FINDING CONDENSED DESCRIPTIONS FOR MULTI-DIMENSIONAL DATA

Nanny WERMUTH, Theo WEHNER and Herbert GÖNNER
Institut für Medizinische Statistik und Dokumentation der Johannes Gutenberg-Universität,
65 Mainz, Langenbeckstr. 1, Federal Republic of Germany

We describe two programs that may be used to find condensed descriptions for data available in a contingency table or in a covariance matrix in the case that these data follow a multinomial or a multivariate normal distribution, respectively. The programs perform a stepwise model search among multiplicative models by computing appropriate likelihood-ratio test statistics.

1. Introduction

The interrelations among several variables are usually difficult to understand unless they can be described in a condensed manner. If there exists a joint probability distribution function, then a fairly compact quantification of such interrelations is available. Condensed descriptions for joint distributions are offered by multiplicative models [6,7]. They are a subclass of log-linear models [1,2], and a subclass of covariance selection models [4] in the case that the joint distribution is a multinomial or a multivariate normal distribution, respectively. Even though multiplicative models for multinomial distributions have already been employed as data-analytic tools in extensive medical studies [3,5], a more widespread use will be likely only after the model search procedures [7] have become well known, and after the corresponding computer programs are easily available. This paper gives description of COVSEL and TASEL, two Fortran IV programs for the model search among multiplicative models.

2. Methods

A multiplicative model states how a joint distribution may be factored into marginal distributions. Suppose that the joint distribution of five variables can be factored as

\[
f(x_1, x_2, x_3, x_4, x_5) = \frac{f(x_1, x_2, x_3) f(x_1, x_4) f(x_1, x_5)}{f(x_1)}
\]

then the notation for this multiplicative model is 123/14/15. Thus, the notation for a multiplicative model shows those subgroups of variables that produce the interrelations among all variables. Furthermore, the notation tells which variable pairs are conditionally independent. They are all those pairs that do not belong jointly to one of the subgroups listed in the notation. In model 123/14/15, for instance, the pairs (2,4), (2,5), (3,4), (3,5), (4,5) are all conditionally independent given the remaining three variables. These independencies imply a relatively simple pattern of association for all variables: given variable 1, the joint variable 23 and the variables 4 and 5 are no longer interrelated, they are independent of each other.

These interpretations may seem rather abstract but they give useful and practical guidelines for the data analysis. Suppose that model 123/14/15 fits a set of data well. Then we know that the variable groups 123, 14 and 15 should be analyzed in more detail because they contain the more important interrelations. In addition, a way to present the data is suggested. The material may be displayed in groups that are homogeneous with respect to variable 1. Then, interrelations are easy to grasp, since there is only one association left, the one for pair (2,3).

Likelihood-ratio tests may be used to evaluate the
fit between the maximum-likelihood estimates for a given model and the observed data. Let \( n_{i_1, \ldots, i_p} \) and \( \hat{m}_{i_1, \ldots, i_p} \) denote the observed and the estimated cell counts for a multinomial distribution, and let \( |S| \) and \( |S'| \) be the determinants of the observed and the estimated covariance matrix for a multivariate normal distribution, and \( n \) be the sample size. Than the chi-square test statistics are [2,4,7]

\[
-2 \ln \prod_{i_1, \ldots, i_p} \left[ \frac{\hat{m}_{i_1, \ldots, i_p}}{n_{i_1, \ldots, i_p}} \right]^{n_{i_1, \ldots, i_p}} \text{, and (2)}
\]

\[
-2 \ln \left[ \frac{|S|^2}{|S'|} \right]^{n/2} , \text{ (3)}
\]

respectively.

To compute these test statistics in the case of multiplicative models, the maximum likelihood estimates need not be evaluated explicitly. Instead, for each model the above statistic may be derived as the sum of test statistics for certain zero partial associations (z.p.a.). As an example, we partition the test statistic in (3) for model 123/14/15. Let \( D_{12345}, D_{123}, D_{14} \) denote the determinants of the observed correlation matrix for all five variables and of the submatrices containing variables (1,2,3) and (1,4), only. Then we can write (3) for model 123/14/15 as

\[
-2 \ln \left[ \frac{D_{123} D_{14} D_{15} D_{1} D_{2}}{D_{12345}} \right]^{n/2} \text{ (4)}
\]

This statistic equals, for instance, the sum of the following five test statistics

\[
-2 \ln \left[ \frac{D_{123} D_{14} D_{15} D_{1} D_{2}}{D_{12345}} \right]^{n/2} + \left( -2 \ln \left[ \frac{D_{123} D_{135} D_{13}}{D_{1235}} \right]^{n/2} \right) + \left( -2 \ln \left[ \frac{D_{123} D_{135} D_{13}}{D_{1235}} \right]^{n/2} \right) + \left( -2 \ln \left[ \frac{D_{13} D_{15} D_{1}}{D_{135}} \right]^{n/2} \right) + \left( -2 \ln \left[ \frac{D_{13} D_{14} D_{1}}{D_{134}} \right]^{n/2} \right) . \text{ (5)}
\]

The first of these statistics evaluates the conditional independence of pair (2,4) in the joint distribution of all five variables (1,2,3,4,5). The second statistic evaluates the conditional independence of pair (2,5) in the marginal distribution of the variables (1,2,3,5). Another interpretation for this type of a sequence of test statistics has been demonstrated [6,7]. They represent (in that order) tests for zero partial association: of pair (2,4) given model 12345; of pair (2,5) given model 1235/1345; of pair (4,5) given model 123/1345; of pair (3,5) given model 123/134/135; of pair (3,4) given model 123/134/15.

The fact that each likelihood-ratio test statistic for a multiplicative model may be partitioned into a sequence of easily computable test statistics for z.p.a. (as in our example) is the basis for the model search procedure performed in COVSEL and TASEL.

3. Description of programs

The programs and flow charts are organized in three parts, (A) the setup for a given set of data, (B) the first selection step and preparations for the second selection step, and (C) the performance of all of the remaining selection steps in one large cycle. The first selection step need not be treated differently than the other steps. We decided to leave it separate though, to save computing time and because we wanted to use the results of the first selection step in other data-analysis programs, as well.

A.

The input cards consist of five different types of cards.

1: Beginning card, AAAA in the first four columns of the first input card;
2: a heading card;
3: information for the data cards (for COVSEL the number of variables (NVAR) and the number of observations (NOBS), for TASEL the number of categories per variable on (ICARD*);
5: the main data (for COVSEL the lower triangular correlation matrix, for TASEL the cell counts in each cell of the contingency table).

In the case that an error in the input cards is encountered, error messages are printed and the next

* A dummy “1” in the second column made this card compatible with L. Goodman’s ECTA-program (University of Chicago).
N. Wermuth et al., Multi-dimensional data description search

C O V S E L

START

CLEAR COUNTERS

READ 1st INPUT-CARD AAAAA INDICATES BEGINNING OF A SET OF DATA

EOF

YES

STOP

NO

AAAA

NO

PRINT FIRST FOUR CHARACTERS OF WRONG CARD

YES

READ AND PRINT HEADING, THE 2nd INPUT-CARD

READ AND PRINT NUMBER OF VARIABLES (NVAR), NUMBER OF OBSERVATIONS (NOBS), THE 3rd INPUT-CARD

NVAR ≤ 9

NO

PRINT ERROR MESSAGE

YES

READ FORMAT FOR DATA MATRIX THE 4th INPUT-CARD

READ AND PRINT: MATI (LOWER TRIANGULAR DATA MATRIX), THE NEXT NVAR INPUT-CARDS

NEGATIVE DIAGONAL ELEMENT

YES

PRINT ERROR MESSAGE

NO

SUBROUTINE READM2

COMPUTE DETERMINANT (DET) AND NEGATIVE INVERSE (MAT) OF MATI

SUBROUTINE: QSW

DET ≤ 10

YES

PRINT ERROR MESSAGE

NO

PRINT DET, MAT

FORM ALL NVAR (NVAR-1)/2 VARIABLE PAIRS

B
COVSEL

FIRST SELECTION STEP STARTS

FORM INDEXCOMBINATION FOR THE FIRST SELECTION STEP

COMPUTE \( \chi^2 \)'s TO TEST Z.P.A. OF EACH VARIABLE PAIR

SELECT VARIABLE PAIR BY FINDING POSITION (INDE) WITH SMALLEST \( \chi^2 \)

SUBROUTINE CHILOW

COMPUTE PROBABILITIES AND PRINT \( \chi^2 \)'s OF FIRST SELECTION STEP

SUBROUTINE: DRUCK1

PRINT \( \chi^2 \)'s OF THE SELECTED PAIR

SUBROUTINE: CDC-DEPENDENT FUNCTION: CHIQ

COMPUTE AND PRINT \( \chi^2 \) TO TEST FIT OF THE SELECTED MODEL

SUBROUTINE: CDPEND

STORE INDEXCOMBINATION OF THE SELECTED PAIR

SUBROUTINE: USPMR

FACTOR INDEXCOMBINATION OF THE SELECTED PAIR

SUBROUTINE: PERM

STORE RESULT OF PERM IN IZA AND INDE

SUBROUTINES USPZ, USPN

PRINT NOTATION FOR THE 1st SELECTED MODEL

CDC-DEPENDENT OVERPRINT

MARK THE SELECTED PAIR (V3(INDE) = 8)

SECOND SELECTION STEP STARTS

MARK ALL VARIABLE PAIRS TO BE EXCLUDED IN 2nd SELECTION STEP (V3(.) = 9)

SUBROUTINE: RED

FIND INDEXCOMBINATION FOR EACH PAIR INCLUDED IN 2nd STEP

SUBROUTINE: TESTIN
C O V S E L

1. **CONTINUE**

2. **COMPUTE \( x^2 \)'S TO TEST FOR ADDITIONAL PAIRS FOR EACH PAIR INCLUDED IN THIS SELECTION PAIR**

3. **SELECT VARIABLE PAIR BY FINDING POSITION (INDE) OF SMALLEST \( x^2 \)**

4. **PRINT \( x^2 \)'S OF THIS SELECTION STEP**

5. **PRINT THE SELECTED PAIR**

6. **COMPUTE AND PRINT \( x^2 \) TO TEST FIT OF THE SELECTED MODEL**

7. **STORE INDEXCOMBINATION OF THE SELECTED PAIR**

8. **FACTOR INDEXCOMBINATION OF THE SELECTED PAIR**

9. **STORE INPUT AND RESULT OF PERM IN IZA AND INE**

10. **MARK THE SELECTED PAIR (V3(INDE) = 9) AND NULLIFY EXCLUSIONS (V9(1) = 9) OF PREVIOUS STEP**

11. **PRINT NOTATION FOR MUTUAL INDEPENDENCE MODEL**

12. **CANCEL ALL INDEXCOMBINATIONS IN IZA THAT ARE CONTAINED IN INE AND PRINT NOTATION FOR THE SELECTED MODEL**

13. **MARK ALL VARIABLE PAIRS TO BE EXCLUDED IN THE NEXT SELECTION STEP (V3(1) = 9)**

14. **FIND INDEXCOMBINATIONS IN IZA FOR EACH PAIR INCLUDED IN THE NEXT SELECTION STEP**

15. **END**
TASEL

START

CLEAR COUNTERS

READ 1st INPUT-CARD
'AAAA' INDICATES BEGINNING OF A SET OF DATA

EOF

YES STOP

NO AAAAA

YES PRINT FIRST FOUR CHARACTERS OF WRONG CARD

READ AND PRINT:
HEADING THE 2nd INPUT-CARD

READ(ICARD) NUMBER OF CATEGORIES FOR EACH VARIABLE, THE 3rd INPUT-CARD

COMPUTE NUMBER OF VARIABLES(NVAR), NUMBER OF CELLS(NZELL) FROM ICARD

NVAR ≤ 9

NO PRINT ERROR MESSAGE

YES NZELL ≤ 1000

NO PRINT ERROR MESSAGE

YES

READ FORMAT FOR THE DATA TABLE, THE 4th INPUT-CARD

READ(TABLE) DATA TABLE, LAST INPUT CARDS OF THIS SET OF DATA

COMPRESS THE DATA TABLE, LAST INPUT CARDS OF THIS SET OF DATA

COMPUTE AND PRINT NUMBER OF OBSERVATIONS(NOBS), PRINT ALSO NVAR, ICARD, TAB

CDC-DEPENDENT VARIABLE FORMAT

FORM ALL NVAR·(NVAR-1)/2 VARIABLE PAIRS

B
TASEL

FIRST SELECTION STEP STARTS

FORM INDEXCOMBINATION FOR THE FIRST SELECTION STEP

COMPUTE \( \chi^2 \)'S TO TEST Z.P.A. OF EACH VARIABLE PAIR

SUBROUTINE: CHICR, CHITA

PRINT \( \chi^2 \)'S OF FIRST SELECTION STEP

SUBROUTINE: DRUCK2

SELECT VARIABLE PAIR BY FINDING POSITION (INDE) WITH HIGHEST PROBABILITY

SUBROUTINE: PRHIGH

MARK THE SELECTED PAIR (V3(INDE)=8)

PRINT THE SELECTED PAIR

PRINT \( \chi^2 \) TO TEST FIT OF THE SELECTED MODEL

SUBROUTINE: CDC-DEPENDENT FUNCTION: CHIQ (COMPUTES PROBABILITIES)

STORE INDEXCOMBINATION OF THE SELECTED PAIR

SUBROUTINE: USPMR

FACTOR INDEXCOMBINATION OF THE SELECTED PAIR

SUBROUTINE: PERM

STORE RESULT OF PERM IN IZA AND INE

SUBROUTINES: USP2, USPN

PRINT NOTATION FOR THE 1ST SELECTED MODEL

SUBROUTINE: CDC-DEPENDENT OVERPRINT

SECOND SELECTION STEP STARTS

MARK ALL VARIABLE PAIRS TO BE EXCLUDED IN 2nd SELECTION STEP (V3(J)=9)

SUBROUTINE: RED

FIND INDEXCOMBINATION FOR EACH PAIR INCLUDED IN 2nd STEP

SUBROUTINE: TESTIN
TASEL

CONTINUE

SUBROUTINE: CHIR, CHITA

SUBROUTINE: DRUCK2

SUBROUTINE: PRHIGH

SELECT VARIABLE PAIR BY FINDING POSITION (INDE) WITH HIGHEST PROBABILITY

MARK THE SELECTED PAIR (V3(INDE)=8)

PRINT THE SELECTED PAIR

SUBROUTINE: USP3MR

SUBROUTINE: PERM

SUBROUTINES: USP, USP3, USPN

STORE INPUT AND RESULT OF PERM IN IZA AND IN E

NULLIFY EXCLUSIONS (V3(I)=9) OF PREVIOUS STEP

CANCEL ALL INDEXCOMBINATIONS IN IZA THAT ARE CONTAINED IN E, AND PRINT NOTATION FOR THE SELECTED MODEL

MARK ALL VARIABLE PAIRS TO BE EXCLUDED IN THE NEXT SELECTION STEP (V3(I)=9)

SUBROUTINE: ZEMLNLEU

SUBROUTINE: RED

FIND INDEXCOMBINATIONS IN IZA THAT BELONG TO EACH PAIR INCLUDED IN THE NEXT SELECTION STEP

SUBROUTINE: CDC-DEPENDENT VARIABLE FORMAT

SUBROUTINE: CDC-DEPENDENT FUNCTION CHIQ COMPUTES PROBABILITIES

999

999 BIG CYCLE STARTS

COMPUTE \( \chi^2 \)' S TO TEST FOR ADDITIONAL Z.P.A. FOR EACH PAIR INCLUDED IN THIS SELECTION STEP

PRINT \( \chi^2 \)' S OF THIS SELECTION STEP
AAA card, that is the next set of data, is being searched for.
B and C.

For a set of data with NVAR variables, there are NVAR (NVAR-1)/2 variable pairs and the same number of selection steps. In each step one variable pair is selected to have zero partial association. At the same time the notation for the corresponding selected multiplicative model is printed, as well as its goodness-of-fit statistic (for COVSEL (3), for TASEL (4)). The decision on the fit or on the lack of fit of a selected model is left to the user (compare [7] and the sample run). The goodness-of-fit statistics are sums of chi-square statistics for z.p.a.

We need some notation to describe the actual computation of a chi-square statistic for z.p.a. [7] at any selection step. Let \((i,j)\) denote the indices of a variable pair and let \((ijK)\) be an index combination in which \(K\) contains indices of some or all other variables. Then, for COVSEL we compute

\[
-2 \ln \left[ 1 - \frac{(r_{ij})^2}{r_{ii} r_{jj}} \right]
\]

where \(r_{ii}, r_{jj}\) are those elements in the inverse correlation matrix of variables \((ijK)\) that correspond to the pair \((i,j)\). For TASEL, we compute

\[
2 \sum_{ijK} n_{ijK} \ln n_{ijK} - \sum_{iK} n_{i,K} \ln n_{i,K} \\
+ \sum_{K} n_{..K} \ln n_{..K} - \sum_{K} n_{..,K} \ln n_{..,K}
\]

where \(n_{ijK}\) denote cell counts in the (marginal) table of variables \((ijK)\) and

\[
n_{..,K} = \sum_{i} n_{i,K} = \sum_{j} n_{ij,K}.
\]

The pair \((i,j)\) is actually selected when it is the most likely pair to have z.p.a. (among all those pairs included in that selection step). After the selection of pair \((i,j)\), the variables \((ijK)\) will no longer be investigated jointly but only in subgroups. The index combination \((ijK)\) will be factored as

\[
(ijK) \rightarrow (K)(ijK) / (K).
\]

This kind of factoring gives the index combinations for subgroups of variables, and it leads to the notation for the selected multiplicative models. After the first selection step, the index combination on the left hand side in (8) and in the denominator are stored in INE, those of the numerator are stored in IZA. After cancelling terms contained in INE as well as in IZA, the content of IZA gives the model notation; the content of INE tells which variable pairs cannot be included in the next selection step (because their z.p.a. would not lead to a multiplicative model).

The vector V3 stores information on the variable pairs. If, for instance pair \((1,2)\) is not yet selected to have z.p.a., then \(V3(1) = 0\), if pair \((1,2)\) is temporarily excluded in a selection step, then \(V3(1) = 9\) and after pair \((1,2)\) has been selected, \(V3(1) = 8\). After all of the NVAR - (NVAR-1)/2 selection steps, all elements of V3 equal 8 and workspaces are cleared for the next set of data.

4. Sample runs

As sample runs we display the two sets of data, that have been used previously [7] to describe the selection algorithm in detail.

For the five indicators on the maturity of a newborn infant (COVSEL) the variable pairs \((4,5)\) and \((2,5)\) are selected in the first two selection steps. Thus, after the second step the selected model is model 1234/135, the corresponding likelihood ratio statistic (3) has a value of 2.78 on 2 degrees of freedom (d.f.). Since this corresponds to a fairly high probability of 0.248 we can regard model 1234/135 as a well-fitting model. But, the z.p.a. of pair \((2,4)\) in the third selection step (and hence model 135/123) is not compatible with the data: a chi-square of 21.00 on 1 degree of freedom indicates with \(p = 0.000\) that pair \((2,4)\) is extremely unlikely to have z.p.a.; similarly, the test statistic for model 135/134/123 with a value of 23.79 on 3 d.f. and \(p = 0.000\) shows the lack of fit of this model. For all of the following selection steps it is assumed, that \((2,4)\) actually has z.p.a., therefore none of the following models should be judged as being acceptable.

For the four symptoms observed on psychiatric patients (TASEL) the variable pairs \((2,3)\), \((3,4)\) and \((1,2)\) are selected in that order in the first three selection steps. The statistics, computed to test their z.p.a., are all small: \(3.39\) on 4 d.f.; \(4.99\) on 2 d.f.; and \(5.49\)
AT. Wermuth et al., Multi-dimensional data description search

1=VALIDITY; 2=SOLIDITY; 3=STABILITY; 4=ACUTE DEPRESSION

NUMBER OF CATEGORIES FOR THE 4 VARIABLES: 1(1) 2(2) 3(2) 4(2)

<table>
<thead>
<tr>
<th>INPUT-TABLE</th>
<th>15</th>
<th>50</th>
<th>9</th>
<th>32</th>
<th>23</th>
<th>22</th>
<th>14</th>
<th>16</th>
<th>25</th>
<th>22</th>
<th>48</th>
<th>27</th>
<th>14</th>
<th>8</th>
<th>47</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUMBER OF CELLS</td>
<td>= 16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NUMBER OF OBSERVATIONS</td>
<td>= 362</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PAIR</th>
<th>INDEXCOMBINATION</th>
<th>CHI-SQUARE</th>
<th>DF</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2</td>
<td>1234</td>
<td>4.7806</td>
<td>4</td>
<td>0.3106</td>
</tr>
<tr>
<td>1,3</td>
<td>1234</td>
<td>12.8693</td>
<td>4</td>
<td>0.0119</td>
</tr>
<tr>
<td>1,4</td>
<td>1234</td>
<td>33.0043</td>
<td>4</td>
<td>0.0000</td>
</tr>
<tr>
<td>2,3</td>
<td>1234</td>
<td>3.3933</td>
<td>4</td>
<td>0.4943</td>
</tr>
<tr>
<td>2,4</td>
<td>1234</td>
<td>22.3829</td>
<td>4</td>
<td>0.0000</td>
</tr>
<tr>
<td>3,4</td>
<td>1234</td>
<td>7.6401</td>
<td>4</td>
<td>0.1057</td>
</tr>
</tbody>
</table>

SELECTED PAIR 2,3

NOTATION FOR THE SELECTED MODELL

<table>
<thead>
<tr>
<th>PAIR</th>
<th>INDEXCOMBINATION</th>
<th>CHI-SQUARE</th>
<th>DF</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2</td>
<td>124</td>
<td>5.4859</td>
<td>2</td>
<td>0.0644</td>
</tr>
<tr>
<td>1,3</td>
<td>134</td>
<td>13.5745</td>
<td>2</td>
<td>0.0011</td>
</tr>
<tr>
<td>2,4</td>
<td>124</td>
<td>19.7331</td>
<td>2</td>
<td>0.0000</td>
</tr>
<tr>
<td>3,4</td>
<td>134</td>
<td>4.9964</td>
<td>2</td>
<td>0.0825</td>
</tr>
</tbody>
</table>

SELECTED PAIR 3,4

CHI-SUM = 8.38366  DF-SUM = 6  P-SUM = 0.21132

NOTATION FOR THE SELECTED MODELL

<table>
<thead>
<tr>
<th>PAIR</th>
<th>INDEXCOMBINATION</th>
<th>CHI-SQUARE</th>
<th>DF</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2</td>
<td>124</td>
<td>5.4859</td>
<td>2</td>
<td>0.0644</td>
</tr>
<tr>
<td>1,3</td>
<td>13</td>
<td>10.0235</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>1,4</td>
<td>124</td>
<td>30.7964</td>
<td>2</td>
<td>0.0000</td>
</tr>
<tr>
<td>2,4</td>
<td>124</td>
<td>19.7331</td>
<td>2</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

SELECTED PAIR 1,2

CHI-SUM = 13.86958  DF-SUM = 8  P-SUM = 0.08523
### NOTATION FOR THE SELECTED MODEL
13/24/14

<table>
<thead>
<tr>
<th>PAIR INDEXCOMBINATION FOR THE SUBTABLE</th>
<th>CHI-SQUARE</th>
<th>DF</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,3 13</td>
<td>10.0235</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>1,4 14</td>
<td>28.0325</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>2,4 24</td>
<td>16.9692</td>
<td>1</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

SELECTED PAIR 2,4
CHIQ-SUM = 30.03674
DF-SUM = 9
P-SUM = 0.00032

### NOTATION FOR THE SELECTED MODEL
13/14/2

<table>
<thead>
<tr>
<th>PAIR INDEXCOMBINATION FOR THE SUBTABLE</th>
<th>CHI-SQUARE</th>
<th>DF</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,3 13</td>
<td>10.0235</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>1,4 14</td>
<td>28.0325</td>
<td>1</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

SELECTED PAIR 1,4
CHIQ-SUM = 58.87125
DF-SUM = 10
P-SUM = 0.00000

### NOTATION FOR THE SELECTED MODEL
13/2/4

<table>
<thead>
<tr>
<th>PAIR INDEXCOMBINATION FOR THE SUBTABLE</th>
<th>CHI-SQUARE</th>
<th>DF</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,3 13</td>
<td>10.0235</td>
<td>1</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

SELECTED PAIR 1,3
CHIQ-SUM = 68.89475
DF-SUM = 11
P-SUM = 0.00000

### NOTATION FOR THE SELECTED MODEL
1/2/3/4
STOP
NUMBER OF VARIABLES = 5
NUMBER OF OBSERVATIONS = 2473

INPUT MATRIX R

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.000000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.431400</td>
<td>1.000000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.514600</td>
<td>0.626300</td>
<td>1.000000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.409100</td>
<td>0.546600</td>
<td>0.703000</td>
<td>1.000000</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.411200</td>
<td>0.260400</td>
<td>0.392600</td>
<td>0.345300</td>
<td>1.000000</td>
</tr>
</tbody>
</table>

NEGATIVE INVERSE OF R

<table>
<thead>
<tr>
<th></th>
<th>-1.531143</th>
<th>0.241026</th>
<th>-1.706053</th>
<th>0.281262</th>
<th>-3.189959</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.273314</td>
<td>0.197754</td>
<td>1.915987</td>
<td>-2.677929</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.362934</td>
<td>-0.038718</td>
<td>0.503717</td>
<td>0.042906</td>
<td>-1.273125</td>
</tr>
</tbody>
</table>

PAIR   INDEXCOMBINATION FOR SUBMATRIX | CHI-SQUARE | DF | P
1,2    12345             | 55.61835  | 1  | 0.00000
1,3    12345             | 40.38187  | 1  | 0.00000
1,4    12345             | 45.13424  | 1  | 0.00000
1,5    12345             | 173.02012 | 1  | 0.00000
2,3    12345             | 313.58623 | 1  | 0.00000
2,4    12345             | 21.25932  | 1  | 0.00000
2,5    12345             | 1.70743   | 1  | 0.19132
3,4    12345             | 1206.42542| 1  | 0.00000
3,5    12345             | 56.01602  | 1  | 0.00000
4,5    12345             | 1.33567   | 1  | 0.24780

SELECTED PAIR: 4,5
CHISUM = 1.33567  DF-SUM = 1  P-SUM = 0.2478

NOTATION FOR THE SELECTED MODEL
1234  /1235

PAIR   INDEXCOMBINATION FOR SUBMATRIX | CHI-SQUARE | DF | P
1,4    1234             | 52.96311  |    |    
1,5    1235             | 180.84899 |    |    
2,4    1234             | 21.00444  |    |    
2,5    1235             | 1.45255   | 1  | 0.22812
3,4    1234             | 1262.04294|    |    
3,5    1235             | 112.43555 |    |    

SELECTED PAIR: 2,5
CHISUM = 2.78823  DF-SUM = 2  P-SUM = 0.2481
### N. Wermuth et al., Multi-dimensional data description search

**VOTATION FOR THE SELECTED MODEL 123/135**

<table>
<thead>
<tr>
<th>PAIR</th>
<th>INDEXCOMBINATION</th>
<th>CHI-SQUARE</th>
<th>DF</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2</td>
<td>1234</td>
<td>54.33437</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,4</td>
<td>1234</td>
<td>52.96311</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,5</td>
<td>135</td>
<td>180.41011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2,3</td>
<td>1234</td>
<td>313.58006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2,4</td>
<td>1234</td>
<td>21.00444</td>
<td>1</td>
<td>0.00000</td>
</tr>
<tr>
<td>3,4</td>
<td>1234</td>
<td>1262.04294</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3,5</td>
<td>135</td>
<td>136.30890</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SELECTED PAIR: 2,4**
CHI-SUM = 22.79267, DF-SUM = 3, P-SUM = 0.00000

**VOTATION FOR THE SELECTED MODEL 135/134/123**

<table>
<thead>
<tr>
<th>PAIR</th>
<th>INDEXCOMBINATION</th>
<th>CHI-SQUARE</th>
<th>DF</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2</td>
<td>123</td>
<td>66.78066</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,4</td>
<td>134</td>
<td>65.40939</td>
<td>1</td>
<td>0.00000</td>
</tr>
<tr>
<td>1,5</td>
<td>135</td>
<td>180.41011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2,3</td>
<td>123</td>
<td>789.06817</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3,4</td>
<td>134</td>
<td>1737.53105</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3,5</td>
<td>135</td>
<td>136.30890</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SELECTED PAIR: 1,4**
CHI-SUM = 69.20206, DF-SUM = 4, P-SUM = 0.00000

**VOTATION FOR THE SELECTED MODEL 135/123/34**

<table>
<thead>
<tr>
<th>PAIR</th>
<th>INDEXCOMBINATION</th>
<th>CHI-SQUARE</th>
<th>DF</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2</td>
<td>123</td>
<td>66.78066</td>
<td>1</td>
<td>0.00000</td>
</tr>
<tr>
<td>1,5</td>
<td>135</td>
<td>180.41011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2,3</td>
<td>123</td>
<td>789.06817</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3,4</td>
<td>34</td>
<td>2340.26333</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3,5</td>
<td>135</td>
<td>136.30890</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SELECTED PAIR: 1,2  
CHI-SUM =  155.98272  DF-SUM =  5  P-SUM =  0.0000

NOTATION FOR THE SELECTED MODEL  
135/34/23

<table>
<thead>
<tr>
<th>PAIR INDEX COMBINATION</th>
<th>CHI-SQUARE</th>
<th>DF</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,3</td>
<td>135</td>
<td>483.11101</td>
<td></td>
</tr>
<tr>
<td>1,4</td>
<td>135</td>
<td>160.41011</td>
<td></td>
</tr>
<tr>
<td>2,3</td>
<td>23</td>
<td>1231.54027</td>
<td></td>
</tr>
<tr>
<td>3,4</td>
<td>34</td>
<td>2348.26333</td>
<td></td>
</tr>
<tr>
<td>3,5</td>
<td>135</td>
<td>136.30890</td>
<td>1</td>
</tr>
</tbody>
</table>

SELECTED PAIR: 3,5  
CHI-SUM =  292.29162  DF-SUM =  6  P-SUM =  0.0000

NOTATION FOR THE SELECTED MODEL  
34/23/15/13

<table>
<thead>
<tr>
<th>PAIR INDEX COMBINATION</th>
<th>CHI-SQUARE</th>
<th>DF</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,3</td>
<td>13</td>
<td>760.77049</td>
<td></td>
</tr>
<tr>
<td>1,5</td>
<td>15</td>
<td>458.56960</td>
<td>1</td>
</tr>
<tr>
<td>2,3</td>
<td>23</td>
<td>1231.54027</td>
<td></td>
</tr>
<tr>
<td>3,4</td>
<td>34</td>
<td>2348.26333</td>
<td></td>
</tr>
</tbody>
</table>

SELECTED PAIR: 1,5  
CHI-SUM =  750.36122  DF-SUM =  7  P-SUM =  0.0000

NOTATION FOR THE SELECTED MODEL  
34/23/15/5

<table>
<thead>
<tr>
<th>PAIR INDEX COMBINATION</th>
<th>CHI-SQUARE</th>
<th>DF</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,3</td>
<td>13</td>
<td>760.77049</td>
<td>1</td>
</tr>
<tr>
<td>2,3</td>
<td>23</td>
<td>1231.54027</td>
<td></td>
</tr>
<tr>
<td>3,4</td>
<td>34</td>
<td>2348.26333</td>
<td></td>
</tr>
</tbody>
</table>

SELECTED PAIR: 1,3  
CHI-SUM =  1511.13171  DF-SUM =  8  P-SUM =  0.0000

NOTATION FOR THE SELECTED MODEL  
34/23/1/5
<table>
<thead>
<tr>
<th>PAIR</th>
<th>INDEXCOMBINATION</th>
<th>CHI-SQUARE</th>
<th>DF</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>2\times3</td>
<td></td>
<td>1231.54027</td>
<td>1</td>
<td>0.00000</td>
</tr>
<tr>
<td>3\times4</td>
<td></td>
<td>2348.26333</td>
<td>1</td>
<td>0.00000</td>
</tr>
</tbody>
</table>

SELECTED PAIR: 2\times3  
CHIG-SUM = 2742.67198  
DF-SUM = 9  
P-SUM = 0.00000

NOTATION FOR THE SELECTED MODEL  
34/1/2/5

---

<table>
<thead>
<tr>
<th>PAIR</th>
<th>INDEXCOMBINATION</th>
<th>CHI-SQUARE</th>
<th>DF</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>3\times4</td>
<td></td>
<td>2348.26333</td>
<td>1</td>
<td>0.00000</td>
</tr>
</tbody>
</table>

SELECTED PAIR: 3\times4  
CHIG-SUM = 5090.93531  
DF-SUM = 10  
P-SUM = 0.00000

NOTATION FOR THE SELECTED MODEL  
1/2/3/4/5
STOP

---

== XCOVSL ==
on 2 d.f.; all with $p > 0.05$. They indicate therefore that the assumptions of z.p.a. are acceptable for these three variable pairs. This result is confirmed by the overall test for the selected model: $p > 0.05$ for the chi-square 14.41 on 8 d.f. The large test statistics for z.p.a. in the following selection steps tell that no simpler multiplicative model fits these data well.

After a model has been selected, it is instructive to look at the expected values implied by this model. More precisely, one wishes to compute the maximum-likelihood estimates for correlations in the case of a given covariance selection model and maximum-likelihood estimates of cell counts in the case of a given log-linear model. Fortran programs are available in both situations [8,2].

5. Hardware and software specifications

The programs have been written in Fortran IV for the CDC 3300 computer at the Computational Center of the University of Mainz. Locations in the program where CDC-dependent functions and formats are used have been identified as such in the flowcharts. Storage requirements are 27,650 words for COVSEL and 31,750 words for TASEL.

6. Mode of availability of the programs

Readers interested in obtaining the programs are invited to contact the first author. The text of the program output is available in German and in English.

Acknowledgements

We thank O. Pietschmann for drawing the flowcharts and H. Bianco for typing the manuscript.

References