respectable task to find and study children who can recite nursery rhyme backwards, or only to estimate the proportion of such an ability?

I do not agree with everything Oakes states about these questions, but they are interesting questions. I do not share all Oakes's grievances about significance tests, but he has got me to see that significance tests should in the future be taught to (social) scientists as a part of data analysis, not as the only vehicle of statistical inference. I do recommend Oakes's book for incitement, but I wish social scientists and psychologists would complement it with Cox and Snell's (1981) more constructive approach to statistical methods. Their spirit is (p. 39): "To summarize, significance tests have a valuable but limited role to play in the analysis of data."

References

Berger, J.O. and Delampady, M. (1987): Testing Precise Hypotheses (with discussion). Statistical Science, 2, pp. 317–352.

Cox, D.R. (1977): The Role of Significance Tests. Scandinavian Journal of Statistics, 4, pp. 49–70.

Cox, D.R. and Snell, E.J. (1981): Applied Statistics: Principles and Examples. Chapman and Hall, London.

Anders Ekholm
University of Helsinki
Helsinki
Finland

Farebrother, R. W., Linear Least Squares Computations. Marcel Dekker, Inc., New York, 1988, ISBN 0-8247-7661-5, xii + 293 pp.

Chapters 1–7 of this book are based on a ten-week course of lectures given to senior undergraduate students in the Faculty of Economic and Social Studies at the University of Manchester. Later chapters contain interpretations and extensions of this material. The main emphasis of the book is computational aspects of the linear least squares problem, which is covered in great detail. A unique feature is the careful attention given the history of least squares computations. The author has taken great pains to locate original references for

the algorithms presented dating back to Legendre, 1805, Gauss, 1809, Cauchy, 1835, and others. I enjoyed this feature of the book very much, and it is worth having only for this!

The book makes no assumption concerning the reader's knowledge of matrix algebra, and starts with an elementary introduction to Gaussian elimination, its matrix representation, and the Cholesky and Doolittle decompositions. Chapter 4 gives a very interesting treatment of methods usually named after Gram-Schmidt. The author points out that the method known as the "modified" Gram-Schmidt orthogonalization procedure is in fact due to Laplace, 1816, and can be viewed as a special case of a method due to Cauchy, 1835, for the system $Z'Xb = Z'y$. Also very useful and not easily found elsewhere is the derivation of hybrid methods based on partial orthogonalization and normal equations. To me this chapter is one of the highlights of the book. Chapter 5 presents several variants of Householder's procedure. Chapter 6 gives a complete treatment of variants of Givens's procedures. Square root free algorithms are introduced and their application to weighted least squares, and the treatment of linear constraints are discussed. A method for imposing constraints by weighting is also discussed here. Stable techniques for adding and deleting regressors and observations by updating the QU decomposition are given in Chapter 7.

A discussion of pseudo-random numbers in Chapter 8 is offered for readers not familiar with statistics, and in Chapter 9 the standard linear statistical model is introduced. At the end of this chapter a project of developing a least squares computer program is offered. A comparison to a list of published programs is suggested. A program I miss in this list is the excellent FORTRAN program implementing modified Gram-Schmidt with iterative refinement by Wampler (1979).

In Chapter 10 a rudimentary treatment of condition numbers is given. An empirical condition number is introduced for changes in the observation set y. Too much emphasis is given to empirical condition numbers, since reliable and cheap methods of estimating condition numbers are now part of many subroutines. It should also have been pointed out that the condition number $(\lambda_1/\lambda_n)^{1/2}$ is the condition number for perturbations to the full data set. Instrumental variable estimates are covered in Chapter 11. Chapter 12 gives a clear and complete presentation of methods for generalized least squares estimation, based on recent work by C.C. Paige. This is an excellent chapter containing material not easily found elsewhere. Iterative refinement of solutions is discussed in Chapter 13. The author here cites some conclusions by R.H. Fletcher, which this reviewer criticized in Björck (1978). Chapter 14 covers canonical expressions for least squares estimators and test statistics. Chapter 15 gives traditional expressions for least squares updating when regressors and observations are added or deleted. It could have been mentioned that these are less stable than those given in Chapter 7. Finally, Chapter 16 discusses least squares estimation subject to linear equality constraints.

This book contains an excellent selection of methods covered in great detail. For some methods several computational variants are presented. I miss, however, the singular value decomposition, first stated in its generality by the statisticians Eckart and Young, 1939.

The aim of the author is that the readers should get a thorough computational grounding by implementing most, if not all, of these algorithms on a computer. The end of each chapter contains implementation exercises and suggests comparisons of timings and accuracy. I suspect that most readers will not have the interest or perseverance to follow these intensions in full. But even doing a small fraction of the exercises should certainly be valuable and illuminating.

In my opinion, perhaps the greatest defect of the book is that readers are supposed to find out for themselves the merits in accuracy and efficiency of the large selection of methods. The book does not guide the reader enough in the choice of algorithm. I

think the author asks a lot from the readers; even many lecturers would be hard put to guide his or her students in a successful way! A lot is known about the numerical properties of the presented algorithms. Why not tell the reader more? I should mention that, as a numerical analyst, I seem to worry more about numerical stability and accuracy than most statisticians! I also think one should not skip so lightly around questions like, for example, when is a small number to be judged equal to zero (page 32), and what is a large number (page 177).

In all the author has done a good job in sharing his vast experience in the computational solution of linear least squares problems. The choice of algorithms is up to date, and the historical notes a delight. This is a very useful book.

References

Björck, Å. (1978): Comment on the Iterative Refinement of Least-Squares Solutions. Journal of the American Statistical Association, 73, pp. 161–166.

Wampler, R.H. (1979): Solutions to Weighted Least Squares Problems by Modified Gram–Schmidt with Iterative Refinement. ACM Transactions of Mathematical Software, 5, pp. 457–465.

Åke Björck
Linköping University
Linköping
Sweden

David, H.A., The Method of Paired Comparisons, Second Edition, Revised. Charles Griffin & Company, London, 1988, ISBN 0-85264-290-3, viii + 188 pp., £21.50.

In the method of paired comparisons several objects are compared, two at a time, by one or more judges. Each comparison results in some relative statement (preference) about the two objects with regard to the variable under consideration. The method is particularly suitable with variables (characteristics) for which no exact method of measurement exists and has a wide range of applications, most notably in psychology, industry, and sports. This book, of which the first edition appeared in 1963, successfully aims to give "a unified picture of the state of the subject" and has been written with both the statistical and non-statistical reader in mind. As such the book contains mostly results and only a few proofs. Familiarity with some of the basic statistical concepts such as maximum likelihood and incomplete block designs is not strictly necessary but it is helpful. For the statistical readers, and especially students, each chapter contains a number of exercises, most of which are of a theoretical nature. This second edition is completely updated with developments over the last 25 years and contains an extensive bibliography, which increased from 141 to 287 items. At the end of the book one may also find five tables to be used for some of the statistical tests of significance, described in the book.

Chapter 1 describes briefly some basic probability models for paired comparisons, including a paragraph on triple comparisons and a historical note, while in Chapter 2 one may find the basic distribution theory of scores, which are the numbers of times certain objects are preferred over other objects. In the second edition some measures of inconsistency are discussed, other than and in addition to the well-known circular triads. Chapter 3 deals with various non-parametric tests for scores in balanced paired comparison experiments. Some of the tests described are: test of one particular object, test of the top score, overall test of equality of the objects, etc. For each test the procedure is given in the form of a recipe and illustrated with a numerical example. The linear model, subject of Chapter 4, is if it holds, particularly appealing for the practical researcher, since it provides numerical estimates of the values of the worths (or merits) of the individual objects. These estimates may be used for further analysis. The best known linear models are discussed, with estimation

methods and tests of significance, including one for the validity of the model.

One of the possible drawbacks of a paired comparison experiment is that, with increasing number of objects, the total number of pairs to be compared becomes too large for a judge to do within a limited time. In such a case one may want to let judges do only a part of the comparisons, leading to issues of experimental design, discussed in Chapter 5. The similarity between paired comparison designs and tournaments is emphasized in Chapter 6. Several types of tournaments are discussed with regard to their suitability for different objectives like selecting the best object or determining a complete ranking. The last chapter is dedicated to miscellaneous subjects. Some of these are selection of judges, treatment of ties and within-pair order effects, Scheffé's method and optimal scaling (Guttman's procedure).

To summarize, this is a very informative book, which, despite its specialized content, can be recommended for any statistical library.

A. van Baaren
Statistics Canada
Ottawa, ON
Canada

Printed in October 1989.