Wells generalizes the notion of an infinite polyhedron to a 'net' in which the only polygons that occur are *skew* polygons. The most famous case (see p. 117) is the 'diamond net' [Hilbert and Cohn-Vossen, 1952, p. 49]; its vertices have integral coordinates (x, y, z) where $x \equiv y \equiv z \pmod 2$ and $x + y + z \equiv 0$ or $1 \pmod 4$. In other words, Wells is looking for *graphs*, of given valency and girth, which can be realized in Euclidean space so as to be symmetrical by translations in three independent directions. The number of possibilities is so great that complete enumeration seems to be out of the question. The search is justified by his observation that such graphs provide structural formulae for more than fifty crystals (which he lists on p. 265). He illustrates many of them by pairs of stereoscopic photographs, and some by very accurate drawings; see especially his pp. 147, 170, 254 and 255.

BIBLIOGRAPHY

A. Andreini 1905, *Sulle reti di poliedri regolari e semi-regolari*, Mem. Soc. Ital. Sec. (3) **14**, 75–129.

W. W. R. Ball and H. S. M. Coxeter 1974, *Mathematical recreations and essays*, 12th ed., Univ. of Toronto Press, Toronto, Ont.

H. S. M. Coxeter 1939, *The regular sponges, or skew polyhedra*, Scripta Math. **6**, 240–244.

______ 1940, *Regular and semiregular polytopes*. I, Math. Z. **46**, 380–407.

______ 1968, *Twelve geometric essays*, Southern Illinois Univ. Press, Carbondale, Ill.

______ 1973, *Regular polytopes*, 3rd ed., Dover, New York.

______ 1974, *Regular complex polytopes*, Cambridge Univ. Press, New York.

______ 1978, *Polytopes in the Netherlands*, Nieuw Arch. Wisk. (3) **26** (to appear).

Keith Critchlow 1969, *Order in space*, Thames and Hudson, London; Viking Press, New York.

J. R. Gott, III, 1967, *Pseudopolyhedrons*, Amer. Math. Monthly **74**, 497–504.

D. Hilbert and S. Cohn-Vossen 1952, *Geometry and the imagination*, Chelsea, New York.

A. V. Shubnikov and V. A. Koptsik 1974, *Symmetry in science and art*, Plenum, New York.

A. Wachman, M. Burt and M. Kleinmann 1974, *Infinite polyhedra*, Technion, Haifa.

Ernst Witt 1941, *Spiegelungsgruppen und Aufzählung halbeinfacher Liescher Ringe*, Abh. Math. Sem. Univ. Hamburg **14**, 289–322.

H. S. M. COXETER

BULLETIN OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 84, Number 3, March 1978

Discrete multivariate analysis: *Theory and practice*, by Yvonne M. M. Bishop, Stephen E. Fienberg, and Paul W. Holland with the collaboration of Richard J. Light and Frederick Mosteller, MIT Press, Cambridge, Mass., 1975. Second printing with corrections, 1976, x + 557 pp., $30.00.

Categorical data arise whenever counts, rather than continuous measurements, are made. Such data are especially important in the social sciences, in which qualitative responses to surveys are frequently a source of information, and in medicine, in which classification of patients, treatments, and/or symptoms and judgements with respect to outcomes are the variables of interest. Analysis of categorical data has a long history, beginning at least with Karl Pearson's famous paper (1900). Contributions of major significance were made by R. A. Fisher (e.g., 1934, 1936, 1941), Bartlett (1935), and Birch (1963). From its inception, the analysis of categorical data has, more than many areas of statistics, emphasized multivariate aspects, although many notable developments such as probit and logit analysis deal with univariate categorical data, usually the dichotomous or binary case (see, for example, Cox, 1970).

Many books and monographs on multivariate analysis have been published over the last two decades since the earliest appeared. In most, the distinction between the analysis of continuous data and categorical data, which are discrete, has not been emphasized. Nearly all analytical treatments have focused on the continuous case, explicitly so, for example, in the case of Dempster (1969).

Discrete multivariate analysis offers a systematic treatment of multivariate categorical data primarily, indeed almost entirely, within the framework of the so-called log-linear probability model, a model on which we comment extensively below. The book is more comprehensive than its competitors, which include Maxwell (1961), Good (1965), Lancaster (1969), Cox (1970), Fleiss (1973), and Plackett (1974), although Plackett at L 3.50 and 159 pages is hard to beat for a cheap, brief, and thorough introduction to the subject. What most distinguishes *Discrete multivariate analysis* is the constant interplay between practical examples and the theoretical structure of the log-linear probability model, which unifies the discussion, so much so that, since the log-linear model is essentially a discrete analogue of the analysis of variance for continuous data, one reviewer was moved to remark that the "... book perhaps would be more aptly titled '*Discrete multivariate analysis of variance*' " (Olkin, 1977).

Consider the familiar 2 × 2 contingency table describing the joint probabilities of occurrence or nonoccurrence of two random categorical events, A_1 and A_2:

	A_2	
A_1	p_{11}	p_{12}
	p_{21}	p_{22}

As is well known, the condition for independence of A_1 and A_2 may be expressed in convenient parametric form as

$$p_{11}p_{22} = p_{12}p_{21}, \tag{1}$$

in this simple case; however, the situation is considerably more complex in contingency tables of higher dimension. It was the problem of formulating the condition for independence in a three-way table which led Bartlett (1935) to a first, but incomplete, development of the log-linear probability model, later fully developed by Birch (1963) in detail for the three-way table. Development for the general case appears in papers by Mosteller (1968), Bishop (1969), and many recent papers by Goodman (1968, 1969, 1970, 1971a, 1971b, 1972a, 1972b, 1972c).

Let $\mathcal{Q} = \{A_1, \ldots, A_q\}$ be a set of categorical random variables, which may take on, respectively, $I_1, \ldots, I_q$ possible values. If we have a sample of

N observations on the q categorical random variables, $\mathscr{Q}$, we might arrange these in an $I_1 \times I_2 \times \cdots \times I_q$ table of counts corresponding to a similar arrangement of the probabilities

(2) $p_{i_1,\ldots,i_q}$, $\quad i_1 = 1, \ldots, I_1, \quad i_2 = 1, \ldots, I_2, \quad \cdots \quad i_q = 1, \ldots, I_q.$

Alternatively, order the logarithms of the

$$Q = \prod_{k=1}^{q} I_k$$

probabilities (2) into a $Q \times 1$ vector by some principle, e.g., lexicographically,

$$\text{(3)} \qquad \log p = \begin{bmatrix} \log p_{1\ldots 1} \\ \vdots \\ \log p_{I_1\cdots I_q} \end{bmatrix}.$$

The vectors $\log p$ may be thought of as points in R^Q. Let M be a linear manifold in R^Q of dimension m, $0 < m \leqslant Q$. The class of models for which the $Q \times 1$ vector U_0 consisting entirely of ones is in M and

$$\text{(4)} \qquad \log p \in M \quad \text{such that} < p, U_0 > = 1,$$

where p is the vector of probabilities corresponding to $\log p$ and $\langle x, y \rangle$ denotes the inner product of x and y, is defined as the class of log-linear probability models (Haberman, 1974, p. 3). Since M is a linear manifold in R^Q, there exist m independent vectors, not necessarily orthogonal, which span M, one of which may be U_0 defined above. Because $\log p$ is contained in M it may be represented in terms of the basis vectors $U_1, \ldots, U_{m-1}$ and $U_m = U_0$. When $m = Q$, so that $M = R^Q$, the model is called *saturated* by Goodman (1968, 1970). Essentially, the saturated model is simply a reparameterization of the joint probabilities on the assumption that they are all strictly positive. While such a reparameterization may be of some intrinsic interest (since, for example, the one discussed below isolates different types of interaction effects), the model becomes considerably more interesting for $m < Q$, since only then are restrictions placed on the Q probabilities other than that they sum to one.

There are clearly many possible choices of a basis for M. The most interesting and useful of these, and the one which leads to the formulation underlying much of *Discrete multivariate analysis*, is the choice, which in the saturated case, allows us to represent the logarithms of the probabilities in a traditional analysis of variance format:

$$\text{(5)} \qquad \begin{aligned} \log p_{i_1,\ldots,i_q} = \mu &+ \alpha_1(i_1) + \cdots + \alpha_q(i_q) \\ &+ \beta_{12}(i_1, i_2) + \cdots + \beta_{q-1,q}(i_{q-1}, i_q) \\ &+ \cdots \\ &+ \omega_{1,\ldots,q}(i_1, \ldots, i_q), \end{aligned}$$

where $\alpha_1(\,) \ldots \omega_{1,\ldots,q}(\,)$ satisfy the usual ANOVA constraints:

$$\begin{gathered}\alpha_1(\cdot\,) = \alpha_2(\cdot\,) = \cdots = \alpha_q(\cdot\,) = 0,\\ \beta_{12}(i_1, \cdot\,) = 0; \quad \beta_{12}(\cdot, i_2) = 0, \quad \ldots, \quad \beta_{q-1,q}(\cdot, i_q) = 0,\\ \omega_{1,\ldots,q}(i_1, \ldots, i_{q-1}, \cdot\,) = 0, \quad \ldots, \quad \omega_{1,\ldots,q}(\cdot, i_2, \ldots, i_q) = 0.\end{gathered} \tag{6}$$

The dot used in place of an index denotes summation over that index. The parameters $\alpha_1(i_1), \ldots, \omega_{1,\ldots,q}(i_1, \ldots, i_q)$ have the usual ANOVA interpretation: μ denotes an overall effect; $\alpha_1(i_1)$ denotes an effect due to A_1 (at "level" i_1); $\beta_{12}(i_1, i_2)$ denotes a second-order interaction effect between A_1 and A_2 (at "levels" i_1 and i_2, respectively); and $\omega_{1,\ldots,q}(i_1, \ldots, i_q)$ denotes a q-order interaction among $A_1, \ldots, A_q$ (at "levels" $i_1, \ldots, i_q$, respectively); etc. Note that although $\log p_{i_1,\ldots,i_q}$ is constrained to lie on the negative axis, μ is not so fixed, and as a result the effects themselves are unconstrained in sign.

Note that the condition $< p, U_0 > = 1$ requires that

$$\mu = -\log \sum_{i_1,\ldots,i_q} \left\{\alpha_1(i_1) + \cdots + \omega_{1,\ldots,q}(i_1, \ldots, i_q)\right\}. \tag{7}$$

Substituting (7) in (5) thus shows that the log-linear probability model in the saturated case is equivalent to the multivariate generalization of the discrete logistic distribution due to Mantel (1966).

It is useful to illustrate the basis vectors for the saturated model in two simple cases. Other models may be derived from the saturated model by deleting some of the basis vectors and the parameters attached to them and thus representing the points $\log p$ in a space of lower dimensionality. Let the collection of basis vectors be represented by the matrix

$$U = [U_0, U_1, \ldots, U_{Q-1}]$$

when $m = Q$. Arrange the "effects" parameters in one long vector in an order corresponding to the ordering of probabilities, e.g.,

$$B = \begin{bmatrix} \mu \\ \alpha_1(1) \\ \vdots \\ \alpha_q(I_q) \\ \vdots \\ \omega_{1,\ldots,q}(I_1, \ldots, I_q) \end{bmatrix}$$

(5) above may be rewritten as

$$\log p = AB, \tag{8}$$

where A is a matrix reflecting the ordering of the parameters $\mu, \ldots, \omega_{1,\ldots,q}(I_1, \ldots, I_q)$ in relation to the ordering of the probabilities. For example in the 2×2 case we have

$$\begin{bmatrix} \log p_{11} \\ \log p_{12} \\ \log p_{21} \\ \log p_{22} \end{bmatrix} = \begin{bmatrix} 1\,1\,0\,1\,0\,1\,0\,0\,0 \\ 1\,1\,0\,0\,1\,0\,1\,0\,0 \\ 1\,0\,1\,1\,0\,0\,0\,1\,0 \\ 1\,0\,1\,0\,1\,0\,0\,0\,1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1(1) \\ \alpha_1(2) \\ \alpha_2(1) \\ \alpha_2(2) \\ \beta_{12}(1,1) \\ \beta_{12}(1,2) \\ \beta_{12}(2,1) \\ \beta_{12}(2,2) \end{bmatrix}.$$

Of course, the representation (8) neglects the restrictions on the parameters imposed in (6). Indeed, there are a great many more parameters in the vector B than the number of probabilities; in the 2×2 case, for example, there are only 4 probabilities (which must sum to one), but 8 elements of B (not counting μ which serves to normalize the sum of probabilities to one). To find the appropriate basis for $M = R^Q$, the restrictions (6) must be imposed on the vector B to yield a vector of parameters θ containing exactly as many components as probabilities. This is most easily accomplished by reparameterizing

(9) $$\theta = LB,$$

where L is of rank Q. Of course, different choices of L generally result in different choices of basis for M.

Suppose, for example, in the 2×2 case we choose

$$L = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{2} & -\frac{1}{2} & 0 & 0 & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ 0 & 0 & 0 & \frac{1}{2} & -\frac{1}{2} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \end{bmatrix}.$$

Then, given the restrictions (6), we have

$$\theta = \begin{bmatrix} \mu \\ \alpha_1(1) \\ \alpha_2(1) \\ \beta_{12}(1,1) \end{bmatrix}.$$

If we know θ, the full set of parameter values B may obviously be recovered from the restrictions. A basis for $M = R^Q$ may now be easily expressed in terms of A and L, for clearly

(10) $$\log p = U\theta = ULB = AB.$$

Thus

(11) $$U = AL'(LL')^{-1}.$$

(See Bock, 1975, p. 239.) For example, in the 2×2 case,

$$U = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

A basis such as U in the above example is often called a *deviation contrast* basis (Bock, 1975, p. 242). Bock (1975, p. 300) gives several other examples of parameterizations and the corresponding bases.

Since each basis vector corresponds to one of the parameters in the ANOVA representation of the logarithms of the probabilities (omitting those which may be determined from the ANOVA restrictions), computations with the log-linear probability model are most conveniently carried out using the parameters θ and the basis vectors U. Suppose, for example, we wish to eliminate the bivariate interaction effect in the 2×2 case (this is equivalent to estimating the probabilities in the 2×2 contingency table under the hypothesis that the events A_1 and A_2 are independent); it suffices to eliminate the final column of U and the least element of θ, thus restricting the vectors $\log p$ to lie in the subspace

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix} \cdot \begin{bmatrix} \mu \\ \alpha_1(1) \\ \alpha_2(1) \end{bmatrix}.$$

The 3×3 case is also revealing. Here

$$\begin{bmatrix} \mu \\ \alpha_1(1) \\ \alpha_1(2) \\ \alpha_1(3) \\ \alpha_2(1) \\ \vdots \\ \beta_{12}(1, 1) \\ \vdots \\ \beta_{12}(3, 3) \end{bmatrix}.$$

The matrix A is lengthy to write out in full but follows in an obvious manner if the probabilities are lexicographically ordered in forming the vector $\log p$. The matrix L for the deviation contrast parameterization is

$$L = \begin{bmatrix}
1 & 1/3 & 1/3 & 1/3 & 1/3 & 1/3 & 1/3 & 1/9 & 1/9 & 1/9 & 1/9 & 1/9 & 1/9 & 1/9 & 1/9 & 1/9 \\
0 & 2/3 & -1/3 & -1/3 & 0 & 0 & 0 & 2/9 & 2/9 & 2/9 & -1/9 & -1/9 & -1/9 & -1/9 & -1/9 & -1/9 \\
0 & -1/3 & 2/3 & -1/3 & 0 & 0 & 0 & -1/9 & -1/9 & -1/9 & 2/9 & 2/9 & 2/9 & -1/9 & -1/9 & -1/9 \\
0 & 0 & 0 & 0 & 2/3 & -1/3 & -1/3 & 2/9 & -1/9 & -1/9 & 2/9 & -1/9 & -1/9 & 2/9 & -1/9 & -1/9 \\
0 & 0 & 0 & 0 & -1/3 & 2/3 & -1/3 & -1/9 & 2/9 & -1/9 & -1/9 & 2/9 & -1/9 & -1/9 & 2/9 & -1/9 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 4/9 & -2/9 & -2/9 & -2/9 & 1/9 & 1/9 & -2/9 & 1/9 & 1/9 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -2/9 & 4/9 & -2/9 & 1/9 & -2/9 & 1/9 & 1/9 & -2/9 & 1/9 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -2/9 & 1/9 & 1/9 & 4/9 & -2/9 & -2/9 & -2/9 & 1/9 & 1/9 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/9 & -2/9 & 1/9 & -2/9 & 4/9 & -2/9 & 1/9 & -2/9 & 1/9
\end{bmatrix}$$

The vector θ in this case is found from the vector B by omitting the certain values for each effect

$$\theta = \begin{bmatrix} \mu \\ \alpha_1(1) \\ \alpha_1(2) \\ \alpha_2(1) \\ \alpha_2(2) \\ \beta_{12}(1, 1) \\ \beta_{12}(1, 2) \\ \beta_{12}(2, 1) \\ \beta_{12}(2, 2) \end{bmatrix},$$

so that $\alpha_1(3)$, $\alpha_2(3)$, $\beta_{12}(1, 3)$, $\beta_{12}(2, 3)$, $\beta_{12}(3, 1)$, $\beta_{12}(3, 2)$ and $\beta_{12}(3, 3)$ are deleted from B to obtain θ. The matrix of basis vectors in this case is

$$U = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & -1 & -1 & -1 & -1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & -1 & -1 & 0 & 0 & -1 & -1 \\ 1 & -1 & -1 & 1 & 0 & -1 & 0 & -1 & 0 \\ 1 & -1 & -1 & 0 & 1 & 0 & -1 & 0 & -1 \\ 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

To eliminate the bivariate interaction effect, we delete the last four columns of U above.

More complicated examples can easily be generated but would serve no useful purpose here. A general method for finding the deviation contrast basis may be derived as an extension of Theorem 1 in Nerlove and Press (1973, p. 13). This, and several other types of bases, which, however, do not necessarily preserve the parameterization in (5), are given by Bock (1975, pp. 528–538).

As indicated above, many unsaturated models may be generated from the saturated case by eliminating one or more basis vectors and suppressing the corresponding parameters. For example, to estimate a model, assuming complete independence of the events $\mathcal{Q}$, we may suppress all interaction effects. Models involving less restrictive forms of independence arise when interactions are selectively eliminated. For example, in the three-variable case, suppressing arguments, we have three main effects, α_1, α_2, and α_3, the three bivariate interactions, β_{12}, β_{13}, β_{23}, and one three-way interaction γ_{123}. If β_{12} and γ_{123} vanish, the events A_1 and A_2 are said to be conditionally independent given A_3. The notion of conditional independence may be easily generalized to sets of events: thus, if two sets of events are conditionally independent given a third set, if we classify by the latter we have independence in the usual sense, but if we do not so classify, the first two sets of events are dependent by virtue of their association with the third classificatory set. (See Nerlove and Press, 1976, pp. 24–25.) An important class of models, called *hierarchical*, is defined as the family such that, if any main or interaction term is zero, all higher-order interaction terms involving the same set of variables are zero as well. For example, in the 4-variable case, if β_{12} is zero and, say, γ_{123}, γ_{124}, and δ_{1234} are also zero, the model is hierarchical. But if β_{12} is zero and, say, γ_{123} is not, the model does not belong to the class of hierarchical models. (See Bishop, Fienberg, and Holland, 1975, p. 34.) Hierarchical models represent a generalization of the notion of conditional independence in a different sense than the result of Nerlove and Press referred to above: Clearly, a hierarchical model with a two-factor effect absent is a model in which those two variables are conditionally independent given the remaining variables; however, higher-order hierarchical models do not imply conditional independence but represent a more complicated structural relationship among variables.

Why do we wish to estimate, and/or test hypotheses about, unsaturated models? Bishop, *et al*, (1975) answer this question at several points (e.g., pp. 11, 45, 122); indeed, Chapter 4 is, in a sense, an extended answer. First, we want to uncover "structure" in the data. Testing for independence between two variables is an elementary example. To do so we set up a model with the bivariate interaction of interest equal to zero then fit a second model including this interaction; the difference in measures of goodness of fit between the two models may be used as a test statistic. Second, models with few parameters play an important role in "smoothing" the data, i.e., in obtaining cell estimates that are more stable than the observed cell counts and, relatedly, in obtaining cell estimates for every elementary cell in a sparse array (containing many cells with zero counts which are not *a priori* empty). "Smoothing" is also related to the understanding of structure, since, "When we fit a log-linear model to multivariate discrete data, the fitted cell estimates provide a smoothed description of the data because the most important structural elements are retained and random sampling fluctuations are damped" (Bishop, *et al*, 1975, p. 123). Third, unsaturated models provide a means of detecting "outliers," i.e., observations which do not really "belong" to the population being described or which are categorized incorrectly. In this case, some cells will appear anomalous with respect to the general pattern suggested by the model. Finally, models aid in providing summary statistics. If, for example, in the three-variable case, events A_1 and A_2 are conditionally independent, the relation between A_1 and A_3 can be safely described by ignoring A_2, i.e., looking at marginal tables involving only A_1 and A_3. This device is called *collapsing* the table (Bishop, *et al*, 1975, pp. 47–48).

Tables which contain cells which are *a priori* zero are called *incomplete*. Such tables are difficult to handle in practice and sometimes are not recognized as such by the investigator. Log-linear models are frequently useful in this connection as demonstrated by Bishop, *et al.*, in a long chapter (pp. 176–228).

Except for two chapters dealing with basic sampling theory for discrete data and the asymptotic results used elsewhere in the book (pp. 435–530) a chapter entitled "Other Methods for Estimation and Testing in Cross-Classifications" (pp. 343–372), and, to a considerable degree, a chapter dealing with measures of association and agreement (pp. 373–400), the bulk of this very large book deals with log-linear probability models and their application to numerous theoretical problems and empirical applications. It represents a collaborative effort on the part of three primary, and two secondary, authors.

The most senior stimulus came from F. Mosteller of the Harvard Statistics Department, whose interests in applying statistics across a very broad spectrum of fields are well known in biology, medicine and the social sciences. Mosteller's paper (1968) on association and estimation in contingency tables is a classic in this field. Bishop, Fienberg and Light were Mosteller's Ph. D. students some years ago (Bishop is now with Harvard's School of Public Health, Fienberg is now with the Department of Applied Statistics at the University of Minnesota, and R. Light is now with Harvard's Graduate School of Education). Paul Holland was formerly a member of Mosteller's department at Harvard and is now at the Educational Testing Service in Princeton. The diversity of fields of interest represented by the authors is reflected in the tremendous variety of empirical examples, discussion of which constitutes a substantial fraction of this book. These examples range over Anthropology, the Bible, Biology, Education, Linguistics, Political Science, Sociology, and Sports to Zoology. Fifty-eight data sets are listed in a special index at the end of the book.

The book contains 14 chapters, the last two of which give background material. The treatment of discrete multivariate analysis is virtually entirely in the context of cross-classifications. Rectangular tables are discussed in Chapter 2, first for 2 × 2 tables, with examples, and then in general. Chapters 3 and 4 discuss maximum likelihood estimation and goodness-of-fit. Chapter 5 deals with the corresponding problems for incomplete tables (i.e., those containing structural zeros). Chapter 6 shows how contingency tables and log-linear models can be used to estimate the size of a closed population. Chapter 7 demonstrates how the same methods may be applied to Markov chains. Chapter 8 treats the special problems of symmetry and marginal homogeneity in square tables. Chapter 9 contains many examples and illustrates how models are built. Chapter 10 is a catch-all for methods of estimation and testing in cross-classifications different from those based on the log-linear model. Chapter 11 summarizes the sequence of four well-known papers of Goodman and Kruskal (1954, 1959, 1963, 1972) on measures of association in contingency tables. The chapter treats only the case of two-dimensional tables with no structural zeros. Chapter 12 adopts a "pseudo-Bayesian" viewpoint of how to deal with sampling zeros (nonstructural zeros) in a contingency table. Chapter 13 is a survey of some of the properties of the basic (generally univariate) discrete distributions, while the final Chapter 14 is a summary of large sample theory including the use of o and O notation, convergence in probability and in distribution, the delta method, and variance stabilizing transformations.

While Bishop, *et al.*, is considerably more comprehensive than its competitors mentioned above, it is not yet the definitive treatment of discrete multivariate analysis its title suggests. Such a book has not been written. Among other things, a truly all-encompassing treatment would include an extensive discussion of logistic models, with all of the log-linear model effects of the basic jointly dependent qualitative variables being permitted to depend upon explanatory variables, so that typical multivariate regression relationships result. It would include analysis of empirical *cdfs*, empirical moments both marginal and joint, an extensive treatment of the multivariate discrete distributions and their properties, mixed discrete and continuous distributions, a treatment of how to handle ordered discrete data, how to handle missing observations other than in a cross-classification context, and available computer programs for analyzing cross-classified data. (There is regrettably almost no discussion of computer programs in the book.) It would also present the Bayesian method of making inferences about discrete data in a unified way. Bishop, *et al.*, mention Bayesian methods almost as an afterthought in Chapter 12, after 400 pages, and then the authors advocate using prior distributions arrived at by first studying the sample data (a practice totally inconsistent with the application of Bayes' theorem). Strictly Bayesian methods of inference are given short shrift, a total of three pages (pp. 405–407) in a book of 557 pages!

Although Bishop, *et al.*, is an excellent book on the log-linear model and a major contribution, it is definitely not a "cook book," in the sense that a cook book is generally arranged in a functional or utilitarian way. Suppose, for example, you are a practitioner and you have collected sample survey data from a control group and an experimental group on a vector of characteristics. You want to know how to test whether there is an effect on the experimental group of various characteristics. That is, you want to do a test of independence in a rectangular $2 \times K$ table. First, you must recognize it is a test of independence that you wish to do. Then, you go to the book and you look up independence in the Subject Index (there is hardly any way to find it in the Table of Contents). We are informed that on pp. 28–29 we can find "independence in a rectangular array" discussed. But on those pages we are not told how to carry out a test. So we go back to the Index and try "independence–as lack of association," pp. 374–380. Now, on page 375 we are told that $(I - 1)(J - 1)$ degrees of freedom are available in a two-dimensional contingency table; still no test. Indeed, there is no one place to which we are directed, as we would be in any elementary text, as to what test to use when parameters are known; what test to use when they are unknown but estimated from grouped data by MLE; and what test to use when parameters are estimated from individual observations by MLE (we must wait until p. 523 to find out what to do in this case). In fact, nowhere in the book can we even find a proof of the simple fact that in a 2×2 table, independence implies and is implied by the equality of the cross products.

In short, it helps to know a good deal of what *Discrete multivariate analysis* is about, before reading the book!

References

M. S. Bartlett, 1935, *Contingency table interactions*, J. Roy. Statist. Soc., Suppl. **2**, 248–252.

M. W. Birch, 1963, *Maximum-likelihood estimation in three-way contingency tables*, J. Roy. Statist. Soc. Ser. B **25**, 220–233.

Y. M. M. Bishop, 1969, *Full contingency tables, logits, and split contingency tables*, Biometrics **25**, 383–400.

R. D. Bock, 1975, *Multivariate statistical methods in behavioral research*, McGraw-Hill, New York.

D. R. Cox, 1970, *The analysis of binary data*, Methuen, London.

A. P. Dempster, 1969, *Elements of continuous multivariate analysis*, Addison-Wesley, Reading, Mass.

R. A. Fisher, 1934, *Statistical methods for research workers*, Oliver and Boyd, Edinburgh.

______, 1936, *Has Mendel's work been rediscovered*?, Ann. Sci. **1**, 115–137.

______, 1941, *The negative binomial distribution*, Ann. Eugenics **11**, 182–187.

J. L. Fleiss, 1973, *Statistical methods for rates and proportions*, Wiley, New York.

I. J. Good, 1965, *The estimation of probabilities*, MIT Press, Cambridge, Mass.

L. A. Goodman, 1968, *The analysis of cross-classified data: independence, quasi-independence, and interaction in contingency tables with or without missing entries*, J. Amer. Statist. Assoc. **63**, 1091–1131.

______, 1969, *On partitioning χ^2 and detecting partial association in three-way contingency tables*, J. Roy. Statist. Soc. Ser. B **31**, 486–491.

______, 1970, *The multivariate analysis of qualitative data: interactions among multiple classifications*, J. Amer. Statist. Assoc. **65**, 226–256.

______, 1971a, *The analysis of multidimensional contingency tables: stepwise procedures and direct estimation methods for building models for multiple classifications*, Technometrics **13**, 33–61.

______, 1971b, *Partitioning of chi-square, analysis of marginal contingency tables and estimation of expected frequencies in multidimensional contingency tables*, J. Amer. Statist. Assoc. **66**, 339–344.

______, 1972a, *A modified multiple regression approach to the analysis of dichotomous variables*, Amer. Sociological Rev. **37**, 28–46.

______, 1972b, *A general model for the analysis of surveys*, Amer. J. Sociology **77**, 1035–1086.

______, 1972c, *Causal analysis of data from panel studies and other kinds of surveys* (unpublished).

L. A. Goodman and W. H. Kruskal, 1954, *Measures of association for cross-classifications*, J. Amer. Statist. Assoc. **49**, 732–764.

______, 1959, *Measures of association for cross-classifications*. II, J. Amer. Statist. Assoc. **54**, 123–163.

______, 1963, *Measures of association for cross-classifications*. III, J. Amer. Statist. Assoc. **58**, 310–364.

______, 1972, *Measures of association for cross-classifications*. IV, J. Amer. Statist. Assoc. **67**, 415–421.

S. J. Haberman, 1974, *The analysis of frequency data*, Univ. of Chicago Press, Chicago, Ill.

H. O. Lancaster, 1969, *The chi-squared distribution*, Wiley, New York.

N. Mantel, 1966, *Models for complex contingency tables and polychotomous response curves*, Biometrics **22**, 83–110.

A. E. Maxwell, 1961, *Analysing qualitative data*, Methuen, London.

F. Mosteller, 1968, *Association and estimation in contingency tables*, J. Amer. Statist. Assoc. **63**, 1–28.

M. Nerlove and S. J. Press, 1976, *Multivariate log-linear probability models for the analysis of qualitative data*, Discussion Paper No. 1, Center for Statistics and Probability, Northwestern Univ., Evanston, Ill.

I. Olkin, 1977, *A statistical technique*, Science **196**, 156–157.

K. Pearson, 1900, *On a criterion that a given system of deviations from the probable in the case of a corrolated system of variables is such that it can reasonably be supposed to have arisen from random sampling*, Philosophical Mag. **50**, 157–175.

R. L. Plackett, 1974, *The analysis of categorical data*, Griffin, London.

MARC NERLOVE

S. JAMES PRESS