M CANO

```
C                   0 - IF IT IS NOT DESIRED
C                   1 - IF IT IS DESIRED
C
      WRITE (6,3) PR1,PR2
      WRITE (6,9) N
      WRITE (6,10) M
      IF(N-M-2) 101,101,102
  101 WRITE(6,21) N,M
      STOP
  102 WRITE (6,11) NS
      WRITE (6,12) PCT
C
C     LOGICAL TAPE 13 IS USED AS INTERMEDIATE STORAGE TO HOLD INPUT
C     DATA.  THE INPUT DATA ARE WRITTEN ON LOGICAL TAPE 13 BY THE
C     SPECIAL INPUT SUBROUTINE NAMED DATA.  THE STORED DATA MAY BE USED
C     FOR RESIDUAL ANALYSIS.
C
      REWIND 13
C
      IO=0
      X=0.0
C
      CALL CORRE (N,M,IO,X,XBAR,STD,RX,R,B,D,T)
C
      REWIND 13
C
C        PRINT MEANS AND STANDARD DEVIATION
C
      WRITE (6,4)
      DO 105 I=1,M
  105 WRITE (6,5) I,XBAR(I),STD(I)
C
C        PRINT CORRELATION MATRIX
C
      WRITE (6,6)
      DO 130 I=1,M
      DO 125 J=1,M
      IF(I-J) 110, 120, 120
  110 K=I+(J*J-J)/2
      GO TO 125
  120 K=J+(I*I-I)/2
  125 T(J)=R(K)
  130 WRITE (6,7) I,(T(J),J=1,M)
C
C        TEST NUMBER OF SELECTIONS
C
      IF(NS) 135, 135, 140
  135 WRITE (6,2)
      GO TO 200
C
C        SAVE THE MATRIX OF SUMS OF CROSS-PRODUCTS OF DEVIATIONS
C
  140 CALL MSTR (RX,R,M,0,1)
C
      NSEL=1
      GO TO 150
C
C        COPY THE MATRIX OF SUMS OF CROSS-PRODUCTS OF DEVIATIONS
C
  145 CALL MSTR (R,RX,M,1,0)
C
C        READ A SELECTION CARD
C
  150 WRITE (6,13) NSEL
      READ (5,8) (IDX(J),J=1,M)
C
C        IN EACH POSITION OF IDX, ONE OF THE FOLLOWING CODES MUST BE
C        SPECIFIED..
C          0 OR BLANK - INDEPENDENT VARIABLE AVAILABLE FOR SELECTION
C          1          - INDEPENDENT VARIABLE TO BE FORCED IN REGRESSION
C          2          - VARIABLE TO BE DELETED
C          3          - DEPENDENT VARIABLE
C
      N35=0
      DO 155 K=1,M
      IF (IDX(K)) 152,153,153
  152 WRITE (6,18) K,NSEL
      GO TO 185
  153 IF (IDX(K)-3) 155,154,152
  154 N35=N35+1
  155 CONTINUE
      IF (N35-1) 156,157,156
  156 WRITE (6,19) NSEL
      GO TO 185
C        CALL THE SUBROUTINE TO PERFORM A STEP-WISE REGRESSION ANALYSIS
C
  157 CALL STPRG (M,N,RX,XBAR,IDX,PCT,NSTEP,ANS,L,B,STD,T,D,IER)
      IF (IER) 158,159,158
  158 WRITE (6,20)
      GO TO 185
C
C        FIND WHETHER TO PRINT THE TABLE OF RESIDUALS
C
  159 IF(NR) 185, 185, 160
C
C        PRINT THE TABLE OF RESIDUALS
C
C
  160 WRITE (6,13) NSEL
      WRITE (6,16)
      WRITE (6,14)
      MM=NSTEP(1)
      DO 180 I=1,N
      READ (13) (D(J),J=1,M)
      YEST=ANS(9)
      K=NSTEP(4)
      DO 170 J=1,K
      KK=L(J)
  170 YEST=YEST+B(J)*D(KK)
      RESI=D(MM)-YEST
  180 WRITE (6,15) I,D(MM),YEST,RESI
      REWIND 13
C
C        TEST TO SEE WHETHER ALL SELECTIONS ARE COMPLETED
C
  185 IF(NSEL-NS) 190, 100, 100
  190 NSEL=NSEL+1
      WRITE (6,17)
      GO TO 145
C
  200 CONTINUE
      END
```

CANONICAL CORRELATION

Problem Description

An analysis of the interrelations between two sets of variables measured on the same subjects is performed by this program. These variables are predictors in one set and criteria in the other set, but it is irrelevant whether the variables in the first set or in the second set are considered as the prediction variables. The canonical correlation, which gives the maximum correlation between linear functions of the two sets of variables, is calculated. χ^2 is also computed to test the significance of canonical correlation.

The sample problem for canonical correlation consists of four variables in the first set (left-hand side) and three variables in the second set (right-hand side) as presented in Table 6. These two sets of measurements have been made on 23 subjects.

Table 6. Sample Data for Canonical Correlation

	First set				Second set		
Observation	X_1	X_2	X_3	X_4	Y_1	Y_2	Y_3
1	191	155	65	19	179	145	70
2	195	149	70	20	201	152	69
3	181	148	71	19	185	149	75
4	183	153	82	18	188	149	86
5	176	144	67	18	171	142	71
6	208	157	81	22	192	152	77
7	189	150	75	21	190	149	72
8	197	159	90	20	189	152	82
9	188	152	76	19	197	159	84
10	192	150	78	20	187	151	72
11	179	158	99	18	186	148	89
12	183	147	65	18	174	147	70
13	174	150	71	19	185	152	65
14	190	159	91	19	195	157	99
15	188	151	98	20	187	158	87
16	163	137	59	18	161	130	63
17	195	155	85	20	183	158	81
18	196	153	80	21	173	148	74
19	181	145	77	20	182	146	70
20	175	140	70	19	165	137	81
21	192	154	69	20	185	152	63
22	174	143	79	20	178	147	73
23	176	139	70	20	176	143	69

Program

Description

The canonical correlation program consists of the main routine named MCANO, a special input subroutine named DATA, and five subroutines from the Scientific Subroutine Package: CORRE, CANOR, MINV, NROOT, and EIGEN.

Capacity

The capacity of the sample program and the format required for data input have been set up as follows:

1. Up to 20 variables, including both the first set of variables (that is, left-hand variables) and the second set of variables (that is, right-hand variables) The number of variables in the first set must be greater than or equal to the number of variables in the second set.
2. Up to 99,999 observations
3. (12F6.0) format for input data cards

Therefore, if a problem satisfies the above conditions it is not necessary to modify the sample program. However, if there are more than 20 variables, dimension statements in the sample main program must be modified to handle the particular problem. Similarly, if input data cards are prepared using a different format, the input format in the input subroutine, DATA, must be modified. The general rules for program modification are described later.

Input

Control Card

One control card is required for each problem and is read by the main program, MCANO. This card is prepared as follows:

Columns	Contents	For Sample Problem
1 - 6	Problem number (may be alphameric)	SAMPLE
7 - 11	Number of observations	00023
12 - 13	Number of variables in the first set (that is, left-hand variables)*	04
14 - 15	Number of variables in the second set (that is, right-hand variables)	03

*The number of variables in the first set must be greater than or equal to the number of variables in the second set.

Leading zeros are not required to be keypunched.

Data Cards

Since input data are read into the computer one observation at a time, each row of data in Table 6 is keypunched on a separate card using the format (12F6.0). This format assumes twelve 6-column fields per card.

If there are more than twelve variables in a problem, each row of data is continued on the second card until the last data point is keypunched. However, each row of data must begin on a new card.

Deck Setup

The deck setup is shown in Figure 47.

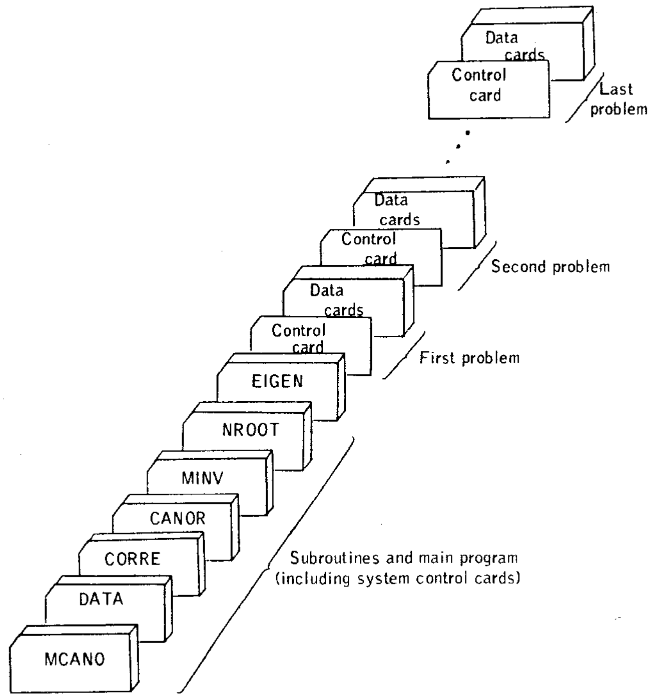

Figure 47. Deck setup (canonical correlation)

Sample

The listing of input cards for the sample problem is presented in Figure 48.

Output

Description

The output of the sample program for canonical correlation includes:

1. Means
2. Standard deviations
3. Correlation coefficients
4. Eigenvalues and corresponding canonical correlation
5. Lambda
6. Chi-square and degrees of freedom
7. Coefficients for left- and right-hand variables

Sample

The output listing for the sample problem is shown in Figure 49 of this sample problem.

```
/DATA                                                          10
SAMPLE000230403                                                20
  191   155    65    19   179   145    70                      30
  195   149    70    20   201   152    69                      40
  181   148    71    19   185   149    75                      50
  183   153    82    18   188   149    86                      60
  176   144    67    18   171   142    71                      70
  208   157    81    22   192   152    77                      80
  189   150    75    21   190   149    72                      90
  197   159    90    20   189   152    82                     100
  188   152    76    19   197   159    84                     110
  192   150    78    20   187   151    72                     120
  179   158    99    18   186   148    89                     130
  183   147    65    18   174   147    70                     140
  174   150    71    19   185   152    65                     150
  190   159    91    19   195   157    99                     160
  188   151    98    20   187   158    87                     170
  163   137    59    18   161   130    63                     180
  195   155    85    20   183   158    81                     190
  196   153    80    21   173   148    74                     200
  181   145    77    20   182   146    70                     210
  175   140    70    19   165   137    81                     220
  192   154    69    20   185   152    63                     230
  174   143    79    20   178   147    73                     240
  176   139    70    20   176   143    69                     250
```

Figure 48. Input card listing (canonical correlation)

Program Modification

Program capacity can be increased or decreased by making changes in dimension statements. Input data in a different format can also be handled by providing a specific format statement. In order to familiarize the user with the program modification, the following general rules are supplied in terms of the sample problem.

1. Changes in the dimension statements of the main program, MCANO:
   a. The dimension of arrays XBAR, STD, CANR, CHISQ, and NDF must be greater than or equal to the total number of variables m (m = p + q, where p is the number of left-hand variables and q is the number of right-hand variables). Since there are seven variables, four on left and three on right, the value of m is 7.

```
CANONICAL CORRELATION.....SAMPLE

  NO. OF OBSERVATIONS          23
  NO. OF LEFT HAND VARIABLES    4
  NO. OF RIGHT HAND VARIABLES   3

MEANS
    185.47826     149.91304     76.86955     19.47826     183.00000     148.82609     75.73912

STANDARD DEVIATIONS
     10.10342       6.31673     10.46338      1.08165       9.84424       6.73965      9.05647

CORRELATION COEFFICIENTS

ROW  1
    1.00000    0.74852    0.37082    0.66441    0.62291    0.66080    0.24683

ROW  2
    0.74852    1.00000    0.63252    0.22590    0.66811    0.72780    0.53194

ROW  3
    0.37082    0.63252    1.00000    0.20657    0.47394    0.60169    0.79684

ROW  4
    0.66441    0.22590    0.20657    1.00000    0.32870    0.34863   -0.10733

ROW  5
    0.62291    0.66811    0.47394    0.32870    1.00000    0.82555    0.39258

ROW  6
    0.66080    0.72780    0.60169    0.34863    0.82555    1.00000    0.47657

ROW  7
    0.24683    0.53194    0.79684   -0.10733    0.39258    0.47657    1.00000


  NUMBER OF      LARGEST     CORRESPONDING                          DEGREES
 EIGENVALUES    EIGENVALUE     CANONICAL      LAMBDA    CHI-SQUARE     OF
   REMOVED      REMAINING     CORRELATION                            FREEDOM

      0          0.79880        0.89376      0.11598    40.93277       12
      1          0.41910        0.64738      0.57644    10.46676        6
      2          0.00767        0.08760      0.99233     0.14636        2

CANONICAL CORRELATION     0.89376

  COEFFICIENTS FOR LEFT HAND VARIABLES
      0.66310      -0.16059       1.05822       -0.56651

  COEFFICIENTS FOR RIGHT HAND VARIABLES
     -0.02133       0.44090       0.89730

CANONICAL CORRELATION     0.64738

  COEFFICIENTS FOR LEFT HAND VARIABLES
      0.09454      -0.83915       0.66309       -0.64892

  COEFFICIENTS FOR RIGHT HAND VARIABLES
     -0.43841      -0.55503       0.70692

CANONICAL CORRELATION     0.08760

  COEFFICIENTS FOR LEFT HAND VARIABLES
      0.02681       0.36055      -0.28827       -0.32496

  COEFFICIENTS FOR RIGHT HAND VARIABLES
      0.70325      -0.70384       0.10028

1F0031217
```

Figure 49. Output listing (canonical correlation)

b. The dimension of array RX must be greater than or equal to the product of m x m. For the sample problem this product is 49 = 7 x 7.

c. The dimension of array R must be greater than or equal to (m + 1) m/2. For the sample problem this number is 28 = (7 + 1)7/2.

d. The dimension of array COEFL must be greater than or equal to the product of p x q. For the sample problem this product is 12 = 4 x 3.

e. The dimension of array COEFR must be greater than or equal to the product of q x q. For the sample problem this product is 9 = 3 x 3.

2. Changes in the input format statement of the special input subroutine, DATA:

Only the format statement for input data may be changed. Since sample data are either two- or three-digit numbers, rather than using six-column fields as in the sample problem, each row of data may be keypunched in seven 3-column fields, and, if so, the format is changed to (7F3.0).

The special input subroutine, DATA, is normally written by the user to handle different formats for different problems. The user may modify this subroutine to perform testing of input data, transformation of data, and so on.

Operating Instructions

The sample program for canonical correlation is a standard FORTRAN program. Special operating instructions are not required. Data set 5 is used for input, and data set 6 is used for output.

Timing

The execution time of this sample program on a System/360, Model 30, using an IBM 2540 Card Reader as input and an IBM 1403, Model 3 as output, is 19 seconds.

```
C
C     ..................................................................
C
C        SAMPLE MAIN PROGRAM FOR CANONICAL CORRELATION - MCANO
C
C        PURPOSE
C           (1) READ THE PROBLEM PARAMETER CARD FOR A CANONICAL
C           CORRELATION, (2) CALL TWO SUBROUTINES TO CALCULATE SIMPLE
C           CORRELATIONS, CANONICAL CORRELATIONS, CHI-SQUARES, DEGREES
C           OF FREEDOM FOR CHI-SQUARES, AND COEFFICIENTS FOR LEFT AND
C           RIGHT HAND VARIABLES, NAMELY CANONICAL VARIATES, AND (3)
C           PRINT THE RESULTS.
C
C        REMARKS
C           THE NUMBER OF LEFT HAND VARIABLES MUST BE GREATER THAN
C           OR EQUAL TO THE NUMBER OF RIGHT HAND VARIABLES.
C
C        SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED
C           CORRE  (WHICH, IN TURN, CALLS THE INPUT SUBROUTINE NAMED
C                  DATA.)
C           CANOR  (WHICH, IN TURN, CALLS THE SUBROUTINES MINV AND
C                  NROOT.  NROOT, IN TURN, CALLS THE SUBROUTINE EIGEN.)
C
C        METHOD
C           REFER TO W. W. COOLEY AND P. R. LOHNES, 'MULTIVARIATE PRO-
C           CEDURES FOR THE BEHAVIORAL SCIENCES', JOHN WILEY AND SONS,
C           1962, CHAPTER 3.
C
C     ..................................................................
C
C     THE FOLLOWING DIMENSIONS MUST BE GREATER THAN OR EQUAL TO THE
C     TOTAL NUMBER OF VARIABLES M (M=MP+MQ, WHERE MP IS THE NUMBER OF
C     LEFT HAND VARIABLES, AND MQ IS THE NUMBER OF RIGHT HAND VARI-
C     ABLES)..
C
      DIMENSION XBAR(20),STD(20),CANR(20),CHISQ(20),NOF(20)
C
C     THE FOLLOWING DIMENSION MUST BE GREATER THAN OR EQUAL TO THE
C     PRODUCT OF M*M..
C
      DIMENSION RX(400)
C
C     THE FOLLOWING DIMENSION MUST BE GREATER THAN OR EQUAL TO
C     (M+1)*M/2..
C
      DIMENSION R(210)
C
C     THE FOLLOWING DIMENSION MUST BE GREATER THAN OR EQUAL TO THE
C     PRODUCT OF MP*MQ..
C
      DIMENSION COEFL(400)
C
C     THE FOLLOWING DIMENSION MUST BE GREATER THAN OR EQUAL TO THE
C     PRODUCT OF MQ*MQ..
C
      DIMENSION COEFR(400)
C
C     ..................................................................
C
C        IF A DOUBLE PRECISION VERSION OF THIS ROUTINE IS DESIRED, THE
C        C IN COLUMN 1 SHOULD BE REMOVED FROM THE DOUBLE PRECISION
C        STATEMENT WHICH FOLLOWS.
C
C     DOUBLE PRECISION XBAR,STD,RX,R,CANR,CHISQ,COEFL,COEFR
C
C        THE C MUST ALSO BE REMOVED FROM DOUBLE PRECISION STATEMENTS
C        APPEARING IN OTHER ROUTINES USED IN CONJUNCTION WITH THIS
C        ROUTINE.
C
C     ..................................................................
C
    1 FORMAT(A4,A2,I5,2I2)
    2 FORMAT(27H1CANONICAL CORRELATION......,A4,A2//22H    NO. OF OBSERVAT
     1IONS,8X,I4/29H    NO. OF LEFT HAND VARIABLES,I5/30H    NO. OF RIGHT
     3HAND VARIABLES,I4/)
    3 FORMAT(6H0MEANS/(8F15.5))
    4 FORMAT(20H0STANDARD DEVIATIONS/(8F15.5))
    5 FORMAT(25H0CORRELATION COEFFICIENTS)
    6 FORMAT(4H0ROW,I3/(10F12.5))
    7 FORMAT(1H0//12H    NUMBER OF,7X,7HLARGEST,7X,13HCORRESPONDING,31X,7
     1HDEGREES/13H  EIGENVALUES,5X,10HEIGENVALUE,7X,9HCANONICAL,7X,6HLAM
     2BDA,5X,10HCHI-SQUARE,7X,2HOF/4X,7HREMOVED,7X,9HREMAINING,7X,11HCOR
     3RELATION,32X,7HFREEDOM/)
    8 FORMAT(1H ,I7,F19.5,F16.5,2F14.5,5X,I5)
    9 FORMAT(1H0/22H CANONICAL CORRELATION,F12.5)
   10 FORMAT(39H0  COEFFICIENTS FOR LEFT HAND VARIABLES/(8F15.5))
   11 FORMAT(40H0  COEFFICIENTS FOR RIGHT HAND VARIABLES/(8F15.5))
C
C     ..................................................................
C
C     READ PROBLEM PARAMETER CARD
C
  100 READ (5,1) PR,PR1,N,MP,MQ
C        PR.......PROBLEM NUMBER (MAY BE ALPHAMERIC)
C        PR1......PROBLEM NUMBER (CONTINUED)
C        N........NUMBER OF OBSERVATIONS
C        MP.......NUMBER OF LEFT HAND VARIABLES
C        MQ.......NUMBER OF RIGHT HAND VARIABLES
C
      WRITE (6,2) PR,PR1,N,MP,MQ
C
      M=MP+MQ
      IO=0
      X=0.0
C
      CALL CORRE (N,M,IO,X,XBAR,STD,RX,R,CANR,CHISQ,COEFL)
C
C     PRINT MEANS, STANDARD DEVIATIONS, AND CORRELATION
C     COEFFICIENTS OF ALL VARIABLES
C
      WRITE (6,3) (XBAR(I),I=1,M)
      WRITE (6,4) (STD(I),I=1,M)
      WRITE (6,5)
      DO 160 I=1,M
      DO 150 J=1,M
      IF(I-J) 120, 130, 130
  120 L=I+(J*J-J)/2
      GO TO 140
  130 L=J+(I*I-I)/2
  140 CANR(J)=R(L)
  150 CONTINUE
  160 WRITE (6,6) I,(CANR(J),J=1,M)
C
      CALL CANOR (N,MP,MQ,R,XBAR,STD,CANR,CHISQ,NOF,COEFR,COEFL,RX)
C
C     PRINT EIGENVALUES, CANONICAL CORRELATIONS, LAMBDA, CHI-SQUARES,
C     DEGREES OF FREEDOMS
C
      WRITE (6,7)
      DO 170 I=1,MQ
      N1=I-1
C
C        TEST WHETHER EIGENVALUE IS GREATER THAN ZERO
C
      IF(XBAR(I)) 165, 165, 170
  165 MM=N1
      GO TO 175
  170 WRITE (6,8) N1,XBAR(I),CANR(I),STD(I),CHISQ(I),NOF(I)
      MM=MQ
C
C     PRINT CANONICAL COEFFICIENTS
C
  175 N1=0
      N2=0
      DO 200 I=1,MM
      WRITE (6,9) CANR(I)
      DO 180 J=1,MP
      N1=N1+1
  180 XBAR(J)=COEFL(N1)
      WRITE (6,10) (XBAR(J),J=1,MP)
      DO 190 J=1,MQ
      N2=N2+1
  190 XBAR(J)=COEFR(N2)
      WRITE (6,11) (XBAR(J),J=1,MQ)
  200 CONTINUE
      GO TO 100
      END
```