Analysis of Data in Square Contingency Tables with Ordered Categories Using the Conditional Symmetry Model and its Decomposed Models

by Sadao Tomizawa*

For the analysis of square contingency tables with ordered categories, three kinds of decompositions for the conditional symmetry model derived by Tomizawa are simply described. Using the conditional symmetry model and its decomposed models, this paper analyzes the data of unaided distance vision of women in Britain first analyzed by Stuart, the data of unaided distance vision of students in a university in Japan, and the data of unaided distance vision of pupils at elementary schools at a city in Tokyo.

Introduction

Table 1 is constructed from the data of the unaided distance vision of 7477 women aged 30–39 employed in Royal Ordnance factories in Britain from 1943 to 1946. The data in Table 1 were first analyzed by Stuart (1,2). Table 2 is constructed from the data of the unaided distance vision of 4746 students aged 18 to about 25, including about 10% of the women of the Faculty of Science and Technology, Science University of Tokyo in Japan examined in April, 1982. Table 3 is constructed from the data of the unaided distance vision of 3168 pupils aged 6–12, including about half the girls at elementary schools in Tokyo, Japan examined in June 1984. In Tables 1, 2, and 3 the row variable is the right eye grade and the column variable is the left eye grade with the categories ordered from the lowest grade (1) to the highest grade (4).

To the data of Tables 1, 2, and 3 it is reasonable to apply models of various kinds of symmetry instead of the statistical independence model. To analyze the data of square contingency tables, the models of symmetry, quasisymmetry and marginal homogeneity are described, for example, in Bishop, Fienberg, and Holland (3), Caussinus (4), and Stuart (2). Caussinus (4) also noted that the symmetry model holds if and only if both quasisymmetry model and marginal homogeneity model hold. McCullagh (5) proposed a conditional symmetry model which is an extension of the symmetry model. Tomizawa (6) derived three kinds of decompositions for the conditional symmetry model. Wall and Lienert (7) proposed a point-symmetry model in J-dimensional contingency cubes. Tomizawa (8,9) proposed models of various kinds of point symmetry in two-dimensional contingency tables and gave their decompositions.

In this paper we analyze the data in Tables 1, 2, and 3 using the conditional symmetry model and its decomposed models.

Table 1. Unaided distance vision of 7477 women aged 30–39 employed in Royal Ordnance factories from 1943 to 1946.

| Right eye grade | Left eye grade | Table 2. Unaided distance vision of 4746 students aged 18 to about 25 including about 10% women in Faculty of Science and Technology, Science University of Tokyo in Japan examined in April 1982. |
|-----------------|----------------|--------------------------------------------------------------------------------------------------|
|                 | Lowest (1)     | Second (2)                                        | Third (3)                                      | Highest (4)       | Total      |
| Lowest (1)      | 492            | 179                                                | 82                                             | 36               | 789        |
| Second (2)      | 205            | 172                                                | 362                                            | 117              | 2456       |
| Third (3)       | 78             | 432                                                | 1512                                           | 234              | 2256       |
| Highest (4)     | 66             | 124                                                | 356                                            | 1520             | 1976       |
| Total           | 841            | 2507                                               | 2222                                           | 1907             | 7477       |

*Department of Information Sciences, Faculty of Science and Technology, Science University of Tokyo, Noda City, Chiba 278 Japan.
Table 3. Unaided distance vision of 3168 pupils aged 6–12 
including about half girls at elementary schools in Tokyo 
examined in June 1984.

| Right eye grade | Lowest (1) | Second (2) | Third (3) | Highest (4) | Total |
|------------------|------------|------------|-----------|-------------|-------|
| Lowest (1)       | 92         | 16         | 7         | 12          | 127   |
| Second (2)       | 15         | 75         | 42        | 10          | 142   |
| Third (3)        | 5          | 33         | 138       | 96          | 272   |
| Highest (4)      | 10         | 21         | 126       | 2470        | 2627  |
| Total            | 122        | 145        | 313       | 2588        | 3168  |

Models and Decompositions

Consider the square $a \times a$ contingency table with 
row variate denoted by $X_1$ and column variate denoted 
by $X_2$. Let $p_{ij}$ denote the probability in the cell in row $i$ 
and column $j$ for $1 \leq i, j \leq a$. 

The models of symmetry, quasisymmetry and 
marginal homogeneity are defined as follows:

$H_S^*: p_{ij} = p_{ji}$ for $1 \leq i < j \leq a$

$H_Q^*: p_{ij} = a_{ij} a_{ij}$ for $1 \leq i, j \leq a$

where $a_i > 0$, $b_j > 0$, $d_{ij} > 0$, and $d_{ij} = d_{ji}$.

$H_M^*: p_{ii} = p_{ii}$ for $1 \leq i \leq a$

where

$$ p_{ii} = \sum_{i=1}^{a} p_{ii} $$

and

$$ p_{ii} = \sum_{i=1}^{a} p_{ii} $$

McCullagh's (5) conditional symmetry model is defined by

$$ H_S^*: p_{ij} = \begin{cases} 0 & \Phi_{ij} (i < j) \\ \Phi_{ii} & (i = j) \\ (2 - \theta) \Phi_{ij} & (i > j) \end{cases} $$

where $\Phi_{ij} = \Phi_{ji}$ and $\Sigma \Phi_{ij} = 1$. This model can be also 
expressed by a log-linear model for the $p_{ij}$ in Tomizawa 
(6).

We next define the extended quasisymmetry model by

$$ H_Q^*: p_{ij} = a_{ij} a_{ij} $$

where $a_i > 0$, $b_j > 0$, $d_{ij} > 0$

$$ d_{ij} = \gamma d_{ij} \text{ for } 1 \leq i < j \leq a $$

and where the parameter $\gamma$ is unspecified. This model is also 
equivalent to the model

$$ H_Q^*: p_{ij} p_{jk} p_{ki} = \gamma \pi_{ij} \pi_{jk} \pi_{ki} \text{ for } 1 \leq i < j < k \leq a $$

where the parameter $\gamma$ is unspecified. Model $H_Q^*$ is also 
expressed by a log-linear model for the $p_{ij}$ in Tomizawa 
(6). A special case of model $H_Q^*$ obtained by putting $\gamma = 1$ is the quasisymmetry model.

We introduce two kinds of modified marginal homogeneity models as follows:

$$ H_m*: p_{ii} = \delta p_{ii} \text{ for } 1 \leq i \leq a - 1 $$

where the parameter $\delta$ is unspecified and where

$$ p_{ii}^{*} = \Pr(X_1 = i, X_1 < X_2) = \sum_{t=1}^{a} p_{it} $$

$\delta p_{ii} = \Pr(X_2 = i, X_1 > X_2) = \sum_{t=1}^{a} p_{it}$

$$ H_{M2}^*: p_{ii} = \delta p_{ii} \text{ for } 2 \leq i \leq a $$

where the parameter $\delta$ is unspecified and where

$$ p_{ii}^{*} = \Pr(X_2 = i, X_1 < X_2) = \sum_{t=1}^{i-1} p_{it} $$

$\delta p_{ii} = \Pr(X_1 = i, X_1 > X_2) = \sum_{t=1}^{i-1} p_{it}$

We define the extended marginal homogeneity model by

$$ H_m*: p_{ii}^{(b)} = p_{ii} \text{ for } 1 \leq i \leq a $$

where the parameter $\delta$ is unspecified and where

$$ p_{ii}^{(b)} = \delta p_{ii} + p_{ii} + p_{ii}^{*} \text{ and } p_{ii}^{(b)} = p_{ii} + p_{ii} + \delta p_{ii} $$

Model $H_m^*$ indicates that the row marginal totals 
summed by multiplying the probabilities for the cells in 
the lower-left triangle of the $a \times a$ table by a common 
weight $\delta$ are equal to the column marginal totals.
summed by the same way. Model $H_{M3}$ also has the property that under $H_{M3}$ the parameter $\delta \geq 1$ is equivalent to $\Pr(X_i \leq i) \approx \Pr(X_k \leq i)$ for every $i$. A special case of model $H_{M3}$ obtained by putting $\delta = 1$ is the marginal homogeneity model.

Finally let

$$R = \sum_{i< j< k} \{p_{ij}p_{jk}/(p_{ij}p_{jk}p_{ik})\}^{(a-1)/3}$$

and introduce three kinds of average models as follows:

- $H_{R1}$: $R = \Delta_1$
  
  where $\Delta_1 = \sum_{i=1}^{a-1} (p_{ii}/p_{ii})(a-1)$

- $H_{R2}$: $R = \Delta_2$

where $\Delta_2 = \sum_{i=2}^{a} (p_{ii}/p_{ii})(a-1)$

- $H_{R3}$: $R = \Delta_3$

where $\Delta_3 = (p_{11}/p_{11} + p_{aa}/p_{aa})/2$

Here $R$ indicates the average of ratio $p_{ij}p_{jk}/(p_{ij}p_{jk}p_{ik})$ for $1 \leq i < j < k \leq a$ in the case that the ratio parameter $\gamma$ in $H_Q$ changes according to each ratio, and $\Delta_1$ indicates the average of $p_{ii}/p_{ii}$ for $1 \leq i \leq a-1$ in the case that the ratio parameter $\delta$ in $H_{M1}$ changes according to each ratio, and also $\Delta_2$ is interpreted similarly, and $\Delta_3$ indicates the average of two ratios in the case that the ratio $p_{ii}/p_{ii}$ is not always equal to the ratio $p_{aa}/p_{aa}$. Therefore model $H_{Rl}$ for $l = 1, 2, 3$ indicates the equilibriums of two kinds of averages $R$ and $\Delta_1$.

We get the decompositions for model $H_S$ as follows.

**Theorem:** for $l = 1, 2, 3$, model $H_S$ holds if and only if all models $H_Q$, $H_{M1}$, and $H_{Rl}$ hold.

The proof of this theorem is given by Tomizawa (6).

We denote special cases of models $H_{Ml}(l = 1, 2)$ obtained by putting $\delta = 1$ by $H_{M1}(l = 1, 2)$. Then we get new decompositions for model $H_S$ as follows:

**Corollary:** for $l = 1, 2$, model $H_S$ holds if and only if both models $H_Q$ and $H_{Ml}$ hold.

### Degrees of Freedom, Estimation, and Test

Let $x_{ij}$ denote the observed frequency in the $ith$ row and $jth$ column of the $a \times a$ contingency table $(1 \leq i, j \leq a)$ where $\Sigma x_{ij} = N$, and let $m_{ij}$ denote the corresponding expected frequency under some model. We assume here that a multinomial distribution applies to the $a \times a$ table.

The degrees of freedom for models $H_{S1}$, $H_{Q1}$, $H_{M1}$, $H_{R1}$ ($l = 1, 2, 3$) and $H_{M2}(l = 1, 2)$ are $(a - 2)(a + 1)/2$, $(a - 3)/2, a - 2, 1$, and $a - 1$, respectively.

The maximum likelihood estimates of $m_{ij}$ under model $H_Q$ can be sought by the iterative procedure in Tomizawa (6) or by the following iterative procedure: as the $(k+1)$th step

$$m^{(k+1)}_{ij} = m^{(k)}_{ij} \frac{\sum_{v=1}^{a} x_{iv}x_{v}^{(k)}[x_{ij} + x_{ij}]}{m^{(k)}_{ij} + m^{(k)}_{ij}} \frac{1}{D} \frac{1}{E}$$

where the initial values are $m^{(0)}_{ij} = 1$ for $1 \leq i, j \leq a$ and where

$$x_{ij} = \sum_{v=1}^{a} x_{ij}$$

$$x_{ij} = \sum_{v=1}^{a} x_{ij}$$

$$m^{(k)}_{ij} = \sum_{v=1}^{a} m^{(k)}_{ij}$$

$$d_{ij} = (a + 1 - (j - i))/(a - 2) \quad (i < j)$$

$$d_{ij} = 1/2 \quad (i \geq j)$$

$$D = (a - 2)N + \sum_{i<j} x_{ij}(a - 2(j - i))$$

$$E = 2(a - 2)N - D$$

$$D^{(k)} = (a - 2) \sum_{i<j} m^{(k)}_{ij} + \sum_{i<j} m^{(k)}_{ij} \{a - 2(j - 2)\}$$

$$E^{(k)} = 2(a - 2) \sum_{i<j} m^{(k)} - D^{(k)}$$

Thus the goodness of fit of models $H_{S1}$, $H_{Q1}$, $H_{M1}$ and $H_{M2}(l = 1, 2)$ can be tested by the Pearson's or the likelihood-ratio chi-squared statistics. The goodness of fit of model $H_{M3}$ can be tested by test statistic $\chi_{M3}^2$ in Tomizawa (6).
Table 4. Chi-square for symmetry models applied to the data in Table 1.

| Symmetry models | Degrees of freedom | Likelihood-ratio chi-square | Pearson's chi-square |
|-----------------|--------------------|-----------------------------|----------------------|
| $H_0$           | 6                  | 19.25                       | 19.11                |
| $H_1$           | 3                  | 7.27                        | 7.26                 |
| $H_{M1}$        | 3                  | 11.97                       | 11.96                |
| $H_{M2}$        | 3                  | 11.99                       | 11.97                |
| $H_2$           | 5                  | 7.35                        | 7.26                 |
| $H_3$           | 2                  | 6.82                        | 6.78                 |
| $H_4$           | 2                  | 0.08                        | 0.08                 |
| $H_5$           | 2                  | 0.99                        | 0.99                 |

Table 5. Chi-square for symmetry models applied to the data in Table 2.

| Symmetry models | Degrees of freedom | Likelihood-ratio chi-square | Pearson's chi-square |
|-----------------|--------------------|-----------------------------|----------------------|
| $H_0$           | 6                  | 16.95                       | 16.87                |
| $H_1$           | 3                  | 5.71                        | 5.78                 |
| $H_{M1}$        | 3                  | 12.52                       | 12.49                |
| $H_{M2}$        | 3                  | 13.94                       | 13.90                |
| $H_2$           | 5                  | 4.98                        | 4.97                 |
| $H_3$           | 2                  | 4.41                        | 4.39                 |
| $H_4$           | 2                  | 0.54                        | 0.54                 |
| $H_5$           | 2                  | 1.96                        | 1.96                 |

Table 6. Chi-square for symmetry models applied to the data in Table 3.

| Symmetry models | Degrees of freedom | Likelihood-ratio chi-square | Pearson's chi-square |
|-----------------|--------------------|-----------------------------|----------------------|
| $H_0$           | 6                  | 9.69                        | 9.58                 |
| $H_1$           | 3                  | 2.81                        | 2.75                 |
| $H_{M1}$        | 3                  | 4.49                        | 4.48                 |
| $H_{M2}$        | 3                  | 6.98                        | 6.95                 |
| $H_2$           | 5                  | 7.83                        | 7.77                 |
| $H_3$           | 2                  | 2.61                        | 2.57                 |
| $H_4$           | 2                  | 2.63                        | 2.63                 |
| $H_5$           | 2                  | 5.12                        | 5.12                 |

Analysis of Table 1

Table 4 presents the likelihood-ratio and the Pearson's chi-squared statistics obtained by applying the models introduced in the previous section to the data in Table 1. The value of test statistic $Q$ in Stuart (2) for testing the goodness of fit of model $H_M$ is 11.96 with 3 degrees of freedom. The value of test statistic $\chi^2_{M3}$ in Tomizawa (6) for testing the goodness of fit of the extended marginal homogeneity model $H_{M3}$ is 0.005 with 2 degrees of freedom. From these values and Table 4, none of models $H_{M1}^*$, $H_{M2}^*$, and $H_M^*$ fits the data well, but all of models $H_{M1}$, $H_{M2}$, and $H_{M3}$ fit the data very well. Moreover, the maximum likelihood estimates of $\delta$ under models $H_{M1}$ and $H_{M2}$ are 0.863. Since this value is less than one, we can say that the left eye is worse than the right eye. Also, the values of chi-square under model $H_Q^*$ lie between the upper 5% and 1% tail values of the $\chi^2$ distribution with 2 degrees of freedom. Under model $H_Q^*$ the estimated value of $\gamma$ obtained by maximum likelihood is 0.929. Also model $H_S$ does not fit the data well, and thus the left eye is not symmetric to the right eye. But model $H_S$ fits adequately and the estimated values of $m_{ij}/m_{ji}$ for $1 \leq i < j \leq 4$ under model $H_S$ are 0.863. Since this value is less than one, we can say again that the left eye is worse than the right eye.

Analysis of Table 2

Table 5 presents the likelihood-ratio and the Pearson's chi-squared statistics obtained by applying various kinds of symmetry models to the data in Table 2. The value of test statistic $Q$ in Stuart (2) for testing the goodness of fit of model $H_M$ is 11.21 with 3 degrees of freedom. This value lies between the upper 5% and 1% tail values of the $\chi^2$ distribution with 3 degrees of freedom. The value of test statistic $\chi^2_{M3}$ in Tomizawa (6) for testing the goodness of fit of the extended marginal homogeneity model $H_{M3}$ is 0.56 with 2 degrees of freedom. From this value and Table 5, neither model $H_{M1}$ nor model $H_{M2}$ fits the data well but all of models $H_{M1}$, $H_{M2}$, and $H_{M3}$ fit the data very well. Moreover, the maximum likelihood estimates of $\delta$ under models $H_{M1}$ and $H_{M2}$ are 1.228, and since this value is greater than one, this value indicates that the left eye is better than the right eye. Both models $H_Q$ and $H_Q^*$ also fit adequately, and under model $H_Q$, the estimated value of $\gamma$ obtained by maximum likelihood is 1.208. Also, since model $H_S$ does not fit the data well, the left eye is not symmetric to the right eye. But model $H_S$ fits the data well, and the estimated values of $m_{ij}/m_{ji}$ for $1 \leq i < j \leq 4$ under model $H_S$ are 1.228. Since this value is greater than one, we can say again that the left eye is better than the right eye.

Analysis of Table 3

Table 6 presents the likelihood-ratio and the Pearson's chi-squared statistics obtained by applying various kinds of symmetry models to the data in Table 3. The value of test statistic $Q$ in Stuart (2) for testing the goodness of fit of model $H_M$ is 6.85 with 3 degrees of freedom, and the value of test statistic $\chi^2_{M3}$ in Tomizawa (6) for testing the goodness of fit of the extended marginal homogeneity model $H_{M3}$ is 4.16 with 2 degrees of freedom. From these values and Table 6, all models fit the data well. Moreover, under model $H_Q$, the estimated value of $\gamma$ obtained by maximum likelihood is 1.136 and the maximum likelihood estimates of $\delta$ under models $H_{M1}$ and $H_{M2}$ are 0.871. We may consider these values close upon one because models $H_Q$, $H_{M1}$, and $H_{M2}$ hold. Also the values of statistics for the goodness of fit of models $H_S$ and $H_{M1}^*(l=1, 2, 3)$ applied to Table 3 are greater than those applied to Table 1 and 2; namely, the goodness of fit of models $H_S$ and $H_{M1}^*(l=1, 2, 3)$ applied to Table 3 are not so good as those applied to Table 1 and 2. By the way, models $H_S$, $H_{M1}$, $H_{M2}$, and $H_M$ applied to Table 1 and 2 did not fit the data well, but those applied to Table 3 fit the data well. Therefore, for the data in Table 3 we can say that the left eye is symmetric to the right eye in the various senses.
The author would like to express his sincere thanks to Professor K. Kunisawa, Mr. M. Jimbo, and Dr. K. Shimizu in Science University of Tokyo for their valuable advice and encouragements.

The author is grateful to Dr. A. Nakajima and Dr. K. Fujiki in Department of Ophthalmology, Juntendo University, for their valuable advice and for giving him the data in Table 3 in this paper.

He is also deeply grateful to Dr. A. Kudo and Dr. T. Yanagawa in Kyushu University who provided him with an excellent opportunity for presenting this paper at the Environmental Risk Assessment and Statistical Methods Conference which is a satellite conference in conjunction with the 12th International Biometric Conference.

REFERENCES

1. Stuart, A. The estimation and comparison of strengths of association in contingency tables. Biometrika 40: 105–110 (1953).
2. Stuart, A. A test for homogeneity of the marginal distributions in a two-way classification. Biometrika 42: 412–416 (1955).
3. Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. Discrete Multivariate Analysis: Theory and Practice. MIT Press, Cambridge, MA. 1975.
4. Caussinus, H. Contribution à l’analyse statistique des tableaux de corrélation. Ann. Fac. Sci. Univ. Toulouse 29: 77–182 (1965).
5. McCullagh, P. A class of parametric models for the analysis of square contingency tables with ordered categories. Biometrika 65: 413–418 (1978).
6. Tomizawa, S. Three kinds of decompositions for the conditional symmetry model in a square contingency table. J. Japan Statist. Soc. 14: 35–42 (1984).
7. Wall, K. D., and Lienert, G. A. A test for point-symmetry in J-dimensional contingency-cubes. Biometrische Z. 18: 259–264 (1976).
8. Tomizawa, S. The decompositions for point symmetry models in two-way contingency tables. Biometrical J., in press.
9. Tomizawa, S. A decomposition for the inclined point-symmetry model in a square contingency table. Biometrical J., in press.