Computer Programs in Biomedicine 6 (1976) 23–38

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.

Model search procedure   multiplicative models   contingency table analysis   analysis of correlation matrices

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 multinominal 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

$$\mathrm{f}(x_1\,x_2\,x_3\,x_4\,x_5) = \frac{\mathrm{f}(x_1\,x_2\,x_3)\,\mathrm{f}(x_1\,x_4)\,\mathrm{f}(x_1\,x_5)}{\mathrm{f}(x_1)\,\mathrm{f}(x_1)} \quad (1)$$

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 $|\Sigma|$ 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]

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

$$-2 \ln \left(\frac{|\hat{\Sigma}|}{|S|}\right)^{-n/2}, \tag{3}$$

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

$$-2 \ln \left[\frac{D_{123}\, D_{14}\, D_{15}/D_1\, D_2}{D_{12345}}\right]^{-n/2} \tag{4}$$

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

$$-2 \ln \left[\frac{D_{1235}\, D_{1345}/D_{135}}{D_{12345}}\right]^{-n/2}$$
$$+\left(-2 \ln \left[\frac{D_{123}\, D_{135}/D_{13}}{D_{1235}}\right]^{-n/2}\right)$$
$$+\left(-2 \ln \left[\frac{D_{134}\, D_{135}/D_{13}}{D_{1345}}\right]^{-n/2}\right)$$
$$+\left(-2 \ln \left[\frac{D_{13}\, D_{15}/D_{1}}{D_{135}}\right]^{-n/2}\right)$$
$$+\left(-2 \ln \left[\frac{D_{13}\, D_{14}/D_{1}}{D_{134}}\right]^{-n/2}\right). \tag{5}$$

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.

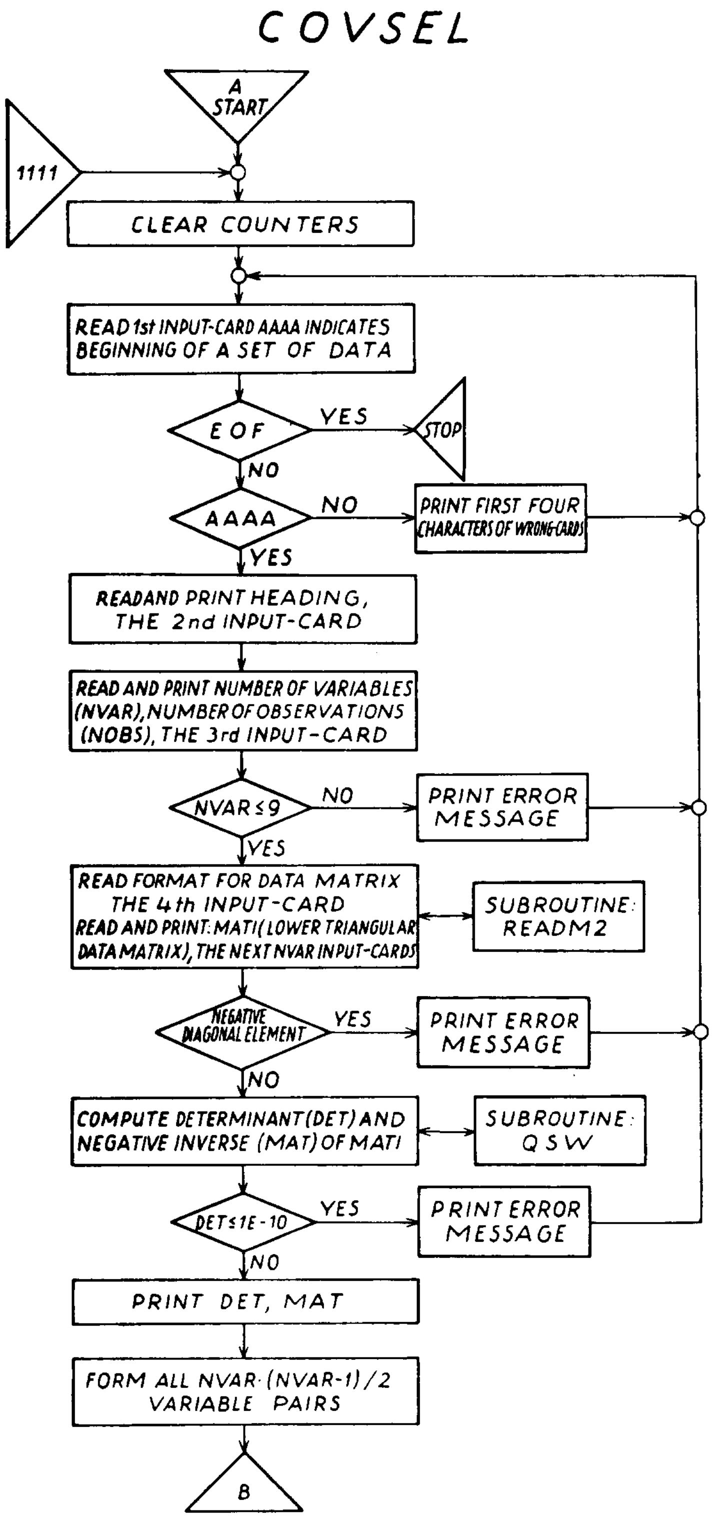
A.

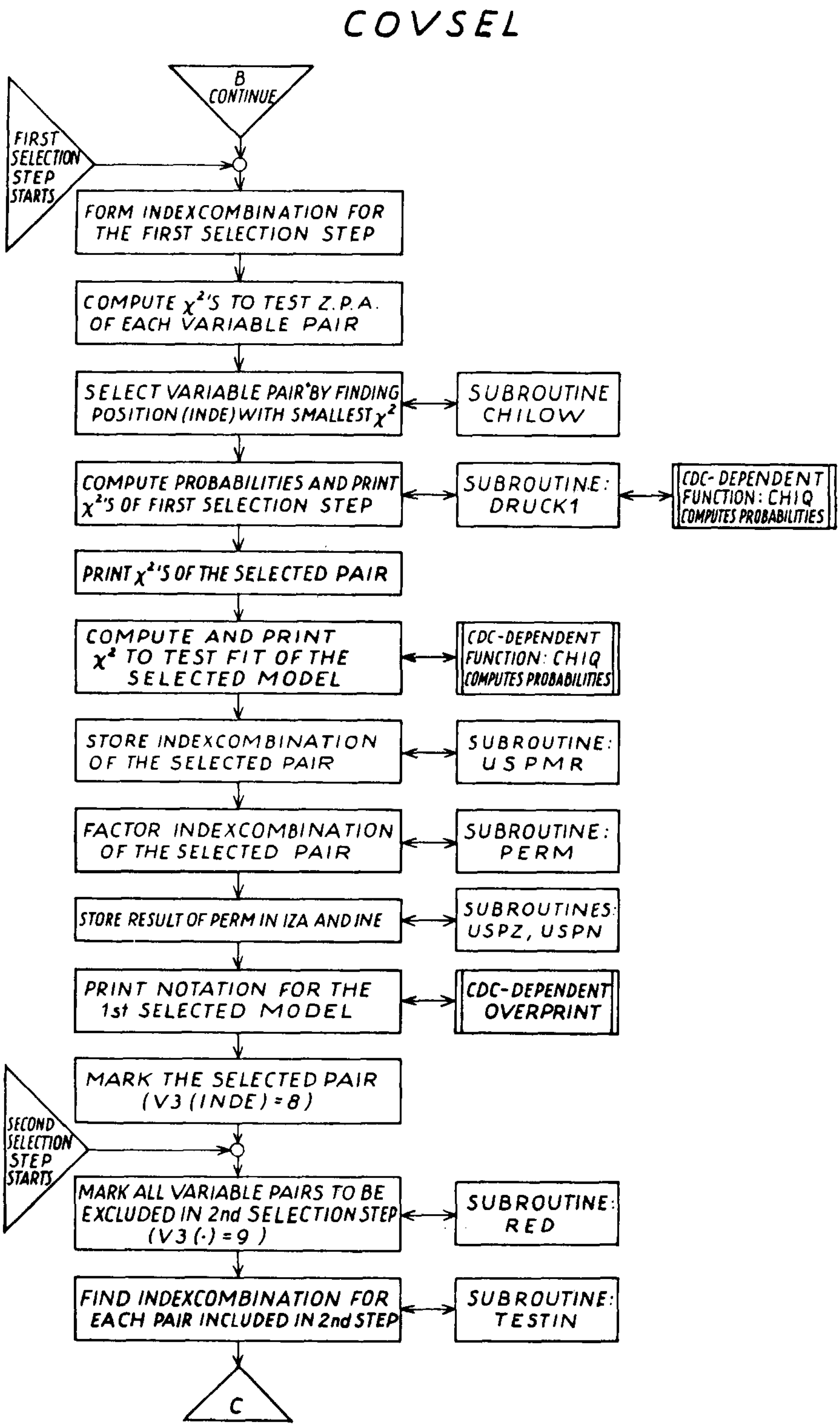
The input cards consist of five different types of cards.

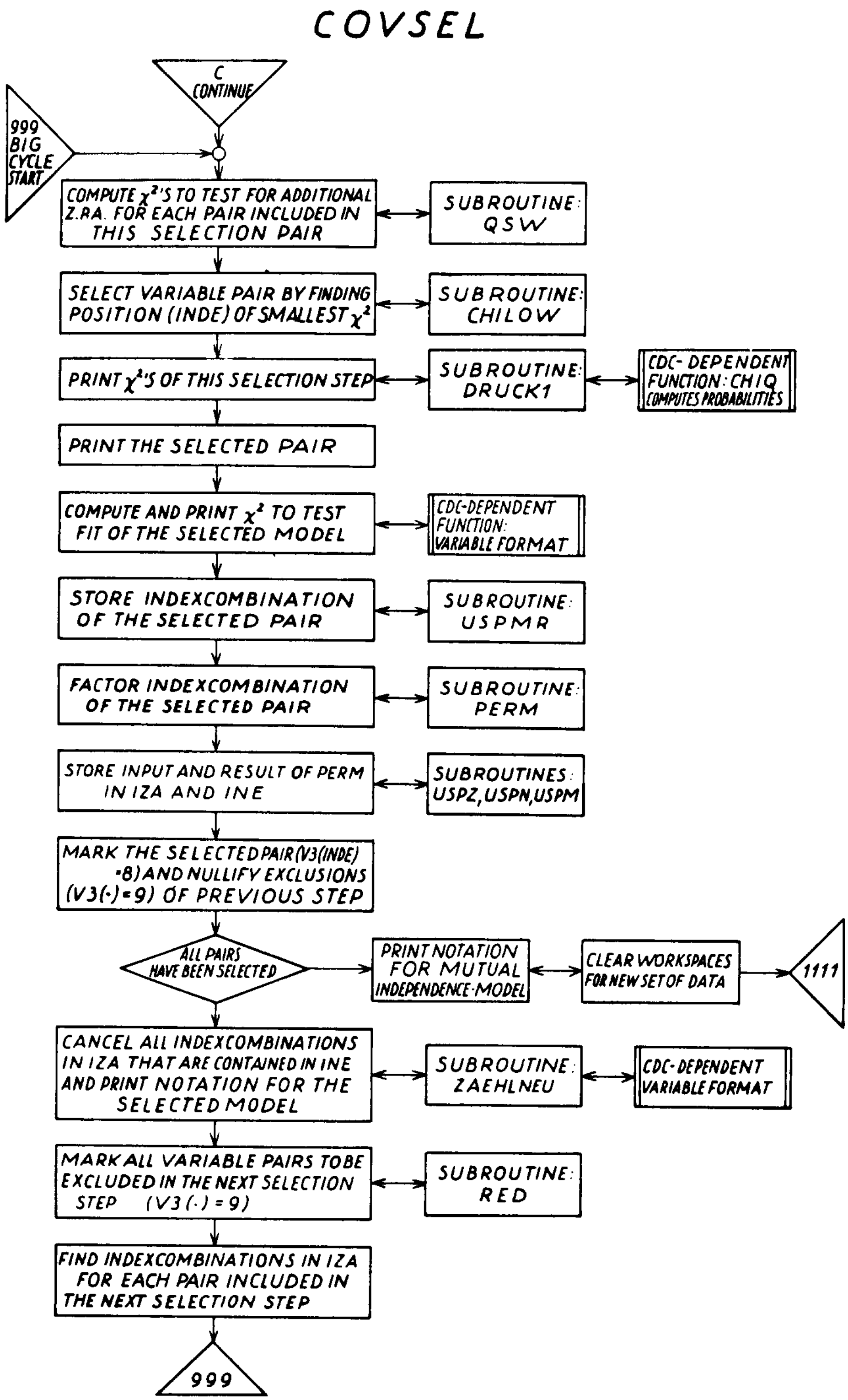
1: Beginning card, AAAA in the first four columns of the first input card;
2: a heading card;
3: information for the data cards (for COVSEL the number of variables (NVAR) and the number of observations (NOBS), for TASEL the number of categories per variable on (ICARD*);
5: the main data (for COVSEL the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

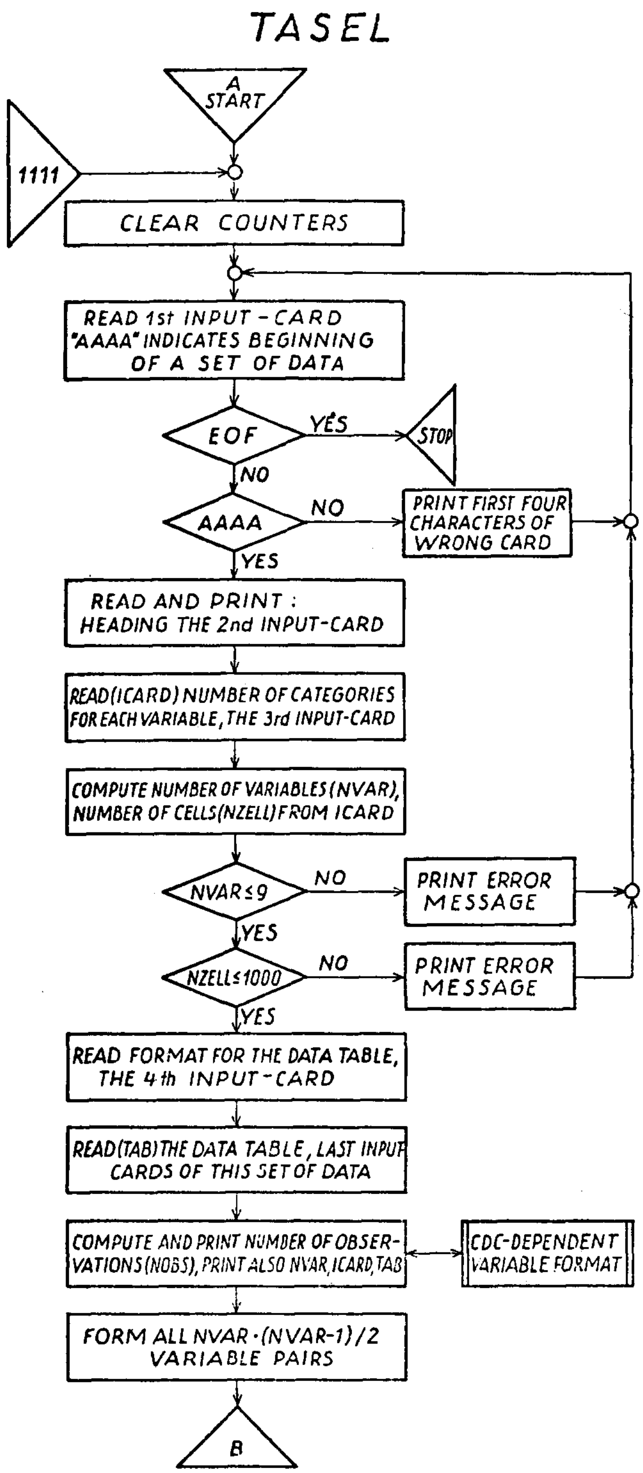
In the case that an error in the input cards is encountered, error messages are printed and the next

*A dummy "1" in the second column made this card compatible with L. Goodman's ECTA-program (University of Chicago).

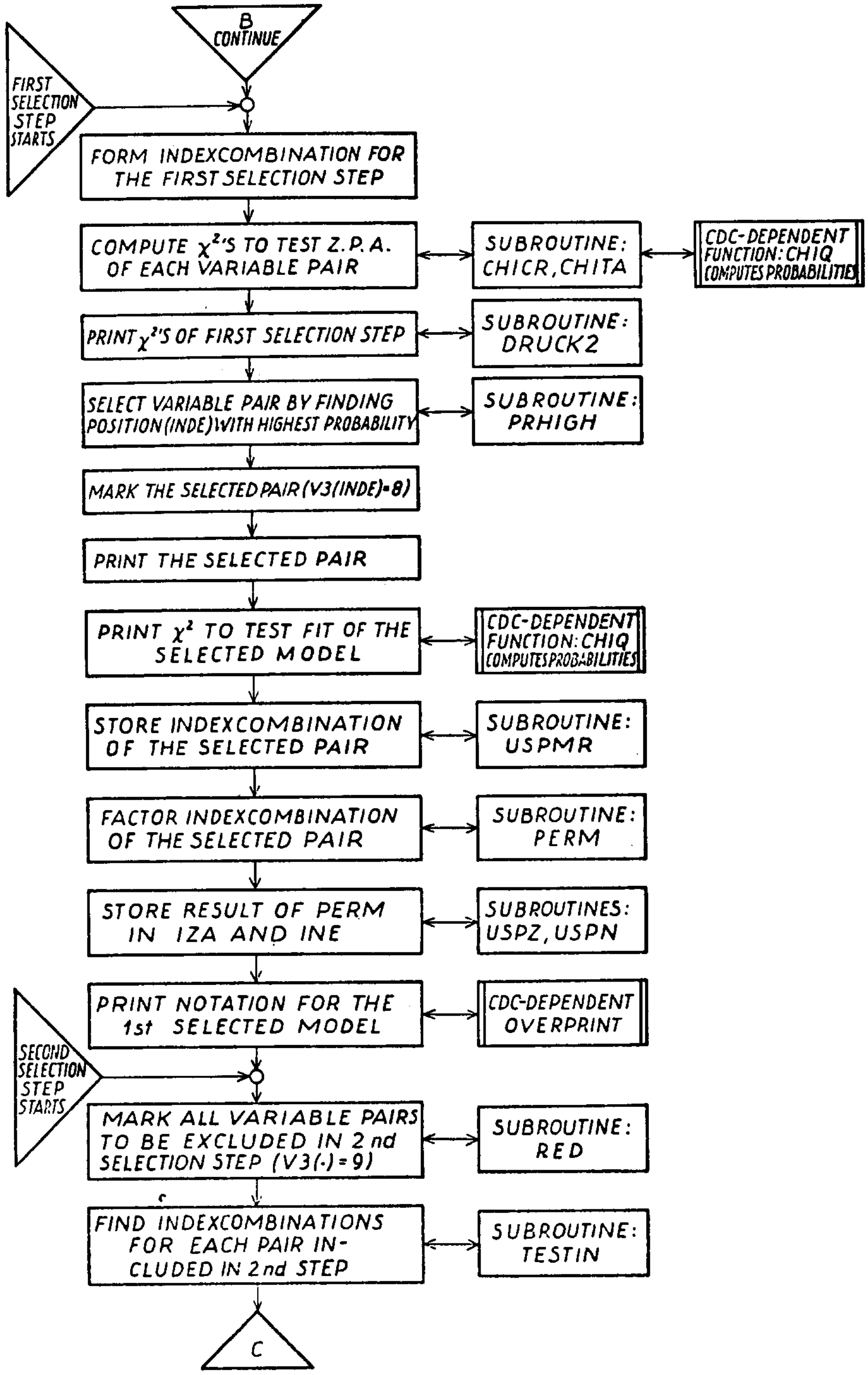

COVSEL
A START
1111
CLEAR COUNTERS
READ 1st INPUT-CARD AAAA INDICATES BEGINNING OF A SET OF DATA
EOF
YES
STOP
NO
AAAA
NO
PRINT FIRST FOUR CHARACTERS OF WRONG-CARDS
YES
READ AND PRINT HEADING, THE 2nd INPUT-CARD
READ AND PRINT NUMBER OF VARIABLES (NVAR), NUMBER OF OBSERVATIONS (NOBS), THE 3rd INPUT-CARD
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
SUBROUTINE: READM2
NEGATIVE DIAGONAL ELEMENT
YES
PRINT ERROR MESSAGE
NO
COMPUTE DETERMINANT (DET) AND NEGATIVE INVERSE (MAT) OF MATI
SUBROUTINE: QSW
DET ≤ 1E-10
YES
PRINT ERROR MESSAGE
NO
PRINT DET, MAT
FORM ALL NVAR·(NVAR-1)/2 VARIABLE PAIRS
B


COVSEL
B CONTINUE
FIRST SELECTION STEP STARTS
FORM INDEXCOMBINATION FOR THE FIRST SELECTION STEP
COMPUTE χ^2'S TO TEST Z.P.A. OF EACH VARIABLE PAIR
SELECT VARIABLE PAIR BY FINDING POSITION (INDE) WITH SMALLEST χ^2
SUBROUTINE CHILOW
COMPUTE PROBABILITIES AND PRINT χ^2'S OF FIRST SELECTION STEP
SUBROUTINE: DRUCK1
CDC-DEPENDENT FUNCTION: CHIQ COMPUTES PROBABILITIES
PRINT χ^2'S OF THE SELECTED PAIR
COMPUTE AND PRINT χ^2 TO TEST FIT OF THE SELECTED MODEL
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: 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

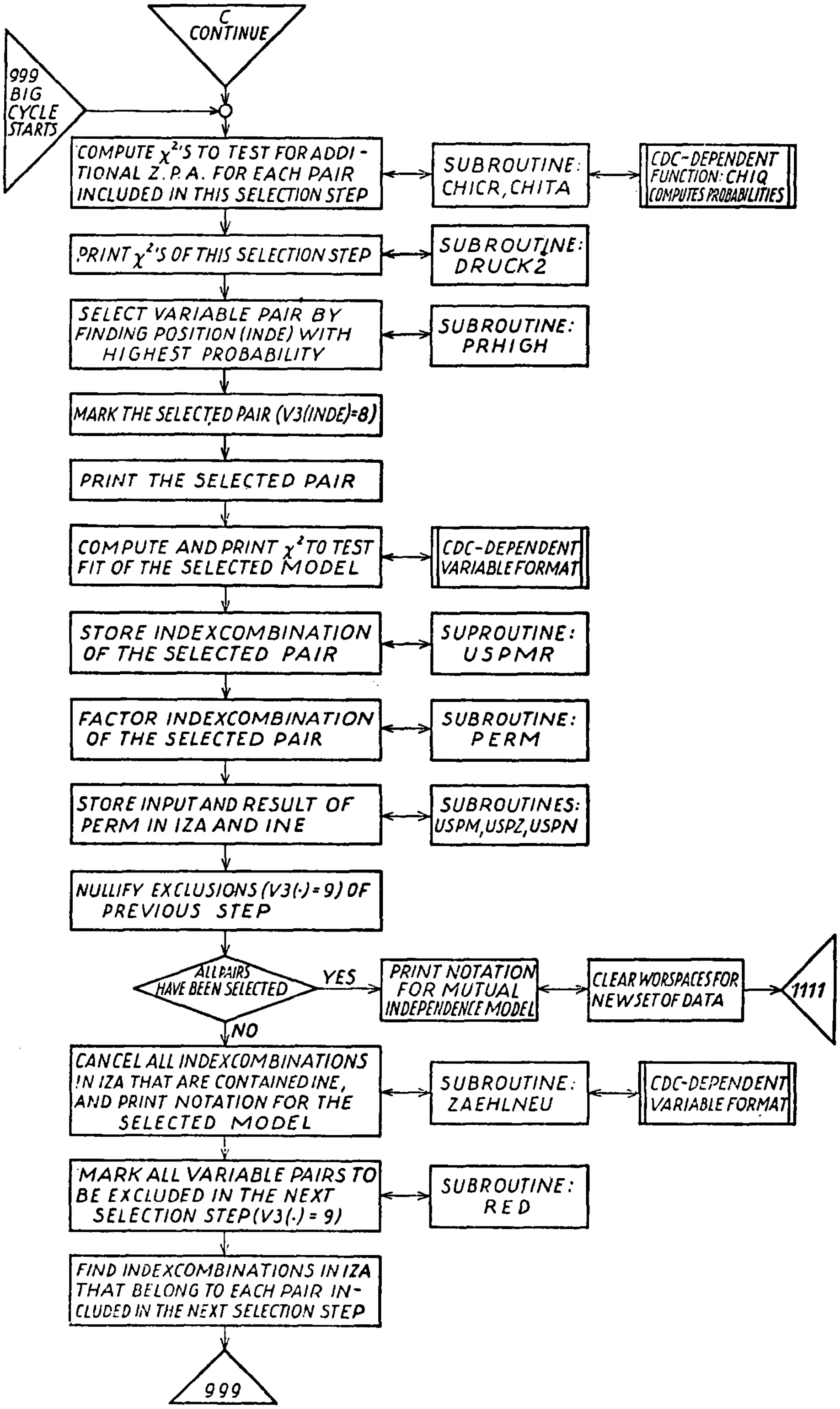

COVSEL
C CONTINUE
999 BIG CYCLE START
COMPUTE χ²'S TO TEST FOR ADDITIONAL Z.RA. FOR EACH PAIR INCLUDED IN THIS SELECTION PAIR
SUBROUTINE: QSW
SELECT VARIABLE PAIR BY FINDING POSITION (INDE) OF SMALLEST χ²
SUBROUTINE: CHILOW
PRINT χ²'S OF THIS SELECTION STEP
SUBROUTINE: DRUCK1
CDC-DEPENDENT FUNCTION: CHIQ COMPUTES PROBABILITIES
PRINT THE SELECTED PAIR
COMPUTE AND PRINT χ² TO TEST FIT OF THE SELECTED MODEL
CDC-DEPENDENT FUNCTION: VARIABLE FORMAT
STORE INDEXCOMBINATION OF THE SELECTED PAIR
SUBROUTINE: USPMR
FACTOR INDEXCOMBINATION OF THE SELECTED PAIR
SUBROUTINE: PERM
STORE INPUT AND RESULT OF PERM IN IZA AND INE
SUBROUTINES: USPZ, USPN, USPM
MARK THE SELECTED PAIR (V3(INDE) =8) AND NULLIFY EXCLUSIONS (V3(·)=9) OF PREVIOUS STEP
ALL PAIRS HAVE BEEN SELECTED
PRINT NOTATION FOR MUTUAL INDEPENDENCE-MODEL
CLEAR WORKSPACES FOR NEW SET OF DATA
1111
CANCEL ALL INDEXCOMBINATIONS IN IZA THAT ARE CONTAINED IN INE AND PRINT NOTATION FOR THE SELECTED MODEL
SUBROUTINE: ZAEHLNEU
CDC-DEPENDENT VARIABLE FORMAT
MARK ALL VARIABLE PAIRS TO BE EXCLUDED IN THE NEXT SELECTION STEP (V3(·) = 9)
SUBROUTINE: RED
FIND INDEXCOMBINATIONS IN IZA FOR EACH PAIR INCLUDED IN THE NEXT SELECTION STEP
999


TASEL
A
START
1111
CLEAR COUNTERS
READ 1st INPUT-CARD
"AAAA" 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(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(TAB) THE DATA TABLE, LAST INPUT
CARDS OF THIS SET OF DATA
COMPUTE AND PRINT NUMBER OF OBSER-
VATIONS (NOBS), PRINT ALSO NVAR, ICARD, TAB
CDC-DEPENDENT
VARIABLE FORMAT
FORM ALL NVAR·(NVAR-1)/2
VARIABLE PAIRS
B

TASEL


B CONTINUE
FIRST SELECTION STEP STARTS
FORM INDEXCOMBINATION FOR THE FIRST SELECTION STEP
COMPUTE χ²'S TO TEST Z.P.A. OF EACH VARIABLE PAIR
SUBROUTINE: CHICR, CHITA
CDC-DEPENDENT FUNCTION: CHIQ COMPUTES PROBABILITIES
PRINT χ²'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 χ² TO TEST FIT OF THE SELECTED MODEL
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: USPZ, USPN
PRINT NOTATION FOR THE 1st SELECTED MODEL
CDC-DEPENDENT OVERPRINT
SECOND SELECTION STEP STARTS
MARK ALL VARIABLE PAIRS TO BE EXCLUDED IN 2nd SELECTION STEP (V3(·)=9)
SUBROUTINE: RED
FIND INDEXCOMBINATIONS FOR EACH PAIR INCLUDED IN 2nd STEP
SUBROUTINE: TESTIN
C

TASEL


C CONTINUE
999 BIG CYCLE STARTS
COMPUTE χ^2'S TO TEST FOR ADDITIONAL Z.P.A. FOR EACH PAIR INCLUDED IN THIS SELECTION STEP
SUBROUTINE: CHICR, CHITA
CDC-DEPENDENT FUNCTION: CHIQ COMPUTES PROBABILITIES
PRINT χ^2'S OF THIS 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
COMPUTE AND PRINT χ^2 TO TEST FIT OF THE SELECTED MODEL
CDC-DEPENDENT VARIABLE FORMAT
STORE INDEXCOMBINATION OF THE SELECTED PAIR
SUPROUTINE: USPMR
FACTOR INDEXCOMBINATION OF THE SELECTED PAIR
SUBROUTINE: PERM
STORE INPUT AND RESULT OF PERM IN IZA AND INE
SUBROUTINES: USPM, USPZ, USPN
NULLIFY EXCLUSIONS (V3(·)=9) OF PREVIOUS STEP
ALL PAIRS HAVE BEEN SELECTED
YES
PRINT NOTATION FOR MUTUAL INDEPENDENCE MODEL
CLEAR WORSPACES FOR NEW SET OF DATA
1111
NO
CANCEL ALL INDEXCOMBINATIONS IN IZA THAT ARE CONTAINED IN E, AND PRINT NOTATION FOR THE SELECTED MODEL
SUBROUTINE: ZAEHLNEU
CDC-DEPENDENT VARIABLE FORMAT
MARK ALL VARIABLE PAIRS TO BE EXCLUDED IN THE NEXT SELECTION STEP (V3(·)=9)
SUBROUTINE: RED
FIND INDEXCOMBINATIONS IN IZA THAT BELONG TO EACH PAIR INCLUDED IN THE NEXT SELECTION STEP
999

AAAA 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

$$-2 \ln \left[1 - \frac{(r^{ij})^2}{r^{ii} r^{jj}}\right] \tag{6}$$

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

$$2 \sum_{ijK} n_{ijK} \ln n_{ijK} - \sum_{iK} n_{i.K} \ln n_{i.K}$$

$$+ \sum_{jK} n_{.jK} \ln n_{.jK} - \sum_{K} n_{..K} \ln n_{..K} \tag{7}$$

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

$$n_{..K} = \sum_i n_{i.K} = \sum_j n_{ijK} \; .$$

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

$$(ijK) \rightarrow \frac{(iK)\,(jK)}{(K)} \; . \tag{8}$$

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

```
1=VALIDITY, 2=SOLIDITY, 3=STABILITY, 4=ACUTE DEPRESSION

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

INPUT-TABLE
   15   30    9   32   23   22   14   16   25   22   46   27   14    8   47   1

NUMBER OF CELLS         = 16

NUMBER OF OBSERVATIONS  = 362


PAIR     INDEXCOMBINATION              CHI-SQUARE      DF            P
         FOR THE SUBTABLE
1,2      1234                              4.7806       4            0.3106

1,3      1234                             12.8693       4            0.0119

1,4      1234                             33.0043       4            0.0000

2,3      1234                              3.3933       4            0.4943

2,4      1234                             22.3829       4            0.0000

3,4      1234                              7.6401       4            0.1057

SELECTED PAIR 2,3

NOTATION FOR THE SELECTED MODELL
124      /134


PAIR     INDEXCOMBINATION              CHI-SQUARE      DF            P
         FOR THE SUBTABLE
1,2      124                               5.4859       2            0.0644

1,3      134                              13.5745       2            0.0011

2,4      124                              19.7331       2            0.0000

3,4      134                               4.9904       2            0.0825

SELECTED PAIR 3,4
CHIQ-SUM =      8.38366      DF-SUM =      6      P-SUM =      0.21132

NOTATION FOR THE SELECTED MODELL
124/13


PAIR     INDEXCOMBINATION              CHI-SQUARE      DF            P
         FOR THE SUBTABLE
1,2      124                               5.4859       2            0.0644

1,3      13                               10.0235       1            0.0000

1,4      124                              30.7964       2            0.0000

2,4      124                              19.7331       2            0.0000

SELECTED PAIR 1,2
CHIQ-SUM =     13.86958      DF-SUM =      8      P-SUM =      0.08523
```

```
NOTATION FOR THE SELECTED MODELL
13/24/14


PAIR     INDEXCOMBINATION             CHI-SQUARE      DF          P
         FOR THE SUBTABLE
1,3      13                             10.0235        1          0.0000

1,4      14                             28.0325        1          0.0000

2,4      24                             16.9692        1          0.0000

SELECTED PAIR 2,4
CHIQ-SUM =     30.83674     DF-SUM =      9      P-SUM =      0.00032

NOTATION FOR THE SELECTED MODELL
13/14/2


PAIR     INDEXCOMBINATION             CHI-SQUARE      DF          P
         FOR THE SUBTABLE
1,3      13                             10.0235        1          0.0000

1,4      14                             28.0325        1          0.0000

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

NOTATION FOR THE SELECTED MODELL
13/2/4


PAIR     INDEXCOMBINATION             CHI-SQUARE      DF          P
         FOR THE SUBTABLE
1,3      13                             10.0235        1          0.0000

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

NOTATION FOR THE SELECTED MODELL
1/2/3/4/
   STOP

== XTASEL ===="""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""

""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""
```

```
1=GESTATION, 2=HEAD CIRC., 3=WEIGHT, 4=LENGTH, 5=CONSTRUCTED INDICATOR

NUMBER OF VARIABLES       = 5

NUMBER OF OBSERVATIONS    = 2473

INPUT MATRIX R
   1.000000
   0.431400   1.000000
   0.514600   0.626300   1.000000
   0.489100   0.546600   0.783000   1.000000
   0.411200   0.260400   0.392600   0.343300   1.000000

NEGATIVE INVERSE OF R
  -1.531143
   0.241026  -1.706053
   0.281262   0.804828  -3.189959
   0.272314   0.197754   1.815987  -2.677929
   0.362934  -0.038718   0.303717   0.042906  -1.273125

PAIR      INDEXCOMBINATION            CHI-SQUARE       DF          P
          FOR SUBMATRIX

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                      313.38623         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                       56.81802         1         0.00000

4,5       12345                        1.33567         1         0.24780

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

NOTATION FOR THE SELECTED MODEL
1234        /1235

PAIR      INDEXCOMBINATION            CHI-SQUARE       DF          P
          FOR SUBMATRIX

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
CHIQ-SUM =      2.78823    DF-SUM =  2    P-SUM =      0.2481
```

NOTATION FOR THE SELECTED MODEL
1234/135

PAIR	INDEXCOMBINATION FOR SUBMATRIX	CHI-SQUARE	DF	P
1,2	1234	54.33437		
1,4	1234	52.96311		
1,5	135	180.41011		
2,3	1234	313.58006		
2,4	1234	21.00444	1	0.00000
3,4	1234	1262.04294		
3,5	135	136.30890		

SELECTED PAIR: 2,4
CHIQ-SUM = 23.79267 DF-SUM = 3 P-SUM = 0.0000

NOTATION FOR THE SELECTED MODEL
135/134/123

PAIR	INDEXCOMBINATION FOR SUBMATRIX	CHI-SQUARE	DF	P
1,2	123	66.78066		
1,4	134	65.40939	1	0.00000
1,5	135	180.41011		
2,3	123	789.06817		
3,4	134	1737.53105		
3,5	135	136.30890		

SELECTED PAIR: 1,4
CHIQ-SUM = 89.20206 DF-SUM = 4 P-SUM = 0.0000

NOTATION FOR THE SELECTED MODEL
135/123/34

PAIR	INDEXCOMBINATION FOR SUBMATRIX	CHI-SQUARE	DF	P
1,2	123	66.78066	1	0.00000
1,5	135	180.41011		
2,3	123	789.06817		
3,4	34	2348.26333		
3,5	135	136.30890		

SELECTED PAIR: 1,2
CHIQ-SUM = 155.98272 DF-SUM = 5 P-SUM = 0.0000

NOTATION FOR THE SELECTED MODEL
135/34/23

PAIR	INDEXCOMBINATION FOR SUBMATRIX	CHI-SQUARE	DF	P
1,3	135	483.11101		
1,5	135	180.41011		
2,3	23	1231.54027		
3,4	34	2348.26333		
3,5	135	136.30890	1	0.00000

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

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

PAIR	INDEXCOMBINATION FOR SUBMATRIX	CHI-SQUARE	DF	P
1,3	13	760.77049		
1,5	15	458.06960	1	0.00000
2,3	23	1231.54027		
3,4	34	2348.26333		

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

NOTATION FOR THE SELECTED MODEL
34/23/13/5

PAIR	INDEXCOMBINATION FOR SUBMATRIX	CHI-SQUARE	DF	P
1,3	13	760.77049	1	0.00000
2,3	23	1231.54027		
3,4	34	2348.26333		

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

NOTATION FOR THE SELECTED MODEL
34/23/1/5

```
PAIR     INDEXCOMBINATION          CHI-SQUARE      DF        P
         FOR SUBMATRIX

2,3      23                        1231.54027      1         0.00000

3,4      34                        2348.26333

SELECTED PAIR: 2,3
CHIQ-SUM =   2742.67198    DF-SUM =  9     P-SUM =      0.0000

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


PAIR     INDEXCOMBINATION          CHI-SQUARE      DF        P
         FOR SUBMATRIX

3,4      34                        2348.26333      1         0.00000

SELECTED PAIR: 3,4
CHIQ-SUM =    5090.93531   DF-SUM = 10     P-SUM =      0.0000

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

[1] M.W. Birch, Maximum likelihood in three-way contingency tables. J, Roy. Statist. Soc B 25 (1963) 220–233.

[2] Y.M.M. Bishop, S. Fienberg and P. Holland, Discrete Multivariate Analysis: Theory and Practice. (M.I.T. Press, Boston, Boston, 1975).

[3] J.P. Bunkers et al. (eds), The National Halothane Study (Natl. Inst. of Health, Bethesda, 1969).

[4] A.P. Dempster, Covariance selection, Biometrics 28 (1972) 157–175.

[5] S. Koller et al., Schwangerschaftsverlauf und Kindesentwicklung (in preparation).

[6] N. Wermuth, Analogies between multiplicative models in contingency tables and covariance selection. Biometrics 32 (to appear March 1976).

[7] N. Wermuth, Model search among multiplicative models. Biometrics 32 (to appear June 1976).

[8] N. Wermuth, E. Scheidt, Fitting a covariance selection model to a matrix, Applied Statistics (submitted 1975 as Algorithm 172).