Journal of Statistical Research *ISSN 0256 - 422 X*
2009, Vol. 43, No. 1, pp. 53-68
Bangladesh

MARGINAL HOMOGENEITY AND DISTANCE SUBSYMMETRY MODELS IN SQUARE CONTINGENCY TABLES WITH ORDERED CATEGORIES

Bayo H. Lawal
School of Arts and Sciences
American University of Nigeria, Yola, Adamawa State, Nigeria
Email: bayo.lawal@aun.edu.ng

Richard A. Sundheim
Department of Business Computer Information System
St. Cloud State University, St. Cloud, MN 56301, USA
Email: sundheim@stcloudstate.edu

summary

In this study, we have employed the GSK and the non-standard log-linear model approaches to fit the class of distance sub-symmetry models to square contingency tables having ordered categories. SAS PROCs CATMOD and GENMOD were employed to implement our models. A macro generates the factor variables to implement models in the latter approach. Except for the DCS-k where no maximum likelihood closed form exists, all other models are easily implemented with the non standard log-linear model approach. The GSK on the other hand readily fits all the models considered in this study. These models are applied to the 5×5 British generational data as well as the 4×4 unaided distance vision data. Both data have received considerable attention and analyses in the literature. Results obtained where applicable agree with those published in previous literature on the subject. The approaches suggest here eliminate any programming that might be required in order to apply these class of models to square contingency tables.

Keywords and phrases: Distance subsymmetry, factor variables, GSK, log-linear, marginal homogeneity, SAS, symmetry

AMS Classification: 62-07

1 Introduction

Tomizawa, Miyamoto and Ouchi (2006) decomposed the symmetry model in a square contingency table having ordinal categories into marginal symmetry and distance subsymmetry models. Decomposed symmetry models are needed to explain and discover why the symmetry models fits poorly the given data. There has been a lot of literature on interpretations of decomposed symmetry models and our goal in this paper is to fit the models considered in Tomizawa *et al.* (2006) paper using the non-standard log-linear model approach discussed in various works in Lawal (2002, 2004) and Lawal and Sundheim (2004) as well as the GSK approach discussed in Grizzle, Starmer and Koch (1969) and Shuster and von Eye (1998). It is hoped that this will make the models considered here more easily and readily applicable by researchers and social scientists working in this area.

For an $I \times I$ square contingency table with ordered categories, if we let π_{ij} to be the joint probability in the i-th row and j-th column respectively, for, $1 \leq (i,j) \leq I$, then, the model of *complete symmetry* (S) has the model formulation:

$$\pi_{ij} = \pi_{ji}, \quad \text{for } 1 \leq i \leq j \leq I \tag{1.1}$$

Following closely the notation employed in Tomizawa *et al.*, let Y_1 and Y_2 relate to the row and column variables respectively. Then, the marginal homogeneity model (MH) is defined as:

$$\Pr(Y_1 = i) = \Pr(Y_2 = i) \quad \text{for } i = 1, 2, \ldots, I$$

In this case, we have

$$\pi_{i+} = \pi_{+i}, \quad \text{for } i = 1, 2, \ldots, I$$

where $\pi_{i+} = \sum_{j=1}^{I} \pi_{ij}$ and $\pi_{+j} = \sum_{i=1}^{I} \pi_{ij}$.

Further, for $i < j$, define

$$G_{ij} = \Pr(Y_1 \leq i, Y_2 \geq j) = \sum_{s=1}^{i} \sum_{t=j}^{I} \pi_{st} \tag{1.2}$$

Similarly, for $i > j$, again define,

$$G^*_{ij} = \Pr(Y_1 \geq i, Y_2 \leq j) = \sum_{s=i}^{I} \sum_{t=1}^{j} \pi_{st} \tag{1.3}$$

With the above formulations in (1.2) and (1.3), the symmetry model (S) is equivalent to:

$$G_{ij} = G^*_{ji} \quad \text{for } i < j \tag{1.4}$$

Similarly, the marginal homogeneity model (MH) can be reformulated as:

$$G_{i,i+1} = G^*_{i+1,i} \quad \text{for } i = 1, 2, \ldots, I-1 \tag{1.5}$$

We know that $MH \subset S$, that is, model S implies the MH model, thus, $S \longrightarrow MH$. Consequently, Tomizawa *et al.* sought decompositions for (1.4) and (1.5) having the structure:

$$G_{ij} = G^*_{ji} \quad \text{for } j - i = 2, 3, \ldots, I-1; \ i < j.$$

The following decompositions are therefore considered by Tomizawa *et al.* (2006).

2 The Distance Cumulative Subsymmetry Model

Following Tomizawa *et al.* (2006), this model has

$$G_{ij} = G^*_{ji} \qquad \text{for } j - i = 2, 3, \ldots, I-1; \quad i < j \tag{2.1}$$

which is equivalent to:

$$\pi_{ij} = \pi_{ji} \qquad \text{for } j - i = 2, 3, \ldots, I-1; \quad i < j$$

Model (2.1) is referred by Tomizawa *et al.* (2006) as the *subsymmetry* (SS) model, which indicates that the probability of an observation falling in cell (i, j), which is one of cells such that the probability from the main diagonal is greater or equal to 2, is equal to the probability that the observation falls in cell (j, i).

If we now consider for fixed k ($k = 2, 3, \ldots, I-1$), the model defined by

$$G_{i,i+k} = G^*_{i+k,i} \qquad \text{for } i = 1, 2, \ldots, I-k \tag{2.2}$$

then, the model in (2.2) is referred to as the *distance cumulative subsymmetry* model with the difference k between the diagonal containing the cut point [i and $i+k$] and the main diagonal. The model is denoted by Tomizawa *et al.* as the DCS-k model.

3 The Distance Subsymmetry Model

Tomizawa *et al.* (2006) described the distance subsymmetry model with distance k (denoted DS-k) as satisfying for fixed k:

$$\pi_{ij} = \pi_{ji} \qquad \text{for } j - i = k; \quad i < j \tag{3.1}$$

Which Tomizawa *et al.* (2006) describes as "this model indicates that the probability that an observation will fall in cell (i, j) with the distance k from the main diagonal, is equal to the probability that the observation falls in cell (j, i) with the same distance.

Just as Tomizawa *et al.* observed, the following models {S, MH, SS, DCS-k, DS-k } which are equivalent are fitted in this paper. Models S, MH, SS, DCS-k and DS-k have, $I(I-1)/2$, $(I-1)$, $(I-1)(I-2)/2$, $(I-k)$ and $(I-k)$ degrees of freedom, respectively.

4 Model Implementation in SAS

We have employed in this paper the two data set employed in Tomizawa *et al.* (2006), namely, the 5×5 occupational mobility data of British fathers and sons (Goodman, 1981) as well as the 4×4 unaided distant vision data of 7477 women aged 30-39 employed in the Royal Ordnance factories in Britain from 1943 to 1946. (Stuart, 1955).

The models discussed in this paper will be implemented via SAS with the GSK and non-standard log-linear modeling approaches. As observed in Tomizawa *et al.* (2006), while

Table 1: Occupational Status for British father-son pairs

Father's Status	Son's Status (1)	(2)	(3)	(4)	(5)	Total
(1)	50	45	8	18	8	129
(2)	28	174	84	154	55	495
(3)	11	78	110	223	96	518
(4)	14	150	185	714	447	1510
(5)	3	42	72	320	411	848
Total	106	489	459	1429	1017	3500

Table 2: Un-aided distant vision as reported in Kendall (1955)

Right Eye Grade	**Left Eye Grade** Best Best (1)	Second Second (2)	Third Third (3)	Worst Worst (4)	Total
Best (1)	1520	266	124	66	1976
Second (2)	234	1512	432	78	2256
Third (3)	117	362	1772	205	2456
Worst (4)	36	82	179	492	789
Total	1907	2222	2507	841	7477

MLE estimates exist in closed forms for the S, SS and the DS-k models, such do not exist for the DCS-k models for $k < (I-1)$. However, both DC-k and DCS-k if $k = (I-1)$ models are equivalent and thus closed form MLE solutions exist for these models.For instance, for a 5×5 table, DC-4 $\equiv$ DCS-4. We present the two approaches employed in this paper in the following sections:

4.1 The GSK Approach

The GSK (which refers to the first names of the authors) was first developed by Grizzle, Starmer and Koch (1969) and uses a weighted least-squares approach which allows vector-valued functions of the type $G(\pi) = 0$ to be tested. A fuller description of this methodology as it applies to our particular application in this paper can be found in Shuster and von Eye (1998). The test statistic for evaluating the hypothesis $G(\pi) = 0$ has been shown, Shuster and von Eye (1998) to be asymptotically distributed χ^2 with the degrees of freedom equal to the number **q** of real-valued functions of $G(\pi)$.

To implement the GSK methodology to our data set, we apply this to the 5×5 table and display how the Symmetry, marginal symmetry, SS, DCS-k and DC-k models are implemented with this approach using PROC CATMOD in SAS. We discuss below the application of the GSK method to fitting each of the models outlined above. However, to illustrate the GSK approach, we present for example, the fitting of the SS model first.

4.1.1 The SS Model:

This is the model in (2.1) defined as:

$$G_{ij} = G^*_{ji} \qquad \text{for } j-i = 2, 3, \dots, I-1; \quad i < j$$

The above for 5×5 tables implies that for $j - i = 2$, we have:

$$G_{13} = G^*_{31}, \quad G_{24} = G^*_{42}, \quad G_{35} = G^*_{53} \tag{4.1}$$

The constraints in (4.1) lead to the following equations:

$$\pi_{13} + \pi_{14} + \pi_{15} = \pi_{31} + \pi_{41} + \pi_{51} \tag{4.2a}$$

$$\pi_{14} + \pi_{15} + \pi_{24} + \pi_{25} = \pi_{41} + \pi_{51} + \pi_{42} + \pi_{52} \tag{4.2b}$$

$$\pi_{15} + \pi_{25} + \pi_{35} = \pi_{51} + \pi_{52} + \pi_{53} \tag{4.2c}$$

Similarly, when $j - i = 3$, we have,

$$G_{14} = G^*_{41}, \quad G_{25} = G^*_{52} \tag{4.3}$$

which again lead to the constraint equations:

$$\pi_{14} + \pi_{15} = \pi_{41} + \pi_{51} \tag{4.4a}$$

$$\pi_{15} + \pi_{25} = \pi_{51} + \pi_{52} \tag{4.4b}$$

Finally, when $j - i = 4$, we have only the constraint:

$$G_{15} = G^*_{51}$$

which leads to the constraint equation:

$$\pi_{15} = \pi_{51} \tag{4.5}$$

Thus, the constraint equation in (4.2a) for instance can be written in form of a contrast as:

$$\pi_{13} + \pi_{14} + \pi_{15} - \pi_{31} - \pi_{41} - \pi_{51} = 0$$

The six constraint equations which constitute our hypothesis relating to $G(\pi) = 0$ can be presented in matrix form as in (4.6).

$$G(\boldsymbol{\pi}) = \begin{bmatrix} G_1(\pi) \\ G_2(\pi) \\ G_3(\pi) \\ G_4(\pi) \\ G_5(\pi) \\ G_6(\pi) \end{bmatrix} = \begin{bmatrix} \pi_{13} + \pi_{14} + \pi_{15} - \pi_{31} - \pi_{41} - \pi_{51} \\ \pi_{14} + \pi_{15} + \pi_{24} + \pi_{25} - \pi_{41} - \pi_{51} - \pi_{42} - \pi_{52} \\ \pi_{15} + \pi_{25} + \pi_{35} - \pi_{51} - \pi_{52} - \pi_{53} \\ \pi_{14} + \pi_{15} - \pi_{41} - \pi_{51} \\ \pi_{15} + \pi_{25} - \pi_{51} - \pi_{52} \\ \pi_{15} - \pi_{51} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{4.6}$$

Again in terms of the 5×5 table, the above six generated contrasts lead to the following implementation in SAS PROC CATMOD:

```
data dps;
  do r=1 to 5;
  do c=1 to 5;
  INPUT COUNT @@;
  output;
  end; end;
  datalines;
  50 45 8 18 8 28 174 84 154 55 11 78 110
  223 96 14 150 185 714 447 3 42 72 320 411
  ;
run;
proc catmod data=dps;
  weight count;
  response
  0 0 1 1 1 0 0 0 0 0 -1 0 0 0 0 -1 0 0 0 0 -1 0 0 0 0,
  0 0 0 1 1 0 0 0 1 1 0 0 0 0 0 -1 -1 0 0 0 -1 -1 0 0 0,
  0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 -1 -1 -1 0 0,
  0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 -1 0 0 0 0 -1 0 0 0 0,
  0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 -1 -1 0 0 0,
  0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -1 0 0 0 0;
  model r*c =/noint;
run;
```

```
                     Analysis of Variance

          Source       DF   Chi-Square    Pr > ChiSq
          ------------------------------------------
          Residual      6         8.49        0.2043
```

The model is based on 6 d.f. because there are six real-valued functions in $G(\boldsymbol{\pi})$.

4.1.2 The Symmetry Model

For implementing the full symmetry model, we again note that for symmetry, we $\pi_{ij} = \pi_{ji}$, which leads for $i \neq j$ to $I(I-1)/2 = 10$ constraints and subsequently ten contrasts for the 5×5 table. These contrasts are employed in the SAS PROC CATMOD procedure and the results are very consistent with those obtained with the likelihood ratio test statistic

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

where n_{ij} are the observed frequencies and $\hat{m}_{ij}$ are the expected frequencies under some model. The model gives a Wald's test statistic 37.62 on 10 d.f.

4.1.3 The Marginal Homogeneity Model

For marginal homogeneity, we have:

$$\pi_{i+} = \pi_{+i} \qquad \text{for } i = 1, 2, \ldots, I-1 \tag{4.7}$$

That is,

$$\pi_{1i} + \pi_{2i} + \pi_{3i} + \pi_{4i} + \pi_{5i} = \pi_{i1} + \pi_{i2} + \pi_{i3} + \pi_{i4} + \pi_{i5} \quad i = 1, 2, 3, 4 \tag{4.8}$$

Thus for a 5×5 table,(4.8) becomes for $i = 1, 2, 3, 4$ respectively:

$$\pi_{21} + \pi_{31} + \pi_{41} + \pi_{51} = \pi_{12} + \pi_{13} + \pi_{14} + \pi_{15} \tag{4.9a}$$
$$\pi_{12} + \pi_{32} + \pi_{42} + \pi_{52} = \pi_{21} + \pi_{23} + \pi_{24} + \pi_{25} \tag{4.9b}$$
$$\pi_{13} + \pi_{23} + \pi_{43} + \pi_{53} = \pi_{31} + \pi_{32} + \pi_{34} + \pi_{35} \tag{4.9c}$$
$$\pi_{14} + \pi_{24} + \pi_{34} + \pi_{54} = \pi_{41} + \pi_{42} + \pi_{43} + \pi_{45} \tag{4.9d}$$

We can re-write (4.9a) for instance as:

$$\pi_{21} + \pi_{31} + \pi_{41} + \pi_{51} - \pi_{12} - \pi_{13} - \pi_{14} - \pi_{15} = 0 \tag{4.10}$$

The above is a contrast and we can write this in terms of the $5 \times 5 = 25$ cells as:

```
0 -1 -1 -1 -1 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0
```

Similar results can be written for the other three subequations and leads to the implementation in PROC CATMOD of the form:

```
proc catmod data=dps;
weight count;
response
0 -1 -1 -1 -1 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0,
0 1 0 0 0 -1 0 -1 -1 -1 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0,
```

```
0 0 1 0 0 0 0 1 0 0 -1 -1 0 -1 -1 0 0 1 0 0 0 0 1 0 0,
0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 -1 -1 -1 0 -1 0 0 0 1 0;
model r*c =/noint ml;
run;
                                Analysis of Variance

                      Source       DF   Chi-Square    Pr > ChiSq
                      ------------------------------------------
                      Residual      4        32.95        <.0001
```

4.1.4 The Quasi-Symmetry Model

The quasi-symmetry model implies (Agresti,1990),

$$\frac{\pi_{ij}}{\pi_{iI}\pi_{Ij}} = \frac{\pi_{ji}}{\pi_{Ii}\pi_{jI}}, \quad \text{for } (i,j) = 1,2,\ldots,I \tag{4.11}$$

Since the model is based on $(I-1)(I-2)/2$ d.f., then the corresponding constraints in terms of (4.11) are:

$$\frac{\pi_{12}}{\pi_{15}\pi_{52}} = \frac{\pi_{21}}{\pi_{51}\pi_{25}} \tag{4.12a}$$
$$\frac{\pi_{13}}{\pi_{15}\pi_{53}} = \frac{\pi_{31}}{\pi_{51}\pi_{35}} \tag{4.12b}$$
$$\frac{\pi_{14}}{\pi_{15}\pi_{54}} = \frac{\pi_{41}}{\pi_{51}\pi_{45}} \tag{4.12c}$$
$$\frac{\pi_{23}}{\pi_{25}\pi_{53}} = \frac{\pi_{32}}{\pi_{52}\pi_{35}} \tag{4.12d}$$
$$\frac{\pi_{24}}{\pi_{25}\pi_{54}} = \frac{\pi_{42}}{\pi_{52}\pi_{45}} \tag{4.12e}$$
$$\frac{\pi_{34}}{\pi_{35}\pi_{54}} = \frac{\pi_{43}}{\pi_{53}\pi_{45}} \tag{4.12f}$$

Taking logarithms of the of both sides and equation to zero, we have respectively:

$$\log\pi_{12} - \log\pi_{15} - \log\pi_{52} - \log\pi_{21} + \log\pi_{51} + \log\pi_{25} = 0 \tag{4.13a}$$
$$\log\pi_{13} - \log\pi_{15} - \log\pi_{53} - \log\pi_{31} + \log\pi_{51} + \log\pi_{35} = 0 \tag{4.13b}$$
$$\log\pi_{14} - \log\pi_{15} - \log\pi_{54} - \log\pi_{41} + \log\pi_{51} + \log\pi_{45} = 0 \tag{4.13c}$$
$$\log\pi_{23} - \log\pi_{25} - \log\pi_{53} - \log\pi_{32} + \log\pi_{52} + \log\pi_{35} = 0 \tag{4.13d}$$
$$\log\pi_{24} - \log\pi_{25} - \log\pi_{54} - \log\pi_{42} + \log\pi_{52} + \log\pi_{45} = 0 \tag{4.13e}$$
$$\log\pi_{34} - \log\pi_{35} - \log\pi_{54} - \log\pi_{43} + \log\pi_{53} + \log\pi_{45} = 0 \tag{4.13f}$$

```
proc catmod data=dps;
weight count;
response
0 1 0 0 -1 -1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 -1 0 0 0,
```

```
0 0 1 0 -1 0 0 0 0 0 -1 0 0 0 1 0 0 0 0 0 1 0 -1 0 0,
0 0 0 1 -1 0 0 0 0 0 0 0 0 0 -1 0 0 0 1 1 0 0 -1 0,
0 0 0 0 0 0 0 1 0 -1 0 -1 0 0 1 0 0 0 0 0 1 -1 0 0,
0 0 0 0 0 0 0 0 1 -1 0 0 0 0 0 0 -1 0 0 1 0 1 0 -1 0,
0 0 0 0 0 0 0 0 0 0 0 0 0 1 -1 0 0 -1 0 1 0 0 1 -1 0
log ;
model r*c =/noint ml;
run;
```

```
                    Analysis of Variance

          Source       DF   Chi-Square    Pr > ChiSq
          -----------------------------------------
          Residual      6         4.63        0.5926
```

4.1.5 The DCS-k Models

The DCS-k models are as described in (2.2) where $k = 2, \dots, (I - k)$, we have,

$$G_{i,i+k} = G^*_{i+k,i} \qquad \text{for } i = 1, 2, \dots, I - k$$

The above leads for a fixed $k = 2$ and $i = 1, 2, 3$ to the three relationships designated as (i) to (iii) respectively in the following:

$$(i) \quad G_{13} = G^*_{31} \quad (ii) \quad G_{24} = G^*_{42}, \quad \text{and} \quad (iii) \quad G_{35} = G^*_{53}$$

Again these lead to the following constraint equations on the cell probabilities,viz,

$$\pi_{13} + \pi_{14} + \pi_{15} = \pi_{31} + \pi_{41} + \pi_{51} \tag{4.14a}$$

$$\pi_{14} + \pi_{15} + \pi_{24} + \pi_{25} = \pi_{41} + \pi_{51} + \pi_{42} + \pi_{52} \tag{4.14b}$$

$$\pi_{15} + \pi_{25} + \pi_{35} = \pi_{51} + \pi_{52} + \pi_{53} \tag{4.14c}$$

Thus DCS-2 model when implemented with PROC CATMOD has a Wald test statistic value of 6.83 and is based on 3 d.f. with a corresponding p-value of 0.0774.

For the case when $k = 3$, again we have,

$$(i) \quad G_{14} = G^*_{41}, \quad \text{and} \quad (iii) \quad G_{25} = G^*_{52} \tag{4.15}$$

which again lead to the constraint equations:

$$\pi_{14} + \pi_{15} = \pi_{41} + \pi_{51} \tag{4.16a}$$

$$\pi_{15} + \pi_{25} = \pi_{51} + \pi_{52} \tag{4.16b}$$

Implementation of the DCS-3 with PROC CATMOD gives a Wald test statistic of 4.23 on 2 d.f. with a p-value of 0.1205.

The DCS-4 model has $k = 4$ and is given by the model,

$$G_{15} = G^*_{51} \tag{4.17}$$

which leads to the single constraint equation,

$$\pi_{15} = \pi_{51}$$

and similar implementation in SAS gives a Wald test statistic value of 2.27 on 1 d.f with a p-value of 0.1315.

4.2 The Non-standard log-linear Model Approach:

The non-standard log-linear approach has been used mostly in recent years by Lawal (2001); Lawal (2004); Lawal and Sundheim (2002); and von Eye and Spiel (1996) amongst others. A non-standard log-linear formulation of the above models can be written in the form (Lawal, 2001):

$$\boldsymbol{\ell} = \mathbf{X}\,\boldsymbol{\lambda} \tag{4.18}$$

where $\ell_{ij} = \ln(m_{ij})$, the log of expected values under some model,and $\mathbf{X}$ is the design matrix consisting of 0s and 1s, which are derived from the indicator or regression variable representing the levels of the factor or regression variable (Lawal and Sundheim, 2002).

The symmetry (S) model in this case is implemented with the non-standard log-linear (NLL) model defined by:

$$l_{ij} = \mu + \lambda_{ij}^{\mathbf{S}} \tag{4.19}$$

where $\mathbf{S}$ relates to the factor variable (a vector) required to fit the symmetry model and proposed in (Lawal, 2001) which has entries that are generated with the following expressions for all (i, j) (Lawal and Sundheim, 2002).

$$\mathbf{S}_{ij} = \begin{cases} (k+1) - (i+1)(\frac{1}{2}i+1) + (I+3)(i+1) - 3 - 2I & \text{if } i \leq j \\ \\ (k+1) - (j+1)(\frac{1}{2}j+1) + (I+3)(j+1) - 3 - 2I & \text{if } i > j \end{cases} \tag{4.20}$$

Thus, for a 5×5 table for instance, the factor variable $\mathbf{S}$ becomes:

$$\mathbf{S} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 6 & 7 & 8 & 9 \\ 3 & 7 & 10 & 11 & 12 \\ 4 & 8 & 11 & 13 & 14 \\ 5 & 9 & 12 & 14 & 15 \end{bmatrix}$$

A similar set of entries can be obtained for the 4×4 Table. The (MH) model is implemented by using the fact that the corresponding conditional marginal homogeneity (CMH) model is related to both the symmetry and quasi-symmetry by the following relation:

$$S = QS \cap CMH \tag{4.21}$$

Hence, a conditional test of marginal homogeneity (CMH) assuming that QS holds is provided by examining the quantity $G^2_{(S)} - G^2_{(QS)}$ which is distributed χ^2 with $I-1$ degrees of freedom. The quasi-symmetry model has the NLL model formulation:

$$l_{ij} = \mu + \lambda_i^R + \lambda_j^C + \lambda_{ij}^{\mathbf{S}}$$

and it is based on $(I-1)(I-2)/2$ degrees of freedom.

4.2.1 Implementing the SS Model

The (SS) model is implemented with the NLL model formulation:

$$l_{ij} = \mu + \lambda_{ij}^{\mathbf{S}} + \gamma_{ij}^{\boldsymbol{\psi}} \tag{4.22}$$

where $\boldsymbol{\psi}$ is a factor variable defined for a general $I \times I$ table as:

$$\boldsymbol{\psi}_{ij} = \begin{cases} 1 & \text{if } i = j \\ a_{ij} = a_{ji} & \text{if } i < j \text{ and } |i-j| \neq 1 \\ b_{ij} & \text{elsewhere} \end{cases} \tag{4.23}$$

where a_{ij} take values $\{2, 3, \ldots, (I^2 - 3I + 4)/2\}$. The remaining $2I - 2$ cells are filled with integers ranging from $(I^2 - 3I + 6)/2$ to $I(I+1)/2$.

Thus, for a 5×5 table, we have from (4.23), the $\{a_{ij} = 2, 3, 4, 5, 6, 7\}$, and hence the $a_{ji} = a_{ij}$ for all $i < j$ provided $|i-j| \neq 1$. The remaining $2I - 2 = 8$, b_{ij} cells designated with dashes $-$ in $\boldsymbol{\psi}'$ are filled with integers ranging from $(I^2 - 3I + 6)/2 = 8$ to $I(I+1)/2 = 15$. Completing the dashes therefore, we have, the factor variable $\boldsymbol{\psi}$ defined as:

$$\boldsymbol{\psi}' = \begin{bmatrix} 1 & - & 2 & 5 & 7 \\ - & 1 & - & 3 & 6 \\ 2 & - & 1 & - & 4 \\ 5 & 3 & - & 1 & - \\ 7 & 6 & 4 & - & 1 \end{bmatrix}, \qquad \boldsymbol{\psi} = \begin{bmatrix} 1 & 8 & 2 & 5 & 7 \\ 12 & 1 & 9 & 3 & 6 \\ 2 & 13 & 1 & 10 & 4 \\ 5 & 3 & 14 & 1 & 11 \\ 7 & 6 & 4 & 15 & 1 \end{bmatrix}$$

4.2.2 The *DS-k* Models

Again for the 5×5 table, to implement the general (DS-k) Model, the corresponding NLL model formulation is:

$$l_{ij} = \mu + \lambda_{ij}^{\mathbf{S}} + \delta_{ij}^{\mathbf{DS}_k} \tag{4.24}$$

Where the $\mathbf{DS}_k$ is a factor variable generated from the following expression for a general $I \times I$ square table with corresponding $k = 1, 2, \ldots, (I-1)$. That is,

$$\text{DS-k} = \begin{cases} 1 & \text{if } i = j \\ i+1 & \text{if } (j-i) = k \\ j+1 & \text{if } (i-j) = k \\ a_h & \text{elsewhere} \end{cases} \tag{4.25}$$

and the remaining $(I^2 - 3I + 2k)$ entries $\{a_h\}$ are randomly filled with consecutive integers ranging from $(I - k + 2)$ to $(I-1)^2 + k$. We display below, the generated factor variables for a 5×5 table for $k = 1, 2, 3, 4$.

$$\text{DS-1} = \begin{bmatrix} 1 & 2 & 6 & 7 & 8 \\ 2 & 1 & 3 & 9 & 10 \\ 11 & 3 & 1 & 4 & 12 \\ 13 & 14 & 4 & 1 & 5 \\ 15 & 16 & 17 & 5 & 1 \end{bmatrix}, \quad \text{DS-2} = \begin{bmatrix} 1 & 5 & 2 & 6 & 7 \\ 8 & 1 & 9 & 3 & 10 \\ 2 & 11 & 1 & 12 & 4 \\ 13 & 3 & 14 & 1 & 15 \\ 16 & 17 & 4 & 18 & 1 \end{bmatrix}$$

$$\text{DS-3} = \begin{bmatrix} 1 & 4 & 5 & 2 & 6 \\ 7 & 1 & 8 & 9 & 3 \\ 10 & 11 & 1 & 12 & 13 \\ 2 & 14 & 15 & 1 & 16 \\ 17 & 3 & 18 & 19 & 1 \end{bmatrix}, \quad \text{DS-4} = \begin{bmatrix} 1 & 3 & 4 & 5 & 2 \\ 6 & 1 & 7 & 8 & 9 \\ 10 & 11 & 1 & 12 & 13 \\ 14 & 15 & 16 & 1 & 17 \\ 2 & 18 & 19 & 20 & 1 \end{bmatrix}$$

For example, when $k = 2$, then from (4.25) we have DS-2 entries generated above are obtained as:

$$\text{DS-2} = \begin{bmatrix} 1 & - & 2 & - & - \\ - & 1 & - & 3 & - \\ 2 & - & 1 & - & 4 \\ - & 3 & - & 1 & - \\ - & - & 4 & - & 1 \end{bmatrix}$$

where the remaining $(I^2 - 3I + 4) = 14$ cell entries designated with dashes $-$ are randomly filled with consecutive integers ranging from $a = 5$ to $a = (I^2 - 2I + 3) = 18$.

Similarly, for the 4×4 table, we have,

$$\text{DS-1} = \begin{bmatrix} 1 & 2 & 5 & 6 \\ 2 & 1 & 3 & 7 \\ 8 & 3 & 1 & 4 \\ 9 & 10 & 4 & 1 \end{bmatrix}, \quad \text{DS-2} = \begin{bmatrix} 1 & 4 & 2 & 5 \\ 6 & 1 & 7 & 3 \\ 2 & 8 & 1 & 9 \\ 10 & 3 & 11 & 1 \end{bmatrix}, \quad \text{DS-3} = \begin{bmatrix} 1 & 3 & 4 & 2 \\ 5 & 1 & 6 & 7 \\ 8 & 9 & 1 & 10 \\ 2 & 11 & 12 & 1 \end{bmatrix}$$

The DS-k factor variables are generated with a SAS macro and can generate such for any sized square ordinal contingency table.

5 Results

We present in Tables 3 and 4, the results of fitting each of the models described above using both the non-standard log-linear model (NLL) approach that are based on the factor variables described in the preceding section and the GSK approach to the 5×5 and 4×4 Tables 1 and 2 respectively. For each of the models fitted, the likelihood ratio test statistic G^2 defined earlier was computed. While the fits for the GSK are based on Wald's test statistic Q (Lawal, 2003), however, conclusions or inferences from the use of both statistics in this case are the same.

Table 3: Results for the 5×5 data in Table 1

Applied models	Degrees of freedom	NLL LRT (G^2)	GSK Wald's (W)
S	10	37.4632	37.62
QS	6	4.6641	4.63
MH	4	32.7991*	32.95
SS	6	8.5757	8.49
DCS-2	3	na	6.83
DCS-3	2	na	4.23
DCS-4	1	2.3583	2.27
DS-1	4	28.8880	28.99
DS-2	3	3.9686	3.96
DS-3	2	2.2488	2.24
DS-4	1	2.3583	2.27

Where the results indicate 'na', it means that no closed form exists for that model and the NLL is not applicable. However, the GSK method was applicable to fitting the DCS-k models. The values obtained are very close to those obtained in Tomizawa *et al.* (2006),

Table 4: Results for the 4×4 data in Table 2

Applied models	Degrees of freedom	NLL LRT (G^2)	GSK Wald's (W)
S	6	19.2492	19.16
QS	3	7.2708*	7.22
MH	3	11.9784	11.98
SS	3	9.2587	9.14
DCS-2	2	na	4.97
DCS-3	1	8.9554	8.83
DS-1	3	9.9905	9.99
DS-2	2	0.3034	0.30
DS-3	1	8.9554	8.83

but this observation should not surprise us as the GSK is not maximum likelihood based, but rather on the weighted least-squares (WLS) approach and thus the results are bound to be slightly different. Where closed form MLE exists, the results obtained from the NLL approach agree with those published in Tomizawa *et al.*. The marginal symmetry model results presented in this paper are the conditional marginal symmetry fits (which assume that the QS model fits), and are asymptotically equivalent to the unconditional test for the MH model (Tomizawa and Tahata (2007). We present in Figure 1, the relationships between the models discussed in this paper.

6 Conclusions

Results obtained in this paper using the non-standard log-linear model approach are consistent with those obtained in Tomizawa *et al.* (2006). The approach does not work for the DCS-k models because they do not have closed form solutions. The exception to this being the DCS-(I-1) which is equivalent to its DS-(I-1) counterpart. We have shown therefore that the class of models in Tomizawa *et al.* can be implemented in SAS PROC GENMOD or PROC CATMOD. We can extend the use to other software such as SPSS PROC GENLOG or GLIM (Katari,1993). The advantage here is that there is no need for extensive programming to obtain the maximum likelihood estimates and hence the expected frequencies under our various models. The symmetry, quasi-symmetry and marginal homogeneity models have received considerable attention and all can easily be implemented in SAS or SPSS (Lawal, 2003, 2004). The macro %DSK makes it possible to generate the factor variables needed to implement the DS-k models.

SAS programs and the macro for implementing all the models in this article are available from the authors.

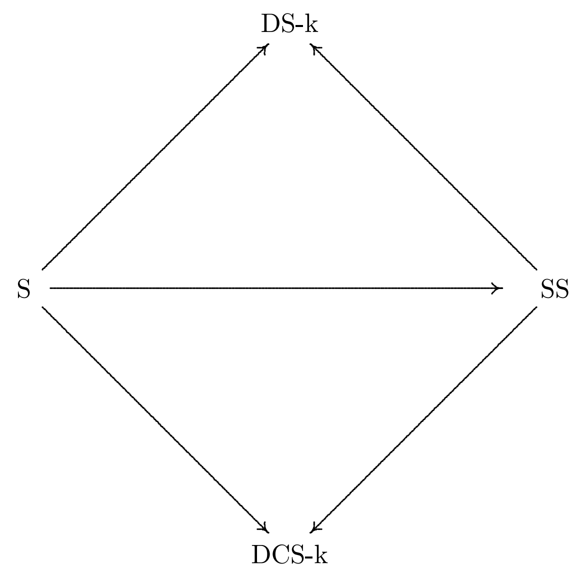

Figure 1: Relationships between the models

A $\longrightarrow$ B indicates that model A implies model B

References

[16] Agresti, A. (1983). A simple diagonals-parameter symmetry and quasi-symmetry model. *Statist. Prob. Letters*, 1, 313-316.

[16] Caussinus, H. (1965). Contributions a l'analyse statistique des tableaux de correlation. *Ann. Fac. Sci. Univ. Toulouse*, 29, 77-182.

[16] von Eye, A. and Spiel, C. (1996). Standard and Non-standard Log-Linear Symmetry Models for Measuring change in categorical Variables. *American Statistician*, 50, 300-305.

[16] Goodman, L.A. (1981). Association models and the bivariate normal for contingency tables with ordered categories. *Biometrics*, **68**, 347-355.

[16] Grizzle, J.E., Starmer, C.F., and Koch, G.G. (1969). Analysis of categorical data by linear models. *Biometrika*, **25**, 489-504.

[16] Katari, M. (1993). Fitting asymmetry and non-independence models with GLIM. *GLIM Newsletter*, 22, 46-50.

[16] Lawal, H. B. (1992). Parsimonious uniform-distance association models for the occupational mobility data. *J. Japan Statist. Soc.*, 22, 126-134.

[16] Lawal, H.B. (2001). Modeling symmetry models in square contingency tables with ordered categories. *Jour. Comput. Statist. and Simul.*, 71, 59-83.

[16] Lawal, H.B. and Sundheim R. (2002). Generating factor variables for Asymmetry, Non-independence and Skew-symmetry models in Square Contingency Tables with a SAS Macro. *Journal of Statistical Software, Vol 7.*,8, 1-23.

[16] Lawal, H.B. (2003). *Categorical Data Analysis with SAS and SPSS Applications.* Lawrence Erlbaum Associates Inc., Mahwah, NJ. May, 2003.

[16] Lawal, H. B. (2004). Using a GLM to decompose the symmetry model in square contingency tables having ordered categories. *Journal of Applied Statistics*, **31**, 279-303.

[16] Lawal, H.B. and Upton, G.J.G. (1995). A computer algorithm for fitting models to $N \times N$ contingency tables having ordered categories. *Communications in Statistics*, Simula, 24(3), 793-805.

[16] Shuster, C. and von Eye A. (1998). Modeling Turn-Over Tables Using the GSK Approach. *Methods of Psychological Research Online*, **3**, 39-53.

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

[16] Tomizawa, S., Miyamoto, N., and Ouchi, M. (2006) Decompositions of symmetry models into marginal homogeneity and distance subsymmetry in square contingency tables with ordered categories. *REVSTAT-Statistical Journal*, **4**, 153-161.

[16] Tomizawa, S. and Tahata, K. (2007), The analysis of symmetry and asymmetry: Orthogonality of decomposition of symmetry into quasi-symmetry and marginal symmetry for multi-way tables. *Journal de la Société Française de Statistique*, **148**, 3-36