Analysis of Unaided Vision Data
using New Decomposition of Symmetry

[1]Kouji Yamamoto and [2]Sadao Tomizawa
[1]Center for Clinical Investigation and Research, Osaka University Hospital,
2-15, Yamadaoka, Suita, Osaka, 565-0871, Japan
[2]Department of Information Sciences, Faculty of Science and Technology,
Tokyo University of Science, Noda City, Chiba, 278-8510, Japan

Abstract: Problem statement: First, this study considers how the structure of symmetry for probabilities is decomposed into two structures. Secondly, this study infers the structure of unknown probabilities which indicates how the right eye is better (or worse) than the left eye for three kinds of data on unaided distance vision of (1) women in Britain, (2) students in an university of Japan and (3) pupils in elementary schools in Tokyo, Japan. This study proposes a new decomposition of symmetry model for probabilities and analyzes these vision data using the decomposition. **Approach:** This study considers a new decomposition theorem that for the probabilities the symmetry model (indicates that the right eye vision is symmetric to the left eye vision) holds. Also this study analyzes the vision data using this decomposition. **Results:** From the statistical approach, we can see that (1) for the vision data of women, the right eye is better than the left eye and the mean of right eye is not equal to the mean of left eye, (2) for the vision data of students, the right eye is worse than the left eye and the mean of right eye is not equal to the mean of left eye and (3) for the vision data of pupils, the right eye is symmetric to the left eye and the mean of right eye is equal to the mean of left eye. **Conclusion:** When the symmetry model fits the data poorly, this new decomposition is useful for seeing which of decomposed two models influences stronger. We can see the structure of asymmetry for vision data in more details.

Key words: Cumulative linear diagonals-parameter symmetry, eye, mean, model, square contingency table, unaided distance vision

INTRODUCTION

Consider three sets of data on unaided distance vision of (1) 7477 women aged 30-39 employed in Royal Ordnance factories in Britain from 1943-1946 (Table 1), (2) 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 (Table 2) and (3) 3168 pupils comprising nearly equal number of boys and girls aged 6-12 at elementary schools in Tokyo, Japan, examined in June 1984 (Table 3). In these data, the row variable is the right eye grade and the column variable is the left eye grade with the categories ordered from the Best (1) to the Worst (4). The data in Table 1 have also been analyzed by many statisticians including Stuart (1955); Bishop *et al*. (2007); McCullagh (1978); Goodman (1979); Agresti (1983); Tomizawa (1985; 1993); Miyamoto *et al*. (2004); Tomizawa *et al*. (2006) and Tomizawa and Tahata (2007). The data in Table 2 have been analyzed by Tomizawa (1984; 1985). The data in

Table 3 have also been analyzed by Tomizawa (1985) and Miyamoto *et al*. (2004).

For these vision data, an individual's right eye grade is strongly associated with his/her left eye grade because many observations concentrate on (or near) the main diagonal cells in each table. Therefore, instead of independence between an individual's right eye grade and his/her left eye grade, we are interested in whether or not an individual's right eye grade is symmetric to his/her left eye grade and in how both eyes are symmetric or asymmetric, for example, whether or not the mean of right eye grade is equal to the mean of left eye grade. In addition, we are interested in seeing the reason by considering a decomposition of structure of symmetry if there is not the probability structure of symmetry between the right eye vision and the left eye vision in each table. Many models of symmetry and asymmetry for probabilities have been proposed by many statisticians. Also decompositions of the symmetry model have been given by some statisticians.

Corresponding Author: Kouji Yamamoto, Center for Clinical Investigation and Research, Osaka University Hospital, 2-15, Yamadaoka, Suita, Osaka, 565-0871, Japan

Table 1: Unaided distance vision of 7477 women aged 30-39 employed in Royal Ordnance factories in Britain from 1943-1946; from Stuart (1955). (The parenthesized values are the maximum likelihood estimates of expected frequencies under the CLDPS model.)

Right eye grade	Left eye grade				
	Best (1)	Sec (2)	Third (3)	Worst (4)	Total
Best (1)	1520	266	124	66	1976
	(1520.00)	(256.31)	(132.83)	(62.32)	
Sec (2)	234	1512	432	78	2256
	(244.68)	(1512.00)	(412.05)	(86.69)	
Third (3)	117	362	1772	205	2456
	(105.99)	(383.42)	(1772.00)	(196.93)	
Worst (4)	36	82	179	492	789
	(40.56)	(71.34)	(187.88)	(492.00)	
Total	1907	2222	2507	841	7477

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, Japan, examined in April 1982; from Tomizawa (1984). (The parenthesized values are the maximum likelihood estimates of expected frequencies under the CLDPS model.)

Right eye grade	Left eye grade				
	Best (1)	Sec (2)	Third (3)	Worst (4)	Total
Best (1)	1291	130	40	22	1483
	(1291.00)	(133.38)	(44.32)	(15.16)	
Sec (2)	149	221	114	23	507
	(145.93)	(221.00)	(115.01)	(21.19)	
Third (3)	64	124	660	185	1033
	(59.69)	(123.11)	(660.00)	(201.30)	
Worst (4)	20	25	249	1429	1723
	(26.40)	(26.22)	(233.29)	(1429.00)	
Total	1524	500	1063	1659	4746

Table 3: Unaided distance vision of 3168 pupils comprising nearly equal number of boys and girls aged 6-12 at elementary schools in Tokyo, Japan, examined in June 1984; from Tomizawa (1985). (Upper and lower parenthesized values are the maximum likelihood estimates of expected frequencies under the S and ME models, respectively.)

Right eye grade	Left eye grade				
	Best (1)	Sec (2)	Third (3)	Worst (4)	Total
Best (1)	2470	126	21	10	2627
	(2470.00)	(111.00)	(15.50)	(11.00)	
	(2470.00)	(120.51)	(19.25)	(8.80)	
Sec (2)	96	138	33	5	272
	(111.00)	(138.00)	(37.50)	(6.00)	
	(100.58)	(138.00)	(31.56)	(4.58)	
Third (3)	10	42	75	15	142
	(15.50)	(37.50)	(75.00)	(15.50)	
	(11.00)	(44.01)	(75.00)	(14.35)	
Worst (4)	12	7	16	92	127
	(11.00)	(6.00)	(15.50)	(92.00)	
	(13.90)	(7.70)	(16.76)	(92.00)	
Total	2588	313	145	122	3168

The purpose of this study is (1) to review some models and decompositions, (2) to give a new decomposition of symmetry and (3) to analyze three sets of vision data using the new decomposition.

MATERIALS AND METHODS

Reviews of models and decompositions: Consider an R×R square contingency table with the same row and column classifications. Let p_{ij} denote the probability that an observation will fall in the ith row and jth column of the table (i =1,…,R; j =1,…,R). The symmetry (S) model is defined by:

$$p_{ij} = p_{ji} \quad (i = 1,…,R; j = 1,…,R)$$

Bowker (1948). For the vision data, this indicates that the probability that an individual's right eye grade is i and left eye grade is j is equal to the probability that his/her right eye grade is j and the left eye grade is i.

The quasi-symmetry (QS) model is defined by:

$$\frac{p_{ij}}{p_{ji}} = \frac{\delta_j}{\delta_i} \ (i \neq j)$$

Caussinus (1965). A special case of this model with $\{\delta_i = \delta\}$ is the S model. Although the detail is omitted, the QS model indicates the symmetry of odds ratios with respect to the main diagonal of a table.

The marginal homogeneity (MH) model is defined by:

$$p_{i\cdot} = p_{\cdot i} \ (i = 1, \ldots, R)$$

where $p_{i\cdot} = \sum_{t=1}^{R} p_{it}$ and $p_{\cdot i} = \sum_{s=1}^{R} p_{si}$ (Stuart, 1955). For the vision data, this indicates that the probability that an individual's right eye grade is i is equal to the probability that his/her left eye grade is i (i = 1,…,4).

Caussinus (1965) gave the decomposition of the S model as follows:

Theorem 1: The S model holds if and only if both the QS and MH models hold.

Each of S, QS and MH models indicates the structure of symmetry for the probabilities in the square table as the vision data. As a model which indicates the structure of asymmetry (instead of symmetry), the linear diagonals-parameter symmetry (LDPS) model is given as:

$$\frac{p_{ij}}{p_{ji}} = \theta^{j-i} \ (i < j)$$

Agresti (1983). For the vision data, this indicates that the probability that an individual's right eye grade is i and left eye grade is j (> i) is θ^{j-i} times higher than the probability that his/her right eye grade is j and left eye grade is i. If $\theta > 1$, then the right eye tends to be better than the left eye. A special case of the LDPS model with $\theta = 1$ is the S model. Also the LDPS model is a special case of the QS model obtained by putting $\{\delta_i = \theta^i\}$.

Let X and Y denote the row and column variables, respectively. For the vision data, X is the right eye grade and Y is the left eye grade. The mean equality (ME) of X and Y model is given by:

$$E(X) = E(Y)$$

where $E(X) = \sum_{i=1}^{R} i \, p_{i\cdot}$ and $E(Y) = \sum_{i=1}^{R} i \, p_{\cdot i}$. For the vision data, this indicates that the mean of right eye vision is equal to the mean of left eye vision.

Yamamoto *et al.* (2007) gave the following theorem:

Theorem 2: The S model holds if and only if both the LDPS and ME models hold.

Let for i<j,

$$G_{ij} = \sum_{s=1}^{i} \sum_{t=j}^{R} p_{st}, \text{ and } G_{ji} = \sum_{s=j}^{R} \sum_{t=1}^{i} p_{st}.$$

For the vision data, (1) G_{ij} for i<j indicates that the cumulative probability that an individual's right eye grade is i or below and his/her left eye grade is j (> i) or above and (2) G_{ji} for i<j indicates that the cumulative probability that an individual's left eye grade is i or below and his/her right eye grade is j (> i) or above.

Miyamoto *et al.* (2004) considered The Cumulative Quasi-Symmetry (CQS) model defined by:

$$\frac{G_{ij}}{G_{ji}} = \frac{\gamma_j}{\gamma_i} \ (i < j)$$

The CQS model is different from the QS model. Miyamoto *et al.* (2004) also considered the cumulative linear diagonals-parameter symmetry (CLDPS) model defined by:

$$\frac{G_{ij}}{G_{ji}} = \Delta^{j-i} \ (i < j)$$

The CLDPS model is different from the LDPS model. The CLDPS model is a special case of the CQS model obtained by putting $\{\gamma_i = \Delta^i\}$. For the vision data, the CLDPS model indicates that the probability that an individual's right eye grade is i or below and his/her left eye grade is j (> i) or above, is Δ^{j-i} times higher than the probability that an individual's left eye grade is i or below and his/her right eye grade is j (> i) or above. If $\Delta > 1$, then the right eye tends to be better than the left eye.

Yamamoto *et al.* (2011) gave the following theorem:

Theorem 3: The S model holds if and only if both the CQS and MH models hold.

A new decomposition of symmetry model: We can obtain a new decomposition of the S model as follows:

Theorem 4: The S model holds if and only if both the CLDPS and ME models hold.

Proof: If the S model holds, then both the CLDPS and ME models hold. Conversely, assuming that the CLDPS and ME models hold and then we shall show that the S model holds. We see:

$$E(X) = \sum_{i=1}^{R} i\, p_{i\cdot}$$
$$= \sum_{s=1}^{R} \sum_{t=s}^{R} p_{t\cdot}$$
$$= \sum_{s=1}^{R} \left(1 - F_{s-1}^{X}\right)$$
$$= R - \sum_{i=1}^{R-1} F_{i}^{X},$$

where $F_{i}^{X} = P(X \le i)$. Similarly we see:

$$E(Y) = R - \sum_{i=1}^{R-1} F_{i}^{Y}$$

where $F_{i}^{Y} = P(Y \le i)$. Thus we see:

$$E(Y) - E(X) = \sum_{i=1}^{R-1} F_{i}^{X} - \sum_{i=1}^{R-1} F_{i}^{Y}$$
$$= \sum_{i=1}^{R-1} G_{i,i+1} - \sum_{i=1}^{R-1} G_{i+1,i}.$$

From the ME model, we obtain:

$$\sum_{i=1}^{R-1} G_{i,i+1} = \sum_{i=1}^{R-1} G_{i+1,i}$$

From the CLDPS model, we obtain:

$$G_{i,i+1} = \Delta G_{i+1,i}.$$

Thus $\Delta = 1$. Namely the S model holds. The proof is completed.

Test of goodness-of-fit of model: Let n_{ij} denote the observed frequency in the (i, j)th cell of the R×R table (i = 1,…, R; j = 1,…,R), with $n = \sum \sum n_{ij}$ and let m_{ij} denote the corresponding expected frequency. Assume that $\{n_{ij}\}$ have a multinomial distribution. The maximum likelihood estimates of expected frequencies $\{m_{ij}\}$ under each model could be obtained, for example, using the Newton-Raphson method to the log-likelihood equations. Each model can be tested for goodness-of-fit by e.g., the likelihood ratio chi-squared statistic G^2 with the corresponding degrees of freedom, defined by:

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

where $\hat{m}_{ij}$ is the maximum likelihood estimate of m_{ij} under the model. The numbers of degrees of freedom for the S, CLDPS and ME models in Theorem 4 are R(R-1)/2, (R+1)(R-2)/2 and 1, respectively. Note that the number of degrees of freedom for the S model equals the sum of those for the CLDPS and ME models.

RESULTS

Consider the vision data in Table 1-3. We shall analyze these data using Theorem 4. Table 4 gives the values of likelihood ratio test statistic G^2 for each model.

Analysis of vision data of women in Table 1: We see from Table 4 that each of QS, LDPS and CLDPS models fits the vision data of women (Table 1) well, however, each of S, MH, CQS and ME models fits these data poorly.

Since the S model does not hold for these data, the probability that a woman's right eye grade is i and her left eye grade is j ($\neq$ i) is not equal to the probability that the woman's right eye grade is j and her left eye grade is i. Namely, a woman's right eye grade is not symmetric to her left eye grade.

The maximum likelihood estimates of parameters $\{\Delta^{j-i}\}$ under the CLDPS model are $\hat{\Delta} = 1.154$, $\hat{\Delta}^2 = 1.332$ and $\hat{\Delta}^3 = 1.537$. Since the CLDPS model holds for these data, the probability that a woman's right eye grade is i or below and her left eye grade is j (>i) or above is estimated to be $\hat{\Delta}^{j-i}$ times higher than the woman's left eye grade is i or below and her right eye grade is j or above. Since $\hat{\Delta} > 1$, a woman's right eye grade is estimated to be better than her left eye grade.

Since the ME model does not hold for these data, the mean of women's right eye grades is not equal to the mean of women's left eye grades. Using the sample proportions $\{\hat{p}_{ij} = n_{ij}/n\}$, we see that the mean of women's right eye grades is estimated to be 2.275 and the mean of women's left eye is estimated to be 2.305. Therefore a woman's right eye grade is expected to be better than her left eye grade in the sense of the mean.

We see from Theorem 4 that the poor fit of the S model is caused by the influence of the lack of structure of the ME model rather than the CLDPS model. Namely the fact that a woman's right eye grade is not symmetric to her left eye grade is caused by the fact that the mean of women's right eye grades is not equal to the mean of women's left eye grades.

Table 4: Values of likelihood ratio chi-squared statistic G^2 for models applied to the data in Table 1, 2 and 3. (*means significant at the 0.05 level)

Models	Degrees of freedom	Table 1	Table 2	Table 3
S	6	19.25*	16.95*	9.69
QS	3	7.27	5.71	2.81
MH	3	11.99*	11.18*	6.87
LDPS	5	7.28	6.95	8.22
CQS	3	8.43*	7.27	1.99
CLDPS	5	8.63	7.90	7.81
ME	1	11.98*	9.94*	1.46

Analysis of vision data of students in Table 2: We see from Table 4 that each of QS, LDPS, CQS and CLDPS models fits the vision data of students (Table 2) well, however, each of S, MH and ME models fits these data poorly.

Since the S model does not hold for these data, the probability that a student's right eye grade is i and his/her left eye grade is j ($\neq$ i) is not equal to the probability that the student's right eye grade is j and his/her left eye grade is i. Namely, a student's right eye grade is not symmetric to his/her left eye grade.

The maximum likelihood estimates of parameters $\{\Delta^{j-i}\}$ under the CLDPS model are $\hat{\Delta} = 0.831$, $\hat{\Delta}^2 = 0.691$ and $\hat{\Delta}^3 = 0.574$. Since the CLDPS model holds for these data, the probability that a student's right eye grade is i or below and his/her left eye grade is j ($>$i) or above is estimated to be $\hat{\Delta}^{j-i}$ times higher than the student's left eye grade is i or below and his/her right eye grade is j or above. Since $\hat{\Delta} < 1$, a student's right eye grade is estimated to be worse than his/her left eye grade.

Since the ME model does not hold for these data, the mean of students' right eye grades is not equal to the mean of students' left eye grades. Using the sample proportions, we see that the mean of students' right eye grades is estimated to be 2.631 and the mean of students' left eye grades is estimated to be 2.602. Therefore a student's right eye grade is expected to be worse than his/her left eye grade in the sense of the mean.

We see from Theorem 4 that the poor fit of the S model is caused by the influence of the lack of structure of the ME model rather than the CLDPS model. Namely the fact that a student's right eye grade is not symmetric to his/her left eye grade is caused by the fact that the mean of students' right eye grades is not equal to the mean of students' left eye grades.

Analysis of vision data of pupils in Table 3: We see from Table 4 that all models fit the vision data of pupils (Table 3) well. Since the S model holds for these data, the probability that a pupil's right eye grade is i and his/her left eye grade is j ($\neq$ i) is equal to the probability that the pupil's right eye grade is j and his/her left eye grade is i. Namely, a pupil's right eye grade is symmetric to his/her left eye grade.

Since the CLDPS model fits these data well, we shall test the hypothesis of $\Delta=1$(i.e., the hypothesis that the S model holds) under the assumption that the CLDPS model holds. It can be tested according to the difference between the likelihood ratio statistic G^2 for the S model and that for the CLDPS model. The difference is 1.88 with 1 degree of freedom. Therefore we can accept the hypothesis of $\Delta=1$in the CLDPS model, at the 0.05 significant level.

Since the ME model fits these data well, we shall also test the hypothesis that the S model holds under the assumption that the ME model holds. The difference between the G^2 value for the S model and that for the ME model is 8.23 with 5 degrees of freedom. Therefore we can accept the hypothesis that the S model holds under the ME model at the 0.05 significant levels. So, for pupils' vision data, we prefer the S model to each of the CLDPS and ME models.

Since the ME model fits these data well, the mean of pupils' right eye grades is equal to the mean of pupils' left eye grades. Under the ME model, the maximum likelihood estimates of the mean of pupils' right (left) eye grades is 1.301. This value is less than 2.5 being the midpoint of categories. Therefore a pupil's right (left) eye grade is expected to be close to Best (1) or Sec. (2), rather than Third (3) or Worst (4).

DISCUSSION

Theorems 1, 2, 3 and 4 may be useful for exploring the reason for the poor fit when the S model fits the data poorly. Especially, Theorem 4 would be useful when we are interested in the asymmetric structure of cumulative probabilities $\{G_{ij}\}$, i $\neq$ j and the symmetry structure of marginal means in the square table as the vision data.

CONCLUSION

We have given a new decomposition in Theorem 4 and have analyzed three sets of data on unaided vision data using Theorem 4. It has been estimated that (1) for the vision data of women, the right eye is better than the left eye and the mean of right eye is not equal to the mean of left eye, (2) for the vision data of students, the right eye is worse than the left eye and the mean of right eye is not equal to the mean of left eye and (3) for the vision data of pupils, the right eye is symmetric to the left eye and the mean of right eye is equal to the mean of left eye which is expected to be close to `Best' or 'Second' rather than `Third' or `Worst'.

REFERENCES

Agresti, A., 1983. A simple diagonals-parameter symmetry and quasi-symmetry model. Stat. Probability Lett., 1: 313-316. DOI: 10.1016/0167-7152(83)90051-2

Bishop, Y.M.M., S.E. Fienberg and P.W. Holland, 2007. Discrete Multivariate Analysis. 1st Edn., Springer, New York, ISBN: 9780387728056, pp: 558.

Bowker, A.H., 1948. A test for symmetry in contingency tables. J. Am. Stat. Assoc., 43: 572-574. PMID: 18123073

Caussinus, H., 1965. Contribution à l'analyse statistique des tableaux de corrélation. Annales Faculté Sci. Toulouse Sér., 29: 77-182. DOI: 10.5802/afst.519

Goodman, L.A., 1979. Multiplicative models for square contingency tables with ordered categories. Biometrika, 66: 413-418. DOI: 10.1093/biomet/66.3.413t

McCullagh, P., 1978. A class of parametric models for the analysis of square contingency tables with ordered categories. Biometrika, 65: 413-418.

Miyamoto, N., W. Ohtsuka and S. Tomizawa, 2004. Linear diagonals-parameter symmetry and quasi-symmetry models for cumulative probabilities in square contingency tables with ordered categories. Biometrical J., 46: 664-674.

Stuart, A., 1955. A test for homogeneity of the marginal distributions in a two-way classification. Biometrika, 42: 412-416.

Tomizawa, S., 1984. Three kinds of decompositions for the conditional symmetry model in a square contingency table. J. Japan Stat. Soc., 14: 35-42.

Tomizawa, S., 1985. Analysis of data in square contingency tables with ordered categories using the conditional symmetry model and its decomposed models. Env. Health Perspectives, 63: 235-239. DOI: 10.1289/ehp.8563235

Tomizawa, S., 1993. Diagonals-parameter symmetry model for cumulative probabilities in square contingency tables with ordered categories. Biometrics, 49: 883-887.

Tomizawa, S., N. Miyamoto and K. Yamamoto, 2006. Decomposition for polynomial cumulative symmetry model in square contingency tables with ordered categories. Metron, 64: 303-314.

Tomizawa, S. and K. Tahata, 2007. The analysis of symmetry and asymmetry: Orthogonality of decomposition of symmetry into quasi-symmetry and marginal symmetry for multi-way tables. J. Société Française Stat., 148: 3-36.

Yamamoto, H., T. Iwashita and S. Tomizawa, 2007. Decomposition of symmetry into ordinal quasi-symmetry and marginal equimoment for multi-way tables. Aus. J. Stat., 36: 291-306.

Yamamoto, K., S. Ando and S. Tomizawa, 2011. Decomposing asymmetry into extended quasi-symmetry and marginal homogeneity for cumulative probabilities in square contingency tables. J. Stat. Adv. Theory Appl., 5: 1-13.