

Applied Psychological Measurement

<http://apm.sagepub.com>

Book Review : Discrete Multivariate Analysis: Theory and Practice

Yvonne M. M. Bishop, Stephen E. Fienberg, Paul W. Holland, Richard J. Light and Frederick Mosteller

Applied Psychological Measurement 1977; 1; 297

DOI: 10.1177/014662167700100218

The online version of this article can be found at:

<http://apm.sagepub.com>

Published by:

 SAGE Publications

<http://www.sagepublications.com>

Additional services and information for *Applied Psychological Measurement* can be found at:

Email Alerts: <http://apm.sagepub.com/cgi/alerts>

Subscriptions: <http://apm.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

BOOK REVIEW*

Peter B. Imrey, Reviewer
University of Illinois

Discrete Multivariate Analysis: Theory and Practice

Yvonne M. M. Bishop, Stephen E. Fienberg, and Paul W. Holland, with
the collaboration of Richard J. Light and Frederick Mosteller.
Cambridge, Massachusetts: The MIT Press, 1975. Pp. 557

Methodology—the body of techniques for measurement, arrangement, reexpression and analysis of information—conducts the tension between theoretical constructs and descriptive data which drives modern science. Methodology determines the direction in which the examination of scientific theory is pursued and the theoretical construct which outruns available methodology for validation soon becomes stale or unproductive. The expressive range of methodological language also shapes the generation of theory and, in much the same manner that practical media and formal structural constraints influence art and literature, the limitations of theoretical and methodological constructs may prove stimulating or stifling to a science at a particular stage of maturity. In any science, a period of primarily methodological rather than substantive development may sometimes be necessary to unblock the logjam created by theories and measurements which cannot effectively interact through existing tools.

Discrete Multivariate Analysis arrives at a time when the various psychosocial disciplines are all suffering, to varying degrees, from attempts to swallow whole those chunks of statistical methodology for continuous data that have been most successful in the natural science disciplines. The movement to quantification of psychological and social research has been motivated, in large measure, by a desire to legitimize behavioral science through application of the “hard science” criteria of objectivity and reproducibility to statements of and data analyses relating to behavioral paradigms. Passionate advocacy of multiple regression and other multivariate analytic tools has been matched by claims that such tools have forced their proponents, through compromises necessary in measurement, data preparation and formal hypothesis construction, to distort and ultimately trivialize their science in order to accommodate the prerequisites of statistical analysis. The intensity of this debate between “traditionalists” and “methodologists” has shown no sign of abating in the last 10 years; indeed, parallel discussions in the area of medical clinical research display the same basic concern as that which troubles academic social scientists. The question underlying debate in both areas is whether the refinement and continuous scaling of information derived from primitive conceptualizations or measurement strategies is likely to prove stimulating or stifling to further scientific growth.

Much of the urgency and stridency of discourse on this issue is certainly due to the perceived disarray of statistical methodology appropriate to categorical information, consisting of observations which fall into nominal, ordinal or scaled classes. The analysis of such discrete data has long been limited in scope and convenience by the basic dependence of available methods on the dimensionality

*Book Review Editor: Howard Wainer, Bureau of Social Science Research

APPLIED PSYCHOLOGICAL MEASUREMENT

Vol. 1, No. 2 Spring 1977 pp. 297-306

© Copyright 1977 West Publishing Co.

297

underlying the definition of the category scheme and on the sampling scheme postulated to generate the observed category counts under analysis. Many procedures have been developed which derive their optimality, or computational tractability, from combinatorial properties peculiar to multidimensional counting arrays of particular sizes, e.g., 2^k or $2 \times k$ contingency tables. These procedures have frequently differed in basic orientation from one another and, as a group, from methodologies which have been quite successful in allowing scientists to deal with continuously scaled measurements. These divergences would seem to be unnecessary, as the basic scientific questions addressed to continuous and categorical data arrays of whatever structure and size are frequently identical.

The need for a unifying analytic approach to categorical data analysis has not gone unnoticed by statisticians, and the last 15 years have seen the development of comprehensive methods in this area emerge as a prominent and increasingly successful field of statistical research. *Discrete Multivariate Analysis* is a large first step in making the fruits of this effort widely available. Quoting the authors:

“In discrete multivariate statistics, however, the excellent guides already available . . . stop short of giving a systematic treatment of large contingency tables, and especially tables that have troublesome irregularities. This book offers such a treatment. . . . The applied statistician or quantitative research worker looking for comprehensive analyses of discrete multivariate data will find here a variety of ways to attack both standard and nonstandard sets of data. As a result, he has available a systematic approach to the analysis of multiway contingency tables. Naturally, new difficulties or constraints raise new problems, but the availability of a flexible approach should strengthen the practitioner’s hand, just as the ready availability of analysis of variance and regression methods has for other data.”

The care devoted throughout the book to examination of a plethora of social, biomedical and psychological data sets emphasizes a broader goal: to provide a framework within which data from these areas can be analyzed on their own terms, without the necessity of tradeoffs between statistical rigor and scientific acuity.

An annotated contents will help convey the book’s scope.

- Chapter 1. *Introduction*—Justification and outline of the presentation.
- Chapter 2. *Structural Models for Counted Data*—Multiplicative models for counts of observations, describing the logarithms of counts as linear functions of parameters chosen according to values of the variables describing the counted events. For multidimensional contingency tables, these models are formally identical to classical partitions of cell means we are all familiar with for analyzing factorial experiments resulting in continuous observations.
- Chapter 3. *Maximum Likelihood Estimates for Complete Tables*—How to estimate the probabilities associated with multidimensional cross-classifications when a log-linear model applies. The optimization used is maximum likelihood, and a class of models (hierarchical) is defined for which the equations resulting under Poisson, multinomial or product multinomial sampling may be solved directly, or by a convenient iterative process.
- Chapter 4. *Formal Goodness of Fit: Summary Statistics and Model Selection*—Global χ^2 fit statistics and local residual analyses for evaluating the utility of a model, or comparing several with a view to selection of the “best.” Partitioning of likelihood and χ^2 to allow evaluation of model features.
- Chapter 5. *Maximum Likelihood Estimates for Incomplete Tables*—When certain cells of a cross-classification cannot occur as a direct consequence of the definitions of the cross-classified variables, the resulting “incomplete tables” are conceptually distinct from possibly identical complete tables in which these cell counts are zero due to the outcome of a probabilistic mechanism. Incomplete tables are no longer analogous to

- factorial design structures, but useful models based on the marginal class schemes may still be applied to those cell counts of stochastic origin. Necessary adjustments in the estimation and evaluation procedures of Chapters 3 and 4 are provided.
- Chapter 6. *Estimating the Size of a Closed Population*—In tag-recapture studies in ecology or multiple-record system studies in demography, the extent of overlap among multiple attempts at observation is used to estimate the number of individuals in a population that have totally eluded detection. This extrapolation is based, however, on the rather tenuous assumption that observation or capture at one time or by one recording system is statistically independent of observation or capture at other times or by other recording systems. When three or more observation times or record systems are available, log-linear models are shown to provide a choice of less rigorous assumptions to replace the full independence model and allow the ultimate estimation of population size from a more realistic framework.
- Chapter 7. *Models for Measuring Change*—Analysis of “repeated measurement” tables, where the dimensions apply the same classification scheme at different times by Markov chain models. Marginal classes are states of the chain, and transition probabilities are reflected in conditional probabilities within slices of marginal tables. Interesting properties of chains are related to log-linear models in the resulting contingency table.
- Chapter 8. *Analyses of Square Tables: Symmetry and Marginal Homogeneity*—For square tables such as those in Chapter 7, interest may center on essentially geometrical considerations, such as whether certain aspects of the table are basically unaltered if rows and columns are interchanged. Such symmetry hypotheses are explored and related, and methods of estimation and testing based on log-linear models are explored.
- Chapter 9. *Model Selection and Assessing Closeness of Fit: Practical Aspects*—Discussion of the considerations in data analysis which direct and modulate iterative application of the formal goodness of fit procedures given earlier (Chapter 4).
- Chapter 10. *Other Methods for Estimation and Testing in Cross-Classifications*—Summary review and commentary on information theoretic, minimum χ^2 , weighted least-squares, logistic, partition of χ^2 , and exact theory approaches to multiple cross-classifications.
- Chapter 11. *Measures of Association and Agreement*—A description and comparative review of the many such measures available for two-dimensional tables. Emphasis on tailoring the measure to the use.
- Chapter 12. *Pseudo-Bayes Estimates for Cell Probabilities*—Estimators of cell probabilities which “shrink” the observed data towards some stable values or hypothesis (such as independence) when the observed data are sparse relative to the number of cells in a one- or two-dimensional table.
- Chapter 13. *Sampling Models for Discrete Data*—What everyone should know about the Poisson, multinomial, hypergeometric and related discrete distributions.
- Chapter 14. *Asymptotic Methods*—A short course (75 pages) in asymptotic concepts and relevant Central Limit Theory.

Because of the general utility of the techniques presented in Chapters 2 through 5, the clarity of expository writing, the brilliant integration into the book of numerous examples of data analysis, and the overall saneness and soundness of the statistical thinking exhibited in the approaches to these examples, this reviewer feels that *Discrete Multivariate Analysis* can be a strong stimulus to a productive reorientation of statistical methodology, particularly in the social and clinical medical sciences. The book is admirably suited for use as a text for both applied statisticians and statistically oriented researchers, and the technical mathematical background necessary is minimal (neither calculus nor

linear algebra is essential). This praise is not meant to imply that the problems of analyzing categorical data are solved here, however, or that *Discrete Multivariate Analysis* is fully comprehensive in terms of what may be done within the basic approach that is chosen. Before remarking on specific features, I will try to set the core (Chapters 2 through 5) of this extensive work in an even broader framework.

If π is a vector of probabilities associated with categories or cells of a contingency table or other categorical data structure, the general log-linear model may be compactly written in matrix terms as

$$\pi = \frac{\exp X\beta}{1' \exp X\beta} \quad [1]$$

where 1 is a vector of units, β is a vector of parameters analogous to those of multiple linear regression, X is a "design matrix," and \exp is an operator which produces the exponential function of every element of the associated argument vector. Equation 1 differs from the usual formulation

$$\mu = X\beta \quad [2]$$

of general linear regression in the use of the exponential operator, corresponding to the choice of a multiplicative rather than additive model, and in the denominator term, simply a normalizing constant requiring the fitted probabilities to sum to unity. A further technical difference is that X in Equation 2 is usually taken to be a full rank, possibly containing a column 1 defining a general mean or an intercept term, while X in Equation 1 must not only be a full rank but is also subject to the identifiability condition that its columns are jointly linearly independent of 1 . When X in Equation 2 is further restricted to contain only zeros and units, we obtain a class of models known as "analysis of variance" designs; when X in Equation 1 is similarly restricted, we obtain analogs for categorical data structures.

When μ in Equation 2 corresponds to the vector of cell means in a factorial cross-classification, the usual statistical practice is to choose X so that elements of μ are described as sums of parameters representing (1) a general mean, (2) main effects corresponding to average effects of single factors, (3) first-order interactions, representing deviations from the predicted values using the main effects, due to simultaneous occurrence of pairs of factors, (4) second-order interactions, representing average deviations from the predicted values using first-order interactions due to combinations of three factors, and so on. In typical data analysis, we sequentially reduce such models by eliminating the parameters reflecting higher-order interactions first; thus, if we allow for the possibility that factor A may have different effects according to the levels of factor B (AB interaction), we always also allow for the possibility that these effects may not cancel over levels of B (A main effect). This restricts us to a special class of models called "hierarchical," because they can only be reached by peeling off outside layers of the model. For unsaturated hierarchical models, the design matrix X has a particular structure: its set of columns X_c can be written as the set theoretic union $\bigcup_{i=1}^k X_i$, for some k , where X_i is the column set of a design matrix X_i designating a hierarchical model for a proper subset of the original factors.

Whenever we are faced with the possibility of interaction between factors in continuous data or marginal classifications in categorical data, multiple parameters may be involved if any factor or classification takes more than two levels. We may deal with the multiple parameters representing a particular combination of variables as a class, including in our model all or none, or on an individual basis by including only those which are scientifically essential, statistically significant, or both. By "classwise inclusion" we refer to the approach of including or deleting multiple interaction parameters on the same variable subset on a groupwise basis. For hierarchical classwise inclusion models, the X , above represents full factorial designs on the corresponding variable sets.

In terms of the above discussion, the models studied by Bishop *et al.* may be characterized as hierarchical log-linear, analysis of variance models with classwise inclusion of parameters. There are a

number of advantages to restricting to such models. [The full textual development given here may be regarded as the culmination of a line of research dating back to Bartlett (1935) and Deming and Stephan (1940) that was developed in the '60s and early '70s mainly by Birch, the authors under the stimulus of Mosteller, and by Bock, Goodman, and his student, Haberman. Due to their elegant work, the mathematical theory of a likelihood-based approach to these models is essentially complete under the assumption of Poisson, multinomial or product multinomial sampling models.] We know that maximum-likelihood estimates of underlying parameters may always be obtained by application of the simple iterative technique of proportional fitting, also called raking, due to Deming and Stephan (1940). We also know when the iterative procedure may be bypassed in favor of direct form expressions for the estimates. The relationship of model features to concepts of complete and conditional independence of variables is clear so that fitted models are often readily interpretable in terms of collapsed, partitioned or marginal tables; hence, they make sense in terms of how we all look at tables in exploratory examinations outside the context of a formal model. Finally, because they are direct translations of popular models for continuous data, they are generally familiar to an audience interested in statistical applications rather than theory.

Limitations, however, are equally clear, and I dwell on these a bit. The initial decision to model cell probabilities is sensible, but quite frequently we are interested in individual cells only as they relate to combinations or functions of their probabilities which reflect particular features of the data structure. In such cases, a model for individual cells is optimal only if the model form and parameters remain interpretable and relevant after the translation process which expresses the ultimate functions of cell probabilities in terms of the model parameters. Under this criterion, the choice of multiplicative models for cell probabilities is highly successful when we are studying independence or related hypotheses which are essentially multiplicative in nature, logistic functions which are intimately related, or separate marginals or other functions preserved by maximum likelihood model-fitting. But the multiplicative model for individual cells can easily get in the way when other types of functions are involved. Thus, in Chapter 8, unconditional inference about marginal homogeneity is unavailable within the log-linear framework, because differences between marginals do not readily translate to multiplicative factors for individual probabilities. Similarly, general measures of association or agreement in Chapter 12 are not dealt with in the log-linear model framework because these functions do not linearize in terms of parameters of a multiplicative model for cells. The analysis of conditional mean scores for categories is also excluded. By restricting to multiplicative models for cells one obtains a comprehensive, unified approach to detecting dependence among variables, but one that is less successful in dealing with questions targeted at specific nonmultiplicative functions of the data array.

The restriction to analysis of variance models within a log-linear framework is somewhat less limiting but also involves sacrifices. For ordinal or scaled categories, it is frequently appropriate to study trends; one way of doing this is by a reexpression of model parameters in terms of linear and curvilinear components relevant to the scaling. Similarly, it may be useful to adjust the model using the mean value of a continuous covariate as it differs across cells. Neither of these problems may be adequately handled within the analysis of variance structure. [Williams and Grizzle (1972) and Haberman (1974a) develop models for trends; the authors would undoubtedly deal with the second problem by categorization of the covariate followed by inclusion into the table.] One advantage of working with only analysis of variance models is that the matrix notation associated with explicit presentation of X becomes avoidable. On the other hand, the use of a little linear algebra here might allow the expression of concepts such as separability and connectivity in terms of rank conditions on matrices and allow uniform computation of degrees of freedom as ranks, eliminating the distraction of messy accounting-type procedures as in Chapters 3 and 5.

The use of hierarchical models takes advantage of the appealing properties that the estimated parameters and fitted probabilities are completely determined by marginals obtained by collapsing

the table in obvious ways, that these marginals are exactly reproduced by the fitted probabilities, and that these are always obtainable through the iterative proportional fitting algorithm. Computational techniques for nonhierarchical models are available, but they involve sophisticated numerical optimization procedures such as cyclic descent or a modified Newton-Raphson technique, and the extent to which these can be programmed for reasonably general classes of nonhierarchical models is yet unclear. That nonhierarchical models are often relevant is clear, though, as they may quite generally be obtained from hierarchical structures by the imposition of simple linear constraints on parameters, e.g., equality of two interaction terms. A nonhierarchical model may frequently be useful as a parsimonious refinement of a hierarchical model; such reduced models particularly suggest themselves for use within multi-stratum multiple-record systems, as discussed in Chapter 6.

Classwise inclusion of parameters was obviously chosen for the notational simplification it allows and is clearly not a restriction necessary to the approach. However, in a $3 \times 3 \times 4$ contingency table, 12 degrees of freedom are associated with second-order interaction. If only one degree of freedom is really operating, an omnibus likelihood ratio test on the 12 is almost certain to miss this submerged parameter. If it does not, however, we are led by casewise inclusion to keep eleven parameters which might otherwise be used to smooth the table. The problem gets worse in tables with more dimensions and categories. In light of these considerations, it seems a shame that the authors did not include even a small section offering guidance in selective reduction within a class of parameters.

Chapter 10 of *Discrete Multivariate Analysis* outlines alternative approaches to categorical data analysis. Some brief supplementary comments are in order. Two general classes of procedures exist. The first, into which the likelihood-based methods here fall, obtains asymptotically efficient (best asymptotically normal) estimates and asymptotically most powerful test statistics based on optimization of some criterion function (likelihood, discrimination information, χ^2 , weighted residual sum of squares). What is known about the small sample behavior of these methods is basically reassuring, but the fact remains that little is known under general conditions. The second class of procedures, of which Fisher's exact test is the simplest example, is based on exact distribution theory and is thus entirely understood for small samples. Unfortunately, such methods are only developed for fairly simple specific situations while theoretical and, particularly, computational problems associated with generalization appear formidable.

Among the asymptotic methods, the various criterion functions fall within a class considered by Haldane (1952). No general preference ordering appears possible. The likelihood function is superior in terms of "second-order efficiency" criteria which have little apparent practical relevance (Rao, 1963; Hoeffding, 1965). Small-sample studies have been neutral (Berkson, 1968) or favorable (Odoroff, 1970) to weighted least-squares. Within the framework of hierarchical log-linear models and of some models associated with combining contingency tables, maximum likelihood is theoretically superior by virtue of satisfying invariance criteria related to operations such as collapsing a table. The discrimination information approach has been used to handle the marginal homogeneity problem successfully (Ireland, Ku & Kullback, 1969). The weighted least-squares approach (Grizzle, Starmer & Koch, 1969) is by far the most general, as it is adaptable to the modeling of general smooth functions of cell probabilities using a general design matrix X . The limitations of hierarchical analysis of variance models for probabilities are avoidable with this approach. The authors' remark, "Even though there are instances where the weighted least-squares approach yields estimates without iteration while maximum likelihood estimates require iteration . . . the reverse situation is usually the case." This statement is erroneous, as the weighted least-squares method uses the unrestricted maximum likelihood estimate of the true weight matrix to obtain direct estimates of parameters in all circumstances. In certain linear problems, however, iteration through the weighted least-squares equations by successive updating of the weight matrix is an efficient numerical scheme for obtaining the maximum likelihood solution to the model.

With respect to the specific presentation of material in *Discrete Multivariate Analysis*, I think most readers would give the authors high marks. What follows are personal impressions, reflections, and comments meant to help in the preparation of a second edition. The explication of the basic maximum likelihood theory in Chapters 2 through 5 is excellent and remarkably successful in conveying a basically physical understanding of how features of a model combine, through the reproduction of various margins, into the fitted table. The approaches to collapsibility and to building higher-order models from lower-order models for pieces of the table are particular examples of how models are clearly related to operations associated with the physical tabular structure. Chapter 3 is marred by an erroneous proof of Theorem 3.4.2 on direct estimation; it is unclear whether a correct proof exists following the line of attack given. The discussion of standardized rates in Chapter 4, with reference to the National Halothane Study, is especially enlightening. Oddly, in this chapter the authors mount a vigorous attack on use of the Mantel-Haenszel test (Mantel & Haenszel, 1959) for detecting dependence or inhomogeneity in a cross-classification of dichotomous variables. This procedure, which has appeared heretofore mostly in the biometric literature, computes an overall test statistic from samples taken in a number of subclasses, or representing different matched sets of observations. The authors argue that the statistic should only be used when association (log odds ratio) is constant across subclasses. In so doing, they ignore the fact that the main use of Mantel-Haenszel is as a general test for partial association, not a specific test for a particular interaction parameter of a log-linear model. While the test may be weak against certain alternative patterns of partial association, it remains an excellent test for situations in which, despite differences in magnitude, association is unidirectional across subclasses. The Mantel-Haenszel and related tests ought to see wider usage, and I hope *Discrete Multivariate Analysis* will not discourage that. I found the discussion of incomplete tables in Chapter 5 fascinating, and the approach to modeling in terms of geometrical rearrangements of tables into canonical structures very attractive. Nevertheless, much is to be gained by viewing these problems from an algebraic standpoint. In particular, the identical treatment of structural and sampling zeros defining separability suggests that a more comprehensive way to look at things might be to consider the relationship between the model X and the observed covariance matrix V of the cell counts. Specific tabular structures might be seen as reflections of matrix canonical forms and linear algebra as a more spacious theater for the generation of results for incomplete arrays. The idea of treating one or more cell counts in a complete table as "outliers" and fitting models to the remaining incomplete array is the particularly significant contribution of this chapter.

The discussion of population size estimation elegantly presents a situation which is tailor-made for the application of log-linear models. I found the ensuing treatment of Markov chains much less attractive. The development is based on an argument that the likelihood equations for the transition matrix of an I state Markov chain, based on n observations at $(T+1)$ times and displayed in an I^{T+1} table, are identical by the Markov property to those based on the $I \times T$ matrix of nT transitions, in which each element of the I^{T+1} table appears T times. If the chain is stationary, similar reduction leads to consideration of the $I \times I$ table obtained by collapsing over time. While the argument is certainly sound, the explanation given is confusing and leaves the impression that we can treat one observation as if it were many by changing the dimensionality of the table. This chapter is frequently ingenious but is distinguished by a certain artificiality not present elsewhere in the book. The examples here left me feeling that the substantive utility of stationary Markov chains in social research is as yet undemonstrated, though such models can easily be used as simplistic, descriptive tools.

Chapter 8 is a particularly useful and skillful examination of various hypotheses of symmetry associated with contingency tables. This piece contains good illustration of the technique (introduced in Chapter 2) of obtaining a simple model through rearrangement of a table; the process of reparameterizing a model design X into simple form is dealt with by redefining dimensions and labels designating the cell counts and applying the same models to different rearrangements generated in this

manner. The results of analyzing a particular table may then be expressed as the rearrangement-model sequence generating the best combination of parsimony and interpretability.

The remaining five chapters might have instead formed an anthology, *Topics in Categorical Data*. Each stands apart from the rest of *Discrete Multivariate Analysis*, and each makes a worthwhile contribution. In Chapter 9 the authors consider general issues of model fitting as reflected in contingency table analysis, and the resulting discussion is a delight. This is the sort of material rarely seen in "text" format, because it deals with the philosophical and "clinical" aspects of data analysis, the extra-mathematical considerations which informally guide the application of highly structured techniques such as those of Chapter 4. Among section titles in this chapter are: Simplicity in Model Building; Searching for Sampling Models; Too Good a Fit; Large Sample Sizes and χ^2 When the Null Model is False; and Data Anomalies and Suppressing Parameters. Chapter 11 is a compendium of association and agreement measures for two-way tables, which sheds light on the relationships of the various competing statistics and the usages and interpretations of each. Chapter 13 is a compact presentation of the properties of basic distributional models for discrete data. Both these chapters are clear as a bell, and serve to provide perspective and handy reference material.

In applying modeling techniques to an array of counts, efficient estimates of cell probabilities are obtained as smooth probabilities calculated under the assumptions of a true model which empirically fits the data. The basic technique espoused in *Discrete Multivariate Analysis* involves moving from the observed cell proportions to more stable estimates based on sufficient marginals under a hierarchical log-linear model. In "Pseudo-Bayes Estimates of Cell Probabilities" a different approach to reduced mean-square error estimation is taken. Observed cell proportions from the data array are combined with, and thereby "shrunk" towards, a set of probabilities which may be either prespecified or estimated from the data under a provisional model. This may be viewed as a compromise between the choice of observed proportions versus modeled ones. The resulting estimators do not fit any particular model, but if the provisional model does not grossly misrepresent the true situation, increased stability may be obtained which more than compensates in the mean-square error sense for some additional bias acquired in the shrinkage process. The approach is recommended particularly for "sparse" tables, in which sampling zeros are common due to small expected values in many cells, and the authors suggest that improved estimation properties may extend to parameters of log-linear or other models, if these are fitted to the pseudo-Bayes estimated proportions rather than those initially observed. Closely related to ridge regression and similar approaches to modeling continuous data, the presentation of this relatively new area of research is parenthetical to Chapters 1 through 11; the explanation of this essentially theoretical material is interesting and stimulating.

Bishop *et al.* follow the ever more common practice of providing basic theoretical background for a methodological text as a final chapter or appendix. Thus, they close with a monograph length treatment of asymptotic theory underlying earlier developments. The discussion of stochastic order is just a bit confusing, but otherwise this "add-on" stands as a highlight of the book and should be useful both as a reference and as teaching material to workers developing or using any of the asymptotic approaches to categorical data. An orderly compilation of this material in one place has long been needed; this reader greatly appreciates the decision to devote so much care and space to it in this book. The content of this last section, unfortunately, puts in perspective a deficiency which pervades the entire treatment of log-linear models: the lack of a coherent practical method of variance-covariance estimation for parameters and fitted probabilities from unsaturated models. In Chapter 6, the orthogonal projection method of Haberman (1974b) is discussed but was apparently not considered sufficiently developed for general practical application. In Chapter 4, it is remarked that variance formulae from models which allow direct estimation of parameters "must be derived *ab initio* for each individual model" and that for models requiring iterative solutions for parameters, "no closed-form variance estimates exist." It is a measure of the pace of research in this field that closed-form consistent estimators of the asymptotic covariance matrices corresponding to the vectors of estimated

parameters and cell-probabilities may now be obtained routinely (Koch, Freeman & Tolley, 1975, building on Causey, 1972), and these considerably enrich our ability to deal efficiently with reduced log-linear models on a parameter-by-parameter basis.

Finally, it is a pleasure to remark that exercises, while not plentiful, are well thought-out and definitely aid in understanding the material. In addition to a standard subject index, an author index and especially an index of data sets used as examples are very helpful to the reader. The list of references is extensive, though readers attempting a literature search will find Killion and Zahn (1974) more useful. A few minor misprints caught the eye:

p. 71: Two references to (3.3.14) should read (3.3.15).

p. 72: Replace J with I in first term of $u_{24(JI)}$.

p. 220: Residuals in Table 5.4.1 were omitted.

p. 249: Missing caret in (6.5.13).

p. 272: Table 7.4.2 does not correspond to textual interpretation.

While much of this review has been devoted to pointing out limitations of *Discrete Multivariate Analysis*, it is the work presented here and related research by V. P. Bhapker, James E. Grizzle, Gary G. Koch, Solomon Kullback and various coauthors that have so widened our perspective and expectations of categorical data analysis. Though more flexible methods are emerging, *Discrete Multivariate Analysis* deserves a long life as a cornerstone guide to an orderly and sensible approach for working with categorized data.

References

- Bartlett, M. S. Contingency table interactions. *Journal of the Royal Statistical Society Supplement*, 1935, 2, 248–252.
- Berkson, J. Application of minimum logit χ^2 estimate to a problem of Grizzle with a notation on the problem of “no interaction”. *Biometrics*, 1968, 24, 75–95.
- Causey, B. D. Sensitivity of ranked contingency table totals to changes in problem conditions. *Annals of Mathematical Statistics*, 1972, 43, 656–658.
- Deming, W. E., & Stephan, F. F. On a least-squares adjustment of a sample frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics*, 1940, 11, 427–444.
- Grizzle, J. E., Starmer, C. F., & Koch, G. G. Analysis of categorical data by linear models. *Biometrics*, 1969, 25, 489–504.
- Haberman, S. J. Log-linear models for frequency tables with ordered classifications. *Biometrics*, 1974, 30, 589–600. (a)
- Haberman, S. J. *The analysis of frequency data*. Chicago, IL: The University of Chicago Press, 1974. (b)
- Haldane, J. B. S. A class of efficient estimates of a parameter. *Bulletin of the International Statistical Institute*, 1953, 33, 231–248.
- Hoeffding, W. Asymptotically optimal tests for multinomial distributions. *Annals of Mathematical Statistics*, 1965, 36, 369–401.
- Ireland, C. T., Ku, H. H., & Kullback, S. Symmetry and marginal homogeneity of an $r \times r$ contingency table. *Journal of the American Statistical Association*, 1969, 64, 1323–1341.
- Killion, R. A., & Zahn, D. A. *A bibliography of contingency table literature: 1900–73*. Florida State University Statistics Report M300, 1974.
- Koch, G. G., Freeman, D. H., & Tolley, H. D. *The asymptotic covariance structure of the estimated parameters from contingency table log-linear models*. University of North Carolina Institute of Statistics, Mimeo Series No. 1046, 1975.

- Mantel, N., & Haenszel, W. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 1970, 22, 719-748.
- Odoroff, C. L. Minimum logit chi-square estimation and maximum likelihood estimation in 2x2x2 and 3x2x2 contingency tables—tests for interaction. *Journal of the American Statistical Association*, 1959, 65, 1617-1631.
- Rao, C. R. Criteria of estimation in large samples. C. R. Rao (Ed.), *Contributions to Statistics*. New York: Pergamon Press, 1963.
- Williams, O. D. and Grizzle, J. E. Analysis of categorical data having ordered response categories. *Journal of the American Statistical Association*, 1972, 67, 55-63.