



DECUS

PROGRAM LIBRARY

DECUS NO.	12-74
TITLE	*REGRES -- MULTIPLE LINEAR REGRESSION
AUTHOR	David C. Howell
COMPANY	University of Vermont Burlington, Vermont
DATE	March 15, 1972
SOURCE LANGUAGE	FOCAL-12 running under the DIAL monitor

*REGRES

Multiple Linear Regression

David C. Howell

University of Vermont

*REGRES is the name for a series of four sequentially linked FOCAL-12 programs which perform a multiple linear regression analysis. This series will handle up to nine predictor variables as written, although this limit can be extended when necessary. The series requires a PDP-12 A or B computer running FOCAL-12 under the DIAL monitor.

This description is divided into five sections: Operating Instructions, Technical Description, Alternative Uses, Program Listing, and Sample Run.

Operating Instructions

No programming knowledge is required of the user once the system is stored on LINCTAPE and FOCAL-12 has been loaded. The following sequence of instructions are intended for someone with minimal familiarity with computer operations.

1. Type L G, *CORDATA, Ø
2. When interrogated by the program, enter the total number of variables (V) (the number of predictors plus the criterion) and then the number of subjects (independent observations) (N).
3. Next enter the data with a space after each data point. In this program one subject's data are entered at a time. The criterion variable is entered first, followed by the predictors.

4. After all of the data have been entered, the program interrogates the user to ask if any changes in the data are desired. If no changes are needed, enter "NO". If some of the data are in error and must be changed, enter a "1". At this point the program will ask for the location of the change (the i^{th} data point) and the corrected entry. These are entered by the user, after which the program asks if any more changes are desired. As before, the user answers "NO" or "1".

5. When no more changes are desired, the program chains to *CORMAT.

6. From this point, the entire set of programs runs almost automatically. At the end of each subprogram, the user is interrogated as to whether he wishes to continue. Under most conditions the answer to this question is affirmative, and the user enters a "1". The only purpose for the interrogation is to allow the program to be halted, should this be desired before it is erased as the next program is entered.

Technical Description

Necessary Tape Files

Before *REGRES can be run for the first time, the user must establish the following files for storage and working areas. The suggested number of blocks per file is indicated along with the file names. Under special conditions the file size may have to be increased or decreased.

<u>File Name</u>	<u>Blocks</u>
CORDATA	15
CORFIL	3
PRED	5
PROD	2

For those who are not familiar with data files, the files are created by entering the instruction L M, b, file name, Ø --where b = the number of blocks. For example: L M, 15, CORDATA, Ø will create CORDATA. These library commands are only entered before the program is run for the first time.

FOCAL-12 Programs

This multiple regression series consists of four independent programs chained together by library instructions. The user has the option of stopping before proceeding to the next program, although normally this would not be done. If he does stop, he can then load and start the next program from the keyboard. In the description below, FOCAL program names are preceded by an asterisk (*), while file names do not have the asterisk.

*CORDATA--This program stores the raw data in the file CORDATA for use in *CORMAT which follows. The program asks for the number of variables (V), the number of subjects (or independent observations) (N) and then the data, which are immediately stored on the data file. It should be pointed out for the benefit of those who may wish to examine the file directly, that the number of variables and the sample size are stored as F1(Ø) and F1(1), respectively. This means that the i^{th} data point is actually stored in location (i + 1), although for purposes of changing data through *CORDATA, it is treated as if it were in location i.

When all of the data have been entered and the necessary corrections made, the program automatically loads and starts *CORMAT.

*CORMAT--This program (a modification of one written by Dr. Lawrence Gordon of the University of Vermont) generates the intercorrelation matrix and stores the information needed by subsequent programs. This

program opens the file CORDATA, reads the data, generates the variance-covariance matrix on the working file CORFIL, and converts this into a matrix of intercorrelations. The program also generates the means and standard deviations of the variables and loads these, the inter-correlations of the predictors, and the correlations of the predictors with the criterion into PRED. Execution time of this program is a function of both N and V. As V, the number of variables, exceeds 9, the working file containing the variance-covariance matrix will be larger than the size of the buffer. Therefore a substantial increase in execution time will result, as the tape transports are continually moving from one block to another.

For the user who wishes to enter the matrix of intercorrelations directly, rather than the raw data, the final contents of PRED are given below, where k = the number of predictors = $(V - 1)$.

<u>Location</u>	<u>Contents</u>
\emptyset	k
1 to k^2	intercorrelations of predictors
(k^2+1) to (k^2+k)	correlations of predictors with criterion
(k^2+1+k) to (k^2+k+V)	standard deviations of all variables
$(k^2+1+k+V)$ to (k^2+k+2V)	means of all variables

With the above information, the user could open PRED, store the relevant information, and go directly to *INVERSE, without having to enter the original data in *CORDATA.

When the next program (*INVERSE) is executed, it will replace the contents of the first k^2 elements of PRED with the inverse of the matrix.

Thus if the user wishes to save the correlations he must stop the program when interrogated, store the matrix in a new file, and then load *INVERSE himself.

*INVERSE--This program calculates the inverse of the matrix of inter-correlations among the predictors (R). It uses the Gauss-Jourdan algorithm, and much of the program has been adapted from a FORTRAN program given by Donald Veldman in Fortran Programming for the Behavior Sciences, 1967. The program prints the inverse matrix (R^{-1}), stores it in PRED, and then interrogates the user as to whether he wishes to proceed to the next program. If so, *BETA is loaded and started.

It should be pointed out that since the PDP-12 only carries 6 significant digits, problems may result if R is very nearly singular. In this case the elements of R^{-1} are quite large, and important information may be lost. It has been the writer's observation that when this happens the resulting multiple correlation coefficient will be so distorted that the problem becomes quite evident (e.g., $R_{\emptyset.12\dots k}^2$ is negative).

*BETA--This final program prints out most of the information in which the user is interested. It first prints a row vector of standardized regression coefficients (β 's), followed by the vector of raw-score regression weights (b's) and the intercept. It then prints the squared multiple correlation coefficient ($R_{\emptyset.12\dots k}^2$) and $R_{\emptyset.12\dots k}$.

This program operates by use of the following formulae.

1. $\underline{\beta} = R^{-1} \cdot \underline{r_{\emptyset i}}$ where $\underline{\beta}$ is a vector of beta weights, $\underline{r_{\emptyset i}}$ is the vector of correlations between the predictors and the criterion.

2. $b_i = \beta_i (s_0/s_i)$ where s_i is the standard deviation of variable i .
3. $a = b_0 = \bar{X}_0 - \sum b_i \bar{X}_i$
4. $R_{0.12\dots k}^2 = \sum \beta_i R_{0i}^2$

Alternative Uses and Modifications

There are several ways in which *REGRES may be modified to allow it to handle special cases. These modifications have not been incorporated into the current version, however, because they tend to make it too unwieldy for the general case, primarily by increasing execution time.

Increasing the Maximum Number of Predictors

Whenever the user wishes to use more than nine predictors some modifications must be made in *CORMAT, and possibly in *INVERSE. Deleting lines 1.85, 1.90, and 1.95 of *CORMAT will allow the user to use at least 14 predictors, and possibly more. When these deletions are made, the execution will end after *CORMAT, and the user will have to start *INVERSE by entering direct commands.

In addition to the above changes, with more than a total of 10 variables the user will have to rewrite the instructions in both programs which print out the matrices. The largest matrix which can now be printed is a 10 x 10. This can be remedied by minor changes.

Although in extreme cases it is possible to increase the number of variables by modifying lines 7.01 and 7.02 of *CORMAT in such a way as to store X(I) and T(I) in files, this is probably not advisable. If the user has enough variables to require this modification, he will have so many variables that execution time will be extremely long.

Stepwise Regression

Stepwise multiple regression deals with running a series of multiple regression analyses, each time adding or deleting a variable. A limited stepwise analysis can be run using this program with only minor changes in the current version, assuming the user knows the order in which he wishes to delete variables from the regression equation, he first enters the predictors in the reverse order under *CORDATA, (i.e. the first variable to be deleted is the last variable entered). Thus if the largest analysis will consist of five predictor variables, he will enter the criterion and all five predictors for the first analysis.

Two modifications in the programs now need to be made. a) Line 7.01 of *CORMAT would be altered to read:

```
7.01 F J=1,V;S X(J)=F2(J+6*(R-1)+1)
```

Note here that the subscript of F2 now has a V replaced by 6, since there are a total of 6 variables in the data matrix, regardless of subsequent values of V. b) The second modification must be made at the end of *BETA. After line 1.90, the user must supply instructions to open CORDATA, reduce the first element of CORDATA (V) by 1, close CORDATA, and load and start *CORMAT.

These changes will have the effect of causing *REGRES to keep recycling, each time reading one variable fewer. It is very important that the user plan carefully the order in which he wishes to have the variables deleted, so that the analysis answers those questions which the user wants answered.

Program Listing

A listing of the four programs which comprise *REGRES follows.

These programs should be entered exactly as written, as space is very limited.

*CORMAT, for example, leaves only 39 storage locations. If this number is decreased to 35 locations, the program will not run with 10 variables.

L L,*CORDATA,

*W

C FOCAL-12

```

Ø1.Ø1 E
Ø1.Ø5 T "THIS PROGRAM LOADS DATA FOR *CORMAT IN CORDATE",!
Ø1.Ø6 T "ENTER ONE SUBJECT'S DATA AT A TIME, CRITERION FIRST",!!
Ø1.1Ø A "NO. OF VARIABLES",V,"NO. OF SUBJECTS",N,;
Ø1.15 L 0,F1,F,CORDATA,Ø
Ø1.2Ø F I-1,N;T !;F J=1,V;A F1(J+V*(I-1)+1)
Ø1.25 T !!
Ø1.3Ø A "DO YOU WISH ANY CHANGES?--(1 OR NO)",ANS,!
Ø1.35 I (ANS-1) 1,6Ø,1.4Ø,1.6Ø
Ø1.4Ø A "ENTER LOCATION AND THEN CORRECT DATA POINT",L,D,!
Ø1.45 S F1(L+1)=D
Ø1.5Ø A "MORE CHANGES?--(1 OR NO)",ANS,!
Ø1.55 I (ANS-1)1.6Ø,1.4Ø,1.6Ø
Ø1.6Ø S F1(Ø)=V;S F1(1)=N
Ø1.95 L C,F1;L G,*CORMAT,Ø
Ø1.99 Q
*
```


L L,*CORMAT,
 *W
 C FOCAL-12

```

Ø1.Ø1 E
Ø1.Ø2 L 0,F1,F,CORFIL,Ø;L 0,F2,F,CORDATA,Ø
Ø1.1Ø S V=F2(Ø);S NN=F2(1);S K=V*V;F I=1,K;S F1(I)=Ø
Ø1.2Ø F R=1,NN;D 7
Ø1.25 L C,F2; L 0,F2,F,PRED,Ø
Ø1.3Ø F I-1,V*V;S F2(I)=Ø
Ø1.5Ø T %5.2,!:,"VARIABLE MEAN SD N=",N
Ø1.55 F I-1,V;T !,I," ",T(I)/N," ",FSQT(F1(I+V*(I-1 ))/(N-1) )
Ø1.6Ø T !:,"R-MATRIX",!," "
Ø1.65 F I=1,V; T %2.Ø,I," "
Ø1.7Ø T %4.3;F I=1,V;T !;F J=1,V;D 6
Ø1.72 F I=2,V;S F2( (V-1) ↑2+I-1)=F1(I)/FSQT(F1(1)*F1(I+V*(I-1) ) )
Ø1.73 F I=1,V;S F2( (V-1) ↑2+V-1+I)=FSQT(f1(I+V*(I-1) ))/(N-1) )
Ø1.74 F I=1,V;S F2( (V-1) ↑2+2*V-1+I)=T(I)/N
Ø1.75 T !:!:!:!:!:;S F2(Ø)=V-1;L C,F1;L C, F2
Ø1.85 A "CONTINUE?--(1 OR NO)",ANS,!
Ø1.9Ø I (ANS-1) 1.99,1.95,1.99
Ø1.95 L G,*INVERSE,
Ø1.99 Q

Ø3.Ø5 IF (N-1) 3.3Ø,3.4Ø,3.3Ø
Ø3.3Ø S F1(J+V*(I-1 ))=F1(J+V*(I-1 ))+(N*X(I)-T(I) )*(N*X(J)-T(J) )/(N ↑2-N)
Ø3.4Ø R

Ø6.2Ø T F1(J+V*(I-1 ))/FSQT(F1(I+V*(I-1) )*F1(J+V*(J-1) ) )," "
Ø6.25 I (I-1) 6.26,6.4Ø,6.26
Ø6.26 I (J-1) 6.3Ø,6.4Ø,6.3Ø
Ø6.3Ø S F2(J-1+(V-1)*(I-2) )=F1(J+V*I-V)/FSQT(F1(I+V*I-V)*F1(J+V*J-V) )
Ø6.4Ø R

Ø7.Ø1 F J=1,V;S X(J)=F2(J+V*(R-1)+1)
Ø7.Ø2 F I=1,V;S T(I)=T(I)+X(I)
Ø7.Ø4 S N=N+1;F I=1,V;F J=1,V;D 3
*
```


L L,*INVERSE,

*W

C FOCAL-12

01.01 E
 01.05 L 0,F1,F,PRED,0;L 0,F2,F,CORFIL,0
 01.20 S $N=F1(0)$
 01.25 F $I=1,N;S F2(I)=0$
 01.30 F $I-1,N;D 2$
 01.45 F $I=1,N;D 7$
 01.50 T !!!
 01.55 T "INVERSE",!!
 01.90 F $I=1,N;D 8$
 01.93 L C,F1;L C,F2
 01.94 T !!
 01.95 A "CONTINUE?--(1 OR NO)",ANS,!
 01.96 I (ANS-1) 1.99,1.97,1.99
 01.97 L G,*BETA,0
 01.99 Q

 02.01 S $T=0$
 02.05 F $J=1,N;D 3$
 02.06 S $F2(IC) = F2(IC)+1;I (IR-IC)2.08,2.20,2.08$
 02.08 F $L=1,N; D 5$
 02.20 S $F2(I+N)=IR;S F2(I+2*N)=IC;S T=F1(IC+N*IC-N);S F1(IC+N*IC-N)=1$
 02.35 F $L=1,N; S F1(L+N*(IC-1)) = F1(L+N*(IC-1))/T$
 02.40 F $M=1,N;D 6$

 03.01 IF $(F2(J)-1) 3.05,3.10,3.05$
 03.05 F $K=1,N;D 4$
 03.10 C

 04.01 I $(F2(K)-1) 4.02,4.10,1.90$
 04.02 I $(T-FABS(F1(K+N*(J-1)))) 4.05,4.05,4.10$
 04.05 S $IR=J;S IC=K;S T=FABS(F1(K+N*(J-1)))$
 04.10 C

 05.01 S $T=F1(L+N*(IR-1));S F1(L+N*(IR-1))=F1(L+N*(IC-1))$
 05.03 S $F1(L+N*(IC-1)) = T$

 06.01 I $(M-IC) 6.02,6.40,6.02$
 06.02 S $T=F1(IC+N*(M-1));S F1(IC+N*(M-1))=0$
 06.05 F $L=1,N;S F1(L+N*(M-1))=F1(L+N*(M-1))-F1(L+N*(IC-1))*T$
 06.40 C

 07.01 S $J=N-I+1;I (F2(J+N)-F2(J+2*N)) 7.03,7.05,7.03$
 07.03 S $K=F2(J+N);S L=F2(J+2*N)$
 07.04 F $M=1,N;S T=F1(K+N*M-N);S F1(K+N*M-N)=F1(L+N*M-N);S F1(L+N*M-N)=T$
 07.05 C

 08.01 F $J=1,N;T F1(J+N*(I-1)), " "$
 08.02 T !
 *

L L,*DETA,
 *W
 C FOCAL-12

01.01 E
 01.05 L 0,F1,F,PRED,0
 01.07 S MC=1;S NC=F1(0);S MR=NC;S NR=NC
 01.12 L 0,F3,F,PROD,0
 01.13 F I=1,81;S F3(I)=0
 01.21 T !
 01.25 F I=1,NC;S Y(I)=F1(NC*NC+I)
 01.26 T !
 01.50 F I=1,NR;D 2
 01.54 T !!
 01.55 T "THE STANDARDIZED REGRESSION COEFF. ARE GIVEN BELOW",!
 01.60 F I=1,NR;D 7
 01.65 T !!!!
 01.67 F I=0,NC;S SD(I)=F1(NC ↑ 2+NC+I+1)
 01.68 T !:
 01.70 T "THE REGRESSION WEIGHTS ARE GIVEN BELOW",!
 01.75 F I=1,NC;T F3(I)*SD(0)/SD(I), " "
 01.76 T !!!
 01.81 F I=0,NC;S M(I)=F1(NC ↑ 2+2*NC+2+I)
 01.82 F I=1,NC;S IN=IN + M(I)*F3(I)*SD(0)/SD(I)
 01.83 T !!, "INTERCEPT =",M(0)-IN,!!
 01.85 F I=1,NC;S R2=R2+F3(I)*Y(I)
 01.87 T "R-SQUARED",R2, " R",FSQT(R2)
 01.90 L C,F3;L C,F1
 01.99 T !!!;Q

 02.01 F J=1,MC;D 3

 03.01 F K=1,NC;D 4

 04.01 S F3(J+MC*(I-1))=F3(J+MC*(I-1))+F1(K+NC*(I-1))*Y(J+MC*(K-1))

 07.01 F J=1,MC;T F3(J+MC*(I-1)), " "
 *

Sample Run

The following example illustrates the use of this program.

Those entries supplied by the user, other than the data themselves, have been underlined.

The data consist of 6 scores for each of 10 subjects. The first entry in each line represents the subject's score on the criterion variable, and the following five entries represent his scores on the five predictors.

The reader will note that a change has been made in the data after they were entered. The 2nd score for subject eight was deliberately entered incorrectly to illustrate the procedure. Since this is the 37th data point (entry), the user entered 37 and then 16 (the correct entry) in response to the interrogation.

LG,*CORDATA,

THIS PROGRAM LOADS DATA FOR *CORMAT IN CORDATA
ENTER ONE SUBJECT'S DATA AT A TIME, CRITERION FIRST

NO. OF VARIABLES:5 NO. OF SUBJECTS:10

```
:0 :3 :0 :2 :3
:2 :6 :3 :1 :1
:3 :9 :7 :0 :5
:6 :7 :4 :4 :4
:7 :11 :2 :6 :2
:9 :14 :8 :3 :6
:5 :15 :9 :8 :0
:10 :15 :1 :7 :3
:11 :12 :11 :9 :4
:12 :15 :10 :8 :4
```

DO YOU WISH ANY CHANGES?--(1 OR NO):1
ENTER LOCATION AND THEN CORRECT DATA POINT:37 :16
MORE CHANGES?--(1 OR NO):NO

VARIABLE	MEAN	SD	N = 10.000
1.0000	6.5000	4.0346	
2.0000	10.8000	4.4171	
3.0000	5.5000	3.9791	
4.0000	4.8000	3.2249	
5.0000	3.2000	1.8135	

R-MATRIX

1	2	3	4	5
1.000	0.804	0.523	0.752	0.349
0.804	1.000	0.518	0.699	0.061
0.523	0.518	1.000	0.424	0.262
0.752	0.699	0.424	1.000	-0.220
0.349	0.061	0.262	-0.220	1.000

CONTINUE?--(1 OR NO):1

INVERSE

2.305	-0.475	-1.486	-0.344
-0.475	1.553	-0.431	-0.473
-1.486	-0.431	2.382	0.728
-0.344	-0.473	0.728	1.305

CONTINUE?--(1 OR NO):1

THE STANDARDIZED REGRESSION COEFF. ARE GIVEN BELOW

0.370 -0.060 0.624 0.480

THE REGRESSION WEIGHTS ARE GIVEN BELOW

0.338 -0.060 0.781 1.068

INTERCEPT $=-3.979$

R-SQUARED 0.903 R 0.550

*

