Journal of the Korean Statistical Society (1999), 28: 3, pp 347–358

Correspondence Analysis of Two-way Contingency Tables with Ordered Column Categories

Kyung-Sook Yang[1] and Myung-Hoe Huh[2]

ABSTRACT

Correspondence analysis is an exploratory method for two-way contingency tables intended to display the association pattern between row and column categories. It has been developed for several decades mainly in France and Japan and, nowadays, is popular worldwide. For the special case, however, that the column consists of ordered categories, correspondence analysis may yield ackward result so that it cannot be fully accommodated. That's because the row and column categories are assumed to be nominal in the ordinary correspondence analysis.

The aim of this study is to develop the correspondence analysis for two-way contingency tables with ordered column categories. It is based on two specific algorithms, "optimal scaling" for the first principal axis and "partial scaling" for the remaining principal axes, that we propose in Sections 2 and 3. As by-product, matrix decomposition for two-way tables is produced in Section 4. A numerical illustration is given and discussed in Section 5.

Keywords: Correspondence Analysis; Ordered Categories; Dual Scaling; Optimal Scaling; Partial Scaling; Goodness-of-Approximation; Quantification Plots.

1. INTRODUCTION

Correspondence analysis was conceived in as early as 1930's and developed from late 1950's mainly by French researchers under the name "l'analyse des correspondances" and by Japanese scholars under the different name "Hayashi's quantification III"(Hayashi 1952). Also, there appeared mathematically equivalent forms of correspondence analysis for two-way contingency tables elsewhere, such as "homogeneity analysis" in the Netherlands and "dual scaling" in Canada (Greenacre 1984, Nishisato 1980).

[1]Institute of Statistics, Korea University, Seoul 136-701, Korea.
[2]Department of Statistics, Korea University, Seoul 136-701, Korea.
E-mail: stat420@kuccnx.korea.ac.kr

For the ordinary correspondence analysis of two-way contingency tables, we use the following notations:

$F = \{f_{ij}\}$: two-way contingency table with n rows and q columns.

$f_{i+} = \sum_{j=1}^{q} f_{ij}, \quad i = 1, \cdots, n$: row margins.

$f_{+j} = \sum_{i=1}^{n} f_{ij}, \quad j = 1, \cdots, q$: column margins.

$N = \sum_{i=1}^{n} \sum_{j=1}^{q} f_{ij}$: total frequency or sample size.

We assume that the row denotes subgroups one of which N units are classified into, and that the column denotes categorical characteristics or responses of each unit. Then, the multinomial composition of the i-th subgroup is represented by row profile vector of size q

$$\boldsymbol{a}_i = (f_{i1}, \cdots, f_{ij}, \cdots, f_{iq})^t / f_{i+}, \quad i = 1, \cdots, n$$

of which q components sum to 1. Hence row profile vectors $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_i, \cdots, \boldsymbol{a}_n$ lie on the $(q-1)$-dimensional simplex embedded in R^q. For the homogeneity of n row profile vectors $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_i, \cdots, \boldsymbol{a}_n$, we may define the index "phi"

$$\phi = \sum_{i=1}^{n} r_i \ \Big(\sum_{j=1}^{q} (a_{ij} - a_{\cdot j})^2 / c_j \Big),$$

where $r_1, \cdots, r_i, \cdots, r_n$ are weights for the rows, $c_1, \cdots, c_j, \cdots, c_q$ are scales for the columns, both of which are to be clarified shortly, $a_{i1}, \cdots, a_{iq}$ are elements of $\boldsymbol{a}_i$ and

$$a_{\cdot j} = \sum_{i=1}^{n} r_i \, a_{ij} / \sum_{i=1}^{n} r_i, \quad j = 1, \cdots, q,$$

the weighted average of a_{ij}'s for $i = 1, \cdots, n$.

For row weights $r_i (i = 1, \cdots, n)$ and column scales $c_j (j = 1, \cdots, q)$, it is natural to set

$$r_i = f_{i+}/N, \quad i = 1, \cdots, n; \quad c_j = f_{+j}/N, \quad j = 1, \cdots, q,$$

as suggested by Lebart et al. (1984) and Greenacre and Hastie (1987). One of the attractive properties by such choice is that Pearson's chi-square statistic

$$\chi^2 = \sum_{i=1}^{n} \sum_{j=1}^{q} \frac{(f_{ij} - f_{i+} f_{+j}/N)^2}{f_{i+} f_{+j}/N},$$

is expressed as

$$\chi^2 = N\phi \ \Big(= N \sum_{i=1}^{n} \sum_{j=1}^{q} \frac{f_{i+}}{N} \frac{(f_{ij}/f_{i+} - f_{+j}/N)^2}{f_{+j}/N} \Big).$$

Therefore, Pearson's χ^2 is a measure of dispersion of row profile vectors $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_i, \cdots, \boldsymbol{a}_n$ from their centroid $\boldsymbol{c} = (c_1, \cdots, c_j, \cdots, c_q)^t$, where

$$\boldsymbol{c} = \boldsymbol{a}_{\cdot} = (a_{\cdot 1}, \cdots, a_{\cdot j}, \cdots, a_{\cdot q})^t.$$

More explicitly,

$$\chi^2/N = \sum_{i=1}^{n} r_i (\boldsymbol{a}_i - \boldsymbol{c})^t D_c^{-1} (\boldsymbol{a}_i - \boldsymbol{c}),$$

where $D_c = \mathrm{diag}(c_1, \cdots, c_j, \cdots, c_q)$ is the $q \times q$ diagonal matrix of column scales (Greenacre and Hastie 1987).

As mentioned before, correspondence analysis can be derived in several modes. In this study, we follow two modes - linear projection and dual scaling. Linear projection method can be formulated as (Lebart et al. 1984, pp. 30-44; Greenacre 1994, pp. 24-31)

1) Project centered row profile vectors $\boldsymbol{a}_i - \boldsymbol{c}$ $(i = 1, \cdots, n)$ on unit-normed $\boldsymbol{v}$ in the scaled Euclidean space, e.g. $\boldsymbol{v}^t D_c^{-1} \boldsymbol{v} = 1$.
2) Maximize the weighted sum of squared projections with respect to $\boldsymbol{v}$, with weights r_i $(i = 1, \cdots, n)$ to the row profiles:

$$max_{\boldsymbol{v}} \parallel B D_c^{-1} \boldsymbol{v} \parallel_{D_r}^2 \quad \text{subject to} \quad \boldsymbol{v}^t D_c^{-1} \boldsymbol{v} = 1 \ ,$$

where $\parallel \boldsymbol{z} \parallel_D^2 \ \equiv \ \boldsymbol{z}^t D \boldsymbol{z}$ and B is the $n \times q$ matrix of centered row profiles:

$$B = \begin{pmatrix} \boldsymbol{a}_1^t & - & \boldsymbol{c}^t \\ & \vdots & \\ \boldsymbol{a}_i^t & - & \boldsymbol{c}^t \\ & \vdots & \\ \boldsymbol{a}_n^t & - & \boldsymbol{c}^t \end{pmatrix} = D_r^{-1}(F/N - \boldsymbol{r}\boldsymbol{c}^t).$$

3) Display the row subgroups and column categories in a lower-dimensional plane.

It turns out that the scaled principal direction vectors $\tilde{\boldsymbol{v}} (= D_c^{-1/2} \boldsymbol{v})$ can be obtained via the eigensystem of $G^t G$, where

$$G = D_r^{1/2} \ B \ D_c^{-1/2} \ .$$

Then, the standard column scores $\boldsymbol{y}$ such that $\boldsymbol{c}^t \boldsymbol{y} = 0$ and $\boldsymbol{y}^t D_c \boldsymbol{y} = 1$ are obtained by

$$\boldsymbol{y} = D_c^{-1/2} \tilde{\boldsymbol{v}} = D_c^{-1} \boldsymbol{v} \ ,$$

And, the principal row scores are given by

$$\boldsymbol{x} = B\boldsymbol{y} \ .$$

Principal row score vector $\boldsymbol{x}$ satisfies $\boldsymbol{r}^t\boldsymbol{x} = 0$ and $\boldsymbol{x}^t D_r \boldsymbol{x} = \lambda$, an eigenvalue of G^tG.

Dual scaling method due to Nishisato(1980) is rather numerical while the linear projection method is geometrical. The steps of dual scaling for principal row scores $\boldsymbol{x}$ and standard column scores $\boldsymbol{y}$ are

0) Assign standard scores $\boldsymbol{y} : q \times 1$ $(\boldsymbol{c}^t\boldsymbol{y} = 0, \ \boldsymbol{y}^t D_c \boldsymbol{y} = 1)$ to the column categories.

1) Then, row subgroups are quantified by principal scores

$$\boldsymbol{x} = B\boldsymbol{y} = D_r^{-1}(F/N)\boldsymbol{y}. \tag{1.1}$$

2) Now, the standard scores $\boldsymbol{y}$ to column categories are adjusted by

$$\boldsymbol{y} \propto D_c^{-1}(F^t/N)\boldsymbol{x}, \quad \text{or} \quad \boldsymbol{y} = D_c^{-1}B^t D_r \boldsymbol{x}/\lambda, \tag{1.2}$$

where

$$\lambda^2 = \boldsymbol{x}^t D_r B D_c^{-1} B^t D_r \boldsymbol{x}.$$

3) Repeat step 1 and step 2 until the principal inertia of $\boldsymbol{x}$, $\boldsymbol{x}^t D_r \boldsymbol{x}$, does not increase further.

It is well known that the two methods yield the same results (Nishisato 1980, pp. 54-73): The standard column scores $\boldsymbol{y}$ and the principal row scores $\boldsymbol{x}$ are given by

$$\boldsymbol{y} = D_c^{-1/2}\tilde{\boldsymbol{v}}, \quad \boldsymbol{x} = B\boldsymbol{y} \ ,$$

where $\tilde{\boldsymbol{v}}$ is an eigenvector of G^tG.

Despite the well-established theory and accumulated experience in empirical research, correspondence analysis may not be acceptable in certain situations when applied to the two-way contingency tables with ordered column categories. That's because the ordinary correspondence analysis does not incorporate the data characteristic such as ordinal nature of the column into the data analysis.

The aim of this study is to develop the special correspondence analysis for two-way contingency tables with ordered column categories. In Sections 2 and 3, the dual scaling method supplemented by linear projections will be proposed to produce order preserving first principal axis for column categories and complementary principal axes for recovery of full information in the two-way contingency table. In Section 4, a special matrix decomposition formula for two-way contingency tables will be produced. In Sections 5 and 6, a numerical example will be given with discussions followed by concluding remarks.

2. OPTIMAL SCALING OF THE FIRST PRINCIPAL AXIS

When column categories of the two-way contingency table are ordered, standard column scores $y_1, \cdots, y_q$, elements of $\boldsymbol{y}$, of the first principal axis should be in compliance with the constraint

$$y_1 \leq \cdots \leq y_q \quad \text{or} \quad y_1 \geq \cdots \geq y_q \,. \tag{2.1}$$

We propose the following algorithm for the optimal scaling of the first principal axis that preserves the order in column categories.

Algorithm OS (Algorithm for the optimal scaling of the first principal axis):

OS-0. Calculate principal row scores $\boldsymbol{x}$ and standard column scores $\boldsymbol{y}$ from the ordinary correspondence analysis of F.

OS-1. If $\boldsymbol{y} = (y_1, \cdots, y_q)^t$ satisfies (2.1), then put $\boldsymbol{x}^{(1)} := \boldsymbol{x}$, $\boldsymbol{y}^{(1)} := \boldsymbol{y}$ and stop. Otherwise, apply the weighted least squares monotone fitting: That is, wherever the ordinal compliance is broken, replace the adjacent scores by their weighted average (with weights proportional to respective marginal frequencies), obtaining $\boldsymbol{y}^\dagger$.

OS-1′. Normalize $\boldsymbol{y}^\dagger$ by applying $\boldsymbol{y}^\dagger := \boldsymbol{y}^\dagger / \sqrt{\boldsymbol{y}^{\dagger t} D_c \boldsymbol{y}^\dagger}$.

OS-2. Substitute $\boldsymbol{y}^\dagger$ for $\boldsymbol{y}$ in (1.1) and compute the principal row scores $\boldsymbol{x}^\dagger$:

$$\boldsymbol{x}^\dagger = B\boldsymbol{y}^\dagger \,.$$

OS-2′. Substitute $\boldsymbol{x}^\dagger$ for $\boldsymbol{x}$ in (1.2) and compute the standard column scores $\boldsymbol{y}^\dagger$:

$$\boldsymbol{y}^* = D_c^{-1} B^t D_r \boldsymbol{x}^\dagger / \lambda^\dagger,$$

where $\lambda^\dagger = \sqrt{\boldsymbol{x}^{\dagger t} D_r B D_c^{-1} B^t D_r \boldsymbol{x}^{\dagger t}}$.

OS-3. Update the standard column scores $\boldsymbol{y}$ by $\boldsymbol{y}^\dagger$.

OS-4. Repeat Steps OS-1 to OS-3 until the principal inertia of $\boldsymbol{x}, \boldsymbol{x}^{\dagger t} D_r \boldsymbol{x}^\dagger$, does not increase and the score vector $\boldsymbol{y}$ changes little. Finally, quit the algorithm after setting

$$\boldsymbol{x}^{(1)} := \boldsymbol{x}^\dagger, \quad \boldsymbol{y}^{(1)} := \boldsymbol{y}^\dagger \ \ (\text{not } \boldsymbol{y}^*),$$

so that the first-axis column scores $\boldsymbol{y}^{(1)}$ satisfy the constraint (2.1).

Algorithm OS is not the first one for optimal scaling of ordered categories. Nishisato and Arri (1975), among others, solved the same problem and proposed

an intuitively appealing algorithm, "Successive Data Modification", which seems basically equivalent to ours. See Nishisato (1980, Section 8.1). But, the main advantage of the Algorithm OS is that it pursues the same geometrical approach as the ordinary correspondence analysis reviewed in Section 1, while Nishisato and Arri's orientation is rather numerical.

3. PARTIAL SCALING OF THE REMAINING AXES

Correspondence analysis, like principal component analysis, is a dimension reduction technique. So, fixing the first principal axis on the $(q-1)$-dimensional simplex, we may pursue further exploration for the remaining principal axes to view multinomial row profiles more completely. For principal component analysis, the second author attacked the similar problem (Suh and Huh 1997).

Since $\boldsymbol{y}^{(1)}$ was obtained from Algorithm OS and

$$\boldsymbol{y}^{(1)} = D_c^{-1/2}\tilde{\boldsymbol{v}}^{(1)} = D_c^{-1}\boldsymbol{v}^{(1)} ,$$

the principal direction vector $\boldsymbol{v}^{(1)}$ in R^q is determined. Then, by projecting the rows of B on $\boldsymbol{v}^{(1)}$, we obtain the "fits" and the "residuals": The fitted matrix $\hat{B}^{(1)}$ and the residual matrix $B^{(2)}$ are, respectively,

$$\hat{B}^{(1)} = BD_c^{-1}\boldsymbol{v}^{(1)}\boldsymbol{v}^{(1)t}, \quad B^{(2)} = B - \hat{B}^{(1)}.$$

Now, to look for the second principal direction vector $\boldsymbol{v}^{(2)}$ that maximizes the sum of squares of projections of the rows of $B^{(2)}$ on it, we formulate

$$max_{\boldsymbol{v}^{(2)}} \parallel B^{(2)}D_c^{-1}\boldsymbol{v}^{(2)} \parallel^2_{D_r} \quad \text{subject to} \quad \boldsymbol{v}^{(2)t}D_c^{-1}\boldsymbol{v}^{(2)} = 1 .$$

Similarly, we may pursue the third, the fourth, and all the remaining principal axes to have nothing left in the matrix B of centered row profiles, leading to the following algorithm.

Algorithm PS (Algorithm for the partial scaling of the remaining axes):

PS-1. Fit $B(\equiv B^{(1)})$ by $\hat{B}^{(1)}$, yielding the residual $B^{(2)}$: $B^{(1)} = \hat{B}^{(1)} + B^{(2)}$. And set $k = 2$.

PS-2. Fit $B^{(k)}$ by $\hat{B}^{(k)}$, yielding the residual $B^{(k+1)}$: $B^{(k)} = \hat{B}^{(k)} + B^{(k+1)}$, where

$$\hat{B}^{(k)} = B^{(k)}D_c^{-1}\boldsymbol{v}^{(k)}\boldsymbol{v}^{(k)t} = B^{(k)}D_c^{-1/2}\tilde{\boldsymbol{v}}^{(k)}\tilde{\boldsymbol{v}}^{(k)t}D_c^{1/2}$$

and $\tilde{\boldsymbol{v}}^{(k)} = D_c^{-1/2}\boldsymbol{v}^{(k)}$ is the principal eigenvector in the eigensystem of $G^{(k)t}G^{(k)}$ for $G^{(k)} = D_r^{1/2}B^{(k)}D_c^{-1/2}$. Increase k by one.

PS-3. Repeat PS-2 for $k = 2$ to $q-1$, where $q-1$ is the assumed rank of B. Hence $B^{(q)} = 0_{n\times q}$.

It is not difficult to show that $\boldsymbol{v}^{(1)}, \boldsymbol{v}^{(2)}, \cdots, \boldsymbol{v}^{(q-1)}$ are orthogonal to each other in the scaled Euclidean space: $\boldsymbol{v}^{(j)t} D_c^{-1} \boldsymbol{v}^{(k)} = 0, \quad j \neq k$. See Yang's dissertation (1998) for details.

From the Algorithm PS, we obtain standard column scores $\boldsymbol{y}^{(k)}$ and the principal row scores $\boldsymbol{x}^{(k)}$, where

$$\begin{aligned} \boldsymbol{y}^{(k)} &= D_c^{-1} \boldsymbol{v}^{(k)} = D_c^{-1/2} \tilde{\boldsymbol{v}}^{(k)} , \\ \boldsymbol{x}^{(k)} &= B \boldsymbol{y}^{(k)} \end{aligned}$$

for $k = 2, \cdots, q-1$. Since

$$B = \hat{B}^{(1)} + \cdots + \hat{B}^{(k-1)} + B^{(k)} = \sum_{l=1}^{k-1} B^{(l)} D_c^{-1} \boldsymbol{v}^{(l)} \boldsymbol{v}^{(l)t} + B^{(k)} ,$$

the principal row scores $\boldsymbol{x}^{(k)}$ is equal to

$$\boldsymbol{x}^{(k)} = \Big(\sum_{l=1}^{k-1} B^{(l)} D_c^{-1} \boldsymbol{v}^{(l)} \boldsymbol{v}^{(l)t} + B^{(k)} \Big) D_c^{-1} \boldsymbol{v}^{(k)} = B^{(k)} \boldsymbol{y}^{(k)} .$$

For later use, define

$$\boldsymbol{u}^{(k)} = \boldsymbol{x}^{(k)} / \sqrt{\lambda_k}, \quad k = 1, 2, \cdots, q-1$$

where $\lambda_k \equiv \boldsymbol{x}^{(k)t} D_r \boldsymbol{x}^{(k)}$. Since λ_k's $(k = 2, \cdots, q-1)$ are principal eigenvalues of $G^{(k)^t} G^{(k)}$ or eigenvalues of $G^{(2)^t} G^{(2)}$,

$$\lambda_2 \geq \cdots \geq \lambda_{q-1} .$$

4. DECOMPOSITION OF TWO-WAY TABLE MATRIX AND GOODNESS-OF-APPROXIMATION

From the Algorithm PS, we obtain the recursive formula

$$B^{(k)} = \hat{B}^{(k)} + B^{(k+1)}, \tag{4.1}$$

for $k = 1, 2, \cdots, q-1$, where $B^{(1)} = B$ and $B^{(q)} = 0_{n\times q}$. Hence

$$B = \hat{B}^{(1)} + \hat{B}^{(2)} + \cdots + \hat{B}^{(q-1)},$$

where

$$\hat{B}^{(k)} = B^{(k)} D_c^{-1/2} \tilde{\boldsymbol{v}}^{(k)} \tilde{\boldsymbol{v}}^{(k)t} D_c^{1/2} = \boldsymbol{u}^{(k)} \sqrt{\lambda_k}\, \boldsymbol{y}^{(k)t} D_c.$$

Therefore, we obtain a matrix decomposition of $B D_c^{-1}$ or

$$D_r^{-1}(F/N - \boldsymbol{r}\boldsymbol{c}^t)D_c^{-1} = \sum_{k=1}^{q-1} \boldsymbol{u}^{(k)}\sqrt{\lambda_k}\,\boldsymbol{y}^{(k)t} = UD_{\sqrt{\lambda}}Y^t, \tag{4.2}$$

where

$$U = (\boldsymbol{u}^{(1)}, \cdots, \boldsymbol{u}^{(q-1)}),\quad D_{\sqrt{\lambda}} = \mathrm{diag}(\lambda_1^{1/2}, \cdots, \lambda_{q-1}^{1/2}),\quad Y = (\boldsymbol{y}^{(1)}, \cdots, \boldsymbol{y}^{(q-1)})\ .$$

Here, we note that

$$\boldsymbol{y}^{(k)t}D_c\,\boldsymbol{y}^{(l)} = \boldsymbol{v}^{(k)t}D_c^{-1}\boldsymbol{v}^{(l)} = \begin{cases} 1 & \text{if} \quad k = l, \\ 0 & \text{if} \quad k \neq l \end{cases} \quad \text{for} \ \ k, l = 1, 2, \cdots, q-1.$$

But, it can be shown that (Yang, 1998)

$$\boldsymbol{u}^{(k)t}D_r\,\boldsymbol{u}^{(l)} = \begin{cases} 1 & \text{if} \quad k = l \quad \text{for} \quad k, l = 1, 2, \cdots, q-1 \\ 0 & \text{if} \quad k \neq l \quad \text{for} \quad k, l = 2, \cdots, q-1 \\ \text{nonzero} & \text{if} \quad k \neq l \quad \text{for} \quad k = 1 \ \text{ or } \ l = 1. \end{cases}$$

in general.

From (4.2), we have the expression for the cell chi-square values of the two-way contingency table:

$$D_r^{-1/2}(F/N - \boldsymbol{r}\boldsymbol{c}^t)D_c^{-1/2} = D_r^{1/2}(UD_{\sqrt{\lambda}}Y^t)D_c^{1/2}\ ,$$

or

$$\frac{f_{ij}/N - r_i c_j}{\sqrt{r}_i\sqrt{c}_j} = \sqrt{r}_i\sqrt{c}_j\Big(\sum_{k=1}^{q-1}\sqrt{\lambda}_k\ u_{ik}\ y_{jk}\Big).$$

When we represent the row profiles in s-dimensional subspace spanned by the columns of $V_s = (\boldsymbol{v}^{(1)}, \cdots, \boldsymbol{v}^{(s)}), s = 1, 2, \cdots, q-1$, we may define the goodness-of-approximation index by

$$\mathrm{GOA}_s = \frac{\parallel BD_c^{-1}(V_s : 0_{q\times q-1-s})\parallel_{D_r}^2}{\parallel BD_c^{-1}V_{q-1}\parallel_{D_r}^2},$$

where $\parallel M \parallel_D^2 = \mathrm{trace}(M^tDM)$. Then,

$$\mathrm{GOA}_s = \frac{\parallel UD_{\sqrt{\lambda}}Y^tD_c(Y_s : 0_{q\times q-1-s})\parallel_{D_r}^2}{\parallel UD_{\sqrt{\lambda}}Y^tD_cY\parallel_{D_r}^2}$$

$$= \frac{\parallel U\mathrm{diag}(\sqrt{\lambda}_1, \cdots, \sqrt{\lambda}_s, 0, \cdots, 0)\parallel_{D_r}^2}{\parallel UD_{\sqrt{\lambda}}\parallel_{D_r}^2} = \frac{\sum_{k=1}^{s}\lambda_k}{\sum_{k=1}^{q-1}\lambda_k},$$

no matter what the values of $\boldsymbol{u}^{(k)t}\ D_r\ \boldsymbol{u}^{(1)}$ for $k = 2, \cdots, q-1$ are.

5. A NUMERICAL ILLUSTRATION

Table 5.1 came from a comparative study of four drugs for effectiveness (Greenacre 1984, p. 263). Each subject from a study panel of 121 patients was randomly assigned to one of drugs, treated with the drug, and the effect was measured. Scale of measurement was five-point Likert scale from Poor(=1) to Excellent(=5). Pearson's χ^2 (d.f. 12) is 47.1 with p-value $\leq 0.01\%$. Hence, we may conclude that there exists some difference in drug effectiveness.

Table 5.1: Comparison of Four Drugs

	Poor	Fair	Good	V. Good	Excel.	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	22	33	20	19	121

Even though the chi-square test tells us statistical significance for the association between drugs and their effectiveness, it does not show the pattern of association. Hence, there is the need for data exploration. Surely, correspondence analysis is one of promising methods for the task.

First, the ordinary correspondence analysis is applied to the 4×5 table. For two-dimensional reduction($s = 2$), the goodness-of-approximation GOA_2 is equal to 98.2 %. Despite of high GOA, the first-axis standard column scores are not monotonic. See Table 5.2. Scores for category 1(=Poor) and 2(=Fair) are inconsistent with the inherent order in the column: Poor $\leq$ Fair $\leq$ Good $\leq$ Very Good $\leq$ Excellent.

Second, the special correspondence analysis is computed by Algorithm OS and Algorithm PS proposed in Sections 2 and 3. Now, the GOA_2 is equal to 96.4 %, a bit smaller than that of ordinary correspondence analysis. In return, as can be seen in Table 5.2, the first-axis standard column scores are congruent with the order in the column, with ties at Poor(=1) and Fair(=2).

Figures 5.1 and 5.2 show two-dimensional quantification plots for principal row scores and standard column scores. Both plots look similar, in that the first principal axis of both plots denotes the general direction of the drug effectiveness. However, the direction is more lucid in Figure 2. Also, the second principal axis of Figure 2 plays some role in showing the U-shape or horse-shoe effect (Maridia,

Kent, and Bibby 1979). One peculiar pattern in this horse-shoe is that "Fair" is at the left end of curve instead of "Poor", indicating that drug C draws the largest frequency of "Poor", even though the general ranking of C is the third, not the fourth.

Table 5.2: Principal Row Scores and Standard Column Scores

		Row Scores			Column Scores	
		Dim 1	Dim 2		Dim 1	Dim 2
Ordinary Correspondence Analysis	A	0.3493	-0.3012	1	-0.4469	-0.2682
	B	0.7040	0.2460	2	-1.1637	1.5946
	C	-0.4549	-0.2480	3	-0.4849	-1.0737
	D	-0.6277	0.3137	4	1.0560	-0.5635
				5	1.7131	0.9927
Correspondence Analysis with Column-Order	A	0.3073	-0.3489	1	-0.7616	-0.3316
	B	0.7159	0.1109	2	-0.7616	1.8289
	C	-0.4993	-0.1707	3	-0.5381	-0.9094
	D	-0.5494	0.4248	4	1.0549	-0.6782
				5	1.7883	0.6470

Note: A=Drug A, B=Drug B, C=Drug C, D=Drug D; 1=Poor, 2=Fair, 3=Good, 4=Very Good, 5=Excellent.

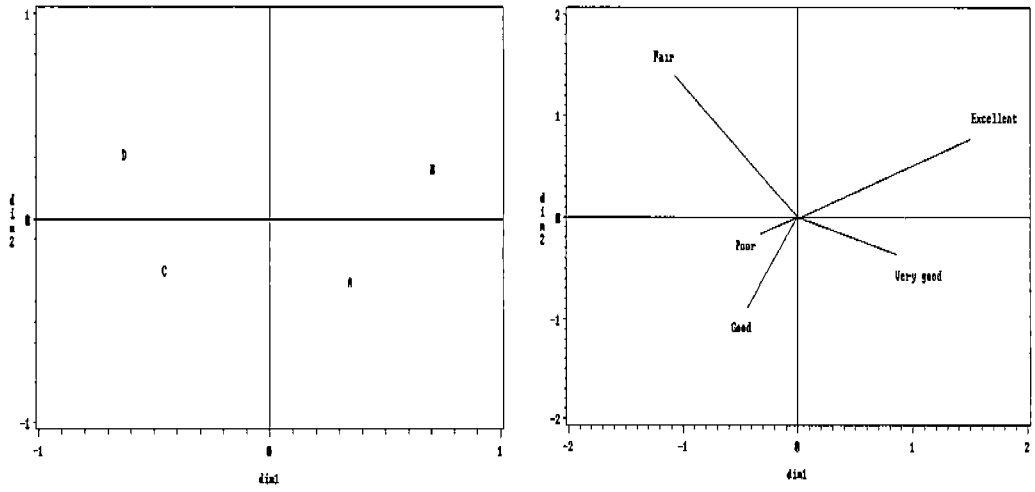

Figure 5.1: Row and Column Plots of Ordinary Correspondence Analysis ($\lambda_1 = 0.305(78.32\%)$, $\lambda_2 = 0.078(19.88\%)$)

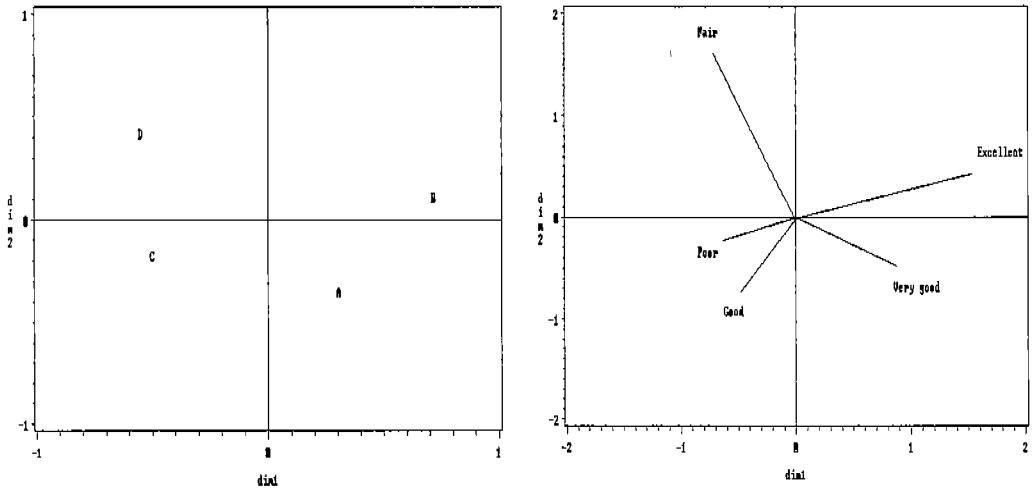

Figure 5.2: Row and Column Plots of Correspondence Analysis with Order-Preserving First Principal Axis($\lambda_1 = 0.291(74.78\%)$, $\lambda_2 = 0.084(21.60\%)$)

6. CONCLUDING REMARKS

For the two-way contingency tables with ordered categories, we developed a special correspondence analysis that quantifies the column categories in given order, using Algorithms OS (Optimal Scaling) and PS (Partial Scaling). When the row, not the column, has ordered categories, the technique proposed in this study can be applied by exchanging the row and the column.

In the future, one may consider the case in that both the row and the column have ordered categories. We speculate that there seems no way that quantifies both the row and column categories in given order by Euclidean geometry. It deserves further research.

REFERENCES

Greenacre, M.J. (1984). *Theory and Applications of Correspondence Analysis.* Academic Press, London.

Greenacre, M.J. (1994). *Correspondence Analysis in Practice.* Academic Press, London.

Greenacre, M.J. and Hastie, T. (1987). "The geometric interpretation of correspondence analysis," *Journal of the American Statistical Association,* **82**, 437-447.

Lebart, L., Morineau, A. and Warwick, K.M. (1984). *Multivariate Descriptive Statistical Analysis.*, Wiley, New York.

Mardia, K.V., Kent, J.T., and Bibby, J.M. (1979). *Multivariate Analysis.* Academic Press, London.

Nishisato, S. (1980). *Analysis of Categorical Data: Dual Scaling and Its Applications.* University of Toronto Press.

Nishisato, S. and Arri, P.S. (1975). "Nonlinear programming approach to optimal scaling of partially ordered categories," *Psychometrika,* **40**, 525-548.

Suh, H.S. and Huh, M.H. (1997). "Partial quantification in principal component analysis," *The Korean Communications in Statistics,* **4**, 637-644.

Yang, K.S. (1998). *Corrrespondence Analyses Specific to Certain Types of Categorical Data.* Ph.D. Dissertation. Department of Statistics, Korea University (written in Korean).