Memoirs of the School of Engineering Okayama University Vol. 10, No. 1, July 1975

Computer Program of Forward Selection and Backward Elimination Procedure in Linear Discriminant Analysis and Test for Differences Between Mean Values of Two Populations

Hirokazu ŌSAKI* and Susumu KIKUCHI*

(Received June, 6, 1975)

Synopsis

In multivariate analysis, the linear discriminant analysis and the test for differences between mean values of two populations are of wide application. It is not essential to increase the variables only in order to increase the degree of accuracy of discrimination or test without evaluating the effect of variables.

Therefore the computer program of selection procedures of variables in these two methods is mentioned in this paper.

1 Introduction

We dealt with the selection procedures in multiple regression analysis in previous paper[1]. If these selection procedures are modified slightly, it will be possible to apply to the linear discriminant analysis and the test for diffrences between mean values of two populations. If the variables which are effective to discrimination are selected from many variables, the discriminant function of using the selected variables will become simple and useful. Further if the selected variables only are used to test for differences between mean values of two populations, the degree of the test criterion will become highest.

---

* Department of Industrial Science

And as these two methods have many common parts logically, the program of the selection procedures for these methods is mentioned in this paper.

2 Analytical method

Suppose Π_1 is one k-variate population according to $N(\mu_1, \Sigma)$ and Π_2 is the other one according to $N(\mu_2, \Sigma)$.

Assume $(x_{1m}, x_{2m}, \ldots, x_{km})'$, $m=1,2,\ldots,N_1$ are the samples from population Π_1 and $(y_{1n}, y_{2n}, \ldots, y_{kn})'$, $n=1,2,\ldots,N_2$ are ones from Π_2.

The sample mean vectors and the sample variance covariance matrices are shown as follows.

Mean vectors

$$\bar{x} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k)' \quad , \quad \bar{x}_i = \sum_{m=1}^{N_1} x_{im} / N_1$$

$$\bar{y} = (\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_k) \quad , \quad \bar{y}_i = \sum_{n=1}^{N_2} y_{in} / N_2$$

Variance covariance matrices

$$S_1 = (\bar{s}_{ij}) \quad , \quad \bar{s}_{ij} = \sum_{m=1}^{N_1} (x_{im} - \bar{x}_i)(x_{jm} - \bar{x}_j)/(N_1 - 1)$$

$$S_2 = (\bar{\bar{s}}_{ij}) \quad , \quad \bar{\bar{s}}_{ij} = \sum_{n=1}^{N_2} (y_{in} - \bar{y}_i)(y_{jn} - \bar{y}_j)/(N_2 - 1)$$

$$i, j = 1,2,\ldots,k$$

Further d and S are calculated from $\bar{x}$, $\bar{y}$, S_1, and S_2.

Difference between mean vectors

$$d = (d_1, d_2, \ldots, d_k)' = \bar{x} - \bar{y} \quad , \quad d_i = \bar{x}_i - \bar{y}_i$$

Pooled variance covariance matrix

$$S = (s_{ij}) \quad , \quad s_{ij} = \{\bar{s}_{ij}(N_1 - 1) + \bar{\bar{s}}_{ij}(N_2 - 1)\}/(N_1 + N_2 - 2)$$

$$i, j = 1,2,\ldots,k$$

2.1 Linear discriminant analysis

Suppose the discriminant function between Π_1 and Π_2 is as follows.

$$Z = a_1z_1+a_2z_2+\cdots+a_kz_k+b_k \quad .....(1)$$

Then the coefficients are given by the following equations[2].

$$a = (a_1,a_2,\ldots,a_k) = d'S^{-1}$$

$Z_{\bar{x}} = \sum_{i=1}^{k} a_i\bar{x}_i$: mean value of Π_1 in discriminant space

$Z_{\bar{y}} = \sum_{i=1}^{k} a_i\bar{y}_i$: mean value of Π_2 in discriminant space

$$b_k = -(Z_{\bar{x}} + Z_{\bar{y}})/2$$

In discriminant space, area of $Z > 0$ shows Π_1 and that of $Z < 0$ does Π_2.

Therefore the probability of miss classification by the discriminant function is calculated by the following equation[3].

$$P(\Pi_1/\Pi_2) = P(\Pi_2/\Pi_1) = \int_c^\infty \exp(-t^2/2)dt \qquad ....(2)$$
$$c = \sqrt{d'S^{-1}d} \,/\, 2$$

$D_k^2 = d'S^{-1}d$ is the sample version of the Mahalanobis Distance

$$D^2 = (\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2).$$

2.2 Test for differences between mean values of two populations

The hypothesis of test is as follows.

$$H_0 : \mu_1 = \mu_2 \quad , \quad H_1 : \mu_1 \neq \mu_2$$

This hypothesis is tested by the following criterion.

$$F_0 = N_1N_2(N_1+N_2-k-1)d'S^{-1}d/ (N_1+N_2)(N_1+N_2-2)k \qquad ....(3)$$

And if $F_0 > F(\alpha,k,N_1+N_2-k-1)$, then H_0 rejects at level $100\alpha\%$. If $F_0 < F(\alpha,k,N_1+N_2-k-1)$, then H_1 accepts.

In equation (2) and (3), the effect of all variables is contained in $d'S^{-1}d$. It is important to determine the subset of variables which are effective to discrimination or test for differences.

Authors were mentioned the Fortran IV program of forward selection and backward elimination procedure in multiple regression analysis. These procedure can be extended to the selection of variables in linear discriminant analysis and test for differences between mean values of two populations[4,5].

3 Procedures

Let us A is the set of used all variables.

$$A = \{x_1, x_2, \ldots, x_k\}$$

And B(m) is a subset consisting of m variables of A.

$$B(m) = \{x_{i_1}, x_{i_2}, \ldots, x_{i_m}\}$$

$D(i_1, i_2, \ldots, i_m) = D(B(m))$ is shown as the value of $d'S^{-1}d$ calculated from variables of B(m).

3.1 Forward selection procedure (FSP)

At the m-th. step , x_{I_m} is put as the selected variable which satisfies the following relations.

$$\max_{i_m} D(B(m-1,k), i_m) = D(B(m-1,k), I_m)$$

$$x_{i_m} \varepsilon \ A - B(m-1,k)$$

where $B(m-1,k) = \{x_{I_1}, x_{I_2}, \ldots, x_{I_{m-1}}\}$

and $B(0,k) = \phi$.

And in this step, $B(m,k) = \{x_{I_1}, x_{I_2}, \ldots, x_{I_m}\}$ is the subset of selected variables until m-th. step. Therefore the discriminant function and the test criterion are as follows.

$$Z = \sum_{j=1}^{m} a_{I_j} x_{I_j} + b_m$$

$$(a_{I_1}, a_{I_2}, \ldots, a_{I_m}) = d_m' S_m^{-1}.$$

where elememts of d_m and S_m are subsets of d and S corresponding to B(m,k).

$$F_{0m} = N_1N_2(N_1+N_2-m-1)D(B(m,k))/m(N_1+N_2)(N_1+N_2-2)$$

$$m = 1,2,\ldots,k.$$

3.2 Backward elimination procedure(BEP)

At the m-th. step, the eliminated variable is denoted by $x_{J_{k-m+1}}$ and satisfies the following relations.

$$\max_{i_m} D(A - C(m-1,k) - i_m) = D(A - C(m-1,k) - J_{k-m+1})$$

$$x_{i_m} \varepsilon A - C(m-1,k)$$

$$\text{where } C(m-1,k) = \{x_{J_k}, x_{J_{k-1}}, \ldots, x_{J_{k-m+2}}\}$$

$$\text{and} \quad C(0,k) = \phi$$

And in this step, $C(m,k) = \{x_{J_k}, x_{J_{k-1}}, \ldots, x_{J_{k-m+1}}\}$ is the subset of eliminated variables until m-th. step. Therefore the discriminant function and test criterion are as follows.

Let us $\{x_{j_1}, x_{j_2}, \ldots, x_{j_{k-m}}\} = A - C(m,k)$.

$$Z = \sum_{l=1}^{k-m} a_{j_l} x_{j_l} + b_{k-m}$$

$$(a_{j_1}, a_{j_2}, \ldots, a_{j_{k-m}}) = d'_{k-m} S^{-1}_{k-m}$$

where elememts of d_{k-m} and S_{k-m} are subsets of d and S corresponding to $\{x_{j_1}, x_{j_2}, \ldots, x_{j_{k-m}}\}$.

$$F_{0m} = N_1N_2(N_1+N_2-k+m-1)D(A - C(m,k))/(k-m)(N_1+N_2)(N_1+N_2-2)$$

$$m = 1,2,\ldots,k.$$

4 Program

These procedures are programmed in Fortran IV and is the form of subroutine[6].

The subroutine name is FSBEDT.

```
      SUBROUTINE FSBEDT(AAA,AA1,MA,BBB,BB1,MB,KKK,NKK,STORE)
```

4.1 Argument list

ARGUMENT	I/O	TYPE	SIZE	DEFINITION	
AAA	INPUT	REAL	50 x 50	unbiased variance covariance matrix	Population Π_1
AA1	INPUT	REAL	50	mean vector	Population Π_1
MA	INPUT	INTEGER	1	number of data	Population Π_1
BBB	INPUT	REAL	50 x 50	unbiased variance covariance matrix	Population Π_2
BB1	INPUT	REAL	50	mean vector	Population Π_2
MB	INPUT	INTEGER	1	number of data	Population Π_2
KKK	INPUT	INTEGER	50	designated variable number	
NKK	INPUT	INTEGER	1	number of variables	
STORE	OUTPUT	REAL	50 x 4	results of two procedures	

4.2 Suggestion on using

4.2.1 NKK $\leqq$ 50

4.2.2 If for some i, AAA(i,i) = 0 and BBB(i,i) = 0, then the computation stops as i-th. column is used in the calculation.

4.2.3 Correspondence between arguments and given data
One group is put as Π_1 and the other Π_2.

$$\text{AAA}(i,j) = \bar{s}_{ij} \qquad \text{BBB}(i,j) = \bar{\bar{s}}_{ij}$$

$$\text{AA1}(i) = \bar{x}_i \qquad \text{BB1}(i) = \bar{y}_i$$

$$\text{MA} = N_1 \qquad \text{MB} = N_2$$

KKK(i) = K_i , K_i: designated variable number (usually $K_i = i$)

NKK = k , k' : number of variables

STORE(i,l) = st_{il} , $st_{i1} = I_i$

$st_{i2} = D(I_1, I_2, \ldots, I_i)$

$st_{i3} = J_i$

$st_{i4} = D(J_1, J_2, \ldots, J_i)$

$i,j = 1,2,\ldots k$, $l = 1,2,3,4$

4.2.4 Subroutine PRINTA and SIMEQS are used in FSBEDT. PRINTA is used to print out of the results. And SIMEQS is used to solve the linear equation. Programs of these subroutines are listed in FSBEDT.

Program list is shown in Table 1.

5 Example

The data in a four-variables(k=4) problem given by R.A. Fisher [7] are used to check the program. These data were used by M.G. Kendall[8] and C.R. Rao[2], too. Given data are shown in Table 2. And the results are shown in Table 3.

Π_1 : *Iris setosa* , Π_2 : *Iris versicolor*

x_1 : SL(sepal length) , x_2 : SW(sepal width)

x_3 : PL(petal length) , x_4 : PW(petal width)

References

[1] H.Ōsaki and S.Kikuchi:Memoirs School Eng., Okayama Univ., 9-2(1974),89.

[2] C.R.Rao:"Linear Statistical Inference and Its Application", John Wiley & Sons,Inc.,(1965),480.

[3] K.Takeuchi:"Suri Tokeigaku", Toyo Keizai Shinpo,(1963),352.

[4] T.Tarumi and A.Kudo:Journ. Japan Statist. Soc.,4-2(1974), 47.

[5] A.Kudo: Memoirs of the Faculty Science, Kyushu University, Series A, 17(1963), 63.

[6] P.S.Dwyer:"Linear Computations", John Wiley & Sons,Inc., (1960).

[7] R.A.Fisher: Ann. Engenics, 7(1936), 179.

[8] M.G.Kendall:"Multivariate Analysis", Discriminant and Classification, Academic Press, (1966),169.

Table 1, Program Listing

```
      SUBROUTINE FSBEDT(AAA,AA1,MA,BBB,BB1,MB,NKK,KKK,STORE)

      SUBROUTINE OF FORWARD AND BACKWARD SELECTION OF VARIABLES
      IN LENEAR DISCRIMINANT ANALYSIS AND TEST FOR DIFFERENCES
      BETWEEN MEAN VALUES OF TWO POPULATIONS

      AAA(50,50)      VARIANCE COVARIANCE MATRIX OF GROUP 1
      AA1(50)         MEAN VECTOR                OF GROUP 1
      MA              NUMBER OF DATA             OF GROUP 1

      BBB(50,50)      VARIANCE COVARIANCE MATRIX OF GROUP 2
      BB1(50)         MEAN VECTOR                OF GROUP 2
      MB              NUMBER OF DATA             OF GROUP 2


      NKK             NUMBER OF DESIGNATED NUMBER OF VARIABLES
      KKK(50)         DESIGNATED NUMBER OF VARIABLES

      STORE(50,4)     RESULTS OF SELECTION PROCEDURES


      DIMENSION AAA(50,50),AA1(50),BBB(50,50),BB1(50),KKK(50),
     1          KKKK(50),CCC(50,50),DIFF(50),STAND(50),KKMOD(50,4),
     2          WORK(50,5),SSTAND(50),STORE(50,4),NSTORE(50)
      MDF=MA+MB-2
      DO 1 I=1,NKK
      DO 2 J=1,NKK
      CCC(I,J)=(AAA(I,J)*FLOAT(MA-1)+BBB(I,J)*FLOAT(MB-1))/FLOAT(MDF)
    2 CONTINUE
      DIFF(I)=AA1(I)-BB1(I)
      SSTAND(I)=SQRT(CCC(I,I))
    1 CONTINUE
      DO 20 I=1,NKK
      DO 21 J=1,NKK
   21 CCC(I,J)=CCC(I,J)/(SSTAND(I)*SSTAND(J))
      CCC(I,I)=1.0
      DIFF(I)=DIFF(I)/SSTAND(I)
      WORK(I,4)=AA1(I)
      WORK(I,5)=BB1(I)
   20 CONTINUE
      WRITE(6,3000)
 3000 FORMAT(1H1,/////1H ,20X,'*** ORDERING OF VARIABLES IN LENEAR ',
     1       'DISCRIMINANT ***',/1H ,20X,'*** ANALYSIS AND TEST FOR ',
     2       'DIFFERENCES BETWEEN     ***',/1H ,20X,'*** MEAN VALUES ',
     3       'OF TWO POPULATIONS                ***',////)
      WRITE(6,3010) NKK
 3010 FORMAT(1H ,30X,'** GIVEN DATA',
     1       ///1H ,20X,'NUMBER OF VARIABLES          = ',I5,
     2       ///1H ,20X,'DESIGNATED VARIABLES NUMBER',
     3       /1H ,25X,'I',10X,'KKK(I)')
      DO 10 I=1,NKK
   10 WRITE(6,3011) I,KKK(I)
 3011 FORMAT(1H ,21X,I5,9X,I5)
```

```
      WRITE(6,3013) MA,MB
3013  FORMAT(///1H ,20X,'NUMBER OF DATA IN GROUP 1  = ',I5,
     1          /1H ,20X,'NUMBER OF DATA IN GROUP 2  = ',I5)
      WRITE(6,3015)
3015  FORMAT(///1H ,20X,'MEAN VECTORS',3X,'GROUP 1',13X,'GROUP 2',
     1       /1H ,25X,'I',9X,'AA1(I)',14X,'BB1(I)')
      DO 15 I=1,NKK
   15 WRITE(6,3017) I,AA1(I),BB1(I)
3017  FORMAT(1H ,21X,I5,4X,E15.7,5X,E15.7)
      WRITE(6,3020)
3020  FORMAT(///1H ,20X,'VARIANCE COVARIANCE MATRICES',
     1       /1H ,35X,'GROUP 1',13X,'GROUP 2',
     2       /1H ,22X,'(I , J)',5X,'AAA(I,J)',12X,'BBB(I,J)')
      DO 17 I=1,NKK
      DO 17 J=I,NKK
   17 WRITE(6,3025) I,J,AAA(I,J),BBB(I,J)
3025  FORMAT(1H ,20X,2I4,1X,E15.7,5X,E15.7)
      WRITE(6,3030)
3030  FORMAT(///1H ,30X,'** CALCULATED DATA',
     1       //1H ,20X,'NORMALIZED DIFFERENCES',
     2       /1H ,25X,'VARIABLE NUMBER',5X,'NORMALIZED DIFFERENCES')
      DO 25 I=1,NKK
   25 WRITE(6,3050) KKK(I),DIFF(I)
3050  FORMAT(1H ,30X,I5,15X,E15.7)
      WRITE(6,3060)
3060  FORMAT(///1H ,20X,'POOLED VARIANCE COVARIANCE MATRIX',
     1       /1H ,25X,'BETWEEN VARIABLES',3X,'VARIANCE COVARIANCE')
      DO 26 I=1,NKK
      DO 26 J=I,NKK
      PV=(AAA(I,J)*FLOAT(MA-1)+BBB(I,J)*FLOAT(MB-1))/FLOAT(MDF)
   26 WRITE(6,3065) KKK(I),KKK(J),PV
3065  FORMAT(1H ,25X,'(',I5,' , ',I5,' )',6X,E15.7)

      FORWARD SELECTION PROCEDURE

      WRITE(6,3070)
3070  FORMAT(1H1,/////1H ,30X,'** FORWARD SELECTION PROCEDURE **',////)
      DO 100 I=1,NKK
  100 KKMOD(I,1)=I
      J=1
      AMAX=0.0
      DO 110 I=1,NKK
      ASQR=DIFF(I)**2
      IF(ASQR-AMAX) 110,110,120
  120 MHAN=I
      AMAX=ASQR
  110 CONTINUE
      KKMOD(J,2)=MHAN
      KKMOD(MHAN,1)=0
      WORK(1,1)=AMAX
      STAND(1)=DIFF(MHAN)/SSTAND(MHAN)
      KKKK(1)=MHAN
      CALL PRINTA (1,MHAN,MA,MB,AMAX,KKK,KKKK,STAND,AA1,BB1,1,NKK)
      IF(NKK .EQ. 1) GO TO 8888
      DO 130 J=2,NKK
      AMAX=0.0
      DO 140 KK=1,NKK
      IF(KKMOD(KK,1).EQ.0) GO TO 140
      KR=KKMOD(KK,1)
      J1=J-1
```

```
      DO 150 L=1,J1
      KP=KKMOD(L,2)
      DO 151 M=1,J1
      KQ=KKMOD(M,2)
      AAA(L,M)=CCC(KP,KQ)
  151 CONTINUE
      AAA(L,J)=CCC(KP,KR)
      AAA(J,L)=AAA(L,J)
      AA1(L)=DIFF(KP)
      BB1(L)=AA1(L)
  150 CONTINUE
      AAA(J,J)=CCC(KR,KR)
      AA1(J)=DIFF(KR)
      BB1(J)=AA1(J)
      CALL SIMEQS(AAA,AA1,J,NCHEC1)
      IF(NCHEC1 .NE. 1) GO TO 162
  161 WRITE(6,3081)
 3081 FORMAT(1H1,/////,20X,'DIAGNAL ELEMENT OF MATRIX IS ZERO')
      RETURN
  162 CONTINUE
      ASQR=0.0
      DO 160 I=1,J
  160 ASQR=ASQR+AA1(I)*BB1(I)
      IF(ASQR-AMAX) 180,180,170
  170 AMAX=ASQR
      MHAN=KK
      DO 175 I=1,J
  175 STAND(I)=AA1(I)
  180 CONTINUE
  140 CONTINUE
      KKMOD(J,2)=MHAN
      KKMOD(MHAN,1)=0
      WORK(J,1)=AMAX
      DO 9510 LL=1,NKK
      AA1(LL)=WORK(LL,4)
      BB1(LL)=WORK(LL,5)
 9510 CONTINUE
      DO 9520 LL=1,J
      KKKK(LL)=KKMOD(LL,2)
      LN=KKKK(LL)
      STAND(LL)=STAND(LL)/SSTAND(LN)
 9520 CONTINUE
      CALL PRINTA(J,MHAN,MA,MB,AMAX,KKK,KKKK,STAND,AA1,BB1,1,NKK)
  130 CONTINUE
      DO 131 I=1,NKK
      STORE(I,1)=KKMOD(I,2)
  131 STORE(I,2)=WORK(I,1)
      STORE(NKK,4)=STORE(NKK,2)
C
C     BACKWARD ELIMINATION PROCEDURE
C
      WRITE(6,3080)
 3080 FORMAT(1H1,/////1H ,30X,'** BACKWARD ELIMINATION PROCEDURE **',
     1      ////)
      DO 200 I=1,NKK
  200 KKMOD(I,1)=I
      NKK1=NKK-1
      DO 210 MMM=1,NKK1
      NCOU=0
```

```
      DO 220 I=1,NKK
      NP=KKMOD(I,1)
      IF(NP .EQ. 0) GO TO 220
      NCOU=NCOU+1
      KKMOD(NCOU,3)=NP
  220 CONTINUE
      AMAX=0.0
      DO 230 I=1,NCOU
      KEF=KKMOD(I,3)
      KKMOD(I,3)=0
      KNNN=0
      DO 240 J=1,NCOU
      KEF=KKMOD(J,3)
      IF(KEF .EQ. 0) GO TO 240
      KNNN=KNNN+1
      KKMOD(KNNN,4)=KEF
  240 CONTINUE
      DO 250 L=1,KNNN
      KP=KKMOD(L,4)
      DO 251 M=1,KNNN
      KQ=KKMOD(M,4)
      AAA(L,M)=CCC(KP,KQ)
  251 CONTINUE
      AA1(L)=DIFF(KP)
      BB1(L)=AA1(L)
  250 CONTINUE
      CALL SIMEQS(AAA,AA1,KNNN,NCHEC2)
      IF(NCHEC2 .EQ. 1) GO TO 161
      ASQR=0.0
      DO 260 L=1,KNNN
  260 ASQR=ASQR+AA1(L)*BB1(L)
      IF(ASQR-AMAX) 271,271,270
  270 AMAX=ASQR
      MHAN=KEE
      DO 275 L=1,KNNN
      STAND(L)=AA1(L)
      KKKK(L)=KKMOD(L,4)
  275 CONTINUE
  271 KKMOD(I,3)=KEE
  230 CONTINUE
      WORK(KNNN,1)=AMAX
      KKMOD(KNNN,2)=MHAN
      DO 9550 MMP=1,NKK
      AA1(MMP)=WORK(MMP,4)
      BB1(MMP)=WORK(MMP,5)
 9550 CONTINUE
      DO 9600 L=1,KNNN
      LN=KKKK(L)
      STAND(L)=STAND(L)/SSTAND(LN)
 9600 CONTINUE
      CALL PRINTA(MMM,MHAN,MA,MB,AMAX,KKK,KKKK,STAND,AA1,BB1,2,NKK)
      KKMOD(MHAN,1)=0
  210 CONTINUE
      DO 295 I=1,NKK1
      STORE(I+1,3)=KKMOD(I,2)
  295 STORE(I,4)=WORK(I,1)
      STORE(1,3)=KKKK(1)

      SELECTION PROCEDURES END
```

```
      WRITE(6,4000)
 4000 FORMAT(1H1,/////1H ,20X,'** LIST OF ORDERED VARIABLES **',
     1       /////1H ,20X,'FORWARD SELECTION PROCEDURE',
     2       //1H ,7X,'STEP',5X,'NUMBER',9X,'DD',17X,'F - TEST')
      DO 4001 I=1,NKK
      NFO=STORE(I,1)
      NFOO=KKK(NFO)
      FT=FLOAT(MA+MB-I-1)*FLOAT(MA*MB)/FLOAT(I*(MA+MB)*(MA+MB-2))
      FT=FT*STORE(I,2)
 4001 WRITE(6,4002) I,NFOO,STORE(I,2),FT
 4002 FORMAT(1H ,5X,I5,5X,I5,5X,E15.7,5X,E15.7)
      WRITE(6,4003)
 4003 FORMAT(////1H ,20X,'BACKWARD ELIMINATION PROCEDURE',
     1       //1H ,7X,'STEP',5X,'NUMBER',9X,'DD',17X,'F - TEST')
      DO 4004 I=1,NKK
      NBA=STORE(I,3)
      NBAA=KKK(NBA)
      FT=FLOAT(MA+MB-I-1)*FLOAT(MA*MB)/FLOAT(I*(MA+MB)*(MA+MB-2))
      FT=FT*STORE(I,4)
 4004 WRITE(6,4002) I,NBAA,STORE(I,4),FT
      WRITE(6,4005)
 4005 FORMAT(/////1H ,20X,'STEP NUMBER WHICH ORDERS DO NOT COINCIDE',/)
      NROT=0
      DO 4010 I=1,NKK
      NUM=0
      DO 4011 J=1,I
      DO 4012 K=1,I
      IF(STORE(J,1) .EQ. STORE(K,3)) GO TO 4013
 4012 CONTINUE
      GO TO 4011
 4013 NUM=NUM+1
 4011 CONTINUE
      IF(NUM .EQ. I) GO TO 4010
      NROT=NROT+1
      NSTORE(NROT)=I
 4010 CONTINUE
      IF(NROT .EQ. 0) GO TO 8888
      DO 4020 I=1,NROT
 4020 WRITE(6,4021) NSTORE(I)
 4021 FORMAT(1H ,20X,I5)
 8888 RETURN
      END
```

```
      SUBROUTINE PRINTA (JP,MPHAN,MAP,MBP,PMAX,KKKP,KPMOD,PSTAND,
     1                   PA1,PB1,MPCOU,MPVAR)
      DIMENSION KKKP(50),KPMOD(50),PSTAND(50),PA1(50),PB1(50)
      WRITE(6,3110) JP
      NFF=KKKP(MPHAN)
      IF(MPCOU.EQ.1) GO TO 9900
      WRITE(6,3121) NFF
      WRITE(6,3130)
      MPDF=MAP+MBP
      JJP=MPVAR-JP
      GO TO 9990
```

```
9900 WRITE(6,3120) NFF
     WRITE(6,3130)
     JJP=JP
     MPDF=MAP+MBP+1
9990 DO 9000 I=1,JJP
     MFF=KPMOD(I)
     MMFF=KKKP(MFF)
     PPPP=PSTAND(I)
     WRITE(6,3140) MMFF,PPPP
9000 CONTINUE
     PMEAN1=0.0
     PMEAN2=0.0
     DO 9010 I=1,JJP
     MFF=KPMOD(I)
     PMEAN1=PMEAN1+PSTAND(I)*PA1(MFF)
     PMEAN2=PMEAN2+PSTAND(I)*PB1(MFF)
9010 CONTINUE
     DISC=(PMEAN1+PMEAN2)/2.0
     DISC=-DISC
     WRITE(6,3150) DISC
     WRITE(6,3155) PMEAN1,PMEAN2
     STAN=SQRT(PMAX)
     ERR=ABS(PMEAN1-DISC)/STAN
     WRITE(6,3170) ERR
     WRITE(6,3160) PMAX,STAN
     FMAB=MAP*MBP*(MAP+MBP-1-JJP)
     FMABC=(MAP+MBP)*(MAP+MBP-2)*JJP
     FFFFF=FMAB*PMAX/FMABC
     MDFAB=MAP+MBP-1-JJP
     WRITE(6,3180) FFFFF
     WRITE(6,3190) JJP,MDFAB
3110 FORMAT(1H ,10X,'* STEP   ( ',I3,' )',/)
3120 FORMAT(1H ,20X,'ENTERING VARIABLE NUMBER . . . X(',I5,' )',/)
3121 FORMAT(1H ,20X,'EXCLUDING VARIABLE NUMBER . . . X(',I5,' )',/)
3130 FORMAT(1H ,20X,'DISCRIMINANT COEFFICIENT',/)
3140 FORMAT(1H ,24X,'B ( ',I5,' )  =  ',E15.7)
3150 FORMAT(1H ,22X,'CONSTANT',10X,E15.7,//)
3155 FORMAT(1H ,20X,'MEAN OF GROUP 1 IN DISCRIMINANT SPACE = ',E15.7,
    1       /1H ,20X,'MEAN OF GROUP 2 IN DISCRIMINANT SPACE = ',E15.7,/)
3160 FORMAT(1H0,20X,'MAHALANOBIS DISTANCE FROM DATA (DD)    = ',E15.7,
    1       /1H ,20X,'SQUARE ROOT OF DD                     = ',E15.7)
3170 FORMAT(1H ,20X,'PROBABILITY OF MISS CLASSIFICATION=',
    1       /1H ,20X,'=NORMAL DISTRIBUTION(Z GRATER THAN ',E15.7,' )')
3180 FORMAT(//1H ,20X,'TEST FOR DIFFERENCES (F - TEST)         = ',E15.7)
3190 FORMAT(1H ,20X,'DEGREE OF FREEDOM  ( ',I4,' , ',I7,' )',//)
     RETURN
     END
```

```
      SUBROUTINE SIMEQS(ABA,DD,NDIM,NCHECK)
      DIMENSION ABA(50,50),DD(50)
      NCHECK=0
      DO 10 K=1,NDIM
      P=ABA(K,K)
      IF(P .EQ. 0.0) GO TO 100
      K1=K+1
      IF(K1 .GT. NDIM) GO TO 21
      DO 20 J=K1,NDIM
   20 ABA(K,J)=ABA(K,J)/P
   21 DD(K)=DD(K)/P
      DO 30 I=1,NDIM
      IF(I .EQ. K) GO TO 30
      P=ABA(I,K)
      DO 40 J=K1,NDIM
   40 ABA(I,J)=ABA(I,J)-ABA(K,J)*P
      DD(I)=DD(I)-DD(K)*P
   30 CONTINUE
   10 CONTINUE
      RETURN
  100 NCHECK=1
      RETURN
      END
```

Table 2, Given Data

Variance covariance matrices

$\bar{s}_{11} = 0.1242490$	$\bar{\bar{s}}_{11} = 0.2664327$
$\bar{s}_{12} = \bar{s}_{21} = 0.0992163$	$\bar{\bar{s}}_{12} = \bar{\bar{s}}_{21} = 0.0851837$
$\bar{s}_{13} = \bar{s}_{31} = 0.0163551$	$\bar{\bar{s}}_{13} = \bar{\bar{s}}_{31} = 0.1828980$
$\bar{s}_{14} = \bar{s}_{41} = 0.0103306$	$\bar{\bar{s}}_{14} = \bar{\bar{s}}_{41} = 0.0557796$
$\bar{s}_{22} = 0.1436898$	$\bar{\bar{s}}_{22} = 0.0984694$
$\bar{s}_{23} = \bar{s}_{32} = 0.0116980$	$\bar{\bar{s}}_{23} = \bar{\bar{s}}_{32} = 0.0826531$
$\bar{s}_{24} = \bar{s}_{42} = 0.0092980$	$\bar{\bar{s}}_{24} = \bar{\bar{s}}_{42} = 0.0412041$
$\bar{s}_{33} = 0.0301592$	$\bar{\bar{s}}_{33} = 0.2208163$
$\bar{s}_{34} = \bar{s}_{43} = 0.0060694$	$\bar{\bar{s}}_{34} = \bar{\bar{s}}_{43} = 0.0731020$
$\bar{s}_{44} = 0.0111061$	$\bar{\bar{s}}_{44} = 0.0391061$

Mean vectors

$\bar{x}_1 = 5.006$	$\bar{y}_1 = 5.936$
$\bar{x}_2 = 3.428$	$\bar{y}_2 = 2.770$
$\bar{x}_3 = 1.462$	$\bar{y}_3 = 4.260$
$\bar{x}_4 = 0.246$	$\bar{y}_4 = 1.326$

Number of data

$N_1 = 50$ $N_2 = 50$

Designated variable number

$K_1 = 1$, $K_2 = 2$ $K_3 = 3$, $K_4 = 4$

Number of variables

$k = 4$

Table 3, Computer Output

```
*** ORDERING OF VARIABLES IN LENEAR DISCRIMINANT ***
*** ANALYSIS AND TEST FOR DIFFERENCES BETWEEN     ***
*** MEAN VALUES OF TWO POPULATIONS                ***


          ** GIVEN DATA

NUMBER OF VARIABLES         =     4

DESIGNATED VARIABLES NUMBER
     I           KKK(I)
     1             1
     2             2
     3             3
     4             4


NUMBER OF DATA IN GROUP 1   =     50
NUMBER OF DATA IN GROUP 2   =     50


MEAN VECTORS    GROUP 1              GROUP 2
     I          AA1(I)               BB1(I)
     1        .5006000E+01         .5936000E+01
     2        .3428000E+01         .2770000E+01
     3        .1462000E+01         .4260000E+01
     4        .2460000E+00         .1326000E+01


VARIANCE COVARIANCE MATRICES
                GROUP 1              GROUP 2
  (I , J)       AAA(I,J)             BBB(I,J)
   1   1      .1242490E+00         .2664327E+00
   1   2      .9921633E-01         .8518368E-01
   1   3      .1635510E-01         .1828980E+00
   1   4      .1033061E-01         .5577960E-01
   2   2      .1436898E+00         .9846939E-01
   2   3      .1169796E-01         .8265307E-01
   2   4      .9297960E-02         .4120408E-01
   3   3      .3015918E-01         .2208163E+00
   3   4      .6069388E-02         .7310204E-01
   4   4      .1110612E-01         .3910612E-01
```

```
        ** CALCULATED DATA

NORMALIZED DIFFERENCES
     VARIABLE NUMBER       NORMALIZED DIFFERENCES
            1                       -.2104197E+01
            2                        .1890995E+01
            3                       -.7898544E+01
            4                       -.6816068E+01
```

(N.B. normalized differnce = $d_i/\sqrt{s_{ii}}$)

```
POOLED VARIANCE COVARIANCE MATRIX
     BETWEEN VARIABLES    VARIANCE COVARIANCE
     (     1 ,      1 )         .1953408E+00
     (     1 ,      2 )         .9220001E-01
     (     1 ,      3 )         .9962654E-01
     (     1 ,      4 )         .3305510E-01
     (     2 ,      2 )         .1210796E+00
     (     2 ,      3 )         .4717551E-01
     (     2 ,      4 )         .2525102E-01
     (     3 ,      3 )         .1254878E+00
     (     3 ,      4 )         .3958572E-01
     (     4 ,      4 )         .2510612E-01


              ** FORWARD SELECTION PROCEDURE **


* STEP   (     1 )

        ENTERING VARIABLE NUMBER . . . X(     3 )

        DISCRIMINANT COEFFICIENT

            B (       3 )  =     -.2229700E+02
          CONSTANT                .6379170E+02

        MEAN OF GROUP 1 IN DISCRIMINANT SPACE =   -.3259821E+02
        MEAN OF GROUP 2 IN DISCRIMINANT SPACE =   -.9498520E+02

        PROBABILITY OF MISS CLASSIFICATION=
        =NORMAL DISTRIBUTION(Z GRATER THAN     .1220350E+02 )

        MAHALANOBIS DISTANCE FROM DATA (DD)   =      .6238699E+02
        SQUARE ROOT OF DD                     =      .7898544E+01

        TEST FOR DIFFERENCES (F - TEST)       =      .1559675E+04
        DEGREE OF FREEDOM   (     1 ,      98 )
```

```
* STEP   (    2 )

        ENTERING VARIABLE NUMBER . . . X(     2 )

        DISCRIMINANT COEFFICIENT

            B  (       3 )   =     -.2851701E+02
            B  (       2 )   =      .1654535E+02
          CONSTANT                  .3031312E+02

        MEAN OF GROUP 1 IN DISCRIMINANT SPACE =     .1502560E+02
        MEAN OF GROUP 2 IN DISCRIMINANT SPACE =    -.7565183E+02

        PROBABILITY OF MISS CLASSIFICATION=
        =NORMAL DISTRIBUTION(Z GRATER THAN     .1605415E+01 )

        MAHALANOBIS DISTANCE FROM DATA (DD)   =     .9067743E+02
        SQUARE ROOT OF DD                     =     .9522470E+01

        TEST FOR DIFFERENCES (F - TEST)       =     .1121902E+04
        DEGREE OF FREEDOM  (     2 ,       97 )

* STEP   (    3 )

        ENTERING VARIABLE NUMBER . . . X(     4 )

        DISCRIMINANT COEFFICIENT

            B  (       3 )   =     -.1966744E+02
            B  (       2 )   =      .1974243E+02
            B  (       4 )   =     -.3186341E+02
          CONSTANT                  .2013141E+02

        MEAN OF GROUP 1 IN DISCRIMINANT SPACE =     .3108485E+02
        MEAN OF GROUP 2 IN DISCRIMINANT SPACE =    -.7134766E+02

        PROBABILITY OF MISS CLASSIFICATION=
        =NORMAL DISTRIBUTION(Z GRATER THAN     .1082260E+01 )

        MAHALANOBIS DISTANCE FROM DATA (DD)   =     .1024325E+03
        SQUARE ROOT OF DD                     =     .1012089E+02

        TEST FOR DIFFERENCES (F - TEST)       =     .8361837E+03
        DEGREE OF FREEDOM  (     3 ,       96 )
```

(N.B. in 3 step, $I_3 = 4$,

$$Z = 19.742x_2 - 19.667x_3 - 31.863x_4 + 20.1314$$

$c = 1.0823$

$F_{03} = 836.184$, d.f.(3,96))

```
* STEP   (    4 )

         ENTERING VARIABLE NUMBER . . . X(     1 )

         DISCRIMINANT COEFFICIENT

             B (       3 )   =      -.2176619E+02
             B (       2 )   =       .1802296E+02
             B (       4 )   =      -.3084417E+02
             B (       1 )   =       .3052770E+01
           CONSTANT                  .1396174E+02

         MEAN OF GROUP 1 IN DISCRIMINANT SPACE =     .3765503E+02
         MEAN OF GROUP 2 IN DISCRIMINANT SPACE =    -.6557851E+02

         PROBABILITY OF MISS CLASSIFICATION=
         =NORMAL DISTRIBUTION(Z GRATER THAN     .2331927E+01 )

         MAHALANOBIS DISTANCE FROM DATA (DD)    =     .1032335E+03
         SQUARE ROOT OF DD                      =     .1016039E+02

         TEST FOR DIFFERENCES (F - TEST)        =     .6254583E+03
         DEGREE OF FREEDOM   (     4 ,       95 )


                      ** BACKWARD ELIMINATION PROCEDURE **


* STEP   (    1 )

         EXCLUDING VARIABLE NUMBER . . . X(     1 )

         DISCRIMINANT COEFFICIENT

             B (       2 )   =       .1974243E+02
             B (       3 )   =      -.1966744E+02
             B (       4 )   =      -.3186341E+02
           CONSTANT                  .2013141E+02

         MEAN OF GROUP 1 IN DISCRIMINANT SPACE =     .3108485E+02
         MEAN OF GROUP 2 IN DISCRIMINANT SPACE =    -.7134766E+02

         PROBABILITY OF MISS CLASSIFICATION=
         =NORMAL DISTRIBUTION(Z GRATER THAN     .1082260E+01 )

         MAHALANOBIS DISTANCE FROM DATA (DD)    =     .1024325E+03
         SQUARE ROOT OF DD                      =     .1012089E+02

         TEST FOR DIFFERENCES (F - TEST)        =     .8361837E+03
         DEGREE OF FREEDOM   (     3 ,       96 )
```

```
* STEP  (    2 )

        EXCLUDING VARIABLE NUMBER . . . X(     4 )

        DISCRIMINANT COEFFICIENT

           B (      2 )  =      .1654535E+02
           B (      3 )  =     -.2851701E+02
          CONSTANT              .3031312E+02

        MEAN OF GROUP 1 IN DISCRIMINANT SPACE =     .1502560E+02
        MEAN OF GROUP 2 IN DISCRIMINANT SPACE =    -.7565183E+02

        PROBABILITY OF MISS CLASSIFICATION=
        =NORMAL DISTRIBUTION(Z GRATER THAN     .1605415E+01 )

        MAHALANOBIS DISTANCE FROM DATA (DD)    =     .9067743E+02
        SQUARE ROOT OF DD                      =     .9522470E+01

        TEST FOR DIFFERENCES (F - TEST)        =     .1121902E+04
        DEGREE OF FREEDOM  (    2 ,      97 )

* STEP  (    3 )

        EXCLUDING VARIABLE NUMBER . . . X(     2 )

        DISCRIMINANT COEFFICIENT

           B (      3 )  =     -.2229700E+02
          CONSTANT              .6379170E+02

        MEAN OF GROUP 1 IN DISCRIMINANT SPACE =    -.3259821E+02
        MEAN OF GROUP 2 IN DISCRIMINANT SPACE =    -.9498520E+02

        PROBABILITY OF MISS CLASSIFICATION=
        =NORMAL DISTRIBUTION(Z GRATER THAN     .1220350E+02 )

        MAHALANOBIS DISTANCE FROM DATA (DD)    =     .6238699E+02
        SQUARE ROOT OF DD                      =     .7898544E+01

        TEST FOR DIFFERENCES (F - TEST)        =     .1559675E+04
        DEGREE OF FREEDOM  (    1 ,      98 )
```

(N.B. in 3 step, $J_2 = 2$ and $J_1 = 3$.

$$Z = -22.287x_3 + 63.792$$

$$c = 12.2035$$

$F_{03} = 1559.675$, d.f.(1,98))

** LIST OF ORDERED VARIABLES **

FORWARD SELECTION PROCEDURE

STEP	NUMBER	DD	F - TEST
1	3	.6238699E+02	.1559675E+04
2	2	.9067743E+02	.1121902E+04
3	4	.1024325E+03	.8361837E+03
4	1	.1032335E+03	.6254583E+03

BACKWARD ELIMINATION PROCEDURE

STEP	NUMBER	DD	F - TEST
1	3	.6238699E+02	.1559675E+04
2	2	.9067743E+02	.1121902E+04
3	4	.1024325E+03	.8361837E+03
4	1	.1032335E+03	.6254583E+03

(N.B. eliminated variables are shown by)
( )
(the order of $J_1, J_2, \ldots, J_k$. DD=D$(J_1, J_2, \ldots, J_m)$)

STEP NUMBER WHICH ORDERS DO NOT COINCIDE

(N.B. if $(I_1, I_2, \ldots, I_m) \neq (J_1, J_2, \ldots, J_m)$,)
( )
(step number m is printed out.)