An Electronic Computer Program for Multiple Regression Analysis

VELPURI SAKUNTALA*

Indian Statistical Institute, Calcutta

Multiple regression is an important statistical tool used in psychology for predicting performance on a criterion. Now that electronic computers are available in India, it may be worthwhile to make a computer program for multiple regression analysis available to psychologists so that they can avail of these computers for predictive and validation research. The computer program described in this paper is designed for an IBM 1401 Electronic Data Processing System with 8,000 positions of core storage and high-low-equal compare, multiply and divide advanced programming features. The program is written in the Fortran Language which is an automatic-coding system designed for scientific computations. It can readily be adapted for use with other types of electronic computers.

The purpose of the present program is to obtain on the computer, for a given set of data,

(1) partial regression coefficients of the criterion or dependent variable Y on each of the three predictor or independent variables X1, X2 and X3,

(2) the standard deviations of the coefficients in (1),

(3) the regression and residual sums of squares,

(4) the multiple correlation coefficient of Y on X1, X2 and X3, and

(5) the test statistics for testing the significance of each of the regression coefficients and of the regression as a whole. There is no restriction on the number N of experimental units on which the variables Y, X1, X2 and X3 are observed.

The following is the listing of the actual program which has been tested at the Computer Science Unit of the Indian Statistical Institute, Calcutta and used (2). The symbols +, —, *, / and SQRT F respectively stand for addition, subtraction, multiplication, division and the positive square root.

Program

FORTRAN COMPILATION—MACHINE SIZE 8000, MODULUS 5, MANTISSA 8

SEQ	STMNT	Fortran Statement
1		SY = 0
2		SX1 = 0
3		SX2 = 0
4		SX3 = 0
5		SYY = 0
6		SX1X1 = 0
7		SX2X2 = 0
8		SX3X3 = 0

*The author wishes to thank Dr. Rhea S. Das for the help and encouragement received in preparing this paper.

Manuscript received on January 22, 1968.

```
SEQ   STMNT      Fortran Statement
 9                SYX1   =   0
10                SYX2   =   0
11                SYX3   =   0
12                SX1X2  =   0
13                SX2X3  =   0
14                SX3X1  =   0
15                Read 1,    0
16          1     FORMAT (I3)
17                DO 5 I =   1, N
18                Read 100, Y, X1, X2, X3
19        100     FORMAT (4X, F3.2, 47X, F2.0, F3.0, F2.0)
20                SY     =   SY + Y
21                SX1    =   SX1 + X1
22                SX2    =   SX2 + X2
23                SX3    =   SX3 + X3
24                SYY    =   SYY + Y * Y
25                SX1X1  =   SX1X1 + X1 * X1
26                SX2X2  =   SX2X2 + X2 * X2
27                SX3X3  =   SX3X3 + X3 * X3
28                SYX1   =   SYX1 + Y * X1
29                SYX2   =   SYX2 + Y * X2
30                SYX3   =   SYX3 + Y * X3
31                SX1X2  =   SX1X2 + X1 * X2
32                SX2X3  =   SX2X3 + X2 * X3
33          5     SX3X1  =   SX3X1 + X3 * X1
34                AN     =   N
35                CSYY   =   SYY — (SY * SY)/AN
36                CSX1X1 =   SX1X1 — (SX1 * SX1)/AN
37                CSX2X2 =   SX2X2 — (SX2 * SX2)/AN
38                CSX3X3 =   SX3X3 — (SX3 * SX3)/AN
39                CSYX1  =   SYX1 — (SY * SX1)/AN
40                CSYX2  =   SYX2 — (SY * SX2)/AN
41                CSYX3  =   SYX3 — (SY * SX3)/AN
42                CSX1X2 =   SX1X2 — (SX1 * SX2)/AN
43                CSX2X3 =   SX2X3 — (SX2 * SX3)/AN
44                CSX3X1 =   SX3X1 — (SX3 * SX1)/AN
45                SEY    =   SQRT F [(CSYY/(AN — 1.)]
46                SEX1   =   SQRT F [(CSX1X1/(AN — 1.)]
47                SEX2   =   SQRT F [(CSX2X2/(AN — 1.)]
48                SEX3   =   SQRT F [(CSX3X3/(AN — 1.)]
49                CRYX1  =   CSYX1/[(AN —1.) * (SEY * SEX1)]
```

```
SEQ     STMNT                Fortran Statement
 50             CRYX2  =  CSYX2/[(AN —1.) * (SEY * SEX2)]
 51             CRYX3  =  CSYX3/[(AN —1.) * (SEY * SEX3)]
 52             CRX1X2 =  CSX1X2/[(AN —1.) * (SEX1 * SEX2)]
 53             CRX2X3 =  CSX2X3/[(AN —1.) * (SEX2 * SEX3)]
 54             CRX3X1 =  CSX3X1/[(AN —1.) * (SEX3 * SEX1)]
 55             C11    =  CSX2X2 * CSX3X3 — CSX2X3 * CSX2X3
 56             C12    =  CSX3X1 * CSX2X3 — CSX1X2 * CSX3X3
 57             C13    =  CSX1X2 * CSX2X3 — CSX3X1 * CSX2X2
 58             C22    =  CSX1X1 * CSX3X3 — CSX3X1 * CSX3X1
 59             C23    =  CSX1X2 * CSX3X1 — CSX1X1 * CSX2X3
 60             C33    =  CSX1X1 * CSX2X2 — CSX1X2 * CSX1X2
 61             D      =  C11 * CSX1X1 + C12 * CSX1X2 + C13 * CSX3X1
 62             B1     =  (C11 * CSYX1 + C12 * CSYX2 + C13 * CSYX3)/D
 63             B2     =  (C12 * CSYX1 + C22 * CSYX2 + C23 * CSYX3)/D
 64             B3     =  (C13 * CSYX1 + C23 * CSYX2 + C33 * CSYX3)/D
 65             A      =  (SY — B1 * SX1 — B2 * SX2 — B3 * SX3)/AN
 66             B1P    =  (B1 * SEX1)/SEY
 67             B2P    =  (B2 * SEX2)/SEY
 68             B3P    =  (B3 * SEX3)/SEY
 69             REG    =  B1 * CSYX1 + B2 * CSYX2 + B3 * CSYX3
 70             RES    =  CSYY — REG
 71             R      =  SQRT F (REG/CSYY)
 72             SEB1   =  SQRT F (RES * C11/[(AN — 4) * D)]
 73             SEB2   =  SQRT F (RES * C22/[(AN — 4) * D)]
 74             SEB3   =  SQRT F (RES * C33/[(AN — 4) * D)]
 75             TB1    =  B1/SEB1
 76             TB2    =  B2/SEB2
 77             TB3    =  B3/SEB3
 78             F      =  REG * (AN — 4)/(RES * 3)
 79             PRINT37, CSYY, CSX1X1, CSX2X2, CSX3X3, CSYX1, CSYX2, CSYX3,
                CSX1X2, CSX2X3, CSX3X1,

                (1) CSX2X3, CSX3X1, SEY, SEX1, SEX2, SEX3, CRYX1, CRYX2,
                    CRYX3, CRX1X2, CRX2X3,

                (2) CRX3X1, B1, B2, B3, A, B1P, B2P, B3P, REG, RES, R, SEB1,
                    SEB2, SEB3, TB1, TB2,

                (3) TB3, F

 80      37     FORMAT (10X, 4F18.6//10X, 6F18.6//10X, 4F18.6//10X, 6F18.6//

                (1) 10X, 4F18.6//10X, 3F18.6//10X, 2F18.6//10X, 4F18.6//10X,
                    4F18.6//)
```

Input of the program :

The input of the program consists of :

(1) program deck ;
(2) control card specifying the size N and
(3) data cards.

The statement with serial number 16 in the program specifies the format of the control card which gives the size N. The statement No. 19 (referenced as 100) gives the format of the data cards.

These two cards are to be prepared depending upon the given data.

Output of the program :

The statement 79 enlists the output printed out at the end of compilation and the statement 80 (referenced as 37) specifies the format in which the output is printed. In the present program the output is printed in the floating point mode upto six decimal places. The printed output consists of the following items in the given order :

First line :

CSYY : corrected sum of squares of Y, the dependent variable,
CSX1X1 : corrected sum of squares of X1, the first independent variable,
CSX2X2 : corrected sum of squares of X2, the second independent variable,
CSX3X3 : corrected sum of squares of X3, the third independent variable.

Second line :

CSYX1 : corrected sum of the cross products of Y and X1.
CSYX2 : corrected sum of the cross products of Y and X2.
CSYX3 : corrected sum of the cross products of Y and X3.
CSX1X2 : corrected sum of the cross products of X1 and X2.
CSX2X3 : corrected sum of the cross products of X2 and X3.
CSX3X1 : corrected sum of the cross products of X3 and X1.

Third line :

SEY : standard deviation of Y over the N observations.
SEX1 : standard deviation of X1 over the N observations.
SEX2 : standard deviation of X2 over the N individuals.
SEX3 : standard deviation of X3 over the N observations.

Fourth line :

CRYX1 : the product moment correlation between Y and X1.
CRYX2 : the product moment correlation between Y and X2.
CRYX3 : the product moment correlation between Y and X3.
CRX1X2 : the product moment correlation between X1 and X2.
CRX2X3 : the product moment correlation between X2 and X3.
CRX3X1 : the product moment correlation between X3 and X1.

Fifth line :

B1 : the partial regression coefficient of Y on X1.
B2 : the partial regression coefficient of Y on X2.
B3 : the partial regression coefficient of Y on X3.
A : the constant in the multiple regression of Y on X1, X2, X3.

Sixth line :

B1P : the partial regression of Y on X1
B2P : the partial regression of Y on X2
B3P : the partial regression of Y on X3
} when Y, X1, X2 and X3 are standardised.

Seventh line :

REG : regression sum of squares.
RES : residual sum of squares.

Eighth line :

R : multiple correlation coefficient of Y on X1, X2, X3.
SEB1 : the standard error of B1.
SEB2 : the standard error of B2.
SEB3 : the standard error of B3.

Ninth line :

TB1 : the student's t value with (N — 4) degrees of freedom for testing the significance of B1,
TB2 : the student's t value with (N — 4) degrees of freedom for testing the significance of B2,
TB3 : the student's t value with (N — 4) degrees of freedom for testing the significance of B3,
and F : the F value with degrees of freedom 3 and N — 4, for testing the significance of the multiple regression equation : $Y = A + B1.X1 + B2.X2 + B3.X3$.

REFERENCES

1. Computer Science ( ) : *Programming the IBM* 1401 *EDPM using Fortran II*, Computer Science Unit, Research & Training School, Indian Statistical Institute, Calcutta.

2. Das, B. C. (1967) : "Age-related Frends of the Amylase, Glyco-protein, Lipaprotein and Serum Protein in Human Blood." *Symposium on Experimental Gerontology* Prague, Czechoslovakia (November 1967).

3. Williams, E. J. (1959) : *Regression Analysis*, John Wiley, New York.