A Non-Iterative Alternative to Ordinal Log-Linear Models

ERIC J. BEH[†]   e.beh@uws.edu.au
School of Quantitative Methods and Mathematical Sciences,
University of Western Sydney, Australia

PAMELA J. DAVY   pam_davy@uow.edu.au
School of Mathematics and Applied Statistics,
University of Wollongong, Australia

Abstract. Log-linear modeling is a popular statistical tool for analysing a contingency table. This presentation focuses on an alternative approach to modeling ordinal categorical data. The technique, based on orthogonal polynomials, provides a much simpler method of model fitting than the conventional approach of maximum likelihood estimation, as it does not require iterative calculations nor the fitting and re-fitting to search for the best model. Another advantage is that quadratic and higher order effects can readily be included, in contrast to conventional log-linear models which incorporate linear terms only.

The focus of the discussion is the application of the new parameter estimation technique to multi-way contingency tables with at least one ordered variable. This will also be done by considering singly and doubly ordered two-way contingency tables. It will be shown by example that the resulting parameter estimates are numerically similar to corresponding maximum likelihood estimates for ordinal log-linear models.

Keywords: Contingency Tables; Ordinal Variables; Orthogonal Polynomials; Scores.

1.   Introduction

Log-linear modeling is a popular statistical tool for the analysis of a contingency table and is especially popular in English speaking countries; see Gower [18] and van der Heijden and de Leeuw [27]. From the analysis of data using log-linear models, the researcher can determine important associations that exist in the data. The conventional method of estimating parameters from a log-linear model has been to use the maximum likelihood estimation technique, such as the Newton-Raphson iterative procedure or iterative proportional scaling (see Agresti [1]). The advantage of the technique lies in the ability to identify multi-level associations in multi-way contingency tables. However, a problem with this approach, as

---

† Requests for reprints should be sent to Eric J. Beh, School of Quantitative Methods and Mathematical Sciences, University of Western Sydney, Australia.

Fienberg [13, p47] points out, is that while there are several authors who have proposed techniques for selecting the optimum log-linear model:

Unfortunately, there is no all-purpose, best method of model selection.

This is because the selection of an optimum log-linear model requires a trial and error approach of fitting and re-fitting a model until optimisation is reached. For very large multi-way data sets, the choice of model can often be very arduous and time consuming for the researcher. The comment of Fienberg's [13] is echoed in Dillon and Goldstein [11, p326] who say:

Unfortunately, no "best" selection strategy exists, since a choice depends in large part on our *a priori* knowledge of the interrelationships amongst the variables. In addition, without some means of effecting a "guided" search, an extremely large number of models may need to be evaluated with no assurance that the selections strategy will yield the best model.

This paper presents a method which relies on orthogonal polynomials based on those in Emerson [12] and described in Beh [4] and Rayner and Best [24]. We review the alternative approach introduced in Beh and Davy [6, 7] which is applicable for ordinal multi-way contingency tables. Parameter estimation using the technique described in this paper is also discussed in Beh [5] where the Pearson chi-squared statistic for a singly ordered two-way contingency table is partitioned using orthogonal polynomials for the ordered variable and generalised basic vectors for the non-ordered variable. The parameter estimation technique described in the articles of Beh and Davy, and further elaborated on in this paper, show that the log-linear model technique commonly employed in the analysis of categorical data can be replaced by the new family of models discussed in this paper.

In this paper we discuss and extend the Beh and Davy methodology in the following seven sections. Section 2 defines the notation for two-way and three-way contingency tables. Section 3 defines the orthogonal polynomials that are used for an ordered variable, while Section 4 defines the association models for a two-way table which form the foundation of our method of parameter estimation. A doubly ordered two-way contingency table is discussed where the row and column variables are of an ordinal nature. A singly ordered table is also considered where only the set of column categories is ordinal. Similar results can be established for the analysis of a two-way contingency table with ordered row categories. These association models are extended to three-way contingency tables in Section 5, where

completely ordered tables are considered. Doubly ordered tables are also investigated where only the row and column variables are ordered. Similar results can be obtained for any combination of two ordered variables. Finally a singly ordered three-way contingency table is considered, where the ordered categories are associated with the row responses. Similar results can be obtained for the ordered column or ordered tube categories. The alternative procedure to parameter estimation is made in Sections 6 and 7 which are applicable to two-way and three-way contingency tables respectively and uses the association models described in Sections 4 and 5. The calculation of the orthogonal polynomials used in this paper require a set of scores that reflect the ordinal ordinal structure of the ordered variable. In many cases natural scores, i.e. 1, 2, ..., are the easiest set of scores to use and give results that compare very well with those calculated using more computationally intensive scores.

A comparison of results using the technique discussed in this paper with those obtained using the conventional log-linear analysis approach is made in Section 8. Two examples are presented - a doubly ordered two-way contingency table and a completely ordered three-way table. It is shown through these examples that the non-iterative parameter estimation approach is an easier and more informative procedure to implement than the conventional maximum likelihood approach.

2. Notation

Consider an $I \times J$ two-way contingency table, N, where the $(i, j)'th$ cell entry is given by n_{ij} for $i = 1, 2, \ldots, I$ and $j = 1, 2, \ldots, J$. It is assumed that such a data set is randomly selected from a population. As the focus of this paper is to consider an analysis of ordinal categorical data, we will be considering situations when N has ordered rows and/or ordered columns. Let the grand total of N be n and the probability matrix be P so that the $(i, j)'th$ cell entry is $p_{ij} = n_{ij}/n$ and $\sum_{i=1}^{I} \sum_{j=1}^{J} p_{ij} = 1$. Define the $i'th$ row marginal proportion as $p_{i\bullet} = \sum_{j=1}^{J} p_{ij}$ and define the $j'th$ column marginal probability as $p_{\bullet j} = \sum_{i=1}^{I} p_{ij}$ so that $\sum_{i=1}^{I} p_{i\bullet} = \sum_{j=1}^{J} p_{\bullet j} = 1$.

For the analysis of multi-level data, suppose we have a three-way contingency table, N, consisting of I rows, J columns and, using the terminology of Kroonenberg [22], K tubes. Kendall and Stuart [21] used the term *layer*.

Let the grand total of N be n where the $(i, j, k)'th$ element is denoted as n_{ijk}. We denote the same element of the matrix P by $p_{ijk} = n_{ijk}/n$. The $i'th$ row, $j'th$ column and $k'th$ tube row marginal proportions are defined as $p_{i \bullet \bullet}$, $p_{\bullet j \bullet}$ and $p_{\bullet \bullet k}$ respectively.

The distinction between population (parameter) values and their estimated (sample) values is made in the conventional way by the placement of a $\wedge$ over the estimated sample value, while the parameter has no $\wedge$ included. For example, the parameter value μ is estimated by the sample value $\hat{\mu}$.

3. Orthogonal Polynomials

For the parameter estimation technique discussed in this paper, we make use of orthogonal polynomials based on the ordinal nature of the categories within a variable. The polynomials used as those considered by Beh [3, 4], Beh and Davy [6, 7], and Rayner and Best [24].

For ordered column categories of a two-way contingency table, say, let $\{b_v(j)\}$ be the set of orthogonal polynomials on $\{p_{\bullet j}\}$, for $v = 1, 2, \ldots, (J-1)$ and $j = 1, 2, \ldots, J$. These polynomials have the property

$$\sum_{j=1}^{J} p_{\bullet j} b_v(j) b_{v'}(j) = \begin{cases} 1, v = v' \\ 0, v \neq v' \end{cases}$$

and are defined so that

$$b_v(j) = A_v^c \left[(s_J(j) - B_v^c) b_{v-1}(j) - C_v^c b_{v-2}(j) \right] \tag{1}$$

where

$$B_v^c = \sum_{j=1}^{J} p_{\bullet j} s_J(j) b_{v-1}^2(j) \tag{2}$$

$$C_v^c = \sum_{j=1}^{J} p_{\bullet j} s_J(j) b_{v-1}(j) b_{v-2}(j) \tag{3}$$

and

$$A_v^c = \left\{ \sum_{j=1}^{J} p_{\bullet j} s_J(j)^2 b_{v-1}^2(j) - (B_v^c)^2 - (C_v^c)^2 \right\}^{-1/2} \tag{4}$$

Note that $b_{-1}(j) = 0$ and $b_0(j) = 1$ for all j. For the calculation of (1), (2), (3) and (4), a set of scores is required to represent the ordinal

nature of the categories. Scores may chosen subjectively, such as natural scores where $s_J(j) = j$, or by more objective means, such as Nishisato scores (Nishisato and Arri [23]), or midrank scores. Beh [4] shows that often natural scores give very similar results to the more computationally intensive scores. For the assignement of scores to ordinal log-linear models, Ishii-Kuntz [20] views the use of natural scores as very simple and are therefore an advantage of the procedure.

Row orthogonal polynomials can be calculated in a similar manner to the column polynomials. For ordered row categories, let $\{a_u(i)\}$ denote the set of orthogonal polynomials on $\{p_{i\bullet}\}$, for $u = 1, 2, \ldots, (I-1)$ and $i = 1, 2, \ldots, I$. These polynomials have the property

$$\sum_{i=1}^{I} p_{i\bullet} a_u(i) a_{u'}(i) = \begin{cases} 1, u = u' \\ 0, u \neq u' \end{cases}$$

When a third variable is included, consisting of K tubes, the set of tube orthogonal polynomials of order w, $\{c_w(k)\}$, are defined so that they are orthogonal across the tube marginal probabilities $\{p_{\bullet\bullet k}\}$ subject to the constraint

$$\sum_{k=K}^{I} p_{\bullet\bullet k} c_w(k) c_{w'}(k) = \begin{cases} 1, w = w' \\ 0, w \neq w' \end{cases}$$

4. Two-Way Contingency Tables

The RC canonical correlation model can be used to measure the departure from independence of two ordered variables from a two-way cross-classification and has been extensively investigated. See for example Gilula [14], Gilula and Haberman [15, 16], and Gilula and Ritov [17]. In this section we consider the correlation models for two-way ordinal cross - classifications as seen in Beh [3], Beh and Davy [6, 7], Rayner and Best [24] and Best and Rayner [8].

4.1. Doubly ordered two-way contingency tables

The RC bivariate correlation model for a doubly ordinal two-way cross-classifications has been discussed by Rayner and Best [24] and Beh [3]. These authors demonstrate that p_{ij} can be reconstituted by the *saturated* model

$$p_{ij} = p_{i\bullet} p_{\bullet j} \left(1 + \sum_{u=1}^{I-1} \sum_{v=1}^{J-1} a_u(i) \hat{\theta}_{uv} b_v(j) \right) \tag{5}$$

where

$$\hat{\theta}_{uv} = \sum_{i=1}^{I} \sum_{j=1}^{J} a_u(i) \, b_v(j) \, p_{ij}$$

The $\sqrt{n}\hat{\theta}_{uv}$ values are asymptotically standard normally distributed random variables and are asymptotically independent, and are calculated using the set of row scores $s_I(i)$ and the set of column scores $s_J(j)$. To determine significant association values we can compare each of these terms with what is expected under the standard normal distribution. Rayner and Best [25] showed that when natural scores are used, $\hat{\theta}_{11}$ is the Pearson product moment correlation. They also showed that when midrank scores are used to calculate this quantity, then it is equivalent to Spearman's rank correlation.

When model (5) is *unsaturated*, it can sometimes give negative p_{ij} values. In this case, Rayner and Best [24] suggested the following alternative exponential model

$$p_{ij} \approx p_{i\bullet} p_{\bullet j} \exp\left\{ \sum_{u=1}^{I-1} \sum_{v=1}^{J-1} a_u(i) \, \hat{\theta}_{uv} b_v(j) \right\} \tag{6}$$

It is with this and similar exponential approximations that we establish a parallel between our non-iterative approach and conventional log-linear models.

4.2.   Singly ordered two-way contingency tables

Suppose now we consider a singly ordered two-way contingency table, where the columns are of an ordinal nature. The saturated bivariate correlation model discussed in Beh [3] and Rayner and Best [25] is

$$p_{ij} = p_{i\bullet} p_{\bullet j} \left(1 + \sum_{v=1}^{J-1} \hat{\theta}_{(i)v} b_v(j) \right) \tag{7}$$

where

$$\hat{\theta}_{(i)v} = \sum_{j=1}^{J} \frac{p_{ij}}{p_{i\bullet}} b_v(j) \tag{8}$$

The exponential approximation of model (7) is

$$p_{ij} \approx p_{i\bullet} p_{\bullet j} \exp\left(\sum_{v=1}^{J-1} \hat{\theta}_{(i)v} b_v(j) \right)$$

5.   Three-Way Contingency Tables

For the analysis of three-way ordinal contingency tables, Beh and Davy [6, 7] presented some models which are extensions of the bivariate models presented in the previous section. These trivariate models can be used to estimate the parameter values from a log-linear model of the data.

5.1.   Completely ordered three-way tables

Suppose that a three-way contingency table consists of three variables which have an ordered categorical responses. Beh and Davy [6] show that the saturated RC model can be defined by

$$p_{ijk} = p_{i\bullet\bullet} p_{\bullet j\bullet} p_{\bullet\bullet k} \left(1 + \sum_{u=1}^{I-1} \sum_{v=1}^{J-1} \hat{\theta}_{uv0} a_u\left(i\right) b_v\left(j\right) \right.$$
$$+ \sum_{u=1}^{I-1} \sum_{w=1}^{K-1} \hat{\theta}_{u0w} a_u\left(i\right) c_w\left(k\right) + \sum_{v=1}^{J-1} \sum_{w=1}^{K-1} \hat{\theta}_{0vw} b_v\left(j\right) c_w\left(k\right) \qquad (9)$$
$$\left. + \sum_{u=1}^{I-1} \sum_{v=1}^{J-1} \sum_{w=1}^{K-1} \hat{\theta}_{uvw} a_u\left(i\right) b_v\left(j\right) c_w\left(k\right) \right)$$

where

$$\hat{\theta}_{uvw} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} a_u\left(i\right) b_v\left(j\right) c_w\left(k\right) p_{ijk}$$

Beh and Davy [6] showed that $\hat{\theta}_{uv0}$, $\hat{\theta}_{u0w}$ and $\hat{\theta}_{0vw}$ are bivariate row-column, row-tube and column-tube correlation values respectively. The $\hat{\theta}_{uvw}$ are extensions of the Pearson product moment correlation when natural scores are used for trivariate data. When midrank scores are used, they are extensions to the Spearman rank correlation for the same data structure.

The term $\hat{\theta}_{uvw}$ is a measure of the deviation of the rows, columns and tubes up to the $(u, v, w)'th$ trivariate moment in the data from might be expected under complete independence. For example, $\hat{\theta}_{111}$ is the linear-by-linear-by-linear association and can be used to assess the trivariate location of the three variables. Refer to Davy, Rayner and Beh [10] for information on the interpretation of $\hat{\theta}_{uvw}$.

The unsaturated form of model of (9) can sometimes calculate negative values in much the same way as the bivariate models can. To avoid this, the value of p_{ijk} can be approximately reconstituted by the exponential

model

$$p_{ijk} \approx p_{i\bullet\bullet}p_{\bullet j\bullet}p_{\bullet\bullet k} \exp \left(\sum_{u=1}^{I-1} \sum_{v=1}^{J-1} \hat{\theta}_{uv0}a_u(i)b_v(j) \right.$$
$$+ \sum_{u=1}^{I-1} \sum_{w=1}^{K-1} \hat{\theta}_{u0w}a_u(i)c_w(k) + \sum_{v=1}^{J-1} \sum_{w=1}^{K-1} \hat{\theta}_{0vw}b_v(j)c_w(k) \qquad (10)$$
$$\left. + \sum_{u=1}^{I-1} \sum_{v=1}^{J-1} \sum_{w=1}^{K-1} \hat{\theta}_{uvw}a_u(i)b_v(j)c_w(k) \right)$$

5.2.   Doubly ordered three-way tables

For a three-way contingency table with ordered row and column categories, Beh and Davy [7] showed that the saturated model associated with such a data structure is

$$p_{ijk} = p_{i\bullet\bullet}p_{\bullet j\bullet}p_{\bullet\bullet k} \left(1 + \sum_{u=1}^{I-1} \hat{\theta}_{u0(k)}a_u(i) + \sum_{v=1}^{J-1} \hat{\theta}_{0v(k)}b_v(j) \right.$$
$$\left. + \sum_{u=1}^{J-1} \sum_{v=1}^{J-1} \hat{\theta}_{uv(k)}a_u(i)b_v(j) \right) \qquad (11)$$

where

$$\hat{\theta}_{uv(k)} = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{p_{ijk}}{p_{\bullet\bullet k}} a_u(i)b_v(j) \qquad (12)$$

The value of $\hat{\theta}_{uv(k)}$ is the $(u,v)'th$ bivariate association between the rows and columns at the $k'th$ tube. For example, $\hat{\theta}_{11(2)}$ is the linear-by-linear association between the rows and column at the second tube category. Beh and Davy [7] defined the $(u,v)'th$ bivariate association for the ordered rows and ordered columns of the three-way contingency table as

$$\hat{\theta}_{uv(\bullet)} = \sum_{k=1}^{K} p_{\bullet\bullet k}\hat{\theta}_{uv(k)} \qquad (13)$$

The value of the $(i,j,k)'th$ cell proportion can be approximately reconstituted using the exponential form of (11)

$$p_{ijk} \approx p_{i\bullet\bullet}p_{\bullet j\bullet}p_{\bullet\bullet k} \exp \left(\sum_{u=1}^{I-1} \hat{\theta}_{u0(k)}a_u(i) + \sum_{v=1}^{J-1} \hat{\theta}_{0v(k)}b_v(j) \right.$$
$$\left. + \sum_{u=1}^{J-1} \sum_{v=1}^{J-1} \hat{\theta}_{uv(k)}a_u(i)b_v(j) \right) \qquad (14)$$

5.3. Singly ordered three-way tables

For the analysis of a three-way table with ordered row categories only (singly ordered), Beh and Davy [7] used

$$p_{ijk} = p_{i\bullet\bullet}p_{\bullet j\bullet}p_{\bullet\bullet k}\left(1 + \sum_{u=1}^{I-1}\hat{\theta}_{u(jk)}a_u\left(i\right)\right)$$

where

$$\hat{\theta}_{u(jk)} = \sum_{i=1}^{I}\frac{p_{ijk}}{p_{\bullet j\bullet}p_{\bullet\bullet k}}a_u\left(i\right) \tag{15}$$

The value of the $(i,j)'th$ cell proportion can be approximately reconstituted using the exponential model

$$p_{ijk} \approx p_{i\bullet\bullet}p_{\bullet j\bullet}p_{\bullet\bullet k}\exp\left(\sum_{u=1}^{I-1}\hat{\theta}_{u(jk)}a_u\left(i\right)\right) \tag{16}$$

6. Two-Way Log-linear Models

Log-linear models are a popular way to determine if there is any relationship between any sets of variables of a cross-classification. In this section we look at a new method of calculating the parameters of a log-linear model and show that it is a far more versatile approach than the conventional maximum likelihood approach discussed in Agresti [1, p80] and Fienberg [13, p54].

6.1. Doubly ordered two-way contingency tables

Suppose that we wish to conduct a log-linear analysis on a two-way contingency table with ordered rows and ordered columns. Agresti [1] showed that the most applicable theoretical log-linear model is

$$\ln m_{ij} = \mu + \alpha_i + \beta_j + \phi\left[s_I\left(i\right) - \mu_I\right]\left[s_J\left(j\right) - \mu_J\right] \tag{17}$$

where μ_I and μ_J are the mean row and column scores respectively and $\sum_{i=1}^{I}\alpha_i = \sum_{j=1}^{J}\beta_j = 0$. The value of m_{ij} is the expected value of the $(i,j)'th$ cell entry in N.

The value of ϕ in (17) describes the association between the two ordinal categorical variables. Note that, just as described by Agresti [1, p77],

when $\phi = 0$, this leads to the independence model, while the quantity $\phi\left[s_I\left(i\right) - \mu_I\right]\left[s_J\left(j\right) - \mu_J\right]$ measures the deviation from independence for $\ln m_{ij}$. Haberman [19] proposed a method of estimating this parameter value using maximum-likelihood estimation, which Agresti [1] describes.

Now, consider the exponential form of the RC canonical correlation model given by (6). Multiplying this model by n and taking the natural logarithm of both sides yields

$$\ln n_{ij} \approx \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \sum_{u=1}^{I-1}\sum_{v=1}^{J-1} a_u\left(i\right)\hat{\theta}_{uv}b_v\left(j\right) \tag{18}$$

where

$$\hat{\mu} = \ln n + \frac{1}{I}\sum_{i=1}^{I}\ln p_{i\bullet} + \frac{1}{J}\sum_{j=1}^{I}\ln p_{\bullet j},$$

$$\hat{\alpha}_i = \ln p_{i\bullet} - \frac{1}{I}\sum_{i=1}^{I}\ln p_{i\bullet} \quad \text{and} \quad \hat{\beta}_j = \ln p_{\bullet j} - \frac{1}{J}\sum_{j=1}^{I}\ln p_{\bullet j}$$

so that $\sum_{i=1}^{I}\hat{\alpha}_i = \sum_{j=1}^{J}\hat{\beta}_i = 0$ and (18) is the ordinal form of the log-linear model presented in van der Heijden et al. [28].

6.1.1. Linear Terms

If only the linear terms of the row and column categories are considered, that is when $u = v = 1$, then (18) simplifies to

$$\ln n_{ij} \approx \mu + \hat{\alpha}_i + \hat{\beta}_j + \hat{\theta}_{11}\frac{\left[s_I\left(i\right) - \hat{\mu}_I\right]}{\hat{\sigma}_I}\frac{\left[s_J\left(j\right) - \hat{\mu}_J\right]}{\hat{\sigma}_J} \tag{19}$$

where $\hat{\mu}_I = \sum_{i=1}^{I}s_I\left(i\right)p_{i\bullet}$ and $\hat{\sigma}_I^2 = \sum_{i=1}^{I}s_I\left(i\right)^2 p_{i\bullet} - \hat{\mu}_I^2$. The parameters $\hat{\mu}_J$ and $\hat{\sigma}_J^2$ are defined similarly.

Comparing model (17) with (19) we obtain an estimate of the parameter ϕ

$$\tilde{\phi} = \frac{\hat{\theta}_{11}}{\hat{\sigma}_I\hat{\sigma}_J} \tag{20}$$

6.1.2. Quadratic Terms

Suppose now we wish to include the quadratic terms with the linear terms in (18) so that we may determine the quadratic nature of the variables and

be more precise with our estimate of ϕ. Then the conventional log-linear model is

$$
\begin{aligned}
\ln m_{ij} &= \mu + \alpha_i + \beta_j + \phi \left[s_I(i) - \mu_I \right] \left[s_J(j) - \mu_J \right] \\
&\quad + \phi_{12} \left[s_I(i) - \mu_I \right]^2 \left[s_J(j) - \mu_J \right] + \phi_{12} \left[s_I(i) - \mu_I \right] \left[s_J(j) - \mu_J \right]^2
\end{aligned}
$$

Therefore, considering the linear and quadratic terms in (18), we obtain,

$$
\begin{aligned}
\ln n_{ij} &= \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\theta}_{11} a_1(i) b_1(j) \\
&\quad + \hat{\theta}_{12} a_1(i) b_2(j) + \hat{\theta}_{21} a_2(i) b_1(j)
\end{aligned}
$$

After a little algebra, the parameters ϕ, ϕ_{12} and ϕ_{21} can be estimated by

$$
\tilde{\phi}_{12} = A_2^c \frac{\hat{\theta}_{12}}{\hat{\sigma}_I \hat{\sigma}_J}, \qquad \tilde{\phi}_{21} = A_2^r \frac{\hat{\theta}_{21}}{\hat{\sigma}_I \hat{\sigma}_J} \quad \text{and}
$$

$$
\tilde{\phi} = \frac{\hat{\theta}_{11}}{\hat{\sigma}_I \hat{\sigma}_J} - \hat{\sigma}_J \hat{\gamma}_J \hat{\phi}_{12} - \hat{\sigma}_I \hat{\gamma}_I \hat{\phi}_{21} \tag{21}
$$

where A_2^r and A_2^c are the values of (4) for the row and column categories respectively and $\hat{\gamma}_I$ and $\hat{\gamma}_J$ denote the (standardised) skewness coefficients for the row and column scores respectively.

Observing the above equations, if the quadratic terms do not contribute to the association in the contingency table, then (21) reduces to (20).

Alternatively, the orthogonal polynomials are constructed to include a (standardised) linear, quadratic and higher order vector and could be included instead. In this case, the log-linear model with the quadratic terms included becomes

$$
\ln m_{ij} = \mu + \alpha_i + \beta_j + \phi a_1(i) b_1(j) + \phi_{21} a_2(i) b_1(j) + \phi_{12} a_1(i) b_2(j)
$$

If the researcher was interested in expressing the model in the conventional form, the row (and column) orthogonal polynomials can be regressed such that

$$
a_u(i) = \sum_{g=0}^{u} A_{ug} \left(\frac{s_I(i) - \hat{\mu}_I}{\hat{\sigma}_I} \right)^g
$$

where A_{ug} are unknown constants. For example, $a_2(i)$ can be expressed as

$$
a_2(i) = A_{20} + A_{21} \left(\frac{s_I(i) - \hat{\mu}_I}{\hat{\sigma}_I} \right) + A_{22} \left(\frac{s_I(i) - \hat{\mu}_I}{\hat{\sigma}_I} \right)^2
$$

Thus, we can model the saturated version of the log-linear model by considering all terms in (18). But for the sake of approximation, it is practical to only consider the first few terms; the linear and quadratic terms, say.

The investigation of quadratic terms will be left for future work.

6.2. Singly ordered two-way contingency tables

The approximation of parameters for singly ordered two-way contingency tables can be made in much the same way as discussed in the previous section.

Suppose that we have a two-way table with ordered column categories. The log-linear model presented in Agresti [1] which is applicable to this table is

$$\ln m_{ij} = \mu + \alpha_i + \beta_j + \tau_{J(i)} [s_J(j) - \mu_J] \tag{22}$$

where $\sum_{i=1}^{I} \alpha_i = \sum_{j=1}^{J} \beta_j = 0$.

The value of $\tau_{J(i)}$, such that $\sum_{i=1}^{I} \tau_{J(i)} = 0$, can be calculated using maximum likelihood estimation, and is the measure of linearity if the ordered column categories on the $j'th$ row category. This quantity is shown to be related to $\hat{\theta}_{(i)1}$ as seen in (8).

Consider the exponential two-way model of (7). Multiplying it by n and taking the natural logarithm of both sides of the expression yields

$$\ln n_{ij} \approx \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \sum_{v=1}^{J-1} \hat{\theta}_{(i)v} b_v(j) \tag{23}$$

Consider the linear component of the ordered column categories. Then (23) reduces to

$$\ln n_{ij} \approx \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\theta}_{(i)1} \frac{[s_J(j) - \hat{\mu}_J]}{\hat{\sigma}_J} \tag{24}$$

Comparing (24) with the classical log-linear model of (22), shows that the parameter $\tau_{J(i)}$ can be approximated by

$$\tilde{\tau}_{J(i)} = \frac{\hat{\theta}_{(i)1}}{\hat{\sigma}_J}$$

7. Three-Way Log-Linear Models

Beh and Davy [6, 7] described a method of parameter estimation for three-way contingency tables with any number of the variables ordered. In this section, we provide some of the derivations in their papers and extend them.

7.1. Completely ordered three-way contingency tables

For a three-way contingency table, where all three variables are ordered, Fienberg [13] and Agresti [1] offer the following log-linear model:

$$
\begin{aligned}
\ln m_{ijk} \;=\; & \mu + \alpha_i + \beta_j + \gamma_k + \phi_{IJ}\left[s_I\left(i\right)-\mu_I\right]\left[s_J\left(j\right)-\mu_J\right] \\
& + \phi_{IK}\left[s_I\left(i\right)-\mu_I\right]\left[s_K\left(k\right)-\mu_K\right] \\
& + \phi_{JK}\left[s_J\left(j\right)-\mu_J\right]\left[s_K\left(k\right)-\mu_K\right] \\
& + \phi_{IJK}\left[s_I\left(i\right)-\mu_I\right]\left[s_J\left(j\right)-\mu_J\right]\left[s_K\left(k\right)-\mu_K\right]
\end{aligned} \tag{25}
$$

where $\sum_{i=1}^{I}\alpha_i = \sum_{j=1}^{J}\beta_j = \sum_{k=1}^{K}\gamma_k = 0$. The value of $s_K(k)$ is the score associated with the $k'th$ ordered tube category. The values ϕ_{IJ}, ϕ_{IK} and ϕ_{JK} describe the bivariate association for each pair of variables. These values correspond to the linear-by-linear associations for each pair of variables, while ϕ_{IJK} is the association term and corresponds to the linear-by-linear-by-linear association. To show this, consider the exponential model (10). Taking the natural logarithm of both sides after multiplying by n yields

$$
\begin{aligned}
\ln n_{ijk} \;=\; & \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \sum_{u=1}^{I-1}\sum_{v=1}^{J-1}\hat{\theta}_{uv0}\,a_u\left(i\right)b_v\left(j\right) \\
& + \sum_{u=1}^{I-1}\sum_{w=1}^{K-1}\hat{\theta}_{u0w}\,a_u\left(i\right)c_w\left(k\right) + \sum_{v=1}^{J-1}\sum_{w=1}^{K-1}\hat{\theta}_{0vw}\,b_v\left(j\right)c_w\left(k\right) \\
& + \sum_{u=1}^{I-1}\sum_{v=1}^{J-1}\sum_{w=1}^{K-1}\hat{\theta}_{uvw}\,a_u\left(i\right)b_v\left(j\right)c_w\left(k\right)
\end{aligned} \tag{26}
$$

where

$$
\hat{\mu} \;=\; \ln n + \frac{1}{I}\sum_{i=1}^{I}\ln p_{i\bullet\bullet} + \frac{1}{J}\sum_{j=1}^{J}\ln p_{\bullet j\bullet} + \frac{1}{K}\sum_{k=1}^{k}\ln p_{\bullet\bullet k},
$$

$$
\hat{\alpha}_i \;=\; \ln p_{i\bullet\bullet} - \frac{1}{I}\sum_{i=1}^{I}\ln p_{i\bullet\bullet},
$$

$$
\hat{\beta}_j \;=\; \ln p_{\bullet j\bullet} - \frac{1}{J}\sum_{j=1}^{J}\ln p_{\bullet j\bullet} \quad \text{and}
$$

$$
\hat{\gamma}_k \;=\; \ln p_{\bullet\bullet k} - \frac{1}{K}\sum_{k=1}^{k}\ln p_{\bullet\bullet k}
$$

so that $\sum_{i=1}^{I}\hat{\alpha}_i = \sum_{j=1}^{J}\hat{\beta}_j = \sum_{k=1}^{K}\hat{\gamma}_k = 0$.

If only the linear components for each variable are considered then (26) simplifies to

$$
\begin{aligned}
\ln n_{ijk} \;=\; & \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \hat{\theta}_{110} \frac{[s_I(i) - \hat{\mu}_I]}{\hat{\sigma}_I} \frac{[s_J(j) - \hat{\mu}_J]}{\hat{\sigma}_J} \\
& + \hat{\theta}_{101} \frac{[s_I(i) - \hat{\mu}_I]}{\hat{\sigma}_I} \frac{[s_K(k) - \hat{\mu}_K]}{\hat{\sigma}_K} \\
& + \hat{\theta}_{011} \frac{[s_J(j) - \hat{\mu}_J]}{\hat{\sigma}_J} \frac{[s_K(k) - \hat{\mu}_K]}{\hat{\sigma}_K} \\
& + \hat{\theta}_{111} \frac{[s_I(i) - \hat{\mu}_I]}{\hat{\sigma}_I} \frac{[s_J(j) - \hat{\mu}_J]}{\hat{\sigma}_J} \frac{[s_K(k) - \hat{\mu}_K]}{\hat{\sigma}_K}
\end{aligned}
$$

Comparing (25) and (26) using natural scores, the parameters from the log-linear model of (25) can be approximated by

$$
\tilde{\phi}_{IJ} = \frac{\hat{\theta}_{110}}{\hat{\sigma}_I \hat{\sigma}_J}, \qquad \tilde{\phi}_{IK} = \frac{\hat{\theta}_{101}}{\hat{\sigma}_I \hat{\sigma}_K},
$$

$$
\tilde{\phi}_{JK} = \frac{\hat{\theta}_{011}}{\hat{\sigma}_J \hat{\sigma}_K} \quad \text{and} \quad \tilde{\phi}_{IJK} = \frac{\hat{\theta}_{111}}{\hat{\sigma}_I \hat{\sigma}_J \hat{\sigma}_K} \tag{27}
$$

7.2. Doubly ordered three-way contingency tables

Suppose now that we are interested in the log-linear model for a three-way contingency table with ordered row and column categories. Then, the theoretical log-linear model for the table is

$$
\begin{aligned}
\ln m_{ijk} \;=\; & \mu + \alpha_i + \beta_j + \gamma_k + \tau_{IK(k)} \left[s_I(i) - \mu_I \right] \tag{28} \\
& + \tau_{JK(k)} \left[s_J(j) - \mu_J \right] + \phi_{IJ(k)} \left[s_I(i) - \mu_I \right] \left[s_J(j) - \mu_J \right]
\end{aligned}
$$

In model (28), $\tau_{IK(k)}$ is the effect the linearity of the row categories has on the $k'th$ tube profile while $\tau_{JK(k)}$ is the effect of the linearity of the columns on the $k'th$ tube category. These values correspond to $\hat{\theta}_{10(k)}$ and $\hat{\theta}_{01(k)}$ respectively from (12). The value of $\phi_{IJ(k)}$ is related to the linear-by-linear association of the ordered rows and columns at the $k'th$ non-ordered tube category. This value differs slightly from the parameter in the model presented in Agresti [1], ϕ_{IJ}, which is the overall linear-by-linear association value. To verify these results, consider the exponential model of association (14). Taking the natural logarithm of this model after multiplying by n

yields

$$
\begin{aligned}
\ln n_{ijk} = {}& \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \sum_{u=1}^{I-1} \hat{\theta}_{u0(k)} a_u(i) \\
& + \sum_{v=1}^{J-1} \hat{\theta}_{0v(k)} b_v(j) + \sum_{u=1}^{I-1} \sum_{v=1}^{J-1} \hat{\theta}_{uv(k)} a_u(i) b_v(j) \qquad (29)
\end{aligned}
$$

Consider the linear terms in model (29). Then the linear model is

$$
\begin{aligned}
\ln n_{ijk} = {}& \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \hat{\theta}_{10(k)} \frac{[s_I(i) - \hat{\mu}_I]}{\hat{\sigma}_I} \qquad (30) \\
& + \hat{\theta}_{01(k)} \frac{[s_J(j) - \hat{\mu}_J]}{\hat{\sigma}_J} + \hat{\theta}_{11(k)} \frac{[s_I(i) - \hat{\mu}_I]}{\hat{\sigma}_I} \frac{[s_J(j) - \hat{\mu}_J]}{\hat{\sigma}_J}
\end{aligned}
$$

Therefore, comparing models (28) and (30), the parameters $\tau_{IK(k)}$, $\tau_{JK(k)}$ and $\phi_{IJ(k)}$ can be approximated by

$$
\tilde{\tau}_{IK(k)} = \frac{\hat{\theta}_{10(k)}}{\hat{\sigma}_I}, \qquad \tilde{\tau}_{JK(k)} = \frac{\hat{\theta}_{01(k)}}{\hat{\sigma}_J} \quad \text{and} \quad \tilde{\phi}_{IJ(k)} = \frac{\hat{\theta}_{11(k)}}{\hat{\sigma}_I \hat{\sigma}_J} \qquad (31)
$$

The overall linear-by-linear association between the rows and column of the doubly ordered three-way contingency table, using (13) and (31), is

$$
\tilde{U}_{11(\bullet)} = \hat{\sigma}_I \hat{\sigma}_J \sum_{k=1}^{K} p_{\bullet\bullet k} \tilde{\phi}_{IJ(k)},
$$

while the parameter ϕ_{IJ} from the model of Agresti [1] and Fienberg [13] can be approximated by

$$
\tilde{\phi}_{IJ} = \frac{\tilde{U}_{11(\bullet)}}{\hat{\sigma}_I \hat{\sigma}_J} = \sum_{k=1}^{K} p_{\bullet\bullet k} \tilde{\phi}_{IJ(k)}
$$

7.3. Singly ordered three-way contingency tables

Consider now a three-way contingency table with ordered row categories. The theoretical log-linear model for such a table is

$$
\begin{aligned}
\ln m_{ijk} = {}& \mu + \alpha_i + \beta_j + \gamma_k + \tau_{IJ(j)} [s_I(i) - \mu_I] \qquad (32) \\
& + \tau_{IK(k)} [s_I(i) - \mu_I] + \tau_{IJK(jk)} [s_I(i) - \mu_I]
\end{aligned}
$$

where $\sum_{i=1}^{I} \alpha_i = \sum_{j=1}^{J} \beta_j = \sum_{k=1}^{K} \gamma_k = 0$. The value $\tau_{IJ(j)}$ is a measure of the linearity of the rows and its effect on the $j'th$ column category. Similarly,

$\tau_{IK(k)}$ is the measure of linearity of the rows and its effect on the $k'th$ tube category while $\tau_{IJK(jk)}$ is the measure of linearity of the rows upon the $j'th$ column and $k'th$ tube category. Therefore, $\tau_{IJK(jk)}$ is related to $\hat{\theta}_{1(jk)}$ using (15). To show this, taking the natural logarithm of both sides of (16) after multiplying by n yields

$$\ln n_{ijk} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \sum_{u=1}^{I-1} \hat{\theta}_{u(jk)} a_u(i) \tag{33}$$

When the linear component of the rows is considered (33) becomes

$$\ln n_{ijk} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \hat{\theta}_{1(jk)} \frac{[s_I(i) - \hat{\mu}_I]}{\hat{\sigma}_I} \tag{34}$$

Comparing models (32) and (34), $\tau_{IJK(jk)}$ can be approximated by

$$\tilde{\tau}_{IJK(jk)} = \frac{\hat{\theta}_{1(jk)}}{\hat{\sigma}_I}$$

8.   Examples

8.1.   Example 1 - A doubly ordered two-way table

Consider the following data set of Table 1 first seen in Srole et al. [26] which is a cross-classification of 1660 patients from mid-town Manhattan according to mental health status and parental socio-economic status. Authors such as Best and Rayner [8], Beh [3, 4] and Weller and Romney [29] have cited this example in their analyses. The parental socio-economic status is designated A through to F in a sequence from highest to lowest position. Therefore, Table 1 is a doubly ordered two-way contingency table, so the log-linear analysis described in sub-section 4.1 is applicable.

Haberman [19] analysed Table 1 by estimating the linear by linear association of the data. Using row scores 3, 1, -1, and -3 and column scores 5, 3, 1, -1, -3 and -5, Haberman [19] concluded that an estimate of the correlation between mental health status and parental socio-economic status is 0.02081, suggesting that as the parental socio-economic status improves, so too does the patient's mental health, although this association is not statistically significant. This is the same conclusion reached in Best and Rayner [8] and Beh [3, 4]. Agresti [2] also analysed Table 1 using unit length natural scores, and obtained an estimate equivalent to $\hat{\phi} = 0.091/4 = 0.02275$ for Haberman's scores. Using equation (20), with Pearson product moment

Table 1. Cross-classification of 1660 Patients from Mid-town Manhattan According to Mental Health Status and Parental Socio-Economic Status.

	A	B	C	D	E	F
Well	64	57	57	72	36	21
Mild Symptom Formation	94	94	105	141	97	71
Moderate Symptom Formation	58	54	65	77	54	54
Impaired	46	40	60	94	78	71

correlation 0.14965, $\hat{\sigma}_I = 2.0873$ and $\hat{\sigma}_J = 3.2230$, the non-iterative estimate $\tilde{\phi} = 0.0222584$ is obtained for Haberman's scoring scheme. Therefore the estimate of the log-linear parameter ϕ using the non-iterative methodology presented in this paper is very close to the estimate obtained using maximum likelihood estimation.

A comparison of the estimates obtained by Agresti, Haberman and equation (20) can be made in terms of the fitted values. A goodness-of-fit test is performed using only the above first order parameter estimates by calculating the chi-squared statistic for each of the three methods. Doing so yields the goodness-of-fit chi-squared statistic using the Haberman and Agresti estimates, $X^2_{HAB} = 11.51938$ and $X^2_{AGR} = 11.44213$ respectively. Using the estimate obtained from the non-iterative procedure outlined here, $X^2_{NI} = 11.40662$. The p-values associated with these statistics, and $(I-1)(J-1)-1 = 14$ degrees of freedom are 0.645, 0.651 and 0.654 respectively. This suggests that the non-iterative approach for Table 1 produces estimates which are slightly better than those obtained from the iterative methods. This single parameter also explains nearly three quarters of the total variation from independence for each of the three estimates, since the Pearson chi-squared statistic is $X^2 = 45.98526$.

8.2. Example 2 - A completely ordered three-way table

Consider Table 2 which is a $4 \times 5 \times 3$ contingency table classifying 1517 people according to the number years of completed schooling, their number of siblings and their general level of happiness. Clogg [9] and Beh and Davy [6, 7] considered this contingency table for their analyses. The three variables can be seen to be of an ordinal nature, and so the log-linear analysis of sub-section 7.1 can be applied. Beh and Davy [6,7] showed

that the the three bivariate relationships were highly significant and that the trivariate was not. Therefore, we only need to concern ourselves with estimating the parameters ϕ_{IJ}, ϕ_{IK} and ϕ_{JK}. Table 3 gives a comparison of each of the parameters calculated using maximum likelihood estimation, denoted by $\hat{\phi}$, and those estimates obtained using (27), denoted by $\tilde{\phi}$.

Table 2. Cross-classification of 1517 People According to their Happiness, Schooling and Number of Siblings.

Years of Schooling	Number of Siblings				
	0-1	2-3	4-5	6-7	8+
Not too Happy					
<12	15	34	36	22	61
12	31	60	46	25	26
13-16	35	45	30	13	8
17+	18	14	3	3	4
Pretty Happy					
<12	17	53	70	67	79
12	60	96	45	40	31
13-16	63	74	39	24	7
17+	15	15	9	2	1
Very Happy					
<12	7	20	23	16	36
12	5	12	11	12	7
13-16	5	10	4	4	3
17+	2	1	2	0	1

Table 3. MLE and Alternative Values of Parameters for Table 2.

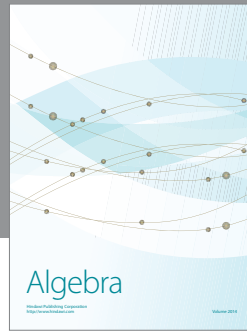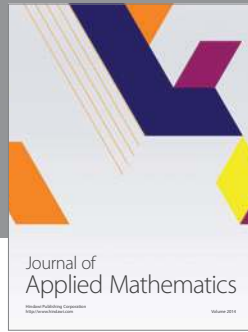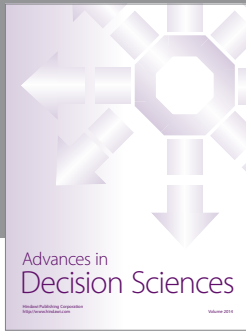| ϕ	ϕ_{IJ}	ϕ_{IK}	ϕ_{JK}
$\hat{\phi}$	-0.3468	-0.2188	0.0726
$\tilde{\phi}$	-0.3298	-0.2140	0.0725

The maximum likelihood estimates compare very well with the alternative values using (27). Thus instead of selecting a log-linear model by "trial and error" as is the case using the conventional approach of parameter estimation for log-linear models, the alternative approach offers an accurate method of parameter estimation.

References

1. A. Agresti. *Analysis of Ordinal Categorical Data*. Wiley, New-York, 1984.
2. A. Agresti. *Categorical Data Analysis*. Wiley, New-York, 1990.
3. E. J. Beh. Simple correspondence analysis of ordinal cross-classifications using orthogonal polynomials. *Biometrical Journal*, 39, 589-613, 1997.
4. E. J. Beh. A comparative study of scores for correspondence analysis of ordered categories. *Biometrical Journal*, 40, 413-429, 1998.
5. E. J. Beh. Partitioning Pearson's chi-squared statistic for a singly ordered two-way contingency table. *The Australian and New Zealand Journal of Statistics*, 43, 327-333, 2001.
6. E. J. Beh and P. J. Davy. Partitioning Pearson's chi-squared statistic for a completely ordered three-way contingency table. *The Australian and New Zealand Journal of Statistics*, 40, 465-477, 1998.
7. E. J. Beh and P. J. Davy. Partitioning Pearson's chi-squared statistic for a partially ordered three-way contingency table. *The Australian and New Zealand Journal of Statistics*, 41, 233-246, 1999.
8. D. J. Best and J. C. W. Rayner. Nonparametric analysis for doubly ordered two-way contingency tables. *Biometrics*, 52, 1153-1156, 1996.
9. C. C. Clogg. Some models for the analysis of association in multiway cross-classifications having ordered categories. *Journal of the American Statistical Association*, 77, 803-815, 1982.
10. P. J. Davy, J. C. W. Rayner and E. J. Beh. Generalised correlations. School of Mathematics and Applied Statistics Preprint 4/2003, University of Wollongong, 2003.
11. W. Dillon and M. Goldstein. *Multivariate Analysis: Methods and Applications*. Wiley, New-York, 1984.
12. P. L. Emerson. Numerical construction of orthogonal polynomials from a general recurrence formula. *Biometrics*, 24, 696-701, 1968.
13. S. E. Fienberg. *The Analysis of Cross-Classified Categorical Data*. MIT Press, Cambridge, 1977.
14. Z. Gilula. Grouping of association in contingency tables : an exploratory canonical correlation approach. *Journal of the American Statistical Association*, 81, 773-779, 1986.
15. Z. Gilula and S. J. Haberman. Canonical analysis of contingency tables by maximum likelihood. *Journal of the American Statistical Association*, 81, 780-788, 1986.
16. Z. Gilula and S. J. Haberman. The analysis of multivariate contingency tables by restricted canonical and restricted association models. *Journal of the American Statistical Association*, 83, 760-771, 1988.
17. Z. Gilula and Y. Ritov. Inferential ordinal correspondence analysis : motivation, derivation and limitations. *International Statistical Review*, 58, 101-108, 1990.

18. J. C. Gower. Discussion of "A combined approach to contingency tables using correspondence analysis and log-linear models". *Applied Statistics*, 38, 249-292, 1989.

19. S. J. Haberman. Log-linear models for frequency tables with ordered classifications. *Biometrics*, 30, 589-600, 1974.

20. M. Ishii-Kuntz. *Ordinal Log-linear Models*. Newury Park, CA: Sage University Paper Series of Quantification Applications in Social Sciences, 07-097, 1994.

21. M. Kendall and A. Stuart. *The Advanced Theory of Statistics: Volume 2*. Charles Griffin, London, 1979.

22. P. M. Kroonenberg. Singular value decomposition of interactions in three-way contingency tables. In R. Coppi and S. Bolasco (Eds.), *Multiway Data Analysis*, 169-184, 1989.

23. S. Nishisato and P. S. Arri. Non-linear programming approach to optimal scaling of partially ordered categories. *Psychometrika*, 40, 525-547, 1975.

24. J. C. W. Rayner and D. J. Best. Smooth extensions of Pearson's product moment correlation and Spearman's Rho. *Statistics and Probability Letters*, 30, 171-177, 1996.

25. J. C. W. Rayner and D. J. Best. A smooth analysis of singly ordered two-way contingency tables. *Journal of Applied Mathematics and Decision Sciences*, 4, 83-98, 2000.

26. L. Srole, T. Langner, S. Michael, M. Opler and T. Rennie. *Mental Health in the Metropolis: The Midtown Manhattan Study*. McGraw-Hill, New-York, 1992.

27. P. van der Heijden and J. de Leeuw. Correspondence analysis used complimentary to log-linear analysis. *Psychometrika*, 50, 429-447, 1985.

28. P. G. M. van der Heijden, A. de Falguerolles and J. de Leeuw. A combined approach to contingency table analysis using correspondence analysis and log-linear analysis. *Applied Statistics*, 38, 249-292, 1989.

29. S. C. Weller and A. K. Romney. *Metric Scaling: Correspondence Analysis*. Newury Park, CA: Sage University Paper Series of Quantification Applications in Social Sciences, 07-075, 1990.