Research Article

Simple Correspondence Analysis of Nominal-Ordinal Contingency Tables

Eric J. Beh

School of Computing and Mathematics, University of Western Sydney, Locked Bag 1797, Penrith South DC, NSW 1797, Australia

Correspondence should be addressed to Eric J. Beh, e.beh@uws.edu.au

The correspondence analysis of a two-way contingency table is now accepted as a very versatile tool for helping users to understand the structure of the association in their data. In cases where the variables consist of ordered categories, there are a number of approaches that can be employed and these generally involve an adaptation of singular value decomposition. Over the last few years, an alternative decomposition method has been used for cases where the row and column variables of a two-way contingency table have an ordinal structure. A version of this approach is also available for a two-way table where one variable has a nominal structure and the other variable has an ordinal structure. However, such an approach does not take into consideration the presence of the nominal variable. This paper explores an approach to correspondence analysis using an amalgamation of singular value decomposition and bivariate moment decomposition. A benefit of this technique is that it combines the classical technique with the ordinal analysis by determining the structure of the variables in terms of singular values and location, dispersion and higher-order moments.

1. Introduction

The analysis of categorical data is a very important component in statistics, and the presence of ordered variables is a common feature. Models and measures of association for ordinal categorical variables have been extensively discussed in the literature, and are the subject of classic texts including Agresti [1], Goodman [2], and Haberman [3].

The visual description of the association between two or more variables is a vital tool for the analyst since it can often provide a more intuitive view of the nature of the association, or interaction, between categorical variables than numerical summaries alone. One such tool is correspondence analysis. However, except in a few cases [4–7], the classical approach to correspondence analysis neglects the presence of ordinal categorical variables

when identifying the structure of their association. One way to incorporate the ordinal structure of categorical variables in simple correspondence analysis is to adopt the approach of Beh [8]. His method takes into account the ordinal structure of one or both variables of a two-way contingency table. At the heart of the procedure is the partition of the Pearson chi-squared statistic described by Best and Rayner [9] and Rayner and Best [10]. However, when there is only one ordered variable, Beh's [8] approach to correspondence analysis does not consider the structure of the nominal variable. This paper does consider the previously neglected nominal variable by using the partition of the Pearson chi-squared statistic described by Beh [11]. The partition involves terms that summarize the association between the nominal and ordinal variables using bivariate moments. These moments are calculated using orthogonal polynomials for the ordered variable and generalized basis vectors of a transformation of the contingency table for the nominal variable.

The correspondence analysis approach described here, referred to as singly ordered correspondence analysis, is shown to be mathematically similar to the doubly ordered approach. The singly ordered and doubly ordered approaches share many of the features that make the classical approach popular. Details of classical correspondence analysis can be found by referring to, for example, Beh [12], Benzécri [13], Greenacre [14], Hoffman and Franke [15], and Lebart et al. [16]. A major benefit of singly ordered correspondence analysis is that nominal row categories and ordinal column categories can be simultaneously represented on a single correspondence plot while ensuring that the structure of both variables is preserved. Constructing such a joint plot for the singly ordered approach of Beh [8] is not possible due to the scaling of coordinates considered in that paper. For the technique described in this paper, the special properties linking the bivariate moments and singular values provide the researcher with an informative interpretation of the association in contingency tables. These numerical summaries also allow, through mechanisms common to correspondence analysis, a graphical interpretation of this association. Hybrid decomposition has also been considered for the nonsymmetrical correspondence analysis of a two-way contingency table by Lombardo et al. [17].

This paper is divided into seven further sections. Section 2 defines the Pearson ratio and various ways in which it can be decomposed to yield numerical and graphical summaries of association. The decompositions considered are (a) singular value decomposition, used in classical correspondence analysis, (b) bivariate moment decomposition, used for the doubly ordered correspondence analysis approach of Beh [8], and (c) hybrid decomposition. This latter technique amalgamates the two former procedures and is important for the singly ordered correspondence analysis technique described in this paper. Section 3 summarizes, by considering the hybrid decomposition of the Pearson ratio, the coordinates needed to obtain a graphical summary of association between the two categorical variables while Section 4 provides an interpretation of the distance between the coordinates in the correspondence plot. Section 5 defines the transition formulae which describe the relationship between the coordinates of the two variables. Various properties of singly ordered correspondence analysis are highlighted in Section 6. The features of the technique are examined using a pedagogical example in Section 7 where it is applied to the data described in Calimlin et al. [18]. Their contingency table summarizes the classification of four analgesic drugs according to their effectiveness judged by 121 hospital patients. The paper concludes with a brief discussion in Section 8.

2. Decomposing Pearson's ratio

2.1. The Pearson ratio

Consider a two-way contingency table N that cross-classifies n units/individuals according to I nominal row categories and J ordered column categories. Denote the (i, j)th element of N by n_{ij}, for $i = 1, 2, \ldots, I$ and $j = 1, 2, \ldots, J$ and the (i, j)th cell relative frequency as $p_{ij} = n_{ij}/n$ so that $\sum_{i=1}^{I} \sum_{j=1}^{J} p_{ij} = 1$. Let the $I \times J$ matrix of these values be denoted as P and let $p_{i\bullet}$ be the ith row marginal proportion of N so that $\sum_{i=1}^{I} p_{i\bullet} = 1$, and D_I the $I \times I$ diagonal matrix where the (i, i)th cell entry is $p_{i\bullet}$. Similarly, let $p_{\bullet j}$ be the jth column marginal proportion so that $\sum_{j=1}^{J} p_{\bullet j} = 1$, and D_J the $J \times J$ diagonal matrix where the (j, j)th cell entry is $p_{\bullet j}$. Define $p_{ij}/p_{i\bullet}$ as the ith row profile and the (i, j)th element of $D_I^{-1}P$, and $p_{ij}/p_{\bullet j}$ the jth column profile and the (i, j)th element of $D_I^{-1}P^T$.

For the (i, j)th cell entry, Goodman [19] described the measure of the departure from independence for row i and column j by the Pearson ratio

$$\alpha_{ij} = \frac{p_{ij}}{p_{i\bullet}p_{\bullet j}}. \tag{2.1}$$

In matrix notation, the Pearson ratio α_{ij} is the (i, j)th cell value of the matrix Δ, where

$$\Delta = D_I^{-1}PD_J^{-1}. \tag{2.2}$$

Independence between the I rows and J columns of N will occur when $\Delta = U$, where U is the $I \times J$ unity matrix where all the values are equal to 1. One can examine where independence does not occur by identifying those Pearson ratios that are statistically significantly different from 1.

A more formal approach to determine whether there exists an association between the row and column categories involves decomposing the matrix of Pearson ratios, Δ. For the correspondence analysis of N, there are a variety of ways in which the decomposition can be performed. Here we will consider three methods of decomposition: singular value decomposition, bivariate moment decomposition, and hybrid decomposition. It is the consideration of the third approach here that is important for the method of correspondence analysis discussed in this paper. The use of hybrid decomposition relies on some basic knowledge of singular value decomposition and bivariate moment decomposition and so these will be described in the following subsections.

2.2. Singular value decomposition

Classically, correspondence analysis involves decomposing the matrix of Pearson ratios using singular value decomposition (SVD) so that

$$\Delta = \tilde{A}\tilde{D}_\lambda\tilde{B}^T, \tag{2.3}$$

where $\tilde{A}$ and $\tilde{B}$ have the property

$$\tilde{A}^T D_I \tilde{A} = I, \qquad \tilde{B}^T D_J \tilde{B} = I, \tag{2.4}$$

respectively where I is an identity matrix. Also, $\tilde{D}_\lambda = \text{diag}(1, \lambda_1, \ldots, \lambda_{M^*})$, where λ_m is the mth largest singular value of α_{ij}, for $m = 1, \ldots, M^*$.

For the decomposition of (2.3), $\tilde{A}$ is an $I \times M$ matrix of left generalized basic vectors, while $\tilde{B}$ is a $J \times M$ matrix of right generalized basic vectors. In both cases, $M = \min(I, J)$ and the first (trivial) singular vector of both matrices has all values equal to one. Let A and B be the matrices $\tilde{A}$ and $\tilde{B}$, respectively, with the trivial singular vector from each is omitted. The matrix $\tilde{D}_\lambda$ is an $M \times M$ diagonal matrix where the (m, m)th cell value is the mth singular value, λ_m, of Δ. These singular values have the property that they are arranged in descending order so that $1 = \lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{M^*} \geq 0$, where $M^* = \min(I, J) - 1$.

Suppose we omit the trivial column vector from $\tilde{A}$ and $\tilde{B}$ to give the $I \times M^*$ matrix A and the $J \times M^*$ matrix B, respectively. Also omit the first row and first column from the matrix $\tilde{D}_\lambda$ (since the $(1, 1)$th element of $\tilde{D}_\lambda$ is equal to 1), obtaining the $M^* \times M^*$ matrix D_λ. Then the SVD of the Pearson ratio becomes the SVD of

$$\Delta - U = A D_\lambda B^T \tag{2.5}$$

whose elements Goodman [19] referred to as Pearson contingencies.

The SVD of these contingencies leads to the Pearson chi-squared statistic being expressed in terms of the sum of squares of the singular values such that

$$X^2 = n \sum_{m=1}^{M^*} \lambda_m^2 = n \, \text{trace}(D_\lambda^2). \tag{2.6}$$

2.3. Bivariate moment decomposition

When a two-way contingency table consists of at least one ordered variable, the ordinal structure of the variable needs to be taken into consideration. Over the past few decades, there have been a number of correspondence analysis procedures developed that take into account the ordinal structure of the variables; see, for example, [4–7]. Generally, these procedures involve imposing ordinal constraints on the singular vectors. Such a procedure therefore forces the position of the points (along the first axis) of the plot to be ordered, thereby imposing what can sometimes lead to unrealistic "correspondences" between row and column categories. A way to overcome this problem is to consider using orthogonal polynomials rather than imposing constraints on the columns of A and B considered in the previous section.

For a doubly ordered two-way contingency table, the correspondence analysis approach of Beh [8] employs the bivariate moment decomposition (BMD) of Pearson ratios so that

$$\Delta = \tilde{A}_* \tilde{Y} \tilde{B}_*^T, \tag{2.7}$$

where

$$\tilde{A}_*^T D_I \tilde{A}_* = I, \qquad \tilde{B}_*^T D_J \tilde{B}_* = I. \tag{2.8}$$

For the decomposition of (2.7), $\tilde{A}_*$ is an $I \times I$ matrix of row orthogonal polynomials, while $\tilde{B}_*$ is a $J \times J$ matrix of column orthogonal polynomials. The (j, v)th element of $\tilde{B}_*$ may be calculated by considering the recurrence relation

$$b_v(j) = S_v \left[(s_J(j) - T_v) b_{v-1}(j) - V_v b_{v-2}(j) \right], \tag{2.9}$$

where

$$T_v = \sum_{j=1}^{J} p_{\bullet j} s_J(j) b_{v-1}^2(j),$$

$$V_v = \sum_{j=1}^{J} p_{\bullet j} s_J(j) b_{v-1}(j) b_{v-2}(j), \qquad (2.10)$$

$$S_v = \left\{ \sum_{j=1}^{J} p_{\bullet j} s_J^2(j) b_{v-1}^2(j) - T_v^2 - C_v^2 \right\}^{-1/2},$$

for $v = 0, 1, \ldots, J-1$. These are based on the general recurrence relation of Emerson [20] and depend on the jth score, $s_J(j)$, assigned to reflect the structure of the column variables. There are many different types of scores that can be considered and Beh [21] discusses the impact of using four different scoring types (two objectively and two subjectively chosen scores) on the orthogonal polynomials. However, for reasons of simplicity and interpretability, we will be considering the use of natural column scores $s_J(j) = j$, for $j = 1, 2, \ldots, J$, and natural row scores in this paper. For both $\tilde{A}_*$ and $\tilde{B}_*$, the first column vector is trivial, having values equal to 1 so that $b_0(j) = 1$ and $a_0(i) = 1$. It is also assumed that $b_{-1}(j) = 0$ and $a_{-1}(i) = 0$, for all i and j.

The matrix $\breve{Y}$ is of size $I \times J$ where the first row and column have values all equal to 1. The nontrivial elements of this matrix are referred to as bivariate moments, or generalized correlations, and describe linear and nonlinear sources of association between the two categorical variables. By omitting these trivial vectors, the decomposition of (2.7) becomes

$$\Delta - U = A_* Y B_*^T, \qquad (2.11)$$

where A_* and B_* are the row and column orthogonal polynomials, respectively, with the first (trivial) column vector omitted. The matrix Y has elements which are the bivariate moments defined by

$$Y = A_*^T P B_*. \qquad (2.12)$$

By considering the BMD (2.11), the Pearson chi-squared statistic can be partitioned into bivariate moments so that

$$X^2 = n \sum_{u=1}^{I-1} \sum_{v=1}^{J-1} Y_{uv}^2 = n \, \text{trace}(Y^T Y) = n \, \text{trace}(Y Y^T), \qquad (2.13)$$

where the elements of Y are asymptotically standard normally distributed. Refer to Best and Rayner [22] and Rayner and Best [23] for a full interpretation of (2.12) and (2.13). An advantage of using BMD is that the (u, v)th element of Y, Y_{uv} has a clear and simple interpretation; it is the (u, v)th bivariate moment between the categories of the row and column variables. As a result, Davy et al. [24] refer to these values as generalized correlations. For example, the linear-by-linear relationship can be measured by

$$Y_{11} = \sum_{i=1}^{I} \sum_{j=1}^{J} p_{ij} \left(\frac{s_I(i) - \mu_I}{\sigma_I} \right) \left(\frac{s_J(j) - \mu_J}{\sigma_J} \right), \qquad (2.14)$$

where $s_I(i)$ and $s_J(j)$ are the set of row and column scores used to construct the orthogonal polynomials, and $\mu_J = \sum_{j=1}^{J} s_J(j) p_{\bullet j}$ and $\sigma_J^2 = \sum_{j=1}^{J} s_J(j)^2 p_{\bullet j} - \mu_J^2$. The quantities μ_I and σ_I^2 are similarly defined. By decomposing the Pearson ratios using BMD when natural scores are used to reflect the ordinal structure of both variables, Y_{11} is equivalent to Pearson's product moment correlation; see Rayner and Best [23]. One can also determine the mean (location) and spread (dispersion) of each of the nonordered row categories across the ordered column categories by calculating $\mu_J(i) = \sum_{j=1}^{J} s_J(j) p_{ij}$ so that $\mu_J = \sum_{i=1}^{I} \mu_J(i)$ and $\sigma_J^2(i) = \sum_{j=1}^{J} s_J(j)^2 p_{ij} - \mu_J(i)^2$, respectively.

2.4. Hybrid decomposition

Another type of decomposition, and one that was briefly discussed by Beh [12], is what is referred to as hybrid decomposition (HD). For a singly ordered contingency table, hybrid decomposition takes into account the ordered variable and nominal variable by incorporating singular vectors from SVD and orthogonal polynomials from BMD such that the Pearson contingencies are decomposed by

$$\Delta - U = AZB_*^T. \tag{2.15}$$

The Z matrix of (2.15) is defined as

$$Z = A^T P B_*. \tag{2.16}$$

The $M^* \times (J-1)$ matrix of Z values, $\{Z_{(u)v} : u = 1, 2, \ldots, M^*, \ v = 1, 2, \ldots, J-1\}$, can be derived by premultiplying (2.15) by $A^T D_I$ and postmultiplying it by $D_J B_*^T$.

If one considers the decomposition of the matrix of Pearson contingencies using the hybrid decomposition of (2.15), then the partition of the Pearson chi-squared statistic can be expressed in terms of the sum of squares of the $Z_{(u)v}$ so that

$$X^2 = n \sum_{u=1}^{M^*} \sum_{v=1}^{J-1} Z_{(u)v}^2 = n \operatorname{trace}(Z^T Z) = n \operatorname{trace}(Z^T Z), \tag{2.17}$$

where the elements of Z are asymptotically standard normal and independent. Refer to Beh [11] for more details on (2.16) and (2.17).

The effect of the column location component on the two-way association in the contingency table is measured by $\sum_{u=1}^{M^*} Z_{(u)1}^2$, while, in general, the vth-order column component is $\sum_{u=1}^{M^*} Z_{(u)v}^2$. The significance of these components can be compared with the chi-squared with M^* degrees of freedom. Testing these column components allows for an examination of the trend of the column categories, the trend being dictated by the vth orthogonal polynomial. For example, the column location component determines if there is any difference in the mean values of the column categories, while the column dispersion component detects if there is any difference in the spread of the columns.

The first-order row location component on the two-way association in the contingency table is measured by $\sum_{v=1}^{J-1} Z_{(1)v}^2$, while in general, the uth-order row component value is equivalent to $\sum_{v=1}^{J-1} Z_{(u)v}^2$. The row location component quantifies the variation in the row categories due to the mean difference in the row categories. Similarly, the row dispersion component

quantifies the amount of variation that is due to the spread in the row categories. Refer to Section 6 for more informative details on the row components.

Partitions of other measures of association using orthogonal polynomials have also been considered. D'Ambra et al. [25] considered the partition of the Goodman-Kruskal tau index. For symmetrically associated multiple categorical random variables, Beh and Davy ([26, 27]) considered the partition of the Pearson chi-squared statistic, while for asymmetrically associated variables Beh et al. [28] considered the partition of the Marcotorchino index [29]. However, the application of extensions to hybrid decomposition will not be considered here.

3. Profile coordinates

One system of coordinates that could be used to visualize the association between the row and column categories is to plot along the kth axis $\{a_{ik}\}$ for the ith row and $\{b_k(j)\}$ for the jth-ordered column. Such coordinates are referred to as standard coordinates. These are analogous to the set of standard coordinates considered by Greenacre [14, page 93].

However, standard coordinates infer that each of the axes is given an equal weight of 1. Thus, while the difference within the row or column variables can be described by the difference between the points, they will not graphically depict the association between the rows and columns. Therefore, alternative plotting systems should be considered.

Analogous to the derivation of profile coordinates in Beh [8] using BMD, the row and column profile coordinates for singly ordered correspondence analysis are defined by

$$F = AZ, \tag{3.1}$$

$$G_* = B_* Z^T, \tag{3.2}$$

respectively. Therefore, by including the correlation quantities, the coordinates (3.1) and (3.2) will graphically depict the linear and nonlinear associations that may exist between the ordered column and nominal row categories.

The relationship between the row (and column) profile coordinates and the Pearson chi-squared statistic can be shown to be

$$X^2 = n \sum_{i=1}^{I} \sum_{v=1}^{J-1} p_{i\bullet} f_{iv}^2 = n \sum_{j=1}^{J} \sum_{u=1}^{M^*} p_{\bullet j} \left(g_{ju}^*\right)^2 \tag{3.3}$$

by substituting the elements of $F^T D_I F$ and $G_*^T D_J G_*$ into (2.17). However, instead of using the Pearson chi-squared statistic as a measure of association in a contingency table, correspondence analysis considers instead X^2/n, referred to as the total inertia. By adopting X^2/n as the measure of association, (3.3) shows that when the profile coordinates are situated close to the origin of the correspondence plot, X^2/n will be relatively small. Thus the hypothesis of independence between the rows and columns will be strong. Profile coordinates far from the origin indicate that the total inertia will be relatively large and the independence hypothesis becomes weak. These conclusions may also be verified by considering the Euclidean distance of a profile coordinate from the origin and other profile coordinates in the correspondence plot; refer to Section 4 for more details.

4. Distances

4.1. Distance from the origin

Consider the ith row profile. The squared Euclidean distance of this profile from the origin is

$$d_I^2(i,0) = \sum_{j=1}^{J} \frac{1}{p_{\bullet j}} \left(\frac{p_{ij}}{p_{i\bullet}} - p_{\bullet j} \right)^2. \tag{4.1}$$

It can be shown that by expressing this in terms of Pearson contingencies, and using (2.15) and (2.4), this distance may be expressed in terms of the sum of squares of the ith row profile coordinate such that

$$d_I^2(i,0) = \sum_{v=1}^{J-1} f_{iv}^2, \tag{4.2}$$

where f_{iv} is the (i,v)th element of F. By substituting (4.2) into (3.3), the Pearson chi-squared statistic can be expressed as

$$X^2 = n \sum_{i=1}^{I} p_{i\bullet} d_I^2(i,0). \tag{4.3}$$

Therefore, row profile coordinates close to the origin support the hypothesis of independence, while those situated far from the origin support its rejection. It can be shown in a similar manner that

$$X^2 = n \sum_{j=1}^{J} p_{\bullet j} d_J^2(j,0), \tag{4.4}$$

where

$$d_J^2(j,0) = \sum_{i=1}^{I} \frac{1}{p_{i\bullet}} \left(\frac{p_{ij}}{p_{\bullet j}} - p_{i\bullet} \right)^2 = \sum_{u=1}^{M^*} \left(g_{ju}^* \right)^2 \tag{4.5}$$

is the squared Euclidean distance of the jth column profile from the origin and g_{ju}^* is the (j,u)th element of (3.2).

4.2. Within variable distances

The squared Euclidean distance between two row profile coordinates, i and i', can be measured by

$$d_I^2(i,i') = \sum_{j=1}^{J} \frac{1}{p_{\bullet j}} \left(\frac{p_{ij}}{p_{i\bullet}} - \frac{p_{i'j}}{p_{i'\bullet}} \right)^2. \tag{4.6}$$

By considering the definition of the row profile coordinates given by (3.1), the squared Euclidean distance between these two profiles can be alternatively be written as

$$d_I^2(i, i') = \sum_{v=1}^{J-1} (f_{iv} - f_{i'v})^2.$$ (4.7)

Therefore, if two row profile coordinates have similar profile, their position in the correspondence plot will be very similar. This distance measure also shows that if two row categories have different profiles, then the position of their coordinates in the correspondence plot will lie at a distance from one another.

Similarly, the squared Euclidean distance between two column profiles, j and j', can be measured by

$$d_J^2(j, j') = \sum_{u=1}^{M^*} (g_{ju}^* - g_{j'u}^*)^2.$$ (4.8)

These results verify the property of distributional equivalence as stated by Lebart et al. [16, page 35], a necessary property for the meaningful interpretation of the distance of profiles in a correspondence plot.

(1) If two profiles having identical profiles are aggregated, then the distance between them remains unchanged.

(2) If two profiles having identical distribution profiles are aggregated, then the distance between them remains unchanged.

The interpretation of the distance between a particular row profile coordinate and a column profile coordinate is a contentious one and an issue that will not be described here, although a brief account is given by Beh [12, page 269].

5. Transition formula

For the classical approach to correspondence analysis, transition formulae allow for the profile coordinates of one variable to be calculated when the profile coordinates of a second variable are known.

To derive the transition formulae for a contingency table with ordered columns and nonordered rows, postmultiply the left- and right-hand sides of (3.1) by Z^T. Doing so leads to

$$FZ^T = AZZ^T = A(A^T P B_*)B_*^T D_J G_*,$$ (5.1)

upon substituting (2.16) and (3.2). Based on the orthogonality properties (2.4) and (2.8), the transition formula becomes

$$FZ^T = D_I^{-1} P G_*.$$ (5.2)

The transition formula (5.2) allows for the row profile coordinates to be calculated when the column profile coordinates are known.

In a similar manner, it can be shown that

$$G_* Z = D_J^{-1} P^T F. \tag{5.3}$$

Beh [30] provided a description of the transition formulae obtained for a doubly ordered correspondence analysis and the configuration of the points in the correspondence plot. For singly ordered correspondence analysis, similar descriptions can be obtained and are summarized in the following propositions.

(i) If the positions of the row profile coordinates are dominated by the first principal axis, then $Z_{(1)2} \approx 0$.

(ii) If the positions of the row profile coordinates are dominated by the second principal axis, then $Z_{(2)1} \approx 0$.

(iii) If the position of the column profile coordinates are dominated by the first principal axis, then $Z_{(2)1} \approx 0$.

(iv) If the positions of the column profile coordinates are dominated by the second principal axis, then $Z_{(1)2} \approx 0$.

However, it is still possible that $Z_{(1)2}$ and/or $Z_{(2)1}$ will be zero if none of the row and column profile coordinates lie along a particular axis. For such a case, it is not possible to determine when this will happen.

For both classical and doubly ordered correspondence analysis, when either the row or column profile positions is situated close to the origin of the correspondence plot, then there is no association between the rows and columns. This is indeed the case too for singly ordered correspondence analysis as indicated by (3.3). The items summarized above show that, in this case, $Z_{(1)2} \approx 0$ and $Z_{(2)1} \approx 0$. It can also be shown that $Z_{(1)1} \approx 0$ and $Z_{(2)2} \approx 0$.

6. Properties

The results above show that the mathematics and characteristics of this approach to singly ordered correspondence analysis are very similar to doubly ordered correspondence analysis and classical simple correspondence analysis. However, there are properties of the singly ordered approach that distinguish it from the other two techniques. This section provides an account of these properties.

Property 1. The row component associated with the mth principal axis is equivalent to the square of the mth largest singular value.

To show this, recall that the total inertia may be written in terms of bivariate moments and in terms of the eigenvalues such that

$$\frac{X^2}{n} = \sum_{u=1}^{M^*} \sum_{v=1}^{J-1} Z_{(u)v}^2 = \sum_{u=1}^{M^*} \lambda_u^2 \tag{6.1}$$

which can be obtained by equating the Pearson chi-squared partitions of (2.6) and (2.17). Therefore, the square of the mth singular value can be expressed by

$$\lambda_m^2 = \sum_{v=1}^{J-1} Z_{(m)v}^2, \tag{6.2}$$

where the right-hand side of (6.2) is just the mth-order row component. For example, the square of the largest singular value may be partitioned so that

$$\lambda_1^2 = Z_{(1)1}^2 + Z_{(1)2}^2 + \cdots + Z_{(1)J-1}^2. \tag{6.3}$$

Therefore, the singly ordered correspondence approach using the hybrid decomposition of (2.16) and (2.17) allows for a partition of the singular values of the Pearson contingencies into components that reflect variation in the row categories in terms of location, dispersion, and higher-order moments. That is, each singular value can be partitioned so that information associated with differences in the mean and spread of the row profiles can be identified. Higher-order moments can also be determined from such a partition.

Property 2. The row component values are arranged in descending order.

This property follows directly from **Property 1**. Since the eigenvalues are arranged in a descending order, so too are the row components.

Property 3. A singly ordered correspondence analysis allows for the inertia associated with a particular axis of a simple correspondence plot (called the principal inertia) to be partitioned in bivariate moments.

Again, this property follows directly from **Property 1**, where the principal inertia of the mth axis is the sum of squares of the bivariate moments when $u = m$.

Property 4. It is possible to identify which bivariate moment contributes the most to a particular squared singular value and hence its associated principal axis.

This is readily seen from **Property 3**.

For classical correspondence analysis, the axes are constructed so that the first axis accounts for most of the information in variation in the categories, the second axis describes accounts for the second most amount of variation, and so on. However, it is unclear what this variation is, or whether it is easily identified as being statistically significant. By considering the partition of the singular values, as described by (6.2), the user is able to isolate important bivariate moments that include variation in terms of location, dispersion, and higher-order components for each principal axis. Therefore, there is more information that is able to be obtained from the axes of the correspondence plot, and the proximity of the points on it, than from a classical correspondence plot.

7. Example

Consider the contingency table given by **Table 1** which was originally seen in Calimlin et al. [18] and analyzed by Beh [11]. The study was aimed at testing four analgesic drugs (randomly assigned the labels A, B, C, and D) and their effect on 121 hospital patients. The patients were given a five-point scale consisting of the categories poor, fair, good, very good, and excellent on which to make their decision.

If only a comparison of the drugs, in terms of the mean value and spread across the different levels of effectiveness, was of interest, attention would be focused on the quantities $\mu_J(i)$ (and $\sigma_J(i)$). These values for Drug A, Drug B, Drug C, and Drug D are 3.3000 (1.2949),

Table 1: Cross-classification of 121 hospital patients according to analgesic drug and its effect.

	Poor	Fair	Good	Very Good	Excellent	Total
Drug A	5	1	10	8	6	30
Drug B	5	3	3	8	12	31
Drug C	10	6	12	3	0	31
Drug D	7	12	8	1	1	29
Total	27	21	33	20	19	121

The column header "Analgesic drug effect" spans the Poor, Fair, Good, Very Good, and Excellent columns.

3.6129 (1.4740), 2.2581 (1.0149), and 2.2069 (0.9606), respectively and were calculated using natural scores for the column categories. Therefore, based on these quantities, it is clear that Drug A and Drug B are very similar in terms of the two components across the different levels of effectiveness. Therefore, these two drugs have a similar effect on the patients. Also, these drugs are different to Drug C and Drug D which are themselves quite similar in effectiveness. However, the association between the Drugs and the different levels of effectiveness is not evident from such measures. This is why correspondence analysis is a suitable analytical tool to graphically depict and summarize the association. It can be seen that Table 1 consists of ordered column categories and nonordered row categories. Therefore, singly ordered correspondence analysis will be used to analyze the effectiveness of the drugs.

The Pearson chi-squared statistic of Table 1 is 47.0712, and with a zero p-value, it is highly statistically significant. Therefore, with a total inertia of 0.3890, there is a significant association between the drugs used and their effect on the patients.

When a classical correspondence analysis is applied, the squared singular values are $\lambda_1^2 = 0.30467$, $\lambda_2^2 = 0.07734$, and $\lambda_3^2 = 0.00701$ and the two-dimensional correspondence plot is given by Figure 1. Here, the first principal axis accounts for $0.30467/0.3890 \times 100 = 78.3\%$ of the total association between the two variables, and the second axis accounts for 19.9%. Therefore, the two-dimensional plot of Figure 1 graphically depicts 98.2% of the association that exists between the analgesic drug being tested and its level of effectiveness.

Figure 1 shows a clear association between the analgesic drug being tested and the effectiveness of that drug. Drug B appears to have an "excellent" effect on the patients that participated in the study, Drug A was rated as "very good," Drug D was deemed only "fair" in its effectiveness and Drug C was judged "good" to "poor." These conclusions are also apparent when eyeballing the cell frequencies of Table 1. However, it is unclear how the profile of each of the four drugs is different, or where they may be similar. By adopting the methodology above, we can determine how these comparisons may be made in terms of differences in location, dispersion, and higher-order components.

The component values that are associated with explaining the variation in the position of the drug coordinates in Figure 2 are $\sum_m Z_{(m)1}^2 = 0.21034$, $\sum_m Z_{(m)2}^2 = 0.08418$, $\sum_m Z_{(m)3}^2 = 0.07268$, and $\sum_m Z_{(m)4}^2 = 0.02452$. Therefore, Figure 2 is constructed using the first (linear) principal axis with a principal inertia value of $\sqrt{0.21034} = 0.45863$, and the second (dispersion) principal axis with a principal inertia value of $\sqrt{0.08148} = 0.28545$, for the four drugs. Together, these two axes contribute to 75% of the variation of the drugs tested, compared with 98.2% of the variation in the patients judgement of the drug. The third (cubic) component contributes to 18.7% of this variation.
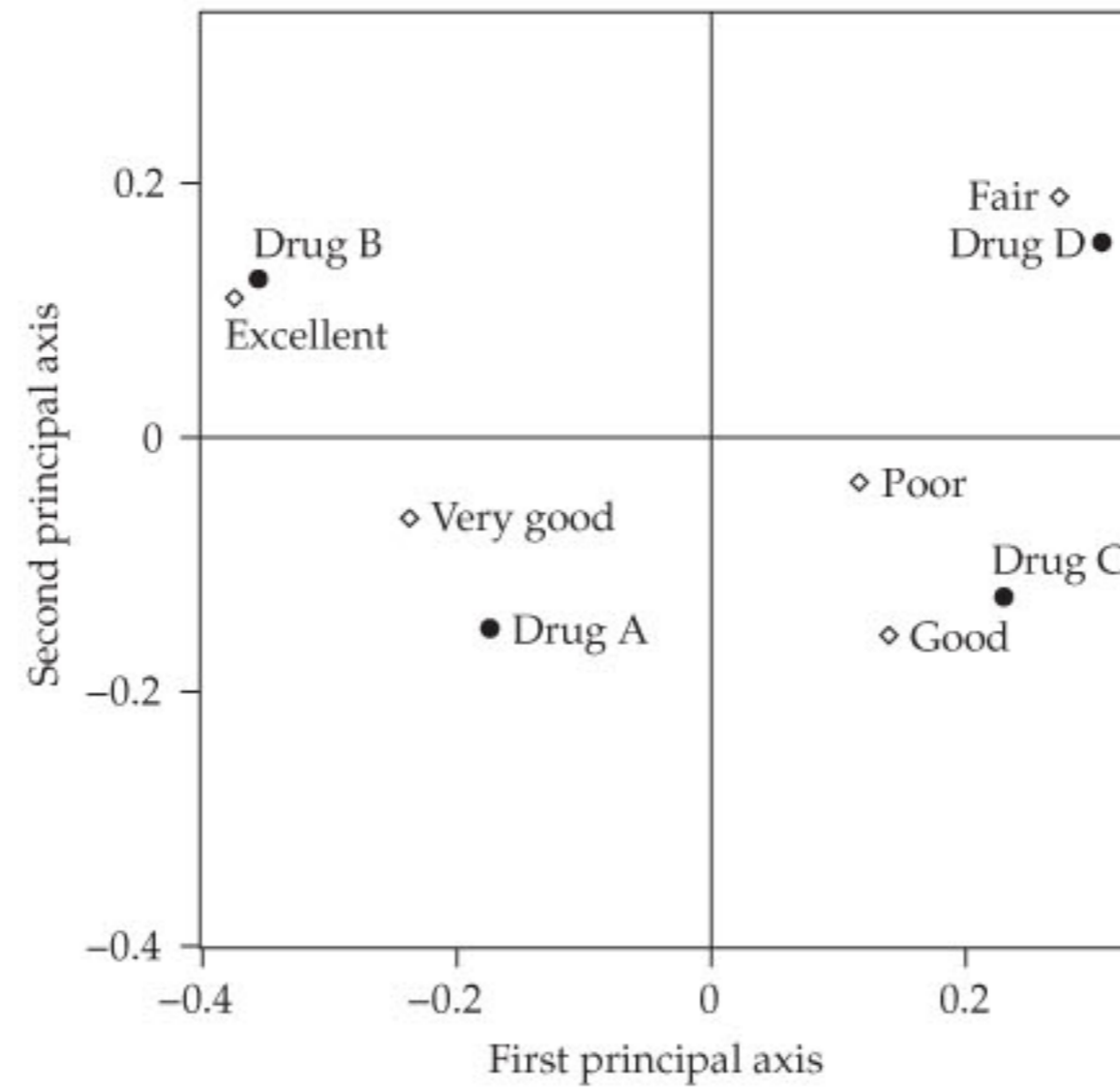
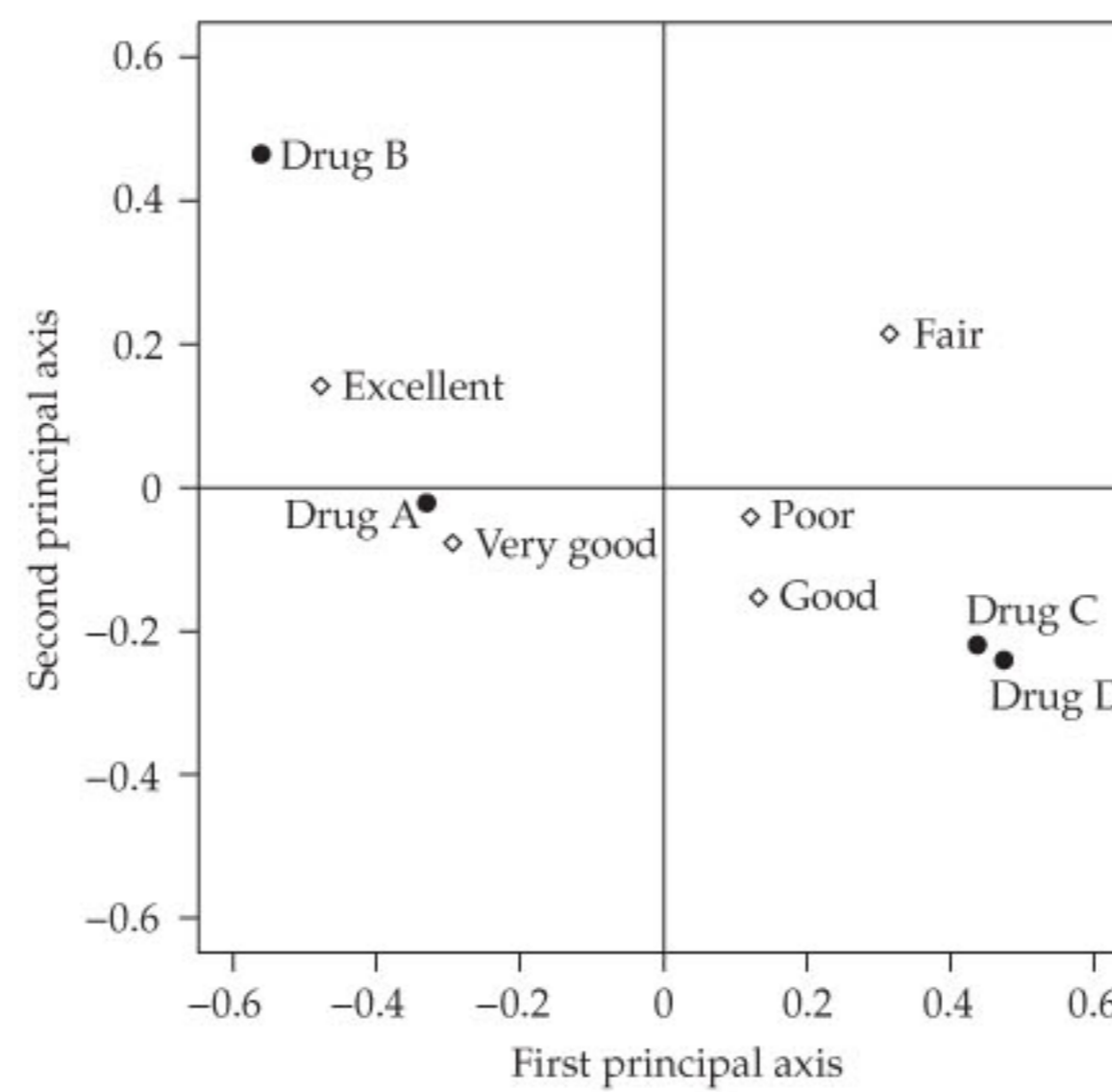**Figure 1:** Classical correspondence plot of Table 1.

Figure 2: Singly ordered correspondence plot of Table 1.

Applying singly ordered correspondence analysis yields $Z_{(1)1} = -0.45648$ and $Z_{(1)2} = -0.26016$. Also, $Z_{(1)3} = 0.16505$ and $Z_{(1)4} = 0.03696$. Therefore, by considering (6.3), we can see that

$$0.3047 = (-0.4565)^2 + (-0.2602)^2 + (0.1651)^2 + (0.0370)^2. \tag{7.1}$$

That is, the dominant source of the first (squared) singular value is due to the linear component of the effectiveness of the drugs. Thus, the location component best describes the variation of the profiles for the drug effectiveness levels along the first principal axis of **Figure 1** (68.4%).

Table 2: Contribution of the drugs tested to each axis of Figure 2.

Drug tested	Principal axis 1		Principal axis 2	
	Contr'n	% Contr'n	Contr'n	% Contr'n
Drug A	0.02705	12.86	0.00011	0.14
Drug B	0.08053	38.29	0.05524	67.79
Drug C	0.04884	23.22	0.01229	15.08
Drug D	0.05392	25.63	0.01384	16.99
Total	0.21034	100	0.08148	100

Table 3: Contribution of the effectiveness of the drugs tested to each axis of Figure 2.

Rating	Principal axis 1		Principal axis 2	
	Contr'n	% Contr'n	Contr'n	% Contr'n
Poor	0.01356	4.46	0.00124	1.60
Fair	0.07502	24.62	0.03576	46.23
Good	0.01953	6.41	0.02432	31.44
Very good	0.05615	18.43	0.00406	5.25
Excellent	0.14040	46.08	0.01197	15.48
Total	0.30466	100	0.07735	100

Figure 2 shows the variation of these drugs in terms of the linear and quadratic components. While Figure 1 indicates that the effectiveness of Drug C and Drug D is different, Figure 2 shows that the positions of Drug C and Drug D are similar across the column responses. This is because the variation between the two drugs exists at moments higher than the dispersion. It is also evident from Figure 2 that these two drugs have quite a different effect than do Drug A and Drug B, which in themselves are different. These conclusions are in agreement with the comments made earlier in the example. Figure 2 also shows that by taking into account the ordinal nature of the column categories, the variation between the drug effectiveness levels may be explored. For example, "good" and "poor" share the very similar first principal coordinate. However, there is slightly more variation (across the drugs) for "good" than there is for "poor."

An important feature of Figure 2 is that it depicts the association between the drugs and the levels of effectiveness. It can be seen from Figure 2, just as Figure 1 concluded, that Drug A and Drug B are more effective in treating pain relief than Drug C and Drug D. However, because of the use of hybrid decomposition, the position of the drug profile coordinates have changed. Figure 1 concluded that Drug D was rated as "fair." This is primarily due to the relatively large cell frequency (with a value of 12) that the two categories share; this feature is a common characteristic of classical correspondence analysis. However, since the drug behaves in a similar manner (in terms of location and spread) when compared with Drug C, its position has shifted to the bottom right quadrant of the plot. Therefore, Drug D is associated more with "poor" and "good" when focusing on these components of the category.

By observing the distance of each category from the origin in Figure 2, Drug B is the furthest away from the origin and so is less likely than the other drugs to contribute to the independence between the drugs and the patients effect. This is because Drug B contributes more

to the row location component (38.29%) than any of the other three drugs in the study, while contributing to 67.79% of the variation in the dispersion component. Further results on the dominance of the drugs to each of the axes in Figure 2 are summarized in Table 2. It shows the contribution, and relative contribution of each drug to each of the two axes. Table 3 provides a similar summary, but for the different effectiveness levels of the drugs.

Recall that Drug C and Drug D are positioned close to one another in Figure 2. Table 2 shows that they contribute roughly the same to the location and dispersion components. Figure 2 also shows that "excellent" is the most dominant of the drug effectiveness categories along the first principal axis and this is reflected in Table 3, accounting for nearly half (46.08%) of the principal inertia for its variable. The second principal axis is dominated by the category "fair" which contributes to 46.23% of the second principal inertia.

8. Discussion

Correspondence analysis has become a very popular method for analyzing categorical data, and has been shown to be applicable in a large number of disciplines. It has long been applied in the analysis of ecological disciplines, and recently in health care and nursing studies [31, 32], environmental management [33], and linguistics [34, 35]. It also has developed into an analytic tool which can handle many data structures of different types such as ranked data [30], time series data [36], and cohort data [37].

The aim of this paper has been to discuss new developments of correspondence analysis for the application to singly ordered two-way contingency tables. Applications of the classical approach to correspondence analysis can be made, although the ordered structure of the variables is often not always reflected in the output. When a two-way table consists of one ordered variable, such as in sociological or health studies where responses are rated according to a Likert scale, the ordinal structure of this variable needs to be considered. The singly ordered correspondence analysis procedure developed by Beh [8] is applicable to singly ordered contingency tables. However, due to the nature of this procedure, only a visualization of the association between the categories of the nonordered variable can be made. Therefore, any between-variable interpretation is not possible. The technique developed in this paper improves upon this singly ordered approach by allowing for the simultaneous representation of the ordered column and nonordered row categories.

References

[1] A. Agresti, *Analysis of Ordinal Categorical Data*, John Wiley & Sons, New York, NY, USA, 1984.

[2] L. A. Goodman, "The analysis of cross-classified data having ordered and/or unordered categories: association models, correlation models, and asymmetry models for contingency tables with or without missing entries," *The Annals of Statistics*, vol. 13, no. 1, pp. 10–69, 1985.

[3] S. J. Haberman, "Log-linear models for frequency tables with ordered classifications," *Biometrics*, vol. 30, no. 4, pp. 589–600, 1974.

[4] A. R. Parsa and W. B. Smith, "Scoring under ordered constraints in contingency tables," *Communications in Statistics. Theory and Methods*, vol. 22, no. 12, pp. 3537–3551, 1993.

[5] Y. Ritov and Z. Gilula, "Analysis of contingency tables by correspondence models subject to order constraints," *Journal of the American Statistical Association*, vol. 88, no. 424, pp. 1380–1387, 1993.

[6] B. F. Schriever, "Scaling of order dependent categorical variables with correspondence analysis," *International Statistical Review*, vol. 51, no. 3, pp. 225–238, 1983.

[7] K.-S. Yang and M.-H. Huh, "Correspondence analysis of two-way contingency tables with ordered column categories," *Journal of the Korean Statistical Society*, vol. 28, no. 3, pp. 347–358, 1999.

[8] E. J. Beh, "Simple correspondence analysis of ordinal cross-classifications using orthogonal polynomials," *Biometrical Journal*, vol. 39, no. 5, pp. 589–613, 1997.

[9] D. J. Best and J. C. W. Rayner, "Analysis of ordinal contingency tables via orthogonal polynomials," Department of Applied Statistics Preprint, University of Wollongong, Australia, 1994.

[10] J. C. W. Rayner and D. J. Best, "Analysis of singly ordered two-way contingency tables," *Journal of Applied Mathematics and Decision Sciences*, vol. 4, no. 1, pp. 83–98, 2000.

[11] E. J. Beh, "Partitioning Pearson's chi-squared statistic for singly ordered two-way contingency tables," *Australian & New Zealand Journal of Statistics*, vol. 43, no. 3, pp. 327–333, 2001.

[12] E. J. Beh, "Simple correspondence analysis: a bibliographic review," *International Statistical Review*, vol. 72, no. 2, pp. 257–284, 2004.

[13] J.-P. Benzécri, *Correspondence Analysis Handbook*, vol. 125 of *Statistics: Textbooks and Monographs*, Marcel Dekker, New York, NY, USA, 1992.

[14] M. J. Greenacre, *Theory and Applications of Correspondence Analysis*, Academic Press, London, UK, 1984.

[15] D. L. Hoffman and G. R. Franke, "Correspondence analysis: graphical representation of categorical data in marketing research," *The American Statistician*, vol. 23, pp. 213–227, 1986.

[16] L. Lebart, A. Morineau, and K. M. Warwick, *Multivariate Descriptive Statistical Analysis*, John Wiley & Sons, New York, NY, USA, 1984.

[17] R. Lombardo, E. J. Beh, and L. D'Ambra, "Non-symmetric correspondence analysis with ordinal variables using orthogonal polynomials," *Computational Statistics & Data Analysis*, vol. 52, no. 1, pp. 566–577, 2007.

[18] J. F. Calimlin, W. M. Wardell, C. Cox, L. Lasagna, and K. Sriwatanakul, "Analgesic efficiency of orally Zomipirac sodium," *Clinical Pharmacology Therapy*, vol. 31, p. 208, 1982.

[19] L. A. Goodman, "A single general method for the analysis of cross-classified data: reconciliation and synthesis of some methods of Pearson, Yule, and Fisher, and also some methods of correspondence analysis and association analysis," *Journal of the American Statistical Association*, vol. 91, no. 433, pp. 408–428, 1996.

[20] P. L. Emerson, "Numerical construction of orthogonal polynomials from a general recurrence formula," *Biometrics*, vol. 24, no. 3, pp. 696–701, 1968.

[21] E. J. Beh, "A comparative study of scores for correspondence analysis with ordered categories," *Biometrical Journal*, vol. 40, no. 4, pp. 413–429, 1998.

[22] D. J. Best and J. C. W. Rayner, "Nonparametric analysis for doubly ordered two-way contingency tables," *Biometrics*, vol. 52, no. 3, pp. 1153–1156, 1996.

[23] J. C. W. Rayner and D. J. Best, "Smooth extensions of Pearsons's product moment correlation and Spearman's rho," *Statistics & Probability Letters*, vol. 30, no. 2, pp. 171–177, 1996.

[24] P. J. Davy, J. C. W. Rayner, and E. J. Beh, "Generalised correlations and Simpson's paradox," in *Current Research in Modelling, Data Mining and Quantitative Techniques*, V. Pemajayantha, R. W. Mellor, S. Peiris, and J. R. Rajasekera, Eds., pp. 63–73, University of Western Sydney, Sydney, Australia, 2003.

[25] L. D'Ambra, E. J. Beh, and P. Amenta, "Catanova for two-way contingency tables with ordinal variables using orthogonal polynomials," *Communications in Statistics. Theory and Methods*, vol. 34, no. 8, pp. 1755–1769, 2005.

[26] E. J. Beh and P. J. Davy, "Partitioning Pearson's chi-squared statistic for a completely ordered three-way contingency table," *Australian & New Zealand Journal of Statistics*, vol. 40, no. 4, pp. 465–477, 1998.

[27] E. J. Beh and P. J. Davy, "Partitioning Pearson's chi-squared statistic for a partially ordered three-way contingency table," *Australian & New Zealand Journal of Statistics*, vol. 41, no. 2, pp. 233–246, 1999.

[28] E. J. Beh, B. Simonetti, and L. D'Ambra, "Partitioning a non-symmetric measure of association for three-way contingency tables," *Journal of Multivariate Analysis*, vol. 98, no. 7, pp. 1391–1441, 2007.

[29] F. Marcotorchino, "Utilisation des Comparaisons par Paires en Statistique des Contingences: Partie III," Tech. Rep. # F-081, IBM, Paris, France, 1985.

[30] E. J. Beh, "Correspondence analysis of ranked data," *Communications in Statistics. Theory and Methods*, vol. 28, no. 7, pp. 1511–1533, 1999.

[31] R. Javalgi, T. Whipple, M. McManamon, and V. Edick, "Hospital image: a correspondence analysis approach," *Journal of Health Care Marketing*, vol. 12, pp. 34–41, 1992.

[32] D. D. Watts, "Correspondence analysis: a graphical technique for examining categorical data," *Nursing Research*, vol. 46, no. 4, pp. 235–239, 1997.

[33] H. Kishino, K. Hanyu, H. Yamashita, and C. Hayashi, "Correspondence analysis of paper recycling society: consumers and paper makers in Japan," *Resources, Conservation and Recycling*, vol. 23, no. 4, pp. 193–208, 1998.

[34] P. J. Hassall and S. Ganesh, "Correspondence analysis of English as an international language," *The New Zealand Statistician*, vol. 31, pp. 24–33, 1996.

[35] A. K. Romney, C. C. Moore, and C. D. Rusch, "Cultural universals: measuring the semantic structure of emotion terms in English and Japanese," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 10, pp. 5489–5494, 1997.

[36] J.-C. Deville and C.-E. Särndal, "Calibration estimators in survey sampling," *Journal of the American Statistical Association*, vol. 87, no. 418, pp. 376–382, 1992.

[37] M. Grassi and S. Visentin, "Correspondence analysis applied to grouped cohort data," *Statistics in Medicine*, vol. 13, no. 23-24, pp. 2407–2425, 1994.