MEMOIRS OF THE SCHOOL OF ENGINEERING OKAYAMA UNIVERSITY Vol. 9, No. 2, DECEMBER 1974

Computer Program of Forward Selection and Backward Elimination Procedure in Multiple Regression Analysis

Hirokazu ŌSAKI* and Susumu KIKUCHI*

(Received December, 10, 1974)

Synopsis

Multiple regression analysis are often used to explain the relation between the dependent variable and the independent variables. In case of that it arises necessity that the important independent variables which are closely correlated with the dependent variable are selected from among all given ones. There are some selection procedures. But these procedures can't be used usefully without using computer.

Therefore two selection procedures that is Forward selection procedure and Backward elimination procedure in multiple regression analysis are programmed by Fortran IV .

1. Introduction

It is important in multiple regression analysis to select or order the independent variables which are closely correlated with the dependent variable in order to explain the phenomenon clearly. There are some procedures of selecting the important independent variables (1,2,3). But these procedures can be applied to the various problems only by using of computer.

In this paper we mention on the Fortran IV program of Forward

* Department of Industrial Science

selection procedure and Backward elimination procedure in multiple regression analysis.

2. Analytical method

Set of independent variables is put as follows.

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

k: number of independent variables

Y is the dependent variable.

$(y_i, x_{1i}, x_{2i}, \ldots, x_{ki})$, $i=1,2,\ldots,n$ are the data of Y and C.

The mean square due to residual variation(S_y^2) is one scale of precision of estimate of the regression line in using independent variables.

$$S_y^2 = \sum_{m=1}^{n} (y_m - \bar{y})^2 - d'S^{-1}d \ /(n - k - 1) \ \ldots\ldots \ (1)$$

where $d' = (d_1, d_2, \ldots, d_k)'$: k x 1 vector

$$d_i = \sum_{m=1}^{n} (y_m - \bar{y})(x_{im} - \bar{x}_i)$$

$S = (s_{ij})$: k x k matrix

$$s_{ij} = \sum_{m=1}^{n} (x_{im} - \bar{x}_i)(x_{jm} - \bar{x}_j)$$

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

The effect of independent variables to the dependent variable is contained in $d'S^{-1}d$. But it is necessary to evaluate the effect of each independent variable or that of subset of independent variables in order to explain clearly the relation between independent variables and dependent one. There are some criterions of evaluating the effect. Therefore the procedures which select or order the independent variables have been proposed by many authors(1,2,3).

In this paper we deal with the procedures that the independent variable is selected or eliminated one by one from among all given independent variables in accordance with the following mentioned criterions. These procedures are called Forward selection procedure and Backward elimination procedure(3).

$D(i_1, i_2, \ldots, i_m)$ is put as the value of $d'S^{-1}d$ which is calculated from independent variables $\{x_{i_1}, x_{i_2}, \ldots, x_{i_m}\}$.

$E(m, i_p)$ is put as the value of $d'S^{-1}d$ which is calculated from independent variables $\{x_{i_1}, x_{i_2}, \ldots, x_{i_m}\} - \{x_{i_p}\}$.

2.1 Forward selection procedure

x_{I_m} is put as the independent variable which satisfies the following relation.

$$\max_{i_m} D(I_1, I_2, \ldots, I_{m-1}, i_m) \qquad \ldots\ldots\ (2)$$
$$\{x_{i_m}\} \subset C - \{x_{I_1}, x_{I_2}, \ldots, x_{I_{m-1}}\}$$

Then the locally best regression line and the mean square due to residual variation are as follows.

$$\hat{Y} = \sum_{j=1}^{m} a_{I_j}(x_{I_j} - \bar{x}_{I_j}) + \bar{y}$$

$$S_{ym}^2 = \left\{ \sum_{i=1}^{n} (y_i - \bar{y})^2 - D(I_1, I_2, \ldots, I_m) \right\} / (n - m - 1)$$

Where $m = 1, 2, 3, \ldots, k$.

2.2 Backward elimination procedure

$x_{J_{k-m+1}}$ is put as the independent variable which satisfies the following relation.

$$\max_{i_p} E(k-m+1, i_p) \quad , \quad \{x_{i_p}\} \subset C - \{x_{J_k}, x_{J_{k-1}}, \ldots, x_{J_{k-m+2}}\} \qquad \ldots\ldots\ (3)$$

Then the locally best regression line and the mean square due to residual variation are as follows.

$$\hat{Y} = \sum_{j=1}^{k-m} b_{i_j}(x_{i_j} - \bar{x}_{i_j}) + \bar{y}$$

$$S_{yk-m}^2 = \left\{ \sum_{i=1}^{n} (y_i - \bar{y})^2 - E(k-m+1, J_{k-m+1}) \right\} / (n-k+m-1)$$

$$\{x_{i_1}, x_{i_2}, \ldots, x_{i_{k-m}}\} = C - \{x_{J_k}, x_{J_{k-1}}, \ldots, x_{J_{k-m+1}}\}$$

Where m = 1,2,...,k-1. And for m=1, $\{x_{J_k}, x_{J_{k-1}}, \ldots, x_{J_{k-m+2}}\} = \phi$, and for m=k-1, $x_{J_1} = C - \{x_{J_k}, x_{J_{k-1}}, \ldots, x_{J_2}\}$.

These two procedures have strong and weak points. But if two procedures are used at the same time, the each strong points can partially make up for each weak point. Therefore these two procedures can be programmed in one program.

3. Program

This program is written by Fortran Ⅳ and is the form of subroutine (4). The subroutine name is FORBAC.

SUBROUTINE FORBAC(AAA,AA1,MA,KKK,NKK,STORE)

3.1 Argument List

ARGUMENT	I / O	TYPE	SIZE	DEFINITION
AAA	INPUT	REAL	50 x 50	unbiased variance covariance matrix
AA1	INPUT	REAL	50	mean vector
MA	INPUT	INTEGER	1	number of data
KKK	INPUT	INTEGER	50	dependent and independent variables number
NKK	INPUT	INTEGER	1	number of independent variables + 1
STORE	OUTPUT	REAL	50 x 4	results of two procedures

3.2 Suggestions on using

3.2.1 NKK $\leq$ 50

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

3.2.3. Correspondence between arguments and given data

$$AAA(i,j) = aa_{ij} \quad , \quad aa_{11} = \sum_{m=1}^{n} (y_m - y)^2 / (n - 1)$$

$$aa_{1i+1} = aa_{i+11} = d_i / (n - 1) \qquad ,i,j=1,2,\ldots,k$$
$$aa_{i+1j+1} = aa_{j+1i+1} = s_{ij} / (n - 1)$$

$AA1(i) = c_i$, $c_1 = \overline{y}$, $c_{i+1} = \overline{x}_i$, $i=1,2,\ldots,k$

MA = n , n : number of data

$KKK(i) = k_i$, k_1 = dependent variable number

k_{i+1} = independent variable number

NKK = k + 1 , k : number of independent variables

$STORE(i,j) = st_{ij}$, $st_{i1} = I_i$

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

$st_{i3} = J_i$

$st_{i4} = E(i+1,J_{i+1})$

3.2.4 Subroutine PRINTA and SIMEQS are used in FORBAC. PRINTA is used to print out the results. And SIMEQS is used to solve the linear equation. These subroutine are listed in FORBAC.

Program listing is shown in Table 1.

4. Example

The data in a four variable(k = 4) problem given by A. Hald(5) are used to check the program. This data were used by N.R.Draper and H.Smith(3) too. Given data are shown in Table 2. And results are shown in Table 3.

References

(1) R.L.Anderson and T.A.Bancroft : "Statistical theory in research", McGraw-Hill Book Co.,(1952)

(2) P.S.Dwyer "Linear computations", John Wiley & Sons,Inc., (1960)

(3) N.R.Draper and H.Smith : "Applied regression analysis", John Wiley & Sons,Inc.,(1966),365

(4) B.Carnahan, H.A.Luther and J.O.Wilkes : "Applied numerical method", John Wiley & Sons,Inc.,(1969),17

(5) A.Hald : "Statistical theory with engineering applications", John Wiley & Sons,Inc., (1952)., 647

Table 1,Program Listing

```
      SUBROUTINE FORBAC(AAA,AA1,MA,KKK,NKK,STORE)
C
C     SUBROUTINE OF FORWARD AND BACKWARD SELECTION OF VARIABLES
C     IN MULTIPLE REGRESSION
C     AAA(50,50)      VARIANCE COVARIANCE MATRIX
C     AA1(50)         MEAN VECTOR
C     MA              NUMBER OF DATA
C     KKK(50)         DESIGNATED NUMBER OF VARIABLES
C     NKK             NUMBER OF DESIGNATED NUMBER OF VARIABLES
C     STORE(50,4)     RESULTS OF SELECTION PROCEDURES
C
C
      DIMENSION AAA(50,50),AA1(50),KKK(50),NSTORE(50),STORE(50,4),
     1          KKMOD(50,4),KKKK(50),CCC(50,50),DIFF(50),
     2            STAND(50),SSTAND(50),WORK(50,2),BB1(50)
      WRITE(6,3000)
 3000 FORMAT(1H1,/////,20X,'**** ORDERING OF VARIABLES IN MULTIPLE ',
     1       'REGRESSION ****',/////)
      WRITE(6,3005) KKK(1)
 3005 FORMAT(1H ,20X,'DEPENDENT VARIABLE NUMBER   =  ',I5,/)
      NKK=NKK-1
      YMEAN=AA1(1)
      YVAR=AAA(1,1)
      DO 22 I=1,NKK
      AA1(I)=AA1(I+1)
   22 KKK(I)=KKK(I+1)
      WRITE(6,3010)
 3010 FORMAT(1H ,20X,'INDEPENDENT VARIABLE NUMBER',/)
      WRITE(6,3020) (KKK(I),I=1,NKK)
 3020 FORMAT(1H ,40X,10I7)
      DO 1 I=1,NKK
      DO 2 J=1,NKK
      CCC(I,J)=AAA(I+1,J+1)*FLOAT(MA-1)
    2 CONTINUE
      DIFF(I)=AAA(1,I+1)*FLOAT(MA-1)
      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,2)=AA1(I)
   20 CONTINUE
      WRITE(6,3040)
 3040 FORMAT(1H ,///,20X,'NORMALIZED COEFFICIENTS',//,
     1       20X,'INDEPENDENT VARIABLE NUMBER    NORMALIZED COEFFICIENT')
      DO 25 I=1,NKK
   25 WRITE(6,3050) KKK(I),DIFF(I)
 3050 FORMAT(1H ,30X,I6,20X,E15.7)
```

```
      WRITE(6,3060)
 3060 FORMAT(1H ,///,20X,'CORRELATION MATRIX',//,20X,
     1      'BETWEEN INDEPENDENT VARIABLES    CORRELATION COEFFICIENT')
      DO 26 I=1,NKK
      DO 26 J=I,NKK
   26 WRITE (6,3065) KKK(I),KKK(J),CCC(I,J)
 3065 FORMAT(1H ,25X,'( ',I5,' , ',I5,' )',15X,E15.7)
C
C     FORWARD SELECTION PROCEDURE
C
      WRITE(6,3070)
 3070 FORMAT(1H1,/////,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
      STORE(1,1)=KKMOD(1,2)
      STORE(1,2)=AMAX
      STAND(1)=DIFF(MHAN)/SSTAND(MHAN)
      KKKK(1)=MHAN
      CALL PRINTA(1,MHAN,MA,AMAX,KKK,KKKK,STAND,AA1,1,NKK,YVAR,YMEAN)
      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
      STORE(J,1)=KKMOD(J,2)
      STORE(J,2)=WORK(J,1)
      DO 9510 LL=1,NKK
 9510 AA1(LL)=WORK(LL,2)
      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,AMAX,KKK,KKKK,STAND,AA1,1,NKK,YVAR,YMEAN)
  130 CONTINUE
      STORE(NKK,4)=STORE(NKK,2)
C     ----------------------------------
C     BACKWARD ELIMINATION PROCEDURE
C     ----------------------------------
      WRITE(6,3080)
 3080 FORMAT(1H1,/////,30X,'** BACKWARD ELIMINATION PROCEDURE **',/////)
      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
      KEE=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
      STORE(KNNN+1,3)=KKMOD(KNNN,2)
      STORE(KNNN,4)=WORK(KNNN,1)
      IF(KNNN .NE. 1) GO TO 9601
      STORE(1,3)=KKKK(1)
 9601 CONTINUE
      DO 9550 MMP=1,NKK
 9550 AA1(MMP)=WORK(MMP,2)
      DO 9600 L=1,KNNN
      LN=KKKK(L)
      STAND(L)=STAND(L)/SSTAND(LN)
 9600 CONTINUE
      CALL PRINTA(MMM,MHAN,MA,AMAX,KKK,KKKK,STAND,AA1,2,NKK,YVAR,YMEAN)
      KKMOD(MHAN,1)=0
  210 CONTINUE
C
C     SELECTION PROCEDURES END
C
      WRITE(6,4000)
 4000 FORMAT(1H1,/////,20X,'*** LIST OF ORDERED VARIABLES ***',
     1       ////1H ,20X,'FORWARD SELECTION PROCEDURE',
     2       //1H ,7X,'STEP',5X,'NUMBER',9X,'SS',18X,'SR')
      DO 4001 I=1,NKK
      NFO=STORE(I,1)
      NFOO=KKK(NFO)
      FOO=(YVAR*FLOAT(MA-1)-STORE(I,2))/(FLOAT(MA-I)-1.0)
 4001 WRITE(6,4002) I,NFOO,STORE(I,2),FOO
 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,'SS',18X,'SR')
      DO 4004 I=1,NKK
      NBA=STORE(I,3)
      NBAA=KKK(NBA)
      BAA=(YVAR*FLOAT(MA-1)-STORE(I,4))/(FLOAT(MA-I)-1.0)
 4004 WRITE(6,4002) I,NBAA,STORE(I,4),BAA
      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(6,4030) NSTORE(I)
4030 FORMAT(1H ,20X,I5)
8888 RETURN
     END


     SUBROUTINE PRINTA(JP,MPHAN,MAP,PMAX,KKKP,KPMOD,PSTAND,PA1,MPCOU,
    1                  MPVAR,YPVAR,YPMEAN)
     DIMENSION KKKP(50),KPMOD(50),PSTAND(50),PA1(50)
     WRITE(6,3110) JP
     NFF=KKKP(MPHAN)
     IF(MPCOU.EQ.1) GO TO 9900
     WRITE(6,3121) NFF
     WRITE(6,3130)
     MPDF=MAP
     JJP=MPVAR-JP
     GO TO 9990
9900 WRITE(6,3120) NFF
     WRITE(6,3130)
     JJP=JP
     MPDF=MAP+1
9990 CONTINUE
     DO 9000 I=1,JJP
     MFF=KPMOD(I)
     MMFF=KKKP(MFF)
     PPPP=PSTAND(I)
     WRITE(6,3140) MMFF,PPPP
9000 CONTINUE
     PMEAN1=0.0
     DO 9010 I=1,JJP
     MFF=KPMOD(I)
     PMEAN1=PMEAN1+PSTAND(I)*PA1(MFF)
9010 CONTINUE
     PMEAN1=YPMEAN-PMEAN1
     WRITE(6,3150) PMEAN1
     WRITE(6,3160) YPMEAN,YPVAR,MAP
     WRITE(6,3190) PMAX
     FFFF=(YPVAR*(FLOAT(MAP)-1.0)-PMAX)/(FLOAT(MAP-JJP)-1.0)
     WRITE(6,3200) FFFF
     STANDD=SQRT(FFFF)
     WRITE(6,3220) STANDD
     MMDF=MAP-JJP-1
     CORRE=SQRT(PMAX/(YPVAR*FLOAT(MAP-1)))
     WRITE(6,3230) CORRE
     WRITE(6,3240) JJP,MMDF
3110 FORMAT(1H ,10X,'* STEP  ( ',I3,' )',/)
3120 FORMAT(1H ,20X,'ENTERING INDEPENDENT VARIABLE NUMBER . . . X( ',
    1       I5,' )',/)
3121 FORMAT(1H ,20X,'EXCLUDING INDEPENDENT VARIABLE NUMBER . . . X( ',
    1       I5,' )',/)
3130 FORMAT(1H ,20X,'REGRESSION COEFFICIENTS',/)
```

```
3140 FORMAT(1H ,24X,▼B ( ▼,I5,▼ )  =  ▼,E15.7)
3150 FORMAT(1H ,22X,▼CONSTANT▼,10X,E15.7,//)
3160 FORMAT(1H ,20X,▼MEAN VALUE OF DEPENDENT VARIABLE           = ▼,E15.7,
   1       /1H ,20X,▼UNBIASED VARIANCE OF DEPENDENT VARIABLE = ▼,E15.7,
   2       /1H ,20X,▼NUMBER OF DATA                           = ▼,I10,/)
3190 FORMAT(1H ,20X,▼SUM OF SQUARES DUE TO REGRESSION (SS)    = ▼,E15.7)
3200 FORMAT(1H ,20X,▼MEAN SQUARE DUE TO RESIDUAL VARIATION(SR)=▼,E15.7)
3220 FORMAT(1H ,20X,▼SQUARE ROOT OF SR                        = ▼,E15.7)
3230 FORMAT(1H ,20X,▼MULTIPLE CORRELATION COEFFICIENT (R)     = ▼,E15.7)
3240 FORMAT(1H0,20X,▼DEGREE OF FREEDOM  ( ▼,I4,▼ , ▼,I7,▼ )▼,//)
     RETURN
     END
```

```
     SUBROUTINE SIMEQS(BBB,DD,NDIM,NCHECK)
     DIMENSION BBB(50,50),DD(50)
     NCHECK=0
     DO 10 K=1,NDIM
     P=BBB(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 BBB(K,J)=BBB(K,J)/P
  21 DD(K)=DD(K)/P
     DO 30 I=1,NDIM
     IF(I .EQ. K) GO TO 30
     P=BBB(I,K)
     IF(K1 .GT. NDIM) GO TO 41
     DO 40 J=K1,NDIM
  40 BBB(I,J)=BBB(I,J)-BBB(K,J)*P
  41 DD(I)=DD(I)-DD(K)*P
  30 CONTINUE
  10 CONTINUE
     RETURN
 100 NCHECK=1
     RETURN
     END
```

Table 2 , Data

Variance covariance matrix (AAA)

$aa_{11} = 226.314$

$aa_{12} = aa_{21} = 64.664$

$aa_{13} = aa_{31} = 191.079$

$aa_{14} = aa_{41} = -51.519$

$aa_{15} = aa_{51} = -206.808$

$aa_{22} = 34.603$

$aa_{23} = aa_{32} = 20.923$

$aa_{24} = aa_{42} = -31.051$

$aa_{25} = aa_{52} = -24.167$

$aa_{33} = 242.141$

$aa_{34} = aa_{43} = -13.878$

$aa_{35} = aa_{53} = -253.417$

$aa_{44} = 41.026$

$aa_{45} = aa_{54} = 3.167$

$aa_{55} = 280.167$

Mean value (AA1)

$c_1 = 95.423$

$c_2 = 7.462$

$c_3 = 48.154$

$c_4 = 11.769$

$c_5 = 30.000$

Dependent and independent variables number (KKK)

$k_1 = 5$

$k_2 = 1$

$k_3 = 2$

$k_4 = 3$

$k_5 = 4$

n = 13 (MA)

k = 4 (NKK : k + 1)

Table 3 , Computer Output

```
**** ORDERING OF VARIABLES IN MULTIPLE REGRESSION ****


DEPENDENT VARIABLE NUMBER  =        5

INDEPENDENT VARIABLE NUMBER

                              1      2      3      4


NORMALIZED COEFFICIENTS

INDEPENDENT VARIABLE NUMBER     NORMALIZED COEFFICIENT
             1                       .3807987E+02
             2                       .4253736E+02
             3                      -.2786328E+02
             4                      -.4280066E+02


CORRELATION MATRIX

BETWEEN INDEPENDENT VARIABLES     CORRELATION COEFFICIENT
        (    1 ,    1 )                 .1000000E+01
        (    1 ,    2 )                 .2285795E+00
        (    1 ,    3 )                -.8241338E+00
        (    1 ,    4 )                -.2454451E+00
        (    2 ,    2 )                 .1000000E+01
        (    2 ,    3 )                -.1392424E+00
        (    2 ,    4 )                -.9729550E+00
        (    3 ,    3 )                 .1000000E+01
        (    3 ,    4 )                 .2953700E-01
        (    4 ,    4 )                 .1000000E+01
```

```
** FORWARD SELECTION PROCEDURE **

* STEP  (  1 )

        ENTERING INDEPENDENT VARIABLE NUMBER . . . X(   4 )

        REGRESSION COEFFICIENTS

          B (     4 )  =   -.7381618E+00
          CONSTANT           .1175679E+03

        MEAN VALUE OF DEPENDENT VARIABLE              =   .9542308E+02
        UNBIASED VARIANCE OF DEPENDENT VARIABLE       =   .2263136E+03
        NUMBER OF DATA                                =        13

        SUM OF SQUARES DUE TO REGRESSION (SS)         =   .1831896E+04
        MEAN SQUARE DUE TO RESIDUAL VARIATION(SR)     =   .8035154E+02
        SQUARE ROOT OF SR                             =   .8963902E+01
        MULTIPLE CORRELATION COEFFICIENT (R)          =   .8213050E+00

        DEGREE OF FREEDOM  (    1 ,     11 )

* STEP  (  2 )

        ENTERING INDEPENDENT VARIABLE NUMBER . . . X(   1 )

        REGRESSION COEFFICIENTS

          B (     4 )  =   -.6139536E+00
          B (     1 )  =    .1439958E+01
          CONSTANT          .1030974E+03

        MEAN VALUE OF DEPENDENT VARIABLE              =   .9542308E+02
        UNBIASED VARIANCE OF DEPENDENT VARIABLE       =   .2263136E+03
        NUMBER OF DATA                                =        13

        SUM OF SQUARES DUE TO REGRESSION (SS)         =   .2641001E+04
        MEAN SQUARE DUE TO RESIDUAL VARIATION(SR)     =   .7476213E+01
        SQUARE ROOT OF SR                             =   .2734267E+01
        MULTIPLE CORRELATION COEFFICIENT (R)          =   .9861395E+00

        DEGREE OF FREEDOM  (    2 ,     10 )
```

(N.B. in step 2, $I_2 = 1$

$$\hat{Y} = -0.6140x_4 + 1.4400x_1 + 103.0974$$

$$s^2_{y2} = 7.4762 \quad , \quad D(4,1) = 2641.001$$
)

```
* STEP  (    3 )

        ENTERING INDEPENDENT VARIABLE NUMBER . . . X(      2 )

        REGRESSION COEFFICIENTS

            B  (       4 )   =      -.2365402E+00
            B  (       1 )   =       .1451938E+01
            B  (       2 )   =       .4161098E+00
          CONSTANT                   .7164830E+02

        MEAN VALUE OF DEPENDENT VARIABLE              =     .9542308E+02
        UNBIASED VARIANCE OF DEPENDENT VARIABLE =     .2263136E+03
        NUMBER OF DATA                                =        13

        SUM OF SQUARES DUE TO REGRESSION (SS)         =     .2667790E+04
        MEAN SQUARE DUE TO RESIDUAL VARIATION(SR)=     .5330305E+01
        SQUARE ROOT OF SR                             =     .2308745E+01
        MULTIPLE CORRELATION COEFFICIENT (R)          =     .9911284E+00

        DEGREE OF FREEDOM    (     3 ,        9 )

* STEP  (    4 )

        ENTERING INDEPENDENT VARIABLE NUMBER . . . X(      3 )

        REGRESSION COEFFICIENTS

            B  (       4 )   =      -.1440606E+00
            B  (       1 )   =       .1551103E+01
            B  (       2 )   =       .5101681E+00
            B  (       3 )   =       .1019099E+00
          CONSTANT                   .6240532E+02

        MEAN VALUE OF DEPENDENT VARIABLE              =     .9542308E+02
        UNBIASED VARIANCE OF DEPENDENT VARIABLE =     .2263136E+03
        NUMBER OF DATA                                =        13

        SUM OF SQUARES DUE TO REGRESSION (SS)         =     .2667899E+04
        MEAN SQUARE DUE TO RESIDUAL VARIATION(SR)=     .5982957E+01
        SQUARE ROOT OF SR                             =     .2446008E+01
        MULTIPLE CORRELATION COEFFICIENT (R)          =     .9911486E+00

        DEGREE OF FREEDOM    (     4 ,        8 )
```

```
** BACKWARD ELIMINATION PROCEDURE **

* STEP  (   1 )

         EXCLUDING INDEPENDENT VARIABLE NUMBER . . . X(      3 )

         REGRESSION COEFFICIENTS

            B (      1 )  =     .1451938E+01
            B (      2 )  =     .4161098E+00
            B (      4 )  =    -.2365402E+00
           CONSTANT             .7164830E+02

         MEAN VALUE OF DEPENDENT VARIABLE               =     .9542308E+02
         UNBIASED VARIANCE OF DEPENDENT VARIABLE =     .2263136E+03
         NUMBER OF DATA                                  =           13

         SUM OF SQUARES DUE TO REGRESSION (SS)    =     .2667790E+04
         MEAN SQUARE DUE TO RESIDUAL VARIATION(SR)=     .5330305E+01
         SQUARE ROOT OF SR                               =     .2308745E+01
         MULTIPLE CORRELATION COEFFICIENT (R)      =     .9911284E+00

         DEGREE OF FREEDOM  (    3 ,        9 )

* STEP  (   2 )

         EXCLUDING INDEPENDENT VARIABLE NUMBER . . . X(      4 )

         REGRESSION COEFFICIENTS

            B (      1 )  =     .1468306E+01
            B (      2 )  =     .6622505E+00
           CONSTANT             .5257735E+02

         MEAN VALUE OF DEPENDENT VARIABLE               =     .9542308E+02
         UNBIASED VARIANCE OF DEPENDENT VARIABLE =     .2263136E+03
         NUMBER OF DATA                                  =           13

         SUM OF SQUARES DUE TO REGRESSION (SS)    =     .2657859E+04
         MEAN SQUARE DUE TO RESIDUAL VARIATION(SR)=     .5790450E+01
         SQUARE ROOT OF SR                               =     .2406335E+01
         MULTIPLE CORRELATION COEFFICIENT (R)      =     .9892817E+00

         DEGREE OF FREEDOM  (    2 ,       10 )
```

(**N.B. in step 2,** $J_3 = 4$

$$\hat{Y} = 1.4683x_1 + 0.6623x_2 + 52.5774$$

$$s_{y2}^2 = 5.7905\ ,\quad E(3,4) = 2657.859$$)

* STEP (3)

EXCLUDING INDEPENDENT VARIABLE NUMBER . . . X(1)

REGRESSION COEFFICIENTS

B (2)	=	.7891248E+00
CONSTANT		.5742368E+02

MEAN VALUE OF DEPENDENT VARIABLE	=	.9542308E+02
UNBIASED VARIANCE OF DEPENDENT VARIABLE	=	.2263136E+03
NUMBER OF DATA	=	13
SUM OF SQUARES DUE TO REGRESSION (SS)	=	.1809427E+04
MEAN SQUARE DUE TO RESIDUAL VARIATION(SR)	=	.8239421E+02
SQUARE ROOT OF SR	=	.9077126E+01
MULTIPLE CORRELATION COEFFICIENT (R)	=	.8162526E+00

DEGREE OF FREEDOM (1 , 11)

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

*** LIST OF ORDERED VARIABLES ***

FORWARD SELECTION PROCEDURE

STEP	NUMBER	SS	SR
1	4	.1831896E+04	.8035154E+02
2	1	.2641001E+04	.7476213E+01
3	2	.2667790E+04	.5330305E+01
4	3	.2667899E+04	.5982957E+01

BACKWARD ELIMINATION PROCEDURE

STEP	NUMBER	SS	SR
1	2	.1809427E+04	.8239421E+02
2	1	.2657859E+04	.5790450E+01
3	4	.2667790E+04	.5330305E+01
4	3	.2667899E+04	.5982957E+01

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

STEP NUMBER WHICH ORDERS DO NOT COINCIDE

(N.B. if $(I_1, I_2, \ldots, I_m) \neq (J_1, J_2, \ldots, J_m)$,

1 **(step number m is printed out)**

2