Journal of Official Statistics
Vol. 5. No. 1, 1989, pp. 93–104

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.

Wainer, H. (Ed.), Drawing Inferences from Self-Selected Samples. Springer-Verlag, New York, 1986. ISBN 0-387-96379-0 (Springer-Verlag New York), ISBN 3-540-96379-0 (Springer-Verlag Berlin). xii + 163 pp., DM 45.00.

Most statisticians analyze data through models that describe an underlying population of interest, for example the iid normal model:

$$y_i \sim_{\text{iid}} \mathrm{N}(\mu, \sigma^2). \tag{1}$$

In practice data come in the form of samples. Let $s_i = 1$ if unit i is selected and 0 otherwise. Then we actually see the distribution of Y given that $S = 1$:

$$y_i | s_i = 1 \sim ? \tag{2}$$

Statistical texts usually ignore the implicit conditioning on S and replace ? in (2) by the population model, such as (1). This is fine if Y and S are independent, a reasonable assumption if we have (or plausibly can pretend we have) a simple random sample. However often this is not reasonable. An important case is *self-selection*, when the subject's choice enters into the inclusion process. *Drawing Inferences from Self-Selected Samples* (DISS) presents analytical approaches to self-selection in four application areas: (i) comparisons of SAT scores when scores are available only to those who choose to take the test (by Howard Wainer); (ii) evaluation of methadone clinics in the treatment of heroin, where clinic patients are volunteers (by Burton Singer); (iii) assessing the effects of job training, where data are available on those who choose to train (by James Heckman and Richard Robb); and (iv) survey nonresponse, where the sample is restricted to individuals who choose to respond (by Robert Glynn, Nan Laird, and Donald Rubin). The book also includes lively contributions from two distinguished discussants, John Hartigan and John Tukey.

I would like to have seen the views of a survey sampling specialist (one who, in my trivial example, treats Y as fixed and bases inference on the distribution of S given Y). The sampler is trained to handle problems of probability sampling where the distribution of S is under control of the sampler but may not correspond to simple random sampling. Sampling theory seems to me of limited help in self-selection, where the distribution of S is not under our control. Nevertheless I suspect a sampler would provide an interesting counterpoint to some of the viewpoints in the book, for example, the remarks on panel vs cross-sectional survey designs in Heckman and Robb's paper. On a point of detail, samplers might also take issue with Heckman and Robb's statement that many social science data sets contain hundreds and thousands of *independent* observations (page 67, italics mine), since large samples usually involve clustering that leads to correlated observations.

The chapters by Wainer on SAT scores and Singer on methadone treatment assessment are well written and relatively nontechnical, although some of the concepts discussed are subtle. They provide a valuable introduction to the more technical material in the papers that follow. Wainer compares mean SAT scores for 21 U.S. states that give primarily SAT tests ("SAT states") with mean SAT scores for 29 states that administer primarily the American College Testing (ACT) Program ("ACT" states). Scores for the latter group are much higher, the distributions barely overlapping. Does the difference (a) reflect superior SAT-taking ability in the ACT states, or (b) are SAT-takers in the ACT states a more select group? Wainer shows that SAT-takers in ACT states rank higher in their college class than SAT-takers in SAT states, thus providing strong evidence in favor of (b).

A common statistical approach to lack of comparability between treatment groups is covariate adjustment on variables thought to capture (or at least reduce) the lack of comparability. Wainer discusses this approach to his problem, using the covariate "percent of high school seniors taking the SAT," or participation rate for short. He points out this approach may yield bias, since "higher-quality schools will yield a greater proportion of their SAT-taking pool," that is, participation rate is a measure of SAT-taking ability as well as of the selection effect. In path analysis terminology, participation rate is not causally prior to SAT-taking ability. Wainer applies an alternative strategy based on converting ACT scores to SAT scores using equating information. This approach almost by magic removes the differences in scores in the SAT and ACT state groups: the adjusted medians are identical! However equating is a not a viable option in other settings. Econometricians would perhaps advocate attempting to fix covariate adjustment via structural equation modeling. It would be interesting to see how results from such an approach compare with the answers from equating, which might plausibly be treated as the gold standard here.

Singer's chapter on Methadone Maintenance Treatment (MMT) programs first reviews heroin abuse and treatment in Hong Kong, Sweden, and New York and presents a pattern of heroin addiction which serves as a baseline for comparing treatments. Singer then discusses some intervention strategies, and the evaluation of MMT programs via performance-based ratings indices. With regard to self-selection, Singer raises the important question of whether the aim is to evaluate and compare programs on the population of volunteers who enter them, or on the target population of all addicts, an issue which is also stressed in the Heckman and Robb paper. Although he views inference to all addicts as desirable, he judges it impractical given the lack of quantitative knowledge to distinguish the two populations. Even if such knowledge were available, it seems to me that considerable extrapolation would be involved.

Even the problem of comparing clinics on the volunteer population is far from easy, given that different clinics may attract different clients. Singer's approach is to develop relatively crude indices based on changes in a patient's activities and behavior before and after treatment, and in essence to compare these indices with historical control treatments deemed to have been successful. Singer discusses the limitations of this approach. The use of change measures from longitudinal data, allowing each subject to act as his or her own control, seems to me a time-honored and powerful way of alleviating the effects of self-selection. In addition, some form of simple covariate adjustment on the baseline variables would seem to be feasible here if they are consistently measured across studies, but maybe that is a big if.

The paper by Heckman and Robb provides an extensive model-based analysis of the selection bias problem in the context of the impact of job training programs. It is not in fact the one presented at the conference, which was published elsewhere, but a revision that takes into account the skeptical reactions of Tukey and Hartigan to the conference paper. (Readers might be able to estimate a Hartigan/Tukey effect by constructing a measure of change from the two versions!) The Hartigan and Tukey discussions are published before the revised Heckman-Robb paper, preserving the chronological order. (Since we all read discussions of papers before papers, maybe this is the right order anyway!) The discussions focus on the sensitivity of results to untestable assumptions in the Heckman and Robb analysis, and illustrate the differing attitudes of statisticians and econometricians towards statistical analysis. Econometricians start with a theoretical model and work towards the data, pruning parameters until the model can be estimated. Statisticians start with the data and work towards a theoretical model, estimating parameters that shed light on, but may only approximate, idealized quantities of econometric theory. Econometricians complain that the statisticians are too atheoretical, or "context-free." Statisticians complain that econometricians are too insensitive to the limitations of their models and the data. If the bridge between data and theory is shaky, as in the self-selection problem, then these two approaches can end up in different places.

Heckman and Robb's paper is long and technical, but contains a helpful introductory section that clarifies its objectives. These include (i) a careful definition of the parameter of interest, the effect of job training; (ii) specification of minimal assumptions needed to identify the parameter, for single cross-section, repeated cross-section and longitudinal designs, under conventional and enriched behavioral models of earnings. They conclude that "although longitudinal data are widely regarded in the social science and statistical communities as a panacea for selection and simultaneity problems, there is no need to use longitudinal data to identify the impact of training on earnings if conventional specifications of earnings functions are adopted. Estimators based on repeated cross-section data for unrelated persons identify the same parameter. This is true for virtually all longitudinal estimators." (Page 65).

This conclusion has created some controversy among advocates of panel surveys, since the implication is that panel surveys are overused in practice. However, the practical ramifications are not clear, since the conclusion results from a mathematical analysis that focusses on the identifiability issue in the context of specific selection models. Survey sampling arguments for panel surveys consider a separate issue, the sampling error of simple estimates of change (such as the difference in means) that ignore selection effects entirely (Cochran (1977, Section 12.10)). The Heckman and Robb analysis ignores sampling error entirely. Also a major reason for social science panel surveys (such as the U.S. Survey of Income and Program Participation) is their ability to measure micro-level transitions that are inestimable from repeated cross-sections.

Heckman and Robb's approach to selection bias is to model the data and the selection mechanism. In the context of job training, a parameter α is introduced to represent the additional earnings from training. Training occurs if an observed variable called index of net benefits (IN) crosses a threshold, say zero. In the behavioral model, IN is viewed as the difference between the expected benefits of training (the gain in future earnings, discounted to some degree) and the expected costs (expenses and loss of earnings during training). Selection bias arises under this model when the propensity to train is related to future earnings in the absence of training, after adjusting for the effects of observed covariates. For example, if (given covariates), those predisposed to be successful are more likely to train, the positive effects of training will be exaggerated by comparing the (adjusted) mean incomes of trainees and non-trainees.

Writing down formal models such as those considered by Heckman and Robb can be a useful way of clarifying thinking. However the purely economic model of training choice seems hard to swallow, as does the assumption of a constant training effect for all individuals: Heckman and Robb do provide a limited discussion of random training effects, but it is mainly directed at defining the parameter of interest. If a distribution of training effects exists, it seems to me that longitudinal data would be needed to estimate it, so the assumption of constant training effect favors the cross-sectional design.

Heckman and Robb's cross-sectional methods for adjusting for selection bias appear to depend crucially on finding instrumental variables that are predictive of the decision to train, but not predictive of earnings. No specific suggestions for variables are offered, and (coming from the statisticians' camp) I have less confidence than Heckman and Robb in the ability of econometric theory to supply them. Purely for illustration, let me propose the variable "distance to training site." This variable might be strongly related to the decision to train, and a plausible econometric story might justify the assumption that "distance to training site" is not related to earnings, after adjustment for other exogenous variables in the model. Human populations are heterogeneous, however, and social science theory does not lead to all-encompassing physical laws. Thus it also seems plausible that the variable *is* related to earnings, particularly if it is acting as a proxy for some unmeasured geographical covariate. This difference of opinion does not matter much for some types of analyses, but it does if a large selection bias adjustment rests on it. For me, these instrumental variables (IV's) often supply blood to a body that is already dead; "minimal identifying assumptions" (MIA's) are too often "missing in action"!

What are the alternatives to selection modeling? One approach is to try to collect as many variables as possible related to the selection process, and then use these variables in a standard covariate adjustment. Here longitudinal surveys may have a distinct advantage over cross-sections, because of superior ability to collect time series information.

Selection models of the type considered by Heckman and Robb have also been applied to survey nonresponse, and it is this application that is the subject of the Glynn, Laird, and Rubin chapter. These authors compare two modeling strategies; let Y denote the outcome variable of interest, X fully-observed covariates and R an indicator for response ($R = 1$) or nonresponse ($R = 0$). *Selection modeling* writes the joint distribution of Y and R in the form

$$f(R, Y|X, \theta, \psi) = f(Y|X, \theta) f(R|Y, X, \psi), \quad (3)$$

where the first component characterizes the distribution of Y given X in the population, and the second component models the incidence of nonresponse as a function of X *and* Y. *Mixture modeling* writes the joint distribution in the alternative form

$$f(R, Y|X, \xi, \omega) = f(Y|X, R, \xi) f(R|X, \omega), \quad (4)$$

where the first distribution characterizes the distribution of Y given X in respondent and nonrespondent strata, and the second component models the incidence of nonresponse as a function of X only. The distribution of Y given X is then a mixture of the distribution of Y given X in the response and nonresponse strata, which explains the name. Selection modeling is natural to econometricians since their models relating Y and X are formulated in the unrestricted population. Mixture modeling is perhaps more natural for statisticians since it is closer to the structure of the observed data. In particular the mixture modeling form (4) emphasizes a basic difficulty inherent with the data; since there are usually no data on Y for nonrespondents, there is no information for estimating the distribution of Y given X, $R = 0$. Rubin (1977) relates the distribution for nonrespondents to that for respondents using a Bayesian prior distribution. The selection modeling form (3) can be estimated without explicit inclusion of prior information relating respondents and nonrespondents. However such a prior specification is implicit, and sensitivity to model specification is an equally serious problem for either version of the model.

Glynn, Laird, and Rubin display sensitivity of the selection approach to model misspecification by simulating results under correctly-specified and misspecified models. They conclude that the method is very unstable unless a covariate is available that is related to *only* one of response or outcome; the variable plays the analogous role to the IV variables in the Heckman and Robb paper. Here as in the job training context, the key question is whether such variables can be found in practice: Glynn, Laird, and Rubin are pessimistic.

The chapter also compares the selection modeling and mixture modeling approaches when a subsample of nonrespondents are available via follow-ups. The assumption is made that the subsample is random. Comparisons are made using simulated data, and real data from a survey on drinking behaviors. They conclude that mixture modeling is more robust than selection modeling to departures from distributional assumptions.

Like any collection of papers, DISS lacks some degree of cohesiveness. The book presents the views of distinguished applied statisticians on a problem that arises in the real world, rather than the artificially constructed world of many mathematical statistics texts. I found the book stimulating and recommend it.

References

Cochran, W. G. (1977): Sampling Techniques. 3rd Edition. John Wiley, New York.

Rubin, D. B. (1977): Formalizing Subjective Notions About the Effects of Nonrespondents in Sample Surveys. Journal of the American Statistical Association, 72, pp. 538–543.

Roderick J. A. Little
U. C. L. A.
Los Angeles, CA
U.S.A.

Turner, C. F. and Martin, E. (Eds.), Surveying Subjective Phenomena, Volume I and Volume II. Russell Sage Foundation, New York, 1984. ISBN 0-87154-882-8 (Vol. I), 0-87154-883-6 (Vol. II), 0-87154-881-X (set). xvi + 495 pp. (Vol. I), xvi + 618 pp. (Vol. II).

It is astounding how much attitude survey research has become a part of academic and political life. It is estimated that a minimum of 100 million survey interviews were conducted between 1971 and 1976 in the United States. More than 28 million survey interviews were conducted by telephone during 1980. It is estimated that 39 % of the British public have been surveyed. In a single one-month period there is documented evidence of the distribution of more than 200 million copies of poll stories in American news media and more than 50 million copies in Britain. More than half of the published articles in the field of sociology report survey data, as do about 30 % of the articles published in political science and economics.

The two volumes reviewed here represent the proceedings of a multi-year panel on "Survey Measurement of Subjective Phenomena," convened in 1980 under the auspices of the United States Committee on National Statistics. The committee was convened because of the discovery of "several instances in which seemingly equivalent (public opinion) survey measurements made at approximately the same time produced surprisingly different results (I:xiii)." The problems all clearly involved nonsampling sources of error (as opposed to sampling errors which are handled by probability theory and confidence testing). The purpose of the panel was to study "the use, reliability, and meaningfulness of survey measurements of attitudes, opinions, and other subjective phenomena (I:xiv)." The work of the panel took more than two years to complete.

In the course of its work, the panel stimulated so much interesting research on the survey profession qua profession that there may be enough material for a unit on subjective measurement in a course on the history of ideas. Volume II Chapter 1 is a fascinating historical sketch of the different kinds of attitude research that developed in the early years – social distance scales, Thurstone scales, Likert's "fast" scaling technique, and so on. A careful reader also finds in this chapter the seeds for debates between academic disciplines about which one studies "real" attitudes and why studies from perspectives other than one's own are to be criticized as "conceptually inadequate." Volume II Chapter 10 is a comparison of the tendency for particular survey houses, e.g., Gallup, Harris, etc., to prefer particular approaches to questionnaire construction, e.g., open-ended, middle response categories, etc., Volume II Chapter 2 is a similarly fascinating sketch of the attempts of economists to define "utility" in a way that is not circular and therefore incapable of independent measurement. The author concludes that "economists have been more concerned with drawing out the implications of utility assumptions based on casual introspection or on an a priori conception of rationality than with attempting to measure utility in practice (Vol. II p. 42)." This level of insight and intellectual honesty in a book not specifically attempting to discredit economic analysis is refreshing.

Developments or events in four specific areas served as catalysts for *Surveying Subjective Phenomena:* (1) Surveys of public confidence in the leaders of national institutions, done at the same time and using allegedly equivalent measures, showed substantial discrepancies in both the levels of reported confidence and the trends across time. (2) Trend studies of "happiness" indicators showed divergent results depending on the survey organization conducting the field work. (3) Surveys of public attitudes toward science were being openly criticized for reifying public opinion, i.e., putting words in the mouths of respondents – on topics for which there was little public information or understanding. (4) Specific surveys were the targets of attack by non-social science university faculty as being based on a methodology that was "ambiguous," "meaningless," and "prejudicial."

Surveying Subjective Phenomena attempts to come to grips with the issues that each of these criticisms raises for the survey measurement profession. Volume I Chapter 1 points out that the problem of fallible measurement is not limited to subjective indicators, survey research, or even social science. There is a fascinating discussion of interlaboratory experiments conducted to achieve replicated measurements of natural science physical constants. The measurements *ought* to have produced the same result but they did not. Experimental studies of the variability among mesurements made by different scientists, by different laboratories, and by different analytical procedures led to "a better understanding of the error structure of such measurements (Vol. I p. 16)." These experiments in the physical sciences are the basis for the panel's recommendation for a coordinated interlaboratory program of measurements for survey research.

In a number of ways the work is a "stiff-upper-lip" exploration of the soft underside of public opinion and survey research. The work reviews past, well-publicized "failures" of survey research, e.g., the Literary Digest poll and the Dewey-Truman polls in 1948. The work reviews, in detail, the recommendations made by blue-ribbon commissions convened to study and make recommendations about those past failures (Mosteller et al. (1948)). The work reviews, chronologically and in detail, attempts made by professional polling associations, other organizations or individuals, and even the Federal government to develop, implement, and enforce standards to ensure quality and consistency in the public opinion or survey product. It was enlightening to me to see that so many of the recommendations and findings of the panel's work have been part of previous efforts as well.

The work raises many of the same questions about nonsampling errors, e.g., intensity of opinion, manufactured responses, question wording bias, questionnaire order effects, respondent understanding, selective reporting of results, and clarity of the concept being measured – as are covered in more polemical tours of the same horizon such as *Lies, Damn Lies, and Statistics* or *The Pollsters* (Wheeler (1976) and Rogers (1954)). But *Surveying Subjective Phenomena* explores the issues more fully and, in my opinion, in a more balanced fashion.

Volume I of *Surveying Subjective Phenomena* contains the panel report per se. Chapters in this volume are compilations of sections contributed by panel members and then subjected to the critical eye of editors and outside reviewers. Some of the sections and chapters in Volume I are designed to review the research and results on various topics in the literature on the reliability and validity of subjective survey measurements. Other sections and chapters in Volume I go considerably astray from this goal and are presented as new ideas or new methodologies that the panel recommends for analyzing subjective survey measures. Finally, there is a 30-page list of panel recommendations to producers and users of subjective survey data.

Volume II consists of individual contributions of authors commissioned to undertake special studies. These chapters were also subject to outside review. Some of these chapters bear directly on the point and mission of the two-volume work – some are in-depth studies, not published elsewhere, of issues in the reliability and validity of subjective survey measures. Volume II Chapter 8, for instance, is a summary of the "non-attitudes debate." Other chapters, however, go considerably far afield from this goal.

As with any published compilation of this scope (and of this panel-based methodology), the work is excessively long; overwritten at some points, underwritten at others; and, overweighted toward the interests (or abilities) of those who happen to have been panel members. Everybody who reads the work will find something that is of great interest, but they also are likely to find a great deal that is not. The two volumes are a good first draft of a book that should be about one third of its 1 145 page length.

On the other hand, I sincerely believe that there is little that has ever been said or written about the reliability and validity of subjective survey data that does not appear somewhere in this work. For this reason, *Surveying Subjective Phenomena, Vols. I & II* rivals the scope of other excellent works with a similar mission (Rossi, Wright, and Anderson (eds.) (1983)) and, therefore, merits attention by practitioners and students of social surveys.

A number of very general observations can be made about the areas of success or failure of the book. These successes and failures, I hope, will define the shape of research on survey methodology in the coming years.

The work makes an important three-way distinction between subjective phenomena, facts, and quasi-facts: (1) Subjective phenomena are those that, in principle, can be directly known, if at all, only by persons themselves – such as expectations (to vote, to have children), satisfactions (happiness, utility), subjective judgements (confidence, fairness), or opinions (for or against something). (2) Factual measurements are in principle verifiable without reference to respondents' interpretations. (3) Quasi-factual measurements allow latitude for the respondent's definition of the criterion for the (factual) behavior or event in question – such as unemployment (whether or not one is actively looking for a job), housing quality (whether a unit is deteriorating or sound), neighborhood quality (what boundaries), crime victimization (whether or not an encounter is judged to be an assault), or ethnicity (judgements based on language, father's lineage, mother's lineage, or other factors.)

The vital issues in survey measurement in this work have to do with measurement of subjective and quasi-factual phenomena. The "true score" models of physical scientists and psychologists are beside the point when one has to consider how to design experiments and calibrate the sources of measurement error for subjective and quasi-factual phenomena. The point of the definition of subjective and quasi-factual phenomena is that there is not an externally verifiable true score. Therefore it is somewhat surprising that the introduction to measurement error in Volume I Chapter 4 is a mechanistic retread of the "true score" model. The panel loses a valuable opportunity to introduce a mathematical notation and language for error models that would contribute significantly to the literature.

Beginning with Volume I Chapter 5 and continuing through much of the rest of the work, *Surveying Subjective Phenomena* concentrates, a section at a time or a chapter at a time, on the work of specific individuals or groups of authors. Volume I Chapter 5, for instance, summarizes early work (e.g., Cantril), contemporary work (e.g., Schuman, Presser, and Associates), and new results showing empirical patterns of disagreement among subjective survey questions. The topics covered in this chapter and in the follow-up piece in Volume II Chapter 7 are not as extensive, nor the analysis as probing, as the book-length treatments of these topics that are summarized in the chapter or that have been published subsequently.

Volume I Chapter 6, on the other hand, is a previously unpublished, mathematically advanced analysis of survey data using the Rasch model for item-centered and respondent-centered analysis. One wonders why this chapter is included and the reason for its placement in Volume I of *Surveying Subjective Phenomena*. The other chapters discussing measurement error do not hint at a Rasch model solution. They are not written in a way to motivate its selection as a tool to manage the complexities of analysis and interpretation that are brought forth. The Rasch model is brought forward to "call attention to some approaches to scientific analysis of survey data that are either novel or underused (Vol. I p. 179)." This is a weak justification, and its presentation seems out of place.

A number of other chapters, notably in Volume II, have this same feeling of being out of place. The reader is struck by how very little relationship there is between the mathematically technical chapters on models of subjective error measurement and the inductive, analytic chapters on patterns of results. The technicians in the subjective error field seem to be stuck on problems that have to do with measuring item and category response metrics (Rasch, latent class, etc.). On the other hand, the inductivists (who actually design and administer a lot more surveys) are stuck on problems of conceptual clarity and definition.

A number of chapters or sections raise, for the record, significant issues in subjective measurement, but have little to say beyond acknowledgement of the problem. Volume I Chapter 7, for instance, raises the issue of conceptual ambiguity in surveys: "if the concepts used in survey questions are not understood in the same way by the survey researcher and the respondent, then responses to the questions are likely to be misinterpreted by the researcher (Vol. I p. 235)." The chapter does not make a clear statement on this issue. It consists of: (1) an extremely general note on the definition of "public opinion"; (2) a lively discussion of what people mean when they use the word "risk"; (3) some examples where interpretations of questions apparently were influenced by the set of response categories; and, (4) an extremely brief description of a technique called ethnographic semantic mapping. The reader, I believe, would be better off with more discussion of this topic, or less. It is as if the editors did not know where else to put these pieces, did not want to leave them out, and were not suffering from the discipline of an overall page limitation.

Volume I Chapter 8 is a review of what is known about the effect of respondent-interviewer social dynamics during an interview. The conclusion is that "variability in the social aspects of the interview situation results in variability in respondents' role expectations and behaviors during interviews (Vol. I p. 273)." But the patterns are inconsistent: "we have only begun to understand how the interview, viewed as a social relationship, influences responses to survey questions (Vol. I p. 274)." Volume II Chapter 9 explores a related terrain: "social desirability ... the notion that some things are good and others are bad, and the notion that respondents want to appear "good" and answer questions in such a manner as to be perceived that way (Vol. II p. 258). The conclusion is similarly vague: "conceptual ambiguities plague the notion of social desirability (Vol. II p. 276)." These chapters, like many others, are notable for their examples but not for their conclusions. After reading several hundred pages like this, one begins to grasp how subjective measurement is; a field rich with rules of thumb about survey design (and other forms of folklore), but a "scientific" understanding of the process may be so complex and so expensive that it will never be achieved.

Volume I Chapter 9 consists of notes on "psychological" sources of bias in the question and answer process. Some of these sources of bias include subtle cues of grammatical structure, affective connotations of particular words, or mood changes induced by positively- or negatively-worded questions. As a sociologically-trained researcher, I find this one of the more fascinating chapters because the point of view on the nature and competence of the respondent is so different from what I am used to. One of the lines of research reviewed in the chapter takes a bald position against the use of any sort of introspective reports because "people do not necessarily have privileged knowledge of their own attitudes, motives, or the causes of their behavior (Vol. I p. 298)." In a year of mud-slinging election advertising in the United States, I find fascinating the suggestion in another study reviewed in this chapter that people often cannot verbalize the reasons for their likes and dislikes because "the salient and notable features of an object are not necessarily the same features that feelings are attached to (Vol. I p. 299)." The conclusion of the chapter attempts to strike the ball right out of the park: "uncritically accepting respondents' stated purposes or motives as valid and basing a full-scale analysis on them is a risky strategy, at best (Vol. I p. 300)." It is too bad that the entire book was not more cohesively constructed so that some of the implications of the statements made in this chapter could be explicitly addressed in other parts of the discussion.

Volume I Chapter 10 contains the 18 recommendations from the panel's multi-year effort. I will not summarize them here because they do not flow directly from the preceding 300 pages nor from the 800 pages of special studies that follow. The recommendations have mostly to do with how the profession of survey research ought to be institutionalized, managed, funded, and monitored in a market economy. Those who want recommendations on how to do surveys better will have to look at the research results in individual chapters and not at the panel's 18 recommendations.

The panel advocates the strongest possible recommendations regarding public education, industry regulation, and subsidies for methodological research. The panel hopes its recommendations will offset some facts about the survey profession: (1) public opinion polling is a competitive industry in the United States and in other countries; (2) large sums of money are not likely to be forthcoming from private sources for methodological research, and, (3) neither national governments nor professional associations are in a position to enforce strict guidelines for the polling profession. Given these facts, it is unclear what effect the panel and its list of recommendations will have on the priorities and conduct of the survey profession.

References
Mosteller, F., Hayman, H., McCarthy, P., Marks, E., and Truman, D. (1949): The Pre-Election Polls of 1948. Social Science Research Council, New York.
Rogers, L. (1954): The Pollsters. Basic, New York.
Rossi, P. H., Wright, J. D., and Anderson, A. B. (eds.) (1983): Handbook of Survey Research. Academic Press, New York.
Wheeler, M. (1976): Lies, Damn Lies, and Statistics. Liveright, New York.

D. Garth Taylor
Chicago Urban Leage
Chicago, IL
U.S.A.

Dey, A., Orthogonal Fractional Factorial Designs. Wiley Eastern Limited, New Delhi, 1985. ISBN 0-85226-165-9. viii + 133 pp., £8.40.

Even the most advanced courses in the design of experiments do not go very deeply into fractional factorials, apart from the traditional 2^{n-k} and 3^{n-k} series. There may perhaps be a passing reference to the 4^5 and 5^6 orthogonal main effects plans based on sets of mutually orthogonal latin squares, but there is rarely time for anything more.

The few topics mentioned above represented the frontier in fractional factorials until the 1960s. Little further progress was made on asymmetrical fractions until the work of Addelman and Kempthorne (1961 a and b) and Margolin in (1968; 1969 a, b, and c; 1972). Since then considerable advances have been made by several statisticians, including Professor Dey himself. Their work has appeared in various journals, among them *Technometrics*.

The interest of *Technometrics* in this work should not be surprising because the past decade has seen a surge in the use of orthogonal fractions by engineers who are involved in modern quality assurance and process improvement. Until lately, they have had available to them only orthogonal main effects plans for two and three factors, whose derivation has often been wrongly attributed to Taguchi. But the main effects plans are not enough: engineers need to have access to good resolution IV designs.

Dey has gathered together the results on orthogonal fractional factorials obtained in the past forty years or so, producing a short but useful synthesis. There is an enormous amount of interesting information, especially about asymmetrical fractions. Reasoning, no doubt, that the interested reader can find the derivations of the procedures in the original papers, the author does not repeat the proofs. However, he does provide some examples of those techniques, including, for example, a helpful discussion of the derivation of the design of Bose and Bush for $3^9//27$ (meaning 9 factors at 3 levels in 27 runs). It is good to see all these methods brought together in one volume.

Unfortunately, retrieval of the information in the book is a problem. The author has added tables, which he calls indexes, that are intended to help the reader find in the text procedures for constructing appropriate designs. I tested them on two examples. First, I tried to find the lattice for $3^7//18$. It was not listed in Table 2.3, "Index of Orthogonal Main-Effect Plans for Symmetrical Factorials," which referred me instead to the 16 run fraction by Stark, but did not tell me where to find it. Happening to know that this lattice can also accomodate a two level factor, I next looked up the $3^7.2$ lattice with 18 points in Table 3.4, "Index of Orthogonal Main-Effect Plans for Asymmetrical Factorials," and was referred to Section 3.3, where I failed to find it. Finally, I tried $6.3^6//18$ and was sent to Section 3.4.3 where I found it on pages 58 and 59. There Dey mentions that it was derived from the lattice of Addelman and Kempthorne (1961 b), but does not tell where to find the derivation of that design. (It is actually derived in a well-written section starting on page 29). Obviously, this book sorely needs a proper index.

My second attempt was to find a resolution IV design for $3.2^4//24$. I found the reference in

Table 4.5, "Index of Orthogonal Resolution IV Designs," which referred me to Section 4.3.1. There I found a procedure for $t.2^{n-1}$ fractions, and could have used some help. After staring at it for a while it dawned upon me that I should attack it as a 6.2^3 problem. The information is there, but it is hard to find.

This is an interesting book and I am glad to have read it, but I wish it were not so condensed. The first chapter is a very short introduction to the topic, marred by several typographical errors. The author's style is so terse that any but the mathematically sophisticated reader will find it hard work indeed. This book could prove helpful to the mathematical statistician who is engaged in experimental design and might be asked sooner or later for an orthogonal resolution IV fraction of an asymmetrical factorial. I am not sure that this book will provide the answer, but it should point the right direction in the literature.

References

Addelman, S. and Kempthorne, O. (1961 a): Orthogonal Main Effect Plans. Aerospace Research Laboratories Report No. 79.

Addelman, S. and Kempthorne, O. (1961 b): Some Main Effect Plans and Orthogonal Arrays of Strength Two. Annals of Mathematical Statistics, 32, pp. 1167–1176.

Margolin, B. H. (1968): Orthogonal Main Effect 2^n 3^m Designs and Two Factor Interaction Aliasing. Technometrics, 10, pp. 559–573.

Margolin, B. H. (1969 a): Results of Factorial Designs of Resolution-IV for the 2^n and 2^n 3^m Series. Technometrics, 11, 431–444.

Margolin, B. H. (1969 b): Resolution-IV Fractional Factorial Designs. Journal of the Royal Statistical Society, Series B, 31, pp. 514–523.

Margolin, B. H. (1969 c): Orthogonal Main Effect Plans Permitting Estimation of All Two Factor Interactions for the 2^n 3^m Series of Designs. Technometrics, 11, pp. 747–762.

Margolin, B. H. (1972): Non-Orthogonal Main Effect Designs for Assymmetrical Factorial Experiments. Journal of the Royal Statistical Society, Series B, 34, pp. 431–440.

Peter W. M. John
University of Texas
Austin, TX
U.S.A.

Kapadia, R. and Andersson, G., Statistics Explained. Basic Concepts and Methods. Ellis Horwood Limited, Chichester, 1987. ISBN B-7458-0053-X, 0-7458-0315-6, 0-470-20966-6. 234 pp., £ 12.95.

In the field of statistics, the bulk of the literature discusses statistical methods and techniques and the intended audience is the producers of statistics. On the other hand, statistical literature written for users of statistics, literature that addresses the users' needs of interpreting, understanding, and critically examining their data is indeed scarce. *Statistics Explained: Basic Concepts and Methods* mainly directs itself to British readers and provides many insights into the possibilities and restrictions of statistics.

From a pedagogical point of view, the book has an excellent presentation. Each chapter is organized in a systematic fashion. First, a certain problem area is introduced using examples. After that comes a discussion of the "what-how-where" of the data; i.e, a discussion of how the data were collected and the questions that the data were meant to answer. This is followed by a short presentation of the statistical theory used in this particular example. And lastly, some interpretation of the data is reached and this interpretation is evaluated. Each chapter concludes with empirical computer exercises written for Minitab; also there are other exercises and solutions.

Because the theoretical parts are interwoven with a rich amount of text and references, the book also becomes accessible to people who have a restricted knowledge of mathematics. The book thus follows the classic Anglo-American tradition of being both easy to read and of interest for a wide circle of readers. It seems to me that this book should be read by everyone who work in media, such as journalists, and even politicians and researchers in other fields.

My personal reflection is that this book paves the way for a follow-up or companion volume, tentatively named *Making Inferences* with a similar organization and presentation. The question of making inferences is so essential from several aspects that it deserves an entire book and not only an abbreviated chapter as in the present book.

Per Nilsson
Statistics Sweden
Stockholm
Sweden

Matérn, B., Spatial Variation. Lecture Notes in Statistics, 36, Springer-Verlag, Berlin, 1986, ISBN 3-540-96365-0 (Springer-Verlag Berlin), ISBN 0-387-96365-0 (Springer-Verlag New York). 151 pp., DM 33.00.

The first edition of *Spatial Variation* was published in 1960 as *Meddelanden från Statens Skogsforskningsinstitut* 49:5 as Matérn's doctoral thesis. Publication in this form gave it limited circulation, and Springer has now issued a facsimile reproduction of the original together with author and subject indices and a three-page Postscript giving a summary of more recent developments.

It is almost unbelievable how far Matérn was ahead of his time. Some of the theoretical work was started in 1948, and by 1960 Matérn had completed an overview of the theory of two-dimensional random fields and point processes and applied these ideas and those of geometrical probability to sampling problems in forestry. It is this emphasis on sampling, and in particular on the precision of sampling designs, which distinguishes *Spatial Variation* from all subsequent books in spatial statistics. One would expect a 1960 research monograph to be completely outdated by now, but in Matérn's case this is far from so. Matérn quotes only a few additional references on spatial sampling in his postscript, and I am aware of only a handful of others. In part this is testimony to the completeness of his approach.

Chapter 1 is (the original) introduction. The second chapter gives an introduction to stationary stochastic processes on R^n, now more often referred to as random fields. Modern readers may find this difficult. It is stated to be a review, but the material (especially characteristic functions) is no longer emphasized in courses and texts on probability theory. (The first place I encountered a Bessel function was whilst a graduate student reading the first edition!) Section 2.6 sketches the idea for random measures, a subject developed in depth in later years by Olav Kallenberg. In his postscript, Matérn seems to confuse this with the random set theories of Kendall and Matheron, which are quite distinct. The countable additivity which Kallenberg added to Matérn's postulates of finite additivity and second-order stationarity is necessary to avoid a measure-theoretic quagmire. It is also a key part of the reduction of moment measures exploited by Krickeberg and the reviewer.

Chapter 3 discusses some specific mechanisms for constructing stochastic processes. These provide one of the most comprehensive catalogues available of stochastic processes with specified correlation functions, which has proved invaluable in these days of extensive simulation. Later sections introduce some widely used models of point processes and random sets. This chapter contains an amazing richness of ideas, many of which are only now being exploited. The one important idea which is not present is that of Gibbsian point processes (Ripley (1988)). These remove the awkwardness of the "more regular than Poisson" processes of § 3.6.

The fourth chapter begins the more practical half of the study by considering what spatial correlograms occur in practice. Matérn uses a few artificial examples to suggest that an exponential correlation function is appropriate, and then in Chapter 5 considers the efficiencies of stratified and systematic sampling schemes under this correlation function. Since exact calculations were too much for the computers of the 1950s, a number of clever approximations are used. Today the exact calculations can be done without difficulty. The actual values are rather different, but the qualitative conclusions (to use systematic sampling on a rectangular grid) remain unchanged. Chapter 6 is a miscellany of calculations related to practical sampling problems in forestry. I have always found this impenetrable. The topics are only very loosely related and there are few firm conclusions. In part it is a commentary on Matérn's 1947 essay (in Swedish, and even less available than the first edition of *Spatial Variation*).

The field of spatial statistics has been changed radically by the computer revolution, so it is no surprise that it is the more general and theoretical Chapters 2 and 3 which have endured best the passage of 25 years. Indeed, they have never been superceded as a reference for general stationary isotropic random fields. The work on Chapter 5 on sampling plans is still a model of what can be done with simple calculations, and has proved to be an inspiration to the geostatistics school. More extensive experience has suggested that the exponential correlation function is less widely applicable than Matérn implies, and that both border effects and long-range correlations need to be taken more seriously than their easy dismissal here.

The postscript is the weakest part. Perhaps it is unfair to expect Matérn (who was by then retired) to be aware of all the recent developments in spatial statistics, but a good deal of the commentary is ill-informed and lacks the incisiveness of the original material. Much more could be made of the developments in geostatistics (Journel and Huijbregts (1978)) and random processes have been quite widely proposed in the design and analysis of field trials (e.g., Bartlett (1978); Besag and Kempton (1986); Wilkinson, Eckert, Hancock, and Mayo (1983)).

Spatial Variation is certainly of historical interest, and a testament to Matérn's vision. But it is considerably more than that. Although not a suitable introduction to spatial statistics, Chapters 2 to 5 are compulsory reading for anyone with aspirations to specialize in the area. I know very little about forestry applications, but suspect that the wisdom on spatial sampling in this volume is still not widely applied outside Sweden.

References

Bartlett, M. S. (1978): Nearest Neighbour Models in the Analysis of Field Experiments. Journal of the Royal Statistical Society, Series B, 40, pp. 147–174.

Besag, J. and Kempton, R. (1986): Statistical Analysis of Field Experiments Using Neighbouring Plots. Biometrics, 42, pp. 231–251.

Journel, A. G. and Huijbregts, C. J. (1978): Mining Geostatistics. Academic Press, London.

Matérn, B. (1947): Metoder att uppskatta noggrannheten vid linje- och provytetaxering. Meddelanden från statens skogsforskningsinstitut, 36, nr 1.

Ripley, B. D. (1988): Statistical Inference for Spatial Processes. Cambridge University Press, London.

Wilkinson, G. N., Eckert, S. R., Hancock, T. W., and Mayo, O. (1983): Nearest Neighbour (NN) Analysis of Field Experiments. Journal of the Royal Statistical Society, Series B, 45, pp. 151–211.

Brian. D. Ripley
University of Strathclyde
Glasgow
U. K.

Stone, M., Coordinate-Free Multivariable Statistics: An Illustrated Geometric Progression from Halmos to Gauss and Bayes. Oxford University Press, Oxford, 1987. ISBN 0-19-852210-X, xiv + 120 pp., £ 12.00.

This is a nice book and I enjoyed reading it. Multivariate linear statistics is developed using vector spaces and transformations. The book is mathematical. The theory is done in the style of Halmos. A vector space of variables is given. The observations are linear real-valued operators on this space. For example, an observed person NN operates on the variable annual income giving the real number 178 000 SEK. Mean values are also operators on the variable space. The variance is an inner product on the same space. The dual space, where the observations lie, is called the evaluator space.

The book is not directly useful for a practising statistician. He or she will not learn any new methods or learn much about the old ones. However, they will gain some insight into the structure of multivariate linear statistics. They will get a new way of looking at multivariate problems, in particular on the proofs of the basic theorems. A special trait is the use of pictures even for complicated high-dimensional situations. The pictures are sometimes integrated into the proofs of the theorems. The book is probably intended for the graduate level.

The book is in another way quite elementary. It does not require much knowledge of statistics. It does not treat more than the basic theory. For example, principal components, factor analysis, cluster analysis, and discriminant analysis are not mentioned. The book does not even consider estimation when the covariance matrix is not proportional to something known (the Behrens-Fisher problem). The theory can thus not be applied directly to subjects like stratified sampling or nested factors.

I would recommend the book to a mathematician who wants to learn multivariate statistics and who is prepared to read at least one more book in the field. A mathematical statistician, who is used to think of linear statistics in terms of vector spaces, should also benefit from the book and may even enjoy it.

Daniel Thorburn
University of Stockholm
Stockholm
Sweden

Printed July 1989