International Journal of Statistics and Probability; Vol. 7, No. 3; May 2018
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Decomposition of Parsimonious Independence Model Using Pearson, Kendall and Spearman's Correlations for Two-Way Contingency Tables

Kiyotaka Iki[1], Shun Sato[2] & Sadao Tomizawa[2]

[1] Faculty of Economics, Nihon University, Japan

[2] Department of Information Sciences, Faculty of Science and Technology, Tokyo University of Science, Japan

Correspondence: Kiyotaka Iki, Faculty of Economics, Nihon University, Chiyoda-ku, Tokyo, Japan.

Received: March 7, 2018 Accepted: March 23, 2018 Online Published: April 27, 2018

doi:10.5539/ijsp.v7n3p105 URL: https://doi.org/10.5539/ijsp.v7n3p105

Abstract

For two-way contingency tables with ordered categories, Tomizawa (1992) considered the parsimonious Linear-by-Linear association model. This model can be described in terms of fewer parameters than the Linear-by-Linear association model (Agresti, 1983). The purpose of this paper is (i) to define the parsimonious independence model, (ii) to show the parsimonious independence model holds if and only if the parsimonious Linear-by-Linear association model holds and the each one of various correlation coefficients is equal to zero, and (iii) show the statistic for testing the parsimonious independence model is asymptotically equivalent to the sum of test statistics for the decomposed models.

Keywords: Kendall's tau-b, Linear-by-Linear association, orthogonal decomposition, Pearson correlation coefficient, Spearman's rho

1. Introduction

For a $r \times c$ contingency table with ordered categories, let X and Y denote the row and column variables, respectively. Also, let $\Pr(X = i, Y = j) = p_{ij}$ for $i = 1, \dots, r; j = 1, \dots, c$. The independence (I) model (Goodman, 1979) is defined by

$$p_{ij} = \mu\alpha_i\beta_j \quad (i = 1, \dots, r; j = 1, \dots, c),$$

where, without loss of generality, $\prod_{i=1}^{r} \alpha_i = \prod_{j=1}^{c} \beta_j = 1$. Suppose that the known scores $\{u_i\}$, $\{v_j\}$ can be assigned to the rows and columns, respectively, where $u_1 < \cdots < u_r$ and $v_1 < \cdots < v_c$. The linear-by-linear (LL) association model (Agresti, 1983) is defined by

$$p_{ij} = \mu\alpha_i\beta_j\theta^{u_i v_j} \quad (i = 1, \dots, r; j = 1, \dots, c).$$

When the scores $\{u_i\}$ and $\{v_j\}$ are equal-interval scores (or the integer scores $\{u_i = i\}$ and $\{v_j = j\}$), the LL association model is identical to the uniform association model (see Goodman, 1979 and Agresti, 1984, p. 78). The odds ratio for rows i and j $(> i)$, and columns s and t $(> s)$ are denoted by $\theta_{(i<j;s<t)}$; thus,

$$\theta_{(i<j;s<t)} = \frac{p_{is}p_{jt}}{p_{it}p_{js}}.$$

Using the log odds ratio, the LL association model can be expressed as

$$\log\theta_{(i<j;s<t)} = (u_j - u_i)(v_t - v_s)\log\theta \quad (1 \le i < j \le r, 1 \le s < t \le c).$$

A special case of the LL association model obtained by letting $\theta = 1$ is the I model.

Let $g_1(i) = u_i$ $(i = 1, \dots, r)$ and $g_2(j) = v_j$ $(j = 1, \dots, c)$. Define the variables U and V by $U = g_1(X)$ and $V = g_2(Y)$. When the I model holds, Pearson correlation coefficient ρ for U and V (denoted by $\rho(U, V)$) is equal to zero, however, the converse does not hold. We are interested in what structure between X and Y is necessary for obtaining the independence, in addition to the structure of the correlation being equal to zero. Instead of the structure that $\rho(U, V)$ is equal to zero, we are also interested in the structure that Kendall's tau-b measure (Kendall, 1945) or Spearman's ρ_s measure (Stuart, 1963) is equal to zero. Let P_C and P_D denote the probability of concordance for a randomly selected pair of observations and the probability of discordance for the pair, respectively, i.e.,

$$P_C = 2\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\sum_{s=1}^{c-1}\sum_{t=s+1}^{c} p_{is}p_{jt} \quad \text{and} \quad P_D = 2\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\sum_{s=1}^{c-1}\sum_{t=s+1}^{c} p_{it}p_{js};$$

see Kendall and Gibbons (1990, p. 6). Kendall's τ_b is defined by

$$\tau_b = \frac{P_C - P_D}{\left[\left(1-\sum_{i=1}^{r} p_{i\cdot}^2\right)\left(1-\sum_{j=1}^{c} p_{\cdot j}^2\right)\right]^{1/2}},$$

where $p_{i\cdot} = \sum_{t=1}^{c} p_{it}$, $p_{\cdot j} = \sum_{s=1}^{r} p_{sj}$. Also, let

$$r_i^X = \sum_{k=1}^{i-1} p_{k\cdot} + \frac{p_{i\cdot}}{2} \ (i = 1, \ldots, r), \quad r_j^Y = \sum_{l=1}^{j-1} p_{\cdot l} + \frac{p_{\cdot j}}{2} \ (j = 1, \ldots, c),$$

where $\{r_i^X\}$ and $\{r_j^Y\}$ are the marginal rigits; see Bross (1958) and Fleiss et al. (2003, pp. 198-205). Let $h_1(i) = r_i^X$ $(i = 1, \ldots, r)$ and $h_2(j) = r_j^Y$ $(j = 1, \ldots, c)$. Define the variables Z_1 and Z_2 by $Z_1 = h_1(X)$ and $Z_2 = h_2(Y)$. Spearman's ρ_s is the correlation coefficient of Z_1 and Z_2, defined by

$$\rho_s = \frac{\sum_{i=1}^{r}\sum_{j=1}^{c}\left(r_i^X - 0.5\right)\left(r_j^Y - 0.5\right)p_{ij}}{\left[\left(\sum_{i=1}^{r}\left(r_i^X - 0.5\right)^2 p_{i\cdot}\right)\left(\sum_{j=1}^{c}\left(r_j^Y - 0.5\right)^2 p_{\cdot j}\right)\right]^{1/2}}.$$

Note that $E(Z_1) = E(Z_2) = 0.5$ although the proof is omitted. Tomizawa et al. (2008) showed the following theorems;

Theorem 1 *The I model holds if and only if Pearson correlation coefficient $\rho(U, V) = 0$ and the LL association model holds.*

Theorem 2 *The I model holds if and only if Kendall's $\tau_b = 0$ and the LL association model holds.*

Theorem 3 *The I model holds if and only if Spearman's $\rho_s = 0$ and the LL association model holds.*

These theorems showed that the structure of the LL association model is necessary for obtaining the independence, in addition to the structure of correlations being equal to zero.

Tomizawa (1992) considered the parsimonious Linear-by-Linear association (PLL) model, defined by

$$p_{ij} = \mu\alpha^{u_i}\beta^{v_j}\theta^{u_i v_j} \quad (i = 1, \ldots, r; j = 1, \ldots, c).$$

Let ω_{ij}^X denotes the local odds of classification in column $j + 1$ instead of j for a fixed row i, i.e., $\omega_{ij}^X = p_{i,j+1}/p_{ij}$ $(i = 1, \ldots, r;\ j = 1, \ldots, c-1)$ and ω_{ij}^Y denotes the local odds of classification in row $i + 1$ instead of i for a fixed column j, i.e., $\omega_{ij}^Y = p_{i+1,j}/p_{ij}$ $(i = 1, \ldots, r-1;\ j = 1, \ldots, c)$. Then using the log odds ratio and the log local odds, the PLL model can be expressed as

$$\begin{aligned}
\log\theta_{(i<j;s<t)} &= (u_j - u_i)(v_t - v_s)\log\theta && (i < j, s < t),\\
\log\omega_{ij}^X &= (v_{j+1} - v_j)\xi_i^X && (i = 1, \ldots, r; j = 1, \ldots, c-1),\\
\log\omega_{ij}^Y &= (u_{i+1} - u_i)\xi_j^Y && (i = 1, \ldots, r-1; j = 1, \ldots, c),
\end{aligned}$$

where ξ_i^X and ξ_j^Y are unspecified. Namely, this model has the restrictions of local row odds and local column odds, in addition to the structure of the LL association model. We are interested in proposing the parsimonious independence model and considering decompositions of the proposed model using the PLL model and correlations.

In this paper, we (i) define the parsimonious independence model, (ii) show the parsimonious independence model holds if and only if the PLL model holds and the each one of $\rho(U, V)$, τ_b and ρ_s equals zero, and (iii) show the goodness-of-fit test statistic for the parsimonious independence model is asymptotically equivalent to the sum of test statistics for the decomposed models. Examples are given.

2. Decompositions of the Model

We define the parsimonious independence (PI) model by

$$p_{ij} = \mu\alpha^{u_i}\beta^{v_j} \quad (i = 1, \ldots, r; j = 1, \ldots, c).$$

The PI model is a special case of the PLL model obtained by letting $\theta = 1$. This model describes that the row and column variables are independent and the local row odds and local column odds have the restrictions, namely,

$$\begin{aligned}
\log\theta_{(i<j;s<t)} &= 0 && (i < j, s < t),\\
\log\omega_{ij}^X &= (v_{j+1} - v_j)\xi_i^X && (i = 1, \ldots, r; j = 1, \ldots, c-1),\\
\log\omega_{ij}^Y &= (u_{i+1} - u_i)\xi_j^Y && (i = 1, \ldots, r-1; j = 1, \ldots, c).
\end{aligned}$$

Tomizawa et al. (2008) showed the following lemma;

Lemma 1 *Pearson correlation coefficient* $\rho(U, V) = 0$ *is equivalent to*

$$\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\sum_{s=1}^{c-1}\sum_{t=s+1}^{c}(u_j - u_i)(v_t - v_s)p_{js}p_{it}(\theta_{(i<j;s<t)} - 1) = 0. \quad (1)$$

From Lemma 1, we obtain the following theorem;

Theorem 4 *The PI model holds, if and only if* $\rho(U, V) = 0$ *and the PLL model holds.*

Proof. Under the PLL model, equation (1) is expressed as

$$\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\sum_{s=1}^{c-1}\sum_{t=s+1}^{c}(u_j - u_i)(v_t - v_s)p_{js}p_{it}(\theta^{(u_j-u_i)(v_t-v_s)} - 1) = 0.$$

Thus, $\rho(U, V) = 0$ holds if and only if $\theta = 1$ (i.e., the PI model holds). The proof is completed.

Tomizawa et al. (2008) gave the following lemma;

Lemma 2 *Kendall's* $\tau_b = 0$ *is equivalent to*

$$\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\sum_{s=1}^{c-1}\sum_{t=s+1}^{c}p_{js}p_{it}(\theta_{(i<j;s<t)} - 1) = 0. \quad (2)$$

From Lemma 2, we obtain the following theorem;

Theorem 5 *The PI model holds, if and only if* $\tau_b = 0$ *and the PLL model holds.*

Proof. Under the PLL model, equation (2) is expressed as

$$\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\sum_{s=1}^{c-1}\sum_{t=s+1}^{c}p_{js}p_{it}(\theta^{(u_j-u_i)(v_t-v_s)} - 1) = 0.$$

Thus, $\tau_b = 0$ holds if and only if $\theta = 1$ (i.e., the PI model holds). The proof is completed.

Tahata et al. (2008) gave the following lemma;

Lemma 3 *Spearman's* $\rho_s = 0$ *is equivalent to*

$$\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\sum_{s=1}^{c-1}\sum_{t=s+1}^{c}(r_j^X - r_i^X)(r_t^Y - r_s^Y)p_{js}p_{it}(\theta_{(i<j;s<t)} - 1) = 0. \quad (3)$$

From Lemma 3, we obtain the following theorem;

Theorem 6 *The PI model holds, if and only if* $\rho_s = 0$ *and the PLL model holds.*

Proof. Under the PLL model, equation (3) is expressed as

$$\sum_{i=1}^{r-1}\sum_{j=i+1}^{r}\sum_{s=1}^{c-1}\sum_{t=s+1}^{c}(r_j^X - r_i^X)(r_t^Y - r_s^Y)p_{js}p_{it}(\theta^{(u_j-u_i)(v_t-v_s)} - 1) = 0.$$

Thus, $\rho_s = 0$ holds if and only if $\theta = 1$ (i.e., the PI model holds). The proof is completed.

3. Orthogonal Decomposition of the PI Model

Let n_{ij} denote the observed frequency in the cell of ith row and jth column of the table ($i = 1, \ldots, r$; $j = 1, \ldots, c$). Assume that a multinomial distribution applies to the $r \times c$ table. The maximum likelihood estimates of expected frequencies under the models can be obtained by using a iterative procedure, for example, the general iterative procedure for log-linear models of Darroch and Ratcliff (1972) or using the Newton-Raphson method to the log-likelihood equations.

Let $G^2(M)$ denote the likelihood ratio chi-squared statistic for testing goodness-of-fit of model M, namely,

$$G^2(M) = 2\sum_{i=1}^{r}\sum_{j=1}^{c}n_{ij}\log\left(\frac{n_{ij}}{\hat{m}_{ij}}\right),$$

where $\hat{m}_{ij}$ is the maximum likelihood estimate of expected frequency m_{ij} under the model M. Each model can be tested for goodness-of-fit by the likelihood ratio chi-squared statistic with the corresponding degrees of freedom (df). The numbers of df for the PI model, PLL model, and $\rho(U,V) = 0$ are $rc-3$, $rc-4$, and 1, respectively. We obtain the following theorem;

Theorem 7 *The test statistic $G^2(PI)$ is asymptotically equivalent to the sum of $G^2(\rho(U,V) = 0)$ and $G^2(PLL)$.*

Proof. Let $\boldsymbol{p} = (p_{11}, \ldots, p_{1c}, p_{21}, \ldots, p_{2c}, \ldots, p_{r1}, \ldots, p_{rc})^t$ denote the $rc \times 1$ vector, where "t" denotes the transposed of vector (or matrix). The structure of $\rho(U,V) = 0$ can be expressed as

$$H_1(\boldsymbol{p}) = \sum_{i=1}^{r}\sum_{j=1}^{c} u_i v_j p_{ij} - \left(\sum_{i=1}^{r} u_i p_{i\cdot} \sum_{j=1}^{c} v_j p_{\cdot j}\right) = 0_{d_1},$$

where $\boldsymbol{0}_s$ is the vector (or scalar) of order $s \times 1$ with all elements zero and $d_1 = 1$. The PLL model can be expressed as

$$\boldsymbol{H}_2(\boldsymbol{p}) = (h_{12}, h_{13}, \ldots, h_{1,c-1}, h_{21}, \ldots, h_{2,c-1}, \ldots, h_{r-1,c-1}, a_2, \ldots, a_{c-1}, b_2, \ldots, b_{r-1})^t = \boldsymbol{0}_{d_2},$$

where

$$\begin{aligned} h_{kl} &= \frac{1}{(u_{k+1}-u_k)(v_{l+1}-v_l)} \log\theta_{(k<k+1;l<l+1)} - \frac{1}{(u_2-u_1)(v_2-v_1)} \log\theta_{(1<2;1<2)}, \\ a_l &= \frac{1}{(v_{l+1}-v_l)} \log\left(\frac{p_{1,l+1}}{p_{1,l}}\right) - \frac{1}{(v_2-v_1)} \log\left(\frac{p_{12}}{p_{11}}\right), \\ b_k &= \frac{1}{(u_{k+1}-u_k)} \log\left(\frac{p_{k+1,1}}{p_{k,1}}\right) - \frac{1}{(u_2-u_1)} \log\left(\frac{p_{21}}{p_{11}}\right), \end{aligned}$$

and $d_2 = rc-4$. From Theorem 4, the PI model can be expressed as

$$\boldsymbol{H}_3(\boldsymbol{p}) = \left(H_1(\boldsymbol{p}), \boldsymbol{H}_2(\boldsymbol{p})^t\right)^t = \boldsymbol{0}_{d_3},$$

where $d_3 = rc-3$. Let $\boldsymbol{h}_s(\boldsymbol{p})$ $(s = 1,2,3)$ denote the $d_s \times rc$ matrix (or vector) of partial derivatives of $\boldsymbol{H}_s(\boldsymbol{p})$ with respect to $\boldsymbol{p}$, i.e.,

$$\boldsymbol{h}_s(\boldsymbol{p}) = \frac{\partial \boldsymbol{H}_s(\boldsymbol{p})}{\partial \boldsymbol{p}^t}.$$

Let $\boldsymbol{\Sigma}(\boldsymbol{p})$ denotes the inverse of information matrix, i.e., $\boldsymbol{\Sigma}(\boldsymbol{p}) = \mathrm{diag}(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}^t$, where $\mathrm{diag}(\boldsymbol{p})$ denotes a diagonal matrix with ith component of $\boldsymbol{p}$ as ith diagonal component. Let $\hat{\boldsymbol{p}}$ denotes $\boldsymbol{p}$ with $\{p_{ij}\}$ replaced by $\{\hat{p}_{ij}\}$, where $\hat{p}_{ij} = n_{ij}/n$ and $n = \sum_i \sum_j n_{ij}$. Using the delta method, $\boldsymbol{H}_3(\hat{\boldsymbol{p}})$ has asymptotically (as $n \to \infty$) a normal distribution with mean $\boldsymbol{H}_3(\boldsymbol{p})$ and covariance matrix;

$$\frac{1}{n}\boldsymbol{h}_3(\boldsymbol{p})\boldsymbol{\Sigma}(\boldsymbol{p})\boldsymbol{h}_3(\boldsymbol{p})^t = \frac{1}{n}\begin{bmatrix} \boldsymbol{h}_1(\boldsymbol{p})\boldsymbol{\Sigma}(\boldsymbol{p})\boldsymbol{h}_1(\boldsymbol{p})^t & \boldsymbol{h}_1(\boldsymbol{p})\boldsymbol{\Sigma}(\boldsymbol{p})\boldsymbol{h}_2(\boldsymbol{p})^t \\ \boldsymbol{h}_2(\boldsymbol{p})\boldsymbol{\Sigma}(\boldsymbol{p})\boldsymbol{h}_1(\boldsymbol{p})^t & \boldsymbol{h}_2(\boldsymbol{p})\boldsymbol{\Sigma}(\boldsymbol{p})\boldsymbol{h}_2(\boldsymbol{p})^t \end{bmatrix}.$$

Then, we can see

$$\boldsymbol{h}_1(\boldsymbol{p})\boldsymbol{\Sigma}(\boldsymbol{p})\boldsymbol{h}_2(\boldsymbol{p})^t = \boldsymbol{0}_{d_2}^t.$$

Namely,

$$\boldsymbol{h}_3(\boldsymbol{p})\boldsymbol{\Sigma}(\boldsymbol{p})\boldsymbol{h}_3(\boldsymbol{p})^t = \begin{bmatrix} \boldsymbol{h}_1(\boldsymbol{p})\boldsymbol{\Sigma}(\boldsymbol{p})\boldsymbol{h}_1(\boldsymbol{p})^t & \boldsymbol{0}_{d_2}^t \\ \boldsymbol{0}_{d_2} & \boldsymbol{h}_2(\boldsymbol{p})\boldsymbol{\Sigma}(\boldsymbol{p})\boldsymbol{h}_2(\boldsymbol{p})^t \end{bmatrix}.$$

Thus, we obtain $\Delta_3(\boldsymbol{p}) = \Delta_1(\boldsymbol{p}) + \Delta_2(\boldsymbol{p})$, where

$$\Delta_s(\boldsymbol{p}) = \boldsymbol{H}_s(\boldsymbol{p})^t[\boldsymbol{h}_s(\boldsymbol{p})\boldsymbol{\Sigma}(\boldsymbol{p})\boldsymbol{h}_s(\boldsymbol{p})^t]^{-1}\boldsymbol{H}_s(\boldsymbol{p}). \tag{4}$$

Under each $\boldsymbol{H}_s(\boldsymbol{p}) = \boldsymbol{0}_{d_s}$ $(s = 1,2,3)$, the Wald statistic $W_s = n\Delta_s(\hat{\boldsymbol{p}})$ has asymptotically a chi-squared distribution with d_s degrees of freedom. From equation (4), we see that $W_3 = W_1 + W_2$. From the asymptotic equivalence of the Wald statistic and the likelihood ratio statistic (Rao, 1973, Sec. 6e. 3; Darroch and Silvey, 1963; Aitchison, 1962), $G^2(PI)$ is asymptotically equivalent to the sum of $G^2(\rho(U,V)=0)$ and $G^2(PLL)$. Note that the numbers of df for testing $\boldsymbol{H}_s(\boldsymbol{p}) = \boldsymbol{0}_{d_s}$ are d_s $(s = 1, 2, 3)$. Thus Theorem 7 is obtained. The proof is completed.

4. Examples

In this section, we use the known integer scores $\{u_i = i\}$, $\{v_j = j\}$ for rows and columns to simplify the problems.

4.1 Example 1

We consider the data in Table 1, obtained in Grizzle et al. (1969). These data have four different operations for treating duodenal ulcer patients correspond to removal of various amounts of the stomach. Operation A1 is drainage and vagotomy, A2 is 25% resection (antrectomy) and vagotomy, A3 is 50% resection (hemigastrectomy) and vagotomy, and A4 is 75% resection. The categories of operation variable have a natural ordering. The dumping severity variable describes the extent of an undesirable potential consequence of the operation (none, slight and moderate), which are also ordered.

When we apply the PLL model for these data, the PLL model fits well with $G^2 = 7.87$ based on df = 8. Also the PI model fits well with $G^2 = 13.61$ based on df = 9. For testing the hypothesis that the PI model holds under the assumption that the PLL model holds, the likelihood ratio statistic $G^2(PI \mid PLL)$ is given as $G^2(PI) - G^2(PLL) = 5.74$ based on df $= 9 - 8 = 1$. Therefore this hypothesis is rejected at 0.05 significance level. Hence we prefer the PLL model to the PI model for the data in Table 1. Also the likelihood ratio statistic $G^2(PLL \mid LL)$ is given as $G^2(PLL) - G^2(LL) = 3.28$ based on df $= 8 - 5 = 3$. Therefore this hypothesis is accepted at 0.05 significance level. Hence we prefer the PLL model to the LL model for the data in Table 1. Under the PLL model, the maximum likelihood estimates of α and β are 0.83 and 0.32, respectively, and the maximum likelihood estimate of θ is 1.16. From Table 3, for any fixed row i, all local odds ω^X_{ij} $(j = 1, 2)$ are estimated to be smaller than 1. Also, the odds ω^X_{i1} (and $\omega^X_{i1}\omega^X_{i2}$) are estimated to increase as the row i increase. Thus it is inferred that the Damping severity tend to worse as the Operation levels increases.

Table 1. Cross-classification of duodenal ulcer patients according to Operation and Dumping Severity; from Grizzle et al. (1969). (The parenthesized values are the maximum likelihood estimates of expected frequencies under the PLL model.)

	Dumping Severity			
Operation	None	Slight	Moderate	Total
A1	61	28	7	96
	(65.66)	(24.35)	(9.03)	
A2	68	23	13	104
	(62.98)	(27.07)	(11.64)	
A3	58	40	12	110
	(60.41)	(30.10)	(15.00)	
A4	53	38	16	107
	(57.95)	(33.47)	(19.34)	
Total	240	129	48	417

Table 2. Likelihood ratio chi-squared values for the testing the models and structures applied to Table 1.

Models	df	G^2
PI	9	13.61
PLL	8	7.87
I	6	10.88
LL	5	4.59
$\rho(U,V) = 0$	1	6.35*
$\tau_b = 0$	1	6.73*
$\rho_s = 0$	1	6.73*

$*$ means significant at 0.05 level.

Table 3. Maximum likelihood estimates of (local) odds of classification in column 2 and column 3 instead of column 1 for a fixed row i $(i = 1, 2, 3, 4)$ under the PLL model, applied to Table 1.

Row i	ω^X_{i1} $(= p_{i2}/p_{i1})$	$\omega^X_{i1}\omega^X_{i2}$ $(= p_{i3}/p_{i1})$
1	0.37	0.14
2	0.43	0.18
3	0.50	0.25
4	0.58	0.33

4.2 Example 2

The data in Table 4, obtained in Fienberg (1980, p. 20), present the relationship between aptitude (as measured at an earlier data by a scholastic aptitude test) and occupation. Occupation level O1 is self-employed, business, O2 is self-employed, professional, O3 is teacher and O4 is salaried, employed. From Table 5 we see that the PI and PLL models fit these data poorly, however, the tests for $\rho(U, V) = 0$, $\tau_b = 0$, and $\rho_s = 0$ are accepted. From Theorems 4, 5 and 6, we see that the poor fit of the PI model is caused by the influence of the lack of structure of the PLL model (not the lack of the $\rho(U, V) = 0$, $\tau_b = 0$, and $\rho_s = 0$).

From Table 5, we see that the I model fits these data poorly. Thus, we can interpret that row and column variables are not independent, although the correlations of row and column variables are equal to zero. These data are one example that when the I model holds, $\rho(U, V) = 0$ is true, however, converse does not always holds.

Table 4. Cross-classification of subjects according to the aptitude and the occupation; from Fienberg (1980, p. 20). (The parenthesized values are the maximum likelihood estimates of expected frequencies under the structure of $\rho(U, V) = 0$.)

	Occupational level				
Aptitude	O1	O2	O3	O4	Totals
(low) A1	122	30	20	472	644
	(119.48)	(29.65)	(19.95)	(475.40)	
A2	226	51	66	704	1047
	(223.92)	(50.74)	(65.93)	(706.23)	
A3	306	115	96	1072	1589
	(306.76)	(115.16)	(96.03)	(1071.10)	
A4	130	59	38	501	728
	(131.88)	(59.47)	(38.06)	(498.59)	
(high) A5	50	31	15	249	345
	(51.34)	(31.45)	(15.04)	(246.82)	
Totals	834	286	235	2998	4353

Table 5. Likelihood ratio chi-squared values for the testing the models and structures applied to Table 4.

Models	df	G^2
PI	17	3086.43*
PLL	16	3086.21*
I	12	37.41*
LL	11	37.20*
$\rho(U, V) = 0$	1	0.22
$\tau_b = 0$	1	0.003
$\rho_s = 0$	1	0.002

$*$ means significant at 0.05 level.

5. Concluding Remarks

When the PI model fits the data poorly, Theorems 4, 5, and 6 may be useful for seeing the reason for the poor fit, namely, which of the lack of the structures $\rho(U, V) = 0$, $\tau_b = 0$ and $\rho_s = 0$ and the lack of the PLL model influences strongly. We point out from Theorem 7 that the statistic for testing the PI model under the assumption that the PLL model holds, i.e., $G^2(PI) - G^2(PLL)$, is asymptotically equivalent to the statistic for testing the $\rho(U, V) = 0$, i.e., $G^2(\rho(U, V) = 0)$. We emphasize that testing the PI model is not equivalent to testing the $\rho(U, V) = 0$. We saw in Example 2 that the structure of $\rho(U, V) = 0$ holds, however, the PI model does not hold.

6. Discussion

Tomizawa (1992) also described the parsimonious uniform (PU) association model. It is a special case of the PLL model obtained by using integer scores $\{u_i = i\}$, $\{v_j = j\}$ or equal interval scores for rows and columns. We may obtain the theorems changed the PLL model into the PU model in a similar manner to this paper.

Acknowledgements

The authors would like to thank the editor and referee for the helpful comments.

References

Agresti, A. (1983). A survey of strategies for modeling cross-classifications having ordinal variables. *Journal of the American Statistical Association, 78*, 184-198.

Agresti, A. (1984). *Analysis of ordinal categorical data.* New York, NY: Wiley.

Aitchison, J. (1962). Large-sample restricted parametric tests. *Journal of the Royal Statistical Society, Series B, 24*, 234-250.

Bross, I. D. J. (1958). How to use ridit analysis. *Biometrics, 14*, 18-38.

Darroch, J. N., & Ratcliff, D. (1972). Generalized iterative scaling for log-linear models. *Annals of Mathematical Statistics, 43*, 1470-1480.

Darroch, J. N., & Silvey, S. D. (1963). On testing more than one hypothesis. *Annals of Mathematical Statistics, 34*, 555-567.

Fienberg, S. E. (1980). *The analysis of cross-classified categorical data* (2nd ed.). Cambridge: The MIT Press.

Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical method for rates and proportions* (3rd ed.). New York, NY: Wiley.

Goodman, L. A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association, 74*, 537-552.

Grizzle, J. E., Starmer, C. F., & Koch, G. G. (1969). Analysis of categorical data by linear models. *Biometrics, 25*, 489-504.

Kendall, M. G. (1945). The treatment of ties in ranking problems. *Biometrika, 33*, 239-251.

Kendall, M. G., & Gibbons, J. D. (1990). *Rank correlation methods* (5th ed.). London: Edward Arnold.

Rao, C. R. (1973). *Linear statistical inference and its applications* (2nd ed.). New York, NY: Wiley.

Stuart, A. (1963). Calculation of Spearman's rho for ordered two-way classifications. *American Statistician, 17*, 23-24.

Tahata, K., Miyamoto, N., & Tomizawa, S. (2008). Decomposition of independence using Pearson, Kendall and Spearman's correlations and association model for two-way classifications. *Far East Journal of Theoretical Statistics, 25*, 273-283.

Tomizawa, S. (1992). More parsimonious linear-by-linear association model in the analysis of cross-classifications having ordered categories. *Biometrical Journal, 34*, 129-140.

Tomizawa, S., Miyamoto, N., & Sakurai, M. (2008). Decomposition of independence model and separability of its test statistic for two-way contingency tables with ordered categories. *Advances and Applications in Statistics, 8*, 209-218.

Copyrights