A Computer Program For Selecting Principal Variables

by

Regina P. Becker and George P. McCabe

Technical Report #84-45

Department of Statistics
Purdue University

October, 1984

PURDUE
UNIVERSITY DEPARTMENT OF STATISTICS

Enclosed is a technical report containing documentation and a listing of a FORTRAN program for performing the computations described in the paper, "Principal Variables," by George P. McCabe, Technometrics, 26, 137-144, May 1984.

The research and programming for this project was supported by the Purdue University Department of Statistics. Accordingly, this work is freely distributed to anyone who requests it.

If, however, you have funds available to support this type of research, we would be very grateful to receive a charitable contribution of $100.00 or more. The enclosed form can be used for this purpose.

Mathematical Sciences Building
West Lafayette, Indiana 47907

Contribution to the

Purdue Research Foundation*

Account No. 7492-08-1551
to support the research of the
Purdue University Statistics Department

Date _____

Amount _____

From _____

_____

_____

_____

*Please make check payable to Purdue Research Foundation and put the above account number in the corner.

Mail to: Department of Statistics
Purdue University
West Lafayette, Indiana 47907

10/26/84 GPM

I. Introduction

This document is a supplement to the paper "Principal Variables" by George P. McCabe, Technometrics 26, 137-144 (May, 1984). The procedure is a variable selection alternative to principal components.

The program runs on the Purdue University Computing Center's CDC 6000 system. Some minor modifications may be needed for other computers.

Two subroutines from the IMSL package are needed to run the program. They are (1) EIGRS for computing eigenvalues and eigenvectors and (2) LEQT1P, a linear equation solver. Usage of the routines is described in comments contained in the program listing.

Users who do not have access to the IMSL package could substitute their own subroutines. It is possible that IMSL would allow a user to obtain copies of these two routines only for use in this program. For information, contact IMSL, NBC Building, 7500 Bellaire Blvd., Houston, Texas 77036; (713) 772-1927.

In Section II, the input specifications are described. The program use is illustrated with data used in the paper, "Predicting Success of Freshmen in a Computer Science Major" by Patricia F. Campbell and George P. McCabe, Communications of the ACM 27, (November 1984). The input is given in Section III and the output in Section IV. The complete program listing is given in Section V.

II. Input

PROGRAM: PVAR

ORDER OF CONTROL CARDS:

A. JOBCARD
B. PFILES(GET,DATAFILE,X=TAPE2) OPTIONAL
C. PFILES(GET,PVAR,ID=DNT)
D. MNF(I=PVAR,L=TRASH)
E. EOR
F. TITLE CARD
G. PARAMETER CARD
H. FORMAT CARD
I. (INPUT)
J. S

PREPARATION OF CARDS SPECIFIC TO THIS PROGRAM:

TITLE CARD	1-80	TITLE FOR THE OUTPUT OR USE BLANK CARD
PARAMETER CARD	1-5	NP= OF VARIABLES (2 LE NP LE 20)
	6-10	NBEST= OF BEST SUBSETS (NBEST LE 20; DEFAULT=10)
	11-15	ITAPE=5 IF DATA ON INPUT ITAPE=2 IF IN PFILES
	16-20	IP=0 IF DATA INPUT IP=1 IF CORRELATION MATRIX INPUT IP=2 IF COVARIANCE MATRIX INPUT
	21-25	IX=0 IF CORRELATION ANALYSIS REQUESTED IX=1 IF COVARIANCE ANALYSIS REQUESTED
	26-30	N=SAMPLE SIZE (NECESSARY ONLY IF IP=0)

FORMAT CARD

COLUMNS 1-80 MAY BE USED TO DESCRIBE THE DATA FORMAT. USE THE USUAL F TYPE FORTRAN FORMAT STATEMENT, OMITTING THE WORD ≠FORMAT≠.

MATRIX INPUT CARDS

IF A MATRIX IS INPUT, THE LOWER TRIANGULAR PORTION SHOULD BE USED. FOR EXAMPLE, IF THERE ARE 4 VARIABLES THE COVARIANCE MATRIX INPUT SHOULD BE

S11
S21 S22
S31 S32 S33
S41 S42 S43 S44

NOTE THAT SIJ IS THE COVARIANCE BETWEEN XI AND XJ

III. Input for Example

```
CS DATA
   11    5    5    1    0
(11F5.0)
1.00
 .376 1.00
 .174 .217 1.00
 .048-.010 .015 1.00
 .268 .119 .078-.018 1.00
 .352 .138 .582-.033 .187 1.00
 .216 .097 .032-.069 .095 .094 1.00
 .179 .243 .599-.066 .110 .531 .150 1.00
-.036 .070-.039 .063-.009 .047 .035 .063 1.00
 .073 .279 .682-.114 .132 .452-.006 .564 .028 1.00
-.183 .026 .287 .089-.023 .194-.206 .122 .084 .354 1.00
```

IV. Output for Example

```
CS DATA


NUMBER OF VARIABLES = 11     NUMBER OF BEST SUBSETS =  5
INPUT FOLLOWS PARAMETER CARD AND FORMAT CARD
IP = 1 :  INPUT IS CORRELATION MATRIX

CORRELATION MATRIX ANALYSIS REQUESTED


   MATRIX W

 1.00000000E+00
 3.76000000E-01  1.00000000E+00
 1.74000000E-01  2.17000000E-01  1.00000000E+00
 4.80000000E-02 -1.00000000E-02  1.50000000E-02  1.00000000E+00
 2.68000000E-01  1.19000000E-01  7.80000000E-02 -1.80000000E-02  1.00000000E+00
 3.52000000E-01  1.38000000E-01  5.82000000E-01 -3.30000000E-02  1.87000000E-01  1.00000000E+00
 2.16000000E-01  9.70000000E-02  3.20000000E-02 -6.90000000E-02  9.50000000E-02  9.40000000E-02  1.00000000
 1.79000000E-01  2.43000000E-01  5.99000000E-01 -6.60000000E-02  1.10000000E-01  5.31000000E-01  1.50000000
 1.00000000E+00
-3.60000000E-02  7.00000000E-02 -3.90000000E-02  6.30000000E-02 -9.00000000E-03  4.70000000E-02  3.50000000
 6.30000000E-02  1.00000000E+00
 7.30000000E-02  2.79000000E-01  6.82000000E-01 -1.14000000E-01  1.32000000E-01  4.52000000E-01 -6.00000000
 5.64000000E-01  2.80000000E-02  1.00000000E+00
-1.83000000E-01  2.60000000E-02  2.87000000E-01  8.90000000E-02 -2.30000000E-02  1.94000000E-01 -2.66000000
 1.22000000E-01  8.40000000E-02  3.54000000E-01  1.00000000E+00


PERCENT VARIATION EXPLAINED BY FIRST  1 PRINCIPAL  COMPONENTS =  28.0019
PERCENT VARIATION EXPLAINED BY FIRST  2 PRINCIPAL  COMPONENTS =  43.0156
PERCENT VARIATION EXPLAINED BY FIRST  3 PRINCIPAL  COMPONENTS =  53.2307
PERCENT VARIATION EXPLAINED BY FIRST  4 PRINCIPAL  COMPONENTS =  62.5381
PERCENT VARIATION EXPLAINED BY FIRST  5 PRINCIPAL  COMPONENTS =  70.9121
PERCENT VARIATION EXPLAINED BY FIRST  6 PRINCIPAL  COMPONENTS =  79.0403
PERCENT VARIATION EXPLAINED BY FIRST  7 PRINCIPAL  COMPONENTS =  85.3768
PERCENT VARIATION EXPLAINED BY FIRST  8 PRINCIPAL  COMPONENTS =  90.8268
PERCENT VARIATION EXPLAINED BY FIRST  9 PRINCIPAL  COMPONENTS =  94.4401
PERCENT VARIATION EXPLAINED BY FIRST 10 PRINCIPAL  COMPONENTS =  97.5933
PERCENT VARIATION EXPLAINED BY FIRST 11 PRINCIPAL  COMPONENTS = 100.0000


PCT VAR        DETERMINANT    1 VARIABLES
EXPLAINED

 13.5318       1.0000000E+00   1
 12.4877       1.0000000E+00      2
 21.1931       1.0000000E+00         3
  9.4368       1.0000000E+00            4
 10.6047       1.0000000E+00               5


PCT VAR        DETERMINANT    2 VARIABLES
EXPLAINED
```

30.9353	9.9996400E-01						7			10
19.9424	9.9991900E-01				5				9	
21.9170	9.9990000E-01		2		4					
30.6588	9.9977500E-01			3	4					
20.0281	9.9967600E-01				4	5				

PCT VAR EXPLAINED DETERMINANT 3 VARIABLES

40.2236	9.9794324E-01						7		9	10
41.1892	9.9614264E-01			3			7		9	
29.3037	9.9564641E-01				4	5			9	
40.0146	9.9421129E-01			3	4				9	
41.1451	9.9392376E-01			3	4		7			

PCT VAR EXPLAINED DETERMINANT 4 VARIABLES

50.1030	9.8777461E-01			3	4	5			9	
50.4810	9.8673919E-01			3	4		7		9	
50.9964	9.8146272E-01			3		5	7		9	
41.6931	9.8118676E-01				4	5			9	11
39.5252	9.8054474E-01				4	5	7		9	

PCT VAR EXPLAINED DETERMINANT 5 VARIABLES

60.2840	9.7205154E-01			3	4	5	7		9	
53.3523	9.6162147E-01		2		4	5			9	11
50.9956	9.5481499E-01		2		4	5	7		9	
58.9425	9.4919426E-01				4	5	7		9	10
58.2285	9.4898451E-01		2		4		6	7	9	

PCT VAR EXPLAINED DETERMINANT 6 VARIABLES

69.8431	9.0396156E-01	2	3	4	5		7		9		
67.1516	9.0209715E-01	2		4	5	6	7		9		
68.5351	8.7822365E-01	2		4	5			8	9		11
68.1789	8.7557054E-01	2		4	5	6			9		11
67.0805	8.7321640E-01	2		4	5		7	8	9		

PCT VAR EXPLAINED DETERMINANT 7 VARIABLES

76.4472	7.8210629E-01		2		4	5	6	7		9		11
76.7044	7.7679094E-01		2		4	5		7	8	9		11
77.9504	7.4723553E-01		2	3	4	5		7		9		11
76.5622	7.0707830E-01	1			4	5		7	8	9		11
77.7637	6.9325224E-01	1	2	3	4	5		7		9		

PCT VAR EXPLAINED DETERMINANT 8 VARIABLES

84.2937	5.7750540E-01	1	2		4	5		7	8	9		11
85.2694	5.4846126E-01	1	2	3	4	5		7		9		11
84.5984	5.2893311E-01		2		4	5	6	7	8	9		11
82.6012	5.1473237E-01	1	2		4	5	6	7		9		11
85.1882	5.1314586E-01		2		4	5	6	7		9	10	11

PCT VAR EXPLAINED DETERMINANT 9 VARIABLES

90.5872	3.4523410E-01	1	2		4	5	6	7	8	9		11
91.0771	3.3184464E-01	1	2		4	5	6	7		9	10	11
91.3669	3.2939410E-01	1	2	3	4	5		7	8	9		11
90.8010	3.2312488E-01	1	2		4	5		7	8	9	10	11
90.8059	3.0805189E-01	1	2	3	4	5	6	7		9		11

PCT VAR EXPLAINED DETERMINANT 10 VARIABLES

96.4037	1.8581055E-01	1	2		4	5	6	7	8	9	10	11
96.1070	1.7164826E-01	1	2	3	4	5	6	7	8	9		11
95.2661	1.4115728E-01	1	2	3	4	5		7	8	9	10	11
95.2236	1.3990163E-01	1	2	3	4	5	6	7		9	10	11
94.1364	1.1396268E-01		2	3	4	5	6	7	8	9	10	11

PCT VAR EXPLAINED DETERMINANT 11 VARIABLES

100.0000	7.3505052E-02	1	2	3	4	5	6	7	8	9	10	11

V. Program Listing

```
      PROGRAM PVAR(INPUT,OUTPUT,TAPE5=INPUT,TAPE6=OUTPUT)
C*************************************************************************
C*                                                                       *
C*    THIS PROGRAM IS AN ALGORITHM FOR THE SUBSET SELECTION              *
C*      PROCEDURE DESCRIBED IN THE PAPER:                                *
C*                                                                       *
C*      MCCABE, GEORGE P. (1984), ~PRINCIPAL VARIABLES,~                 *
C*      TECHNOMETRICS, 26, 137-144.                                      *
C*                                                                       *
C*            DEPARTMENT OF STATISTICS                                   *
C*              PURDUE UNIVERSITY                                        *
C*            WEST LAFAYETTE, IN 47907                                   *
C*                                                                       *
C*  PROGRAMMER- REGINA BECKER                                            *
C*  DEPARTMENT OF STATISTICS                                             *
C*  PURDUE UNIVERSITY                                                    *
C*  JUNE 1984                                                            *
C*                                                                       *
C*************************************************************************
C
C
C     ***INPUT INFORMATION****
C
C    VARIABLES: NP = NUMBER OF INDEPENDENT VARIABLES (MUST BE BETWEEN
C                    2 AND 20 INCLUSIVELY)
C
C               NBEST = THE NUMBER OF BEST SUBSETS TO BE PRINTED
C
C               ITAPE = THE INFORMATIONAL INPUT SOURCE
C                       5 IF INFORMATION IS ON INPUT
C                       2 IF INFORMATION IS ON TAPE
C
C               IP = INFORMATIONAL INPUT CODE
C                    0 IF DATA IS INPUT
C                    1 IF CORRELATION MATRIX IS INPUT
C                    2 IF COVARIANCE MATRIX IS INPUT
C
C              IX = TYPE OF ANALYSIS REQUESTED
C                   0 IF CORRELATION ANALYSIS IS REQUESTED
C                   1 IF COVARIANCE ANALYSIS IS REQUESTED
C
C               N = SAMPLE SIZE (NECESSARY ONLY IF IP=0)
C
C    ORDER OF INPUT:
C
C     1. TITLE CARD              TITLE FOR OUTPUT OR USE BLANK CARD
C
C     2. PARAMETER CARD          NP, NBEST, ITAPE, IP, IX, N
C                                USE (6I5) FORMAT
C
C     3. FORMAT CARD             COLUMNS 1-80 MAY BE USED TO DESCRIBE THE
C                                INFORMATIONAL INPUT FORMAT. USE THE USUAL
C                                F TYPE FORMAT STATEMENT, OMITTING THE
C                                WORD FORMAT.
C
C     4. MATRIX INPUT CARDS
C
C        IF A MATRIX IS INPUT THE LOWER TRIANGULAR PORTION SHOULD BE
```

```
C        USED.  FOR EXAMPLE, IF THERE ARE 4 VARIABLES THE COVARIANCE
C        MATRIX INPUT SHOULD BE:
C
C              S11
C              S21 S22
C              S31 S32 S33
C              S41 S42 S43 S44
C
C       NOTE: SIJ IS THE COVARIANCE BETWEEN XI AND XJ
C*************************************************************************
C
C       THE FOLLOWING SUBROUTINES ARE CALLED BY THE PROGRAM:
C
C       1. DISC -   DETERMINES BEST SUBSET AND PCT VARIANCE EXPLAINED
C
C       2. EIGRS -  FROM THE IMSL LIBRARY, THIS SUBROUTINE CALCULATES
C                   EIGENVALUES AND EIGENVECTORS
C
C       3. LEQT1P - FROM THE IMSL LIBRARY, THIS SUBROUTINE SOLVES THE
C                   SYSTEM OF LINEAR EQUATIONS A*X=B FOR X.
C*************************************************************************
C
C       SUBROUTINE DISC
C
C       USAGE: CALL DISC(W,NP,INDBLK,INDROW NBEST)
C
C              W    IS THE ARRAY WHICH CONTAINS THE CORRELATION OR
C                   COVARIANCE MATRIX AND ITS SUBMATRICES
C
C              NP   IS THE NUMBER OF INDEPENDENT VARIABLES
C
C              INDBLK(I) IS THE ENDING POSITION OF MATRIX (I-1)
C                   IN ARRAY W. THE SUBMATRICES ARE STORED IN
C                   SYMMETRIC STORAGE MODE. THE STARTING
C                   POSITION OF MATRIX I IS INDBLK(I)+1.
C
C              INDROW(I) IS THE NUMBER OF ELEMENTS NEEDED TO STORE
C                   SUBMATRIX I IN SYMMETRIC STORAGE MODE
C
C              NBEST  IS THE NUMBER OF BEST SUBSETS TO BE PRINTED
C*************************************************************************
C
C       SUBROUTINE EIGRS
C
C       USAGE: CALL EIGRS(W,NP,JOBN,D,Z,IZ,WK,IER)
C
C              W     INPUT REAL SYMMETRIC MATRIX OF ORDER NP, WHOSE
C                    EIGENVALUES ARE TO BE COMPUTED
C
C              NP    INPUT ORDER OF THE MATRIX W
C
C              JOBN  INPUT OPTION PARAMETER. SET TO 1 TO COMPUTE
C                    EIGENVALUES AND EIGENVECTORS.
C
C              D     OUTPUT VECTOR OF LENGTH NP CONTAINING THE
C                    EIGENVALUES OF W.
C
```

```
C
C               Z      OUTPUT NP BY NP MATRIX CONTAINING EIGENVECTORS
C                      OF W (STORED IN COLUMNS)
C
C               IZ     INPUT ROW DIMENSION OF MATRIX Z AS SPECIFIED
C                      IN DIMENSION STATEMENT OF CALLING PROGRAM (IZ=20) .
C
C               WK     WORK AREA NEEDED. WK IS DIMENSIONED NP.
C
C               IER    ERROR PARAMETER
C*****************************************************************************
C
C      SUBROUTINE LEQT1P
C
C       USAGE: CALL LEQT1P(WP,M,N,BB,IB,IDGT,D1,D2,IER)
C
C              WP     INPUT VECTOR OF LENGTH N(N+1)/2 CONTAINING THE
C                     N BY N COEFFICIENT MATRIX OF THE EQUATION AX=B.
C                     WP IS STORED IN SYMMETRIC STORAGE MODE AND IS
C                     POSITIVE DEFINITE.
C
C              M      NUMBER OF RIGHT-HAND SIDES (COLUMNS IN B)
C
C              N      ORDER OF A AND NUMBER OF ROWS IN B
C
C              BB     INPUT MATRIX OF DIMENSION N BY M CONTAINING THE
C                     RIGHT HAND SIDES OF THE EQUATION AX=B
C
C              IB     ROW DIMENSION OF BB EXACTLY AS SPECIFIED IN
C                     CALLING PROGRAM (IB=20)
C
C              IDGT   THE ELEMENTS OF A ARE ASSUMED TO BE CORRECT TO
C                     IDGT DECIMAL DIGITS
C
C              D1,D2  COMPONENTS OF THE DETERMINANT OF A.
C
C              IER    ERROR PARAMETER
C*****************************************************************************

      DIMENSION W(1540),INDBLK(20),INDROW(20)
      DIMENSION XMSUM(20,20),XSUM(20),NM(20),X(20)
      DIMENSION FMT(8),A(80),SXY(20,20),SX(20)
      DIMENSION Z(20,20),D(20),WK(20),AA(210)

      DATA INDBLK/0,1,4,10,20,35,56,84,120,165,220,286,364,455,560,680,
     1 816,969,1140,1330/
      DATA INDROW/0,1,3,6,10,15,21,28,36,45,55,66,78,91,105,120,136,
     1153,171,190/

C     READ TITLE CARD
      READ(5,400) (A(I),I=1,80)
400   FORMAT(80A1)
      WRITE(6,410) (A(I),I=1,80)
410   FORMAT(≠1≠,80A1//)

C     READ PARAMETER CARD
```

```
      READ(5,600) NP,NBEST,ITAPE,IP,IX,N
600   FORMAT(6I5)

C     READ FORMAT CARD
      READ(5,601) FMT
601   FORMAT(8A10)

      IF(NP.LE.1.OR.NP.GT.20)THEN
      WRITE(6,700)
700   FORMAT(1X,≠ERROR-NUMBER OF VARIABLES MUST BE BETWEEN 2 AND 20≠//)
      GOTO 80
      ENDIF

      IF(ITAPE.NE.5.AND.ITAPE.NE.2)THEN
      WRITE(6,710)
710   FORMAT(1X,≠ERROR - INCORRECT INPUT MEDIUM SPECIFIED≠/)
      GOTO80
      ENDIF

      IF(NBEST.LT.1.OR.NBEST.GT.20)NBEST=10
      IF(IP.LT.0.OR.IP.GT.2)THEN
      WRITE(6,720)
720   FORMAT(1X,≠ERROR ON PARAMETER CARD - DATA INPUT SPECIFICATION≠/)
      GOTO 80
      ENDIF

      IF(IX.NE.0.AND.IX.NE.1)THEN
      WRITE(6,730)
730   FORMAT(1X,≠ERROR ON PARAMETER CARD-ANALYSIS SPECIFICATION≠/)
      GOTO 80
      ENDIF

      IF(IX.EQ.1.AND.IP.EQ.1)THEN
      WRITE(6,740)
740   FORMAT(1X,≠DATA OR COVARIANCE MATRIX NEEDED FOR COVARIANCE ANALYSI
     *S≠/)
      GOTO 80
      ENDIF

      WRITE(6,405) NP,NBEST
405   FORMAT(1X,≠NUMBER OF VARIABLES = ≠,I2,5X,≠NUMBER OF BEST SUBSETS =
     * ≠,I2)
      IF(ITAPE.EQ.5)THEN
      WRITE(6,406)
406   FORMAT(1X,≠INPUT FOLLOWS PARAMETER CARD AND FORMAT CARD≠)
      ELSE
      WRITE(6,407)
407   FORMAT(1X,≠INPUT FROM PFILES VIA TAPE2≠)
      ENDIF

      IF(IP.EQ.0)WRITE(6,408)
      IF(IP.EQ.1) WRITE(6,409)
      IF(IP.EQ.2)WRITE(6,411)
408   FORMAT(1X,≠IP = 0 :  DATA INPUT≠/)
409   FORMAT(1X,≠IP = 1 :  INPUT IS CORRELATION MATRIX ≠/)
411   FORMAT(1X,≠IP = 2 :  INPUT IS COVARIANCE MATRIX≠/)

      IF(IX.EQ.0)THEN
      WRITE(6,750)
750   FORMAT(1X,≠CORRELATION MATRIX ANALYSIS REQUESTED≠//)
```

```
      ENDIF
      IF(IX.EQ.1)THEN
      WRITE(6,760)
760   FORMAT(1X,≠COVARIANCE MATRIX ANALYSIS REQUESTED≠//)
      ENDIF


      IA=INDBLK(NP)

      IF(IP.EQ.0)THEN

      DO 425 I=1,NP
      DO 425 J=1,NP
      SXY(I,J)=0.0
      SX(I)=0.0
425   CONTINUE

C     READ INPUT
      DO 460 I=1,N
      READ(ITAPE,FMT) (X(J),J=1,NP)
      DO 460 J=1,NP
      SX(J)=SX(J)+X(J)
      DO 460 II=1,J
      SXY(J,II)=X(J)*X(II)+SXY(J,II)
460   CONTINUE

      DO 480 I=1,NP
      DO 480 J=1,I
      XMSUM(I,J)=(SXY(I,J)-(SX(I)*SX(J)/FLOAT(N)))/FLOAT(N-1)
480   CONTINUE

      IF(IX.EQ.1)GO TO 342
      IF(IX.EQ.0)GOTO 322
      ENDIF


      DO 320 I=1,NP
      READ(ITAPE,FMT) (XMSUM(I,J),J=1,I)
320   CONTINUE

      IF(IP.EQ.2.AND.IX.EQ.1)GO TO 342
      IF(IP.EQ.2.AND.IX.EQ.0)GO TO 322
      IF(IP.EQ.1) GO TO 342
322   DO 330 I=2,NP
      DO 330 J=1,I-1
      XMSUM(I,J)=XMSUM(I,J)/SQRT(XMSUM(I,I)*XMSUM(J,J))
330   CONTINUE

      DO 340 I=1,NP
      XMSUM(I,I)=1.0
340   CONTINUE


342   DO 350 I=1,NP
      DO 350 J=1,I
      IBR=IA+INDROW(I)
      W(IBR+J)=XMSUM(I,J)
      AA(INDROW(I)+J)=W(IBR+J)
350   CONTINUE
```

```
      WRITE(6,800)
  800 FORMAT(≠     MATRIX W≠/)

      DO 110 I=1,NP
      IBR = IA + INDROW(I)
      WRITE(6,810) (W(IBR+J),J=1,I)
  810 FORMAT(7E16.8)
  110 CONTINUE
      WRITE(6,111)
  111 FORMAT(1X,///)

C     CALCULATE EIGENVALUES AND EIGENVECTORS.  CURRENTLY, THE
C     EIGENVECTORS ARE NOT USED.

      JOBN=1
      IZ=20
      CALL EIGRS(AA,NP,JOBN,D,Z,IZ,WK,IER)

C     D RETURNS EIGENVALUES, SMALLEST TO LARGEST
C     Z CONTAINS EIGENVECTORS CORRESPONDING TO D

      IF (IER.EQ.0)GO TO 56
      J=MOD(IER,128)
      WRITE(6,820) J
  820 FORMAT(1X,≠EIGENVALUE ROUTINE FAILS TO CONVERGE ON EIGENVALUE ≠,
     *I2, ≠.  PROGRAM STOPS.≠//)
      GO TO 80

C     REORDER EIGENVALUES AND EIGENVECTORS
   56 KK=NP/2
      DO 352 I=1,KK
      TEMP=D(I)
      D(I)=D(NP-I+1)
      D(NP-I+1)=TEMP

      DO 341 J=1,NP
      TEMP=Z(J,I)
      Z(J,I)=Z(J,NP-I+1)
      Z(J,NP-I+1)=TEMP
  341 CONTINUE
  352 CONTINUE


C     CALCULATE PCT VARIANCE EXPLAINED BY FIRST K PRINCIPAL COMPONENTS

      SUM=0.0
      DO 360 I=1,NP
      SUM=SUM+D(I)
  360 CONTINUE

      PCSUM=0.0
      DO 370 I=1,NP
      PCSUM=PCSUM+D(I)
      PCVAR=100.*PCSUM/SUM

      WRITE(6,380) I, PCVAR
  380 FORMAT(1X,≠PERCENT VARIATION EXPLAINED BY FIRST ≠,I2,≠ PRINCIPAL
     *COMPONENTS = ≠,F8.4)
  370 CONTINUE
      CALL DISC(W,NP,INDBLK,INDROW,NBEST)
```

```
80    STOP
      END


      SUBROUTINE DISC(W,N,INDBLK,INDROW,NBEST)
      DIMENSION W(1),T(1),INDBLK(1),INDROW(1),SAVE(20),NT(21)
      DIMENSION NX(20),USAVE(200),ISAVE(200),IPW(20),NAL(20)
      DIMENSION  BB(20,20), WP(210)
      DIMENSION BT(20,20)
      INTEGER BLANK
      DATA BLANK/≠ ≠/
      DATA NAL/≠1≠,≠2≠,≠3≠,≠4≠,≠5≠,≠6≠,≠7≠,≠8≠,≠9≠,≠10≠,
     *         ≠11≠,≠12≠,≠13≠,≠14≠,≠15≠,≠16≠,≠17≠,≠18≠,≠19≠,≠20≠/
      DO 3 I=1,N
      SAVE(I)=0.0
3     CONTINUE
      NX(N)=-10
      KOUNT=0
      NM=N-1
      NA=N+1
      NT(NA)=1
      DO 1 I=1,200
      USAVE(I)=0.0
    1 ISAVE(I)=0
      DO 5 I=1,N
    5 NT(I)=0
   10 DO 20 I=1,NM
      IF(NT(I).EQ.0)GO TO 21
      NT(I)=0
   20 CONTINUE
      I=N
      IF(NT(I).NE.0) GO TO 100
      NT(I)=1
      KOUNT=KOUNT+1
      IM=I-1
      GO TO 23
   21 NT(I)=1
      KOUNT=KOUNT+1
      K=I+1
      IM=I-1
      DO 22 J=K,N
      IF(NT(J).EQ.1)GO TO 30
   22 CONTINUE
   23 JM=N
      IA=INDBLK(JM)
      IC=IA+INDROW(I)+I
      U=W(IC)
      GO TO 29
   30 JM=J-1
      IA=INDBLK(JM)
      IC=IA+INDROW(I)+I
      U=SAVE(JM)*W(IC)
   29 INUM=NX(JM)+10
      IF(USAVE(INUM+10).GE.U) GO TO 70
      DO 50 LOOPA=1,9
      LOOP=11-LOOPA
      ID=INUM+LOOP
      IB=ID-1
      IF(U.LE.USAVE(IB)) GO TO 65
```

```
      ISAVE(ID)=ISAVE(IB)
   50 USAVE(ID)=USAVE(IB)
      ID=IB
   65 USAVE(ID)=U
      ISAVE(ID)=KOUNT
   70 IF(I.EQ.1) GO TO 10
      SAVE(IM)=U
      NX(IM)=NX(JM)+10
      IB=INDBLK(IM)
      ID=IA+INDROW(I)
      DO 31 L=1,IM
      IF=ID+L
      A=W(IF)/W(IC)
      LP=INDROW(L+1)
      IH=IA+LP
      II=IB+LP
      DO 31 M=L,IM
      IJ=ID+M
      W(II)=W(IH)-A*W(IJ)
      II=II+M
   31 IH=IH+M
      GO TO 10
  100 N10=10*N-9
      IA=INDBLK(N)

      SUM=0.0
      DO 362 II=1,N
      IW=INDROW(II)+II+IA
      SUM=SUM+W(IW)
  362 CONTINUE

      DO 120 I=1,N10,10
      IUSE=I-1
      II = IUSE/10+1
      WRITE(6,601) II
  601 FORMAT(///2X,≠PCT VAR≠,7X,≠DETERMINANT  ≠,I2, ≠ VARIABLES≠/
     *2X,≠EXPLAINED≠//)


      DO 115 J=1,NBEST
C     BLANK OUT IPW FOR FORMATTING

      DO 555 KK=1,20
        IPW(KK)=BLANK
  555 CONTINUE

      KOUNT=ISAVE(IUSE+J)
      IF(KOUNT.EQ.0) GO TO 120
      L=0
      DO 110 K=1,N
      IF(KOUNT-KOUNT/2*2.EQ.0) GO TO 110
      L=L+1
      NT(L)=K
  110 KOUNT = KOUNT/2

C     SET THE BEST SUBSET INDICATORS IN IPW FOR PRINTING
      DO 558 KK=1,L
       IPW(NT(KK))=NAL(NT(KK))
  558 CONTINUE
```

```
      IF(L.EQ.N)THEN
      RES=100.0
      GOTO 115
      ENDIF
      KL=1
      KK=1
      DO 830 II=1,N
      IF(NT(KK).EQ.II)THEN
      KK=KK+1
      ELSE
      NT(L+KL)=II
      KL=KL+1
      ENDIF
830   CONTINUE

C     LOAD SIGMA 11 IN WP IN SYMMETRIC STORAGE MODE

      DO 840 II=1,L
      DO 840 JJ=1,II
      IWP=INDROW(II)+JJ
      IW=INDROW(NT(II))+NT(JJ)+IA
      WP(IWP)=W(IW)
840   CONTINUE
C     LOAD SIGMA12 IN B IN FULL STORAGE MODE

      DO 850 II=1,L
      DO 850 JJ=L+1,N
      IF(NT(JJ).GT.NT(II))GOTO 849
      IW=INDROW(NT(II))+NT(JJ)+IA
      BB(II,JJ-L)=W(IW)
      BT(JJ-L,II)=BB(II,JJ-L)
      GOTO 850
849   IW=INDROW(NT(JJ))+NT(II)+IA
      BB(II,JJ-L)=W(IW)
      BT(JJ-L,II)=BB(II,JJ-L)
850   CONTINUE

      M=N-L

      CALL LEQT1P(WP,M,L,BB,20,10,D1,D2,IER)

C     MULTIPLY BT*BB TO GET SIGMA22.1, THEN TAKE ITS TRACE

      TR=0.0
      DO 930 II=1,M
      DO 930 JJ=1,L
      TR=TR+BT(II,JJ)*BB(JJ,II)
930   CONTINUE

      DO 950 II=1,L
      IW=INDROW(NT(II))+NT(II)+IA
      TR=TR+W(IW)
950   CONTINUE

C     COMPUTE RES, THE PERCENT VARIANCE EXPLAINED
      RES=100.*TR/SUM
  115 WRITE(6,600) RES,USAVE(IUSE+J),(IPW(KK),KK=1,20)
  600 FORMAT(1X,F8.4,5X,E14.7,2X,20(X,A2))
  120 CONTINUE
      RETURN
```