Journal of Official Statistics
Vol. 4, No. 2, 1988, pp. 191–199

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.

Gelfand, A.E. (Ed.), Contributions to the Theory and Application of Statistics: A Volume in Honor of Herbert Solomon. Academic Press, Inc., Boston, 1987, ISBN 0-12-279450-8, xxviii + 544 pp., $ 59.95.

In 1984, Herbert Solomon celebrated his 65th birthday and his 25th year at Stanford University. To his honor, this volume of contributed papers was prepared and published. Like most books of this type, the contents are heterogeneous. This also reflects the diversity of Solomon's own interests and contributions, including, for instance, operations research, geometrical probability, multivariate analysis, and "jurimetrics." After a biographical sketch and a list of Solomon's publications, the book provides twenty papers grouped into four sections. The contents are well summarized by the following list of informative titles.

1. Operations Research and Applied Probability

Inequalities for Distributions With Increasing Failure Rate (M. Brown).
A Markov Decision Approach to Nuclear Materials Safeguards (H. Chernoff & Y. Yao).
On the Persistent Release of Particles in a Fluid Flow (J. Gani & P. Todorovic).
Statistical Inference for Random Parameter Markov Population Process Models (D. Gaver & J.P. Lehoczky).

2. Distribution Theory and Geometric Probability

Probabilistic-Geometric Theorems Arising From the Analysis of Contingency Tables (P. Diaconis & B. Efron).
Some Remarks on Exchangeable Normal Variables With Applications (S. Geisser).
Asymptotics for the Ratio of Multiple t-Densities (S.J. Press & A.W. Davis).
Periodogram Testing Based on Spacings (A.F. Siegel & J. Beirlant).
Tests for Uniformity Arising From a Series of Events (M.A. Stephens).
Spatial Classification Error Rates Related to Pixel Size (P. Switzer & A. Venetoulias).

3. Applications

The Use of Peremptory Challenges in Jury Selection (M.H. DeGroot).
An Information-Processing Model Based on Reaction Times in Solving Linear Equations (J.B. Kadane, J.H. Larkin & R.E. Mayer).
Diagnostic Errors and Their Impact on Disease Trends (M.A. Kastenbaum).

Hypothesis Testing in the Courtroom (D.H. Kaye).
Multivariate Discrimination of Depressive Groups Across Cultures (J.E. Mezzich & E.S. Raab).

4. Inference Methodology

Estimation in Parametric Mixture Families (A.E. Gelfand).
Multiple Shrinkage Generalizations of the James-Stein Estimator (E.I. George).
The Analysis of a Set of Multidimensional Contingency Tables Using Log-Linear Models, Latent-Class Models, and Correlation Models: The Solomon Data Revisited (L.A. Goodman).
Selection Procedure for Multinomial Populations with Respect to Diversity Indices (M.H. Rizvi, K. Alam & K.M.L. Saxena).
Confidence Intervals for the Common Variance of Equicorrelated Normal Random Variables (S. Zacks & P.F. Ramig).

Gani and Todorovic discuss interesting applications of probability, although the paper is of less importance than previous work on the same topic by the authors.

Gaver and Lehoczky study transitions in a population between "colonies" or "compartments," where each colony has an unknown migration rate parameter. The likelihood for a "simple Markov population process" (SMPP) is easily seen to be of multivariate independent Poisson type. The authors also allow heterogeneity by assuming that several realizations of an SMPP are observed, corresponding to rate vectors independently drawn from a superpopulation. First, Bayesian posterior distributions are calculated for the realized rate parameters under a few completely specified superpopulations. Next, a bit more realistic, these superpopulations are allowed to depend on unknown parameters, simultaneously estimated by empirical Bayes methods. Formulas are derived, but the paper does not provide much insight.

Diaconis and Efron show relations between some classical probabilistic theorems. This is an entertaining paper in the spirit of Feller's books.

Stephens contributes a discussion of tests for the Poisson process property of a series of events. He reviews and compares conventional and special tests against various types of alternatives. The paper appears to be a valuable addition to the literature on this subject.

Kastenbaum discusses data quality in medical diagnoses and death certificates, considering, in particular, cancers in the United States. Misclassifications and varying coding practice affect official vital statistics and can make conclusions about disease or mortality rates questionable. The rate of autopsy confirmation in the United States has decreased, which renders assessment of error frequencies even more difficult. As his main example, Kastenbaum discusses the differences between the 1968 and 1977 U.S. lung cancer death rates. Of particular value is the concluding chronological bibliography on the subject, containing more than a hundred references from 1955 to 1984.

The contribution by Gelfand concerns admissibility, Bayes estimation, and empirical Bayes estimation in parametrically specified mixtures of a distribution family. A typical example of the paper is obtained by scaling a chi-square distribution by a scaling factor that has a Poisson mixing distribution with an unknown intensity parameter. The paper is unconvincing about the method's potential for applications.

The paper by Goodman is the longest paper in this collection. Starting from a 2^5 cross-classification data set analyzed by Solomon in 1961, Goodman reviews some models for categorical data, of log-linear, linear, and latent-class type. The presentation is elementary and in some respects detailed. It provides references to previous papers by the author rather than new ideas.

Behind this collection is a distinguished group of authors. However, the heterogeneity of subjects in combination with a specialized and technical character makes the book a natural purchase only for statistical libraries. Several contributions are reports of the current status of ongoing research and are likely soon to be outdated by more comprehensive or conclusive papers by the same authors in other publications. Unfortunately, the text is only of typewriter quality.

Rolf Sundberg
University of Stockholm
Stockholm
Sweden

Kish, L., Statistical Design for Research. John Wiley & Sons, Inc., New York, 1987, ISBN 0-471-08359-3, xxii + 267 pp., $ 34.95.

This book provides an excellent description of issues that must be considered in developing or evaluating designs for empirical investigations. Three categories of designs are assessed, namely, experimental designs, sample surveys, and observational studies. The material is kept nontechnical and concise with abundant references to other sources for technical details. This book can be read by any researcher regardless of his or her statistical training.

In Chapter 1, criteria are developed to compare design options based upon what should be the 3 R's of good design: Representation, Randomization, and Realism. These criteria aim at creating the design closest to the ideal, i.e., one that facilitates unbiased estimation and testing. The three basic design categories are described in relation to this ideal. I particularly appreciated Kish's distinction between the goals and aims of survey sampling and those of experimental design.

Chapter 2 expands on these concepts in describing the design features of survey samples. Four population levels are defined and the effects of nonresponse and undercoverage are distinguished for each level. Key to sample design is the use of design effects. Kish presents a thorough description of this concept for estimates based upon both the total sample and on subclasses. Then, design characteristics such as clustering and stratification are described and their impact on the design effect summarized.

Observational studies are often used when probability sampling is impractical. Five basic designs for observational studies are presented in Chapter 3. Models are developed to contrast these designs in terms of variance and bias for a fixed total cost. Four major types of bias are distinguished and the ability of each design to reduce or eliminate each type of bias is discussed. While I liked the model-based comparison of the designs, I thought the presentation was not as straightforward as it could have been nor was it as easy to read as the rest of the book. Total cost is set to what would be achieved by the one-shot case study. Thus, other designs include sample size reduction factors in their variance expression to reflect the additional costs associated with design features such as control groups, pretests, etc. Many of the studies that I have been involved with have demanded fixed precision instead of cost. To use the material presented here would require reworking the variance expression and developing an explicit cost model.

Unfortunately, the author does not include a review of the conceptual principles underlying experimental designs similar to the presentations in Chapters 2 and 3 for surveys and observational studies. While the mathematical underpinnings of experimental designs are treated in numerous texts, a discussion of the thought processes that lead to choosing one design over another is often lacking in these texts. Kish's book would have benefitted from such a discussion, using the same unified approach as he uses elsewhere in his book.

Extraneous variables can disguise treatment effects or population characteristics of interest. Chapter 4 provides an excellent summary of methods for controlling the biasing or variance-inflating effects of extraneous variables. These methods include design features such as stratification in surveys, or blocking in experimental designs, and analysis procedures such as post-stratification and standardization. Guidelines are presented for choosing the best set of control variables and for deciding whether to control by design or by analysis.

Samples and censuses are contrasted in Chapter 5. Various spinoffs of these designs are discussed, for instance, registers and samples associated with censuses. The presentation is not adequately specific, which limits the usefulness of this chapter. For instance, small area estimation is alluded to but not fully addressed.

Sample designs measuring population characteristics that vary with time is the topic of Chapter 6. The characteristics of panel, repeated, periodic and overlapping surveys are described in terms of variance reduction and bias control. The presentation is detailed and, to my knowledge, not readily available elsewhere.

Chapter 7 consists of miscellaneous material the author thought too technical to present

elsewhere. There are some good nuggets of information that make the chapter worthwhile. For instance, the practical requirements for a measurable survey design are summarized and strategies for satisfying multiple survey objectives are explained.

Overall, I found this book highly readable. It presents concepts of design that researchers have had to discover by trial and error or through discussions with more experienced colleagues. I believe this book would be valuable reading for practicing statisticians. Professors teaching classes in experimental design, sampling, or social research are encouraged to review these concepts with their students in addition to materials traditionally presented. In short, this book may not be perfect but it is very good and I recommend it.

Brenda G. Cox
Research Triangle Institute
Research Triangle Park, NC
U.S.A.

Little, R.J.A. and Rubin, D.B., Statistical Analysis with Missing Data. John Wiley & Sons, Inc., New York, 1987, ISBN 0-471-80254-9, xiv + 278 pp., $ 34.95.

Standard methods for the analysis of statistical data usually assume that all intended observations are available. Hence, the observations can be stored in a rectangular matrix, the rows representing records or cases, and the columns variables. However, for various reasons values may be missing in this matrix; respondents in a sample survey may refuse to answer some or all of the questions, or controlled experiments may break down due to mechanical failures.

The aim of the book is to survey current methods for the treatment of missing data. Particularly in survey sampling, missing data caused by nonresponse is a serious problem. Since nonresponse rates tend to increase, the problem becomes more and more serious. Therefore, this book is important for those who are involved in both theoretical research and practical applications of survey sampling.

The first part of the book covers the history of the missing data problem in three important areas of statistics: analysis of variance, survey sampling, and multivariate analysis. The introductory chapter starts with a general overview. One approach is to discard all incomplete records and use only complete records in the analysis. For a small number of missing values this approach may work satisfactorily. However, if the number of missing values is substantial a serious bias may be introduced. A second approach is to use imputation, i.e., to insert fictitious values where observed values are missing. Since imputed values have a different status from observed values, standard analysis must be modified in order to account for this difference. A third approach is to apply some kind of weighting technique. Weights are assigned to available observations so that the weighted observations are adjusted for missing observations. A fourth and final approach is to define a model for the mechanism that causes missing observations. Inference based on such a model can account for the missing values, but the validity of the model should, of course, be checked.

Vital for proper treatment of missing data is whether or not there is a relation between the missing data mechanism and the variable to be investigated (Y). It is very important to have auxiliary variables (X) which are not affected. In the first part of the book three cases are distinguished: (1) the missing data mechanism depends on Y, and possibly X as well, (2) the mechanism depends on X but not on Y, and (3) the mechanism is independent of X and Y. Case (3) is denoted by "missing completely at random" (MCAR), and case (2) by "missing at random (MAR) within subclasses of X." This case corresponds to what is called an "ignorable" missing data mechanism.

Chapter 2 reviews methods to take care of missing values in the dependent variable in ANOVA. A straightforward method would be to discard cases with missing data and carry out the analysis on the remaining, complete cases. This approach has the disadvantage that balanced designs become unbalanced, and this increases the computational effort considerably. A better approach is to estimate the missing values, and to carry out the analysis on all cases. Various techniques are discussed.

Chapter 3 concentrates on multivariate analysis. Here, a complete case analysis works well only under MAR, otherwise estimators may have a severe bias. Furthermore, the number of discarded cases may be so substantial that hardly any are left for analysis. Available case methods use all available cases for a particular estimator. A consequence is that different estimators may be based on different sample sizes and different cases, which complicates the comparison of results. Furthermore, estimates of correlations may be outside [–1, 1], and covariance matrices may not be positive definite. Another approach is to fill in missing values (imputation). Mean imputation and regression imputation may produce unbiased estimators, but estimated variances are often not correct. The authors do not recommend these methods. Their performance is unreliable, and the required adjustments are too ad-hoc to yield satisfactory results.

Chapter 4 discusses methods to deal with nonresponse in survey sampling. Two different ways of nonresponse modeling are used interchangeably, which is a little confusing. Sometimes the "fixed response approach" is used in which the finite population is divided in a response and a nonresponse stratum, whereas at other points each element in the population is assumed to have a certain, unknown probability of response (the "random response approach"). Two approaches to nonresponse adjustment are considered: weighting cell adjustment and imputation. A special but important case of weighting, raking, receives little attention. Recent results with respect to the variance of raking estimators are not mentioned. Several types of imputation are discussed (mean, random, regression, hot deck, cold deck). Many of the imputation methods have problems with variance estimation. Therefore, some attention is paid to variance estimation based on ultimate clusters. One practical problem with imputed data sets is not discussed, namely, that other users of the data set may believe that they have a complete data set. Analysis of such data sets, particularly on certain selected subsamples or variables, may produce erroneous results. Another approach to tackling nonresponse is not mentioned here: resampling of nonrespondents. Resampling produces indications of whether estimators are biased, and these results can be used to correct for bias.

The second part of the book presents a systematic approach to the analysis of data with missing values. Inference is based on maximum likelihood (ML) techniques derived from statistical models for the data and the mechanism causing data to be missing. Chapter 5 starts with a summary of the maximum likelihood theory for the complete case. It is shown that for large samples and under assumed asymptotic normality, this approach also produces valid results for the incomplete case under MAR. Another approach is to treat the missing data as parameters which can be estimated using ML methods. In practice, this approach is only useful if the fraction of missing values tends to zero.

Chapter 6 discusses methods based on factoring the likelihood so that each factor corresponds to the likelihood for a complete case or for an easier missing data problem. An interesting result is that for the case of bivariate normality, this method is the same as regression estimation, a common estimation technique in survey sampling.

Since in general, missing data patterns are not simple, ML estimation is not always possible. Chapter 7 considers iterative methods to obtain estimates. Special attention is paid to the expectation maximization (EM) algorithm. The idea is to alternatively estimate the missing data and the parameters of the model until convergence. A practical problem in applying this algorithm is the convergence speed, which is proportional to the fraction of missing data.

In Chapters 8, 9, and 10, ML and EM techniques are applied to various problems involving incomplete data. Chapter 9 concentrates on continuous variables, under the assumption of multivariate normality. Applications considered are factor analysis, variance components, linear regression, and time series. Chapter 9 deals with categorical variables, under multinomial and log-linear models. Chapter 10 discusses mixed normal and categorical data. In all cases, the assumption of ignorability proves vital.

The final chapter is devoted to a model-based approach to nonresponse in sample surveys. Special attention is paid to multiple imputation.

The book is not restricted to the missing data problem in survey sampling, and there-

fore, not all theories and applications within this field are discussed. For those concerned with sample surveys only, Madow, Nisselson, Olkin, and Rubin (1983) is an important additional source. Still, the book gives a very good overview of the missing data problem in general, and the state of the art in dealing with missing data. It would be valuable, however, to see the techniques implemented and tried on real data. It is not a simple book. A good background in many aspects of statistics is necessary to understand all subjects treated. Many estimation procedures are presented and discussed, but practical statisticians may need more concrete algorithms or references to existing computer programs. The book should be studied in the statistical methods department in every statistical agency.

References

Madow, W.G., Nisselson, H., Olkin, I., and Rubin, D.B., Eds. (1983): Incomplete Data in Sample Surveys (Vols. 1–3). Academic Press, New York.

Jelke G. Bethlehem
Netherlands Central Bureau of Statistics
Voorburg
The Netherlands

McLachlan, G.J. and Basford, K.E., Mixture Models: Inference and Applications to Clustering. Marcel Dekker, Inc., New York, 1988, ISBN 0-8247-7691-7, xi + 253 pp., $ 69.75 (U.S. and Canada), $ 83.50 (all other countries).

In the last twenty years the quantity of literature on cluster analysis has increased dramatically and a number of books have appeared on the subject including those of Hartigan (1975), Everitt (1980), and Gordon (1981). Each of these has given brief descriptions of an approach to clustering based on finite mixture distributions, particularly mixtures of normal distributions. The text by McLachlan and Basford provides a more comprehensive account of this approach to clustering.

After a brief introduction to the history of mixture models, the authors quickly start their discussion of estimation via the expectation-maximization (EM) algorithm, identifiability, and tests for number of components. Chapter 1 also contains an account of the properties of maximum likelihood estimators including the computation of the associated information matrix. Later chapters cover normal mixtures in some detail. Other types of mixtures such as latent class distributions are covered in somewhat less detail, as are the estimation of mixing proportions, the partitioning of treatment means in ANOVA, and the maximum likelihood approach to the clustering of three way data. Much of the material is technically demanding; however, the many interesting numerical examples provided are helpful in clarifying the more difficult points. A topic not covered but one which would be of practical importance is mixture models for data containing both continuous and categorical variables. The text contains a valuable bibliography with approximately 450 references, most of which are relatively recent, and a number of FORTRAN listings which may be of interest to researchers.

On the whole, the book is well written and well produced and contains much that will be of interest to applied and research statisticians and to those in other disciplines involved in the application of clustering techniques. The text's use for students is limited both by the level of the material and by the lack of suitable exercises.

There are two other texts covering mixture distributions which might be considered competitors – Everitt and Hand (1981), and Titterington, Smith, and Makov (1985). McLachlan and Basford's book is, however, different from both, since it concentrates on the use of mixture distributions as models for the classification process. The book would be a useful addition to the library of those statisticians interested in the theory or application of clustering.

References

Everitt, B.S. (1980): Cluster Analysis. Second edition, Gower Press, London.

Everitt, B.S. and Hand, D.J. (1981): Finite Mixture Distributions. Chapman and Hall, London.

Gordon, A.D. (1981): Classification. Chapman and Hall, London.
Hartigan, J.A. (1975): Clustering Algorithms. Wiley, New York.
Titterington, D.M., Smith, A.F.M., and Makov, U.E. (1985): Statistical Analysis of Finite Mixture Distributions. Wiley, New York.

Brian S. Everitt
University of London
London
U.K.

Searle, S.R., Linear Models for Unbalanced Data. John Wiley & Sons, Inc., New York, 1987, ISBN 0-471-84096-3, xxiv + 536 pp., $ 49.95.

Searle's popular and successful (1971) text has introduced a generation of students to the linear models which underlie factorial experiments. It advocates use of the overparameterized singular linear model, which has the virtue of containing easily interpreted individual parameters for all main effects and interactions, in exchange for considering such concepts as generalized inverses of matrices and estimable vs. non-estimable parameters. Other distinguishing characteristics of the (1971) book are its devotion to the analysis of unbalanced data (unequal numbers of observations in the sub-most cells), and an unusually clear and meticulous (though at times verbose) writing style. The estimation and testing procedures in the widely-used SAS-GLM computing procedures owe a heavy debt to Searle's approach in the (1971) text.

Not surprisingly, the (1971) book has become dated by subsequent developments. Despite its many strong points, its continued use as a text has become difficult to justify. Chief among these newer findings has been the promotion since the mid 1970s by R. R. Hocking and others of the superiority in most contexts of the full-rank cell means approach to linear modeling.

With the publication of Hocking's (1985) linear models text, there was at last an attractive up-to-date alternative to Searle (1971). However, the former is written at a somewhat higher level of mathematical sophistication and abstraction. Searle's new text fills the gap. It is current, and as expected, contains his lucid writing throughout. And Searle now also generally favors the cell means model.

The present text is, in a sense, two books in one. The first five chapters cover modeling and analysis of one-way and two-way fixed effects layouts having unequal (and possibly zero) cell sizes, using elementary algebra and no matrix notation. Analyses with covariables in the one-way set-up are studied in Chapter 6, and Chapter 7 discusses results in matrix algebra and quadratic forms used in the remainder of the book. To this point the presentation is less demanding and sophisticated than that in Searle (1971), but beginning with Chapter 8, which introduces the General Linear Model, the level of the material roughly parallels that of Searle's earlier text. Chapter 9 considers the analysis of the two-way layout using various overparameterized models, while Chapter 10 examines the cell means model for general layouts as well as various special cases, including the possibility of empty cells and the nonexistence of selected interaction terms. General models containing covariables are explored in Chapter 11. Chapter 12, containing 27 pages, relates the material presented to the output from several computing packages, including BMDP, GENSTAT, SAS and SPSS-X. The final chapter is a 33-page survey of mixed models.

Most chapters conclude with a good selection of exercises. Many of these require the completion of algebraic details from the text, while others involve playing through the algebra with artificial integer data selected to minimize computational difficulties. A solutions manual for the exercises is not yet available.

At times, some unfamiliar but sensible matrix notation is employed. As an example, $\{_r u_i\}$ denotes a row vector having ith entry u_i, while $\{_d A_i\}$ represents a block diagonal matrix whose ith block is the matrix A_i.

For purposes of interval estimation and tests of hypotheses, only normally distributed errors are considered. Except for the brief final chapter, the presentation is limited to fixed effects, and Searle does not discuss Generalized Linear Models in the sense of McCullagh and Nelder (1983).

This text has a number of distinctive fea-

tures. Its devotion to the analysis of unbalanced data is appropriate in view of the prevalence of such data in statistical practice. The material is current with a number of newer and recent results, many of them Searle's own contributions. The book is very well written, though some readers already familiar with the material may complain that, especially in the earlier chapters, it is, at times, exactingly clear and painfully redundant. However, my experience from using the (1971) text is that students appreciate and learn from such presentations.

Searle is especially disturbed that statistical computing packages supply a great many statistical tests without precisely displaying the hypotheses being tested. Throughout this book, for all conceivable hypotheses of interest with the models studied, he associates test statistics with hypotheses. It is shown that many frequently available tests involve the cell sample sizes, and it is argued that such data-dependent hypotheses are usually inappropriate.

Section 5.6 discusses how to sensibly select interaction contrasts of likely interest when there are empty cells in two-way layouts. In Chapter 6, the presentation of models with covariables in the one-way case includes tests of the equality of the mean response over all groups, where the groups may have differing mean values of the covariable. The authoritative discussion in Chapter 10 of analyses with more than two factors where all factors are crossed, (as opposed to some nested), goes well beyond both Searle (1971) and Hocking (1985). For example, a theorem (p.407) indicates a necessary and sufficient condition for estimability of a general linear function of cell means in a model with linear restrictions (e.g., those deleting selected interactions) and some empty cells. This is a matter of some delicacy because the incidence of the empty cells affects which interaction contrasts can be estimated. Chapter 11 provides, in great detail, procedures for testing every imaginable hypothesis in models with covariables, and likewise, precisely what is being tested by every conceivable test statistic in commercial software appropriate for analyzing such models. (Hocking (1985) contains relatively little material dealing with covariables.) Chapter 13 is a good current summary of procedures for point estimation in mixed models without delving as deeply into this area as Hocking's book.

While the text material is interesting and appropriate, there are a number of other things that could have been added to this book that would have increased both its quality and length.

Most importantly, the presentation lacks real (or at least realistic) data sets and case studies–the data are all artificial. As a result, the reader cannot completely learn from this text how to proceed to select an appropriate model and how to undertake the fine detail work involved in exploring the sources of significant main effects and interactions. Test results are described as either significant or non-significant, with no reference to α level or P-value; the weight of evidence of test results is not considered. Nor are problems of simultaneous inference dealt with adequately. There is minimal discussion of analysis of simple effects and interpretation of interactions; it is difficult to communicate the essentials without real examples. There is an over-emphasis on hypothesis testing and point estimation at the expense of interval or regional estimation. Instructors who use this book as a course text should address these shortcomings.

There is almost no commentary on the desirability to check for the aptness of assumptions. Many users of these models (and software developers) do not appreciate that the residuals from least squares fits to the types of models considered in this book can and often should be subjected to the same kinds of residual analyses now customarily undertaken in multiple regression contexts: plotting and other techniques to check for outliers, normality, independence, homoscedasticity, etc. Corresponding remedies (transformations, alternative robust procedures, nonparametric or Bayesian approaches) are likewise not considered here.

As a result of the minimal attention given to mixed models, important topics usually considered under the heading "linear models", such as repeated measures and split plots, are not found in this presentation.

Though Searle's command of the algebra is impressive, algebraic details are overemphasized. Many of those contained in the first six chapters are uninteresting, and mostly unnecessary for those who will proceed on to the

general matrix formulations in the later chapters.

On the other hand, there is practically no mention of the underlying geometry of least squares and linear models, much of which requires minimal mathematical background and is helpful for a thorough appreciation of the statistical analyses.

A curious exception to these impressions is Searle's examination, for additive models, of whether the pattern of empty cells creates a disconnected layout. He presents the geometric approach of Weeks and Williams (1964), which is fine for small two-way layouts though unhelpful in larger situations, but forgoes the opportunity to present the more generally useful algebraic condition involving the rank of the coefficient matrices of reduced sets of normal equations. Hocking's (1985) approach to disconnectedness, which uses a certain canonical form of the model, is a better approach than Searle's for addressing this issue.

The notion of the basis of a vector space is rather elementary. Thus it is disappointing that Searle talks about there being at most rank (X) linearly independent estimable functions of the parameter vector without referring to the notion of a *basis set of estimable functions*. On page 354, he gives an example where knowledge of a basis set of interaction contrasts can bypass the need to invert a certain matrix. In general, if all else fails, the Gram-Schmidt orthogonalization can be used to construct such a basis set. Furthermore, Section 12.3, which discusses how SAS-GLM generates basis sets of estimable functions, could have been shortened with the use of basis terminology.

From my reading, I suspect that this book is as laden with typographical errors as were the early printings of Searle (1971); these will undoubtedly be corrected in future printings. Not all such errors are trivial–there are two different errors in what is arguably the most important formula in the book, (146) on p. 291, which gives the F statistic for a general linear hypothesis.

I recommend consideration of this text for a one or two semester course in linear models for factorial experiments. If you have successfully used Searle (1971) for this purpose, here is its replacement. Many errors and misconceptions in the earlier book have been cleaned up, and the new version is current, correct, and well-written. However, as noted above, the text material needs to be supplemented in certain areas, especially with examples that thoroughly discuss the analysis of data from some real experiments.

The book is also an important reference for anyone needing to analyze designed experiments with continuous responses and unequal numbers of observations in the sub-most cells where these models may additionally include covariables. It is the best available source for learning about the correspondence between the test statistics and hypotheses inherent in the generation of computer output and data analysis for such experimental situations.

Chapters 10 and 11 contain much new and recent material that suggests ideas for research in linear model theory. However, for students intending to do doctoral-level research in the areas of linear models dicussed in this book, assuming that only a single textbook is to be adopted, Hocking (1985) would be a better choice.

References

Hocking, R.R. (1985): The Analysis of Linear Models. Brooks/Cole Publishing Company, Monterey, CA.

McCullagh, P. and Nelder, J.A. (1983): Generalized Linear Models. Chapman and Hall, London.

Searle, S.R. (1971): Linear Models. John Wiley and Sons, New York.

Weeks, D.L. and Williams, D.R. (1964): A Note on the Determination of Connectedness in an N-Way Cross Classification. Technometrics, 6, pp. 319–324.

Burt S. Holland
Temple University
Philadelphia, PA
U.S.A.

Printed September 1988.