

# Visualizations for Matched Pairs Models Using Modified Correspondence Analysis

Chanyoon Lee<sup>a</sup>, Yong-Seok Choi<sup>1,a</sup>

<sup>a</sup>Department of Statistics, Pusan National University, Korea

---

## Abstract

Matched pairs are twice continuously measured data with the same categories. They can be represented as the square contingency tables. We can also consider symmetry and marginal homogeneity. Moreover, we can infer the matched pairs models; the symmetry model, the quasi-symmetry model, and the ordinal quasi-symmetry model. These inferences are involved in assumptions for special distributions. In this study, we visualize matched pairs models using modified correspondence analysis. Modified correspondence analysis can be used when square contingency tables are given; consequently, it is involved in the square and asymmetric correspondence matrix. This technique does not need assumptions for special distributions and is more helpful than the correspondence analysis to visualize matched pairs models.

**Keywords:** Square contingency tables, matched pairs models, asymmetric correspondence matrix, modified correspondence analysis.

---

## 1. Introduction

In categorical data, matched pairs are twice continuously measured data with similar categories (Jeong and Choi, 2009). They yield square contingency tables since they are dependent and their row and column categories are equal. In the square contingency tables, we consider the symmetry and marginal homogeneity. We can consider the matched pairs models related the marginal homogeneity. These models are the symmetry model, the quasi-symmetry model, and the ordinal quasi-symmetry model. Presently, we can approach them using the marginal homogeneity test; however, this method needs assumptions for special distributions.

Correspondence analysis is a multivariate graphical method to visualize the relationships between row and column categories of a two-way contingency table (Choi, 2001). It does not need to assumptions for special distributions. In general, the correspondence analysis is involved in the rectangular correspondence matrix; consequently, the correspondence matrix is not square. The correspondence matrix is a square and asymmetric matrix in the matched pairs that is known as the asymmetric correspondence matrix in this study.

Greenacre (2000) proposed correspondence analysis involved in the asymmetric correspondence matrix. His idea of this correspondence analysis is derived from the decomposition of the asymmetric correspondence matrix provided by Gower (1977). This decomposition gives the symmetric and skew-symmetric correspondence matrices (or the symmetric and skew-symmetric parts), respectively. We then perform the correspondence analysis using the symmetric and skew-symmetric correspondence matrices, respectively. In this study, we call this correspondence analysis the modified correspondence analysis. In the asymmetric correspondence matrix, a modified correspondence analysis

---

This work was supported by a 2-Year Research Grant of Pusan National University.

<sup>1</sup> Corresponding author: Department of Statistics, Pusan National University, 2, Busandaehak-ro 63beon-gil, Geumjeong-gu, Busan 609-735, Korea. E-mail: yschoi@pusan.ac.kr

can visualize the relationships between row and column categories of the square contingency tables; in addition, the information of modified correspondence analysis is larger than the correspondence analysis (Greenacre, 2000). The modified correspondence analysis can be more helpful than the correspondence analysis in the visualizing matched pairs models.

Here, in the visualizing the matched pairs models, it is possible to discuss the appropriateness of modified correspondence analysis instead of the multidimensional scaling. Recently, Huh and Lee (2012) developed multidimensional scaling for the asymmetric distance matrix using 'altitude' conceptions, and illustrated an example for the matched pairs. Their idea is also derived from the decomposition proposed by Gower (1977). They facilitated to interpret the skew-symmetric part in the multidimensional scaling. However, they only considered the frequencies of square contingency table in the asymmetric distance matrix regardless of the marginal frequencies. So, in the skew-symmetric part, it is difficult to visualize the relationships between the row and column categories of the square contingency table using the proposed methods. In opposite sides, it is possible that the modified correspondence analysis displays these relationships in the skew-symmetric part. That is, a modified correspondence analysis can visualize the matched pairs models.

In Section 2, we describe the conceptions for the matched pairs models. In Section 3, we introduce the modified correspondence analysis. In Section 4, we give illustrations of visualizations for matched pairs models using two matched pairs data. Finally, we provide a conclusion in Section 5.

## 2. Matched Pairs Models

### 2.1. Matched pairs models

Suppose that there are categorical responses of two samples between each subject in one sample pair with an observation in the other sample. These types are usually provided from mobility tables, the migration tables, and the first and second order service usage satisfaction surveys. These two samples have equal categories and they are called matched pairs; in addition, they are dependent and yield the square contingency tables. (Agresti, 2007)

In the square contingency tables, we denote  $n_{ij}$  by the frequency of  $i^{\text{th}}$  row and  $j^{\text{th}}$  row column, where  $i, j = 1, 2, \dots, I$  and  $n = \sum_{i=1}^I \sum_{j=1}^I n_{ij}$  is the total frequency. And let  $\pi_{ij}$  be the population proportion for  $i^{\text{th}}$  row and  $j^{\text{th}}$  column. Denote that  $\pi_{i+} = \sum_{j=1}^I \pi_{ij}$  and  $\pi_{+j} = \sum_{i=1}^I \pi_{ij}$ .

For all  $i$ ,  $\pi_{i+} = \pi_{+i}$  implies that the marginal homogeneity is satisfied, and  $\pi_{ij} = \pi_{ji}$  implies that the symmetry is satisfied, where  $i = 1, 2, \dots, I$ . If the symmetry is satisfied, then the marginal homogeneity is also satisfied. But the marginal homogeneity can be satisfied without symmetry (Agresti, 2007; Jeong and Choi, 2009).

There are three kinds of matched pairs models; the symmetry model, the quasi-symmetry model, and the ordinal quasi-symmetry model. The symmetry model implies that symmetry is satisfied; however, the others imply that the marginal homogeneity is not satisfied. Especially, the ordinal quasi-symmetry model can be considered when the categories are ordinal scores.

For all  $i, j = 1, 2, \dots, I$ , assume that  $\pi_{ij} > 0$ . The equation

$$\log\left(\frac{\pi_{ij}}{\pi_{ji}}\right) = 0 \quad (2.1)$$

implies to the symmetry model. The equation

$$\log\left(\frac{\pi_{ij}}{\pi_{ji}}\right) = \beta_i - \beta_j \quad (2.2)$$

implies to the quasi-symmetry model, where  $\beta_i$ 's are model parameters. The equation

$$\log\left(\frac{\pi_{ij}}{\pi_{ji}}\right) = \beta(u_i - u_j) \quad (2.3)$$

implies to the ordinal quasi-symmetry model, where  $u_1 \leq u_2 \leq \dots \leq u_I$  are ordered scores and  $\beta$  is the model parameter. The Equation (2.3) is the special case of the Equation (2.2) when all model parameters in the Equation (2.2) have a linear trend.

We introduce the marginal homogeneity tests according to Agresti (2007). In the marginal homogeneity test for the nominal categories, the null hypothesis is the symmetry model and the alternative hypothesis is the quasi-symmetry model. In this hypothesis test, the likelihood ratio statistic

$$G^2(S|QS) = G^2(S) - G^2(QS) \quad (2.4)$$

is used as the test statistic, where  $G^2(S)$  and  $G^2(QS)$  are the goodness-of-fit statistics for the symmetry and quasi-symmetry models, respectively. The test statistic  $G^2(S|QS)$  has the chi-square distribution with the degrees of freedom  $I - 1$ , in  $I \times I$  contingency tables. In the marginal homogeneity test for the ordinal categories, the null hypothesis is the symmetry model and the alternative hypothesis is the ordinal quasi-symmetry model. In this hypothesis test, the likelihood ratio statistic

$$G^2(S|OQS) = G^2(S) - G^2(OQS) \quad (2.5)$$

is used as the test statistic, where  $G^2(OQS)$  are the goodness-of-fit statistics for the ordinal quasi-symmetry model. The test statistic  $G^2(S|OQS)$  has the chi-square distribution with the degrees of freedom 1, in  $I \times I$  contingency tables. We can evaluate  $G^2(S)$ ,  $G^2(QS)$ , and  $G^2(OQS)$  using SAS/PROC GENMOD.

### 3. Modified Correspondence Analysis and Its Advantages

#### 3.1. Conceptions

In the matched pairs, we know that  $\pi_{ij}$  is the population proportion. Because  $p_{ij}$  is the maximum likelihood estimator of  $\pi_{ij}$  So, we use the sample proportion defined as  $p_{ij} = n_{ij}/n$ , instead of  $\pi_{ij}$  (Agresti, 2007). Also,  $E[p_{ij}] = \pi_{ij}$  is satisfied; therefore,  $p_{ij}$  is the unbiased estimator of  $\pi_{ij}$ .

Here, if  $\mathbf{P} = \{p_{ij}\}$  is a general correspondence matrix of a two-way contingency table, the correspondence analysis can be used to visualize the relationship between row categories and column categories of this. However,  $\mathbf{P}$  is the square and asymmetric matrix and we can call this an asymmetric correspondence matrix. Greenacre (2000) proposed a modified correspondence analysis for this type; consequently, we can visualize the matched pairs models using the modified correspondence analysis.

We decompose the asymmetric correspondence matrix  $\mathbf{P} = \{p_{ij}\}$  into two matrices (or parts) according to Gower (1977), such that

$$\mathbf{P} = \mathbf{S} + \mathbf{T}, \quad (3.1)$$

where  $\mathbf{S}$  is the symmetric correspondence matrix (or symmetric part) defined as  $\mathbf{S} = (1/2)(\mathbf{P} + \mathbf{P}')$  and  $\mathbf{T}$  is the skew-symmetric correspondence matrix (or skew-symmetric part) defined as  $\mathbf{T} = (1/2)(\mathbf{P} - \mathbf{P}')$  and  $i, j = 1, 2, \dots, I$ .

Generally, for the centering to the origin point, we consider that the centered asymmetric correspondence matrix using  $\mathbf{r}$  and  $\mathbf{c}$ , defined as  $\hat{\mathbf{P}} = \mathbf{P} - \mathbf{rc}^t$ , where  $\mathbf{P}\mathbf{1}_I = \mathbf{r}$  and  $\mathbf{P}^t\mathbf{1}_I = \mathbf{c}$  are called the row and column profiles, respectively, and  $\mathbf{1}_I = [1, 1, \dots, 1]^t$ . From (3.1), the equation

$$\mathbf{P} - \mathbf{rc}^t = \mathbf{S} - \mathbf{rc}^t + \mathbf{T} \quad (3.2)$$

can be established. But, unfortunately, there are two problems in the right-hand side of (3.2).

One of these problems is the elements of  $\mathbf{T}$ . Since  $\mathbf{T}$  has positive and negative elements,  $\mathbf{T}$  can not be considered as the correspondence matrix which has elements from 0 to 1. So, Greenacre (2000) suggests that  $\mathbf{T}$  is considered 'already centered matrix'. This is reasonable since  $\sum_{i=1}^I \sum_{j=1}^I t_{ij} = 0$ .

Another problem is the asymmetric matrix  $\mathbf{S} - \mathbf{rc}^t$ . Since  $\mathbf{S}$  is the symmetric matrix, we wish  $\mathbf{S} - \mathbf{rc}^t$  is the symmetric matrix, also. But  $\mathbf{rc}^t$  is not the symmetric matrix. Greenacre (2000) provides that we make the same row and column marginal sums by letting  $\mathbf{w} = (\mathbf{r} + \mathbf{c})/2$ . So, we use  $\mathbf{w}$  instead of  $\mathbf{r}$  and  $\mathbf{c}$ .

So, let  $\hat{\mathbf{P}}^* = \mathbf{P} - \mathbf{ww}^t$  be the recentered asymmetric correspondence matrix. This can be expressed as

$$\mathbf{P} - \mathbf{ww}^t = \mathbf{S} - \mathbf{ww}^t + \mathbf{T}. \quad (3.3)$$

Let  $\mathbf{M}^* = \mathbf{D}_w^{-1/2} \hat{\mathbf{P}}^* \mathbf{D}_w^{-1} \hat{\mathbf{P}}^* \mathbf{D}_w^{-1/2}$ , where  $\mathbf{D}_w = \text{diag}\{\mathbf{w}\}$ . Then, the total inertia with respect to  $\hat{\mathbf{P}}^*$  is evaluated and decomposed as

$$\text{tr}[\mathbf{M}^*] = \sum_{i=1}^I \sum_{j=1}^I \frac{(s_{ij} - w_i w_j)^2}{w_i w_j} + \sum_{i=1}^I \sum_{j=1}^I \frac{t_{ij}^2}{w_i w_j}, \quad i, j = 1, 2, \dots, I, \quad (3.4)$$

where  $\sum_{i=1}^I \sum_{j=1}^I t_{ij} = 0$ . Hence, the Equation (3.3) and (3.4) are corresponded each other. The first term of right-hand side of (3.4) is the total inertia with respect to  $\mathbf{S} - \mathbf{ww}^t$ , and the second term of right-hand side of (3.4) is the total inertia with respect to  $\mathbf{T}$ .

Greenacre (2000) provided that the generalized singular value decomposition of  $\mathbf{S} - \mathbf{ww}^t$  and  $\mathbf{T}$  for finding the principal coordinates of the symmetric and skew-symmetric parts, as the following. Suppose that the  $2I \times 2I$  asymmetric correspondence matrix

$$\tilde{\mathbf{P}} = \begin{pmatrix} \mathbf{P} & \mathbf{P}^t \\ \mathbf{P}^t & \mathbf{P} \end{pmatrix}. \quad (3.5)$$

Suppose that  $\tilde{\mathbf{P}}$  has

$$\tilde{\mathbf{w}} = \frac{1}{2} \begin{pmatrix} \mathbf{w} \\ \mathbf{w} \end{pmatrix}$$

and we assume that the generalized singular value decompositions of  $\mathbf{S} - \mathbf{ww}^t$  and  $\mathbf{T}$  are

$$\mathbf{S} - \mathbf{ww}^t = \mathbf{X}_{(S)} \mathbf{D}_{\lambda(S)}^* \mathbf{X}_{(S)}^t \quad \text{and} \quad \mathbf{T} = \mathbf{X}_{(T)} \mathbf{D}_{\lambda(T)}^* \mathbf{X}_{(T)}^t, \quad (3.6)$$

respectively, where  $\mathbf{D}_{\lambda(S)}^*$  is the diagonal matrix whose elements are singular values of  $\mathbf{D}_w^{-1/2}(\mathbf{S} - \mathbf{ww}^t)\mathbf{D}_w^{-1/2}$ ,  $\mathbf{D}_{\lambda(T)}^*$  is the diagonal matrix whose elements are singular values of  $\mathbf{D}_w^{-1/2}\mathbf{T}\mathbf{D}_w^{-1/2}$ , where  $\mathbf{X}_{(S)} \mathbf{D}_{\lambda(S)}^{-1} \mathbf{X}_{(S)}^t = \mathbf{I}_I = \mathbf{X}_{(T)} \mathbf{D}_{\lambda(T)}^{-1} \mathbf{X}_{(T)}^t$ . Here,

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

Table 1: Cross-citations among eight Economics journals (Stigler, 1994)

Journals	Cited journals								Total
	AER	EJ	Ecmca	JET	JPE	QJE	RESt	RESd	
AER	1,090	108	423	136	652	275	120	209	3,013
EJ	440	369	458	103	375	160	55	208	2,168
Ecmca	157	44	894	269	170	92	35	229	1,890
JET	123	22	669	754	126	100	5	252	2,051
JPE	373	61	321	108	646	155	75	155	1,894
QJE	301	43	204	98	230	220	32	114	1,242
RESt	416	66	416	25	310	98	574	97	2,002
RESd	140	40	379	189	166	94	12	291	1,311
Total	3,040	753	3,704	1,682	2,675	1,194	908	1,555	15,571

is a block diagonal matrix and composed of  $2 \times 2$  blocks and in the final position if  $I$  is odd. The matrix  $\mathbf{J}$  represents a rotation of planar points by  $90^\circ$  such that

$$\mathbf{J}^t = -\mathbf{J}. \quad (3.7)$$

From (3.1), (3.3), (3.6) and (3.7), we get

$$\tilde{\mathbf{P}} - \mathbf{w}\mathbf{w}^t = \begin{pmatrix} \mathbf{X}_{(S)}/2 & \mathbf{X}_{(T)}/2 \\ \mathbf{X}_{(S)}/2 & -\mathbf{X}_{(T)}/2 \end{pmatrix} \begin{pmatrix} \mathbf{D}_{\lambda(S)}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{\lambda(T)}^* \end{pmatrix} \begin{pmatrix} \mathbf{X}_{(S)}/2 & \mathbf{X}_{(T)}\mathbf{J}/2 \\ \mathbf{X}_{(S)}/2 & -\mathbf{X}_{(T)}\mathbf{J}/2 \end{pmatrix} \quad (3.8)$$

and the Equation (3.8) implies that the generalized singular value decomposition of  $\tilde{\mathbf{P}} - \mathbf{w}\mathbf{w}^t$  is established from the symmetric and skew-symmetric parts of  $\mathbf{P} - \mathbf{w}\mathbf{w}^t$ .

Finally, suppose that the principal coordinates of the row and column for the symmetric part are  $\mathbf{R}_{(S)}^*$  and  $\mathbf{C}_{(S)}^*$ , respectively, they can be evaluated as  $\mathbf{R}_{(S)}^* = \mathbf{D}_w^{-1}\mathbf{X}_{(S)}\mathbf{D}_{\lambda(S)}^* = \mathbf{C}_{(S)}^*$  from the generalized singular value decomposition. In similarly,  $\mathbf{R}_{(T)}^*$  and  $\mathbf{C}_{(T)}^*$  are the principal coordinates of the row and column for the skew-symmetric part, respectively, and can be calculated as  $\mathbf{R}_{(T)}^* = \mathbf{D}_w^{-1}\mathbf{X}_{(T)}\mathbf{D}_{\lambda(T)}^*$  and  $\mathbf{C}_{(T)}^* = -\mathbf{D}_w^{-1}\mathbf{X}_{(T)}\mathbf{D}_{\lambda(T)}^*$ . So, we can get the correspondence map for symmetric and skew-symmetric parts, and they are called the symmetry and skew-symmetry map, respectively.

### 3.2. Interpretations

In the situation of the modified correspondence analysis, to display in distinct the categories of rows and columns onto the symmetry and skew-symmetry map is not meaningful, so each pair of points is replaced by one point expressing a common categories. Greenacre (2000) suggested that the symmetry and skew-symmetry map should be scaled equally for a comparison of two maps. So, we overlay the symmetry and skew-symmetry map, and call this map the lapping map.

To recognize the degree and direction of deviation from symmetry, we need to concentrate on the interpretations of the skew-symmetry map since the skew-symmetry map has a definitive influence on the visualizations for the matched pairs models. Section 4 describes these facts. We use Table 1 for this section to show the cross-citation among eight economics journals. The eight journals are 'ER', 'EJ', 'Ecmca', 'JET', 'JPE', 'QJE', 'RESt', 'RESd'. In the modified correspondence analysis for Table 1, note that the symbol '□' is a point of the symmetry map and symbol '▲' is a point of the skew-symmetry map. Section 4 adequately describes these facts.

Figure 2 is the visualizations of degrees of symmetric and skew-symmetric parts for Table 1. The total inertia of symmetric part is 0.4063(75.39%), and the total inertia of skew-symmetric part is 0.1326(24.61%). The information of symmetric part is larger than the information of skew-symmetric

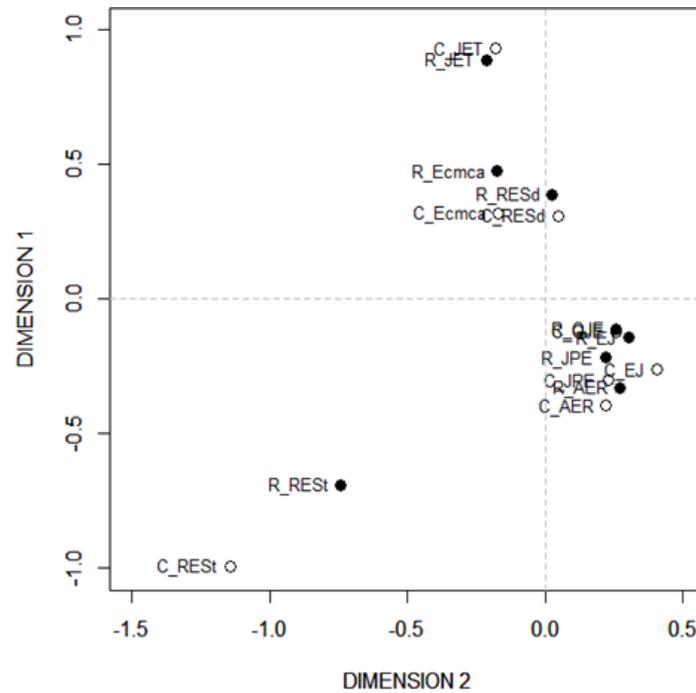


Figure 1: Correspondence map for Table 1 (○ : row category, ● : column category)

part. In each point related the journals citations of Figure 2, the point 'REST' and 'JET' of skew-symmetry map are nearer (from the origin point) than the symmetry map. The point 'Ecmca', 'RESd', 'AER', 'EJ', 'QJE' and 'JPE' of skew-symmetry map are as near as the symmetry map. So, in 'REST' and 'JET', the information of symmetric part is larger than the information of the skew-symmetric part and in 'Ecmca', 'RESd', 'AER', 'EJ', 'QJE' and 'JPE', the information of skew-symmetric part is as large as the symmetric part information.

The three solid lines and the arc shaped arrow in Figure 2 can help to interpret the skew-symmetry map. The interpretations for skew-symmetry map are initiated by the selections of two points on the skew-symmetry map. We can form the triangle using the selected two points and the origin point. These points are linked by the three solid lines in Figure 2. We can also consider the rotation direction on the quadrant between the selected two points. This is associated with the arc shaped arrow direction, in Figure 2.

In the formed triangle, the wider the area of triangle, the greater the deviation from the symmetry between the selected two points (Constantine and Gower, 1978). If the selected two points are close to the origin point, then it indicates the small deviation from the symmetry (Izenman, 2008). In addition, if a point related the  $i^{\text{th}}$  category rotate clockwise to a point related the  $j^{\text{th}}$  category as the arc shaped arrow in Figure 2, then the  $i^{\text{th}}$  row category has the positive-deviation for the  $j^{\text{th}}$  column category, that is  $n_{ij} > n_{ji}$ , where  $i, j = 1, 2, \dots, I$ .

A point related the  $j^{\text{th}}$  category rotate counter-clockwise to a point related the  $i^{\text{th}}$  category, then the  $j^{\text{th}}$  row category has a negative-deviation for the  $i^{\text{th}}$  column category as the converse of an arc shaped direction in Figure 2 (Greenacre, 2000). It seems that the widest triangle, formed from two points and the origin point, is the triangle formed from point 'EJ', point 'Ecmca', and the origin point. So,

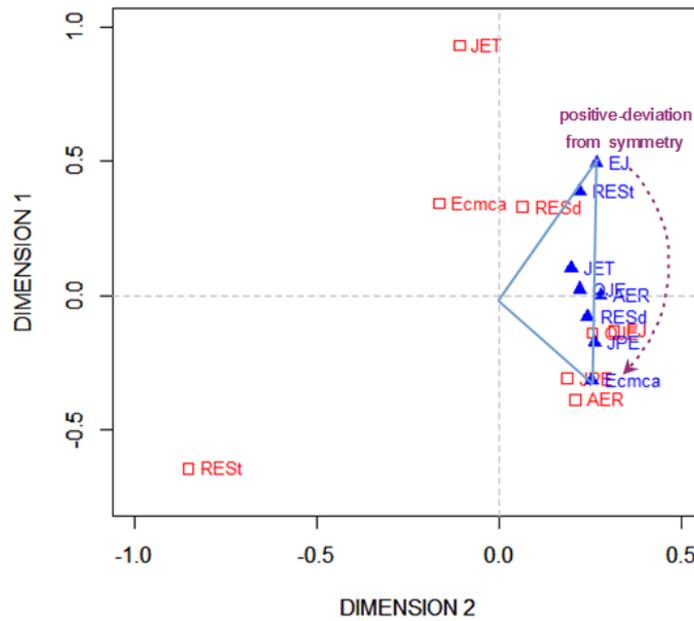


Figure 2: Lapping map for Table 1

‘EJ’ has the largest positive-deviation for ‘Ecmca’, and ‘Ecmca’ has the largest negative-deviation for ‘EJ’.

Figure 1 represents the corresponding relationships for Table 1. First, the total inertia is 0.4643. It is smaller than the inertia of modified correspondence analysis, evaluated as  $0.4063 + 0.1326 = 0.5389$  from the equation of (3.4). So, in the aspects of inertia, the modified correspondence analysis is better than the correspondence analysis. This fact proved by Greenacre (2000). In the corresponding relationships, we only know that each journal cites itself, respectively, and that each journal is cited by itself. Unfortunately, we are difficult to know the degree of symmetry and the skew-symmetry clearly using the correspondence analysis. Therefore, the modified correspondence is better to visualize the matched pairs models.

To recognize the degree and direction of deviations numerically, Greenacre (2000) proposed the normalized frequencies of skew-symmetric part, defined as

$$\Delta = (\delta_{ij}) = n\mathbf{D}_w^{-\frac{1}{2}}\mathbf{T}\mathbf{D}_w^{-\frac{1}{2}}, \tag{3.9}$$

where  $i, j = 1, 2, \dots, I$ .

For  $\delta_{ij} > 0$ , the greater  $\delta_{ij}$ 's, the more the positive-deviation from the  $i^{th}$  row category to the  $j^{th}$  column category, and for  $\delta_{ij} < 0$ , the smaller  $\delta_{ij}$ 's, the more the negative-deviation from the  $i^{th}$  category in row to the  $j^{th}$  category in column, and for  $\delta_{ij} = 0$ , there is no deviation between the  $i^{th}$  and  $j^{th}$  category. And the larger  $|\delta_{ij}|$ 's, the wider the area of triangle, which is formed from the origin point, the two points related the  $i^{th}$  and  $j^{th}$  category respectively. In Table 2, the largest positive-deviation is ‘EJ’ for ‘Ecmca’ and the largest negative-deviation is ‘Ecmca’ for ‘EJ’. The smallest positive-deviation is ‘JET’ for ‘QJE’ and the smallest negative-deviation is ‘QJE’ for ‘JET’. The largest deviation is more than 10% of the total frequencies, 15,571.

Table 2: Tabulated values from  $\Delta$  for Table 1

Journals	Cited journals							
	AER	EJ	Ecmca	JET	JPE	QJE	RESt	RESd
AER	0	-1229.43	708.00	42.58	826.09	-105.43	-1098.19	257.95
EJ	1229.43	0	1586.26	381.95	1338.35	682.96	-58.75	904.11
Ecmca	-708.00	-1586.26	0	-1355.72	-462.60	-469.91	-1462.57	-580.22
JET	-42.58	-381.95	1355.72	0	67.87	10.33	-94.49	299.91
JPE	-826.09	-1338.35	462.60	-67.87	0	-350.05	-1003.52	-47.33
QJE	105.43	-682.96	469.91	-10.33	350.05	0	-385.99	117.86
RESt	1098.19	58.75	1462.57	94.49	1003.52	385.99	0	458.30
RESd	-257.95	-904.11	580.22	-299.91	47.33	-117.86	-458.30	0

Table 3: Opinions on premarital and extramarital sex (Agresti, 2007)

Premarital Sex	Extramarital Sex				Total
	$u_1 = 1$	$u_2 = 2$	$u_3 = 3$	$u_4 = 4$	
$u_1 = 1$	144	2	0	0	146
$u_2 = 2$	33	4	2	0	39
$u_3 = 3$	84	14	6	1	105
$u_4 = 4$	126	29	25	5	185
Total	387	49	33	6	475

## 4. Illustrations

### 4.1. Economics journals cross-citations Data

This data was already referred in Section 3.2. Also, we gave interpretations for the modified correspondence analysis from Table 1. We can know that the symmetry model cannot be satisfied since the information of skew-symmetric part as large as the information of symmetric part, for six journals in Figure 2. Table 1 has the nominal categories. Therefore, we guess the quasi-symmetry model.

In the marginal homogeneity test as (2.4), the goodness-of-fit statistics are evaluated as  $G^2(S) = 2501.32$  and  $G^2(QS) = 25.47$ . The test statistic for the test of (2.4) is  $G^2(S|QS) = 2475.85$ , which has the chi-square distribution with the degrees of freedom 7 and  $p$ -value is less than 0.001, so the marginal homogeneity is not satisfied under the 5% significance level. Therefore, we can consider the quasi-symmetry model from Table 1.

### 4.2. Opinions on premarital and extramarital sex data

Table 3 is composed of four ordered scores and shows that the opinions on sex before the marriage (premarital sex) and sex after marriage (extramarital sex). The ordinal scores ' $u_1 = 1$ ', ' $u_2 = 2$ ', ' $u_3 = 3$ ', ' $u_4 = 4$ ' represent 'Very Wrong', 'Wrong', 'Wrong only Sometimes', and 'Never Wrong', respectively.

Figure 3 is the visualizations of degrees of symmetric and skew-symmetric parts for Table 3. The total inertia of symmetric part is 0.0600(7.78%), and the total inertia of skew-symmetric part is 0.7122(92.22%). The information of the skew-symmetric part is larger than the information of symmetric part. In each point related to the opinions on premarital and extramarital sex described in of Figure 3, the four points 'NW'(Never Wrong), 'SW'(Wrong only Sometimes), 'W'(Wrong), and 'VW'(Very Wrong) of symmetry map are nearer (from the origin point) than those of the skew-symmetry map. So, in 'Never Wrong', 'Wrong only Sometimes', 'Wrong', and 'Very Wrong', the information of skew-symmetric part is larger than the information of the symmetric part.

It seems that the widest triangle, formed from two points and the origin point, is the triangle formed from point 'NW'(Never Wrong), point 'VW'(Very Wrong), and the origin point. The point

Table 4: Tabulated values from  $\Delta$  for Table 3

Premarital Sex	Extramarital Sex			
	$u_1 = 1$	$u_2 = 2$	$u_3 = 3$	$u_4 = 4$
$u_1 = 1$	0	-67.99	-147.12	-187.58
$u_2 = 2$	67.99	0	-51.72	-106.25
$u_3 = 3$	147.12	51.72	0	-70.22
$u_4 = 4$	187.58	106.25	70.22	0

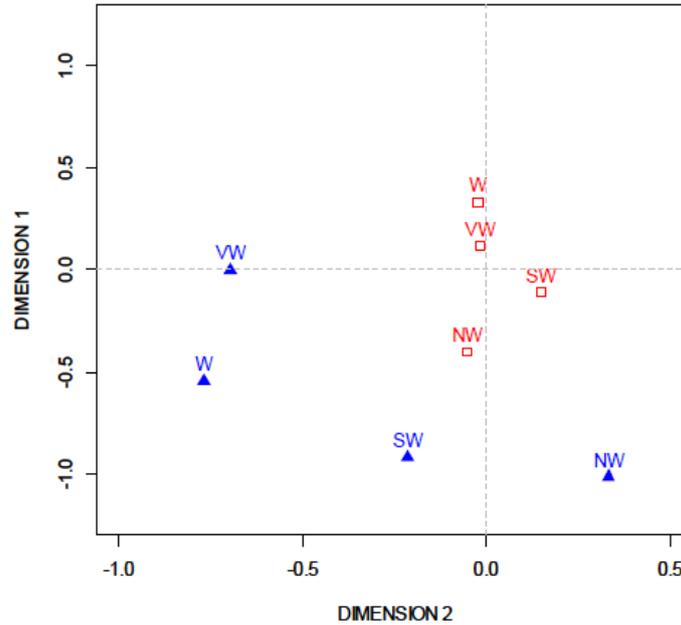


Figure 3: Lapping map for Table 3 (VW : Very Wrong, W : Wrong, SW : Wrong only Sometimes, NW : Never Wrong)

‘NW’(Never Wrong) has the positive-deviation for other points, and the point ‘W’ has the positive-deviation for point ‘SW’(Wrong only Sometimes) and for ‘NW’(Never Wrong), point ‘SW’(Wrong only Sometimes) has the positive-deviation for point ‘NW’(Never Wrong). Table 4 indicates that the largest positive-deviation is ‘Never Wrong’ for ‘Very Wrong’ and the largest negative-deviation is ‘Very Wrong’ for ‘Never Wrong’. The smallest positive-deviation is ‘Wrong’ for ‘Very Wrong’ and the smallest negative-deviation is ‘Wrong’ for ‘Very Wrong’. The largest deviation is 40% of the total frequencies, 475, and the smallest deviation is about 14% of the total frequencies. The pairs of two ordered scores whose difference is one: the pair of  $u_1$  and  $u_2$ , the pair of  $u_2$  and  $u_3$ , the pair of  $u_3$  and  $u_4$ . The deviations are similar. The pairs of two ordered scores whose difference is two: the pair of  $u_1$  and  $u_3$ , the pair of  $u_2$  and  $u_4$ . Those of the deviations are similar, too, and they are larger than the deviations of the pairs of the ordered scores with one difference. Finally, the pairs of two ordered scored whose difference is three, the pair of  $u_1$  and  $u_4$  has the largest deviation.

We can know that the symmetry model cannot be satisfied since the information of skew-symmetric part is larger than the information of symmetric part, from Figure 3. Table 3 has the ordinal categories and all deviations have a linear trend. Therefore, we guess the ordinal quasi-symmetry model. In the marginal homogeneity test as (2.5), the goodness-of-fit statistics for the test of (2.5) are evaluated

as  $G^2(S) = 402.20$  and  $G^2(OQS) = 2.10$ . The test statistic is  $G^2(S|OQS) = 400.10$ , which has the chi-square distribution with the degrees of freedom 1. The  $p$ -value is less than 0.001; consequently, the marginal homogeneity is not satisfied under 5% significance level. Therefore, we can consider the ordinal quasi-symmetry model from Table 3.

## 5. Conclusion

We can know that the symmetry model cannot be satisfied if the information of skew-symmetric part is larger than (or as large as) the information of symmetric part for the majority of categories in the lapping map. In these situations, we can guess the quasi-symmetry models if the square contingency tables have nominal categories. We can guess the ordinal quasi-symmetry models if the square contingency tables have ordinal categories and all deviations have a linear trend. We can confirm a linear trend from the points of the skew-symmetric part in the lapping map or the tabulated values from  $\Delta$ . These visualizations can be useful methods if the square contingency tables relatively have many categories; therefore, we can guess the matched pairs models with no assumption for special distributions.

## References

- Agresti, A. (2007). *An Introduction to Categorical Data Analysis*, Wiley, New York.
- Choi, Y. S. (2001). *SAS Correspondence Analysis*, Free Academy, Seoul.
- Constantine, A. G. and Gower, J. C. (1978). Graphical representation of asymmetric matrices, *Journal of Applied Statistics*, **27**, 297–304.
- Gower, J. C. (1977). The analysis of asymmetry and orthogonality, *Recent Developments in Statistics*, 109–123.
- Greenacre, M. J. (1984). *Theory and Applications of Correspondence Analysis*, Academic Press, New York.
- Greenacre, M. J. (2000). Correspondence analysis of square asymmetric matrices, *Journal of Applied Statistics*, **49**, 297–310.
- Huh, M. H. and Lee, Y. G. (2012). Multidimensional scaling of asymmetric distance matrices, *Korean Journal of Applied Statistics*, **25**, 613–620.
- Izenman, A. J. (2008). *Modern Multivariate Statistical Techniques*, Springer, New York.
- Jeong, K. M. and Choi, Y. S. (2009). *Categorical Data Analysis Using SAS*, Free Academy, Seoul.
- Stigler, S. (1994). Citation patterns in the Journals of Statistics and Probability, *Statistical Science*, **9**, 94–108.