Statistics Canada Statistique Canada

Methodology Branc

Social Survey
Methods Division

Canada

[redacted]

WORKING PAPER NO. SSMD-88-032 E

METHODOLOGY BRANCH

2-190E
C.2

MODEL SELECTION FOR MULTIWAY CATEGORICAL DATA USING PARTITIONING OF CHI-SQUARE:

Part II - Selected Examples and Program Documentation

SSMD-88-032 E

A.C. Singh and N. Ferragne
Social Survey Methods Division, Statistics Canada, Ottawa

ABSTRACT

This report contains several examples on log-linear model selection for multiway categorical data taken from the literature and their analyses by the RX^2 method (Singh, 1988). The RX^2 method provides a fast screening procedure for model selection within the class of hierarchical models. It uses ranking of all parameters in a decreasing order of importance as measured by P-values. RX^2 provides an alternative to the ranking method based on the well-known partitioning of the likelihood ratio test statistic, G^2 which can be quite expensive computationally. Use of RX^2 for the problem of polytomous logit model selection is also considered. Detailed computer program codes (in SAS-IML) are documented in the appendix.

RÉSUMÉ

Ce rapport contient plusieurs exemples de sélection de modèles log-linéaires pour des données discrètes mutidimensionelles provenant de la littérature et analysés à l'aide de la méthode RX^2 (Singh, 1988). La méthode RX^2 est une procédure simple et rapide pour la sélection de modèles parmi la classe de modèles hiérarchiques. Cette procédure utilise le "RANKING" des paramètres du modèle en ordre d'importance décroissant tel qu'évalué par les "p-values". La méthode RX^2 fournit une alternative à la méthode du "Ranking" des paramètres reposant sur la partition bien connue du test statistique du rapport de vraisemblance, G^2 qui peut être très coûteux à déterminer. L'utilisation de la méthode RX^2 pour résoudre le problème de sélection de modèles logistiques "polytomous" est également considérée. La documentation détaillée du programme informatique (en SAS-IML) est disponible en Appendice.

Some Typographical Errors for: **WORKING PAPER NO. SSMD-88-032 E**

Page 4: $\sum_{i'} \sum_{k'} u^{AB}_{i'k'} \; x^{A}_{ii'} \; x^{B}_{kk'}$

Should read as:

$\sum_{i'} \sum_{k'} u^{AC}_{i'k'} \; x^{A}_{ii'} \; x^{C}_{kk'}$

Page 12: "The selected model as reported in Agresti (1984) is [A] [BC]. A glance at the table 1.1 gives us the model [A] [BC] as an initial model."

Should read as:

"The selected model as reported in Agresti (1984) is [AB] [BC]. A glance at the table 1.1 gives us the model [AB] [BC] as an initial model."

Page 15: In the Observed Counts table: line 2,1,2 86

Should read as: line 2,1,1 86

Page 19: "The model selected by Goodman's forward or backward stepwise procedure (see Fienberg, 1977) at 5% level was found to be [AC] [BC] [CD]."

Should read as:

"The model selected by Goodman's forward or backward stepwise procedure (see Fienberg, 1977) at 5% level was found to be [AC] [BD] [CD]."

Page 23: In the Observed Counts table: line 2,1,2,3 38

Should read as: line 2,1,2,2 38

Page 27 "The model selected in Fienberg (1977) is [AB] [AC] [AE] [BD] [BD] [CDE]."

Should read as:

"The model selected in Fienberg (1977) is [AB] [AC] [AE] [BD] [BE] [CDE]."

1. INTRODUCTION

The RX^2 method (Singh, 1988) produces a ranking of hierarchical log-linear model parameters in a decreasing order of importance as measured by P-values. The constraints on parameter inclusion imposed by sampling designs can be easily incorporated in the ranking. The method is useful as a screening procedure for model selection in which consequence of adding or deleting parameters becomes apparent in the process of model building. Goodman's method of partitioning the likelihood ratio statistic, G^2, could in principle be used for the ranking problem and provides, in general, a slightly different but optimal ranking RG^2. In our experience with several empirical examples, it was found that the ranking RX^2 agreed well with RG^2 and any discrepancy present did not affect the model selection process. However, the RG^2 method entails a large number of model fits corresponding to various possible parameter combinations. This number increases very rapidly as the number of cross-classification variables increases. For instance, when the number q of variables is respectively 2,3,4,5, and 6, the corresponding number of required model fits is 3,11,42,141, and 493. The general formula with q variables is $(\frac{1}{2})\ \Sigma_{i=1}$ $\binom{q}{i}\ (\binom{q}{i}+1)$. Clearly, RG^2 would be very costly when q is 4 or more. For exploratory purposes, it would be desirable to have a quick and simple alternative method. The method RX^2, although not optimal, can be used as an alternative to RG^2.

Section 2 contains a brief review of RX^2 and in section 3, a stepwise description of the implementation of RX^2 is presented. The application of RX^2 to various examples given in section 4 is based on a sample software developed by the second author. The computer program in SAS-IML is documented in the appendix. There are fifteen examples taken from the literature and they range from 3-dimensional to 6-dimensional tables. Examples of polytomous logit models are also considered indirectly via log-linear models. The case of dichotomous logit model is treated directly by a simple modification of RX^2. The results from RX^2 analyses of various examples were found to be consistent with those reported in the literature in almost all cases. In many examples, RG^2 ranking was also computed and comparison with RX^2 showed a good agreement.

The fifteen examples along with their source for reference are listed below.

Example 1 (3-dim) Death Penalty Data (Agresti, 1984, p. 51 and p. 110)
- log-linear and logit (dichotomous)

Example 2 (3-dim) Food Poisoning Data (Bishop, Fienberg, and Holland, 1975, p. 90)
- log-linear

Example 3 (3-dim) Race, Sex, and Homicide weapon (Haberman, 1978, p. 162 and p. 293)
- log-linear and logit (dichotomous)

Example 4 (3-dim) Woman's Place, Sex, and Education (Haberman, 1978, p. 183)
- log-linear

Example 5 (3-dim) Dumping Severity Data (Agresti, 1984, p. 66)
- log-linear

Example 6 (3-dim) Heart Disease Data (Agresti, 1984, p. 111)
- log-linear and logit (dichotomous)

Example 7 (3-dim) Sex of Respondent, Degree of Husband, and Degree of Wife (Haberman, 1978, p.227)
- log-linear

Example 8 (4-dim) Detergent Preference Data (Feinberg, 1977, p. 59 and p. 79)
- log-linear and logit

Example 9 (4-dim) Symptoms of Psychiatric Patients (Benedetti and Brown, 1978)
- log-linear

Example 10 (4-dim) Year of Survey, Sex, Education, and Place (Haberman, 1978, p. 255)
- log-linear

Example 11 (4-dim) Year of Survey, Religion, Education, and Attitude toward Non-therapeutic Abortion (Haberman, 1978, p. 262)
- log-linear

Example 12 (4-dim) Armed Forces Qualification Test (Fienberg, 1977, p. 89)
- logit (polytomous)

Example 13 (5-dim) Jamaican Lizards (Bishop, Fienberg, and Holland, 1975, p. 164)
- log-linear

Example 14 (5-dim) Cancer Knowledge Data (Fienberg 1977, p. 73 and p. 80)
- log-linear and logit (dichotomous)

Example 15 (6-dim) Risk Factors for Coronary Heart Disease (Edwards and Havránek, 1985)
- log-linear

2. RX^2 METHOD : A BRIEF REVIEW

For convenience, we shall describe RX^2 using a 3-dimensional table $I \times J \times K$ of counts $\{X_{ijk}\}$ corresponding to three factors A, B and C. For the problem of ranking $T(=IJK)$ parameters of the saturated log-linear model, we will assume that we are not interested in discriminating effects of different levels within the same combination of factors. It will be sufficient to rank overall factor effects only e.g. for a 3-dimensional table, we would like to rank in decreasing under of importance (as measured by P-values) first the two-factor effects (AB, AC, BC), and then the one-factor or main effects (A, B, C). We start with an initial ranking $R(0)$ of T parameters in the lexicographic order such that it respects certain constraints of hierarchical models and sampling designs. There will be no design constraints for Poisson or multinomial sampling. However, for product multinomial sampling as well as for logit models, suitable constraints on the ranking of factor effects within certain classes will be needed.

Next, we compute Pearson-Fisher's X^2 statistics for a few hypotheses. For a three-way table, we consider three hypotheses $H_1 \subset H_2 \subset H_3$ where H_3 denotes the hypothesis of all three and higher-order factor effects are zero and H_2, H_1 are defined similarly. We then use a representation of $X^2(H_i)$ for each i as a sum of r_i non-negative components C_h^2, $h=T-r_i+1$ to T where r_i is the number of test parameters under H_i. The components C_h^2 termed "Conditional Contributions" have asymptically χ_1^2 distributions and represent contributions of test parameters to the total X^2 measure after controlling successively for all those parameters with higher priority according to a tentatively chosen ranking. The C_h^2 components are used to obtain a final ranking of parameters when the importance is measured in terms of low P-values. Using this criterion for judging importance may lead sometimes to a tie between effects. In those situations, some ad hoc ranking such as the one used in the initial ranking $R(0)$ can be done for breaking the ties.

For a given log-linear hypothesis under consideration, the conditional contributions C_h^2's such that $X^2 = \Sigma C_h^2$ can be defined as follows. Following Haberman (1978, p.208) write the saturated log-linear model for the expected counts $\{m_{ijk}\}$ for a three-dimensional table as given below. Let $i' = 1$ to $I-1$, $j' = 1$ to $J-1$, and $k' = 1$ to $K-1$, then

$$\log m_{ijk} = u + \sum_{i'} u^{A}_{i'} x^{A}_{ii'} + \sum_{j'} u^{B}_{j'} x^{B}_{jj'} + \sum_{k'} u^{C}_{k'} x^{C}_{kk'} +$$

$$\sum_{i'}\sum_{j'} u^{AB}_{i'j'} x^{A}_{ii'} x^{B}_{jj'} + \sum_{i'}\sum_{k'} u^{AB}_{i'k'} x^{A}_{ii'} x^{B}_{kk'} +$$

$$\sum_{j'}\sum_{k'} u^{BC}_{j'k'} x^{B}_{jj'} x^{C}_{kk'} + \sum_{i'}\sum_{j'}\sum_{k'} u^{ABC}_{i'j'k'} x^{A}_{ii'} x^{B}_{jj'} x^{C}_{kk'},$$

where $x^{A}_{ii'} = 1$ for $i=i'$; 0 for $i \neq i'$ and $i \neq I$; -1 for $i=I$, and so on.

Let M denote the TxT matrix with (t,h) th element $(\partial \log m_t / \partial u_h)$, h, t=1, ...T where t corresponds to the cell (i,j,k) as a result of the lexicographic ordering of cell indices and h corresponds to the order of a tentatively chosen ranking of u-parameters. For a 2x2x2 table, we have under the lexicographic ranking R(0) for u-parameters,

$$M = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \end{bmatrix}$$

The u-parameters themselves can be expressed as a linear function of log m-parameters (Haberman, 1978, p. 249). For a three-way table, we have

$$\begin{aligned}
Tu &= \Sigma\Sigma\Sigma \log m_{ijk} \\
Tu_{i'} &= \Sigma\Sigma\Sigma\, (I\,\delta_{ii'} - 1) \log m_{ijk} \\
Tu_{j'} &= \Sigma\Sigma\Sigma\, (J\,\delta_{jj'} - 1) \log m_{ijk} \\
Tu_{k'} &= \Sigma\Sigma\Sigma\, (K\,\delta_{kk'} - 1) \log m_{ijk} \\
Tu_{i'j'} &= \Sigma\Sigma\Sigma\, (I\,\delta_{ii'} - 1)(J\,\delta_{jj'} - 1) \log m_{ijk} \\
Tu_{i'k'} &= \Sigma\Sigma\Sigma\, (I\,\delta_{ii'} - 1)(K\,\delta_{kk'} - 1) \log m_{ijk} \\
Tu_{j'k'} &= \Sigma\Sigma\Sigma\, (J\,\delta_{jj'} - 1)(K\,\delta_{kk'} - 1) \log m_{ijk} \\
Tu_{i'j'k'} &= \Sigma\Sigma\Sigma\, (I\,\delta_{ii'} - 1)(J\,\delta_{jj'} - 1)(K\,\delta_{kk'} - 1) \log m_{ijk},
\end{aligned}$$

where $\delta_{xy} = 1$ if $x=y$ and 0 otherwise, and $T=IJK$.

Now if U denotes the TxT matrix whose (h,t) th element is $(\partial u_h/\partial \log m_t)$. Then clearly, M^{-1} is U and it has a closed form expression. For example, for a 2x2x2 table under the lexicographic ranking R(0), we have

$$U = (1/8)\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix}$$

We next define the T-vector

$$\underset{\sim}{f} = M' (\underset{\sim}{x} - \underset{\sim}{m}),$$

and the TxT covariance matrix

$$S(T) = M' \; \text{Diag} (\underset{\sim}{m}) \; M$$

Let S(h) for h=1 to T, denote the principal submatrices of S(T). Then, the inverse matrix G(h) of S(h) can be easily obtained from $S(T)^{-1}$ by using the lemma on inverse of partitioned matrices in a reverse fashion in successive steps. We will use the term "devolution" to denote this operation of finding G(h-1) from G(h) successively for h=T, T-1,... Note that the matrix $S^{-1}(T)$ or G(T) does not require inversion because it can be expressed as

$$G(T) = S^{-1}(T) = U \; \text{Diag} (\underset{\sim}{m})^{-1} \; U'.$$

Let $\hat{m}$ be the maximum likelihood estimate of $\underset{\sim}{m}$ under a given null hypothesis with r degrees of freedom. Then

$$X^2 = (\underset{\sim}{x} - \underset{\sim}{m})' \operatorname{Diag}(\underset{\sim}{m})^{-1} (\underset{\sim}{x} - \underset{\sim}{m}) \Big|_{m = \hat{m}}$$

$$= \underset{\sim}{f}' G(T) \underset{\sim}{f} \Big|_{u(m) = u(\hat{m})}$$

$$= \sum_{h=T-r+1}^{T} C_h^2 ,$$

where

$$C_h^2 = \{(g_{h1}, \ldots, g_{hh})(f_1, \ldots, f_h)'\}^2 / g_{hh} ,$$

all terms are evaluated at $\hat{m}$ and g_{hi}'s form the last row of the hxh matrix $G(h)$.

For each hypothesis H_i, the components C_h^2 are calculated for various possible parameter rankings within the class of i-factor effects in order to find the ranking according to smallest P-values. For a three-way table, we start with H_2, and rank the two-factor effects which would give a revised version R(I) of the initial ranking R(0). Next, under H_1, the one-factor effects are ranked resulting in a revised version R(II) of R(I). The R(II) gives the desired ranking RX^2 which can be viewed as a "composite" ranking based on a few selected groups of test parameters. The number of groups or steps required in obtaining the composite ranking RX^2 is simply one less the number of variables.

Although parameter ranking for logit models (dichotomous or polytomous) can be obtained indirectly by first ranking log-linear model parameters under suitable constraints, the case of dichotomous logit models can be treated directly by a simple modification of RX^2 method. The details can be found in Singh (1988).

3. IMPLEMENTATION OF THE RX^2 METHOD

We will now present the computer implementation aspects of the RX^2 method. This will be done in a systematic manner using a series of steps. The implementation steps make reference to various modules of a sample software developed for the RX^2 method. The software is given in the appendix.

We will consider two cases. The first one will be the ranking of factor effects of a log-linear model (with the option of a fixed marginal total corresponding to a polytomous logit model) and the second one will be the ranking of factor effects of a logit model corresponding to a dichotomous "response" variable. The case of dichotomous logit model is considered separately because in this case, it is convenient to modify RX^2 to get directly the desired ranking.

3.1 Ranking of factor effects of a log-linear model

There is a total of 28 steps (repetitive in nature) performed sequentially as listed below.

STEP 1: Specification of the number of categorical variables and the number of levels for each of these variables (see the module MAIN contained in appendix A.1).

STEP 2: Indication of whether or not the design constraint of a fixed marginal total is required (see the module MAIN in appendix A.1).

STEP 3: Specification of the response variable. Requesting a log-linear model having a fixed marginal total induces a logit model. The response variable must be specified. If we do not wish to consider a log-linear model having a fixed marginal total, the default option for response variable is the last categorical variable (see the module MAIN in appendix A.1). In this sample program, the number of variables in the fixed marginal table is restricted to be one less the total number of the variables. This restriction will be removed in a future revision of the program.

STEP 4: Specification of the total number of u-parameters in the model and the number of effects for each class of factor effects (number of main effects, number of 2-factor effects, number of 3-factor effects, etc.; see the module MAIN in appendix A.1).

STEPS 5: Specification of the initial ranking R(0) of the factor effects and the degrees of freedom associated with the factor effects. The initial ranking is based on a "partial" lexicographic arrangement of the factor effects. We say "partial" because the initial ranking of some of the effects may not be in the lexicographic arrangement depending on the response variable selected (see the module MAIN in appendix A.1).

STEP 6: Creation of the initial M matrix using an algorithmic approach under the initial ranking R(0) (see the module M contained in appendix A.2).

STEP 7: Creation of the initial U matrix using an algorithmic approach under the initial ranking R(0) (see the module U contained in appendix A.3).

We may now proceed with the actual ranking of the factor effects. Each class of factor effects is ranked separately. We start by ranking the class of highest-order factor effects, then we rank the next class of highest-order factor effects, and so on until we get to the main effects. Suppose that for a class of factor effects, the initial ranking is A B C. Ranking for this class is done in the following fashion: we first compute the contribution of C and the associated P-value, then we successively exchange A and B with C to determine their P-values. The factor with the largest P-value computed will be given the rank in the third position. The same process is repeated until all the factor effects have been ranked.

STEP 8: Select the class of factor effects corresponding to the highest interaction (this occurs when the SETUSED variable is equal to the total number of categorical variables). For this class of factor effects no ranking is needed since there is only one member in the class (see the module MAIN in appendix A.1).

STEP 9: Specification of the observed frequencies. The observed counts must be in the lexicographic arrangement (see the module SET in appendix A.4).

STEP 10: Specification of the expected frequencies. The expected frequencies are estimated by the maximum likelihood estimates (MLE's) obtained from fitting a log-linear model in which all factor effects in the given class and higher-order ones are omitted (the MLE's can be obtained using the categorical modelling procedure CATMOD in SAS). The expected frequencies must also be in the lexicographic arrangement (this is indicated in the module SET in appendix A.4).

STEP 11: Computation of the f vector and G matrix (see module SET in appendix A.4).

STEP 12: Computation of the chi-square component value of all the factor effects. To compute the chi-square component value of a factor effect, we must add chi-square component values of all the levels of a factor effect (see the module CONTRIB in appendix A.5 and in the module CSQUARE in appendix A.6).

STEP 13: Save the results for this part of the ranking process (this is indicated in the module MAIN in appendix A.1).

We repeat steps 14 to 26 until all the classes of factor effects have been ranked.

STEP 14: Select the class of factor effects corresponding to the next highest order of factor effects we wish to rank. This is indicated by the SETUSED variable (see the module MAIN in appendix A.1).

STEP 15: Specification of the "start" and "finish" positions between which ranking is to be done. For example, if we wish to rank four-factor effects in positions 1 2 3 4, we start by determining which effect will go in position 4, then which one will go in position 3 followed by which one will be in position 2. The effect for position 1 is readily determined since by then there is only one member left for ranking. Therefore, the "start" position would be 4 and the "finish" position would be 2. If we are considering a loglinear model having a fixed marginal total (or a logit model), the class of second highest-order factor effects will have one less factor effect to rank since the position for that effect is pre-determined. This is indicated in module MAIN in appendix A.1.

We repeat steps 16 to 23 for all the positions from the "start" to the "finish" position. The term "target position" will be used for the position currently being considered for ranking.

STEP 16: Specification of some counter variables for processing: the position of the factor effect following the target position (NEXTTERM), the total number of u-parameters starting with the ones associated with NEXTTERM up to the last one associated with the highest-order factor effect, the total number of factor effects remaining to be ranked in the class of factor effects we wish to rank, and the total number of u-parameters that precede the class of factor effects we wish to rank. This is indicated in the module RANKING in appendix A.7.

We now determine which factor effect has the largest P-value in a target position. This is indicated in steps 17 to 23.

STEP 17: Specification of the expected frequencies. This is similar to step 10.

STEP 18: Computation of the f vector and G matrix as in step 11.

STEP 19: Devolve the G matrix and reduce the f vector corresponding to all the factor effects following the target position (this is indicated in module RANKING in appendix A.7). In executing this step, we also compute the chi-square component values of all the factor effects following the target position (see module CSQUARE in appendix A.6).

STEP 20: Computation of the contribution of the factor effect which is in the target position as indicated in module RANKING in appendix A.7. This is similar to step 12.

STEP 21: Verify if the factor effect in the target position is the one with the largest P-value (this is indicated in module RANKING in appendix A.7).

STEP 22: Specification of the next unranked factor effect to be moved to the target position in order to determine its contribution (this is indicated in module RANKING in appendix A.7).

STEP 23: In order to compute the contribution in the target position for each unranked factor effect, move the columns of the M matrix and rows of the U matrix corresponding to the current ordering of the factor effects. This is indicated in the module RANKING in appendix A.7. The actual moving of columns and rows is indicated in the module MOVE in appendix A.8. We then return to step 17.

STEP 24: At this stage, the factor effect having the largest P-value in the target position has been determined. We modify the ranking of the factor effects, the columns of the M matrix, and the rows of the U matrix to reflect the ordering just obtained for the current target position (this is indicated in module RANKING in appendix A.7). We then return to step 16 for the next target position.

STEP 25: When all the factor effects within a given class have been ranked, we compute the f vector and the G matrix reflecting the current ordering (this is indicated in module SET in appendix A.4).

STEP 26: Computation of the chi-square components for all the factor effects as in step 12.

STEP 27: Save the results of this part of the ranking process as in step 13. We now return to STEP 14 to start ranking the next highest-order factor effects class.

STEP 28: Print the results of the RX^2 analysis (this is indicated in the module PRINT in appendix A.9).

3.2 Ranking of factor effects of a logit model (for a dichotomous response variable)

The ranking of factor effects of a logit model having a dichotomous response variable basically requires the same steps as those presented in section 3.1 except that the option of specifying a marginal total in step 3 is no longer required. The modules are also the same except for the module SET which is now given in appendix A.10.

4. SELECTED EXAMPLES

We present 15 numerical examples of model selection from data sets which have been analysed previously in the literature. The selected examples come from medical and social fields and vary from three to six-way cross-classifications. Under RX^2 method of ranking factor effects, it is easy to obtain as a byproduct the P-value for each effect. These P-values can then be used to select a few hierarchical models for further investigation. For this purpose, significant i-factor effects are determined from corresponding P-values under H_i, $i=1,2,\ldots$. It is interesting to note that in almost all of our numerical examples, tentative models selected by the RX^2 method at the usual 1% and 5% levels were consistent with the ones selected by some other methods, as reported in the literature. Some of the examples were also analysed with the optimal ranking RG^2 method. It should be noted that the rankings RX^2 or RG^2 are unaffected by the choice of significance level for parameter inclusion.

Example 1 (3 dim): DEATH PENALTY DATA (Agresti, 1984, p.51 and p.110)

This example is based on the data concerning the effects of racial characteristics on the decision regarding whether to impose the death penalty after an individual is convicted for a homicide. The variables considered are

A: defendant's race (white, black),

B: victim's race (white, black),
C: death penalty (yes, no).

The data under the lexicographic arrangement are as follows:

Observed Counts
Cell (i,j,k), x_{ijk}

1,1,1	19	2,1,1	11
1,1,2	132	2,1,2	52
1,2,1	0	2,2,1	6
1,2,2	9	2,2,2	97

The selected model as reported in Agresti (1984) is [A] [BC]. A glance at the table 1.1 gives us the model [A] [BC] as an initial model. The optimal ranking obtained with the RG^2 method was C B A AB BC AC ABC which happens to be identical with the one obtained with the RX^2 method. It can be seen that the RG^2 method also leads us to the same model.

Table 1.1 RX^2 for Death Penalty Data (Log-Linear)

ORDERING	DF	H3	P-VALUE	H2	P-VALUE	H1	P-VALUE
C	1.0000	5.1E-10	1.0000	1.1E-11	1.0000	197.9	0
B	1.0000	8.9E-12	1.0000	5.5E-26	1.0000	31.9141	1.6E-08
A	1.0000	2.1E-12	1.0000	9.6E-25	1.0000	0.1104	0.7397
AB	1.0000	4.1E-11	1.0000	115.0	0	101.6	0
BC	1.0000	1.6E-12	1.0000	5.6149	0.0178	8.9448	.0027826
AC	1.0000	4.8E-10	1.0000	0.2214	0.6379	0.3067	0.5797
ABC	1.0000	0.3755	0.5400	1.5529	0.2127	72.7485	1.4E-17
TOTAL	7.0000	0.3755	0.9998	122.4	0	413.5	0

Suppose, we regard the death penalty verdict (variable C) as the response variable (dichotomous response) and consider the defendant's race (variable A) and the victim's race (variable B) as the explanatory variables. Under this logit approach, two models namely [A] [B] and [B] were found to fit the data well, see Agresti (1984). Using the RX^2 method modified for the logit model, the model [B] would be selected as an initial model and then possibly [A] [B] as the next model, see table 1.2.

Table 1.2 RX^2 for Death Penalty Data (Logit)

ORDERING	DF	H2	P-VALUE	H1	P-VALUE
B	1.0000	2.1E-10	1.0000	5.6149	0.0178
A	1.0000	6.8E-11	1.0000	1.3561	0.2442
AB	1.0000	0.3755	0.5400	.0063795	0.9363
TOTAL	3.0000	0.3755	0.9452	6.9773	0.0726

Example 2 (3-dim): FOOD POISONING DATA (Bishop, Fienberg, and Holland, 1975 p. 90)

The data from this example comes from an epidemiologic study following an outbreak of food poisoning that occurred at an outing held for the personnel of an insurance company. We have a three-dimensional contingency table of 304 observations having the following cross-classification variables:

A: consumer's illness (presence or absence),
B: potato salad (eaten, not eaten),
C: crabmeat (eaten, not eaten).

The data under the lexicographic arrangement are as follows:

Observed Counts
Cell (i,j,k), x_{ijk}

1,1,1	120	2,1,1	80
1,1,2	22	2,1,2	24
1,2,1	4	2,2,1	31
1,2,2	0	2,2,2	23

Table 2. RX^2 for Food Poisoning Data (Log-Linear)

ORDERING	DF	H3	P-VALUE	H2	P-VALUE	H1	P-VALUE
B	1.0000	2.8E-32	1.0000	7.4E-12	1.0000	116.3	0
C	1.0000	3.9E-32	1.0000	6.8E-11	1.0000	90.6447	0
A	1.0000	1.8E-11	1.0000	3.9E-11	1.0000	0.4737	0.4913
AB	1.0000	2.1E-15	1.0000	48.5754	3.2E-12	25.4737	4.5E-07
BC	1.0000	3.1E-11	1.0000	11.7473	6.1E-04	66.3289	3.9E-16
AC	1.0000	9.7E-11	1.0000	9.3179	.0022692	4.7500	0.0293
ABC	1.0000	1.7026	0.1919	5.9817	0.0145	6.9605	.0083328
TOTAL	7.0000	1.7026	0.9745	75.6223	1.1E-13	310.9	0

The selected model as reported in Bishop, Fienberg and Holland (1975) is [AB] [AC] [BC]. Table 2 clearly indicates that the factor effects AB, BC, and AC should be included in our model so that RX^2 method also gives [AB] [BC] [AC] as an initial model. The optimal ranking obtained with the RG^2 method was B C A AB BC AC ABC which is identical with the one obtained with the RX^2 method. It can be seen that RG^2 also leads to the same initial model.

Example 3: RACE, SEX, AND HOMICIDE WEAPON (Haberman, 1978, p. 162 and p. 293)

We are now dealing with a three-dimensional contingency table of 13832 observations which contains a cross-classification of homicide victims in 1970 by race, sex, and type of assault. These variables have the following levels

A: race of victim (white, black),
B: sex of victim (male, female),
C: type of assault (firearms and explosives, cutting and piercing instruments).

The data under the lexicographic arrangement are as follows:

Observed Counts
Cell (i,j,k), x_{ijk}

1,1,1	3910	2,1,1	5218
1,1,2	808	2,1,2	1385
1,2,1	1050	2,2,1	929
1,2,2	234	2,2,2	298

The selected model as reported in Haberman (1978, p. 181) is [AB] [AC] [BC]. A glance at table 3.1 would lead us to select the same model as an initial model. The optimal ranking obtained with the RG^2 method was B C A AB AC BC ABC which shows a slight discrepancy in the ordering of the main effects in comparison to the one obtained with the RX^2 method. However, this does not change the choice of the tentative models by RG^2.

Table 3.1 RX^2 for Data on Race, Sex, and Homicide Weapons (Log-Linear)

ORDERING	DF	H3	P-VALUE	H2	P-VALUE	H1	P-VALUE
A	1.0000	7.3E-33	1.0000	3.0E-09	1.0000	241.6	0
B	1.0000	1.6E-30	1.0000	1.6E-09	1.0000	5611.3	0
C	1.0000	1.3E-30	1.0000	1.2E-09	1.0000	5079.4	0
AB	1.0000	1.9E-31	1.0000	74.8791	0	272.7	0
AC	1.0000	1.7E-33	1.0000	36.6943	1.4E-09	21.5526	3.4E-06
BC	1.0000	1.9E-31	1.0000	4.2828	0.0385	2177.4	0
ABC	1.0000	1.0754	0.2997	1.1672	0.2800	60.6605	6.8E-15
TOTAL	7.0000	1.0754	0.9935	117.0	0	13464.6	0

When we treat the type of assault (variable C) as being the dependent variable and the race of victim (variable A) and sex of victim (variable B) as the independent or explanatory variables we have a dichotomous logit model. The logit model corresponding to the above log-linear model of no three-factor interaction is [A] [B] and from table 3.2, we see that the RX^2 method would also select this model as an initial one for further evaluation.

Table 3.2 RX^2 for Data on Race, Sex, and Homicide Weapons (Logit)

ORDERING	DF	H2	P-VALUE	H1	P-VALUE
A	1.0000	2.8E-31	1.0000	36.6935	1.4E-09
B	1.0000	3.5E-30	1.0000	6.3607	0.0117
AB	1.0000	1.0754	0.2997	1.5820	0.2085
TOTAL	3.0000	1.0754	0.7830	44.6362	1.1E-09

Example 4 (3-dim): WOMAN'S PLACE, SEX, AND EDUCATION (Haberman, 1978, p. 183)

This examples deals with data from the 1975 General Social Survey. The subjects sampled were cross-classified by attitude toward women staying at home, sex of respondent, and education of respondent. The levels of the variables under consideration were

A: sex of respondent (male, female),
B: education of respondent (less than 9 years, between 9 and 12 years, more than 12 years),
C: response (agree, disagree).

The data under the lexicographic arrangement are as follows:

Observed Counts
Cell (i,j,k), x_{ijk}

1,1,1	72	2,1,2	86
1,1,2	47	2,1,2	38
1,2,1	110	2,2,1	173
1,2,2	196	2,2,2	283
1,3,1	44	2,3,1	28
1,3,2	179	2,3,2	187

The total sample size was 1443. A plausible model as given in Haberman (1978 p. 188) is [AB] [BC]. As seen from table 4 the RX^2 method would also lead us to select the same model. The optimal ranking obtained with the RG^2 method was B C A BC AB AC ABC which is identical with the one obtained with the RX^2 method. It can be seen that the two methods lead to the same model.

Table 4 RX^2 for Data on Woman's Place, Sex, and Education (Log-Linear)

ORDERING	DF	H3	P-VALUE	H2	P-VALUE	H1	P-VALUE
B	2.0000	8.5E-12	1.0000	1.7E-10	1.0000	285.8	0
C	1.0000	1.4E-11	1.0000	6.8E-10	1.0000	120.5	0
A	1.0000	2.4E-11	1.0000	7.3E-11	1.0000	14.9751	1.1E-04
BC	2.0000	1.5E-11	1.0000	162.8	0	150.1	0
AB	2.0000	8.9E-10	1.0000	14.9569	5.7E-04	31.9875	1.1E-07
AC	1.0000	1.4E-10	1.0000	0.2335	0.6289	0.4331	0.5105
ABC	2.0000	5.9478	0.0511	0.8352	0.6586	3.0617	0.2164
TOTAL	11.0000	5.9478	0.8768	178.8	0	606.9	0

Example 5 (3-dim): DUMPING SEVERITY DATA (Agresti, 1984, p. 66)

We consider a three-way cross-classification of operation, dumping severity, and hospital. The three variables have the following levels

A: operation (a, b, c, or d),
B: dumping severity (none, slight, moderate)
C: hospital (1, 2, 3, or 4).

The data under the lexicographic arrangement are as follows:

Observed Counts
Cell (i,j,k), x_{ijk}

1,1,1	23	2,1,1	86	3,1,1	20	4,1,1	24
1,1,2	18	2,1,2	18	3,1,2	13	4,1,2	9
1,1,3	8	2,1,3	12	3,1,3	11	4,1,3	7
1,1,4	12	2,1,4	15	3,1,4	14	4,1,4	13
1,2,1	7	2,2,1	10	3,2,1	13	4,2,1	10
1,2,2	6	2,2,2	6	3,2,2	13	4,2,2	15
1,2,3	6	2,2,3	4	3,2,3	6	4,2,3	7
1,2,4	9	2,2,4	3	3,2,4	8	4,2,4	6
1,3,1	2	2,3,1	5	3,3,1	5	4,3,1	2
1,3,2	1	2,3,2	2	3,3,2	2	4,3,2	2
1,3,3	3	2,3,3	4	3,3,3	2	4,3,3	4
1,3,4	1	2,3,4	2	3,3,4	3	4,3,4	4

For this data set, the total sample size was 417. Two possible models as given in Agresti (1984) are [A] [B] [C] and [AB] [C]. The RX^2 method gives us a subset of these two models as an initial model: [B] [C] (see table 5). The next tentative model suggested by the RX^2 method would include the term AB, i.e. the model [AB] [C]. Thus the results are consistent with the earlier analysis. The optimal ranking obtained with the RG^2 method was B C A AB BC AC ABC which is identical to the one obtained with the RX^2 Method. The tentative models selected by RG^2 were found to be identical to those under RX^2.

Table 5 RX^2 for Dumping Severity Data (Log-Linear)

ORDERING	DF	H3	P-VALUE	H2	P-VALUE	H1	P-VALUE
B	2.0000	2.5E-10	1.0000	1.1E-10	1.0000	133.7	0
C	3.0000	2.0E-11	1.0000	9.8E-12	1.0000	29.0911	2.1E-06
A	3.0000	8.8E-11	1.0000	1.4E-10	1.0000	1.0432	0.7908
AB	6.0000	2.5E-10	1.0000	10.5419	0.1036	9.2230	0.1614
BC	6.0000	1.3E-10	1.0000	8.1179	0.2296	20.8297	.0019684
AC	9.0000	3.1E-11	1.0000	0.9671	0.9995	1.2110	0.9988
ABC	18.0000	12.6444	0.8122	12.8546	0.8001	13.2998	0.7735
TOTAL	47.0000	12.6444	1.0000	32.4815	0.9470	208.4	0

Example 6 (3-dim): HEART DISEASE DATA (Agresti, 1984, p. 111)

This example deals with the male residents of Framingham, Massachusetts, aged 40 to 59, classified by blood pressure and serum cholesterol levels. During a six-year follow-up period, they were also classified according to whether they developed coronary heart disease. The variables have the following levels.

A: coronary heart disease (absent, present)
B: serum cholesterol in mg/100cc (less than 200, between 200 and 219, between 220 and 259, more than 259)
C: systolic blood pressure in mm Hg (less than 127, between 127 and 146, between 147 and 166).

The data under the lexicographic arrangement are as follows:

Observed Counts
Cell (i,j,k), x_{ijk}

1,1,1	2	1,3,1	8	2,1,1	117	2,3,1	119
1,1,2	3	1,3,2	11	2,1,2	121	2,3,2	209
1,1,3	3	1,3,3	6	2,1,3	47	2,3,3	68
1,1,4	4	1,3,4	6	2,1,4	22	2,3,4	43
1,2,1	3	1,4,1	7	2,2,1	85	2,4,1	67
1,2,2	2	1,4,2	12	2,2,2	98	2,4,2	99
1,2,3	0	1,4,3	11	2,2,3	43	2,4,3	46
1,2,4	3	1,4,4	11	2,2,4	20	2,4,4	33

The model [AB] [AC] [BC] was found to provide an adequate fit to the data. Table 6.1 would lead us to select the same model under the ranking obtained by the RX^2 method. The optimal ranking obtained with the RG^2 method was A C B AB AC BC ABC which is similar (except for main effect terms) to the one obtained with the RX^2 method. This discrepancy does not affect the choice of tentative models by RG^2.

Table 6.1 RX^2 for Heart Disease Data (Log-Linear)

ORDERING	DF	H3	P-VALUE	H2	P-VALUE	H1	P-VALUE
A	1.0000	1.0E-10	1.0000	2.2E-13	1.0000	986.5	0
B	3.0000	3.1E-10	1.0000	1.7E-10	1.0000	82.5064	0
C	3.0000	9.7E-10	1.0000	3.0E-10	1.0000	310.8	0
AB	3.0000	4.1E-11	1.0000	35.0285	1.2E-07	72.2129	1.4E-15
AC	3.0000	8.4E-11	1.0000	28.9660	2.3E-06	299.1	0
BC	9.0000	5.1E-10	1.0000	25.2145	.0027428	62.1527	5.1E-10
ABC	9.0000	6.5645	0.6824	12.7957	0.1721	52.0760	4.4E-08
TOTAL	31.0000	6.5645	1.0000	102.0	1.7E-09	1865.3	0

We will now treat the dichotomous variable A (whether or not the residents developed a coronary heart disease) as response variable and the serum cholesterol level (variable B) and systolic blood pressure (variable C) as explanatory variables. The selected model, under this logit approach, was [B] [C]. From table 6.2, it is clear that the RX^2 method would lead us to select the same model.

Table 6.2 RX^2 for Heart Disease Data (Logit)

ORDERING	DF	H2	P-VALUE	H1	P-VALUE
B	3.0000	4.0E-11	1.0000	35.0285	1.2E-07
C	3.0000	1.7E-10	1.0000	23.3741	3.4E-05
BC	9.0000	6.5645	0.6824	6.3756	0.7018
TOTAL	15.0000	6.5645	0.9686	64.7781	3.7E-08

Example 7 (3-dim): SEX OF RESPONDENT, DEGREE OF HUSBAND, AND DEGREE OF WIFE (Haberman, 1978, p. 227)

This example deals with data on married subjects in a 1974 General Social Survey, cross-classified by sex of respondent, highest degree attained by husband, and highest degree attained by wife. The levels of each variable are

A: sex of respondent (male, female)
B: husband's highest degree (less than high school diploma, high school diploma or junior college degree, bachelor's degree, graduate degree),
C: wife's highest degree (less than high school diploma, high school diploma or junior college degree, at least bachelor's degree).

The data under the lexicographic arrangement are as follows:

Observed Counts
Cell (i,j,k), x_{ijk}

1,1,1	135	1,3,1	4	2,1,1	124	2,3,1	1
1,1,2	60	1,3,2	35	2,1,2	63	2,3,2	24
1,1,3	1	1,3,3	12	2,1,3	1	2,3,3	26
1,2,1	43	1,4,1	2	2,2,1	39	2,4,1	0
1,2,2	151	1,4,2	24	2,2,2	219	2,4,2	17
1,2,3	19	1,4,3	23	2,2,3	18	2,4,3	14

Haberman (1978) used this example to illustrate use of residual analysis for checking deviations from the model [A] [BC]. It was observed that the behaviour of residuals did not support this model. Using RX^2, we see from Table 7 that at 1%, the initial model will be [A] [BC] and at 5%, it will be [BC] [AB]. It is assumed here that the main effects are necessarily included. Thus the results based on RX^2 are consistent with the earlier analysis. It may be noted that RG^2 method leads to the same ranking and conclusion as those obtained under RX^2.

Table 7 RX^2 for Data on Sex, Degree of Husband, Degree of Wife (Log-Linear)

ORDERING	DF	H3	P-VALUE	H2	P-VALUE	H1	P-VALUE
B	3.0000	2.5E-09	1.0000	1.7E-10	1.0000	474.4	0
C	2.0000	2.9E-11	1.0000	1.5E-09	1.0000	326.3	0
A	1.0000	9.4E-12	1.0000	1.6E-10	1.0000	1.2976	0.2546
BC	6.0000	1.5E-10	1.0000	498.8	0	819.7	0
AB	3.0000	1.2E-11	1.0000	11.0491	0.0115	15.2218	.0016366
AC	2.0000	9.1E-12	1.0000	4.7348	0.0937	7.8730	0.0195
ABC	6.0000	11.3499	0.0781	30.7235	2.9E-05	35.1071	4.1E-06
TOTAL	23.0000	11.3499	0.9793	545.3	0	1679.9	0

Example 8 (4-dim): DETERGENT PREFERENCE DATA (Fienberg, 1977, p. 59 and p. 79)

Consider the four-dimensional table of counts based on a sample of 1008 consumers cross-classified according to

A: the softness of the laundry water used (soft, medium, hard)
B: the previous use of detergent brand M (previous user or not)
C: the temperature of the laundry water used (high, low)
D: the preference for detergent brand X over brand M in a consumer blind trial (X or M)

The data under the lexicographic arrangement are as follows:

Observed Counts
Cell (i,j,k), x_{ijk}

1,1,1,1	19	2,1,1,1	23	3,1,1,1	24
1,1,1,2	29	2,1,1,2	47	3,1,1,2	43
1,1,2,1	57	2,1,2,1	47	3,1,2,1	37
1,1,2,2	49	2,1,2,2	55	3,1,2,2	52
1,2,1,1	29	2,2,1,1	33	3,2,1,1	42
1,2,1,2	27	2,2,1,2	23	3,2,1,2	30
1,2,2,1	63	2,2,2,1	66	3,2,2,1	68
1,2,2,2	53	2,2,2,2	50	3,2,2,2	42

The model selected by Goodman's forward or backward stepwise procedure (see Fienberg, 1977) at 5% level was found to be [AC] [BC] [CD]. It can be seen from table 8.1 that the same model would be suggested by the RX^2 method using 5% level of significance. The optimal ranking obtained with the RG^2 method was C B A D BD CD AC BC AB AD ABD BCD ABC ACD ABCD. It is seen that the RG^2 method also leads to the same model.

Table 8.1 RX^2 for Detergent Preference Data (Log-Linear)

ORDERING	DF	H4	P-VALUE	H3	P-VALUE	H2	P-VALUE	H1	P-VALUE
C	1.0000	9.0E-34	1.0000	1.4E-10	1.0000	2.1E-11	1.0000	72.3214	1.4E-17
B	1.0000	3.5E-33	1.0000	7.9E-12	1.0000	1.1E-11	1.0000	1.9206	0.1658
A	2.0000	1.4E-31	1.0000	2.4E-11	1.0000	1.1E-10	1.0000	0.5000	0.7788
D	1.0000	4.0E-34	1.0000	8.0E-12	1.0000	9.8E-12	1.0000	0.0635	0.8011
BD	1.0000	3.1E-33	1.0000	2.2E-11	1.0000	20.5122	5.9E-06	20.5714	5.7E-06
CD	1.0000	1.7E-32	1.0000	1.6E-10	1.0000	4.3583	0.0368	4.3214	0.0376
AC	2.0000	3.3E-34	1.0000	9.8E-12	1.0000	6.0822	0.0478	5.0238	0.0811
BC	1.0000	1.3E-33	1.0000	1.4E-11	1.0000	1.2535	0.2629	2.0992	0.1474
AB	2.0000	4.3E-32	1.0000	9.1E-12	1.0000	1.0753	0.5841	1.0556	0.5899
AD	2.0000	9.0E-33	1.0000	1.5E-11	1.0000	0.3953	0.8207	0.3889	0.8233
ABD	2.0000	5.3E-33	1.0000	5.3400	0.0693	5.2334	0.0730	5.4524	0.0655
BCD	1.0000	8.3E-33	1.0000	2.2470	0.1339	2.8546	0.0911	0.0992	0.7528
ABC	2.0000	7.3E-34	1.0000	1.3548	0.5079	1.6081	0.4475	1.0079	0.6041
ACD	2.0000	7.2E-10	1.0000	0.1953	0.9070	0.0871	0.9574	0.0952	0.9535
ABCD	2.0000	0.7379	0.6915	0.7336	0.6929	0.4423	0.8016	0.7937	0.6725
TOTAL	23.0000	0.7379	1.0000	9.8706	0.9920	43.9023	.0054031	115.7	2.4E-14

We now consider the detergent preference data but this time we view the variable preference (variable D) as the response variable (dichotomous response) and softness (variable A), use (variable B), and temperature (variable C) as explanatory variables. The simplest logit model was found to be [B], (see Fienberg, 1977, p. 79). Table 8.2 shows that the RX^2 method would lead us to models [B], and [B] [C] which are clearly consistent with the one obtained earlier.

Table 8.2 RX^2 for Detergent Preference Data (Logit)

ORDERING	DF	H3	P-VALUE	H2	P-VALUE	H1	P-VALUE
B	1.0000	4.2E-34	1.0000	8.1E-11	1.0000	20.5121	5.9E-06
C	1.0000	5.0E-33	1.0000	3.5E-11	1.0000	3.7216	0.0537
A	2.0000	1.6E-31	1.0000	6.3E-11	1.0000	0.2105	0.9001
AB	2.0000	2.8E-33	1.0000	5.1238	0.0772	5.0049	0.0819
BC	1.0000	1.4E-32	1.0000	2.1949	0.1385	2.1100	0.1463
AC	2.0000	7.2E-10	1.0000	0.1606	0.9229	0.1896	0.9095
ABC	2.0000	0.7379	0.6915	0.7447	0.6891	0.7206	0.6975
TOTAL	11.0000	0.7379	1.0000	8.2240	0.6931	32.4694	6.4E-04

Example 9 (4-dim): SYMPTOMS OF PSYCHIATRIC PATIENTS (Benedetti and Brown, 1978).

We consider the data on symptoms of patients receiving psychiatric treatment. We have a four-dimensional table (2x2x2x2) of counts based on a sample size of 362 patients. The four indices of the variables are

A: validity (psychasthenic, energetic),
B: solidity (hysteric, rigid),
C: stability (extroverted, introverted),
D: depression (no, yes).

The data under the lexicographic arrangement are as follows:

Observed Counts
Cell (i,j,k), x_{ijk}

1,1,1,1	12	2,1,1,1	47
1,1,1,2	16	2,1,1,2	14
1,1,2,1	27	2,1,2,1	46
1,1,2,2	32	2,1,2,2	9
1,2,1,1	8	2,2,1,1	14
1,2,1,2	22	2,2,1,2	23
1,2,2,1	22	2,2,2,1	25
1,2,2,2	30	2,2,2,2	15

The model selected by Benedetti and Brown is [ABD] [AC] [CD]. From table 9 we see that the initial model under the RX^2 method would be [ABD] [AC]. The next possible term for inclusion in the two factor effects would be CD. Thus the results are consistent with those of the earlier analysis.

Table 9 RX^2 for Data on Symptoms of Psychiatric Patients (Log-Linear)

ORDERING	DF	H4	P-VALUE	H3	P-VALUE	H2	P-VALUE	H1	P-VALUE
C	1.0000	4.9E-15	1.0000	1.9E-11	1.0000	2.5E-09	1.0000	6.9061	.0085903
B	1.0000	1.5E-14	1.0000	3.1E-10	1.0000	2.7E-09	1.0000	5.3481	0.0207
D	1.0000	2.6E-14	1.0000	3.7E-11	1.0000	1.8E-09	1.0000	4.4199	0.0355
A	1.0000	3.6E-15	1.0000	7.4E-11	1.0000	2.7E-09	1.0000	1.5912	0.2072
AD	1.0000	9.7E-13	1.0000	3.2E-11	1.0000	27.7236	1.4E-07	28.7403	8.3E-08
BD	1.0000	1.6E-12	1.0000	1.5E-11	1.0000	16.8893	4.0E-05	18.5746	1.6E-05
AC	1.0000	3.2E-12	1.0000	4.7E-12	1.0000	9.9513	.0016074	8.6630	.0032474
AB	1.0000	1.2E-12	1.0000	4.0E-14	1.0000	2.7206	0.0991	3.1934	0.0739
CD	1.0000	3.5E-12	1.0000	8.7E-12	1.0000	1.4403	0.2301	2.1657	0.1411
BC	1.0000	5.7E-12	1.0000	2.2E-11	1.0000	0.1055	0.7453	3.6E-32	1.0000
ABD	1.0000	4.3E-14	1.0000	5.2120	0.0224	3.5036	0.0612	8.6630	.0032474
BCD	1.0000	6.9E-14	1.0000	2.3128	0.1283	3.6238	0.0570	1.3370	0.2476
ABC	1.0000	3.2E-15	1.0000	0.4670	0.4944	0.6533	0.4189	0.8950	0.3441
ACD	1.0000	1.0E-14	1.0000	0.3982	0.5280	0.2306	0.6311	0.8950	0.3441
ABCD	1.0000	1.2E-05	0.9973	6.8E-04	0.9792	0.4292	0.5124	0.1768	0.6741
TOTAL	15.0000	1.2E-05	1.0000	8.3907	0.9072	67.2711	1.4E-08	91.5691	5.1E-13

Example 10 (4-dim): YEAR OF SURVEY, SEX, EDUCATION, AND PLACE (Haberman, 1978, p. 255)

We now look at a four-dimensional contingency table of subjects in General Social Surveys in which the respondents were asked if they agreed with the following statement: "Women should take care of running their homes and leave running of the country up to men". The variables under consideration are

A: year of survey (1974, 1975),
B: sex of respondents (male, female),
C: education of respondent (less than 9 years, between 9 and 12 years, more than 12 years),
D: response (agree, disagree).

The data under the lexicographic arrangement are as follows:

Observed Counts
Cell (i,j,k,l), x_{ijkl}

1,1,1,1	89	1,2,1,1	83	2,1,1,1	72	2,2,1,1	86
1,1,1,2	43	1,2,1,2	29	2,1,1,2	47	2,2,1,2	38
1,1,2,1	102	1,2,2,1	152	2,1,2,1	110	2,2,2,1	173
1,1,2,2	182	1,2,2,2	284	2,1,2,2	196	2,2,2,2	283
1,1,3,1	48	1,2,3,1	33	2,1,3,1	44	2,2,3,1	28
1,1,3,2	193	1,2,3,2	190	2,1,3,2	179	2,2,3,2	187

The marginal table AB is fixed under the sampling procedure. Two possible nested models [AB] [BC] [CD] and [AB] [BCD] were suggested by Haberman (1978) although the latter one seemed preferable. From table 10, RX^2 method leads us to select the initial model as [AB], [BCD]. The term AB is included because of the sampling constraint. The RX^2 method can be modified, although not done in this report, so that the fixed marginal AB is accorded the priority rank among all the two factor effects.

Table 10 RX^2 for Data on Year of Survey, Sex, Education, and Place (Log-Linear)

ORDERING	DF	H4	P-VALUE	H3	P-VALUE	H2	P-VALUE	H1	P-VALUE
C	2.0000	3.5E-12	1.0000	9.0E-10	1.0000	1.7E-10	1.0000	522.0	0
D	1.0000	3.0E-12	1.0000	3.8E-11	1.0000	1.5E-10	1.0000	240.5	0
B	1.0000	8.2E-12	1.0000	3.3E-10	1.0000	2.1E-10	1.0000	23.7273	1.1E-06
A	1.0000	1.2E-11	1.0000	2.3E-10	1.0000	1.6E-11	1.0000	0.0784	0.7795
CD	2.0000	4.3E-10	1.0000	7.6E-11	1.0000	357.0	0	335.9	0
BC	2.0000	4.2E-12	1.0000	1.8E-10	1.0000	39.3506	2.9E-09	72.5162	1.7E-16
AC	2.0000	1.2E-11	1.0000	9.3E-10	1.0000	1.8635	0.3939	2.4723	0.2905
AB	1.0000	1.2E-11	1.0000	8.4E-10	1.0000	0.3515	0.5533	0.3793	0.5380
BD	1.0000	1.2E-11	1.0000	2.6E-10	1.0000	0.0114	0.9150	2.2853	0.1306
AD	1.0000	5.1E-11	1.0000	1.9E-10	1.0000	6.8E-04	0.9791	.0031348	0.9554
BCD	2.0000	9.0E-10	1.0000	8.6721	0.0131	0.2725	0.8726	6.1087	0.0472
ACD	2.0000	1.5E-09	1.0000	2.3807	0.3041	2.4604	0.2922	1.0930	0.5790
ABC	2.0000	1.9E-11	1.0000	1.2517	0.5348	1.2017	0.5483	0.3824	0.8259
ABD	1.0000	9.8E-11	1.0000	0.1727	0.6777	0.6275	0.4283	0.3347	0.5629
ABCD	2.0000	0.4004	0.8186	0.3656	0.8329	1.3120	0.5189	0.8701	0.6472
TOTAL	23.0000	0.4004	1.0000	12.8428	0.9553	404.4	0	1208.6	0

Example 11 (4-dim): YEAR OF SURVEY, RELIGION, EDUCATION, AND ATTITUDE TOWARD NONTHERAPEUTIC ABORTION (Haberman, 1978, p. 262)

This example deals with information concerning factors associated with attitudes toward nontherapeutic abortions and temporal changes of these attitudes. This is an example of a 3x3x3x3 contingency table with the following variables

A: year of survey (1972, 1973, 1974),
B: religion of respondent (north protestant, south protestant, catholic),
C: ducation of respondent (less than 9 years, between 9 and 12 years, more than 12 years),
D: attitude (positive, mixed, negative).

The data under the lexicographic arrangement are as follows:

Observed Counts
Cell (i,j,k,l), x_{ijkl}

1,1,1,1	9	2,1,1,1	17	3,1,1,1	23
1,1,1,2	16	2,1,1,2	17	3,1,1,2	13
1,1,1,3	41	2,1,1,3	42	3,1,1,3	32
1,1,2,1	85	2,1,2,1	102	3,1,2,1	106
1,1,2,2	52	2,1,2,3	38	3,1,2,2	50
1,1,2,3	105	2,1,2,3	84	3,1,2,3	88
1,1,3,1	77	2,1,3,1	88	3,1,3,1	79
1,1,3,2	30	2,1,3,2	15	3,1,3,2	21
1,1,3,3	38	2,1,3,3	31	3,1,3,3	31
1,2,1,1	8	2,2,1,1	14	3,2,1,1	5
1,2,1,2	8	2,2,1,2	11	3,2,1,2	15
1,2,1,3	46	2,2,1,3	34	3,2,1,3	37
1,2,2,1	35	2,2,2,1	61	3,2,2,1	38
1,2,2,2	29	2,2,2,2	30	3,2,2,2	39
1,2,2,3	54	2,2,2,3	59	3,2,2,3	54
1,2,3,1	37	2,2,3,1	49	3,2,3,1	52
1,2,3,2	15	2,2,3,2	11	3,2,3,2	12
1,2,3,3	22	2,2,3,3	19	3,2,3,3	32
1,3,1,1	11	2,3,1,1	6	3,3,1,1	8
1,3,1,2	14	2,3,1,2	16	3,3,1,2	10
1,3,1,3	38	2,3,1,3	26	3,3,1,3	24
1,3,2,1	47	2,3,2,1	60	3,3,2,1	65
1,3,2,2	35	2,3,2,2	29	3,3,2,2	39
1,3,2,3	115	2,3,2,3	108	3,3,2,3	89
1,3,3,1	25	2,3,3,1	31	3,3,3,1	37
1,3,3,2	21	2,3,3,2	18	3,3,3,2	18
1,3,3,3	42	2,3,3,3	50	3,3,3,3	43

The total sample size was 3181. The model [AD] [BCD] was selected by Haberman (1978 p. 269). As seen from table 11, the RX^2 method leads us to the selection of the same model as an initial choice.

Table 11 RX^2 for Data on Attitudes toward Nontherapeutic Abortion (Log-Linear)

ORDERING	DF	H4	P-VALUE	H3	P-VALUE	H2	P-VALUE	H1	P-VALUE
B	2.0000	5.2E-10	1.0000	2.1E-10	1.0000	4.6E-11	1.0000	121.5	0
C	2.0000	1.9E-10	1.0000	2.0E-10	1.0000	5.3E-10	1.0000	648.2	0
D	2.0000	3.5E-11	1.0000	3.0E-10	1.0000	9.8E-11	1.0000	292.4	0
A	2.0000	5.6E-10	1.0000	2.6E-10	1.0000	5.0E-10	1.0000	0.0572	0.9718
CD	4.0000	6.4E-10	1.0000	1.8E-10	1.0000	157.6	0	179.0	0
BD	4.0000	4.1E-10	1.0000	1.3E-10	1.0000	70.5947	1.7E-14	97.3952	0
BC	4.0000	2.9E-10	1.0000	3.0E-10	1.0000	22.5478	1.6E-04	62.2949	9.5E-13
AD	4.0000	1.7E-10	1.0000	4.0E-10	1.0000	22.3191	1.7E-04	23.9296	8.3E-05
AB	4.0000	2.1E-10	1.0000	4.6E-10	1.0000	3.2179	0.5220	2.7494	0.6006
AC	4.0000	7.7E-10	1.0000	1.6E-10	1.0000	2.3445	0.6727	1.5838	0.8117
BCD	8.0000	1.0E-09	1.0000	27.2497	6.4E-04	41.3586	1.8E-06	71.8988	2.1E-12
ABC	8.0000	9.2E-10	1.0000	10.6723	0.2210	8.8999	0.3508	8.5338	0.3831
ABD	8.0000	1.6E-10	1.0000	10.0869	0.2590	9.1090	0.3332	9.7787	0.2809
ACD	8.0000	2.2E-10	1.0000	7.6744	0.4659	9.0826	0.3354	13.1569	0.1066
ABCD	16.0000	11.4522	0.7808	11.8963	0.7511	9.6738	0.8831	8.8777	0.9184
TOTAL	80.0000	11.4522	1.0000	67.5796	0.8376	356.7	0	1541.4	0

Example 12 (4-dim): ARMED FORCES QUALIFICATION TEST (Fienberg, 1977, p. 89)

The data for this example deals with a national sample of 2294 young males rejected for military service because of failure to pass the Armed Forces Qualification Test. The data form a four-dimensional cross-classification with the variables having the following levels

A: respondant's education (grammar school, some high school, high school graduate),
B: father's education (grammar school, some high school, high school graduate, not available),
C: age of respondent (less than 22, 22 or more),
D: race of respondent (white, black),

The data under the lexicographic arrangement are as follows:

Observed Counts
Cell (i,j,k,l), x_{ijkl}

1,1,1,1	39	2,1,1,1	29	3,1,1,1	8
1,1,1,2	19	2,1,1,2	40	3,1,1,2	19
1,1,2,1	231	2,1,2,1	115	3,1,2,1	51
1,1,2,2	110	2,1,2,2	133	3,1,2,2	103
1,2,1,1	4	2,2,1,1	8	3,2,1,1	1
1,2,1,2	5	2,2,1,2	17	3,2,1,2	7
1,2,21	17	2,2,2,1	21	3,2,2,1	13
1,2,2,2	18	2,2,2,2	38	3,2,2,2	25
1,3,1,1	11	2,3,1,1	9	3,3,1,1	6
1,3,1,2	2	2,3,1,2	14	3,3,1,2	3
1,3,2,1	18	2,3,2,1	28	3,3,2,1	45
1,3,2,2	11	2,3,2,2	25	3,3,2,2	18
1,4,1,1	48	2,4,1,1	17	3,4,1,1	8
1,4,1,2	49	2,4,1,2	79	3,4,1,2	24
1,4,2,1	197	2,4,2,1	111	3,4,2,1	35
1,4,2,2	178	2,4,2,2	206	3,4,2,2	81

In Fienberg (1977) this data was analysed by viewing the respondent's education as the response variable (polytomous response) and the remaining three variables as explanatory. In this logit approach the three-dimensional marginal table corresponding to B, C and D is treated as fixed. The simplest hierarchical model providing a good fit to the data was found to be [BCD] [ABD] [AC]. When we use the RX^2 method with the appropriate constraint, we get the same model as an initial choice, see table 12.

Table 12 RX^2 for Armed Force Qualification Data (Log-Linear)

ORDERING	DF	H4	P-VALUE	H3	P-VALUE	H2	P-VALUE	H1	P-VALUE
A	2.0000	2.2E-13	1.0000	6.3E-11	1.0000	4.6E-08	1.0000	200.9	0
B	3.0000	6.3E-12	1.0000	4.7E-10	1.0000	3.4E-07	1.0000	1085.4	0
C	1.0000	1.5E-13	1.0000	6.9E-11	1.0000	5.3E-09	0.9999	808.7	0
D	1.0000	2.1E-10	1.0000	8.3E-10	1.0000	3.6E-08	0.9998	10.3383	0.001303
AB	6.0000	2.2E-11	1.0000	5.1E-10	1.0000	93.2515	0	246.4	0
AD	2.0000	9.6E-13	1.0000	7.0E-11	1.0000	101.4	0	105.4	0
BD	3.0000	2.2E-10	1.0000	3.5E-10	1.0000	54.0460	1.1E-11	71.3601	2.2E-15
CD	1.0000	2.9E-11	1.0000	5.1E-11	1.0000	9.3259	.0022594	0.2947	0.5872
AC	2.0000	2.7E-10	1.0000	5.4E-10	1.0000	12.1842	.0022607	62.2258	3.1E-14
BC	3.0000	1.8E-11	1.0000	4.7E-10	1.0000	9.7626	0.0207	420.5	0
BCD	3.0000	7.4E-11	1.0000	5.3834	0.1458	5.1007	0.1646	103.9	0
ABD	6.0000	1.9E-10	1.0000	16.8965	.0096713	24.7111	3.9E-04	54.8293	5.0E-10
ABC	6.0000	1.5E-10	1.0000	9.4123	0.1517	32.8393	1.1E-05	89.2831	4.2E-17
ACD	2.0000	4.7E-11	1.0000	0.1612	0.9226	0.4721	0.7897	1.7036	0.4267
ABCD	6.0000	4.7302	0.5788	5.5358	0.4771	9.3069	0.1570	44.7768	5.2E-08
TOTAL	47.0000	4.7302	1.0000	37.3892	0.8409	352.4	0	3305.9	0

Example 13 (5-dim): JAMAICAN LIZARDS (Bishop, Fienberg, and Holland, 1975, p. 164)

We now look at the data on the structural habitat of Grahami and Opalinus lizards from Whitehouse, Jamaica. The data consist of observed counts for perch height, perch diameter, insolation, and time-of-day categories for both grahami and opalinus lizards. The levels of each of these variables are

A: height (less than 5 feet, 5 feet or more),
B: diameter (2 inches or less, more than 2 inches),
C: insolation (sun, shade),
D: time (early, midday, late).
E: species (grahami, opalinus).

The data under the lexicographic arrangement are as follows:

Observed Counts
Cell (i,j,k,l,m), x_{ijklm}

1,1,1,1,1	20	1,2,1,1,1	8	2,1,1,1,1	13	2,2,1,1,1	6
1,1,1,1,2	2	1,2,1,1,2	3	2,1,1,1,2	0	2,2,1,1,2	0
1,1,1,2,1	8	1,2,1,2,1	4	2,1,1,2,1	8	2,2,1,2,1	0
1,1,1,2,2	1	1,2,1,2,2	1	2,1,1,2,2	0	2,2,1,2,2	0
1,1,1,3,1	4	1,2,1,3,1	5	2,1,1,3,1	12	2,2,1,3,1	1
1,1,1,3,2	4	1,2,1,3,2	3	2,1,1,3,2	0	2,2,1,3,2	1
1,1,2,1,1	34	1,2,2,1,1	17	2,1,2,1,1	31	2,2,2,1,1	12
1,1,2,1,2	11	1,2,2,1,2	15	2,1,2,1,2	5	2,2,2,1,2	1
1,1,2,2,1	69	1,2,2,2,1	60	2,1,2,2,1	55	2,2,2,2,1	21
1,1,2,2,2	20	1,2,2,2,2	32	2,1,2,2,2	4	2,2,2,2,2	5
1,1,2,3,1	18	1,2,2,3,1	8	2,1,2,3,1	13	2,2,2,3,1	4
1,1,2,3,2	10	1,2,2,3,2	8	2,1,2,3,2	3	2,2,2,3,2	4

The model as reported in Bishop, Fienberg, and Holland (1975) was found to be [CD] [ABE]. The RX^2 method from table 13 leads us to select the model [CD] [AE] [BE] [AB] as an initial model using 1% level. At 5%, the terms CE and DE should also be added. However, the model [CD] [AE] [BE] [AB] can be accepted because its P-value is seen to be **0.0796.** Some diagnostic analysis can be done to choose an appropriate model.

Table 13 Data on Jamaican Lizards (Log-Linear)

ORDERING	DF	H5	P-VALUE	H4	P-VALUE	H3	P-VALUE	H2	P-VALUE	H1	P-VALUE
C	1.0000	5.4E-14	1.0000	4.8E-11	1.0000	4.8E-12	1.0000	3.4E-11	1.0000	224.7	0
D	2.0000	1.0E-14	1.0000	2.3E-11	1.0000	9.8E-12	1.0000	3.3E-12	1.0000	96.8085	0
E	1.0000	1.3E-13	1.0000	7.7E-12	1.0000	5.3E-12	1.0000	1.6E-11	1.0000	157.5	0
A	1.0000	5.3E-15	1.0000	5.0E-11	1.0000	4.7E-11	1.0000	1.9E-11	1.0000	48.8582	2.8E-12
B	1.0000	1.2E-13	1.0000	2.4E-12	1.0000	1.1E-10	1.0000	2.5E-12	1.0000	28.1489	1.1E-07
CD	2.0000	2.9E-14	1.0000	2.0E-10	1.0000	2.1E-12	1.0000	45.7326	1.2E-10	128.8	0
AE	1.0000	2.1E-13	1.0000	1.3E-11	1.0000	6.9E-12	1.0000	24.6689	6.8E-07	0.1135	0.7362
BE	1.0000	6.6E-15	1.0000	1.8E-10	1.0000	3.4E-11	1.0000	18.8929	1.4E-05	40.9645	1.6E-10
AB	1.0000	4.8E-13	1.0000	2.4E-11	1.0000	2.8E-11	1.0000	16.2150	5.7E-05	4.7943	0.0286
CE	1.0000	8.8E-14	1.0000	7.2E-11	1.0000	8.0E-13	1.0000	5.9351	0.0148	39.8936	2.7E-10
DE	2.0000	5.3E-14	1.0000	1.0E-10	1.0000	4.9E-11	1.0000	6.7757	0.0338	45.1206	1.6E-10
BC	1.0000	2.0E-15	1.0000	6.8E-11	1.0000	5.9E-11	1.0000	3.4880	0.0618	3.7518	0.0528
BD	2.0000	3.4E-14	1.0000	3.5E-11	1.0000	7.6E-12	1.0000	3.7278	0.1551	1.5319	0.4649
AD	2.0000	1.1E-12	1.0000	5.5E-11	1.0000	1.6E-11	1.0000	2.3164	0.3141	18.4397	9.9E-05
AC	1.0000	3.8E-14	1.0000	3.4E-12	1.0000	2.3E-11	1.0000	0.9568	0.3280	26.3901	2.8E-07
ACE	1.0000	2.1E-13	1.0000	5.2E-11	1.0000	1.8332	0.1757	0.2298	0.6317	1.3E-28	1.0000
BCD	2.0000	1.9E-13	1.0000	6.8E-11	1.0000	2.4216	0.2980	0.3502	0.8394	0.2270	0.8927
ACD	2.0000	4.4E-15	1.0000	1.1E-10	1.0000	1.1784	0.5548	1.6986	0.4277	19.4610	5.9E-05
CDE	2.0000	1.2E-13	1.0000	9.8E-11	1.0000	1.6410	0.4402	3.3808	0.1844	46.7660	7.0E-11
ABD	2.0000	1.4E-11	1.0000	4.6E-12	1.0000	1.2523	0.5347	0.9883	0.6101	4.0993	0.1288
ABC	1.0000	6.0E-13	1.0000	8.4E-11	1.0000	0.5635	0.4529	0.1591	0.6900	1.8156	0.1778
ABE	1.0000	2.8E-12	1.0000	1.2E-10	1.0000	0.2434	0.6218	.0083271	0.9273	0.8582	0.3543
ADE	2.0000	7.5E-13	1.0000	1.4E-13	1.0000	0.9676	0.6164	0.1818	0.9131	1.7589	0.4150
BCE	1.0000	7.5E-13	1.0000	1.8E-11	1.0000	0.0555	0.8137	0.0127	0.9104	8.1986	.0041923
BDE	2.0000	2.8E-12	1.0000	2.7E-12	1.0000	0.0211	0.9895	0.0684	0.9664	4.4823	0.1063
ABDE	2.0000	1.0E-06	1.0000	6.3570	0.0416	6.4631	0.0395	4.3965	0.1110	2.9929	0.2239
ABCE	1.0000	1.8E-06	0.9989	1.1318	0.2874	0.7210	0.3958	1.3395	0.2471	.0070922	0.9329
ABCD	2.0000	1.6E-12	1.0000	3.0387	0.2189	2.4997	0.2865	6.6460	0.0360	5.7163	0.0574
BCDE	2.0000	1.3E-06	1.0000	0.6577	0.7197	0.2443	0.8850	1.2681	0.5305	3.8014	0.1495
ACDE	2.0000	2.6E-22	1.0000	0.3239	0.8505	0.1316	0.9363	6.2474	0.0440	2.0426	0.3601
ABCDE	2.0000	6.3E-13	1.0000	0.3498	0.8395	0.7042	0.7032	2.1917	0.3342	1.2908	0.5245
TOTAL	47.0000	4.1E-06	1.0000	11.8590	1.0000	20.9416	0.9996	157.9	6.6E-14	969.3	0

Example 14 (5-dim): CANCER KNOWLEDGE DATA (Fienberg, 1977, p. 73)

Here we have a five-dimensional table of counts based on a sample of 1729 individuals who were cross-classified according to whether they

A: read newspapers (yes, no),
B: listen to the radio (yes, no),
C: do "solid" reading (yes, no),
D: attend lecture (yes, no),
E: have good or poor knowledge regarding cancer.

The data under the lexicographic arrangement are as follows:

Observed Counts
Cell (i,j,k,l,m), x_{ijklm}

1,1,1,1,1	23	1,2,1,1,1	27	2,1,1,1,1	1	2,2,1,1,1	3
1,1,1,1,2	8	1,2,1,1,2	18	2,1,1,1,2	3	2,2,1,1,2	8
1,1,1,2,1	102	1,2,1,2,1	201	2,1,1,2,1	16	2,2,1,2,1	67
1,1,1,2,2	67	1,2,1,2,2	177	2,1,1,2,2	16	2,2,1,2,2	83
1,1,2,1,1	8	1,2,2,1,1	7	2,1,2,1,1	4	2,2,2,1,1	2
1,1,2,1,2	4	1,2,2,1,2	6	2,1,2,1,2	3	2,2,2,1,2	10
1,1,2,2,1	35	1,2,2,2,1	75	2,1,2,2,1	13	2,2,2,2,1	84
1,1,2,2,2	59	1,2,2,2,2	156	2,1,2,2,2	50	2,2,2,2,2	393

The model selected in Fienberg (1977) is [AB] [AC] [AE] [BD] [BD] [CDE]. From table 14.1, the model to be selected for further investigation would be the one with all two-factor effects and possibly CDE effect. We see that the model selected using the RX^2 method is consistent with the earlier analysis except for the terms BC and AD. This should be investigated further. The optimal ranking obtained with the RG^2 method was D, B, E, A, C, AC, CE, AB, AE, CD, BD, BE, DE, AD, BC, CDE, ADE, ACE, BCD, ABD, BDE, ABC, ACD, ABE, BCE, ACDE, BCDE, ABCD, ABCE, ABDE, ABCDE. Inspite of some discrepancies between RX^2 and RG^2 in the two-factor terms and main effects, the two methods lead to the same conclusion.

Table 14.1 Cancer Knowledge Data (Log-Linear)

ORDERING	DF	H5	P-VALUE	H4	P-VALUE	H3	P-VALUE	H2	P-VALUE	H1	P-VALUE
B	1.0000	1.7E-10	1.0000	3.2E-11	1.0000	7.4E-10	1.0000	1.0E-09	1.0000	473.7	0
D	1.0000	6.7E-10	1.0000	1.9E-10	1.0000	1.1E-09	1.0000	8.3E-13	1.0000	1231.2	0
E	1.0000	3.5E-10	1.0000	1.5E-10	1.0000	1.9E-11	1.0000	1.2E-12	1.0000	89.3285	0
A	1.0000	1.2E-11	1.0000	5.1E-12	1.0000	4.3E-11	1.0000	1.6E-12	1.0000	27.2348	1.8E-07
C	1.0000	1.8E-10	1.0000	2.3E-10	1.0000	1.3E-10	1.0000	1.8E-10	1.0000	4.5813	0.0323
AC	1.0000	7.1E-10	1.0000	9.6E-12	1.0000	1.1E-09	1.0000	246.0	0	233.2	0
AE	1.0000	8.2E-10	1.0000	6.3E-11	1.0000	5.2E-11	1.0000	103.3	0	74.5408	0
CE	1.0000	1.1E-12	1.0000	3.5E-12	1.0000	1.6E-11	1.0000	148.5	0	152.2	0
AB	1.0000	3.6E-14	1.0000	4.9E-12	1.0000	2.5E-10	1.0000	71.1937	2.8E-17	19.3690	1.1E-05
BE	1.0000	3.5E-11	1.0000	2.9E-11	1.0000	3.1E-10	1.0000	24.6477	6.9E-07	82.2030	0
CD	1.0000	4.5E-11	1.0000	8.1E-12	1.0000	1.5E-10	1.0000	23.4476	1.3E-06	19.3690	1.1E-05
BD	1.0000	1.1E-09	1.0000	1.5E-12	1.0000	1.6E-10	1.0000	21.0975	4.4E-06	418.9	0
BC	1.0000	3.2E-11	1.0000	3.0E-11	1.0000	6.3E-12	1.0000	21.0698	4.4E-06	25.2637	5.0E-07
AD	1.0000	4.6E-10	1.0000	1.5E-11	1.0000	9.1E-13	1.0000	20.4549	6.1E-06	3.9844	0.0459
DE	1.0000	3.2E-12	1.0000	1.6E-11	1.0000	2.5E-11	1.0000	17.6835	2.6E-05	103.5	0
CDE	1.0000	4.3E-13	1.0000	1.6E-11	1.0000	4.1290	0.0422	0.3524	0.5528	130.5	0
ADE	1.0000	4.1E-11	1.0000	1.0E-11	1.0000	2.9896	0.0838	3.6577	0.0558	43.1053	5.2E-11
ACE	1.0000	2.4E-10	1.0000	3.5E-12	1.0000	3.1011	0.0782	1.1752	0.2783	12.4980	4.1E-04
BCD	1.0000	1.9E-11	1.0000	3.1E-12	1.0000	2.6598	0.1029	.0077004	0.9301	33.0370	9.0E-09
ABC	1.0000	5.2E-10	1.0000	2.0E-13	1.0000	1.5769	0.2092	2.1728	0.1405	83.0775	0
BDE	1.0000	3.1E-12	1.0000	6.3E-12	1.0000	1.3346	0.2480	7.5294	.0060701	64.9075	7.8E-16
ABD	1.0000	4.0E-12	1.0000	5.8E-12	1.0000	1.1633	0.2808	0.1325	0.7159	20.6599	5.5E-06
ACD	1.0000	1.2E-09	1.0000	2.7E-10	1.0000	0.0757	0.7832	0.0123	0.9118	159.4	0
ABE	1.0000	4.1E-11	1.0000	8.7E-15	1.0000	.0038144	0.9508	0.1000	0.7519	28.7617	8.2E-08
BCE	1.0000	3.9E-12	1.0000	8.6E-13	1.0000	2.2E-05	0.9963	0.0721	0.7884	53.8028	2.2E-13
ACDE	1.0000	1.2E-12	1.0000	1.1490	0.2838	1.0546	0.3045	9.9037	.0016495	19.7947	8.6E-06
BCDE	1.0000	5.3E-11	1.0000	0.3140	0.5753	0.7609	0.3831	1.2658	0.2606	5'.7068	6.4E-13
ABCD	1.0000	1.1E-11	1.0000	0.4635	0.4960	0.4521	0.5013	5.4526	0.0195	73.7126	1.4E-17
ABDE	1.0000	2.4E-11	1.0000	0.1896	0.6632	0.1201	0.7289	5.2710	0.0217	27.2348	1.8E-07
ABCE	1.0000	5.1E-11	1.0000	0.1935	0.6600	0.2394	0.6246	14.9016	1.1E-04	27.7392	1.4E-07
ABCDE	1.0000	1.0073	0.3155	1.0059	0.3159	1.5477	0.2135	1.9078	0.1672	23.3667	1.3E-06
TOTAL	31.0000	1.0073	1.0000	3.3155	1.0000	21.2087	0.9063	751.3	0	3811.8	0

NOTE: H5: U(5 FACTOR EFFECTS) = 0
H4: U(4 FACTOR EFFECTS) = 0
H3: U(3 FACTOR EFFECTS) = 0
H2: U(2 FACTOR EFFECTS) = 0

We will now treat variable E (knowledge of cancer) as a response variable (dichotomous response) and condition on the remaining four explanatory variables (A: newspaper, B: radio, C: reading, and D: lectures). The appropriate logit model was found to be [A] [B] [CD], see Fienberg (1977, p. 80). It is clear from table 14.2 that the RX^2 method would permit us to select a logit model having all the main effects (A, B, C, and D) and possibly CD effect. This is consistent with the earlier analysis.

Table 14.2 Cancer Knowledge Data (Logit)

ORDERING	DF	H4	P-VALUE	H3	P-VALUE	H2	P-VALUE	H1	P-VALUE
C	1.0000	1.9E-32	1.0000	1.3E-09	1.0000	1.5E-09	1.0000	148.5	0
A	1.0000	4.2E-09	0.9999	1.2E-10	1.0000	7.4E-11	1.0000	36.1431	1.8E-09
B	1.0000	7.6E-13	1.0000	4.1E-10	1.0000	5.8E-10	1.0000	6.9119	.0085626
D	1.0000	2.6E-11	1.0000	8.2E-11	1.0000	1.3E-10	1.0000	4.6835	0.0305
CD	1.0000	1.5E-09	1.0000	2.7E-10	1.0000	3.3953	0.0654	2.6066	0.1064
AD	1.0000	4.8E-14	1.0000	5.4E-10	1.0000	3.1104	0.0778	4.0337	0.0446
AC	1.0000	3.8E-12	1.0000	7.9E-11	1.0000	2.8444	0.0917	0.8310	0.3620
BD	1.0000	2.1E-09	1.0000	5.3E-11	1.0000	1.3493	0.2454	1.3515	0.2450
BC	1.0000	1.6E-10	1.0000	7.5E-11	1.0000	.0015372	0.9687	0.0559	0.8132
AB	1.0000	1.4E-11	1.0000	2.5E-11	1.0000	1.7E-04	0.9897	0.0188	0.8910
ACD	1.0000	1.5E-10	1.0000	0.9646	0.3260	0.8762	0.3493	0.6050	0.4367
BCD	1.0000	5.7E-11	1.0000	0.4966	0.4810	0.4734	0.4914	0.6280	0.4281
ABC	1.0000	1.5E-13	1.0000	0.1859	0.6664	0.2343	0.6284	0.0916	0.7622
ABD	1.0000	6.7E-11	1.0000	0.1461	0.7023	0.1356	0.7127	0.1243	0.7244
ABCD	1.0000	1.0073	0.3155	1.0477	0.3060	1.1864	0.2761	0.9489	0.3300
TOTAL	15.0000	1.0073	1.0000	2.8410	0.9997	13.6068	0.5555	207.5	0

Example 15 (6-dim): RISK FACTORS FOR CORONARY HEART DISEASE (Edwards and Havránek, 1985)

We now look at information collected at the beginning of a 15 year follow-up study of probable risk factors for coronary thrombosis, comprising data on all men employed in a car factory. We have a six-dimensional contingency table with the following variables

A: smoking (no, yes),
B: strenuous mental work (no, yes)
C: strenuous physical work (no, yes)
D: systolic blood pressure (less than 140, 140 or more),
E: ratio of beta and alpha lipoproteins (less than 3, 3 or more).
F: family anamnesis of coronary heart disease (negative, positive).

The data under the lexicographic arrangement are as follows:

Observed Counts
Cell (i,j,k,l,m), x_{ijklm}

1,1,1,1,1,1	44	1,2,1,1,1,1	112	2,1,1,1,1,1	40	2,2,1,1,1,1	67
1,1,1,1,1,2	5	1,2,1,1,1,2	21	2,1,1,1,1,2	7	2,2,1,1,1,2	9
1,1,1,1,2,1	23	1,2,1,1,2,1	70	2,1,1,1,2,1	32	2,2,1,1,2,1	66
1,1,1,1,2,2	7	1,2,1,1,2,2	14	2,1,1,1,2,2	3	2,2,1,1,2,2	14
1,1,1,2,1,1	35	1,2,1,2,1,1	80	2,1,1,2,1,1	12	2,2,1,2,1,1	33
1,1,1,2,1,2	4	1,2,1,2,1,2	11	2,1,1,2,1,2	3	2,2,1,2,1,2	8
1,1,1,2,2,1	24	1,2,1,2,2,1	73	2,1,1,2,2,1	25	2,2,1,2,2,1	57
1,1,1,2,2,2	4	1,2,1,2,2,2	13	2,1,1,2,2,2	0	2,2,1,2,2,2	11
1,1,2,1,1,1	129	1,2,2,1,1,1	12	2,1,2,1,1,1	145	2,2,2,1,1,1	23
1,1,2,1,1,2	9	1,2,2,1,1,2	1	2,1,2,1,1,2	17	2,2,2,1,1,2	4
1,1,2,1,2,1	50	1,2,2,1,2,1	7	2,1,2,1,2,1	80	2,2,2,1,2,1	13
1,1,2,1,2,2	9	1,2,2,1,2,2	2	2,1,2,1,2,2	16	2,2,2,1,2,2	3
1,1,2,2,1,1	109	1,2,2,2,1,1	7	2,1,2,2,1,1	67	2,2,2,2,1,1	9
1,1,2,2,1,2	14	1,2,2,2,1,2	5	2,1,2,2,1,2	17	2,2,2,2,1,2	2
1,1,2,2,2,1	51	1,2,2,2,2,1	7	2,1,2,2,2,1	63	2,2,2,2,2,1	16
1,1,2,2,2,2	5	1,2,2,2,2,2	4	2,1,2,2,2,2	14	2,2,2,2,2,2	4

The model search procedure used by Edwards and Havránek led to the following two hierarchical models [AC] [AD] [AE] [BC] [DE] [BE] [F] and [AC] [AD] [AE] [BC] [DE] [CE] [F]. With the RX^2 method, it is seen from table 15 (a) and (b) that at 1% level, the initial model would be [BC] [AC] [BE] [AE] [CE] [DE] [AD] [AB] [F]. This corresponds to the union of the two models selected earlier except for the extra term AB. Some simple analysis based on the conditional G^2 statistic can be performed to further simplify the initial model.

Table 15(a) RX^2 for Risk Factors for Coronary Heart Disease Data (Log-Linear)

ORDERING	DF	H6	P-VALUE	H5	P-VALUE	H4	P-VALUE
F	1.0000	5.6E-10	1.0000	1.3E-10	1.0000	1.1E-10	1.0000
B	1.0000	4.9E-10	1.0000	5.3E-10	1.0000	8.5E-11	1.0000
E	1.0000	1.0E-10	1.0000	2.0E-10	1.0000	5.3E-11	1.0000
D	1.0000	7.7E-10	1.0000	2.8E-11	1.0000	1.4E-11	1.0000
A	1.0000	6.3E-12	1.0000	2.2E-13	1.0000	2.8E-12	1.0000
C	1.0000	3.1E-13	1.0000	4.5E-11	1.0000	6.3E-11	1.0000
BC	1.0000	7.7E-12	1.0000	1.1E-11	1.0000	1.0E-11	1.0000
AC	1.0000	2.6E-11	1.0000	2.4E-11	1.0000	1.0E-13	1.0000
BE	1.0000	1.7E-09	1.0000	2.7E-11	1.0000	6.6E-11	1.0000
AE	1.0000	3.3E-10	1.0000	1.3E-10	1.0000	4.0E-11	1.0000
CE	1.0000	1.3E-10	1.0000	5.3E-12	1.0000	6.0E-11	1.0000
DE	1.0000	9.0E-12	1.0000	2.9E-11	1.0000	2.3E-11	1.0000
AD	1.0000	1.1E-12	1.0000	6.3E-10	1.0000	2.1E-10	1.0000
AB	1.0000	3.3E-10	1.0000	2.8E-10	1.0000	3.8E-11	1.0000
BF	1.0000	5.9E-11	1.0000	3.9E-12	1.0000	3.5E-16	1.0000
EF	1.0000	1.3E-10	1.0000	3.5E-10	1.0000	1.3E-10	1.0000
DF	1.0000	7.4E-11	1.0000	8.6E-11	1.0000	5.8E-11	1.0000
AF	1.0000	9.3E-12	1.0000	1.5E-14	1.0000	4.4E-14	1.0000
BD	1.0000	2.5E-10	1.0000	1.9E-11	1.0000	4.8E-11	1.0000
CF	1.0000	5.1E-11	1.0000	1.1E-10	1.0000	1.9E-11	1.0000
CD	1.0000	9.5E-13	1.0000	6.4E-13	1.0000	3.5E-12	1.0000
ABC	1.0000	3.7E-11	1.0000	4.8E-11	1.0000	1.6E-11	1.0000
CDF	1.0000	3.4E-10	1.0000	2.3E-10	1.0000	1.0E-10	1.0000
ADE	1.0000	1.6E-11	1.0000	1.1E-10	1.0000	3.5E-12	1.0000
DEF	1.0000	6.2E-14	1.0000	2.6E-12	1.0000	5.3E-11	1.0000
ACF	1.0000	3.2E-10	1.0000	1.1E-10	1.0000	2.2E-10	1.0000
BDE	1.0000	8.2E-12	1.0000	7.1E-11	1.0000	2.3E-11	1.0000
BDF	1.0000	3.6E-12	1.0000	5.3E-14	1.0000	3.2E-12	1.0000
AEF	1.0000	1.1E-09	1.0000	8.2E-11	1.0000	1.4E-11	1.0000
ADF	1.0000	2.3E-10	1.0000	4.9E-11	1.0000	2.4E-10	1.0000
BCF	1.0000	6.1E-11	1.0000	5.8E-11	1.0000	3.6E-12	1.0000
CEF	1.0000	4.2E-11	1.0000	1.2E-11	1.0000	1.3E-10	1.0000
ABD	1.0000	4.0E-12	1.0000	3.4E-13	1.0000	7.7E-13	1.0000
BCE	1.0000	5.3E-12	1.0000	1.4E-11	1.0000	9.1E-11	1.0000
ABF	1.0000	3.9E-12	1.0000	2.5E-10	1.0000	1.1E-10	1.0000
ACE	1.0000	1.7E-12	1.0000	1.5E-10	1.0000	4.2E-11	1.0000
BEF	1.0000	2.7E-11	1.0000	1.6E-10	1.0000	7.8E-12	1.0000
BCD	1.0000	8.2E-12	1.0000	1.5E-12	1.0000	5.1E-11	1.0000
ACD	1.0000	1.9E-11	1.0000	7.9E-11	1.0000	1.8E-11	1.0000
CDE	1.0000	3.0E-11	1.0000	5.2E-11	1.0000	1.1E-11	1.0000
BDEF	1.0000	1.4E-10	1.0000	3.2E-10	1.0000	1.1597	0.2815
BCDE	1.0000	5.5E-12	1.0000	4.9E-12	1.0000	0.8298	0.3623
BCEF	1.0000	1.6E-12	1.0000	6.4E-11	1.0000	0.6650	0.4148
ABCD	1.0000	5.1E-11	1.0000	7.2E-13	1.0000	0.6666	0.4142
ABDE	1.0000	6.6E-12	1.0000	1.6E-14	1.0000	0.6386	0.4242
ABCE	1.0000	4.4E-13	1.0000	2.1E-11	1.0000	0.4577	0.4987
ACDF	1.0000	3.0E-13	1.0000	3.5E-13	1.0000	0.1791	0.6721
ABDF	1.0000	1.2E-10	1.0000	1.3E-11	1.0000	0.2319	0.6301
ADEF	1.0000	1.2E-12	1.0000	1.2E-11	1.0000	0.1155	0.7340
ACDE	1.0000	9.8E-16	1.0000	5.4E-12	1.0000	0.0645	0.7995
CDEF	1.0000	2.1E-11	1.0000	2.3E-10	1.0000	0.0429	0.8358
BCDF	1.0000	2.1E-12	1.0000	2.1E-11	1.0000	0.0428	0.8361
ABCDF	1.0000	8.8E-12	1.0000	2.6155	0.1058	2.4516	0.1174
ABCEF	1.0000	5.8E-11	1.0000	1.9090	0.1671	1.8424	0.1747
ACDEF	1.0000	1.1E-11	1.0000	2.2068	0.1374	1.9121	0.1667
ABCDE	1.0000	6.6E-11	1.0000	0.6082	0.4354	0.7438	0.3884
BCDEF	1.000	2.2E-14	1.0000	0.1812	0.6703	0.2099	0.6468
ABDEF	1.0000	8.9E-11	1.0000	0.1573	0.6916	0.1133	0.7364
ABCDEF	1.0000	0.2651	0.6066	6.3E-05	0.9937	0.0280	0.8671
TOTAL	63.0000	0.2651	1.0000	7.6781	1.0000	19.6099	1.0000

Table 15(b) RX^2 for Risk Factors for Coronary Heart Disease Data (Log-Linear)

ORDERING	DF	H3	P-VALUE	H2	P-VALUE	H1	P-VALUE
F	1.0000	4.0E-10	1.0000	5.1E-08	0.9998	947.9	0
B	1.0000	5.5E-13	1.0000	3.9E-08	0.9998	44.1200	3.1E-11
E	1.0000	1.1E-11	1.0000	5.5E-08	0.9998	42.8903	5.8E-11
D	1.0000	3.4E-12	1.0000	7.2E-08	0.9998	38.7230	4.9E-10
A	1.0000	1.7E-11	1.0000	4.0E-08	0.9998	3.5638	0.0591
C	1.0000	1.3E-11	1.0000	3.4E-07	0.9995	0.0918	0.7619
BC	1.0000	8.4E-11	1.0000	636.0	0	618.4	0
AC	1.0000	4.9E-10	1.0000	27.4168	1.6E-07	27.4986	1.6E-07
BE	1.0000	1.1E-10	1.0000	17.9498	2.3E-05	26.5296	2.6E-07
AE	1.0000	6.0E-11	1.0000	17.3830	3.1E-05	19.4030	1.1E-05
CE	1.0000	4.8E-11	1.0000	16.6399	4.5E-05	15.8832	6.7E-05
DE	1.0000	2.4E-10	1.0000	12.8267	3.4E-04	19.8159	8.5E-06
AD	1.0000	1.8E-17	1.0000	11.0102	9.1E-04	9.0391	.0026426
AB	1.0000	1.0E-11	1.0000	9.6498	.0018937	7.6920	.0055466
BF	1.0000	1.2E-12	1.0000	4.7757	0.0289	39.3053	3.6E-10
EF	1.0000	4.1E-11	1.0000	3.0276	0.0819	34.7686	3.7E-09
DF	1.0000	2.5E-11	1.0000	1.1268	0.2885	27.0120	2.0E-07
AF	1.0000	2.5E-13	1.0000	1.0713	0.3007	4.3026	0.0381
BD	1.0000	7.4E-12	1.0000	0.5008	0.4792	2.7382	0.0980
CF	1.0000	1.9E-11	1.0000	0.1708	0.6794	.0048886	0.9443
CD	1.0000	4.6E-12	1.0000	0.0952	0.7577	0.1222	0.7266
ABC	1.0000	5.9617	0.0146	3.3802	0.0660	4.4981	0.0339
CDF	1.0000	3.4837	0.0620	3.3561	0.0670	1.0999	0.2943
ADE	1.0000	2.8167	0.0933	2.7512	0.0972	1.7648	0.1840
DEF	1.0000	2.8333	0.0923	2.8898	0.0891	21.5106	3.5E-06
ACF	1.0000	1.9858	0.1588	1.8085	0.1787	7.9527	.0048015
BDE	1.0000	1.6550	0.1983	1.3876	0.2388	0.0657	0.7977
BDF	1.0000	1.5207	0.2175	.0038393	0.9506	2.4383	0.1184
AEF	1.0000	1.1916	0.2750	0.8025	0.3703	15.5139	8.2E-05
ADF	1.0000	0.9832	0.3214	1.0539	0.3046	7.4356	.0063945
BCF	1.0000	0.8397	0.3595	0.1638	0.6857	329.6	0
CEF	1.0000	0.7284	0.3934	0.8235	0.3641	12.3851	4.3E-04
ABD	1.0000	0.6717	0.4125	1.4957	0.2213	4.3026	0.0381
BCE	1.0000	0.5433	0.4611	0.1215	0.7274	22.3840	2.2E-06
ABF	1.0000	0.2661	0.6060	1.0896	0.2966	1.1999	0.2733
ACE	1.0000	0.1191	0.7301	0.0455	0.8312	0.7436	0.3885
BEF	1.0000	0.0943	0.7588	4.8E-04	0.9824	17.0174	3.7E-05
BCD	1.0000	0.0746	0.7847	0.1365	0.7118	10.1950	.0014082
ACD	1.0000	0.0553	0.8142	0.2882	0.5914	1.6431	0.1999
CDE	1.0000	0.0284	0.8661	0.3473	0.5557	.0048886	0.9443
BDEF	1.0000	1.2248	0.2684	1.2055	0.2722	1.4128	0.2346
BCDE	1.0000	1.0139	0.3140	2.2323	0.1352	3.9245	0.0476
BCEF	1.0000	0.2378	0.6258	1.8074	0.1788	19.8159	8.5E-06
ABCD	1.0000	0.7924	0.3734	1.5636	0.2111	1.0043	0.3163
ABDE	1.0000	0.5433	0.4611	1.1308	0.2876	2.2949	0.1298
ABCE	1.0000	0.3631	0.5468	7.2767	.0069854	8.7610	.0030774
ACDF	1.0000	0.1257	0.7229	0.0182	0.8927	0.3960	0.5292
ABDF	1.0000	0.6434	0.4225	0.2412	0.6234	3.2205	0.0727
ADEF	1.0000	0.1043	0.7467	0.1353	0.7130	1.3042	0.2535
ACDE	1.0000	0.0698	0.7916	2.1573	0.1419	2.7382	0.0980
CDEF	1.0000	.0057937	0.9393	0.4289	0.5125	0.4568	0.4991
BCDF	1.0000	0.0739	0.7858	0.0825	0.7740	7.4356	.0063945
ABCDF	1.0000	2.2834	0.1308	2.9388	0.0865	2.7382	0.0980
ABCEF	1.0000	2.0946	0.1478	0.6359	0.4252	3.0554	0.0805
ACDEF	1.0000	1.0054	0.3160	0.4744	0.4910	0.5220	0.4700
ABCDE	1.0000	1.2793	0.2580	.0017907	0.9662	0.0136	0.9072
BCDEF	1.0000	0.1829	0.6689	0.4880	0.4848	4.6980	0.0302
ABDEF	1.0000	0.0358	0.8499	0.1363	0.7120	0.5915	0.4418
ABCDEF	1.0000	0.2078	0.6485	0.3443	0.5574	0.1961	0.6579
TOTAL	63.0000	45.0390	0.9575	809.5	0	2466.7	0

5. CONCLUDING REMARKS

Based on several numerical examples considered in this report, the RX^2 method appears to be quite useful as a screening device for selecting a small set of tentative models. The final selection should be based on optimal tests such as G^2. The method RX^2 has a systematic simple structure and is computationally fast because it involves fairly simple operations. The sample computer program listed here represents only an initial version. It is planned to revise it in future which would allow more flexibility in fixing marginals and would apply to higher dimensional tables. When dealing with categorical survey data, it may be noted that corrections to chi-square such as those of Rao-Scott should also be incorporated in the RX^2 method. This, however, needs further investigation.

ACKNOWLEDGEMENTS

This work was partially supported by a grant from National Sciences and Engineering Research Council of Canada. The work of the second author was done at Statistics Canada during his co-op term from University of Ottawa. The authors would like to thank Christine Larabie and Judy Clarke for their efficient manuscript processing.

REFERENCES

AGRESTI, A. (1984). *Analysis of Ordinal Categorical Data.* New York, John Wiley & Sons.

BENEDETTI, J.K., and Brown, M.B. (1978). Strategies for the selection of log-linear models. *Biometrics*, Vol. 34, pp 680-686.

BISHOP, Y.M., FIENBERG, S.E., and HOLLAND, P.W. (1975). *Discrete Multivariate Analysis.* Cambridge, The MIT Press.

EDWARDS, D. and HAVRÁNEK, T. (1985). A fast procedure for model search in multi-dimensional contingency tables. *Biometrika*, Vol, 72, 339-351.

FIENBERG, S.E. (1977). *The Analysis of Cross-classified Categorical Data.* Cambridge, The MIT Press.

HABERMAN, S.J. (1978). *Analysis of Qualitative Data.* New York, Academic Press.

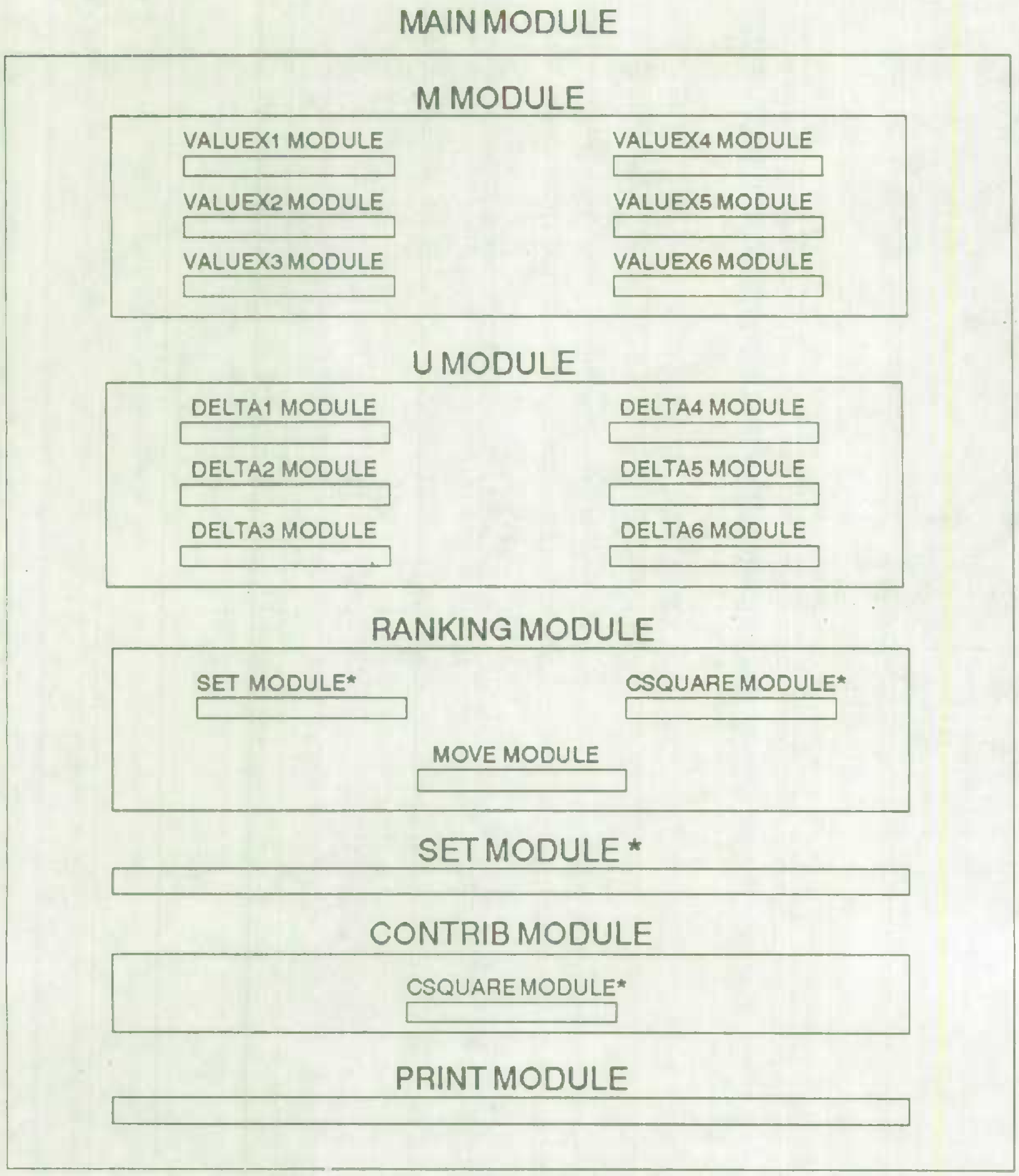
SINGH, A.C. (1988). Model selection for multiway categorical data using partitioning of chi-square: Part I - Theory and methods, *Methodology Branch Working Paper No SSMD-88-031E,* Statistics Canada; another version to appear in *Proc. Soc. Statist. Sec, American Statistical Assocation,* 1988.

APPENDIX

We present a sample software that was developed by the second author to implement the RX^2 method. It is based on the Interactive Matrix Language (IML) of SAS (Statistical Analysis System) software system. Although this particular software was developed to manipulate contingency tables having a maximum of six dimensions, it can be easily modified to handle higher dimensional tables.

The program designed for the RX^2 method consists of 21 modules having various interfaces. We present the hierarchy of these modules and a general program logic flowchart as well as the actual program code in the appendices A.1 to A.10.

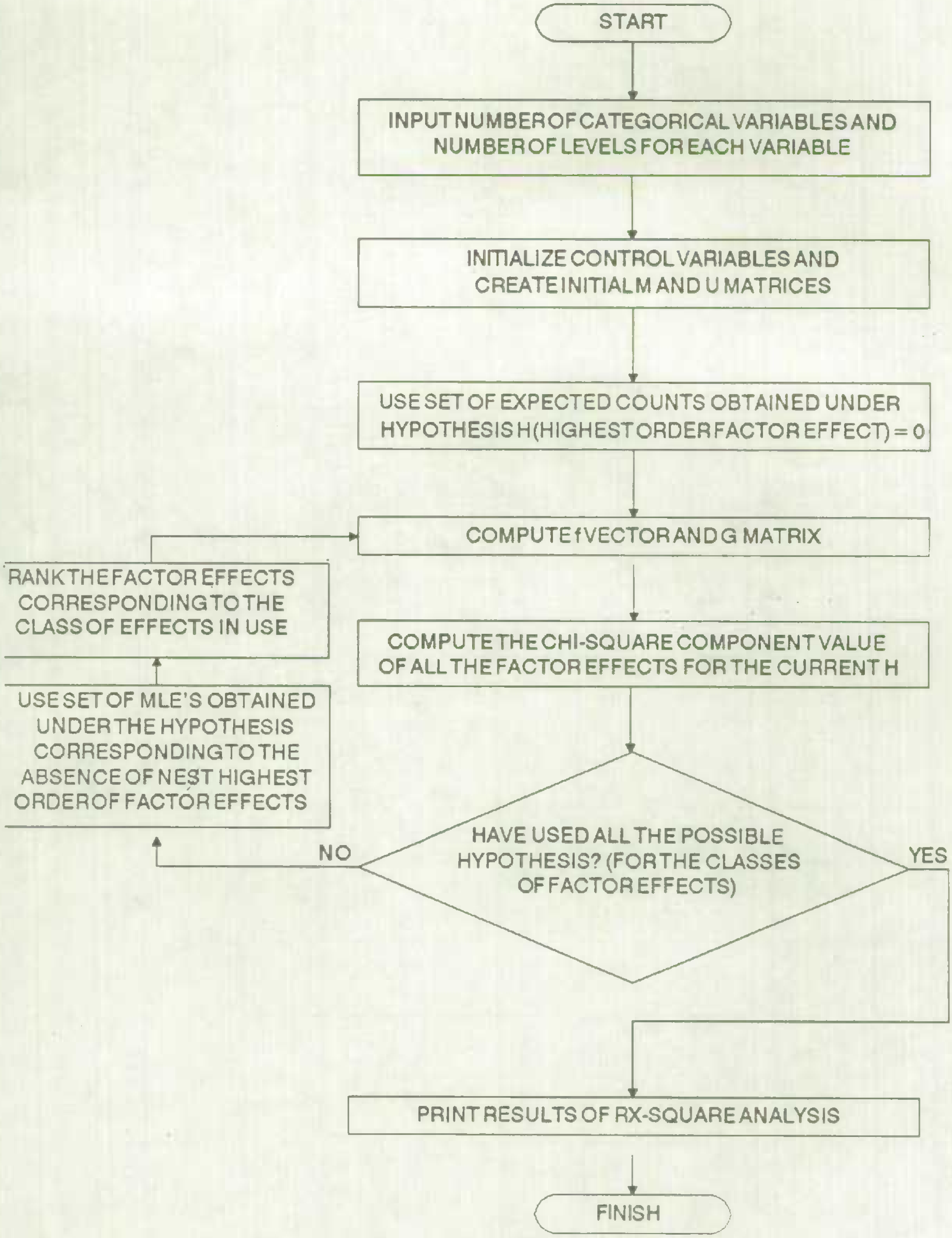
PROGRAM MODULES HIERARCHY



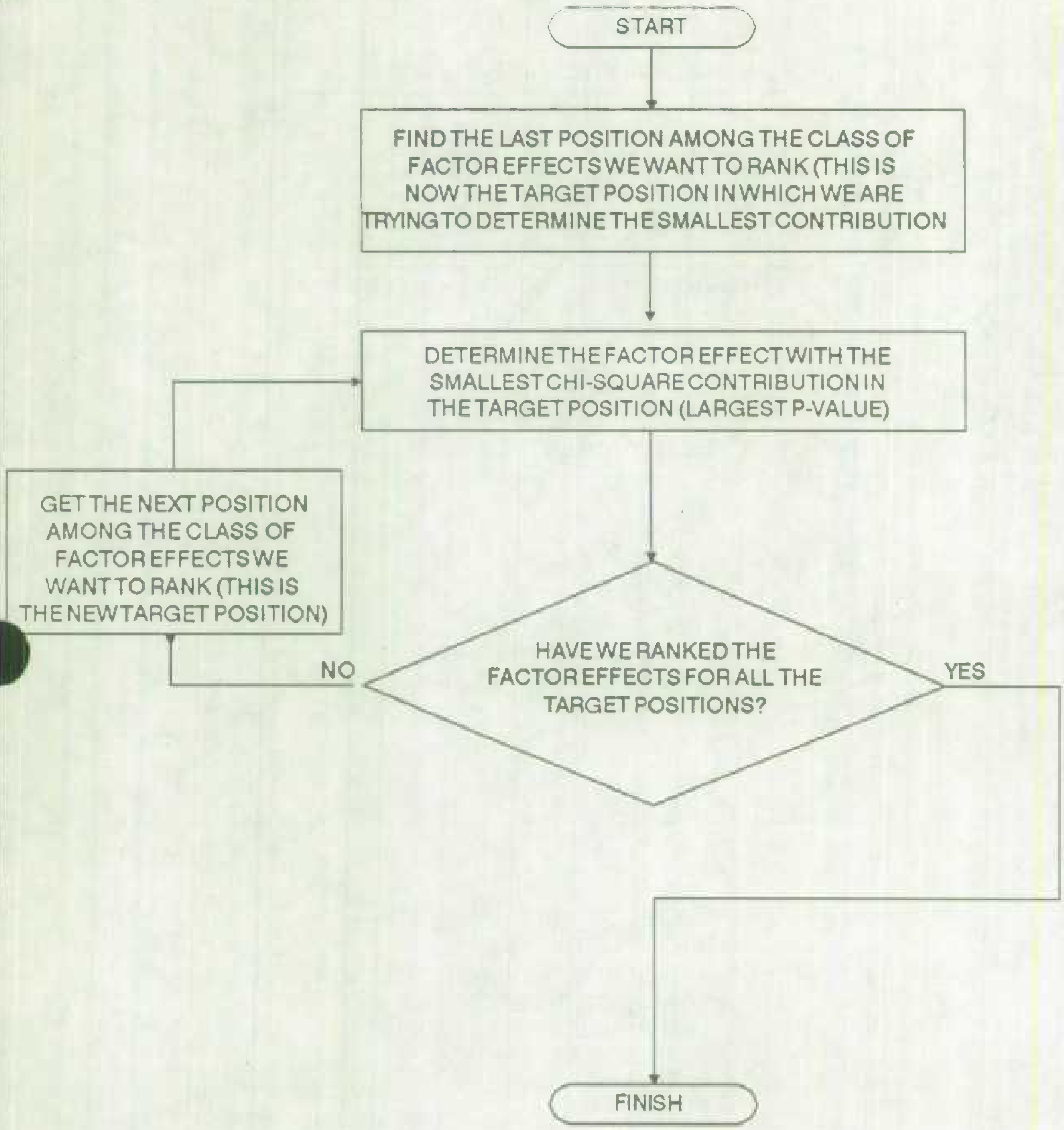
*: INDICATES A REPEATED OCCURENCE

RX-SQUARE METHOD

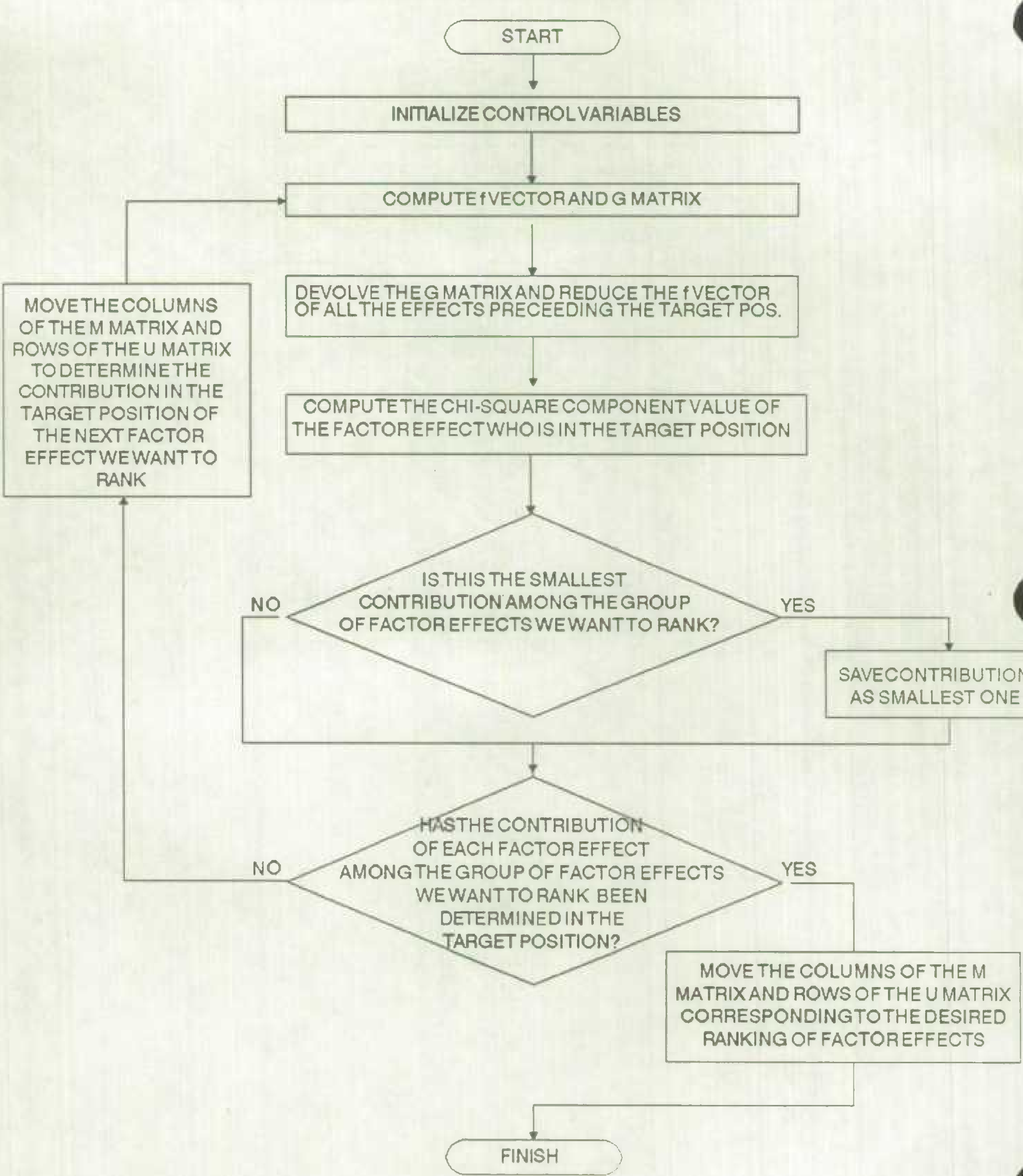
PROGRAM LOGIC FLOWCHART: LEVEL 1



LEVEL 2: RANKING OF THE FACTOR EFFECTS



LEVEL 3: DETERMINATION OF THE SMALLEST CHI-SQUARE CONTRIBUTION



Appendix A.1: Main module

MODULE : MAIN.

PROCESSING : This is the main module for the running of the program. It sets up the initial ordering of the factor effects and ranks them according to their contribution to the total chi-square value.

CALLED BY : none.

MODULES CALLED : M, RANKING, SET, CONTRIB, PRINT and U.

VARIABLES USED :

- A : Number of levels of variable 1.
- B : Number of levels of variable 2.
- C : Number of levels of variable 3.
- D : Number of levels of variable 4.
- E : Number of levels of variable 5.
- F : Number of levels of variable 6.
- COUNTERS: Vector containing the number of factor effects for each class of factor effects.
- DEV : Vector containing the chi-square component value of each factor effect.
- FIXEDMAR: Indicator telling if we are fitting a log-linear model with a fixed marginal total (FIXEDMAR = 'YES' or 'NO')
- M : M matrix.
- NU : Total number of u-parameters remaining to be ordered.
- ORDER : Vector containing the degrees of freedom of each factor effect.
- ORDER2 : Vector containing the ordering of the factor effects.
- ORDERING: Matrix containing the results of the analysis
- POSITION: Column indicator for the position in which we are trying to rank the factor effects (relative to all the factor effects)(for example, if our factor effects are A B C AB AC BC ABC, then ABC is in POSITION=1, BC in POSITION=2, and so on up to A where POSITION=7).
- PVAL : Vector containing the P-value corresponding to the chi-square component values of each factor effect.
- RESPONSE: Indicator for the response variable (for logit models)
- SETUSED : Indicator for the current set of expected counts in use.
- START : Indicators for the position variable
- FINISH :
- U : U matrix.
- UTOT : Total number of u-parameters.
- VARS : Total number of variables.

```
*********************** START OF MODULE MAIN *************************;
START MAIN;

********************************************************************;
*********************** START OF USER INPUT ************************;
********************************************************************;

PRINT 'ARMED FORCES QUALIFICATION DATA',
      '(FIENBERG, P.89)';

*-- NUMBER OF CATEGORICAL VARIABLES ------------------------------*;
  VARS = 4;
*-- TYPE OF MODEL WE WANT TO FIT ---------------------------------*;
  FIXEDMAR = 'YES;
*-- RESPONSE VARIABLE (IF WE WANT TO FIT A LOGLINEAR MODEL WITHOUT ---*;
*-- A FIXED MARGINAL TOTAL, WE GIVE THE LAST CATEGORICAL VARIABLE  ---*;
*--  AS A RESPONSE VARIABLE - FOR NOTATIONAL PURPOSES ONLY) ----------*;
*-- FOR EXAMPLE, FOR A 4 DIMENSIONAL TABLE, RESPONSE = 'D' AND -------*;
*-- FIXEDMAR = 'NO' WHEN WE WANT TO FIT A LOGLINEAR MODEL WITHOUT ----*;
*-- A FIXED MARGINAL TOTAL ---------------------------------------*;
  RESPONSE = 'A';
*-- NUMBER OF LEVELS OF EACH CATEGORICAL VARIABLE -------------------*;
  A = 3;
  B = 4;
  C = 2;
  D = 2;
  E = 0;
  F = 0;

********************************************************************;
***********************   END  OF USER INPUT  **********************;
********************************************************************;

  COUNTERS = J(6,1,0);
  IF VARS = 2 THEN DO;
    UTOT = A * B;
    COUNTERS(|2|) = 1;
    COUNTERS(|1|) = 2;
    ORDER = (A-1) // (B-1) // (A-1)*(B-1);
    ORDER2 = {'A', 'B', 'AB', 'TOTAL'};
    END;
  ELSE IF VARS = 3 THEN DO;
    UTOT = A * B * C;
    COUNTERS(|3|) = 1;
    COUNTERS(|2|) = 3;
    COUNTERS(|1|) = 3;
    IF RESPONSE = 'A' THEN DO;
      FACT1 = (B-1)*(C-1) // (A-1)*(B-1) // (A-1)*(C-1);
      FACT2 = {'BC', 'AB', 'AC'};
      END;
    ELSE IF RESPONSE = 'B' THEN DO;
      FACT1 = (A-1)*(C-1) // (A-1)*(B-1) // (B-1)*(C-1);
      FACT2 = {'AC', 'AB', 'BC'};
      END;
    ELSE IF RESPONSE = 'C' THEN DO;
      FACT1 = (A-1)*(B-1) // (A-1)*(C-1) // (B-1)*(C-1);
      FACT2 = {'AB', 'AC', 'BC'};
```

```
      END;
    ORDER = (A-1) // (B-1) // (C-1) // FACT1 //
            (A-1)*(B-1)*(C-1);
    ORDER2 = {'A', 'B', 'C'} // FACT2 // {'ABC', 'TOTAL'};
    END;
  ELSE IF VARS = 4 THEN DO;
    UTOT = A * B * C * D;
    COUNTERS(|4|) = 1;
    COUNTERS(|3|) = 4;
    COUNTERS(|2|) = 6;
    COUNTERS(|1|) = 4;
    IF RESPONSE = 'A' THEN DO;
      FACT1 = (B-1)*(C-1)*(D-1) // (A-1)*(B-1)*(C-1) //
              (A-1)*(B-1)*(D-1) // (A-1)*(C-1)*(D-1);
      FACT2 = {'BCD', 'ABC', 'ABD', 'ACD'};
      END;
    ELSE IF RESPONSE = 'B' THEN DO;
      FACT1 = (A-1)*(C-1)*(D-1) // (A-1)*(B-1)*(C-1) //
              (A-1)*(B-1)*(D-1) // (B-1)*(C-1)*(D-1);
      FACT2 = {'ACD', 'ABC', 'ABD', 'BCD'};
      END;
    ELSE IF RESPONSE = 'C' THEN DO;
      FACT1 = (A-1)*(B-1)*(D-1) // (A-1)*(B-1)*(C-1) //
              (A-1)*(C-1)*(D-1) // (B-1)*(C-1)*(D-1);
      FACT2 = {'ABD', 'ABC', 'ACD', 'BCD'};
      END;
    ELSE IF RESPONSE = 'D' THEN DO;
      FACT1 = (A-1)*(B-1)*(C-1) // (A-1)*(B-1)*(D-1) //
              (A-1)*(C-1)*(D-1) // (B-1)*(C-1)*(D-1);
      FACT2 = {'ABC', 'ABD', 'ACD', 'BCD'};
      END;
    ORDER = (A-1) // (B-1) // (C-1) // (D-1) //
            (A-1)*(B-1) // (A-1)*(C-1) // (A-1)*(D-1) //
            (B-1)*(C-1) // (B-1)*(D-1) // (C-1)*(D-1) //
            FACT1 //
            (A-1)*(B-1)*(C-1)*(D-1);
    ORDER2 = {'A', 'B', 'C', 'D', 'AB', 'AC', 'AD', 'BC', 'BD', 'CD'} //
             FACT2 // {'ABCD', 'TOTAL'};
    END;
  ELSE IF VARS = 5 THEN DO;
    UTOT = A * B * C * D * E;
    COUNTERS(|5|) = 1;
    COUNTERS(|4|) = 5;
    COUNTERS(|3|) = 10;
    COUNTERS(|2|) = 10;
    COUNTERS(|1|) = 5;
    IF RESPONSE = 'A' THEN DO;
      FACT1 = (B-1)*(C-1)*(D-1)*(E-1) // (A-1)*(B-1)*(C-1)*(D-1) //
              (A-1)*(B-1)*(C-1)*(E-1) // (A-1)*(B-1)*(D-1)*(E-1) //
              (A-1)*(C-1)*(D-1)*(E-1);
      FACT2 = {'BCDE', 'ABCD', 'ABCE', 'ABDE', 'ACDE'};
      END;
    ELSE IF RESPONSE = 'B' THEN DO;
      FACT1 = (A-1)*(C-1)*(D-1)*(E-1) // (A-1)*(B-1)*(C-1)*(D-1) //
              (A-1)*(B-1)*(C-1)*(E-1) // (A-1)*(B-1)*(D-1)*(E-1) //
              (B-1)*(C-1)*(D-1)*(E-1);
      FACT2 = {'ACDE', 'ABCD', 'ABCE', 'ABDE', 'BCDE'};
      END;
```

```
    ELSE IF RESPONSE = 'C' THEN DO;
      FACT1 = (A-1)*(B-1)*(D-1)*(E-1) // (A-1)*(B-1)*(C-1)*(D-1) //
              (A-1)*(B-1)*(C-1)*(E-1) // (A-1)*(C-1)*(D-1)*(E-1) //
              (B-1)*(C-1)*(D-1)*(E-1);
      FACT2 = {'ABDE', 'ABCD', 'ABCE', 'ACDE', 'BCDE'};
      END;
    ELSE IF RESPONSE = 'D' THEN DO;
      FACT1 = (A-1)*(B-1)*(C-1)*(E-1) // (A-1)*(B-1)*(C-1)*(D-1) //
              (A-1)*(B-1)*(D-1)*(E-1) // (A-1)*(C-1)*(D-1)*(E-1) //
              (B-1)*(C-1)*(D-1)*(E-1);
      FACT2 = {'ABCE', 'ABCD', 'ABDE', 'ACDE', 'BCDE'};
      END;
    ELSE IF RESPONSE = 'E' THEN DO;
      FACT1 = (A-1)*(B-1)*(C-1)*(D-1) // (A-1)*(B-1)*(C-1)*(E-1) //
              (A-1)*(B-1)*(D-1)*(E-1) // (A-1)*(C-1)*(D-1)*(E-1) //
              (B-1)*(C-1)*(D-1)*(E-1);
      FACT2 = {'ABCD', 'ABCE', 'ABDE', 'ACDE', 'BCDE'};
      END;
    ORDER = (A-1) // (B-1) // (C-1) // (D-1) // (E-1) //
            (A-1)*(B-1) // (A-1)*(C-1) // (A-1)*(D-1) // (A-1)*(E-1) //
            (B-1)*(C-1) // (B-1)*(D-1) // (B-1)*(E-1) // (C-1)*(D-1) //
            (C-1)*(E-1) // (D-1)*(E-1) //
            (A-1)*(B-1)*(C-1) // (A-1)*(B-1)*(D-1) //
            (A-1)*(B-1)*(E-1) // (A-1)*(C-1)*(D-1) //
            (A-1)*(C-1)*(E-1) // (A-1)*(D-1)*(E-1) //
            (B-1)*(C-1)*(D-1) // (B-1)*(C-1)*(E-1) //
            (B-1)*(D-1)*(E-1) // (C-1)*(D-1)*(E-1) //
            FACT1 //
            (A-1)*(B-1)*(C-1)*(D-1)*(E-1);
    ORDER2 = {'A', 'B', 'C', 'D', 'E', 'AB', 'AC', 'AD', 'AE', 'BC',
              'BD', 'BE', 'CD', 'CE', 'DE', 'ABC', 'ABD', 'ABE', 'ACD',
              'ACE', 'ADE', 'BCD', 'BCE', 'BDE', 'CDE'} // FACT2 //
             {'ABCDE', 'TOTAL'};
    END;
  ELSE IF VARS = 6 THEN DO;
    UTOT = A * B * C * D * E * F;
    COUNTERS(|6|) = 1;
    COUNTERS(|5|) = 6;
    COUNTERS(|4|) = 15;
    COUNTERS(|3|) = 20;
    COUNTERS(|2|) = 15;
    COUNTERS(|1|) = 6;
    IF RESPONSE = 'A' THEN DO;
      FACT1 = (B-1)*(C-1)*(D-1)*(E-1)*(F-1) //
              (A-1)*(B-1)*(C-1)*(D-1)*(E-1) //
              (A-1)*(B-1)*(C-1)*(D-1)*(F-1) //
              (A-1)*(B-1)*(C-1)*(E-1)*(F-1) //
              (A-1)*(B-1)*(D-1)*(E-1)*(F-1) //
              (A-1)*(C-1)*(D-1)*(E-1)*(F-1);
      FACT2 = {'BCDEF', 'ABCDE', 'ABCDF', 'ABCEF', 'ABDEF', 'ACDEF'};
      END;
    ELSE IF RESPONSE = 'B' THEN DO;
      FACT1 = (A-1)*(C-1)*(D-1)*(E-1)*(F-1) //
              (A-1)*(B-1)*(C-1)*(D-1)*(E-1) //
              (A-1)*(B-1)*(C-1)*(D-1)*(F-1) //
              (A-1)*(B-1)*(C-1)*(E-1)*(F-1) //
              (A-1)*(B-1)*(D-1)*(E-1)*(F-1) //
              (B-1)*(C-1)*(D-1)*(E-1)*(F-1);
      FACT2 = {'ACDEF', 'ABCDE', 'ABCDF', 'ABCEF', 'ABDEF', 'BCDEF'};
      END;
```

```
ELSE IF RESPONSE = 'C' THEN DO;
  FACT1 = (A-1)*(B-1)*(D-1)*(E-1)*(F-1) //
          (A-1)*(B-1)*(C-1)*(D-1)*(E-1) //
          (A-1)*(B-1)*(C-1)*(D-1)*(F-1) //
          (A-1)*(B-1)*(C-1)*(E-1)*(F-1) //
          (A-1)*(C-1)*(D-1)*(E-1)*(F-1) //
          (B-1)*(C-1)*(D-1)*(E-1)*(F-1);
  FACT2 = {'ABDEF', 'ABCDE', 'ABCDF', 'ABCEF', 'ACDEF', 'BCDEF'};
  END;
ELSE IF RESPONSE = 'D' THEN DO;
  FACT1 = (A-1)*(B-1)*(C-1)*(E-1)*(F-1) //
          (A-1)*(B-1)*(C-1)*(D-1)*(E-1) //
          (A-1)*(B-1)*(C-1)*(D-1)*(F-1) //
          (A-1)*(B-1)*(D-1)*(E-1)*(F-1) //
          (A-1)*(C-1)*(D-1)*(E-1)*(F-1) //
          (B-1)*(C-1)*(D-1)*(E-1)*(F-1);
  FACT2 = {'ABCEF', 'ABCDE', 'ABCDF', 'ABDEF', 'ACDEF', 'BCDEF'};
  END;
ELSE IF RESPONSE = 'E' THEN DO;
  FACT1 = (A-1)*(B-1)*(C-1)*(D-1)*(F-1) //
          (A-1)*(B-1)*(C-1)*(D-1)*(E-1) //
          (A-1)*(B-1)*(C-1)*(E-1)*(F-1) //
          (A-1)*(B-1)*(D-1)*(E-1)*(F-1) //
          (A-1)*(C-1)*(D-1)*(E-1)*(F-1) //
          (B-1)*(C-1)*(D-1)*(E-1)*(F-1);
  FACT2 = {'ABCDF', 'ABCDE', 'ABCEF', 'ABDEF', 'ACDEF', 'BCDEF'};
  END;
ELSE IF RESPONSE = 'F' THEN DO;
  FACT1 = (A-1)*(B-1)*(C-1)*(D-1)*(E-1) //
          (A-1)*(B-1)*(C-1)*(D-1)*(F-1) //
          (A-1)*(B-1)*(C-1)*(E-1)*(F-1) //
          (A-1)*(B-1)*(D-1)*(E-1)*(F-1) //
          (A-1)*(C-1)*(D-1)*(E-1)*(F-1) //
          (B-1)*(C-1)*(D-1)*(E-1)*(F-1);
  FACT2 = {'ABCDE', 'ABCDF', 'ABCEF', 'ABDEF', 'ACDEF', 'BCDEF'};
  END;
ORDER = (A-1) // (B-1) // (C-1) // (D-1) // (E-1) // (F-1) //
        (A-1)*(B-1) // (A-1)*(C-1) // (A-1)*(D-1) // (A-1)*(E-1) //
        (A-1)*(F-1) // (B-1)*(C-1) // (B-1)*(D-1) // (B-1)*(E-1) //
        (B-1)*(F-1) // (C-1)*(D-1) // (C-1)*(E-1) // (C-1)*(F-1) //
        (D-1)*(E-1) // (D-1)*(F-1) // (E-1)*(F-1) //
        (A-1)*(B-1)*(C-1) // (A-1)*(B-1)*(D-1) //
        (A-1)*(B-1)*(E-1) // (A-1)*(B-1)*(F-1) //
        (A-1)*(C-1)*(D-1) // (A-1)*(C-1)*(E-1) //
        (A-1)*(C-1)*(F-1) // (A-1)*(D-1)*(E-1) //
        (A-1)*(D-1)*(F-1) // (A-1)*(E-1)*(F-1) //
        (B-1)*(C-1)*(D-1) // (B-1)*(C-1)*(E-1) //
        (B-1)*(C-1)*(F-1) // (B-1)*(D-1)*(E-1) //
        (B-1)*(D-1)*(F-1) // (B-1)*(E-1)*(F-1) //
        (C-1)*(D-1)*(E-1) // (C-1)*(D-1)*(F-1) //
        (C-1)*(E-1)*(F-1) // (D-1)*(E-1)*(F-1) //
        (A-1)*(B-1)*(C-1)*(D-1) // (A-1)*(B-1)*(C-1)*(E-1) //
        (A-1)*(B-1)*(C-1)*(F-1) // (A-1)*(B-1)*(D-1)*(E-1) //
        (A-1)*(B-1)*(D-1)*(F-1) // (A-1)*(B-1)*(E-1)*(F-1) //
        (A-1)*(C-1)*(D-1)*(E-1) // (A-1)*(C-1)*(D-1)*(F-1) //
        (A-1)*(C-1)*(E-1)*(F-1) // (A-1)*(D-1)*(E-1)*(F-1) //
        (B-1)*(C-1)*(D-1)*(E-1) // (B-1)*(C-1)*(D-1)*(F-1) //
        (B-1)*(C-1)*(E-1)*(F-1) // (B-1)*(D-1)*(E-1)*(F-1) //
        (C-1)*(D-1)*(E-1)*(F-1) // FACT1 //
        (A-1)*(B-1)*(C-1)*(D-1)*(E-1)*(F-1);
```

```
    ORDER2 = {'A', 'B', 'C', 'D', 'E', 'F',
              'AB', 'AC', 'AD', 'AE', 'AF', 'BC', 'BD', 'BE', 'BF',
              'CD', 'CE', 'CF', 'DE', 'DF', 'EF',
              'ABC', 'ABD', 'ABE', 'ABF', 'ACD', 'ACE', 'ACF', 'ADE',
              'ADF', 'AEF', 'BCD', 'BCE', 'BCF', 'BDE', 'BDF', 'BEF',
              'CDE', 'CDF', 'CEF', 'DEF',
              'ABCD', 'ABCE', 'ABCF', 'ABDE', 'ABDF', 'ABEF', 'ACDE',
              'ACDF', 'ACEF', 'ADEF', 'BCDE', 'BCDF', 'BCEF', 'BDEF',
              'CDEF'} // FACT2 //
              {'ABCDEF', 'TOTAL'};
    END;
  DEV = J(2**VARS,1,0);
  PVAL = J(2**VARS,1,0);
  ORDERING = J(2**VARS,1,0);
*-- CREATE INITIAL M AND U MATRICES USING THE INITIAL RANKING OF ----*;
*-- FACTOR EFFECTS -------------------------------------------------*;
  RUN M;
  RUN U;
  TEMPM = M;
  TEMPU = U;
*-- RANKS THE CLASSES OF FACTOR EFFECTS STARTING WITH THE HIGHEST ---*;
*-- EFFECT DOWN TO THE MAIN EFFECTS ---------------------------------*;
*-- NOTE: IF WE FIT A LOGLINEAR MODEL WITH A FIXED MARGINAL TOTAL ---*;
*--        WE HAVE ONE LESS FACTOR EFFECT TO RANK -------------------*;
  DO SETUSED = VARS TO 1 BY -1;
    IF (SETUSED = 5 & VARS ¬= 5) THEN DO;
      START = COUNTERS(|6|) + 1;
      FINISH = SUM(COUNTERS(|5:6|)) - 1;
      IF (VARS = 6 & FIXEDMAR = 'YES') THEN FINISH = FINISH - 1;
      END;
    ELSE IF (SETUSED = 4 & VARS ¬= 4) THEN DO;
      START = SUM(COUNTERS(|5:6|)) + 1;
      FINISH = SUM(COUNTERS(|4:6|)) - 1;
      IF (VARS = 5 & FIXEDMAR = 'YES') THEN FINISH = FINISH - 1;
      END;
    ELSE IF (SETUSED = 3 & VARS ¬= 3) THEN DO;
      START = SUM(COUNTERS(|4:6|)) + 1;
      FINISH = SUM(COUNTERS(|3:6|)) - 1;
      IF (VARS = 4 & FIXEDMAR = 'YES') THEN FINISH = FINISH - 1;
      END;
    ELSE IF (SETUSED = 2 & VARS ¬= 2) THEN DO;
      START = SUM(COUNTERS(|3:6|)) + 1;
      FINISH = SUM(COUNTERS(|2:6|)) - 1;
      IF (VARS = 3 & FIXEDMAR = 'YES') THEN FINISH = FINISH - 1;
      END;
    ELSE IF (SETUSED = 1) THEN DO;
      START = SUM(COUNTERS(|2:6|)) + 1;
      FINISH = SUM(COUNTERS(|1:6|)) - 1;
      END;
*-- IF THE SET IN USE CORRESPONDS TO THE HYPOTHESIS H: U(HIGHEST ----*;
*-- FACTOR EFFECT) = 0 THEN NO RANKING NEEDS TO BE DONE SINCE WE ----*;
*-- ONLY HAVE ONE FACTOR EFFECT. OTHERWISE, DETERMINES THE RANKING --*;
*-- OF THE FACTOR EFFECTS FOR THE CLASS OF FACTOR EFFECTS -----------*;
*-- CORRESPONDING TO THE HYPOTHESIS IN USE --------------------------*;
    IF SETUSED ¬= VARS THEN DO;
      DO POSITION = START TO FINISH;
*-- EXECUTES MODULE RANKING -----------------------------------------*;
        RUN RANKING;
        END;
      END;
```

```
    NU = UTOT;
*-- EXECUTES MODULES SET AND CONTRIB --------------------------------*;
    RUN SET;
*-- COMPUTES THE CHI-SQUARE COMPONENT VALUE OF ALL THE FACTOR EFFECTS *;
    RUN CONTRIB;
*-- SAVES THE RESULTS OBTAINED UNOER A SINGLE HYPOTHESIS -------------*;
    OROERING = ORDERING || DEV;
    OROERING = ORDERING || PVAL;
    END;
*-- PRINTS THE RESULTS OF THE RX-SQUARE ANALYSIS ---------------------*;
*-- EXECUTES MODULE PRINT -------------------------------------------*;
  RUN PRINT;
  FINISH;
************************  END OF MODULE MAIN  *************************;
```

Appendix A.2:
M module and auxiliary modules (VALUEX1, VALUEX2, VALUEX3, VALUEX4, VALUEX5, AND VALUEX6.)

MODULE	: M.		
PROCESSING	: This module is used to generate the initial M matrix. The representation is based on the method presented by Haberman (see Haberman, 1978, p.208).		
CALLED BY	: MAIN module.		
MODULES CALLED	: VALUEX1, VALUEX2, VALUEX3, VALUEX4, VALUEX5, VALUEX6.		
VARIABLES USED :	A	:	Number of levels of variable 1.
	B	:	Number of levels of variable 2.
	C	:	Number of levels of variable 3.
	D	:	Number of levels of variable 4.
	E	:	Number of levels of variable 5.
	F	:	Number of levels of variable 6.
	CINDEX1	:	
	CINDEX2	:	
	CINDEX3	:	Indices indicating which columns of
	CINDEX4	:	the M matrix we are modifying.
	CINDEX5	:	
	CINDEX6	:	
	COLNUM	:	Column number indicator.
	INDICES	:	Vector of index variables for processing
	LOOP1	:	
	LOOP2	:	
	LOOP3	:	Variables to keep track of processing.
	LOOP4	:	
	LOOP5	:	
	LOOP6	:	
	M	:	M matrix.
	ORDER	:	Vector containing the degrees of freedom of each factor effect.
	RINDICES	:	Vector of row index variables for processing.
	RINDEX1	:	
	RINDEX2	:	
	RINDEX3	:	Indices indicating which rows of the
	RINDEX4	:	M matrix we are modifying.
	RINDEX5	:	
	RINDEX6	:	
	ROWNUM	:	Row number indicator.
	UTOT	:	Total number of u-parameters.
	VARS	:	Total number of variables.
	X1	:	
	X2	:	
	X3	:	Numerical variables having possible
	X4	:	values -1, 0 or 1.
	X5	:	
	X6	:	

```
***********************   START OF MODULE M  ****************************;
START M;
*-- INITIAL M MATRIX CONSISTING ENTIRELY OF ONES ---------------------*;
  M = J(UTOT,UTOT,1);
*-- INITIALIZE THE ROW NUMBER ----------------------------------------*;
  ROWNUM = 1;
*-- INITIALIZE THE INDEX VARIABLES -----------------------------------*;
  INDICES = J(6,1,0);
  INDICES(|1|) = A;
  INDICES(|2|) = B;
  IF VARS > 2 THEN INDICES(|3|) = C; ELSE INDICES(|3|) = 1;
  IF VARS > 3 THEN INDICES(|4|) = D; ELSE INDICES(|4|) = 1;
  IF VARS > 4 THEN INDICES(|5|) = E; ELSE INDICES(|5|) = 1;
  IF VARS > 5 THEN INDICES(|6|) = F; ELSE INDICES(|6|) = 1;
*-- CHANGE THE CONTENT OF THE APPROPRIATE CELLS ----------------------*;
  RINDICES = J(6,1,0);
  DO RINDEX1 = 1 TO INDICES(|1|); RINDICES(|1|) = RINDEX1;
    DO RINDEX2 = 1 TO INDICES(|2|); RINDICES(|2|) = RINDEX2;
      DO RINDEX3 = 1 TO INDICES(|3|); RINDICES(|3|) = RINDEX3;
        DO RINDEX4 = 1 TO INDICES(|4|); RINDICES(|4|) = RINDEX4;
          DO RINDEX5 = 1 TO INDICES(|5|); RINDICES(|5|) = RINDEX5;
            DO RINDEX6 = 1 TO INDICES(|6|); RINDICES(|6|) = RINDEX6;
*-- INITIALIZE THE COLUMN NUMBER -------------------------------------*;
              COLNUM = 1;
*-- MODIFIES THE CONTENT OF THE CELLS CORRESPONDING TO THE -----------*;
*-- MAIN EFFECTS -----------------------------------------------------*;
              DO LOOP1 = 1 TO VARS;
                DO CINDEX1 = 1 TO ORDER(|LOOP1|);
*-- EXECUTES MODULE VALUEX1 ------------------------------------------*;
                  RUN VALUEX1;
*-- INCREMENTS TO THE NEXT COLUMN ------------------------------------*;
                  COLNUM = COLNUM + 1;
                  M(|ROWNUM,COLNUM|) = X1;
                  END;
                END;
*-- MODIFIES THE CONTENT OF THE CELLS CORRESPONDING TO THE -----------*;
*-- TWO FACTOR EFFECTS -----------------------------------------------*;
              DO LOOP1 = 1 TO (VARS - 1);
                DO LOOP2 = (LOOP1 + 1) TO VARS;
                  DO CINDEX1 = 1 TO ORDER(|LOOP1|);
                    DO CINDEX2 = 1 TO ORDER(|LOOP2|);
*-- EXECUTES MODULES VALUEX1 AND VALUEX2 -----------------------------*;
                      RUN VALUEX1;
                      RUN VALUEX2;
*-- INCREMENTS TO THE NEXT COLUMN ------------------------------------*;
                      COLNUM = COLNUM + 1;
                      M(|ROWNUM,COLNUM|) = X1*X2;
                      END;
                    END;
                  END;
                END;
*-- MODIFIES THE CONTENT OF THE CELLS CORRESPONDING TO THE -----------*;
*-- THREE FACTOR EFFECTS ---------------------------------------------*;
              IF VARS > 2 THEN DO;
                DO LOOP1 = 1 TO (VARS - 2);
                  DO LOOP2 = (LOOP1 + 1) TO (VARS - 1);
                    DO LOOP3 = (LOOP2 + 1) TO VARS;
```

```
                        DO CINDEX1 = 1 TO ORDER(|LOOP1|);
                          DO CINDEX2 = 1 TO ORDER(|LOOP2|);
                            DO CINDEX3 = 1 TO ORDER(|LOOP3|);
*-- EXECUTES MODULES VALUEX1, VALUEX2 AND VALUEX3 -------------------*;
                              RUN VALUEX1;
                              RUN VALUEX2;
                              RUN VALUEX3;
*-- INCREMENTS TO THE NEXT COLUMN ----------------------------------*;
                              COLNUM = COLNUM + 1;
                              M(|ROWNUM,COLNUM|) = X1*X2*X3;
                              END;
                            END;
                          END;
                        END;
                      END;
                    END;
                  END;
*-- MODIFIES THE CONTENT OF THE CELLS CORRESPONDING TO THE -----------*;
*-- FOUR FACTOR EFFECTS ----------------------------------------------*;
              IF VARS > 3 THEN DO;
                DO LOOP1 = 1 TO (VARS - 3);
                  DO LOOP2 = (LOOP1 + 1) TO (VARS - 2);
                    DO LOOP3 = (LOOP2 + 1) TO (VARS - 1);
                      DO LOOP4 = (LOOP3 + 1) TO VARS;
                        DO CINDEX1 = 1 TO ORDER(|LOOP1|);
                          DO CINDEX2 = 1 TO ORDER(|LOOP2|);
                            DO CINDEX3 = 1 TO ORDER(|LOOP3|);
                              DO CINDEX4 = 1 TO ORDER(|LOOP4|);
*-- EXECUTES MODULES VALUEX1, VALUEX2, VALUEX3 AND VALUEX4 -----------*
                                RUN VALUEX1;
                                RUN VALUEX2;
                                RUN VALUEX3;
                                RUN VALUEX4;
*-- INCREMENTS TO THE NEXT COLUMN ------------------------------------*
                                COLNUM = COLNUM + 1;
                                M(|ROWNUM,COLNUM|) = X1*X2*X3*X4;
                                END;
                              END;
                            END;
                          END;
                        END;
                      END;
                    END;
                  END;
                END;
*-- MODIFIES THE CONTENT OF THE CELLS CORRESPONDING TO THE -----------*
*-- FIVE FACTOR EFFECTS ----------------------------------------------*
              IF VARS > 4 THEN DO;
                DO LOOP1 = 1 TO (VARS - 4);
                  DO LOOP2 = (LOOP1 + 1) TO (VARS - 3);
                    DO LOOP3 = (LOOP2 + 1) TO (VARS - 2);
                      DO LOOP4 = (LOOP3 + 1) TO (VARS - 1);
                        DO LOOP5 = (LOOP4 + 1) TO VARS;
                          DO CINDEX1 = 1 TO ORDER(|LOOP1|);
                            DO CINDEX2 = 1 TO ORDER(|LOOP2|);
                              DO CINDEX3 = 1 TO ORDER(|LOOP3|);
                                DO CINDEX4 = 1 TO ORDER(|LOOP4|);
                                  DO CINDEX5 = 1 TO ORDER(|LOOP5|);
```

```
*-- EXECUTES MODULES VALUEX1, VALUEX2, VALUEX3, VALUEX4 AND VALUEX5 --*
                                 RUN VALUEX1;
                                 RUN VALUEX2;
                                 RUN VALUEX3;
                                 RUN VALUEX4;
                                 RUN VALUEX5;
*-- INCREMENTS TO THE NEXT COLUMN -----------------------------------*
                                 COLNUM = COLNUM + 1;
                                 M(|ROWNUM,COLNUM|)=
                                 X1*X2*X3*X4*X5;
                                 END;'
                               END;
                             END;
                           END;
                         END;
                       END;
                     END;
                   END;
                 END;
               END;
             END;
*-- MODIFIES THE CONTENT OF THE CELLS CORRESPONDING TO THE -----------*
*-- SIX FACTOR EFFECTS - ---------------------------------------------*
           IF VARS > 5 THEN DO;
             DO LOOP1 = 1 TO (VARS - 5);
               DO LOOP2 = (LOOP1 + 1) TO (VARS - 4);
                 DO LOOP3 = (LOOP2 + 1) TO (VARS - 3);
                   DO LOOP4 = (LOOP3 + 1) TO (VARS - 2);
                     DO LOOP5 = (LOOP4 + 1) TO (VARS - 1);
                       DO LOOP6 = (LOOP5 + 1) TD VARS;
                         DO CINDEX1 = 1 TO ORDER(|LOOP1|);
                           DO CINDEX2 = 1 TO ORDER(|LOOP2|);
                             DO CINDEX3 = 1 TO ORDER(|LOOP3|);
                               DO CINDEX4 = 1 TO ORDER(|LOOP4|);
                                 DO CINDEX5 = 1 TO ORDER(|LOOP5|);
                                   DO CINDEX6 = 1 TO ORDER(|LOOP6|);
*-- EXECUTES MODULES VALUEX1, VALUEX2, VALUEX3, VALUEX4, -------------*;
*-- VALUEX5 AND VALUEX6 ----------------------------------------------*;
                                     RUN VALUEX1;
                                     RUN VALUEX2;
                                     RUN VALUEX3;
                                     RUN VALUEX4;
                                     RUN VALUEX5;
                                     RUN VALUEX6;
*-- INCREMENTS TO THE NEXT COLUMN ------------------------------------*;
                                     COLNUM = COLNUM + 1;
                                     M(|ROWNUM,COLNUM|)=
                                     X1*X2*X3*X4*X5*X6;;
                                     END;
                                   END;
                                 END;
                               END;
                             END;
                           END;
                         END;
                       END;
                     END;
                   END;
                 END;
               END;
             END;
```

```
*-- INCREMENTS TO THE NEXT ROW ---------------------------------------------*;
                ROWNUM = ROWNUM + 1;
                END;
              END;
            END;
          END;
        END;
      END;
*-- FREE WORKSPACE ---------------------------------------------------------*;
  FREE ROWNUM COLNUM LOOP1 LOOP2 LOOP3 LOOP4 LOOP5 LOOP6
       INDICES RINDICES
       RINDEX1 RINDEX2 RINDEX3 RINDEX4 RINDEX5 RINDEX6
       CINDEX1 CINDEX2 CINDEX3 CINDEX4 CINDEX5 CINDEX6
       X1 X2 X3 X4 X5 X6;
  FINISH;
********************   END OF MODULE M  ********************************;
```

MODULES : VALUEX1, VALUEX2, VALUEX3, VALUEX4, VALUEX5, VALUEX6.

PROCESSING : These modules are used in the creation of the M matrix. The representation is based on the method presented by Haberman (see Haberman, 1978, p.208).

CALLED BY : M module.

MODULES CALLED : none.

VARIABLES USED :

AUX1	:	Auxiliary variables for processing.
AUX2	:	
CINDEX1	:	
CINDEX2	:	
CINDEX3	:	Indices indicating which columns of
CINDEX4	:	the M matrix we are modifying.
CINDEX5	:	
CINDEX6	:	
INDICES	:	Vector of index variables for processing.
LOOP1	:	
LOOP2	:	
LOOP3	:	Variables to keep track of processing.
LOOP4	:	
LOOP5	:	
LOOP6	:	
RINDICES	:	Vector of row index variables for processing.
X1	:	
X2	:	
X3	:	Numerical variables having possible
X4	:	values -1, 0 or 1.
X5	:	
X6	:	

```
*******************   START OF MODULE VALUEX1  ************************;
START VALUEX1;
  AUX1 = RINDICES(|LOOP1|);
  AUX2 = INDICES(|LOOP1|);
  IF AUX1 = CINDEX1 THEN X1 = 1;
  ELSE IF AUX1 = AUX2 THEN X1 = -1;
  ELSE X1 = 0;
*-- FREE WORKSPACE -----------------------------------------------------*;
  FREE AUX1 AUX2;
  FINISH;
********************   END OF MODULE VALUEX1  ************************;

*******************   START OF MODULE VALUEX2  ************************;
START VALUEX2;
  AUX1 = RINDICES(|LOOP2|);
  AUX2 = INDICES(|LOOP2|);
  IF AUX1 = CINDEX2 THEN X2 = 1;
  ELSE IF AUX1 = AUX2 THEN X2 = -1;
  ELSE X2 = 0;
*-- FREE WORKSPACE -----------------------------------------------------*;
  FREE AUX1 AUX2;
  FINISH;
********************   END OF MODULE VALUEX2  ************************;

*******************   START OF MODULE VALUEX3  ************************;
START VALUEX3;
  AUX1 = RINDICES(|LOOP3|);
  AUX2 = INDICES(|LOOP3|);
  IF AUX1 = CINDEX3 THEN X3 = 1;
  ELSE IF AUX1 = AUX2 THEN X3 = -1;
  ELSE X3 = 0;
*-- FREE WORKSPACE -----------------------------------------------------*;
  FREE AUX1 AUX2;
  FINISH;
********************   END OF MODULE VALUEX3  ************************;

*******************   START OF MODULE VALUEX4  ************************;
START VALUEX4;
  AUX1 = RINDICES(|LOOP4|);
  AUX2 = INDICES(|LOOP4|);
  IF AUX1 = CINDEX4 THEN X4 = 1;
  ELSE IF AUX1 = AUX2 THEN X4 = -1;
  ELSE X4 = 0;
*-- FREE WORKSPACE -----------------------------------------------------*;
  FREE AUX1 AUX2;
  FINISH;
********************   END OF MODULE VALUEX4  ************************;

*******************   START OF MODULE VALUEX5  ************************;
START VALUEX5;
  AUX1 = RINDICES(|LOOP5|);
  AUX2 = INDICES(|LOOP5|);
  IF AUX1 = CINDEX5 THEN X5 = 1;
  ELSE IF AUX1 = AUX2 THEN X5 = -1;
  ELSE X5 = 0;
*-- FREE WORKSPACE -----------------------------------------------------*;
  FREE AUX1 AUX2;
```

```
  FINISH;
********************   END OF MODULE VALUEX5  *************************;

*******************   START OF MODULE VALUEX6  ***********************;
START VALUEX6;
  AUX1 = RINDICES(|LOOP6|);
  AUX2 = INDICES(|LOOP6|);
  IF AUX1 = CINDEX6 THEN X6 = 1;
  ELSE IF AUX1 = AUX2 THEN X6 = -1;
  ELSE X6 = 0;
*-- FREE WORKSPACE ---------------------------------------------------*;
  FREE AUX1 AUX2;
  FINISH;
********************   END OF MODULE VALUEX6  *************************;
```

Appendix A.3:
U module and auxiliary modules (DELTA1, DELTA2, DELTA3, DELTA4, DELTA5, and DELTA6)

MODULE : U.

PROCESSING : This module is used to generate the initial U matrix. The representation is based on the method presented by Haberman (see Haberman, 1978, p.249). (Note: the U matrix created is the U matrix presented by Haberman multiplied by the factor UTOT).

CALLED BY : MAIN module.

MODULES CALLED : DELTA1, DELTA2, DELTA3, DELTA4, DELTA5, DELTA6.

VARIABLES USED :

A	: Number of levels of variable 1.
B	: Number of levels of variable 2.
C	: Number of levels of variable 3.
D	: Number of levels of variable 4.
E	: Number of levels of variable 5.
F	: Number of levels of variable 6.
CINDICES	: Vector of column index variables for processing.
CINDEX1	:
CINDEX2	:
CINDEX3	: Indices indicating which columns of
CINDEX4	: the U matrix we are modifying.
CINDEX5	:
CINDEX6	:
COLNUM	: Column number indicator.
DELTA	: Numerical variable having possible value 0 or 1.
DELTA1	:
DELTA2	:
DELTA3	: Numerical variables for the creation
DELTA4	: of the U matrix.
DELTA5	:
DELTA6	:
INDICES	: Vector of index variables for processing.
LOOP1	:
LOOP2	:
LOOP3	: Variables to keep track of processing.
LOOP4	:
LOOP5	:
LOOP6	:
ORDER	: Vector containing the degrees of freedom of each factor effect.
RINDEX1	:
RINDEX2	:
RINDEX3	: Indices indicating which rows of the
RINDEX4	: U matrix we are modifying.
RINDEX5	:
RINDEX6	:
ROWNUM	: Row number indicator.
U	: U matrix.
UTOT	: Total number of u-parameters.
VARS	: Total number of variables.

```
*********************   START OF MODULE U  **************************;
START U;
*-- INITIAL MATRIX CONSISTING ENTIRELY OF ONES -----------------------*;
  U = J(UTOT,UTOT,1);
*-- INITIALIZE THE COLUMN NUMBER -------------------------------------*;
  COLNUM = 1;
*-- INITIALIZE THE INDEX VARIABLES -----------------------------------*;
  INDICES = J(6,1,0);
  INDICES(|1|) = A;
  INDICES(|2|) = B;
  IF VARS > 2 THEN INDICES(|3|) = C; ELSE INDICES(|3|) = 1;
  IF VARS > 3 THEN INDICES(|4|) = D; ELSE INDICES(|4|) = 1;
  IF VARS > 4 THEN INDICES(|5|) = E; ELSE INDICES(|5|) = 1;
  IF VARS > 5 THEN INDICES(|6|) = F; ELSE INDICES(|6|) = 1;
*-- CHANGE THE CONTENT OF THE APPROPRIATE CELLS ----------------------*;
  CINDICES = J(6,1,0);
  DO CINDEX1 = 1 TO INDICES(|1|); CINDICES(|1|) = CINDEX1;
    DO CINDEX2 = 1 TO INDICES(|2|); CINDICES(|2|) = CINDEX2;
      DO CINDEX3 = 1 TO INDICES(|3|); CINDICES(|3|) = CINDEX3;
        DO CINDEX4 = 1 TO INDICES(|4|); CINDICES(|4|) = CINDEX4;
          DO CINDEX5 = 1 TO INDICES(|5|); CINDICES(|5|) = CINDEX5;
            DO CINDEX6 = 1 TO INDICES(|6|); CINDICES(|6|) = CINDEX6;
*-- INITIALIZE THE ROW NUMBER  ---------------------------------------*;
              ROWNUM = 1;
*-- MODIFIES THE CONTENT OF THE CELLS CORRESPONDING TO THE -----------*;
*-- MAIN EFFECTS -----------------------------------------------------*;
              DO LOOP1 = 1 TO VARS;
                DO RINDEX1 = 1 TO ORDER(|LOOP1|);
*-- EXECUTES MODULE DELTA1 -------------------------------------------*;
                  RUN DELTA1;
*-- INCREMENTS TO THE NEXT ROW ---------------------------------------*;
                  ROWNUM = ROWNUM + 1;
                  U(|ROWNUM,COLNUM|) = DELTA1;
                  END;
                END;
*-- MODIFIES THE CONTENT OF THE CELLS CORRESPONDING TO THE -----------*;
*-- TWO FACTOR EFFECTS -----------------------------------------------*;
              DO LOOP1 = 1 TO (VARS - 1);
                DO LOOP2 = (LOOP1 + 1) TO VARS;
                  DO RINDEX1 = 1 TO ORDER(|LOOP1|);
                    DO RINDEX2 = 1 TO ORDER(|LOOP2|);
*-- EXECUTES MODULES DELTA1 AND DELTA2 -------------------------------*;
                      RUN DELTA1;
                      RUN DELTA2;
*-- INCREMENTS TO THE NEXT ROW ---------------------------------------*;
                      ROWNUM = ROWNUM + 1;
                      U(|ROWNUM,COLNUM|) = DELTA1*DELTA2;
                      END;
                    END;
                  END;
                END;
*-- MODIFIES THE CONTENT OF THE CELLS CORRESPONDING TO THE -----------*;
*-- THREE FACTOR EFFECTS ---------------------------------------------*;
              IF VARS > 2 THEN DO;
                DO LOOP1 = 1 TO (VARS - 2);
                  DO LOOP2 = (LOOP1 + 1) TO (VARS - 1);
                    DO LOOP3 = (LOOP2 + 1) TO VARS;
```

```
                        DO RINDEX1 = 1 TO ORDER(|LOOP1|);
                          DO RINDEX2 = 1 TO ORDER(|LOOP2|);
                            DO RINDEX3 = 1 TO ORDER(|LOOP3|);
*-- EXECUTES MODULES DELTA1, DELTA2 AND DELTA3 ------------------------*;
                              RUN DELTA1;
                              RUN DELTA2;
                              RUN DELTA3;
*-- INCREMENTS TO THE NEXT ROW ----------------------------------------*;
                              ROWNUM = ROWNUM + 1;
                              U(|ROWNUM,COLNUM|) =
                              DELTA1*DELTA2*DELTA3;
                              END;
                            END;
                          END;
                        END;
                      END;
                    END;
                  END;
*-- MODIFIES THE CONTENT OF THE CELLS CORRESPONDING TO THE -----------*;
*-- FOUR FACTOR EFFECTS -----------------------------------------------*;
                IF VARS > 3 THEN DO;
                  DO LOOP1 = 1 TO (VARS - 3);
                    DO LOOP2 = (LOOP1 + 1) TO (VARS - 2);
                      DO LOOP3 = (LOOP2 + 1) TO (VARS - 1);
                        DO LOOP4 = (LOOP3 + 1) TO VARS;
                          DO RINDEX1 = 1 TO ORDER(|LOOP1|);
                            DO RINDEX2 = 1 TO ORDER(|LOOP2|);
                              DO RINDEX3 = 1 TO ORDER(|LOOP3|);
                                DO RINDEX4 = 1 TO ORDER(|LOOP4|);
*-- EXECUTES MODULES DELTA1, DELTA2, DELTA3 AND DELTA4 ---------------*;
                                  RUN DELTA1;
                                  RUN DELTA2;
                                  RUN DELTA3;
                                  RUN DELTA4;
*-- INCREMENTS TO THE NEXT COLUMN -------------------------------------*;
                                  ROWNUM = ROWNUM + 1;
                                  U(|ROWNUM,COLNUM|) =
                                  DELTA1*DELTA2*DELTA3*DELTA4;
                                  END;
                                END;
                              END;
                            END;
                          END;
                        END;
                      END;
                    END;
                  END;
*-- MODIFIES THE CONTENT OF THE CELLS CORRESPONDING TO THE -----------*;
*-- FIVE FACTOR EFFECTS -----------------------------------------------*;
                IF VARS > 4 THEN DO;
                  DO LOOP1 = 1 TO (VARS - 4);
                    DO LOOP2 = (LOOP1 + 1) TO (VARS - 3);
                      DO LOOP3 = (LOOP2 + 1) TO (VARS - 2);
                        DO LOOP4 = (LOOP3 + 1) TO (VARS - 1);
                          DO LOOP5 = (LOOP4 + 1) TO VARS;
                            DO RINDEX1 = 1 TO ORDER(|LOOP1|);
                              DO RINDEX2 = 1 TO ORDER(|LOOP2|);
                                DO RINDEX3 = 1 TO ORDER(|LOOP3|);
                                  DO RINDEX4 = 1 TO ORDER(|LOOP4|);
                                    DO RINDEX5 = 1 TO ORDER(|LOOP5|);
```

```
*-- EXECUTES MODULES DELTA1, DELTA2, DELTA3, DELTA4 AND DELTA5 -------*;
                                         RUN DELTA1;
                                         RUN DELTA2;
                                         RUN DELTA3;
                                         RUN DELTA4;
                                         RUN DELTA5;
*-- INCREMENTS TO THE NEXT COLUMN ------------------------------------*;
                                         ROWNUM = ROWNUM + 1;
                                         U(|ROWNUM,COLNUM|)=
                                         DELTA1*DELTA2*DELTA3*DELTA4*DELTA5;
                                         END;
                                       END;
                                     END;
                                   END;
                                 END;
                               END;
                             END;
                           END;
                         END;
                       END;
                     END;
*-- MODIFIES THE CONTENT OF THE CELLS CORRESPONDING TO THE -----------*;
*-- SIX FACTOR EFFECTS - ---------------------------------------------*;
                   IF VARS > 5 THEN DO;
                     DO LOOP1 = 1 TO (VARS - 5);
                       DO LOOP2 = (LOOP1 + 1) TO (VARS - 4);
                         DO LOOP3 = (LOOP2 + 1) TO (VARS - 3);
                           DO LDOP4 = (LOOP3 + 1) TO (VARS - 2);
                             DO LOOP5 = (LOOP4 + 1) TO (VARS - 1);
                               DO LOOP6 = (LOOP5 + 1) TO VARS;
                                 DO RINDEX1 = 1 TO ORDER(|LOOP1|);
                                   DO RINDEX2 = 1 TO ORDER(|LOOP2|);
                                     DO RINDEX3 = 1 TO ORDER(|LOOP3|);
                                       DO RINDEX4 = 1 TO ORDER(|LOOP4|);
                                         DO RINDEX5 = 1 TO ORDER(|LOOP5|);
                                           DO RINDEX6 =1 TO ORDER(|LOOP6|);
*-- EXECUTES MODULES DELTA1, DELTA2, DELTA3, DELTA4, -----------------*;
*-- DELTA5 AND DELTA6 ------------------------------------------------*;
                                             RUN DELTA1;
                                             RUN DELTA2;
                                             RUN DELTA3;
                                             RUN DELTA4;
                                             RUN DELTA5;
                                             RUN DELTA6;
*-- INCREMENTS TO THE NEXT COLUMN ------------------------------------*;
                                             ROWNUM = ROWNUM + 1;
                                             U(|ROWNUM,COLNUM|)=
                                             DELTA1*DELTA2*DELTA3*
                                             DELTA4*DELTA5*DELTA6;
                                             END;
                                           END;
                                         END;
                                       END;
                                     END;
                                   END;
                                 END;
                               END;
                             END:
```

```
                      END;
                    END;
                  END;
                END;
*-- INCREMENTS TO THE NEXT COLUMN ------------------------------------*;
              COLNUM = COLNUM + 1;
              END;
            END;
          END;
        END;
      END;
    END;
*-- FREE WORKSPACE ---------------------------------------------------*;
  FREE ROWNUM COLNUM LOOP1 LOOP2 LOOP3 LOOP4 LOOP5 LOOP6
       INDICES CINDICES
       RINDEX1 RINDEX2 RINDEX3 RINDEX4 RINDEX5 RINDEX6
       CINDEX1 CINDEX2 CINDEX3 CINDEX4 CINDEX5 CINDEX6
       DELTA1 DELTA2 DELTA3 DELTA4 DELTA5 DELTA6;
  FINISH;
**********************   END OF MODULE U  ****************************;
```

MODULES : DELTA1, DELTA2, DELTA3, DELTA4, DELTA5, DELTA6.

PROCESSING : These modules are used in the creation of the U matrix. The representation is based on the method presented by Haberman (see Haberman, 1978, p.249).

CALLED BY : U module.

MODULES CALLED : none.

VARIABLES USED : AUX1 : Auxiliary variables for processing.
AUX2 :
CINDICES: Vector of column index variables for processing.
DELTA : Numerical variables having possible value 0 or 1.
DELTA1 :
DELTA2 :
DELTA3 : Numerical variables for the creation
DELTA4 : of the U matrix.
DELTA5 :
DELTA6 :
INDICES : Vector of index variables for processing.
LOOP1 :
LOOP2 :
LOOP3 : Variables to keep track of processing.
LOOP4 :
LOOP5 :
LOOP6 :
RINDEX1 :
RINDEX2 :
RINDEX3 : Indices indicating which rows of the
RINDEX4 : U matrix we are modifying.
RINDEX5 :
RINDEX6 :

```
******************   START OF MODULE DELTA1   ************************;
START DELTA1;
  AUX1 = CINDICES(|LOOP1|);
  AUX2 = INDICES(|LOOP1|);
  IF AUX1 = RINDEX1 THEN DELTA = 1;
  ELSE DELTA = 0;
  DELTA1 = AUX2 * DELTA - 1;
*-- FREE WORKSPACE ---------------------------------------------------*;
  FREE AUX1 AUX2 DELTA;
  FINISH;
*******************   END OF MODULE DELTA1   ************************;

******************   START OF MODULE DELTA2   ************************;
START DELTA2;
  AUX1 = CINDICES(|LOOP2|);
  AUX2 = INDICES(|LOOP2|);
  IF AUX1 = RINDEX2 THEN DELTA = 1;
  ELSE DELTA = 0;
  DELTA2 = AUX2 * DELTA - 1;
*-- FREE WORKSPACE ---------------------------------------------------*;
  FREE AUX1 AUX2 DELTA;
  FINISH;
*******************   END OF MODULE DELTA2   ************************;

******************   START OF MODULE DELTA3   ************************;
START DELTA3;
  AUX1 = CINDICES(|LOOP3|);
  AUX2 = INDICES(|LOOP3|);
  IF AUX1 = RINDEX3 THEN DELTA = 1;
  ELSE DELTA = 0;
  DELTA3 = AUX2 * DELTA - 1;
*-- FREE WORKSPACE ---------------------------------------------------*;
  FREE AUX1 AUX2 DELTA;
  FINISH;
*******************   END OF MODULE DELTA3   ************************;

******************   START OF MODULE DELTA4   ************************;
START DELTA4;
  AUX1 = CINDICES(|LOOP4|);
  AUX2 = INDICES(|LOOP4|);
  IF AUX1 = RINDEX4 THEN DELTA = 1;
  ELSE DELTA = 0;
  DELTA4 = AUX2 * DELTA - 1;
*-- FREE WORKSPACE ---------------------------------------------------*;
  FREE AUX1 AUX2 DELTA;
  FINISH;
*******************   END OF MODULE DELTA4   ************************;

******************   START OF MODULE DELTA5   ************************;
START DELTA5;
  AUX1 = CINDICES(|LOOP5|);
  AUX2 = INDICES(|LOOP5|);
  IF AUX1 = RINDEX5 THEN DELTA = 1;
  ELSE DELTA = 0;
  DELTA5 = AUX2 * DELTA -1 ;
*-- FREE WORKSPACE ---------------------------------------------------*;
  FREE AUX1 AUX2 DELTA;
```

```
  FINISH;
********************   END OF MODULE DELTA5   *************************;

******************   START OF MODULE DELTA6   ************************;
START DELTA6;
  AUX1 = CINDICES(|LOOP6|);
  AUX2 = INDICES(|LOOP6|);
  IF AUX1 = RINDEX6 THEN DELTA = 1;
  ELSE DELTA = 0;
  DELTA6 = AUX2 * DELTA - 1;
*-- FREE WORKSPACE ---------------------------------------------------*;
  FREE AUX1 AUX2 DELTA;
  FINISH;
********************   END OF MODULE DELTA6   *************************;
```

D

Appendix A.4: SET module
(for loglinear models with or without a fixed marginal total)

MODULE : SET

PROCESSING : Computes the f vector and G matrix for a given null hypothesis.

CALLED BY : RANKING module, MAIN module.

MODULES CALLED : none.

VARIABLES USED :

OBS	:	Vector of observed counts .
MLE	:	Vector of expected counts (MLE's).
f	:	Score function for u-parameters.
G	:	Inverse of the covariance matrix of f.
SETUSED	:	Indicator for the current set of expected counts in use.
TEMPM	:	Temporary M matrix representing the current ranking of U-parameters.
TEMPU	:	Temporary U matrix representing the inverse of the TEMPM matrix multiplied by the factor UTOT.
UTOT	:	Total number of u-parameters.

```
************************  START OF MODULE SET  ************************;
START SET;

*********************************************************************;
************************  START OF USER INPUT  ************************;
*********************************************************************;

*********************************************************************
***                                                               ***
***                 ARMED FORCES QUALIFICATION DATA               ***
***                        (FIENBERG, P.89)                       ***
***                  (I = 3, J = 4, K = 2, L = 2)                 ***
***                                                               ***
*********************************************************************;
*-- MATRIX OF OBSERVED FREQUENCIES --*;
OBS  = {39,  19, 231, 110,   4,   5,  17,  18,  11,   2,
        18,  11,  48,  49, 197, 178,  29,  40, 115, 133,
         8,  17,  21,  38,   9,  14,  28,  25,  17,  79,
       111, 206,   8,  19,  51, 103,   1,   7,  13,  25,
         6,   3,  45,  18,   8,  24,  35,  81};
*-- MATRIX OF EXPECTED FREQUENCIES FOR H: U(4 FACTOR EFFECT) = 0 --*;
IF SETUSED = 4 THEN
     MLE = {41.2699, 16.7301, 228.73,  112.27,  4.33939, 4.66061,
            16.6606, 18.3394, 9.52291, 3.47709, 19.4771, 9.52291,
            46.8678, 50.1322, 198.132, 176.868, 26.7321, 42.2679,
            117.268, 130.732, 6.66539, 18.3346, 22.3346, 36.6654,
            10.2036, 12.7964, 26.7964, 26.2036, 19.3989, 76.6011,
            108.601, 208.399, 7.99805, 19.002,  51.002,  102.998,
            1.99522, 6.00478, 12.0048, 25.9952, 6.27347, 2.72653,
            44.7265, 18.2735, 6.73325, 25.2667, 36.2667, 79.7333};
*-- MATRIX OF EXPECTED FREQUENCIES FOR H: U(3 FACTOR EFFECTS) = 0 --*;
IF SETUSED = 3 THEN
     MLE = {38.0956, 25.4529, 223.575, 111.877, 4.66545, 5.02795,
            18.9839, 15.3227, 6.9142,  2.46822, 25.7366, 6.88094,
            43.9459, 50.4298, 203.083, 174.541, 24.6599, 39.9453,
            114.041, 138.354, 6.41531, 16.7621, 20.5699, 40.2527,
            11.3435, 9.81755, 33.272,  21.5669, 27.5124, 76.5439,
            100.186, 208.758, 8.86986, 16.9765, 63.7586, 91.3951,
            2.2336,  6.8956,  11.1319, 25.7389, 7.14745, 7.30907,
            32.5862, 24.9573, 6.19691, 20.371,  35.0755, 86.3566};
*-- MATRIX OF EXPECTED FREQUENCIES FOR H: U(2 FACTOR EFFECTS) = 0 --*;
IF SETUSED = 2 THEN
     MLE = {35.4563, 40.5594, 139.086, 159.104, 6.87782, 7.86771,
            26.9799, 30.863,  7.51026, 8.59118, 29.4608, 33.701,
            40.8321, 46.7089, 160.174, 183.227, 32.974,  37.7198,
            129.349, 147.965, 6.3963,  7.31689, 25.0991, 28.7023,
            6.98446, 7.9897,  27.3983, 31.3416, 37.9734, 43.4388,
            148.96,  170.399, 16.5611, 18.9447, 64.965,  74.3151,
            3.21252, 3.67489, 12.6019, 14.4156, 3.50793, 4.01281,
            13.7607, 15.7412, 19.072,  21.817,  74.8148, 85.5825};

*********************************************************************;
************************   END  OF USER INPUT  ************************;
*********************************************************************;

*-- MATRIX OF EXPECTED FREQUENCIES FOR H: U(MAIN EFFECTS) = 0 -------*;
  IF SETUSED = 1 THEN MLE = J(UTOT,1,1)*(SUM(OBS)/UTOT);
```

```
*-- COMPUTES THE F VECTOR AND THE G MATRIX -------------------------*;
  F = TEMPM` * (OBS - MLE);
  G = TEMPU * INV(DIAG(MLE)) * TEMPU`;
*-- FREE WORKSPACE ---------------------------------------------------*;
  FREE OBS MLE;
  FINISH;
************************  END OF MODULE SET  *************************;
```

Appendix A.5: CONTRIB module

MODULE : CONTRIB.

PROCESSING : Computes the chi-square component value (csquare value) of all the factor effects.

CALLED BY : MAIN module.

MODULES CALLED : CSQUARE.

VARIABLES USED :

COUNT1	:	Counters for processing.
COUNT2	:	
CSQUTOT	:	Total CSQUARE value of a factor effect.
DEV	:	Vector containing the chi-square component value of each factor effect.
DF	:	Degrees of freedom of a factor effect.
ORDER	:	Vector containing the degrees of freedom of each factor effect.
PVAL	:	Vector containing the P-value corresponding to the chi-square component value of each factor effect.
VARS	:	Total number of variables.

```
*********************  START OF MODULE CONTRIB  **********************;
START CONTRIB;
*-- COMPUTES THE CHI-SQUARE COMPONENT VALUE OF EACH FACTOR EFFECT ---*;
  DO COUNT1  = 1 TO (2**VARS)-1;
*-- INITIALIZE THE TOTAL CSQUARE VALUE AND DEGREES OF FREEDOM -------*;
*-- OF THE CURRENT FACTOR EFFECT ------------------------------------*;
    CSQUTOT = 0;
    DF = 0;
*-- COMPUTES THE CHI-SQUARE COMPONENT VALUE OF THE CURRENT ----------*;
*-- FACTOR EFFECT ADDING THE CSQUARE VALUE FOR EACH LEVEL OF THE-----*;
*-- FACTOR EFFECT ---------------------------------------------------*;
    DO COUNT2 = 1 TO ORDER(|(2**VARS)-COUNT1|);
*-- EXECUTES THE CSQUARE MODULE -------------------------------------*;
      RUN CSQUARE;
      END;
*-- SAVES THE CHI-SQUARE COMPONENT VALUE IN THE DEVOLUTION ----------*;
*-- VECTOR AND THE CORRESPONDING P-VALUE IN THE PVALUE VECTOR -------*;
    DEV(|(2**VARS)-COUNT1|) = CSQUTOT;
    PVAL(|(2**VARS)-COUNT1|) = 1 - PROBCHI(CSQUTOT,DF);
    END;
*-- FREE WORKSPACE --------------------------------------------------*;
  FREE COUNT1 COUNT2 CSQUTOT DF;
  FINISH;
**********************  END OF MODULE CONTRIB * *********************;
```

Appendix A.6: CSQUARE module

MODULE : CSQUARE.

PROCESSING : Computes the CSQUARE value of a u-parameter.

CALLED BY : CONTRIB module, RANKING module.

MODULES CALLED : none.

VARIABLES USED :

CSQUTOT	:	Total CSQUARE value of a factor effect.
DF	:	Counter for the degrees of freedom of a factor effect.
f	:	Score function for u-parameters.
G	:	Inverse of the covariance matrix of f.
NU	:	Number of u-parameters remaining to be ordered.
UTOT	:	Total number of u-parameters.

```
********************  START OF MODULE CSQUARE  ************************;
START CSQUARE;
*-- INCREMENTS THE TOTAL CSQUARE VALUE OF THE CURRENT EFFECT ---------*;
  CSQUTOT = CSQUTOT + ((G(|NU,|) * F) ** 2) / (G(|NU,NU|)*(UTOT**2));
*-- INCREMENTS THE DEGREES OF FREEDOM ---------------------------------*;
  DF = DF + 1;
*-- DEVOLUTES THE G MATRIX --------------------------------------------*;
  G = G(|1:NU-1,1:NU-1|) -
      (INV(G(|NU,NU|)) * (G(|1:(NU-1),NU|) * G(|NU,1:(NU-1)|)));
*-- REDUCES THE F VECTOR ----------------------------------------------*;
  F = F(|1:NU-1,|);
*-- REDUCES THE NUMBER OF U-PARAMETERS REMAINING TO BE ORDERED -------*;
  NU = NU - 1;
  FINISH;
********************   END OF MODULE CSQUARE  ************************;
```

Appendix A.7: RANKING module

MODULE : RANKING.

PROCESSING : Ranks the factor effects belonging to a given class of factor effects.

CALLED BY : MAIN module.

MODULES CALLED : SET module, CSQUARE module, MOVE module.

VARIABLES USED :

COUNTER : Total number of u-parameters + constant term that preceed the class of factor effects we are trying to rank.

COUNTERS: Vector containing the number of factor effects for each level of factor effects.

COUNT1 : Counter variables for processing.

COUNT2 :

CSQUTOT : Total CSQUARE value of a factor effect

DF : Degrees of freedom of a factor effect.

LARGPVAL: Indicator of the current largest P-value.

LARGEST : Indicator for the factor effect having the largest P-value (smallest contribution).

M : M matrix.

NEXTTERM: Position of the factor effect following the position in which we are trying to rank the factor effects (relative to all the factor effects)(for example, if our effects are A B C AB AC BC ABC, then if POSITION=3 (AC) then NEXTTERM=6 (BC)).

NUMTERMS: Total number of terms (factor effects) remaining to be ranked in the class of factor effects we are trying to rank.

NU : Total number of u-parameters remaining to be ordered.

ORDER : Vector containing the degrees of freedom of each factor effect.

POSITION: Column indicator for the position in which we are trying to rank the factor effects (relative to all the factor effects)(for example, if our factor effects are A B C AB AC BC ABC, then ABC is in POSITION=1, BC in POSITION=2, and so on up to A where POSITION=7).

POSMOVE : Column indicator for the position of the effect we want to move (relative to all the other factor effects)(for example, if our factor effects are A B C AB AC BC ABC, then A is in POSMOVE=1, B in POSMOVE=2, and so on up to ABC where POSMOVE=7).

PVALTEMP: Temporary P-value.

SETUSED : Indicator for the current set of expected counts in use.

TEMP1 : Temporary variables for processing.

TEMP2 :

TEMPM : Temporary M matrix representing the current ranking of u-parameters.

TEMPU : Temporary U matrix representing the inverse of the TEMPM matrix multiplied by the factor UTOT.

TOTNEXTU: Total number of u-parameters starting with the ones associated with NEXTTERM up to the last one associated with the highest factor effect.

U : U matrix.

UTOT : Total number of u-parameters.

VARS : Total number of variables.

```
****************** START OF MODULE RANKING ***********************;
START RANKING;
*-- NOTE: IF WE FIT A LOG-LINEAR MODEL WITH A FIXED MARGINAL TOTAL ---*;
*--       WE HAVE ONE LESS FACTOR EFFECT TO RANK --------------------*;
  NEXTTERM = 2**VARS +1 - POSITION;
  TOTNEXTU = SUM(ORDER(|NEXTTERM:(2**VARS - 1)|));
  IF SETUSED = 1 THEN DO;
    NUMTERMS = SUM(COUNTERS(|1:6|)) + 1 - POSITION;
    COUNTER = 1;
    END;
  ELSE IF SETUSED = 2 THEN DO;
    IF (VARS = 3 & FIXEDMAR = 'YES') THEN DO;
      NUMTERMS = SUM(COUNTERS(|2:6|)) - POSITION;
      COUNTER = 1 + SUM(ORDER(|1:COUNTERS(|1|)+1|));
      END;
    ELSE DO;
      NUMTERMS = SUM(COUNTERS(|2:6|)) + 1 - POSITION;
      COUNTER = 1 + SUM(ORDER(|1:COUNTERS(|1|)|));
      END;
    END;
  ELSE IF SETUSED = 3 THEN DO;
    IF (VARS = 4 & FIXEDMAR = 'YES') THEN DO;
      NUMTERMS = SUM(COUNTERS(|3:6|)) - POSITION;
      COUNTER = 1 + SUM(ORDER(|1:SUM(COUNTERS(|1:2|))+1|));
      END;
    ELSE DO;
      NUMTERMS = SUM(COUNTERS(|3:6|)) + 1 - POSITION;
      COUNTER = 1 + SUM(ORDER(|1:SUM(COUNTERS(|1:2|))|));
      END;
    END;
  ELSE IF SETUSED = 4 THEN DO;
    IF (VARS = 5 & FIXEDMAR = 'YES') THEN DO;
      NUMTERMS = SUM(COUNTERS(|4:6|)) - POSITION;
      COUNTER = 1 + SUM(ORDER(|1:SUM(COUNTERS(|1:3|))+1|));
      END;
    ELSE DO;
      NUMTERMS = SUM(COUNTERS(|4:6|)) + 1 - POSITION;
      COUNTER = 1 + SUM(ORDER(|1:SUM(COUNTERS(|1:3|))|));
      END;
    END;
  ELSE IF SETUSED = 5 THEN DO;
    IF (VARS = 6 & FIXEDMAR = 'YES') THEN DO;
      NUMTERMS = SUM(COUNTERS(|5:6|)) - POSITION;
      COUNTER = 1 + SUM(ORDER(|1:SUM(COUNTERS(|1:4|))+1|));
      END;
    ELSE DO;
      NUMTERMS = SUM(COUNTERS(|5:6|)) + 1 - POSITION;
      COUNTER = 1 + SUM(ORDER(|1:SUM(COUNTERS(|1:4|))|));
      END;
    END;
  TEMPM = M;
  TEMPU = U;
  LARGPVAL = -1;
*-- DETERMINES WHICH FACTOR EFFECT HAS THE SMALLEST CONTRIBUTION-----*;
*-- IN A GIVEN POSITION ---------------------------------------------*;
  DO COUNT1 = 1 TO NUMTERMS;
    NU = UTOT;
*-- EXECUTES MODULE SET ----------------------------------------------*;
    RUN SET;
    CSQUTOT = 0;
    DF = 0;
```

```
*-- DEVOLUTES THE G MATRIX AND REDUCES THE F VECTOR OF ALL THE ------*;
*-- FACTOR EFFECTS PRECEEDING THE TARGET POSITION IN WHICH WE ARE ---*;
*-- TRYING TO DETERMINE THE SMALLEST CONTRIBUTION -------------------*;
    DO COUNT2 = 1 TO TOTNEXTU;
*-- EXECUTES MODULE CSQUARE -----------------------------------------*;
      RUN CSQUARE;
      END;
    CSQUTOT = 0;
    DF = 0;
*-- COMPUTES THE CONTRIBUTION OF THE FACTOR EFFECT WHO IS IN THE -----*;
*-- POSITION IN WHICH WE ARE TRYING TO RANK THE FACTOR EFFECTS -------*;
    DO COUNT2 = 1 TO ORDER(|NEXTTERM-COUNT1|);
*-- EXECUTES MODULE CSQUARE -----------------------------------------*;
      RUN CSQUARE;
      END;
    PVALTEMP = 1 - PROBCHI(CSQUTOT,DF);
*-- VERIFIES IF THE FACTOR EFFECT IS PRESENTLY THE ONE WITH THE ------*;
*-- SMALLEST CONTRIBUTION -------------------------------------------*;
    IF PVALTEMP > LARGPVAL THEN DO;
      LARGPVAL = PVALTEMP;
      LARGEST = COUNT1;
      END;
*-- GETS THE NEXT UNRANKED FACTOR EFFECT TO BE MOVED TO THE POSITION -*;
*-- IN WHICH WE ARE TRYING TO RANK THE FACTOR EFFECTS ----------------*;
    POSMOVE = NEXTTERM - COUNT1 - 1;
    TEMPM = M;
    TEMPU = U;
*-- EXECUTES MODULE MOVE IF WE HAVE NOT COMPUTED THE CONTRIBUTION ----*;
*-- FOR EACH UNRANKED TERM IN THE POSITION IN WHICH WE ARE TRYING ----*;
*-- TO RANK THE FACTOR EFFECTS --------------------------------------*;
    IF COUNT1 ¬= NUMTERMS THEN RUN MOVE;
    END;
*-- IF THE FACTOR EFFECT WITH THE SMALLEST CONTRIBUTION WAS THE ------*;
*-- FIRST ONE COMPUTED, NO CHANGE IN THE INITIAL RANKING HAS TO BE ---*;
*-- DONE. OTHERWISE WE MODIFY THE RANKING IN THE M AND U MATRICES ----*;
*-- AND IN THE ORDER AND ORDER2 VECTORS (ACCORDING TO THE EFFECT -----*;
*-- WITH THE SMALLEST CONTRIBUTION)----------------------------------*;
  IF LARGEST ¬= 1 THEN DO;
    POSMOVE = (2**VARS) + 1 - POSITION - LARGEST;
    TEMP1 = ORDER(|POSMOVE|);
    TEMP2 = ORDER2(|POSMOVE|);
*-- EXECTUES MODULE MOVE --------------------------------------------*;
    RUN MOVE;
    M = TEMPM;
    U = TEMPU;
    COUNT1 = POSMOVE;
    DO COUNT2 = 1 TO (LARGEST - 1);
      ORDER(|COUNT1|) = ORDER(|COUNT1+1|);
      ORDER2(|COUNT1|) = ORDER2(|COUNT1+1|);
      COUNT1 = COUNT1 + 1;
      END;
    ORDER(|COUNT1|) = TEMP1;
    ORDER2(|COUNT1|) = TEMP2;
    END;
*-- FREE WORKSPACE --------------------------------------------------*;
    FREE NEXTTERM TOTNEXTU NUMTERMS COUNTER COUNT1 COUNT2
         LARGPVAL LARGEST PVALTEMP TEMP1 TEMP2 CSQUTOT DF;
  FINISH;
**********************  END OF MODULE RANKING  ***********************;
```

Appendix A.8: MOVE module

MODULE : MOVE.

PROCESSING : Moves the columns of the M matrix and rows of the U matrix corresponding to the desired ranking of u-parameters.

CALLED BY : RANKING module.

MODULES CALLED : none.

VARIABLES USED :

AUXM	:	Auxiliary variable for moving columns of the matrix M.
AUXU	:	Auxiliary variable for moving rows of the matrix U.
COUNTA	:	Counter variables for processing.
COUNtB	:	
COUNTER	:	Total number of u-parameters + constant term that preceed the class of factors we are trying to rank.
COUNTERS	:	Vector containing the number of factor effects for each level of factor effects.
ORDER	:	Vector containing the degrees of freedom of each factor effect.
POS1	:	Column indicator for the position of the factor effect we want to move (relative to the class of factor effects we are working with).
POS2	:	Column indicator for the position we we want to move the factor effect to (relative to the class of factor effects we are working with)(for example, if we are working with the class of 1-factor effects A B C and we want to move A in position C, then POS1 = 1 and POS2 = 3).
POSITION	:	Column indicator for the position in which we are trying to rank the factor effects (relative to all the factor effects)(for example, if our factor effects are A B C AB AC BC ABC, then ABC is in POSITION=1, BC in POSITION=2, and so on up to A where POSITION=7).
POSMOVE	:	Column indicator for the position of the effect we want to move (relative to all the other factor effects)(for example, if our factor effects are A B C AB AC BC ABC, then A is in POSMOVE=1, B in POSMOVE=2, and so on up to ABC where POSMOVE=7).
SETUSED	:	Indicator for the current set of expected counts in use.
TEMPM	:	Temporary M matrix representing the current ranking of u-parameters.
TEMPORD	:	Vector containing the degrees of freedom corresponding to the class of factor effects we are trying to order.
TEMPU	:	Temporary U matrix representing the inverse of the TEMPM matrix multiplied by the factor UTOT.

```
********************** START OF MODULE MOVE ************************;
START MOVE;
*-- DETERMINE THE POSITION TO MOVE THE FACTOR EFFECT TO (THE ---------*;
*-- POSITION IS RELATIVE TO THE CLASS OF FACTOR EFFECTS WE ARE -------*;
*-- CURRENTLY USING --------------------------------------------------*;
*-- NOTE: IF WE FIT A LOG-LINEAR MODEL WITH A FIXED MARGINAL TOTAL, --*;
*--       WE HAVE ONE LESS FACTOR EFFECT TO RANK ---------------------*;
  IF SETUSED = 1 THEN DO;
    TEMPORD = ORDER(|1:COUNTERS(|1|)|);
    COUNTA = 1;
    POS2 = NROW(TEMPORD) - POSITION + SUM(COUNTERS(|2:6|)) + 1;
    END;
  ELSE IF SETUSED = 2 THEN DO;
    IF (VARS = 3 & FIXEDMAR = 'YES') THEN DO;
      TEMPORD = ORDER(|2+COUNTERS(|1|):SUM(COUNTERS(|1:2|))|);
      COUNTA = 2 + COUNTERS(|1|);
      END;
    ELSE DO;
      TEMPORD = ORDER(|1+COUNTERS(|1|):SUM(COUNTERS(|1:2|))|);
      COUNTA = 1 + COUNTERS(|1|);
      END;
    POS2 = NROW(TEMPORD) - POSITION + SUM(COUNTERS(|3:6|)) + 1;
    END;
  ELSE IF SETUSED = 3 THEN DO;
    IF (VARS = 4 & FIXEDMAR = 'YES') THEN DO;
      TEMPORD = ORDER(|2+SUM(COUNTERS(|1:2|)):SUM(COUNTERS(|1:3|))|);
      COUNTA = 2 + SUM(COUNTERS(|1:2|));
      END;
    ELSE DO;
      TEMPORD = ORDER(|1+SUM(COUNTERS(|1:2|)):SUM(COUNTERS(|1:3|))|);
      COUNTA = 1 + SUM(COUNTERS(|1:2|));
      END;
    POS2 = NROW(TEMPORD) - POSITION + SUM(COUNTERS(|4:6|)) + 1;
    END;
  ELSE IF SETUSED = 4 THEN DO;
    IF (VARS = 5 & FIXEDMAR = 'YES') THEN DO;
      TEMPORD = ORDER(|2+SUM(COUNTERS(|1:3|)):SUM(COUNTERS(|1:4|))|);
      COUNTA = 2 + SUM(COUNTERS(|1:3|));
      END;
    ELSE DO;
      TEMPORD = ORDER(|1+SUM(COUNTERS(|1:3|)):SUM(COUNTERS(|1:4|))|);
      COUNTA = 1 + SUM(COUNTERS(|1:3|));
      END;
    POS2 = NROW(TEMPORD) - POSITION + SUM(COUNTERS(|5:6|)) + 1;
    END;
  ELSE IF SETUSED = 5 THEN DO;
    IF (VARS = 6 & FIXEDMAR = 'YES') THEN DO;
      TEMPORD = ORDER(|2+SUM(COUNTERS(|1:4|)):SUM(COUNTERS(|1:5|))|);
      COUNTA = 2 + SUM(COUNTERS(|1:4|));
      END;
    ELSE DO;
      TEMPORD = ORDER(|1+SUM(COUNTERS(|1:4|)):SUM(COUNTERS(|1:5|))|);
      COUNTA = 1 + SUM(COUNTERS(|1:4|));
      END;
    POS2 = NROW(TEMPORD) - POSITION + COUNTERS(|6|) + 1;
    END;
*-- DETERMINE THE POSITION OF THE FACTOR EFFECT WE WANT TO MOVE ------*;
```

```
*-- (THE POSITION IS RELATIVE TO THE CLASS OF FACTOR EFFECTS WE ARE --*;
*-- WORKING WITH) ---------------------------------------------------*;
  POS1 = POSMOVE - COUNTA + 1;
  TEMPORD = TEMPORD(|1:POS2|);
  IF POS1 = 1 THEN COUNTB = 0;
  ELSE COUNTB = SUM(TEMPORD(|1:POS1-1|));
*-- MOVE THE CORRESPONDING COLUMNS OF THE M MATRIX TO REPRESENT THE --*;
*-- DESIRED RANKING OF THE U-PARAMETERS ----------------------------*;
  AUXM = TEMPM(|,(COUNTER+1+COUNTB):
                 (COUNTER+SUM(TEMPORD(|1:POS1|)))|);
  TEMPM(|,(COUNTER+1+COUNTB):
           (COUNTER+COUNTB+SUM(TEMPORD(|(POS1+1):NROW(TEMPORD)|)))|) =
  TEMPM(|,(COUNTER+1+SUM(TEMPORD(|1:POS1|))):
           (COUNTER+SUM(TEMPORD))|);
  TEMPM(|,(COUNTER+1+COUNTB+SUM(TEMPORD(|POS1+1:NROW(TEMPORD)|))):
           (COUNTER+SUM(TEMPORD))|) = AUXM;
*-- MOVE THE CORRESPONDING ROWS OF THE U MATRIX TO REPRESENT THE -----*;
*-- DESIRED RANKING OF THE U-PARAMETERS ----------------------------*;
*-- NOTE: MOVING COLUMNS IN THE M MATRIX MUST BE REFLECTED IN THE U --*;
*-------- MATRIX BY MOVING ITS ROWS ACCORDINGLY ---------------------*;
  AUXU = TEMPU(|(COUNTER+1+COUNTB):
                 (COUNTER+SUM(TEMPORD(|1:POS1|))),|);
  TEMPU(|(COUNTER+1+COUNTB):
           (COUNTER+COUNTB+SUM(TEMPORD(|(POS1+1):NROW(TEMPORD)|))),|) =
  TEMPU(|(COUNTER+1+SUM(TEMPORD(|1:POS1|))):
           (COUNTER+SUM(TEMPORD)),|);
  TEMPU(|(COUNTER+1+COUNTB+SUM(TEMPORD(|POS1+1:NROW(TEMPORD)|))):
           (COUNTER+SUM(TEMPORD)),|) = AUXU;
*-- FREE WORKSPACE --------------------------------------------------*;
  FREE TEMPORD COUNTA COUNTB POS1 POS2 AUXM AUXU;
  FINISH;
**********************   END  OF MODULE MOVE  ************************;
```

Appendix A.9: PRINT module

MODULE : PRINT.

CALLED BY : MAIN module.

MODULES CALLED : none.

PROCESSING : Prints the result of the analysis giving the final ordering of the factor effects, their degrees of freedom, their CSQUARE value under each hypothesis and their corresponding P-value.

VARIABLES USED :

- ORDER : Vector containing the degrees of freedom of each factor effect.
- ORDER2 : Vector containing the ordering of the factor effects.
- ORDERING: Matrix containing the results of the RX^2 analysis.
- VARS : Total number of variables.

```
********************** START OF MODULE PRINT ************************;
START PRINT;
*-- THE FIRST COLUMN OF THE DEVOLUTION TABLE CONTAINS THE ORDER ------*;
*-- OF THE EFFECTS (THE NUMBER OF LEVELS OF EACH EFFECT CORRESPONDS---*;
*-- TO THE DEGREES OF FREEDOM) ---------------------------------------*;
  ORDERING(|,1|) = ORDER;
*-- THE LAST ELEMENT OF THE COLUMN CORRESPONDS TO THE TOTAL DEGREES --*;
*-- OF FREEDOM -------------------------------------------------------*;
  ORDERING(|2**VARS,1|) = SUM(ORDERING(|,1|));
*-- THE LAST ELEMENT OF THE COLUMN CORRESPONDS TO THE TOTAL CSQUARE --*;
  ORDERING(|2**VARS,2|) = SUM(ORDERING(|,2|));
*-- THE LAST ELEMENT OF THE COLUMN CORRESPONDS TO THE P-VALUE OF -----*;
*-- TOTAL CHI-SQUARE VALUE OF THE PREVIOUS COLUMN -------------------*;
  ORDERING(|2**VARS,3|) =
    1 - PROBCHI(ORDERING(|2**VARS,2|),ORDERING(|2**VARS,1|));
  ORDERING(|2**VARS,4|) = SUM(ORDERING(|,4|));
  ORDERING(|2**VARS,5|) =
    1 - PROBCHI(ORDERING(|2**VARS,4|),ORDERING(|2**VARS,1|));
  IF VARS > 2 THEN DO;
    ORDERING(|2**VARS,6|) = SUM(ORDERING(|,6|));
    ORDERING(|2**VARS,7|) =
      1 - PROBCHI(ORDERING(|2**VARS,6|),ORDERING(|2**VARS,1|));
    END;
  IF VARS > 3 THEN DO;
    ORDERING(|2**VARS,8|) = SUM(ORDERING(|,8|));
    ORDERING(|2**VARS,9|) =
      1 - PROBCHI(ORDERING(|2**VARS,8|),ORDERING(|2**VARS,1|));
    END;
  IF VARS > 4 THEN DO;
    ORDERING(|2**VARS,10|)= SUM(ORDERING(|,10|));
    ORDERING(|2**VARS,11|)=
      1- PROBCHI(ORDERING(|2**VARS,10|),ORDERING(|2**VARS,1|));
    END;
  IF VARS > 5 THEN DO;
    ORDERING(|2**VARS,12|)= SUM(ORDERING(|,12|));
    ORDERING(|2**VARS,13|)=
      1- PROBCHI(ORDERING(|2**VARS,12|),ORDERING(|2**VARS,1|));
    END;
*-- PRINT THE FINAL DEVOLUTION TABLE ---------------------------------*;
  PRINT ' -------------------------------------------------------- ',
        '|  RX-SQUARE RANKING AND CHI-SQUARE COMPONENT VALUES  |',
        ' -------------------------------------------------------- ',,;
  IF VARS = 6 THEN PRINT
    'RANKING OF TERMS BASED ON COMBINED H1, H2, H3, H4, H5 & H6',
    '----------------------------------------------------------',,,,
    ORDERING(|ROWNAME=ORDER2 COLNAME={'   DF   ' '   H6   ' 'P-VALUE'
                 '   H5   ' 'P-VALUE' '   H4   ' 'P-VALUE' '   H3   '
                 'P-VALUE' '   H2   ' 'P-VALUE' '   H1   ' 'P-VALUE'}|),,,
    'NOTE:    H6: U(6 FACTOR EFFECTS) = 0',
    '         H5: U(5 FACTOR EFFECTS) = 0',
    '         H4: U(4 FACTOR EFFECTS) = 0',
    '         H3: U(3 FACTOR EFFECTS) = 0',
    '         H2: U(2 FACTOR EFFECTS) = 0',
    '         H1: U(MAIN EFFECTS) = 0    ',,,,,,;
  ELSE IF VARS = 5 THEN PRINT
    'RANKING OF TERMS BASED ON COMBINED H1, H2, H3, H4 & H5',
    '------------------------------------------------------',,,,
```

```
   ORDERING(|ROWNAME=ORDER2 COLNAME={'   DF   ' '   H5   ' 'P-VALUE'
              '  H4   ' 'P-VALUE' '  H3   ' 'P-VALUE' '  H2   '
              'P-VALUE' '  H1   ' 'P-VALUE'}|),,,
   'NOTE:     H5: U(5 FACTOR EFFECTS) = 0',
   '          H4: U(4 FACTOR EFFECTS) = 0',
   '          H3: U(3 FACTOR EFFECTS) = 0',
   '          H2: U(2 FACTOR EFFECTS) = 0',
   '          H1: U(MAIN EFFECTS) = 0    ',,,,,,;
  ELSE IF VARS = 4 THEN PRINT
   'RANKING OF TERMS BASED ON COMBINED H1, H2, H3 & H4',
   '--------------------------------------------------',,,,
   ORDERING(|ROWNAME=ORDER2 COLNAME={'   DF   ' '   H4   ' 'P-VALUE'
              '  H3   ' 'P-VALUE' '  H2   ' 'P-VALUE' '  H1   '
              'P-VALUE'}|),,,
   'NOTE:     H4: U(4 FACTOR EFFECTS) = 0',
   '          H3: U(3 FACTOR EFFECTS) = 0',
   '          H2: U(2 FACTOR EFFECTS) = 0',
   '          H1: U(MAIN EFFECTS) = 0    ',,,,,,;
  ELSE IF VARS = 3 THEN PRINT
   'RANKING OF TERMS BASED ON COMBINED H1, H2 & H3',
   '----------------------------------------------',,,,
   ORDERING(|ROWNAME=ORDER2 COLNAME={'   DF   ' '   H3   ' 'P-VALUE'
              '  H2   ' 'P-VALUE' '  H1   ' 'P-VALUE'}|),,,
   'NOTE:     H3: U(3 FACTOR EFFECTS) = 0',
   '          H2: U(2 FACTOR EFFECTS) = 0',
   '          H1: U(MAIN EFFECTS) = 0    ',,,,,,;
  ELSE IF VARS = 2 THEN PRINT
   'RANKING OF TERMS BASED ON COMBINED H1 & H2',
   '------------------------------------------',,,,
   ORDERING(|ROWNAME=ORDER2 COLNAME={'   DF   ' '   H2   ' 'P-VALUE'
              '  H1   ' 'P-VALUE'}|),,,
   'NOTE:     H2: U(2 FACTOR EFFECTS) = 0',
   '          H1: U(MAIN EFFECTS) = 0    ',,,,,,;
*-- FREE WORKSPACE -----------------------------------------------------*;
  FREE ORDER ORDER2 VARS ORDERING;
  FINISH;
**********************  END OF MODULE PRINT  ***************************;
```

Appendix A.10: SET module
(for logit models having a dichotomous response variable)

MODULE : SET (for logit models)

PROCESSING : Computes the f vector and G matrix for a given null hypothesis.

CALLED BY : RANKING module, MAIN module.

MODULES CALLED : none.

VARIABLES USED :

- OBS : Vector of observed counts .
- MLE : Vector of expected counts (MLE's).
- F : Score function for u-parameters.
- G : Inverse of the covariance matrix of F.
- NT : Vector of sample totals.
- PTHAT : Vector of expected proportions.
- SETUSED : Indicator for the current set of expected counts in use.
- TEMPM : Temporary M matrix representing the current ranking of U-parameters.
- TEMPU : Temporary U matrix representing the inverse of the TEMPM matrix multiplied by the factor UTOT.
- UTOT : Total number of u-parameters.

```
*************************  START OF MODULE SET  ************************
START SET;

**********************************************************************
*************************  START OF USER INPUT  ************************
**********************************************************************

**********************************************************************
***                                                                ***
***                 DETERGENT PREFERENCE DATA                      ***
***                    (USING LOGIT MODEL)                         ***
***                     (FIENBERG, P.79)                           ***
***                   (I = 3, J = 2, K = 2)                        ***
***                                                                ***
**********************************************************************
*-- MATRIX OF OBSERVED FREQUENCIES --*;
OBS  = {19, 57, 29, 63, 23, 47, 33, 66, 24, 37, 42, 68};
*-- MATRIX OF SAMPLE TOTALS --*;
NT = { 48, 106,  56, 116,  70, 102,  56, 116,  67, 89, 72, 110};
*-- MATRIX OF EXPECTED FREQUENCIES FOR H: U(3 FACTOR EFFECTS) = 0 --*;
*-- FROM FITTING LOGIT MODEL --*;
IF SETUSED = 3 THEN DO;
MLE = {19.3165, 56.6835, 28.6835, 63.3165, 24.1379, 45.8621,
       31.8621, 67.1379, 22.5455, 38.4545, 43.4545, 66.5455};
PTHAT = MLE # (NT ## -1);
END;
*-- MATRIX OF EXPECTED FREQUENCIES FOR H: U(2 FACTOR EFFECTS) = 0 --*;
*-- FROM FITTING LOGIT MODEL --*;
IF SETUSED = 2 THEN DO;
MLE = {19.1313, 48.9065, 30.1738, 69.7884, 26.7236, 45.2763,
       29.1887, 67.8114, 26.3658, 40.5965, 38.4169, 65.6209};
PTHAT = MLE # (NT ## -1);
END;

**********************************************************************
*************************   END  OF USER INPUT  ************************
**********************************************************************

*-- MATRIX OF EXPECTED FREQUENCIES FOR H: U(MAIN EFFECTS) = 0 -------*;
  IF SETUSED = 1 THEN DO;
    PTHAT = J(UTOT,1,1) * (SUM(OBS)/SUM(NT));
    MLE = NT # PTHAT;
    END;
*-- COMPUTES THE F VECTOR AND THE G MATRIX -------------------------*;
  F = TEMPM` * (OBS - MLE);
  G = TEMPU * INV(DIAG(MLE#(1-PTHAT))) * TEMPU`;
*-- FREE WORKSPACE -------------------------------------------------*;
  FREE OBS MLE;
  FINISH;
************************  END OF MODULE SET  ************************;
```