REVSTAT – Statistical Journal
Volume 13, Number 3, November 2015, 233–243

# MARGINAL HOMOGENEITY MODEL FOR ORDERED CATEGORIES WITH OPEN ENDS IN SQUARE CONTINGENCY TABLES

Authors: Serpil Aktas
– Department of Statistics, Hacettepe University, Beytepe, Ankara, Turkey
spxl@hacettepe.edu.tr

Song Wu
– Department of Applied Mathematics and Statistics, Stony Brook University, NY, USA
song.wu@stonybrook.edu

Received: August 2013 Revised: December 2013 Accepted: May 2014

Abstract:

- A marginal homogeneity model tests whether the row and column distributions of a square contingency table have the same sample margins. However, for variables with ordered categories, the marginal homogeneity model does not take into account the ordering information, leading to significant loss of power. Score-based tests have been proposed for ordinal variables. In this paper, we extend the idea of scores and propose a new method based on the standardized scores to test the marginal homogeneity for ordered categories with open ends. Our simulation studies demonstrate that our proposed scores is more powerful than the usual scores.

Key-Words:

- *marginal homogeneity model; square contingency table; score.*

AMS Subject Classification:

- 49A05, 78B26.

## 1. MARGINAL HOMOGENEITY MODEL

Square contingency tables often arise in social, behavioral sciences and medical studies [1], and are used to display joint responses of two variables that have the same category levels. For example, to examine whether there exists any difference in unaided distance vision between right and left eyes in Royal Ordnance factories in Britain, 7477 employed women aged 30–39 were sampled between 1943–1946. Based on the data, a square contingency table can be constructed, with the row variable for the right eye vision grade and the column variable for the left eye vision grade. The right and left eyes have exactly the same vision grades, which are ordered based on the same criterion, e.g., from the best to the worst. Williams (1952) investigated the possibility of assessing association in a two-way table based on scores, which may be assigned to one or both of the row and column variables.

To study the symmetry of such square contingency tables, several models have been developed, including the complete symmetry model, the quasi-symmetry model, and the marginal homogeneity model [1, 3, 4, 5]. Consider a two-way $r\times r$ square contingency table with the same row and column classifications, let $p_{ij}$ denote the cell probability that an observation falls in the $i$th row and $j$th column of the table ($i = 1, 2, ..., r; j = 1, 2, ..., r$). Stuart (1955) proposed the hypothesis of marginal homogeneity that can be expressed in terms of the marginal cell probabilities

$$H_0 : \ p_{i+} = p_{+i}\,, \quad i = 1, ..., r \ ,$$

which is equivalent to the hypothesis that

$$H_0 : \ \text{The two samples have the same marginal distribution}\,.$$

The $p_{i+}$ and $p_{+i}$ are the marginal probabilities of $i$th row and $i$th column, respectively. This hypothesis is tested with the test statistic

$$Q = n\,\mathbf{d}'\hat{V}^{-1}\mathbf{d} \ , \tag{1.1}$$

where $\mathbf{d}' = (d_1, ..., d_r)$ with $d_i = p_{i+} - p_{+i}$ , $n$ is sample size, and $\hat{V}$ is the maximum likelihood estimate of the covariance matrix. The elements of $\hat{V}$ are

$$\hat{v}_{ij} = -(p_{ij} + p_{ji}) \quad \text{for} \ \ i \neq j \qquad \text{and} \qquad \hat{v}_{ii} = p_{i+} + p_{+i} - 2p_{ii} \ . \tag{1.2}$$

$Q$ has the chi-square distribution with $r - 1$ degrees of freedom. Later, Bhapkar (1979) proposed a similar type of test by taking the elements of covariance matrix as

$$\begin{aligned} \hat{v}_{ij} \ &= \ -(p_{ij} + p_{ji}) - (p_{i+} - p_{+i})\,(p_{j+} - p_{+j}) \qquad \text{for} \quad i \neq j \ , \\ \hat{v}_{ii} \ &= \ p_{i+} + p_{+i} - 2p_{ii} - (p_{i+} - p_{+i})^2 \ . \end{aligned} \tag{1.3}$$

A main disadvantage of the marginal homogeneity model is that it does not take into account the ordering information for ordinal variables, *i.e.* marginal homogeneity hypothesis is invariant to change of orders in the variable categories. However, one may be interested in whether one marginal distribution is stochastically larger or smaller than the other. If the distribution of cases across the row categories is the same as the distribution of cases across the column categories, these margins can be referred to as homogeneous and the distributions are the same; Otherwise, margins would be referred to as heterogeneous.

Agresti (1983) considered testing marginal homogeneity for ordinal categorical variables by using some fixed scores to weigh the marginal probability differences between corresponding row and column categories. The test statistics is:

$$d = \sum_{i=1}^{r} w_i (p_{i+} - p_{+i}) \tag{1.4}$$

for some fixed scores $\{w_i\}$. The estimated variance of $d$ is

$$\hat{S}_d^2 = \frac{1}{n} \left( \sum_{i,j} (w_i - w_j)^2 p_{ij} - d^2 \right). \tag{1.5}$$

Then for large $n$

$$z = \frac{d}{\hat{S}_d} \tag{1.6}$$

has approximately standard normal distribution under the null hypothesis [8]. Fleiss and Everitt (1971) also considered different forms of marginal homogeneity test. Relation of Marginal Homogeneity model with the other models were given in [1, 10, 11]. Caussinus' quasi-symmetry model can hold true in contingency tables in which the row and column marginals are not homogeneous [12].

## 2. STANDARDIZED SCORES FOR OPEN ENDED CATEGORIES

In the score-based methods, different score choices will lead to different test statistics and consequently provide different conclusions. Score can be assigned either based on distributional assumptions or based on prior knowledge [11]. Agresti (1983) defined some score sets, for example, for four categories, the scores $3, 1, -1, -3)$ to detect differences in location and $1, -1, -1, 1$ to detect differences in dispersion. In well-balanced data, the impact of the choice of scores on the final inference is minimal [8]. However, when the data are very unbalanced or in an open-ended form, results may significantly change with respect to choice of scores [8].

Several scores have been suggested in the literatures that they could handle two-way contingency tables with open-ended categories, in which one or both of categories may take the form of "greater than" or "less than". For example, Graubard and Korn (1987) discussed the equally spaced scores for $2 \times C$ contingency tables, and found that equally spaced scores might yield conservative results. In general, midrank is a very useful score choice when there are large differences in marginal counts. However, if the distribution is highly skewed within an interval, midpoints are also poor estimates of the true values. Particularly, midrank scores can be very unreasonable in applications when the marginals are far from uniform. Other ways to estimate scores include using latent root analysis that maximizes the correlation between the two sets [2], and optimizing conventional scores for a particular set of variables by comparing the squared correlation coefficients with the monotonic correlation ratios [14]. However, all these scores ignored the open-ended features of the categories.

To overcome the limitations of the equally spaced scores when variables have open ended categories, here we propose standardized $z$-scores based on semi-interquartile range for the row and column variables. The median is not only an appropriate measure for ordinal, interval, and ratio scale data, but also is well known to be the most convenient measure of location for the open ended categories. The standardized $z$-scores are defined as follows:

$$z_i = \frac{s_i - Q_2}{0.5\, IQR}\ , \tag{2.1}$$

$$z_j = \frac{s_j - Q_2}{0.5\, IQR}\ , \tag{2.2}$$

where,

$IQR$**:** interquartile range, $IQR = Q_3 - Q_1$

$Q_1$**:** first quartile

$Q_2$**:** second quartile (median)

$Q_3$**:** third quartile

$z_i$**:** $i$th row scores

$z_j$**:** $j$th row scores

$s_i$ & $s_j$**:** midpoints of row and column categories calculated as

$$s_i = \frac{LL_i - UL_i}{2}\ , \qquad s_j = \frac{LL_j - UL_j}{2}$$

$LL$ & $UL$**:** the lower and upper limits of a class, respectively.

Note that the semi-interquartile range is a good measure of spread for skewed distributions.

## 3. SIMULATION STUDY

Simulation studies have been carried out to investigate the statistical properties of our newly proposed standardized scores. We used two different schemes to simulate the data. The first one assumes that the row and column variables have the same marginal distribution, whereas the second one assumes that the row and column variables have different marginal distributions. The combinations of different sample size ($n = 50, 100, 500$), different table dimensions ($R = 5, 8$), and different levels of association between row and column variables ($\rho = 0.0, 0.5, 0.9$) were considered. For each simulation scenario, 1000 replications were performed. The simulated data were analyzed by both the standardized-score methods and the usual-score method defined in Agresti (1983), which allows the comparisons between the two methods. Matlab and SAS softwares were used for generating and analyzing data sets.

Table 1 shows the comparison of Type I errors between the usual and standardized scores, under the assumption that the row and column variables have the same marginal distribution. To construct the square contingency table, we applied the similar simulation strategy in Yang et al (2012). We first generated random numbers from a bivariate normal distribution with the same means of 25 and common variance of 36, and then the bivariate samples were cross tabulated into a two-way contingency table. Different correlation coefficient were assumed ($\rho = 0.0, 0.5, 0.9$) to evaluate how the association strength may affect the tests. Random samples were classified into $5\times5$ or $8\times8$ cross tables with the following categories:

$$X, Y(5\times5) \;:\; < 9.9, 10\text{–}14.9, 15\text{–}19.9, 20\text{–}24.9, > 25\,; \quad \text{or}\,,$$
$$X, Y(8\times8) \;:\; < 4.9, 5\text{–}9.9, 10\text{–}14.9, 15\text{–}19.9, 20\text{–}24.9, 25\text{–}29.9, 30\text{–}34.9, > 35\,.$$

Note that $X$ and $Y$ denote the row and column variables, and the first and last class intervals are open-ended classes. Hence, $R\times R$ contingency tables with open-ended ordinal variables were constructed. The row and column marginals are expected to be the same. In each simulation scenario, 1000 replication runs were performed to estimate the Type I error rates for both methods. Three different significance levels ($\alpha = 0.01, 0.05, 0.10$) were considered. As shown in Table 1, the actual Type I errors are very close to the nominal levels in each case, suggesting the validation of both score methods.

Next, we made the power comparison between the standardized and usual score methods. Similarly, we generated random numbers from a bivariate normal distribution with different means of 25 and 36, and common variance of 36, and the random samples were classified into square tables. In this way, the row and column variables have the different marginal distributions. Let $\alpha = 0.05$ denote the nominal level of significance of the tests, the empirical power of the tests can be calculated as the proportion of the test statistic is greater than the critical value, which is given by $P(X^2 > C)/t$, where $t$ is the number of replications in the simulation study, and $C$ is the critical value of the chi-square distribution for $\alpha = 0.05$ with associated degrees of freedom $R - 1$.

**Table 1**: Comparison of Type I errors between the usual and standardized scores.

| $\alpha$ | Usual Scores | | | | Standardized Scores | | | |
|---|---|---|---|---|---|---|---|---|
| | $\rho$ | $n$ | $R=5$ | $R=8$ | $\rho$ | $n$ | $R=5$ | $R=8$ |
| 0.01 | 0 | 50 | 0.0062 | 0.0105 | 0 | 50 | 0.0082 | 0.0088 |
| | | 100 | 0.0106 | 0.0118 | | 100 | 0.0098 | 0.0086 |
| | | 500 | 0.008 | 0.0096 | | 500 | 0.0101 | 0.0186 |
| | 0.5 | 50 | 0.0038 | 0.0048 | 0.5 | 50 | 0.008 | 0.0016 |
| | | 100 | 0.0062 | 0.0054 | | 100 | 0.0116 | 0.0101 |
| | | 500 | 0.0084 | 0.0089 | | 500 | 0.0094 | 0.0102 |
| | 0.9 | 50 | 0.0072 | 0.0104 | 0.9 | 50 | 0.0074 | 0.0112 |
| | | 100 | 0.0076 | 0.0094 | | 100 | 0.0988 | 0.0093 |
| | | 500 | 0.0082 | 0.009 | | 500 | 0.0099 | 0.0099 |
| 0.05 | 0 | 50 | 0.0352 | 0.0334 | 0 | 50 | 0.0398 | 0.0339 |
| | | 100 | 0.0398 | 0.0395 | | 100 | 0.0458 | 0.0444 |
| | | 500 | 0.0421 | 0.044 | | 500 | 0.0482 | 0.0468 |
| | 0.5 | 50 | 0.0383 | 0.0323 | 0.5 | 50 | 0.0418 | 0.0482 |
| | | 100 | 0.0406 | 0.0456 | | 100 | 0.0433 | 0.0506 |
| | | 500 | 0.0456 | 0.0456 | | 500 | 0.0439 | 0.0494 |
| | 0.9 | 50 | 0.0392 | 0.3297 | 0.9 | 50 | 0.0483 | 0.0456 |
| | | 100 | 0.0438 | 0.0431 | | 100 | 0.0489 | 0.0467 |
| | | 500 | 0.0439 | 0.0437 | | 500 | 0.0499 | 0.0494 |
| 0.10 | 0 | 50 | 0.0676 | 0.0668 | 0 | 50 | 0.0709 | 0.0701 |
| | | 100 | 0.069 | 0.0687 | | 100 | 0.0974 | 0.0974 |
| | | 500 | 0.0768 | 0.0754 | | 500 | 0.091 | 0.0905 |
| | 0.5 | 50 | 0.072 | 0.0651 | 0.5 | 50 | 0.0756 | 0.0756 |
| | | 100 | 0.068 | 0.0954 | | 100 | 0.0829 | 0.0829 |
| | | 500 | 0.0826 | 0.1028 | | 500 | 0.0965 | 0.0965 |
| | 0.9 | 50 | 0.0712 | 0.0808 | 0.9 | 50 | 0.0876 | 0.0877 |
| | | 100 | 0.0823 | 0.0843 | | 100 | 0.0943 | 0.0949 |
| | | 500 | 0.0966 | 0.0985 | | 500 | 0.0997 | 0.0996 |

As shown in Table 2, the proposed standardized scores has much higher power than the usual score procedures. For both methods, the power is substantially greater for larger sample size and correlation, that is, the power for detecting the marginal heterogeneity is the lowest for $\rho = 0$, highest for $\rho = 0.90$, and lowest for $n = 50$, highest for $n = 500$.

**Table 2**: Empirical power comparison between the usual and standardized scores for $\alpha = 0.05$.

| Usual Scores | | | | Standardized Scores | | | |
|---|---|---|---|---|---|---|---|
| $\rho$ | $n$ | $R=5$ | $R=8$ | $\rho$ | $n$ | $R=5$ | $R=8$ |
| 0 | 50 | 0.5754 | 0.5308 | 0 | 50 | 0.7782 | 0.7689 |
| | 100 | 0.6106 | 0.6326 | | 100 | 0.7234 | 0.7316 |
| | 500 | 0.63 | 0.6596 | | 500 | 0.7301 | 0.7886 |
| 0.5 | 50 | 0.6388 | 0.6768 | 0.5 | 50 | 0.7465 | 0.79 |
| | 100 | 0.7762 | 0.7754 | | 100 | 0. 8065 | 0.8112 |
| | 500 | 0.8884 | 0.8109 | | 500 | 0.9008 | 0.9102 |
| 0.9 | 50 | 0.7172 | 0.7103 | 0.9 | 50 | 0.8874 | 0.8812 |
| | 100 | 0.8876 | 0.8594 | | 100 | 0.897 | 0.8711 |
| | 500 | 0.9092 | 0.9008 | | 500 | 0.9976 | 0.9576 |

## 4. NUMERICAL EXAMPLES

A hypothetical $5 \times 5$ square contingency table with both row and column having open-ended categories is generated to illustrate the utilization and efficiency of the standardized scores (Table 3).

**Table 3**: A simulated $5 \times 5$ table.

| $R/C$ | $\leq 9.9$ | 10–14.9 | 15–19.9 | 20–24.9 | $\geq 25$ | Total |
|---|---|---|---|---|---|---|
| $\leq 9.9$ | 24 | 23 | 34 | 12 | 45 | 138 |
| 10–14.9 | 37 | 7 | 5 | 25 | 32 | 106 |
| 15–19.9 | 48 | 11 | 17 | 37 | 22 | 135 |
| 20–24.9 | 28 | 9 | 7 | 17 | 13 | 74 |
| $\geq 25$ | 6 | 13 | 15 | 5 | 8 | 47 |
| Total | 143 | 63 | 78 | 96 | 120 | 500 |

The sample quartiles are calculated as:
For row variable: $Q_1 = 9.47$, $Q_2 = 15.17$, $Q_3 = 19.80$.
For column variable: $Q_1 = 9.32$, $Q_2 = 17.77$, $Q_3 = 24.68$.

Interquartile ranges for the row and column variables are $IQR = 19.80 - 9.47 = 10.33$; $IQR = 24.68 - 9.32 = 15.36$, respectively.
Using Equations (2.1) and (2.2), standardized scores are displayed in Table 4. The scores in Equation (1.4) would be $z_i \times z_j$ due to the nature of open ends.

**Table 4**: Standardized scores for row and column variables.

| Standardized row scores ($z_i$) | Standardized column scores ($z_j$) |
|---|---|
| −1.49468 | −1.34375 |
| −0.52662 | −0.69271 |
| 0.44143 | −0.04167 |
| 1.40948 | 0.609375 |
| 2.377541 | 1.260417 |

Using the standardized scores we get,
$d = 0.085, \quad \hat{S}_d^2 = 0.0053, \quad \hat{S}_d = 0.0728, \quad z = \frac{0.085}{0.0728} = 1.1675.$
Therefore, the null hypothesis of marginal homogeneity is not rejected.

When we utilize the usual scores as: $w_i = w_j = \{-3, -1, 0, 1, 3\}$, we get,
$d = 0.538, \quad \hat{S}_d^2 = 0.01875, \quad \hat{S}_d = 0.13693, \quad z = \frac{0.538}{0.13693} = 3.929.$
The result indicates that null hypothesis of marginal homogeneity ($p_{i+} = p_{+i}$) is rejected at 0.05 significance level.

## 5. CONCLUSIONS

Ordinal variables are common in many research areas. Marginal homogeneity model tests that the marginal frequencies do not differ significantly between the row and column variables. Marginal homogeneity model requires assigning scores through row and column variables. The problem for open ended categories is to assign the proper scoring. The simplest scoring method is admittedly integer scoring. The standardized scores employing the marginal homogeneity test in the presence of an open-ended category is proposed in this paper. Ordinal models require assigning scores to levels of ordinal variables. When responses are ordered categories, it is usually important to test the hypotheses of marginal homogeneity using ordinal information. When the variation of the between variable levels in contingency tables are large, standardized scores will be appropriate. Different choices of the row and the column scores can lead to different conclusion concerning association of the rows and columns. When we employ different scores in the modeling, inferences derived from the analyses would be dependent on the scoring system.

The proposed scores give better results than the usual scores with respect to their statistical power values in detecting marginal heterogeneity. Results also show that the use of ordinal approach becomes relatively more efficient as correlation coefficient and the sample size increase. We showed that the usual score and standardized score methods can achieve similar type I errors when data were simulated under null hypothesis, while the standardized score method has larger power than usual score method when data were simulated under alternative hypothesis. When ordinal variables in a two contingency table are a discretized form of continuous variables, it is reasonable to use the standardized scores based on sample quartiles. Our simulation suggests that the proposed method competes well with alternative.

## ACKNOWLEDGMENTS

We acknowledge the valuable suggestions from the referees.

## REFERENCES

[1] Bishop, Y.M.; Fienberg, S. and Holland, P.W. (1975). *Discrete Multivariate Analysis*, MIT Press, Cambridge, MA.

[2] Williams, E.J. (1952). Use of scores for the analysis of association in contingency tables, *Biometrika*, **39**, 274–289.

[3] Ireland, C.T.; Ku, H.H. and Kullback, S. (1969). Symmetry and marginal homogeneity of r×r contingency tables, *J. Amer. Statist. Assoc.*, **64**, 1323–1341.

[4] Tomizawa, S.; Miyamoto, N. and Ouchi, M. (2006). Decompositions of symmetry model into marginal homogeneity and distance subsymmetry in square contingency tables with ordered categories, *REVSTAT – Statistical Journal*, **4**, 2.

[5] Yang Z.; Sun X. and Hardin, J.W. (2012). Testing marginal homogeneity in clustered matched-pair data, *J. Stat. Plan. Infer.*, **141**, 1313–1318.

[6] Stuart, A. (1955). A test for homogeneity of the marginal distribution in a two-way classification, *Biometrika*, **42**(11), 412–416.

[7] Bhapkar, V.P. (1979). On tests of symmetry when higher order interactions are absent, *J. Indian Statist. Assoc.*, **17**, 17–26.

[8] Agresti, A. (1983). Testing marginal homogeneity for ordinal categorical variables, *Biometrics*, **39**(2), 505–510.

[9] Fleiss, J.R. and Everitt, B.S. (1971). Comparing the marginal totals of square contingency tables, *British Journal of Mathematical and Statistical Psychology*, **24**, 117–123.

[10] Agresti, A. (2013). *Categorical Data Analysis*, Wiley, New York.

[11] Lawal, H.B. (2012). *Categorical Data Analysis with SAS and SPSS Applications*, LEA, Mahwah, New Jersey.

[12] Goodman, L.A. (2002). Contributions to the statistical analysis of contingency tables: Notes on quasi-symmetry, quasi-independence, log-linear models, log-bilinear models and correspondence analysis model, *Ann. Fac. Sci. Toulouse Math*, **6**(11), 525–540.

[13] Graubard, B.I. and Korn, E.L. (1987). Choice of column scores for testing independence in ordered 2×k contingency tables, *Biometrics*, **43**, 471–476.

[14] Allen, M.P. (1976). Conventional and optimal interval scores for ordinal variables, *Sociological Methods & Research*, **4**, 475–494.