A Survey of Strategies for Modeling Cross-Classifications Having Ordinal Variables

Author(s): Alan Agresti

Source: *Journal of the American Statistical Association*, Vol. 78, No. 381 (Mar., 1983), pp. 184-198
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: https://www.jstor.org/stable/2287128
Accessed: 20-11-2024 18:38 UTC

A Survey of Strategies for Modeling Cross-Classifications Having Ordinal Variables

ALAN AGRESTI*

A survey is given of the main strategies for modeling cross-classifications that contain ordinal variables. Models are described for two-way tables in which one or both classifications are ordered and for multidimensional tables in which at least one classification is ordered. Primary emphasis is given to construction, interpretation, and implications of loglinear and logit models.

KEY WORDS: Loglinear models; Logit models; Odds ratios; Ordered categorical data; Association; Tests of independence.

1. INTRODUCTION

In the past two decades, methods for analysis of categorical data have become increasingly well developed. This is reflected by the many books that have recently appeared on this topic, particularly since publication of the outstanding compendium by Bishop, Fienberg, and Holland (1975). Most of these books, however, pay little attention to methods for ordinal categorical data. Specialized models for ordinal variables have been developed more recently and in a somewhat fragmented manner. Hence, standard loglinear and logit models designed for cross-classifications of nominal variables are routinely applied in practice to tables that contain ordinal variables.

The purpose of this article is to describe, in an integrated manner, the main strategies that exist for modeling cross-classifications that contain ordinal variables. Many of these methods have been formulated here differently than in the original sources in order to clarify their relationships with the standard models for nominal variables and to make comparisons among them easier. We place special emphasis on providing interpretations and implications of the models. Throughout the article we suggest descriptive measures derived from model parameters that aid in interpreting the modeled associations and that can be used in conjunction with the traditional model-free measures of association.

We assume that the reader has a basic familiarity with loglinear and logit models. Most of this article concerns the construction of these types of models for ordinal variables. Section 2 deals with loglinear models, first for the case of a two-way table with one ordinal and one nominal variable, then for the case of a two-way table with two ordinal variables, and finally for the case of a multidimensional table having at least one ordinal variable. Section 3 deals with logit models for the same three settings. Unlike loglinear models, the logit models require the identification of a response variable, which we assume to be ordinal. The same examples are used in Section 3 as in Section 2 in order to facilitate comparisons of the models and the interpretations of their parameters. In Section 4 we discuss briefly some alternative models for ordinal variables. Methods for fitting the models of Sections 2 through 4 are summarized in Section 5. In the final section we make a critical comparison of the various model types.

In what follows, G^2 denotes the likelihood ratio statistic

$$G^2 = 2 \sum n \log(n/\hat{m})$$

for testing the goodness of fit of a model by comparing observed frequencies $\{n\}$ to maximum likelihood estimates $\{\hat{m}\}$ of expected frequencies satisfying the model. Pearson chi-squared values are similar to the G^2 values for all examples and are therefore omitted. All inferential statements assume one of the usual sampling models for categorical data—full multinomial, independent multinomial, or independent Poisson sampling.

To motivate the study of specialized models for ordinal variables, consider the data in Table 1. These data, reported by Grizzle, Starmer, and Koch (1969), were taken from a study comparing four operations for treating duodenal ulcer. The operations correspond to removal of various amounts of the stomach and thus have a natural ordering. The dumping severity variable describes the extent of an undesirable potential consequence of the operation. The categories of this variable are also ordered, with "none" representing the most desirable result. The operations were performed at four hospitals. Sampling details are not provided by Grizzle, Starmer, and Koch (1969), but the near uniformity of operation frequencies for each hospital suggests that the two-way operation-hospital marginal counts should be treated as fixed. Hence, we shall treat the cell counts in Table 1 as outcomes of independent multinomial samples taken at the 16 combinations of hospital and operation.

* Alan Agresti is Associate Professor, Department of Statistics, University of Florida, Gainesville, FL 32611. Computer support was provided by the Northeast Regional Data Center at the University of Florida. Part of this research was completed while the author was visiting the Mathematics Department at Imperial College. The author is grateful to Dr. Jane Pendergast and Abbas Kezouh for computing assistance, and to Dr. Clifford Clogg and Dr. Peter McCullagh for providing personal computer programs.

© Journal of the American Statistical Association
March 1983, Volume 78, Number 381
Invited Essay Review

When we fit the standard hierarchical loglinear models to Table 1, we observe that the model under which dumping severity is jointly independent of operation and hospital fits fairly well. The G^2 goodness-of-fit statistic equals 31.64, based on 30 degrees of freedom (df), yielding a P value of $P = .385$. More complex standard models do not provide marked improvements in fit. These standard models ignore the ordinal nature of two of the variables, however. Their parameters do not describe well the types of departures from independence addressed, for example, by the question "Does dumping severity tend to increase when more of the stomach is removed?" Using various analyses in this article, we shall, in fact, obtain strong evidence of this type of association in Table 1.

2. LOGLINEAR MODELS

In the standard construction of loglinear models given in basic references such as Bishop, Fienberg, and Holland (1975), all variables are treated as if they are nominal in scale. That is, the parameter estimates and chi-squared statistics are invariant under reorderings of categories of any variable. More appropriate models are available when one uses the extra information provided by the natural orderings of the categories of the ordinal variables in the cross-classification.

For example, consider the two-dimensional table, with row variable denoted by X and column variable denoted by Y. Let m_{ij} denote the expected frequency in the cell in row i and column j, $1 \le i \le r$, $1 \le j \le c$. For this table we do not usually expect the independence model

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y \tag{2.1}$$

to give an adequate fit. However, in the usual hierarchical scheme, the model of next greater complexity is the saturated one having an additional $(r - 1)(c - 1)$ independent λ_{ij}^{XY} parameters. If one or both variables are ordinal, though, simple models exist that are more complex and realistic than the independence model, yet that are not saturated.

Similar remarks apply to higher-dimensional tables. Consider the model for a three-dimensional $r \times c \times l$ table,

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}. \tag{2.2}$$

This model, for which df $= (r - 1)(c - 1)(l - 1)$, includes partial association terms for each pair of variables but no three-factor interaction term. The model is quite general as it assumes no structure for the form of the two-factor associations. When at least one of the variables is ordinal, more parsimonious models can be constructed that posit certain simple types of associations. Also, unlike the standard models for three nominal variables, unsaturated models exist having three-factor interaction terms, in which a simple structure is posited for the nature of that interaction.

2.1 Loglinear Model for Ordinal-Nominal Table

Suppose that the column variable of a two-dimensional table is ordinal and that scores $\{v_j\}$ are assigned to the columns, where $v_1 < v_2 < \ldots < v_c$. In many applications the choice of scores will reflect assumed distances between categories for an underlying interval scale. In the absence of such information, the choice of equal-interval scores results in the simplest interpretation for the following model (assuming that it fits with that choice of scores). Further discussion of the implications of having to assign these scores is given in Section 6.2.

A loglinear model that is more complex than the independence model (2.1) and uses the ordinal information is given by

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y + \tau_i^X(v_j - \bar{v}), \tag{2.3}$$

where without loss of generality the model parameters satisfy

$$\sum \lambda_i^X = \sum \lambda_j^Y = \sum \tau_i^X = 0.$$

For loglinear models the residual df for testing goodness of fit equals the difference between the number of cells in the table and the number of linearly independent pa-

Table 1. Cross-Classification of Duodenal Ulcer Patients According to Operation, Hospital, and Dumping Severity (Operation A is drainage and vagotomy, B is 25 percent resection and vagotomy, C is 50 percent resection and vagotomy, D is 75 percent resection)

	Dumping Severity											
	Hospital I			Hospital II			Hospital III			Hospital IV		
Operation	None	Slight	Mod.	None	Slight	Mod.	None	Slight	Mod.	None	Slight	Mod.
A	23	7	2	18	6	1	8	6	3	12	9	1
B	23	10	5	18	6	2	12	4	4	15	3	2
C	20	13	5	13	13	2	11	6	2	14	8	3
D	24	10	6	9	15	2	7	7	4	13	6	4

Source: Grizzle, Starmer, and Koch (1969).

rameters in the model. For model (2.3), therefore,

$$\begin{aligned} \text{df} &= rc - [1 + (r - 1) + (c - 1) + (r - 1)] \\ &= (r - 1)(c - 2), \end{aligned}$$

and the model is unsaturated when $c > 2$.

The independence model is the special case of model (2.3) with all $\tau_i^X = 0$. Note that the association term $\tau_i^X (v_j - \bar{v})$ depends on the ordinal variable Y only through its scores $\{v_j\}$. This term reflects a departure of $\log m_{ij}$ from independence, which is linear in Y for fixed X. If $\tau_i^X > 0$ ($\tau_i^X < 0$), the probabilities in row i are higher above $\bar{v}$ (below $\bar{v}$) than would be expected if the variables were independent.

The $\{\tau_i^X\}$ can also be interpreted through log odds and log odds ratios. For a pair of rows i and i',

$$\log m_{ij} - \log m_{i'j} = (\lambda_i^X - \lambda_{i'}^X) + (\tau_i^X - \tau_{i'}^X)(v_j - \bar{v}). \quad (2.4)$$

That is, on a log scale, the difference between the proportions in any two rows is a linear function of the ordinal variable Y, with slope equal to the difference in the relevant tau parameters. If $\tau_i^X > \tau_{i'}^X$, the conditional Y distribution at the ith level of X is stochastically larger than the conditional Y distribution at the i'th level of X. For an arbitrary pair of rows i and i' and an arbitrary pair of columns $j > j'$,

$$\log \left[\frac{m_{ij}/m_{ij'}}{m_{i'j}/m_{i'j'}} \right] = (\tau_i^X - \tau_{i'}^X)(v_j - v_{j'}). \quad (2.5)$$

Hence the log odds ratio is proportional to the distance between the columns and is always positive if $\tau_i^X > \tau_{i'}^X$. For the integer scores $\{v_j = j\}$, the log odds ratio takes on the constant value $\tau_i^X - \tau_{i'}^X$, for all $c - 1$ pairs of adjacent columns.

We will refer to the $\{\tau_i^X\}$ as row effects and to model (2.3) as the loglinear row-effects model. Model (2.3) or special cases of it have been discussed by Simon (1974, Formulation A), Haberman (1974), Goodman (1979a), Andrich (1979), Duncan (1979), Duncan and McRae (1979), and Fienberg (1980, pp. 61–64). Goodman's row-effects model is the special case of model (2.3) in which the $\{v_j\}$ are equal-interval, so that the log odds ratio for adjacent columns depends only on the rows involved.

The 3×3 cross-classification in Table 2 gives the joint distribution of the ordinal variable "political ideology" and the nominal variable "party affiliation" for a sample of voters taken by Hedlund (1978) in the 1976 presidential primaries in Wisconsin. In analyzing these data, we might wish to discern whether members of one party tend to be more liberal or tend to be more conservative than members of another party. The first set of parenthesized values in Table 2 consists of maximum likelihood estimates of expected frequencies for model (2.3) with integer scores $\{v_j = j\}$. This model gives a reasonably good fit,

Table 2. Cross-Classification of Voters in 1976 Wisconsin Primaries According to Party Affiliation and Political Ideology (Parenthesized values are maximum likelihood estimates of expected frequencies under loglinear model (2.3) and logit model (3.2), respectively)

Party Affiliation	Political Ideology: Conservative	Moderate	Liberal	Total
Democrat	100 (93.6,92.2)	156 (168.7,170.4)	143 (136.6,136.4)	399
Independent	141 (145.8,144.3)	210 (200.4,203.3)	119 (123.8,122.5)	470
Republican	127 (128.6,129.1)	72 (68.9,65.0)	15 (16.6,19.9)	214
Total	368	438	277	1083

Source: Hedlund (1978).

with $G^2 = 2.81$ based on df $= 2$. Maximum likelihood methods for fitting the row-effects model and other ordinal loglinear models are described in Section 5.

For the row-effects model applied to Table 2, the association parameter estimates are $\hat{\tau}_1^X = .495$, $\hat{\tau}_2^X = .224$, and $\hat{\tau}_3^X = -.719$. This indicates that in the sample, the Democrats tend to be the most liberal group, and the Republicans tend to be much more conservative than the other two groups. This model predicts constant log odds ratios for adjacent columns of political ideology. For example, $\hat{\tau}_1^X - \hat{\tau}_3^X = 1.214$ means that the odds of being classified liberal instead of moderate and the odds of being classified moderate instead of conservative are $\exp(1.214) = 3.37$ times higher for Democrats than for Republicans.

The independence model yields $G^2 = 105.66$ based on df $= 4$ for these data. The difference in likelihood ratio statistics for model (2.1) and model (2.3) is 102.85, based on df $= 4 - 2 = 2$. Given that the row-effects model holds, this difference is a chi-squared test statistic for testing independence (H_0: $\tau_1^X = \tau_2^X = \tau_3^X = 0$). For these data it indicates strong evidence of an association. If the row-effects model holds, this conditional test will be asymptotically more powerful at detecting an association than the test with df $= (r - 1)(c - 1)$ based on model (2.1), which ignores the ordinal nature of Y (see Goodman 1981c). The conditional chi-squared test has df $= r - 1$ and is analogous in intent to the Kruskal-Wallis test for comparing r groups on an ordinal response. The test statistic there (corrected for ties on the response) also has a null asymptotic chi-squared distribution with df $= r - 1$. For these data the Kruskal-Wallis statistic equals 96.59.

2.2 Loglinear Model for Ordinal-Ordinal Table

Suppose now that both the column and the row variables of a two-dimensional table are ordinal. We assume

Table 3. Likelihood Ratio Statistics for Standard Loglinear Models Fit to Data of Table 1 (D denotes dumping severity, O denotes operation, H denotes hospital)

Fitted Marginals	G^2	DF	P-value
(D,O,H)[a]	32.61	39	.755
(DO,H)[a]	21.73	33	.933
(DH,O)[a]	24.51	33	.857
(OH,D)	31.64	30	.385
(DO,DH)[a]	13.63	27	.985
(OD,OH)	20.76	24	.653
(HD,HO)	23.54	24	.488
(DO,DH,OH)	12.50	18	.820

[a] These models are appropriate if the *H* margin or *O* margin alone (or neither margin) are fixed, but not if the *O–H* marginal distribution is considered fixed.

here that scores $\{u_i\}$ and $\{v_j\}$ are assigned to the rows and columns, respectively, where $u_1 < u_2 < \ldots < u_r$ and $v_1 < v_2 < \ldots < v_c$. A simple loglinear model that uses the ordinal information but that has only one more parameter than the independence model is given by

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y + \beta^{XY}(u_i - \bar{u})(v_j - \bar{v}). \quad (2.6)$$

Here $\sum \lambda_i^X = \sum \lambda_j^Y = 0$, and df $= (r - 1)(c - 1) - 1$.

In model (2.6) the association term $\beta^{XY}(u_i - \bar{u})(v_j - \bar{v})$ reflects a deviation from independence that is linear in X for fixed Y and linear in Y for fixed X. We will refer to this model as the linear-by-linear association model. The independence model is the special case of $\beta^{XY} = 0$. If $\beta^{XY} > 0$ we expect more observations to have large X and large Y values or small X and small Y values than if X and Y are independent. The magnitude of β^{XY} can be interpreted as follows. For an arbitrary pair of rows $i < i'$ and an arbitrary pair of columns $j < j'$, note that

$$\log[m_{ij}m_{i'j'}/m_{ij'}\, m_{i'j}] = \beta^{XY}(u_{i'} - u_i)(v_{j'} - v_j). \quad (2.7)$$

That is, the log odds ratio is directly proportional to the product of the distance between the rows and the distance between the columns. Hence, β^{XY} is the log odds ratio per unit distances $u_{i'} - u_i = v_{j'} - v_j = 1$ on X and Y.

Birch (1965), Nelder and Wedderburn (1972), Haberman (1974), and Goodman (1979a) have suggested this model in various forms. Goodman studied it for the special case $\{u_i = i\}$, $\{v_j = j\}$ in which the local odds ratio $\theta_{ij} = m_{ij}m_{i+1,j+1}/m_{i,j+1}m_{i+1,j}$ for adjacent rows i and $i+1$ and adjacent columns j and $j+1$ is uniformly exp (β^{XY}). He referred to that special case as the uniform association model.

We now reconsider the data in Table 1 on dumping severity for ulcer operations. Table 3 contains results of fitting standard loglinear models. We mentioned in the introduction that G^2[(D, OH)] = 31.64 based on df = 30, where (D, OH) symbolizes the model where dumping severity (D) is jointly independent of operation type (O) and hospital (H) (that is, the fitted marginals are D and O-H). The best-fitting model at the next level of complexity in the standard hierarchy is symbolized by (OD, OH), and has $G^2 = 20.76$ based on df = 24. The statistic

$$G^2[(\text{D, OH})] - G^2[(\text{OD, OH})] = 10.88$$

based on df = 30 − 24 = 6 gives a test of the independence of dumping severity and operation, under the assumption that dumping severity is independent of hospital for each operation. This test is identical to the likelihood ratio test of independence for the marginal table (see Table 4) relating dumping severity and operation, since for the structure (OD, OH) we can collapse over the hospital dimension in studying the association between the other two variables. The P value here of about .10 provides only weak evidence that dumping severity and operation are associated, particularly since we used the maximum improvement in fit in guiding our choice of model.

In analyzing the association between dumping severity and operation, we have not used their ordinal nature. We can do this by fitting the linear-by-linear association model (2.6) to Table 4. The first set of parenthesized values in Table 4 consists of maximum likelihood estimates of expected frequencies for the uniform association (integer scores) version of that model. The model gives a good fit, with $G^2 = 4.59$ based on df = 5. The estimated constant value of the local log odds ratio is $\hat{\beta}^{XY} = .163$, reflecting a tendency for dumping to be more severe when more of the stomach is removed.

The difference in G^2 values between the two-dimensional independence model and the uniform association model is 10.88 − 4.59 = 6.29, based on df = 6 − 5 = 1. Hence, we obtain fairly strong evidence ($P < .02$) of a positive association when we use the ordinal nature of the two variables, and we may reach a different conclusion than we would with a cursory use of the standard models. This single-degree-of-freedom test of $H_0: \beta^{XY} = 0$ is similar in spirit to tests that have been proposed for other measures of association in ordinal-ordinal tables. For example, see Yates (1948).

Table 4. Cross-Classification of Duodenal Ulcer Patients According to Operation and Dumping Severity (Parenthesized values are maximum likelihood estimates of expected frequencies corresponding to loglinear model (2.6) and logit model (3.5), respectively)

	Dumping Severity			
Operation	None	Slight	Moderate	Total
A	61 (62.5,63.2)	28 (26.2,24.9)	7 (7.3,7.9)	96
B	68 (62.9,63.1)	23 (30.9,30.4)	13 (10.2,10.5)	104
C	58 (61.0,60.7)	40 (35.3,35.7)	12 (13.7,13.6)	110
D	53 (53.7,53.1)	38 (36.6,37.9)	16 (16.7,16.0)	107
Total	240	129	48	417

Source: Grizzle, Starmer, and Koch (1969).

2.3 Generalized Model for Two Dimensions

Model (2.6) is appropriate if the linear-by-linear association property holds for the sets of scores chosen by the researcher. Alternatively, one can treat the scores $\{u_i\}$ and $\{v_j\}$ as parameters $\{\mu_i\}$ and $\{\nu_j\}$ and obtain estimated scores under which that property is best approximated. Such a model can be expressed as

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y + \beta^{XY}\mu_i\nu_j, \qquad (2.8)$$

where without loss of generality $\sum \mu_i = \sum \nu_j = 0$ and $\sum \mu_i^2 = \sum \nu_j^2 = 1$.

Owing to the constraints on the score parameters df $= rc - [1 + (r - 1) + (c - 1) + 1 + (r - 2) + (c - 2)] = (r - 2)(c - 2)$, and we need a table with dimensions at least 3×3 for this model to be unsaturated. This model has been discussed by Andersen (1980, p. 211) and by Goodman (1979a, 1981a,b), who referred to it as "Model II" and as the "RC model."

Model (2.8) can be interpreted like model (2.6) through the log odds ratio (2.7) if we replace the fixed scores $\{u_i\}$ and $\{v_j\}$ by the parameters $\{\mu_i\}$ and $\{\nu_j\}$. This model differs from the previous one, though, in being invariant under interchanges of rows or columns. Hence the $\{\mu_i\}$ and $\{\nu_j\}$ need not be monotonic. Since $\log \theta_{ij} = \beta^{XY}(\mu_{i+1} - \mu_i)(\nu_{j+1} - \nu_j)$, lack of monotonicity in the scores indicates nonmonotonic associations, in the sense that local associations are positive in some locations and negative in other locations. Also, unlike model (2.6), this model is not loglinear in its parameters. It is, however, a special case of the general loglinear model for two dimensions with association term $\lambda_{ij}^{XY} = \beta^{XY}\mu_i\nu_j$.

Since model (2.8) is invariant under changes of row and column orderings, we can also use it when the row or column variable is nominal. For example, model (2.8) may be regarded as a generalization of the loglinear row effects model (2.3) if we identify the $\{\mu_i\}$ as row effects and the $\{\nu_j\}$ as parameter versions of the column scores. Generally, the $\{\mu_i\}$ may be regarded as row effects and the $\{\nu_j\}$ may be regarded as column effects. We will refer to model (2.8) as the multiplicative row- and column-effects (RC) model.

The RC model has its simplest interpretation for ordinal variables when the parameter scores are monotonic. To illustrate, suppose that Y is ordinal and that $\nu_1 < \ldots < \nu_c$. Then if $\mu_i > \mu_{i'}$, the conditional distribution of Y in row i is stochastically larger than the conditional distribution of Y in row i'. If we expect stochastic orderings, it would seem natural to fit the RC model under the added constraint that the $\{\nu_j\}$ are monotonic. For the resulting model, G^2 would be invariant only to reversals in column orderings, and the model would be truly ordinal. This approach does not seem to have been considered in the literature.

For the political ideology data (Table 2), maximum likelihood estimation of the RC model gives $G^2 = 1.67$ based on df = 1. There is little improvement in fit over the row-effects model, for which $G^2 = 2.81$ based on df = 2. We obtain $\hat{\nu}_1 = -.664$, $\hat{\nu}_2 = -.079$, and $\hat{\nu}_3 = .743$ for the estimated scores on political ideology and $\hat{\mu}_1 = .545$, $\hat{\mu}_2 = .254$, $\hat{\mu}_3 = -.799$ for the party affiliation row effects. The $\{\hat{\nu}_j\}$ are nearly evenly spaced, which illustrates why the simpler row-effects model having equal-interval column scores fits almost as well. To make the $\{\hat{\tau}_i^X\}$ from the row-effects model comparable to the $\{\hat{\mu}_i\}$ from the RC model, we scale the former so that $\sum (\hat{\tau}_i^X)^2 = 1$. We then obtain $\hat{\tau}_1^X = .549$, $\hat{\tau}_2^X = .249$, and $\hat{\tau}_3^X = -.798$, very similar to the $\{\hat{\mu}_i\}$. Since the $\{\hat{\nu}_j\}$ are monotone increasing and since $\hat{\mu}_1 > \hat{\mu}_2 > \hat{\mu}_3$ in the RC model, we again conclude that Democrats are stochastically more liberal than Independents, who are themselves stochastically much more liberal than Republicans.

Similarly, the RC model does not fit much better than the uniform association model for the dumping severity data (Table 4), as $G^2 = 2.85$ with df = 2 provides a reduction in G^2 of only 1.74 based on df = 5 − 2 = 3. To make the estimated association parameter comparable for the two models, we need to use the same scaling for the fixed scores in (2.6) as for the parameter scores in (2.8). A scaling that results in a meaningful interpretation for $\hat{\beta}^{XY}$ is to use scores having means of zero and standard deviations of one with respect to the marginal distributions; that is,

$$\sum u_iP_{i+} = \sum v_jP_{+j} = \sum \hat{\mu}_iP_{i+} = \sum \hat{\nu}_jP_{+j} = 0,$$
$$\sum u_i^2P_{i+} = \sum v_j^2P_{+j} = \sum \hat{\mu}_i^2P_{i+} = \sum \hat{\nu}_j^2P_{+j} = 1,$$

where $P_{ij} = n_{ij}/n$. This rescaling yields $\hat{\beta}^{XY} = .124$ for model (2.6) and $\hat{\beta}^{XY} = .140$ for model (2.8). Hence, both models suggest that the odds ratio equals approximately $\exp(.13) = 1.14$ for distances of one standard deviation on both dumping severity and operation.

Goodman (1981b) pointed out that for these standardized scores, models (2.6) and (2.8) for expected frequencies have the same form as the bivariate normal density if we identity β^{XY} with $\rho/(1 - \rho^2)$, where ρ is the Pearson correlation. The $\hat{\beta}^{XY}$ values obtained for Table 4 correspond to correlation values of .122 and .137. The actual correlation values $\sum \sum u_iv_jP_{ij}$ and $\sum \sum \hat{\mu}_i\hat{\nu}_jP_{ij}$ are .122 and .138, respectively. Haberman (1981) introduced a test of $H_0\colon \beta^{XY} = 0$ for the RC model. He showed that under the null hypothesis of independence, the test statistic is asymptotically equivalent to one based on a canonical correlation analysis.

2.4 Loglinear Models for Multidimensional Table

The models of Sections 2.1 through 2.3 can be readily generalized to multidimensional tables having at least one ordinal variable. We will illustrate for the $r \times c \times l$ cross-classification of three variables X, Y, and Z having expected frequencies $\{m_{ijk}\}$.

The hierarchical loglinear models of usual interest for three dimensions range from the simple mutual independence model

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z \qquad (2.9)$$

to the model (2.2), which contains all the partial association terms but no three-factor interaction term. The next most complex model beyond (2.2) has the $(r-1)(c-1)(l-1)$ independent three-factor interaction parameters $\{\lambda_{ijk}^{XYZ}\}$ and is of little interest because it is saturated. If one or more of the variables in the cross-classification are ordinal, though, we can construct a richer hierarchy of models that includes (a) partial association models that are more parsimonious and simpler to interpret than model (2.2), and (b) three-factor interaction models that are unsaturated and also easily interpretable.

In interpreting these models we refer to two types of odds ratios. The $(r-1)(c-1)$ odds ratios

$$\theta_{ij(k)} = m_{ijk}m_{i+1,j+1,k}/m_{i+1,j,k}m_{i,j+1,k}, \qquad 1 \le i \le r-1,\ 1 \le j \le c-1$$

describe the local conditional association between X and Y within a fixed level k of Z. Similar odds ratio $\theta_{i(j)k}$ and $\theta_{(i)jk}$ describe the local conditional associations between X and Z within levels of Y and between Y and Z within levels of X. The ratio of odds ratios

$$\theta_{ijk} = \theta_{ij(k+1)}/\theta_{ij(k)} = \theta_{i(j+1)k}/\theta_{i(j)k} = \theta_{(i+1)jk}/\theta_{(i)jk}$$

is used for describing local three-factor interaction. There is an absence of three-factor interaction if all $(r-1)(c-1)(l-1)$ of the θ_{ijk} equal one.

As in the case of two-way tables, a simple way to construct loglinear models is to let the ordinal variables contribute to the association terms through linear departures of $\log m_{ijk}$ from independence. Table 5 lists these models and their residual degrees of freedom for the cases in which all three, two, or one of the variables are ordinal.

To illustrate, suppose that X, Y, and Z are all ordinal. Let $\{u_i\}$, $\{v_j\}$, and $\{w_k\}$ represent scores for the levels of X, Y, and Z, respectively. The model

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \beta^{XY}(u_i - \bar{u})(v_j - \bar{v}) + \beta^{XZ}(u_i - \bar{u})(w_k - \bar{w}) + \beta^{YZ}(v_j - \bar{v})(w_k - \bar{w}) \qquad (2.10)$$

has an association term for each pair of variables, yet has only three more parameters than the independence model (2.9) and is always unsaturated.

Model (2.10) assumes no three-factor interaction. For the special case of the integer scores $\{u_i = i\}$, $\{v_j = j\}$, and $\{w_k = k\}$,

$$\log \theta_{ij(k)} = \beta^{XY}, \quad \log \theta_{i(j)k} = \beta^{XZ}, \quad \log \theta_{(i)jk} = \beta^{YZ}. \qquad (2.11)$$

Thus, the conditional association is uniform for each pair of variables, and the strength of association is homogeneous across the levels of the third variable. That model can be described as a homogeneous uniform association model.

The other models in Table 5 can be interpreted in a similar manner. To illustrate, consider the model for two ordinal variables Y and Z. For integer scores, the parameter β^{YZ} pertains to the uniform association between Y and Z that is homogeneous across levels of X. The $\{\tau_{1i}^X\}$ represent row effects of X on the X-Y association that are homogeneous across levels of Z. Thus, the ordering of their values indicates how the levels of X are stochastically ordered with respect to the conditional Y distributions (within each level of Z). Similarly, the $\{\tau_{2i}^X\}$ represent row effects of X on the X-Z association that are homogeneous across levels of Y. For $2 \times 2 \times 2$ tables, the models in Table 5 are equivalent to the standard model (2.2). For larger tables, though, they are more parsimonious and simpler to interpret because of the structured association terms.

The models in Table 5 can also be generalized to allow for three-factor interaction. For example, a simple model for the case of three ordinal variables is

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \beta^{XY}(u_i - \bar{u})(v_j - \bar{v}) + \beta^{XZ}(u_i - \bar{u})(w_k - \bar{w}) + \beta^{YZ}(v_j - \bar{v})(w_k - \bar{w}) + \beta^{XYZ}(u_i - \bar{u})(v_j - \bar{v})(w_k - \bar{w}). \qquad (2.12)$$

This model has only one more parameter than model (2.10), so df $= rcl - r - c - l - 2$. For integer scoring, the local interaction takes the constant value β^{XYZ}, and we can refer to the model as a uniform interaction model. The association between X and Y is then uniform within each layer of Z but with strength changing linearly across the levels of Z. Hence, model (2.12) with integer scores is a type of heterogeneous uniform association model.

More general uniform interaction models have less structured associations for some or all pairs of variables. For example, if the term $\beta^{XYZ}u_i v_j w_k$ is added to the standard loglinear model (2.2), then $\log \theta_{ijk} = \beta^{XYZ}$ for integer scores, but the conditional associations are not uniform. See Goodman (1981c) for further details.

We illustrate multidimensional models using Table 6,

Table 5. Association Terms and DF for Loglinear Models in Three Dimensions with Linear Ordinal Effects

	Association Terms			
Ordinal Variables	X − Y	X − Z	Y − Z	DF
X,Y,Z	$\beta^{XY}(u_i - \bar{u})(v_j - \bar{v})$	$\beta^{XZ}(u_i - \bar{u})(w_k - \bar{w})$	$\beta^{YZ}(v_j - \bar{v})(w_k - \bar{w})$	$rc\ell - r - c - \ell - 1$
Y,Z	$\tau_{1i}^X(v_j - \bar{v})$	$\tau_{2i}^X(w_k - \bar{w})$	$\beta^{YZ}(v_j - \bar{v})(w_k - \bar{w})$	$rc\ell - 3r - c - \ell + 3$
Z	λ_{ij}^{XY}	$\tau_i^X(w_k - \bar{w})$	$\tau_j^Y(w_k - \bar{w})$	$rc\ell - rc - r - c - \ell + 3$

Table 6. Cross-Classification of Detroit Residents According to Rated Performance of Radio and TV Networks, Race, and Year of Survey (The parenthesized values are estimated expected frequencies corresponding to loglinear models (2.2) and (2.10) and logit model (3.6), respectively)

		Rated Performance of Radio and TV Networks			
Year	Race	Poor	Fair	Good	Total
1959	White	54 (53.5,49.2,50.8)	253 (246.2,254.0,250.0)	325 (332.3,328.9,331.2)	632
	Black	4 (4.5,3.6,4.6)	23 (29.8,32.3,29.7)	81 (73.7,72.0,73.7)	108
1971	White	158 (158.5,165.7,160.2)	636 (642.8,629.2,640.5)	600 (592.7,599.1,593.3)	1394
	Black	24 (23.5,21.5,24.4)	144 (137.2,140.5,135.7)	224 (231.3,230.0,231.8)	392
Total		240	1056	1230	2526

Source: Duncan and McRae (1979).

a $2 \times 2 \times 3$ cross-classification of X = race, Y = year, and Z = rated performance of radio and TV networks, taken from Duncan and McRae (1979). This table is based on surveys taken in the Detroit area in 1959 and 1971 to study social change in metropolitan communities. The standard no three-factor interaction model (2.2) fits these data fairly well, as evidenced by $G^2 = 3.57$ based on df $= 2$, but it does not use the ordinal nature of Z. All models listed in Table 5 are equivalent for these data, since the nominal variables X and Y each have two categories. The homogeneous uniform association (integer scores) version of model (2.10) provides an adequate fit, with $G^2 = 5.58$ based on df $= 4$ and the difference in G^2 values for the two models equaling 2.01 based on df $= 4 - 2 = 2$. The estimated conditional local log odds ratios are $\hat{\beta}^{XY} = .561$, $\hat{\beta}^{XZ} = .543$, and $\hat{\beta}^{YZ} = -.307$. Thus, performance of the networks tends to be rated higher by blacks than by whites and lower in 1971 than in 1959.

The uniform interaction version of model (2.12) gives a slightly better fit than either of the above-mentioned models, with $G^2 = 2.22$ based on df $= 3$. It also allows a more detailed interpretation of the data. The estimated parameters $\hat{\beta}^{XY} = .812$, $\hat{\beta}^{XZ} = .671$, $\hat{\beta}^{YZ} = -.462$, and $\hat{\beta}^{XYZ} = -.405$ reflect the fact that the tendency of blacks to rate performance of the networks higher than do whites is greater in 1959 than in 1971. That is, the two conditional X-Z local log odds ratios are predicted to be $\hat{\beta}^{XZ} - .5\,\hat{\beta}^{XYZ} = .874$ in 1959 and $\hat{\beta}^{XZ} + .5\,\hat{\beta}^{XYZ} = .469$ in 1971. Similarly, the negative shift from 1959 to 1971 in rated performance of the networks tends to be more substantial for blacks than for whites.

The models discussed in this section can be suitably modified when certain marginal distributions are fixed by the sampling design. For example, the $\{\hat{m}\}$ for model (2.10) will satisfy for all i, j, k,

$$\hat{m}_{i++} = n_{i++},\ \hat{m}_{+j+} = n_{+j+},\ \hat{m}_{++k} = n_{++k},$$
$$\sum u_i v_j \hat{m}_{ij+} = \sum u_i v_j n_{ij+},$$
$$\sum u_i w_k \hat{m}_{i+k} = \sum u_i w_k n_{i+k},$$
$$\sum v_j w_k \hat{m}_{+jk} = \sum v_j w_k n_{+jk}.$$

If we also wanted to constrain, say, all $\hat{m}_{ij+} = n_{ij+}$, we would substitute the general term λ_{ij}^{XY} for the X-Y association term. To illustrate, a simple model for Table 1 has the form

$$\log m_{ijk} = \mu + \lambda_i^{\mathrm{O}} + \lambda_j^{\mathrm{D}} + \lambda_k^{\mathrm{H}} + \lambda_{ik}^{\mathrm{OH}} + \beta^{\mathrm{OD}}(u_i - \bar{u})(v_j - \bar{v}). \quad (2.13)$$

This model treats the O-H marginal table as fixed, assumes a linear-by-linear association for dumping severity and operation, and assumes conditional independence be-

Table 7. Analysis of Association for Table 1

Association Terms						
O − H	O − D	H − D	G^2	DF	Difference in G^2	Difference in DF
λ_{ik}^{OH}	λ_{ij}^{OD}	λ_{jk}^{DH}	12.50	18		
					4.57	5
λ_{ijk}^{OH}	$\beta^{OD}u_iv_j$	λ_{jk}^{DH}	17.07	23		
					5.38	3
λ_{ik}^{OH}	$\beta^{OD}u_iv_j$	$\tau_k^{H}v_j$	22.45	26		
					2.90	3
λ_{ik}^{OH}	$\beta^{OD}u_iv_j$	—	25.35	29		
					6.29	1
λ_{ik}^{OH}	—	—	31.64	30		

tween hospital and dumping. For integer scores we obtain $G^2 = 25.35$ based on df = 29.

As with standard loglinear models, hierarchical comparisons are often useful. Table 7 summarizes the results of fitting several nested loglinear models to Table 1. This analysis of association reveals that model (2.13) is a good one for Table 1. The difference in G^2 values of 6.29 between structure (D, OH) and model (2.13) is the same as we obtained in Section 2.2 for the analogous comparison with the marginal two-way table (Table 4). This is because model (2.13) implies that we can collapse over the hospital dimension in studying the D-O association. Both analyses indicate fairly strongly that there is a positive D-O association ($\hat{\beta}^{OD} = .163$).

Generalizations of the above models to more than three dimensions are easily formulated. For example, a simple model that includes all pairwise associations has linear effects of ordinal variables through terms of the form $\beta^{XY}(u_i - \bar{u})(v_j - \bar{v})$ for pairs of ordinal variables and terms of the form $\tau_i^X(v_j - \bar{v})$ for pairings of ordinal with nominal variables, and it has general association terms for pairs of nominal variables. Models that are more complex than this can be constructed to permit parameter scores, to allow for interaction, or to allow nonlinear effects of ordinal variables. Clogg (1982) and Agresti and Kezouh (1983) give further details on models for multidimensional tables.

3. LOGIT MODELS

Let P_i denote the probability that a randomly selected observation falls in the ith category of a variable, $i = 1, \ldots, c$. For dichotomous variables, the log odds or "logit" $\log(P_2/P_1)$ is commonly formed for a response when one wishes to model that variable as a function of qualitative or quantitative explanatory variables (see, e.g., Cox 1970; Fienberg 1980, Ch. 6). When $c > 2$ there are several ways of forming a set of $c - 1$ logits (see Fienberg 1980, p. 110). For ordinal variables the most meaningful logits are those that take category order into account. Some examples of these are the "accumulated" logits

$$L_j = \log[\sum_{i>j} P_i / \sum_{i \le j} P_i], j = 1, 2, \ldots, c - 1,$$

the "continuation ratio" logits

$$\log[P_{j+1} / \sum_{i \le j} P_i], j = 1, 2, \ldots, c - 1,$$

and the "adjacent categories" logits

$$\log[P_{j+1}/P_j], j = 1, 2, \ldots, c - 1.$$

A nice feature of the continuation-ratio logits is that the results of fitting models to separate logits are independent. The accumulated logits use all categories for each logit, and are necessarily monotone since they are logits of distribution function values. The difference between two groups on this logit scale will be constant if the underlying continuous distributions are of the logistic form, since the logit transform of the logistic distribution function $F(x) = [1 + \exp(-(x + \tau))]^{-1}$ is an additive function of the location parameter τ. We shall use this type of logit in the models in this section, though the models still make sense with the other types.

Logit models are of special interest when one variable is a response variable. In this section we will describe logit models for ordinal response variables, first for the case of a two-way table where the other variable is nominal or ordinal and then for a multidimensional table.

3.1 Logit Model for Ordinal-Nominal Table

Suppose that the column variable of a two-dimensional table is ordinal, and let L_{ij} denote the jth accumulated logit within row i; that is,

$$L_{ij} = \log\left(\frac{m_{i,j+1} + \cdots + m_{ic}}{m_{i1} + \cdots + m_{ij}}\right). \quad (3.1)$$

A simple additive model for the logits is

$$L_{ij} = \mu_j + \tau_i^X, i = 1, \ldots, r, j = 1, \ldots, c - 1, \quad (3.2)$$

where $\sum \tau_i^X = 0$. For logit models df is obtained by subtracting the number of linearly independent parameters in the model from the number of logits formed within the table. For model (3.2) there are $r(c - 1)$ logits and $(c - 1) + (r - 1)$ independent parameters, so df = $(r - 1)(c - 2)$, the same as for the loglinear model for this setting (model (2.3)). Among the authors who have suggested model (3.2) in various forms are Snell (1964), Williams and Grizzle (1972), Simon (1974, Formulation B), Clayton (1974), Bock (1975, pp. 544–546), and McCullagh (1979,1980).

The parameters in model (3.2) are quite easy to interpret. Note that $\sum_i L_{ij}/r = \mu_j$, so the $\{\mu_j\}$ are average logits and are monotonic decreasing. The $\{\tau_i^X\}$ are row-effect parameters that specify the nature of the association. For an arbitrary pair of rows i and i', the difference in logits

$$L_{ij} - L_{i'j} = \tau_i^X - \tau_{i'}^X \quad (3.3)$$

is constant for all $c - 1$ logits. If $\tau_i^X > \tau_{i'}^X$, then the conditional Y distribution is stochastically larger in row i than in row i'. Also note that

$$L_{ij} - L_{i'j} = \log\left[\frac{(m_{i'1} + \cdots + m_{i'j})/(m_{i',j+1} + \cdots + m_{i'c})}{(m_{i1} + \cdots + m_{ij})/(m_{i,j+1} + \cdots + m_{ic})}\right] \quad (3.4)$$

is the log odds ratio for the 2×2 table formed by taking rows i and i' of the table and dichotomizing the response. The logit model (3.2) assumes that all $c - 1$ of these collapsings yield the same log odds ratio.

The independence model is the special case in which all $\tau_i^X = 0$; that is, each of the $c - 1$ logits takes on the same value for every row. Given that model (3.2) holds, its reduction in G^2 relative to the independence model gives a test of independence ($H_0: \tau_1^X = \ldots = \tau_r^X = 0$). As in the corresponding test for the loglinear model (2.3), the test statistic has asymptotically a chi-squared distribution with df $= r - 1$ under H_0.

We fit model (3.2) to the data in Table 2 on Y = political ideology and X = party affiliation, using methods described in Section 5. The second set of parenthesized values in Table 2 consists of maximum likelihood estimates of expected frequencies. For this fit, $G^2 = 4.70$ based on df = 2. The estimates of the average logits are $\hat{\mu}_1 = .532$ and $\hat{\mu}_2 = -1.325$ and the estimates of the effects of party affiliation are $\hat{\tau}_1^X = .670$, $\hat{\tau}_2^X = .282$, and $\hat{\tau}_3^X = -.952$.

The model predicts constant differences between pairs of rows in the two logits. The differences are also constant predicted log odds ratios for the two collapsings of each pair of rows into 2×2 tables. For example, $\hat{\tau}_1^X - \hat{\tau}_3^X = 1.622$ means that the odds of being classified liberal instead of moderate or conservative and the odds of being classified liberal or moderate instead of conservative are $\exp(1.622) = 5.06$ times higher for Democrats than for Republicans. This model gives strong evidence of an association, as the reduction in G^2 from the independence model is 100.96 based on df = 2.

The results obtained here are very similar to those obtained using the corresponding loglinear model (2.3). For that model with integer column scores, the differences in tau parameters reflect constant predicted log odds ratios for the $c - 1$ 2×2 tables formed using only adjacent columns. Figure 1 contrasts the two types of constant odds ratio upon which the loglinear and logit strategies are based. When both models fit reasonably well (as with Table 2), the $| \hat{\tau}_i^X - \hat{\tau}_{i'}^X |$ will tend to be smaller for the

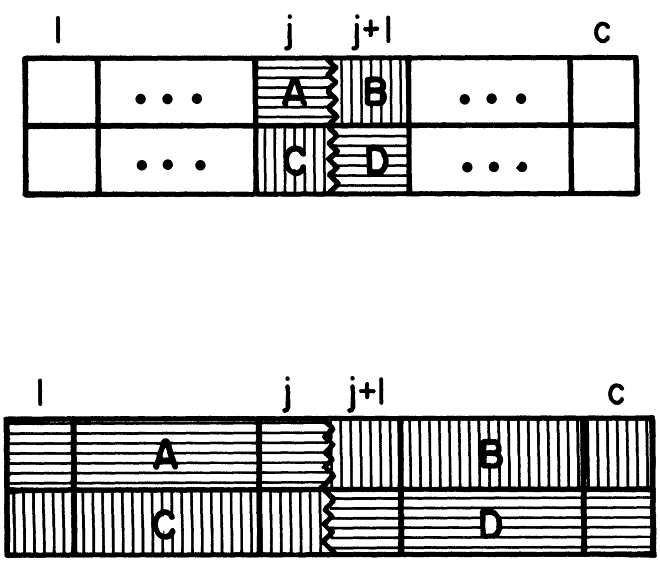

Figure 1. Odds ratios AD/BC that are constant for $c - 1$ cutpoints of column variables. Loglinear model (upper section): constant odds ratios for adjacent columns. Logit model (lower section): constant odds ratios for dichotomized response.

loglinear model since local associations will tend to be weaker than associations involving two aggregates of columns. For example, suppose that the underlying conditional density f_i of Y in row i is $N(\mu_i, \sigma^2)$. Then

$$| \log f_i(v_1) f_{i'}(v_2) / f_i(v_2) f_{i'}(v_1) | = \sigma^{-2} | (\mu_i - \mu_{i'})(v_1 - v_2) |$$

is a monotonic increasing function of $| v_1 - v_2 |$.

3.2 Logit Model for Ordinal-Ordinal Table

Suppose again that the column variable is an ordinal response, but that now the row variable is also ordinal. As with the loglinear model (2.6) for this setting, we assume that scores $\{u_i\}$ are assigned to the rows. It will not be necessary to assign scores to the levels of the response variable, since the accumulated logits (3.1) within each row again form the responses. A simple linear model for these logits is

$$L_{ij} = \mu_j + \beta^X (u_i - \bar{u}), \quad i = 1, \ldots, r, j = 1, \ldots, c - 1. \tag{3.5}$$

For this model df $= (r - 1)(c - 1) - 1$, the same as for the loglinear model (2.6).

As in model (3.2), μ_j represents the average across the r rows of the values of the jth accumulated logit. Each of the $(c - 1)$ logits is linearly related to the row variable, with slope β^X being the same for all logits. Hence for an arbitrary pair of rows $i < i'$, the difference in logits

$$L_{i'j} - L_{ij} = \beta^X (u_{i'} - u_i)$$

is proportional to the distance between the rows. If $\beta^X > 0$, the logit increases as X increases, which implies that the conditional Y distributions are stochastically larger at larger values of X. For integer row scores, β^X represents the constant value of the log odds ratio for the $(r - 1)(c - 1)$ 2×2 tables obtained by taking all pairs of adjacent rows and all dichotomous collapsings of the response. The independence model is the special case of model (3.5) in which $\beta^X = 0$.

We fit model (3.5) with integer row scores to the data in Table 4 on Y = dumping severity and X = operation. The second set of parenthesized values in Table 4 consists of maximum likelihood estimates of expected frequencies corresponding to that fit. The fit is very good, with $G^2 = 4.27$ based on df = 5. The estimates of the average logits are $\hat{\mu}_1 = -.320$ and $\hat{\mu}_2 = -2.074$, and the estimate of the linear effect of operation on the logit of dumping severity is $\hat{\beta}^X = .225$. Hence, the odds that dumping severity is above a certain point rather than below it are $\exp(.225) = 1.25$ times higher for operation $i + 1$ than for operation i, $i = 1, 2, 3$. This model also gives moderately strong evidence of an association, as the reduction in G^2 from the independence model is 6.61 based on df = 1.

Again the results and substantive interpretations that follow from this model agree with those made with the corresponding simple loglinear model (2.6). The estimated expected frequencies are very similar. For model (2.6), though, the association parameter estimate $\hat{\beta}^{XY} = .163$ reflects constant predicted log odds ratios for 2×2 tables for which the columns as well as the rows are adjacent. When the loglinear and logit models with integer scores fit reasonably well (as with Table 4), $|\hat{\beta}^{XY}|$ will tend to be smaller than $|\hat{\beta}^{X}|$ since β^{XY} refers to local association in both dimensions.

3.3 Logit Models for Multidimensional Table

Logit models for multidimensional tables resemble multiple regression models for quantitative response variables. To illustrate, we consider the three-dimensional table in which the layer variable Z is an ordinal response having l categories. Within each combination of X and Y, there are $(l - 1)$ accumulated logits

$$L_{ijk} = \log\left(\frac{m_{ij,k+1} + \cdots + m_{ijl}}{m_{ij1} + \cdots + m_{ijk}}\right), k = 1, \ldots, l - 1.$$

Table 8 lists simple models having linear effects of ordinal variables. The effects in these models are homogeneous across the $l - 1$ ways of forming the logits. The parameter interpretations are the partial analogs of those given in Sections 3.1 and 3.2. The residual df for testing the models are also given in Table 8. For these to be the same as for the corresponding loglinear models in Table 5, we must use general terms for partial associations among explanatory variables in the loglinear models. That is, the X-Y association term in the first two models in Table 5 must be replaced by $\lambda_{ij}{}^{XY}$. Interaction terms can be added to the models in Table 8 as in multiple regression, analysis of covariance, and analysis of variance models, respectively.

A notable feature of the logit models discussed thus far is the assumption that the effects of the explanatory variables are the same for the different ways of forming the accumulated logits. These models may be generalized to include nonhomogeneous logit effects. For example, the model for nominal variables X and Y may be generalized to

$$L_{ijk} = \mu_k + \tau_{ik}{}^X + \tau_{jk}{}^Y, \tag{3.6}$$

where $\sum_i \tau_{ik}{}^X = \sum_j \tau_{jk}{}^Y = 0$, $k = 1, \ldots, l - 1$, and where df $= (r - 1)(c - 1)(l - 1)$. The nice interpretation possessed by the simpler model (that the levels of X, or Y, are constant location shifts of each other on an accumulated logit scale) is lost with model (3.6). This model with nonstructured differences in logits resembles the general loglinear model (2.2) for no three-factor interaction, except for the asymmetry of identifying a response. For the two-dimensional table the nonhomogeneous logit model $L_{ij} = \mu_j + \tau_{ij}{}^X$ is a saturated model.

Table 8. Association Terms and DF for Logit Models in Three Dimensions with Ordinal Response Z and Linear Ordinal Effects.

Ordinal Variables	Association Terms			DF
	X − Y	X − Z	Y − Z	
X,Y,Z	—	$\beta^X(u_i - \bar{u})$	$\beta^Y(v_j - \bar{v})$	$rc\ell$-rc-ℓ-1
Y,Z	—	$\tau_i{}^X$	$\beta^Y(v_j - \bar{v})$	$rc\ell$-rc-ℓ-c + 1
Z	—	$\tau_i{}^X$	$\tau_j{}^Y$	$rc\ell$-rc-r-c-ℓ + 3

For Table 6 relating Z = rated performance of radio and TV networks to X = race and Y = year, we fit the first logit model in Table 8 (which is equivalent to the other models in the table since $r = c = 2$) using integer scores. The model fits reasonably well, with $G^2 = 3.66$ based on df = 4. The estimated effect parameters of $\hat{\beta}^X = .669$ and $\hat{\beta}^Y = -.396$ reflect the tendency of rated performance to be higher for blacks than for whites and lower in 1971 than in 1959. Again the results are similar to those obtained with the corresponding loglinear model (2.10).

4. ALTERNATIVE MODELS

The loglinear and logit models are only two of many types of models that have been suggested for ordinal variables. In this section we briefly describe several alternative types of models.

4.1 Mean Response

Many researchers do not find parameters based on odds measures or log transformations to be as easily interpretable as those in standard regression models. If we regard the process of ranking observations on an ordinal scale as being more similar to the measurement process with interval scales than with nominal scales, then we can argue that models for ordinal response variables should resemble regression models for interval response variables more than loglinear and logit models for nominal variables. Indeed, ordinal variables are treated like interval variables once we assign scores to their levels in the models of Sections 2 and 3. To obtain a regression-type model having easily interpretable parameters, we can model the conditional mean of the response variable.

To illustrate, consider the two-way cross-classification of ordinal variables X and Y having scores $\{u_i\}$ and $\{v_j\}$. Within level i of X, the conditional mean of Y is $M_i = \sum_j v_j m_{ij}/n_{i+}$, $i = 1, \ldots, r$. The usual linear regression model is

$$M_i = \mu + \beta^X(u_i - \bar{u}), \quad i = 1, \ldots, r. \tag{4.1}$$

The parameter μ is the average of the conditional means, and β^X is the change in the conditional mean per unit change in X.

Williams and Grizzle (1972) (see also Bhapkar 1968 and Grizzle, Starmer, and Koch 1969) used the weighted least squares method to fit this type of response function. For model (4.1) there are r responses and 2 parameters, so df $= r - 2$ and we need $r \geq 3$ to obtain an unsaturated model. Corresponding models for multidimensional tables are equally easy to construct and interpret. The model for a single nominal explanatory variable is saturated, though.

We fit model (4.1) to the data in Table 4 on Y = dumping severity and X = operation, obtaining $\hat{\mu} = 1.537$ and $\hat{\beta}^X = .075$ for a weighted least squares solution when integer scores are used. The predicted increase in the mean dumping severity is .075 categories for every additional 25 percent of stomach removal. The test of $H_0: \beta^X = 0$ yields a chi-squared statistic of 6.37 based on df = 1. The model fits adequately, as evidenced by a residual chi-squared of .23 based on df = 2. Model (4.1), like the loglinear and logit models fit earlier to these data, gives evidence of a relatively weak positive association between operation and dumping severity.

4.2 Alternate Response Functions

Models like (4.1) can be formulated just as simply with response functions other than the mean. For example, let Y^* denote the response on Y for a randomly selected person from the combined population, and let Y_i denote the response for a randomly selected person in level i of X. The response $P(Y_i > Y^*) + \frac{1}{2}P(Y_i = Y^*)$, which is equivalently a mean ridit score for Y in level i of X, is not dependent on assigning scores to the levels of Y (see Bross 1958). A weighted least squares analysis for this response function was given by Semenya and Koch (1979).

Suitable response functions may sometimes be suggested by the form of an assumed underlying response distribution. For example, let $\{F_{ij}, j = 1, \ldots, c\}$ denote the distribution function of Y within level i of X. When it is reasonable to assume an exponential-type underlying distribution for Y, or an underlying distribution of the type used in survival analysis (see Cox 1972), we expect log $(1 - F_{ij})$ to be approximately the same constant multiple of log $(1 - F_{i'j})$ for $1 \leq j \leq c - 1$. We might therefore pose the model with a constant difference between levels of X on the log-log scale of the complement of the distribution function. McCullagh (1979,1980) suggested the response function $\log[-\log(1 - F_{ij})]$ in addition to the logit model (3.5), and he argued that it would be appropriate for a wide class of underlying distributions that includes the Pareto and Weibull.

Yet other approaches have been suggested for ordinal response variables. McKelvey and Zavoina (1975) assumed an underlying normal distribution and generalized the probit model. Hawkes (1971) and Ploch (1974) suggested a linear model for sign scores for differences between pairs of observations. Schollenberger, Agresti, and Wackerly (1979) suggested a logit model for the probability of concordance for pairs of observations. For the case of two ordinal variables, Clayton (1974) and Wahrendorf (1980) estimated an assumed common odds ratio for all $(r - 1)(c - 1)$ collapsings of X and of Y into a 2×2 table. Their approach assumes that the joint distribution is contained in a class proposed by Plackett (1965). Goodman (1981b) gave examples for which this assumption does not seem as appropriate as the assumption of constant local association that is inherent with the uniform association loglinear model. Semenya and Koch (1980, pp. 103–118) defined more general models in terms of the odds ratios for the collapsed tables.

Many specialized models have also been proposed for square tables with ordered categories. These models are most naturally applied when the two sets of categories are identical, as occurs in matched-pairs experiments. Discussions of these models are presented by Bishop, Fienberg, and Holland (1975, pp. 285–286), Goodman (1972,1979a,b), and McCullagh (1977,1978).

5. ESTIMATION

Maximum likelihood (ML) and weighted least squares (WLS) methods can be used to fit the models introduced in Sections 2 through 4. Computer packages and programs are available for obtaining ML and WLS estimates for most of these models.

5.1 Maximum Likelihood

ML estimates for loglinear models are identical for the three standard sampling schemes, provided that the model fits fixed marginal counts. If the estimates are finite, they are unique, which is always the case when none of the cell counts equals zero. The likelihood equations imply that certain functions of expected cell frequencies are equated to sufficient statistics that are the same functions of observed cell counts. For ordinal loglinear models, the sufficient statistics include sample conditional or joint moments and low-order marginal distributions.

To illustrate, for independent Poisson sampling in the $r \times c$ table, the kernel of the log-likelihood function is $\sum n_{ij} \log m_{ij} - \sum m_{ij}$. For the row-effects model (2.3), the likelihood equations are

$$\hat{m}_{i+} - n_{i+} = 0, \qquad i = 1, \ldots, r,$$

$$\hat{m}_{+j} - n_{+j} = 0, \qquad j = 1, \ldots, c,$$

$$\sum_j \hat{m}_{ij} v_j - \sum_j n_{ij} v_j = 0, \quad i = 1, \ldots, r.$$

For the linear-by-linear association model (2.6), the third set of equations is replaced by the single equation

$$\sum \hat{m}_{ij} u_i v_j - \sum n_{ij} u_i v_j = 0.$$

The likelihood equations can be solved using various types of iterative methods. Under the Poisson sampling model, loglinear models are special cases of Nelder and Wedderburn's (1972) generalized linear models. That is, the cell counts are independently distributed according to a member (Poisson) of the exponential family, and the natural parameter (log m) is linked directly to independent variables through a linear model. The computer package GLIM (Baker and Nelder (1978)) is designed for fitting generalized linear models, and we have found it very simple to apply to the ordinal models of Section 2. GLIM uses the Newton-Raphson procedure, which corresponds to iterative calculations of a weighted least squares form. The asymptotic covariance matrix of the model parameter estimates is produced as a by-product of this process. Haberman (1979) contains a Fortran program (FREQ) that also fits loglinear models using the Newton-Raphson method and that reports adjusted cell residuals; page 377 gives Haberman's formulation of the row effects model and page 385 the linear-by-linear association model.

An iterative scaling method for fitting ordinal loglinear models follows from Theorem 1 of Darroch and Ratcliff (1972). Simon (1974) and Fienberg (1980, pp. 61–64) apply it to the row-effects model (2.3). In that case, a single cycle has three steps. The approximation $m_{ij}^{(t)}$ for $\hat{m}_{ij}$ at the tth stage is multiplied by $n_{i+}/m_{i+}^{(t)}$ to obtain the next approximation, $m_{ij}^{(t+1)}$. The $\{m_{ij}^{(t+1)}\}$ satisfy $m_{i+}^{(t+1)} = n_{i+}$, $i = 1, \ldots, r$. Next, $m_{ij}^{(t+1)}$ is multiplied by $n_{+j}/m_{+j}^{(t+1)}$ for all i and j to obtain values $\{m_{ij}^{(t+2)}\}$ that satisfy $m_{+j}^{(t+2)} = n_{+j}$, $j = 1, \ldots, c$. At the third step, $m_{ij}^{(t+3)}$ equals $m_{ij}^{(t+2)}$ multiplied by

$$[\sum_k v_k^* n_{ik} / \sum_k v_k^* m_{ik}^{(t+2)}]^{v_j^*} \\ [\sum_k (1 - v_k^*) n_{ik} / \sum_k (1 - v_k^*) m_{ik}^{(t+2)}]^{1 - v_j^*},$$

where the $\{v_j^*\}$ are a linear rescaling of the column scores $\{v_j\}$ that satisfy $0 \le v_j^* \le 1$. For the linear-by-linear association model (2.6), the row scores $\{u_i\}$ are also rescaled to satisfy $0 \le u_i^* \le 1$, and the multiplicative factor in the third step is

$$[\sum_{k,l} u_k^* v_l^* n_{kl} / \sum_{k,l} u_k^* v_l^* m_{kl}^{(t+2)}]^{u_i^* v_j^*} \\ \times \left[\frac{\sum_{k,l} (1 - u_k^* v_l^*) n_{kl}}{\sum_{k,l} (1 - u_k^* v_l^*) m_{kl}^{(t+2)}} \right]^{1 - u_i^* v_j^*}.$$

Duncan and McRae (1979) show how to fit ordinal loglinear models through iterative use of the computer program ECTA of Fay and Goodman (1975), which is based on iterative proportional fitting. Goodman (1979a) and Clogg (1982) describe another iterative approach in which each step is a unidimensional application of the Newton approximation method. The calculations in these methods and in the iterative scaling method just described are very simple compared with those in the Newton-Raphson method, since no matrix inversion is necessary. However, a very large number of cycles may be needed for adequate convergence, and a separate inversion of the negative of the information matrix is required to obtain the asymptotic covariance matrix of the model parameter estimates. Let $\mathbf{m}$ be a vector having the expected frequencies as elements, let $\boldsymbol{\beta}$ be a vector having the model parameters as elements, and let $\mathbf{X}$ be the design matrix for the formulation of the model as

$$\log \mathbf{m} = \mathbf{X}\boldsymbol{\beta}$$

with the sum of the expected frequencies equal to n. Then for full multinomial sampling, the estimated information matrix is

$$-\mathbf{X}'[\mathbf{D}_{\hat{\mathbf{m}}} - \hat{\mathbf{m}}\hat{\mathbf{m}}'/n]\mathbf{X},$$

where $\mathbf{D}_{\hat{\mathbf{m}}}$ is a diagonal matrix with the expected frequency estimates on the main diagonal.

If one set of parameter scores is known, then the multiplicative model (2.8) becomes loglinear. ML estimates for multiplicative models can be obtained through repeated application of any of the above procedures, in which alternate sets of scores are fixed. For model (2.8), in each cycle one treats the column scores as known and estimates the row scores (as in a row-effects model), then one treats those estimates as fixed and estimates the column scores.

McCullagh (1980) shows that the accumulated logit models comprise a multivariate analog of the generalized linear model. In the appendix to his paper he shows how to obtain ML estimates for a class of models for monotonic transformations of the distribution function of the response. He reports fast convergence using the Newton-Raphson method, even for poor initial estimates. See also Simon (1974) and Bock (1975, pp. 544–546) for Newton-Raphson approaches for ML fitting of the accumulated logit models. The computer package MULTIQUAL (Bock and Yates 1973) can be used to fit the accumulated logit models.

5.2 Weighted Least Squares

The logit models, loglinear models, and alternative response models can be fit simply using the WLS approach for categorical data as described by Grizzle, Starmer, and Koch (1969). The ordinal loglinear and logit models can be expressed in the form

$$\mathbf{K} \log (\mathbf{Am}) = \mathbf{X}\boldsymbol{\beta}.$$

To illustrate, for the accumulated logit model (3.2) applied to Table 2,

$$\boldsymbol{\beta}' = (\mu_1, \mu_2, \tau_1^X, \tau_2^X), \quad \mathbf{m}' = (m_{11}, m_{12}, \ldots, m_{33}),$$

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 \\ 0 & 1 & -1 & -1 \end{pmatrix}$$

$\mathbf{K} = \mathbf{I} \otimes \mathbf{K}^*$ and $\mathbf{A} = \mathbf{I} \otimes \mathbf{A}^*$, where

$$\mathbf{A}^* = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{K}^* = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 1 \end{pmatrix},$$

and $\mathbf{I}$ is the 3×3 identity matrix. For loglinear models $\mathbf{K} \log(\mathbf{Am})$ can be taken as the vector of local log odds ratios.

Let $\mathbf{V}$ denote the sample covariance matrix of the observed data, **n**. For example, suppose that independent multinomial samples of sizes $\{n_{1+}, \ldots, n_{r+}\}$ are selected at r combinations of levels of explanatory variables. Then, $\mathbf{V}$ is a block diagonal matrix whose ith block $\mathbf{V}_i$ satisfies

$$n_{i+}\mathbf{V}_i = \begin{pmatrix} n_{i1}(n_{i+} - n_{i1}) & -n_{i1}n_{i2} & \ldots & -n_{i1}n_{ic} \\ -n_{i2}n_{i1} & n_{i2}(n_{i+} - n_{i2}) & \ldots & -n_{i2}n_{ic} \\ \vdots & \vdots & & \vdots \\ -n_{ic}n_{i1} & -n_{ic}n_{i2} & \ldots & n_{ic}(n_{i+} - n_{ic}) \end{pmatrix}.$$

By the delta method, the estimated asymptotic covariance matrix of the vector of sample responses $\mathbf{F} = \mathbf{K} \log(\mathbf{An})$ is

$$\mathbf{S} = (\mathbf{KD}^{-1}\mathbf{A})\mathbf{V}(\mathbf{KD}^{-1}\mathbf{A})',$$

where $\mathbf{D}$ is a diagonal matrix with element d_{ii} equal to the ith row of $\mathbf{A}$ multiplied by $\mathbf{n}$. The WLS estimate of $\boldsymbol{\beta}$ is $\mathbf{b} = (\mathbf{X}'\mathbf{S}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{S}^{-1}\mathbf{F}$, its estimated covariance matrix is $(\mathbf{X}'\mathbf{S}^{-1}\mathbf{X})^{-1}$, and the goodness-of-fit statistic is $(\mathbf{F} - \mathbf{Xb})'\mathbf{S}^{-1}(\mathbf{F} - \mathbf{Xb})$.

The WLS approach is especially useful for models, such as the mean response model, for which the ML approach cannot be easily applied using existing computer packages. It has the advantage (relative to ML) of not requiring iteration, but this is of minor importance as efficient computer programs become more widely available for the ML approach. The computer program GENCAT (Landis et al. 1976) can be used to provide WLS fits for the models considered in this article. See Williams and Grizzle (1972) and Semenya and Koch (1980) for further details about WLS estimation for ordinal models.

6. COMPARISON OF MODELS

The major theme of this article has been the importance of using the quantitative nature of ordinal variables in analyzing categorical data. When doing so, (a) we have available a greater variety of models, most of which are more parsimonious and have simpler parameter interpretations than do the standard models for nominal variables, (b) we can describe the data using parameters that are similar to those used in ordinary regression for continuous variables (e.g., correlations, slopes), and (c) we obtain greater power for detecting important alternatives to null hypotheses of independence, conditional independence, or no interaction. For the data sets analyzed in this article, we have obtained similar substantive results whether the ordinal method being used is a loglinear model, a logit model, or a model having an alternative type of response function. However, some distinctions can be made that may help us in selecting a strategy for analyzing ordinal data.

6.1 Comparison of Responses

The mean response models have the advantage of closely resembling regression models for continuous response variables. Fitting them is a reasonable strategy if the categorical nature of the response variable is due to crude measurement of inherently continuous variables.

For the loglinear, logit, and log-log models, cell probabilities are determined by the model parameters. This is not the case for the mean (or mean ridit) models. Hence, we cannot easily use these models to make conclusions about structural aspects such as stochastic orderings on the response. If model (4.1) holds, for example, $\beta^X = 0$ is not equivalent to independence. The loglinear and logit models directly reflect the actual discrete way we have measured the variables, and special cases of those models correspond to conditions such as independence.

6.2 Logit vs. Loglinear Models

When there are only two response categories, the ordinal logit models are special cases of corresponding ordinal loglinear models. The association parameters in the logit models then equal those in the corresponding loglinear models, when the two response scores are one unit apart in the loglinear model. This equivalence does not occur when the number of response categories exceeds two, and neither model is then a special case of the other.

Table 9 summarizes the results of fitting the two types of models to Tables 2, 4, and 6. The similarity of results observed here will probably occur often in practice, since both model types imply that different levels of variables are stochastically ordered on ordinal response variables. On structural grounds, neither model is clearly preferable to the other, and the implications for odds ratio behavior

Table 9. Analyses of Association for Loglinear and Logit Models Fit to Tables 2, 4, and 6

Table	Model	DF	G^2	Association Parameter Estimates (standard errors in parentheses)
2	Independence	4	105.66	zero
	Loglinear	2	2.83	$\hat{\tau}_1^X = .495(.062)$, $\hat{\tau}_2^X = .224(.059)$, $\hat{\tau}_3^X = -.719(.080)$
	Logit	2	4.70	$\hat{\tau}_1^X = .670(.083)$, $\hat{\tau}_2^X = .282(.079)$, $\hat{\tau}_3^X = -.952(.102)$
4	Independence	6	10.88	zero
	Loglinear	5	4.59	$\hat{\beta}^{XY} = .163(.065)$
	Logit	5	4.27	$\hat{\beta}^X = .225(.088)$
6	Independence	7	84.37	zero
	Loglinear	4	5.58	$\hat{\beta}^{XZ} = .543(.084)$, $\hat{\beta}^{YZ} = -.307(.070)$
	Logit	4	3.66	$\hat{\beta}^X = .669(.101)$, $\hat{\beta}^Y = -.396(.086)$

seem theoretically reasonable for both model types. More precisely, just as the logit model will tend to fit well if the response variable has an underlying logistic distribution, so the loglinear models will tend to fit well if underlying distributions are approximately normal and if appropriate scores are used. For example, if an underlying joint distribution is bivariate normal with density f, then

$$\log[f(u_1, v_1)f(u_2, v_2)/f(u_1, v_2)f(u_2, v_1)] = (u_2 - u_1)(v_2 - v_1)\rho/(1 - \rho^2)\sigma_X\sigma_Y$$

is analogous to property (2.7) for loglinear model (2.6). See Goodman (1981b) for an example of the excellent fit that loglinear model (2.6) can give to a bivariate normal distribution.

McCullagh (1980, pp. 121–122) argues that the logit model is preferable to the loglinear model because with the logit model it is easier to state conclusions about association parameters without reference to the groupings of response categories. If the logit model holds for the true cell proportions, then if response categories are combined it will still hold and the association parameters will not change values. If a loglinear model holds for the true cell proportions, on the other hand, it may not hold when categories of any variable are combined. Even if it does hold, the association parameters may not be of the same order of magnitude.

In practice, this behavior is probably not a serious deficiency of the loglinear model, since (a) the real underlying distributions are unlikely to be such that *either* model will fit for all categorizations, and (b) scoring systems such as the standardized scores can be used to make the association parameters in the loglinear model less dependent on the choice of categories. To illustrate, suppose that the "slight" and "moderate" categories of dumping severity are combined in Table 4, so that dumping is measured as "none" or "some." For the resulting 4×2 table, the ML estimate of the association parameter in loglinear model (2.6) and logit model (3.5) (which are identical since $c = 2$) is .229. As expected, this is similar to the value $\hat{\beta}^X = .225$ for the logit model for the original table, but it is not similar to $\hat{\beta}^{XY} = .163$ for the loglinear model for Table 4. If we use standardized scores in the loglinear model, however, we get $\hat{\beta}^{XY} = .124$ for Table 4 and $\hat{\beta}^{XY} = .125$ for the collapsed table.

It is advantageous that the logit model does not require the assignment of scores to levels of a response variable that is inherently ordinal in scale. Even with the logit model, though, scores must be assigned to levels of ordinal explanatory variables. For both model types, however, it is not really necessary to treat these scores as reasonable scalings of the ordinal variables in order for the models to be valid. For example, consider loglinear models (2.6) and (2.8). The row and column scores indicate how far apart the rows and columns must be judged to be in order for the association to be linear-by-linear. If the model fits when a particular pair of row scores are relatively close, this tells us that the association is relatively weaker in that part of the table. This information is useful even if we do not truly believe that those two rows would be close together for an underlying interval-scale measurement of that variable. In fact, such an underlying scale has monotone scores, whereas score estimates for model (2.8) will tend to be nonmonotonic when associations change directions over various parts of the table. Thus, we can use such scores to provide information about the nature of the association without needing to regard them as indices of how far apart the ordered levels truly are.

In summary, both the logit and loglinear models provide sensible strategies for analyzing ordinal categorical data. The main issue that dictates the choice of one model over the other is whether it is important to identify a response variable. As can be seen by comparing Tables 5 and 8, for multidimensional tables it is easier to formulate the logit model than the loglinear model, since association patterns among the explanatory variables do not have to be considered.

[*Received February 1981. Revised September 1982.*]

REFERENCES

AGRESTI, ALAN, and KEZOUH, ABBAS (1983), "Association Models for Multidimensional Cross-Classifications of Ordinal Variables," *Communications in Statistics*, to appear.

ANDERSEN, E.B. (1980), *Discrete Statistical Models with Social Science Applications*, Amsterdam: North-Holland.

ANDRICH, DAVID (1979), "A Model for Contingency Tables Having an Ordered Response Classification," *Biometrics*, 35, 403–415.

BAKER, R.J., and NELDER, J.A. (1978), *The GLIM System. Release 3. Generalised Linear Interactive Modelling Manual*, Oxford: NAG.

BHAPKAR, V.P. (1968), "On the Analysis of Contingency Tables With a Quantitative Response," *Biometrics*, 24, 329–338.

BIRCH, M.W. (1965), "The Detection of Partial Association, II: The General Case," *Journal of the Royal Statistical Society*, Ser. B, 27, 111–124.

BISHOP, Y.M.M., FIENBERG, S.E., and HOLLAND, P.W. (1975), *Discrete Multivariate Analysis*, Cambridge, Mass.: MIT Press.

BOCK, R. DARRELL (1975), *Multivariate Statistical Methods in Behavioral Research*, New York: McGraw-Hill.

BOCK, R.D., and YATES, G. (1973), "Log-linear Analysis of Nominal or Ordinal Qualitative Data by the Method of Maximum Likelihood," Chicago: National Education Resources.

BROSS, IRWIN D.J. (1958), "How to Use Ridit Analysis," *Biometrics*, 14, 18–38.

CLAYTON, D.G. (1974), "Some Odds Ratio Statistics for the Analysis of Ordered Categorical Data," *Biometrika*, 61, 525–531.

CLOGG, CLIFFORD C. (1982), "Some Models for the Analysis of Association in Multi-Way Cross-Classifications Having Ordered Categories," *Journal of the American Statistical Association*, 77, 803–815.

COX, DAVID R. (1970), *The Analysis of Binary Data*, London: Chapman and Hall.

——— (1972), "Regression Models and Life Tables (with Discussion)," *Journal of the Royal Statistical Society*, Ser. B, 34, 187–220.

DARROCH, J.N., and RATCLIFF, D. (1972), "Generalized Iterative Scaling for Log-Linear Models," *The Annals of Mathematical Statistics*, 43, 1470–1480.

DUNCAN, OTIS D. (1979), "How Destination Depends on Origin in the Occupational Mobility Table," *American Journal of Sociology*, 84, 793–803.

DUNCAN, OTIS D., and McRAE, JAMES A., JR. (1979), "Multiway Contingency Analysis With a Scaled Response or Factor," *Sociological Methodology 1979*, ed. Karl F. Schuessler, San Francisco: Jossey-Bass.

FAY, ROBERT, and GOODMAN, LEO A. (1975), "ECTA Program:

Description for Users," University of Chicago, Dept. of Statistics, Chicago, Illinois.
FIENBERG, STEPHEN (1980), *The Analysis of Cross-Classified Categorical Data* (2nd ed.), Cambridge, Mass.: The MIT Press.
GOODMAN, LEO A. (1972), "Some Multiplicative Models for the Analysis of Cross-Classified Data," *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1), eds. L. LeCam et al., Berkeley: University of California Press, 649-696.
——— (1979a), "Simple Models for the Analysis of Association in Cross-Classifications Having Ordered Categories," *Journal of the American Statistical Association*, 74, 537–552.
——— (1979b), "Multiplicative Models for Square Contingency Tables With Ordered Categories," *Biometrika*, 66, 413–418.
——— (1981a), "Association Models and Canonical Correlation in the Analysis of Cross-Classifications Having Ordered Categories," *Journal of the American Statistical Association*, 76, 320–334.
——— (1981b), "Association Models and the Bivariate Normal Distribution in the Analysis of Cross-Classifications Having Ordered Categories," *Biometrika*, 68, 347–355.
——— (1981c), "Three Elementary Views of Loglinear Models for the Analysis of Cross-Classifications Having Ordered Categories," in *Sociological Methodology 1981*, ed. Samuel Leinhardt, San Francisco: Jossey-Bass Publishers, 193–239.
GRIZZLE, J.E., STARMER, C.F., and KOCH, G.G. (1969), "Analysis of Categorical Data by Linear Models," *Biometrics*, 25, 489–504.
HABERMAN, SHELBY J. (1974), "Loglinear Models for Frequency Tables With Ordered Classifications," *Biometrics*, 30, 589–600.
——— (1979), *Analysis of Qualitative Data, Volume 2: New Developments*, New York: Academic Press.
——— (1981), "Tests for Independence in Two-Way Contingency Tables Based on Canonical Correlation and on Linear-By-Linear Interaction," *The Annals of Statistics*, 9, 1178–1186.
HAWKES, ROLAND K. (1971), "The Multivariate Analysis of Ordinal Measures," *American Journal of Sociology*, 76, 908–926.
HEDLUND, RONALD D. (1978), "Cross-over Voting in a 1976 Open Presidential Primary," *Public Opinion Quarterly*, 41, 498–514.
LANDIS, J.R., STANISH, W.M., FREEMAN, J.L., and KOCH, G.G. (1976), "A Computer Program for the Generalized Chi-Square Analysis of Categorical Data Using Weighted Least Squares," *Computer Programs in Biomedicine*, 6, 196–231.
McCULLAGH, PETER (1977), "A Logistic Model for Paired Comparisons With Ordered Categorical Data," *Biometrika*, 64, 449–453.
——— (1978), "A Class of Parametric Models for the Analysis of Square Contingency Tables with Ordered Categories," *Biometrika*, 65, 413–418.
——— (1979), "The Use of the Logistic Function in the Analysis of Ordinal Data," *Bulletin of the International Statistical Institute*.
——— (1980), "Regression Models for Ordinal Data," *Journal of the Royal Statistical Society*, Ser. B, 42, 109–142.
McKELVEY, RICHARD D., and ZAVOINA, WILLIAM (1975), "A Statistical Model for the Analysis of Ordinal Level Dependent Variables," *Journal of Mathematical Sociology*, 4, 103–120.
NELDER, J.A., and WEDDERBURN, R.W.M. (1972), "Generalized Linear Models," *Journal of the Royal Statistical Society*, Ser. A, 135, 370–384.
PLACKETT, ROBIN L. (1965), "A Class of Bivariate Distributions," *Journal of the American Statistical Association*, 60, 516–522.
PLOCH, DONALD R. (1974), "Ordinal Measures of Association and the General Linear Model," in *Measurement in the Social Sciences*, Chapter 12, ed. Hubert M. Blalock, Chicago: Aldine.
SCHOLLENBERGER, JOHN, AGRESTI, ALAN, and WACKERLY, DENNIS (1979), "Measuring Association and Modelling Relationships Between Interval and Ordinal Variables," *Proceedings of the Social Statistics Section, American Statistical Association*, 624–626.
SEMENYA, KOFI, and KOCH, GARY G. (1979), "Linear Models Analysis for Rank Functions of Ordinal Categorical Data," *Proceedings of the Statistical Computing Section, American Statistical Association*, 271–276.
——— (1980), "Compound Function and Linear Model Methods for the Multivariate Analysis of Ordinal Categorical Data," University of North Carolina Institute of Statistics Mimeo Series No. 1323.
SIMON, GARY (1974), "Alternative Analyses for the Singly-Ordered Contingency Table," *Journal of the American Statistical Association*, 69, 971–976.
SNELL, E.J. (1964), "A Scaling Procedure for Ordered Categorical Data," *Biometrics*, 20, 592–607.
WAHRENDORF, JÜRGEN (1980), "Inference in Contingency Tables With Ordered Categories Using Plackett's Coefficient of Association for Bivariate Distributions," *Biometrika*, 67, 15–21.
WILLIAMS, O. DALE, and GRIZZLE, JAMES E. (1972), "Analysis of Contingency Tables Having Ordered Response Categories," *Journal of the American Statistical Association*, 67, 55–63.
YATES, FRANK (1948), "The Analysis of Contingency Tables With Groupings Based on Quantitative Characteristics," *Biometrika*, 35, 176–181.