

Generating Factor variables for Asymmetry, Non-independence and Skew-symmetry models in Square Contingency Tables using SAS

H. Bayo Lawal *
Department of Statistics
St. Cloud State University
St. Cloud, MN 56301
lawal@stcloudstate.edu

Richard A. Sundheim
Department of BCIS
St. Cloud State University
St. Cloud, MN 56301
sundheim@stcloudstate.edu

Abstract

In this paper, a SAS program (macro) is written to generate factor and regression variables required for implementing asymmetry, non-independence, non-symmetry + independence models as well as skew-symmetry models in discussed in square $a \times a$ contingency tables having nominal or ordinal categories. While several authors have developed similar factor variables for use with GLIM, we have extended this to the non-independence and the non-symmetry+independence models. The former includes both the fixed and variable distance models as well as the quasi-ordinal symmetry model. Further, our implementation of the asymmetry model in terms of the required factor variable is different from those defined for implementation of same in GLIM. Most of the models described in this paper however assume ordinal categories for the contingency table. The SAS macro developed can be applied to any square table of dimension a .

We apply the models discussed in this paper to the 5×5 Danish mobility data that have been widely analyzed in various literatures.

Keywords: Log-linear, Ordinal, Symmetry, Square Tables

*H. Bayo Lawal is Professor of Statistics and Richard Sundheim is Professor of Statistics in the Departments of Statistics and Business Computer Information Systems respectively, St. Cloud State University, St. Cloud, MN 56301.

1 Introduction

Lawal (2002) has utilized the non-standard log-linear model approach to model asymmetry, non-independence, skew-symmetry and non-symmetry + independence models in square two-way contingency tables. To implement the non-standard log-linear model approach, we need to generate the relevant factor or regression variables required for such model. Kateri (1993) and Kutylowski (1989) have discussed the generation of factor variables required for the implementation of some of the models being considered in this paper in GLIM. Our implementation of the symmetry model in this paper for instance is consistent with the procedure proposed in Friedl (1995) except that our factor variable for the null symmetry model is defined differently. Both ours and Friedl have $\frac{I(I+1)}{2}$ levels. The implementation of the null asymmetry model therefore involves only this single factor variable whereas, the approach by Kateri and Kutylowski involves two such factor variables designated as SC₋ and SS₋ in both their papers. Further, their programs are written for GLIM.

The class of models considered and implemented in this paper are those described in Goodman (1985) namely, the asymmetry, the non-independence and the non-symmetry+independence models. The Skew-symmetry models are discussed in Yamagushi (1990) and we have presented in Appendix B, SAS Macro %Factors(A) for generating the relevant factor and regression variables necessary for implementing the models discussed in this paper. We demonstrate the use of this macro for the 5×5 Danish social mobility data which has been analyzed by various authors as an example. The macro can be utilized to fit the models discussed here to any square table of any dimension. All the cell entries for the factor or regression variables generated to implement the models described in the following sections are presented in C for the 5×5 Danish table example. In Appendix B, also, we present the SAS Macro %Analysis(A,count,combined) which utilizes the SAS macro %Factors(A) to implement all the models discussed in the present paper for a square table with dimension A, the argument in the macro.

2 Asymmetry Models

Asymmetry models (Goodman, 1985) are those models having the symmetry model as their baseline. That is, these models measure deviations from the complete symmetry model S. Models belonging to this group are described in (1a) to (1f) below.

(1a) The complete symmetry model represents the baseline or null asymmetry model (O) for this group.

The model is described in Goodman (1985) as the null-asymmetry model. To implement the null-symmetry model (S) for instance, Lawal (2001) suggested generating the factor variable required to fit this model from the recurrence relation for an $a \times a$ table as:

$$S_h^k = S_{h-1}^k + (a + 2 - h), \quad \text{for } h = 2, \dots, (a - k) \quad (1)$$

where $k = |i - j| = 0, 1, \dots, (a - 1)$, k is the k -th diagonal and $S_1^k = k + 1$. For a 5×5 table for instance, $k = |i - j| = 0, 1, 3, 4$. The main diagonal elements have $k = 0$ and $h = 2, 3, 4, 5$.

In the programming in SAS, the above recurrence relation and hence the entries for the factor variable S are generated with the following expressions for all (i, j) :

$$S_{ij} = \begin{cases} (k + 1) - (i + 1)(\frac{1}{2}i + 1) + (a + 3)(i + 1) - 3 - 2a & \text{if } i \leq j \\ (k + 1) - (j + 1)(\frac{1}{2}j + 1) + (a + 3)(j + 1) - 3 - 2a & \text{if } i > j \end{cases} \quad (2)$$

where k and a are as defined above. We note here that when $i = j$, then $k = 0$ in (2). The S factor variable has levels that equal $a(a + 1)/2 = 15$ for $a = 5$.

Hence, the resulting vector (this is indicated as a factor variable in SAS) necessary for implementing the complete symmetry or null asymmetry model which is generated from the above expression for a 5×5 table is:

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

We note here that the factor variable defined for S has entries that do not exactly match those generated in Friedl (1995), but the common feature of both vectors here is that both have 15 levels as expected. The S' in Friedl (1995) is generated from the expression:

$$S' = 2^{(i-1)} + 2^{(j-1)} \quad \text{for } (i, j) = 1, 2, \dots, a$$

The null asymmetry model (S) is based on $(a-1)(a-2)/2$ degrees of freedom.

- (1b) The *triangle asymmetry* (T) model is described in Goodman (1985). This model is the composite model S+T where T is a regression variable defined as:

$$T_{ij} = \begin{cases} 1 & \text{if } i < j \\ 2 & \text{if } i > j \\ 3 & \text{if } i = j \end{cases} \quad (3)$$

This model is equivalent to the conditional symmetry model described in McCullagh (1978), and is based on $(a+1)(a-2)/2$ degrees of freedom.

- (1c) The diagonals-parameter symmetry model (DPS) is described in Agresti (1983). The model is the composite model S+D where D is a factor variable generated such that we have:

$$D_{ij} = \begin{cases} |i-j| & \text{if } i < j \\ |i-j| + a - 1 & \text{if } i > j \\ 2a - 1 & \text{if } i = j \end{cases} \quad (4)$$

The model is based on $(a-1)(a-2)/2$ degrees of freedom and is described in Goodman (1985) as the *diagonals asymmetry* model (D).

- (1d) The linear diagonals-parameter symmetry model (LDPS) is described in Agresti (1983) and Tomizawa (1992). It is the composite model S+F, where F is a regression variable pertaining to the fixed distance model (Haberman, 1978; Lawal, 1996) and is generated as:

$$F_{ij} = \begin{cases} |i-j| + 1 & \text{if } i < j \\ 1 & \text{elsewhere} \end{cases} \quad (5)$$

The model is equivalent to the *ordinal quasi symmetry* (OQS) model described in Agresti (1996). The model is based on $(a + 1)(a - 2)/2$ degrees of freedom and will be described here as the *asymmetry fixed distance* (F) model.

- (1e) The odds-symmetry model types I & II (OS1 & OS2) are fully described in Tomizawa (1985) and are composite models S+OS1 and S+OS2 respectively where OS1 and OS2 are factor variables generated as:

$$(OS1)_{ij} = \begin{cases} i & \text{if } i < j \\ (a - 1) + j & \text{if } i > j \\ 2a - 1 & \text{for } i = j \end{cases} \quad (6)$$

and

$$(OS2)_{ij} = \begin{cases} (2a - j) & \text{if } i < j \\ (a + 1) - i & \text{if } i > j \\ 2a - 1 & \text{for } i = j \end{cases} \quad (7)$$

respectively. Both models are based on $(a - 1)(a - 2)$ degrees of freedom. Both models will be designated as the asymmetry odds I and II (OS1) & (OS2) models.

- (1f) Lawal (2002) has described the *2-ratio parameter symmetry* (2RPS) model introduced in Tomizawa (1987) as the *2-ratio parameter asymmetry* (2RPA) model which has the multiplicative formulation:

$$\pi_{ij} = \gamma \pi_{ji} \delta^{j-i} \quad \text{for } (i \neq j) \quad (8)$$

This model is the composite model $(S) + (T) + (F)$ and is based on $(a^2 - a - 4)/2$ degrees of freedom and reduces to the asymmetry triangles model when $\delta = 1$ in equation (8) above.

3 Skew-Symmetry Models

These class of models have the quasi-symmetry (QS) model as its baseline model. The models are proposed in Yamagushi (1990). Again, these models measure deviations from the baseline QS model. Also belonging to this group are the following models:

(2a) The quasi-symmetry model (QS) which is the null (O) model for this class of asymmetry models.

The model is described by Goodman (1985) as the *asymmetry* (RC) model. Yamagushi described it as the *null skew-asymmetry* model. The model is the composite R+C+S, where R and C refer respectively to rows and column factor variables, where

$$R_{ij} = i \quad \text{for } (i, j) = 1, 2, \dots, a \quad (9)$$

and,

$$C_{ij} = j \quad \text{for } (i, j) = 1, 2, \dots, a \quad (10)$$

The model is based on $(a-1)(a-2)/2$ degrees of freedom.

(2b) The quasi conditional symmetry (QCS) model (Tomizawa, 1992) is equivalent to the *uniform skew-symmetry level* model in Yamagushi (1990) and is the composite model QS+T, where T is the regression variable defined in (3). The model is equivalent to the *triangles-parameter skew-symmetry* (SP_SK) model in Yamagushi (1990).

(2c) The quasi-diagonals parameter symmetry model (QDPS) is the model described in Yamagushi as the *diagonals-parameter skew-symmetry* model, designated as the (DP_SK) model. The model is the composite model QS+D, with D being a factor variable defined in (4). The model has $(a(a-3)/2)$ degrees of freedom.

(2d) The quasi-odds symmetry (QOS) model (Tomizawa, 1985) is described in Yamagushi as the *middle-value-effect skew-symmetry* model and is designated as the (M_SK) model. The model is described in Bishop *et. al.* (1975) as the *adjusted quasi-symmetry* (AOS) model. It is the composite model QS+OS1 or QS+OS2. The model is based on $(a-2)(a-3)/2$ degrees of freedom.

4 Non-Independence Models

This class of models have the independence model as its baseline or null. Consequently, all the models described in this section models the interaction structure or deviations from the independence model in

the table. Belonging to this category are the following:

(3a) The null or independence model (O) has:

$$\pi_{ij} = \alpha_i \beta_j, \quad \text{for all } (i, j) \quad (11)$$

where α_i and β_j relate to the row and column marginals respectively. Thus, the null (O) model is the composite model R+C where R and C are the rows and column factor variables defined in (9) and (10) respectively.

(3a) The fixed and variable distance models (F and V). Both the fixed distance and variable distance models are described in Haberman (1978). The model has been implemented in Lawal & Upton (1990, 1995) and Lawal (1992). The models are often designated as (F) and (V) respectively (Lawal and Upton, 1990). Both models are composite models: O+F and O +V¹ - V^(a-1), where F is a regression variable defined in (5) and V^h = {V¹, ..., V^(a-1)} are factor or regression variables defined for $h = 1, 2, \dots, (a - 1)$ as:

$$V_{ij}^h = \begin{cases} 2 & \text{for } (i, j) = 1, \dots, h \quad \text{and} \quad (i, j) = h + 1, \dots, a \\ 1 & \text{otherwise} \end{cases} \quad (12)$$

Both models (F) and (V) belong to the principal diagonal class models (Upton, 1985) satisfying

$$\Phi_{ij} = 0 \quad \text{unless } i = j.$$

where Φ_{ij} is the log-odds ratio under this model.

(3b) The quasi-independence model (Q), the non-independence triangles model (T) and the loyalty model (L) belong to the diagonal band models discussed in Upton (1985). For this class of models, the log-odds ratios satisfy:

$$\Phi_{ij} = 0 \quad \text{unless } |i - j| = 1$$

The models are composite models O+Q, O+T and O+L respectively where Q and L are factor and regression variables defined as:

$$Q_{ij} = \begin{cases} i & \text{if } i = j \\ (a+1) & \text{if } i \neq j \end{cases} \quad (13)$$

and,

$$L_{ij} = \begin{cases} 1 & \text{if } i = j \\ 2 & \text{if } i \neq j \end{cases} \quad (14)$$

respectively. T is considered as a factor variable in this case. Model (T) is the triangles (Goodman, 1972) model while (L) is the uniform loyalty model discussed in Upton & Sarlvik (1981) in the context of political election studies. The non-independence models (L), (T) and (Q) have respectively, the degrees of freedom $a(a-2)$, (a^2-2a-1) and (a^2-3a+1) .

(3C) The non-independence diagonals-parameter model (D) and the non-independence absolute diagonals-parameter model (DA) are described in Goodman (1972, 1985) as the *asymmetric minor diagonal* and *symmetric minor diagonal* models respectively. The models are the composite models O+D and O+DA respectively, where D is a factor variable defined in (4) and DA is also a factor variable defined as:

$$DA_{ij} = \begin{cases} a & \text{if } i = j \\ |i-j| & \text{if } i \neq j \end{cases} \quad (15)$$

Models (D) and (DA) have respectively $(a-2)^2$ and $(a-1)(a-2)$ degrees of freedom. The diagonals-absolute triangle (DAT) nonindependence model is the composite model O+DA+T and is based on (a^2-3a+1) degrees of freedom.

5 Non-symmetry + Independence Models

This class of models has as its baseline, the null non-symmetry+independence model defined in Goodman (1985) as:

$$\pi_{ij} = \alpha_i \alpha_j, \quad \text{for all } (i, j) \quad (16)$$

The model is the model in (11) with $\beta_j = \alpha_j$. The model defined in (16) is the familiar Halfway (H) model described in Hope (1982). This model is generalized in Goodman (1985) to the triangle non-symmetry+independence model (T) which is a composite model H+T where H comprises of $(a - 1)$ regression variables such that H_i is defined (Hope, 1982) as:

$$f_i = \begin{cases} 1 & \text{if cell in row } i \\ 0 & \text{elsewhere} \end{cases}, \quad \text{and} \quad s_i = \begin{cases} 1 & \text{if cell in column } i \\ 0 & \text{elsewhere} \end{cases}$$

Thus,

$$H_i = f_i + s_i, \quad i = 1, 2, \dots, (I - 1).$$

Similarly the non-symmetry+independence diagonals (D), diagonals-absolute (DA) and diagonals-absolute triangles (DAT) are composite models H+D, H+DA and H+DA+T respectively. The non-symmetry+independence models (O), (T), (D), (DA) and (DAT) have degrees of freedom $a(a-1)$, $(a-1)(a-2)$, $(a+1)(a-2)$, $(a-1)^2$, and $a(a-2)$ respectively.

6 Applications

The data below is the well analyzed 5×5 Danish social mobility data which gives the cross-classification of father's and son's occupational status categories in Denmark (Bishop *et al.*, 1975).

TABLE 1 about here

The results of applying all the models discussed in the previous sections to the Danish Social mobility data in Table 1 are presented in Table 2, where G^2 and X^2 refer to the likelihood ratio and Pearson's (Goodness-of-fit: GOF) test statistics defined respectively as:

$$G^2 = 2 \sum_{ij} n_{ij} \ln \left(\frac{n_{ij}}{m_{ij}} \right), \quad X^2 = \sum_{ij} \frac{(n_{ij} - m_{ij})^2}{m_{ij}}$$

TABLE 2 here

7 Conclusions

Results obtained from the implementation of this macro to the 5×5 Danish social mobility data agree with results previously published in various literature. Composite models can easily be implemented with the macro. Further, models with main diagonal deleted as in Goodman (1985) can similarly be easily implemented from the SAS program in Appendix B. A detailed program description is provided in appendix A.

Acknowledgments

The author would like to express their thanks to Dr. Sandra Keith of the Department of Mathematics for generating the expression in (2) in MAPLE.

APPENDICES

A Program Description

This SAS program has 3 steps.

Step 1. Compile macro program FACTORS locally or
store the compiled macro in a permanent SAS catalog

This macro generates the factor and regression variables for
Asymmetry, Non-independence and Skew-symmetry models
in Square Contingency Tables.

The macro has three parameters
FACTORS(A, Lib=work, FactorDSN=generate)

A = dimension of the square contingency table

LIB = the SAS library name where you wish to
store the factor variables SAS data set

It is a Keyword so the default value is WORK
which means the factor variables data set will
be stored in the work library unless you provide
a different value in the macro statement after
the equals sign. Note: If you accept the default
value, then you only need to specify the
dimension parameter A.

FactorDSN = SAS data set name you wish to give the
factor variables data set. The default value is
GENERATE.

Example1: %Factors(5)
or %Factors(5, Lib=work, FactorDSN=generate)
Specifies a 5x5 contingency table will be analyzed.
Default values Work and Generate will be used for
the Library and Data set name for the Factors.

Example2: %Factors(5, Lib=C, FactorDSN=Table)
Specifies dimension 5 again.
The factor variables data set will be stored in a
library named C and be given the name TABLE.
This assumes that a Library C has been formed in your
SAS session. We usually do this in Step 3 below.

Step 2. Compile macro program ANALYSIS locally or
store the compiled macro in a permanent SAS catalog

This macro merges the Count data (from the contingency
table) with the Factor variables generated by the macro
in Step 1.

The macro has 5 parameters:

`Analysis(A, DepVar, CombinedDSN, Lib=work, FactorDSN=generate)`

A = dimension of the square contingency table

DepVar = name of your dependent variable in your count data from your contingency table. This data is read into a SAS data set in Step 3 below.

CombinedDSN = name of the SAS data set after merging the count data with the factor variables.

Lib = the SAS library name where you wish to store the combined SAS data set

It is a Keyword so the default value is WORK which means the factor variables data set will be stored in the work library unless you provide a different value in the macro statement after the equals sign. Note: If you accept the default value then you only need to specify the dimension parameter A.

FactorDSN = SAS data set name you wish to give the factor variables data set. The default value is GENERATE.

Example 1. `%Analysis(5, Count, Combined)`

or `%Analysis(5, Count, Combined, Lib=work, FactorDSN=generate)`

Specifies dimension 5 for the contingency table.

Indicates that the dependent variable in our contingency table data is named COUNT

The SAS data set that merges the factors with the count data will be named COMBINED

The Library where COMBINED will be stored is the default Work library.

The factors variable SAS data set was named GENERATE by default in Step 1.

Example 2. `%Analysis(5, Count, Combined, Lib=C, FactorDSN=Table)`

The first three parameters are the same as in Example 1.

The last two parameters are the same as Example 2 of Step 1.

Step 3. Read in count data --- name the SAS data set with the macro variable DataDSN. You can use any valid SAS data set name. We use the name Data1 in the program below.

Store the data in a SAS library of your choice. The library name is to be specified by the macro variable DataLib. We use the Work library in the program below.

We make both of these macro variables global so their values can be referred to in the macro program Analysis in Step 2 above.

Two Libname statements are optionally given.

The first Libname statement sets up the library where the count data set (DataDSN) will be stored. If you use the temporary Work library then you can leave this statement commented out. If you want to store the data permanently you will need to provide the correct path to the library in the libname statement.

The second Libname statement sets up the library where the Factors data set and the Combined data set will be stored.

Again you can leave this commented out if you use the temporary Work library. If you want to store the data permanently you will need to provide the correct path to the library and the library name that you intend to use when executing the FACTORS and ANALYSIS macros. (we have AAAAA in the statement below just as a place holder where the name you want will go)

Our example data set has dimension 5 (5x5 contingency table).

The data is input into one column in the following order:

The first row elements, second row elements, etc giving us a data set with one column and 25 rows.

Finally we execute the two macros to get our analyses.

B SAS macros for implementing the models

```

*****;

*** Step 1 ***
    Compile the macro FACTORS (see below).

*** Step 2 ***
    Compile the macro ANALYSIS (see below).

*** Step 3 ***
    Run the following program;

options nodate nonumber ls=85 ps=66;
%let DataLib = work;
%let DataDSN=data1;
%global DataLib DataDSN;

* libname &DataLib 'Provide the correct path here';
* libname AAAAA 'Provide the correct path here';

data &DataLib..&DataDSN;
    INPUT COUNT @@;
datalines;
18 17 16 4 2
24 105 109 59 21
23 84 289 217 95
8 49 175 348 198
6 8 69 201 246
;
run;

%Factors(5);
%Analysis(5,count,combined);

***** STEP 1 Macro *****;

%macro Factors(a,lib=work,FactorDSN=generate);

data generate1;
    do R=1 to &a;
        do C=1 to &a;

*****
create vectors for
S symmetry
L vertical mobility
    T Triangles
    Q Quasi-Independencefactor variable
    DA Diagonal Absolute
    D Diagonal factor variable
    F Fixed distance
OS1 Odds symmetry I
OS2 Odds symmetry II
*****;

            if R lt C then
            do;
                k=abs(R-C);
                S=k+1-(R+1)*((0.5*R)+1)+(&a+3)*(R+1)-3-2*&a;
                L=1;
                T=1;
                Q=&a+1;
                DA=abs(R-C);
                D=abs(R-C);
                F=abs(R-C)+1;
                OS1=R;
                OS2=2*&a-C;
            end;

            if R gt C then
            do;
                k=abs(R-C);
                S=(k+1)-(C+1)*(.5*C+1)+(&a+3)*(C+1)-3-2*&a;
                L=1;
                T=2;
                Q=&a+1;
                DA=abs(R-C);
                D=abs(R-C)+(&a-1);

```

```

        F=1;
        OS1=(&a-1)+C;
        OS2=(&a+1)-R;
    end;

    if R eq C then
    do;
        S=1-(C+1)*(.5*C+1)+(&a+3)*(C+1)-3-2*&a;
        L=2;
        T=3;
        Q=R;
        DA=&a;
        D=2*&a-1;
        F=1;
        OS1=2*&a-1;
        OS2=2*&a-1;
    end;

    U=R*C;
    output;
    drop k;
end; end;
run;

* generate FF,SS, H, and DD vectors;
* use ID variable to help in merging with the V vectors;

data generate2;
set generate1;
ID=_N_;
ARRAY FF(&a);
ARRAY SS(&a);
ARRAY H(&a);
ARRAY DD(&a);

do i=1 TO &a;
    if R=i then FF(i)=1;
    else FF(i)=0;
    if C=i then SS(i)=1;
    else SS(i)=0;
    H(i)=FF(i)+SS(i);
    DD(i)=FF(i)-SS(i);
end;
drop i;
run;

*** create the V vectors;
%global h;
%do h=1 %to %eval(&a-1);
    data vdata&h;
    ID=0;
    %do i=1 %to &a;
        %do j=1 %to &a;
            ID=ID+1;
            %if &i <=%eval(&h) and &j <=%eval(&h) %then %do;
                v&h=2; %end;
            %else %if &i >=%eval(&h+1) and &j >=%eval(&h+1)
                %then %do; v&h=2; %end;
            %else %do; v&h=1; %end;
        end;
    output;
    %end;
%end;
proc sort data=vdata&h; by ID;
%end;
run;

*** combine v vectors with other factors ---;

data &lib..&FactorDSN(drop=ID);
merge generate2
    %do k=1 %to %eval(&a-1);
        vdata&k
    %end;
;
by ID;
run;
%mend factors;

***** STEP 2 Macro *****

```

```

%macro analysis(a, depvar, CombinedDSN, lib=work, FactorDSN=generate);

    data &lib..&CombinedDSN;
    merge &DataLib..&DataDSN &lib..&FactorDSN;

    *** Fit asymetry models O, F, T, D, OS1, OS2, 2RPA, LDPS respectively ***;
    %let b = %eval (&a-1);
    proc genmod data=&lib..&CombinedDSN;
        class s;
        model &depvar= s /dist=poi;
        title 'Asymetry Model';
        title2 'O Model';

    proc genmod data=&lib..&CombinedDSN;
        class s;
        model &depvar= s f/dist=poi;
        title2 'LDPS Model';

    proc genmod data=&lib..&CombinedDSN;
        class s t;
        model &depvar= s t /dist=poi;
        title2 'T Model';

    proc genmod data=&lib..&CombinedDSN;
        class s d;
        model &depvar= s d /dist=poi;
        title2 'D Model';

    proc genmod data=&lib..&CombinedDSN;
        class s os1;
        model &depvar= s os1 /dist=poi;
        title2 'OS1 Model';

    proc genmod data=&lib..&CombinedDSN;
        class s os2;
        model &depvar= s os2 /dist=poi;
        title2 'O Model';

    proc genmod data=&lib..&CombinedDSN;
        class s t;
        model &depvar= s t f /dist=poi;
        title2 '2RPA Model';
run;

    *** Fit skew-symmetry models O, T, D, QOS respectively ***;

    proc genmod data=&lib..&CombinedDSN;
        class r c s;
        model &depvar= r c s /dist=poi;
        title 'Skew-symmetry Model';
        title2 'O Model';

    proc genmod data=&lib..&CombinedDSN;
        class r c s t;
        model &depvar= r c s t /dist=poi;
        title2 'T Model';

    proc genmod data=&lib..&CombinedDSN;
        class r c s d;
        model &depvar= r c s d /dist=poi;
        title2 'D Model';

    proc genmod data=&lib..&CombinedDSN;
        class r c s os1;
        model &depvar= r c s os1 /dist=poi;
        title2 'QOS Model';
run;

    *** Fit non-independence models as they appear respectively in Table 2 ***;

    proc genmod data=&lib..&CombinedDSN;
        class r c;
        model &depvar= r c /dist=poi;
        title 'Non-Independence Model';
        title2 'O Model';

    proc genmod data=&lib..&CombinedDSN;
        class r c;
        model &depvar= r c f/dist=poi;
        title2 'F Model';

```



```

proc genmod data=&lib..&CombinedDSN;
  class r c;
  model &depvar= r c u/dist=poi;
  title2 'U Model';

proc genmod data=&lib..&CombinedDSN;
  class r c v1-v&b;
  model &depvar= r c v1-v&b /dist=poi;
  title2 'V Model';

proc genmod data=&lib..&CombinedDSN;
  class r c;
  model &depvar= r c L /dist=poi;
  title2 'L Model';

proc genmod data=&lib..&CombinedDSN;
  class r c t;
  model &depvar= r c t /dist=poi;
  title2 'T Model';

proc genmod data=&lib..&CombinedDSN;
  class r c q;
  model &depvar= r c q /dist=poi;
  title2 'Q Model';

proc genmod data=&lib..&CombinedDSN;
  class r c d;
  model &depvar= r c d /dist=poi;
  title2 'D Model';

proc genmod data=&lib..&CombinedDSN;
  class r c da;
  model &depvar= r c da /dist=poi;
  title2 'DA Model';

proc genmod data=&lib..&CombinedDSN;
  class r c da t;
  model &depvar= r c da t /dist=poi;
  title2 'DAT Model';

proc genmod data=&lib..&CombinedDSN;
  class r c;
  model &depvar= r c u f /dist=poi;
  title2 'UF Model';

*** Fit Non-symmetry + independence models ***;
*** 0, F, U, T, D, DA, and DAT respectively ***;

proc genmod data=&lib..&CombinedDSN;
  model &depvar= h1-h&b /dist=poi;
  title 'Non-symmetry and Independence Model';
  title2 '0 Model';

proc genmod data=&lib..&CombinedDSN;
  model &depvar= h1-h&b f /dist=poi;
  title2 'F Model';

proc genmod data=&lib..&CombinedDSN;
  model &depvar= h1-h&b u /dist=poi;
  title2 'U Model';

proc genmod data=&lib..&CombinedDSN;
  class t;
  model &depvar= h1-h&b t /dist=poi;
  title2 'T Model';

proc genmod data=&lib..&CombinedDSN;
  class d;
  model &depvar= h1-h&b d /dist=poi;
  title2 'D Model';

proc genmod data=&lib..&CombinedDSN;
  class da;
  model &depvar= h1-h&b da /dist=poi;
  title2 'DA Model';

proc genmod data=&lib..&CombinedDSN;
  class da t;

```

```
model &depvar= h1-h&b da t /dist=poi;  
title2 'DAT Model';  
run;  
%mend analysis;
```

C Vectors generated for fitting the Models

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

$$T = \begin{bmatrix} 3 & 1 & 1 & 1 & 1 \\ 2 & 3 & 1 & 1 & 1 \\ 2 & 2 & 3 & 1 & 1 \\ 2 & 2 & 2 & 3 & 1 \\ 2 & 2 & 2 & 2 & 3 \end{bmatrix} \quad OS1 = \begin{bmatrix} 9 & 1 & 1 & 1 & 1 \\ 5 & 9 & 2 & 2 & 2 \\ 5 & 6 & 9 & 3 & 3 \\ 5 & 6 & 7 & 9 & 4 \\ 5 & 6 & 7 & 8 & 9 \end{bmatrix} \quad OS2 = \begin{bmatrix} 9 & 8 & 7 & 6 & 5 \\ 4 & 9 & 7 & 6 & 5 \\ 3 & 3 & 9 & 6 & 5 \\ 2 & 2 & 2 & 9 & 5 \\ 1 & 1 & 1 & 1 & 9 \end{bmatrix}$$

$$F = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad Q = \begin{bmatrix} 1 & 6 & 6 & 6 & 6 \\ 6 & 2 & 6 & 6 & 6 \\ 6 & 6 & 3 & 6 & 6 \\ 6 & 6 & 6 & 4 & 6 \\ 6 & 6 & 6 & 6 & 5 \end{bmatrix} \quad V1 = \begin{bmatrix} 2 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & 2 \\ 1 & 2 & 2 & 2 & 2 \\ 1 & 2 & 2 & 2 & 2 \\ 1 & 2 & 2 & 2 & 2 \end{bmatrix}$$

$$V2 = \begin{bmatrix} 2 & 2 & 1 & 1 & 1 \\ 2 & 2 & 1 & 1 & 1 \\ 1 & 1 & 2 & 2 & 2 \\ 1 & 1 & 2 & 2 & 2 \\ 1 & 1 & 2 & 2 & 2 \end{bmatrix} \quad V3 = \begin{bmatrix} 2 & 2 & 2 & 1 & 1 \\ 2 & 2 & 2 & 1 & 1 \\ 2 & 2 & 2 & 1 & 1 \\ 1 & 1 & 1 & 2 & 2 \\ 1 & 1 & 1 & 2 & 2 \end{bmatrix} \quad V4 = \begin{bmatrix} 2 & 2 & 2 & 2 & 1 \\ 2 & 2 & 2 & 2 & 1 \\ 2 & 2 & 2 & 2 & 1 \\ 2 & 2 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 & 2 \end{bmatrix}$$

$$R = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 & 3 \\ 4 & 4 & 4 & 4 & 4 \\ 5 & 5 & 5 & 5 & 5 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix} \quad U = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 6 & 8 & 10 \\ 3 & 6 & 9 & 12 & 15 \\ 4 & 8 & 12 & 16 & 20 \\ 5 & 10 & 15 & 20 & 25 \end{bmatrix}$$

$$H1 = \begin{bmatrix} 2 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \quad H2 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} \quad H3 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 2 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$H4 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 2 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

References

- Agresti, A. (1983). "A simple diagonals-parameter symmetry and quasi-symmetry model". *Statist. Prob. Letters*, 1, 313-316.
- Agresti, A. (1996). *An Introduction to Categorical Data Analysis* John Wiley and Sons, New York.
- Bishop, Y.M; Fienberg, S. & Holland, P.W. (1975). *Discrete Multivariate Analysis*. MIT Press.
- Caussinus, H. (1965). "Contributions a l'analyse statistique des tableaux de correlation". *Ann. Fac. Sci. Univ. Toulouse*, 29, 77-182.
- Clogg, C. C. (1995). "Latent class models". In *Handbook of Statistical Modeling for the Social and Behavioral Sciences*. Edited by Armiger, G, Clogg, C. C. & Sobel, M. E. 311-359.
- Clogg, C.C., Eliason, S.R. & Grego, J.M. (1990). "Models for the Analysis of Change in Discrete Variables". In *Statistical Methods in Longitudinal Research*, Vol 2, Ed. A. Von Eye, San Diego: Academic Press, 409-441.
- von Eye, A. & Spiel, C. (1996). "Standard and Non-standard Log-Linear Symmetry Models for Measuring change in categorical Variables". *American Statistician*, 50, 300-305.
- Firth D. & Treat, B. R. (1988). "Square contingency tables and GLIM". *GLIM Newsletter*, 16, 16-20.
- Friedl, H. (1995). "A note on generating the factors for symmetry parameters in log-linear models". *GLIM Newsletter*, 24, 33-36.
- Goodman, L.A. (1972). "Some multiplicative models for the analysis of cross-classified data". In Le carm *et al.. Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, 649-695. Berkeley: University of California Press.
- Goodman, L.A. (1979). "Multiplicative models for the analysis of occupational mobility tables and other kinds of cross-classification tables". *Amer. Jour. Sociology*, 84, 804-829.
- Goodman, L. A. (1985). "The analysis of cross-classified data having ordered and/or unordered categories: Association models, correlation models, and asymmetry models for contingency tables with or without missing entries". *The Annals of Statistics*, 13, 10-69.
- Haberman, S. J. (1978). *Analysis of Qualitative Data*, Vol 2 New Developments. Academic Press, New York.
- Hope K. (1982). "Vertical and non vertical class mobility in three countries". *American Sociological Review*, 47, 99-113.
- Katari, M. (1993). "Fitting asymmetry and non-independence models with GLIM". *GLIM Newsletter*, 22, 46-50.
- Kutylowski, A. J. (1989). "Analysis of symmetric cross-classifications". In *Lecture Notes in Statistics*, 57, 188-197.
- Lawal, H. B. (1992). "Parsimonious uniform-distance association models for the occupational mobility data". *J. Japan Statist. Soc.*, 22, 126-134.
- Lawal H. B. (1996). "Using SAS GENMOD procedure to fit diagonal class models to square contingency tables having ordered categories". *Proceedings of the Midwest SAS Users Group*, 149-160.
- Lawal, H.B. (2001). "Modeling symmetry models in square contingency tables with ordered categories". *Jour. Comput. Statist. and Simul.*, 71, 59-83.
- Lawal, H.B. (2002). "Application of the non-standard log-linear models to symmetry and diagonal parameter models in square contingency tables". Submitted.

- Lawal H. B. & Upton G.J.G. (1990). "Alternative interaction structures in square contingency tables having ordered classificatory variables". *Quality & Quantity*, 24, 107-127.
- Lawal, H.B. & 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.
- McCullagh, P. (1978). "A class of parametric models for the analysis of square contingency tables with ordered categories". *Biometrika*, 65, 413-418.
- Tomizawa, S. (1985). "Decompositions for odds-symmetry models in a square contingency table with ordered categories". *J. Japan Statist. Soc.*, 15, 151-159.
- Tomizawa, S. (1987). "Decomposition for 2-Ratios-Parameter symmetry model in square contingency tables with ordered categories". *Biom. J.*, 1, 45-55.
- Tomizawa, S. (1992). "Quasi-diagonals-parameter symmetry model for square contingency tables with ordered categories". *Calcutta Statist. Assoc. Bull.*, 39, 53-61.
- Upton, G. J. G. (1985). "A survey of log linear models for ordinal variables in an $I \times J$ contingency table". *Guru Nanak Jour. Sociology*, 6, 1-18.
- Upton G.J.G. & Sarlvik, B. (1981). "A loyalty-distance model for voting change". *Jour. R. Statist. Soc. A*, 144, 247-259.
- Yamagushi, K. (1990) "Some models for the analysis of asymmetric association in square contingency tables with ordered categories". In *Sociological Methodology 1990*, ed. C.C. Clogg, Oxford: Basil Blackwell, pp. 181-212.

TABLE 1
Danish Occupational mobility data, Svalastoga (1959).

Father's Status	Son's Status					Total
	(1)	(2)	(3)	(4)	(5)	
(1)	18	17	16	4	2	57
(2)	24	105	109	59	21	318
(3)	23	84	289	217	95	708
(4)	8	49	175	348	198	778
(5)	6	8	69	201	246	530
Total	79	263	658	829	562	2391

where:

- (1) Professional and high administrative
- (2) Managerial, executive and high supervisory
- (3) Low inspectional and supervisory
- (4) Routine non-manual and skilled manual
- (5) Semi- and unskilled manual

TABLE 2

Results of fitting all models discussed in this paper to the data in Table 1

Models	Degrees of Freedom	GOF Statistics		P-value
		G^2	X^2	
Asymmetry models:				
S (O)	10	24.8021	24.4211	0.0057
S+F (F)	9	19.0543	18.9374	0.0247
S+T (T)	9	18.8188	18.5221	0.0268
S+D (D)	6	14.8378	14.8837	0.0216
S+OS1 (OS1)	6	4.5557	4.3781	0.6019
S+OS2 (OS2)	6	15.9891	15.6910	0.0138
S+T+F (2RPA)	8	18.6809	18.4421	0.0167
Skew-asymmetry models:				
QS (O)	6	6.4681	6.3070	0.3728
QS+T (T)	5	6.4001	6.2287	0.2692
QS+D (D)	3	4.4401	4.3971	0.2177
QS+OS1 (QOS)	3	3.2280	3.1016	0.3578
Non-independence models:				
O	16	654.2073	754.1043	0.0000
F	15	41.8886	38.7366	0.0002
U	15	47.9921	68.4557	0.0000
V	12	32.8302	29.9336	0.0010
L	15	350.5954	332.6755	0.0000
T	14	349.7620	331.1906	0.0000
Q	11	248.6958	270.2523	0.0000
D	9	10.2278	10.1403	0.3324
DA	12	12.4137	12.2081	0.4131
DAT	11	12.3294	12.1606	0.3394
UF	14	17.7791	19.6632	0.2170
Non-symmetry + Independence models:				
O	20	667.3834	758.9443	0.0000
F	19	555.6246	546.1145	0.0000
U	19	64.7302	85.2490	0.0000
T	18	360.1853	342.1155	0.0000
D	12	20.1404	19.7887	0.0645
DA	16	30.1046	29.5336	0.0175
DAT	15	24.1213	23.6868	0.0631

where in the Table above, the P-value is computed as $P(G^2 \geq \chi^2_\nu)$ and the QS model is implemented as R+C+S.