P.R.L.

TECHNICAL NOTE

TN-83-33

A PROGRAMMING SYSTEM ON MULTIPLE REGRESSION ANALYSIS

By

C.S.R.Murty

March, 1983

PHYSICAL RESEARCH LABORATORY
AHMEDABAD

CONTENTS

A PROGRAMMING SYSTEM ON MULTIPLE REGRESSION ANALYSIS

C.S.R.MURTY
Physical Research Laboratory
Ahmedabad - 380009

ABSTRACT

This note describes briefly the multiple regression analysis technique. A general programming system for doing the regression analysis has been developed in FORTRAN IV and tested using IBM 360/44. The package is quite versatile and fits any model which can be brought as a linear model. Various statistical tests are provided. The note describes to a large extent the control card information for using the package. The input and output for a sample problem is supplied.

Key Words:

1. Regression analysis
2. Curve fitting
3. Multiple regression analysis
4. Polynomial fitting.

1. Introduction:-

This note describes the Multiple Regression Analysis which is a powerful statistical tool generally used by many a scientist. The principle used underlying this tool is the 'principle of least squares'. A brief mathematical description of the technique is presented in section 2. This is followed by presentation of programming system details in section 3. Section 4 and 5 respectively contain the control card information for using the program and a sample problem alongwith its input and output.

2. Mathematical description:-

The multiple regression analysis is used to fit a linear relationship between a dependent variable y and several, say k, independent variables x_1, $x_2 \dots x_k$ when a sample of n observations on them are available. Let the linear relationship between them be of the form.

$$y = A_0 + A_1x_1 + A_2x_2 + \dots + A_kx_k + u \qquad \dots(1)$$

where A_i for i=0, 1....k are the k+1 regression coefficients to be determined and u is the error in estimating the observed value of y from the model using corresponding values of the independent variables. Equation (1) is true for all the n given observations and as such we have

$$y_i = A_0 + A_1 x_{1i} + A_2 x_{2i} + \ldots + A_k x_{ki} + u_i$$

$$\text{for } i = 1, \ldots n \qquad \ldots (2)$$

These equations can be written in matrix notation as follows:

$$Y = X A + U \qquad \ldots (3)$$

where

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{bmatrix}$$

is a column vector of n observations on y

$$X = \begin{bmatrix} X_0, & X_1, & \ldots & X_k \end{bmatrix}$$

$$= \begin{bmatrix} 1 & x_{11} & x_{21} & \ldots & x_{k1} \\ 1 & x_{12} & x_{22} & \ldots & x_{k2} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 1 & x_{1n} & x_{2n} & \ldots & x_{kn} \end{bmatrix}$$

is an n by (k+1) matrix

$$X_0 = \begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}$$

is a column vector of n unit elements and is required for estimating the intercept A_0

$$X_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \cdot \\ \cdot \\ \cdot \\ x_{in} \end{bmatrix}$$

is a column vector of n observations on x_i for $i = 1, \ldots k$

$$U = \begin{bmatrix} u_1 \\ \cdot \\ \cdot \\ \cdot \\ u_n \end{bmatrix}$$

is a column vector of n disturbances

$$A = \begin{bmatrix} A'_0 \\ A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_k \end{bmatrix}$$

is a column vector of k+1 coefficients

The A coefficients and the distribution parameters of U are unknown and are to be estimated. The following assumptions are generally made for estimating and statistically testing the parameters, exactly.

U is $N\ (0, \sigma^2 I_n)$ ... (4)

where N stands for normal and I_n the n^{th} order unit matrix.

X is a set of fixed numbers ... (5)

and X has rank $k+1 < n$. ... (6)

The assumptions (4) are necessary for statistical inference (or testing the hypotheses), while (5) and (6) are necessary for estimating the parameters of the model or coefficients A_i of the model.

The parameters of (1) are estimated using the principle of least squares.

Let $\hat{A} = (\hat{A}_0, \hat{A}_1 \ldots \hat{A}_k)$ represent a column vector of the estimates of A.

Hence $Y = X\hat{A} + e$

where e denotes the column vector of n residuals $(Y - X\hat{A})$.

According to the principle of least squares the coefficients $\hat{A}$ are obtained by minimizing the sum of the squares of the n residuals.

Thus we have to

$$Min\ Z = \sum_{i=1}^{n} e_i^2 = e^1 e$$

(where the superfix 1 indicates prime)

$$= (Y - X\hat{A})^1 \ (Y - X\hat{A})$$

$$= Y^1 Y - 2\hat{A}^1 X^1 Y + \hat{A} X^1 X\hat{A} \qquad \ldots (7)$$

($\because$ $Y^1 X\hat{A}$ is a scaler)

To find the coefficient vector $\hat{A}$ which minimizes Z, one has to differentiate Z with respect to the parameter vector $\hat{A}$ and then equate the same to zero.

$$\frac{\partial Z}{\partial \hat{A}} = \frac{\partial}{\partial \hat{A}} (e^1 e) = -2X^1 Y + 2X^1 X\hat{A} = 0$$

$$\therefore \quad X^1 X\hat{A} = X^1 Y$$

Hence $\hat{A} = (X^1 X)^{-1} X^1 Y$ (by virtue of (6)) ...(8)

Equation (8) represents a system of simultaneous equations generally known as normal equations.

It can be shown that

a) $E(\hat{A}) = A$, where E denotes the expected value which indicates that the least squares estimators are unbiased.

b) $Var(\hat{A}) = \sigma^2 (X^1 X)^{-1}$

c) The least squares estimators $\hat{A}_i$ are best linear unbiased estimators of the unknowns A_i.

d) $S^2 = e^1 e/(n-k-1)$ is an unbiased estimator of the disturbance variance.

e) $\hat{A}_i$ is $N(A_i, a_{ii}^{-1}\sigma^2)$ where a_{ii}^{-1} is the i^{th} diagonal element of the matrix $(X^1X)^{-1}$.

f) $\sum e_i^2/\sigma^2$ follows a chi-square distribution with $(n-k-1)$ degrees of freedom.

g) The test statistic 't' given by

$$t = (\hat{A}_i - B_i) / ((\sum e_i^2 / (n-k-1))^{1/2} (a_{ii}^{-1})^{1/2})$$

follows a 't' distribution with $(n-k-1)$ degrees of freedom, where B_i is any hypothetical value. The hypothetical value $B_i = 0$, is the one in which one is generally interested as it tests whether y is influenced by x_i or not linearly. Thus

$$t = \hat{A}_i / ((\sum e_i^2 / (n-k-1))^{1/2} (a_{ii}^{-1})^{1/2})$$

is generally provided as output in a regression package. Further 100 $(1-\alpha)$ percent confidence interval for A_i is given by

$$\hat{A}_i \pm t_{\alpha/2} (\sum e_i^2 / (n-k-1))^{1/2} (a_{ii}^{-1})^{1/2}$$

where t_α is the table value at α level.

h) To test whether any linear combination of the coefficients $\hat{A}_i$ is significantly different from a constant B, the 't' statistic is used and is given by

$$t = (L^1 A - B) \;/\; S\,(L^1 (X^1 X)^{-1} L)^{0.5}$$

where L is a column vector of constants which makes the linear combination.

i) To test the hypothesis that $\hat{A}_i = 0$ for $i = 1$ to k which is a test of whether $x_1, x_2, \ldots x_k$ have any influence on y the following F statistic is used.

$$F = \frac{1}{k}\left(\hat{A}^1 X^1 Y - \frac{1}{n}(\sum y_i)^2\right) \Big/ \left(\frac{e^1 e}{n-k-1}\right)$$

follows an F-distribution with k and n-k-1 degrees of freedom. The above F statistic is also equivalent to

$F = (R^2/k) \;/\; ((1-R^2) \;/\; (n-k-1))$, where R^2 is the coefficient of determination.

j) The coefficient of determination which is the square of the multiple correlation is given by

$$R^2 = RSS/TSS$$

where explained (or regression) sum of squares denoted by RSS is given by

$$RSS = \hat{A}^1 X^1 Y - (1/n)\,(\sum y_i)^2$$

and total sum of squares denoted by TSS is given by

$$TSS = Y^1 Y - (1/n)\,(\sum y_i)^2$$

and error sum of squares denoted by ESS is given by

$$ESS = TSS - RSS$$

$$\text{Mean } ESS = ESS/(n-k-1)$$

$$\text{Std.error} = (\text{Mean } ESS)^{0.5}$$

k) $\bar{R}^2$ (R bar-square) which is also known as adjusted R^2 (adjusted for degrees of freedom) is given by

$$\bar{R}^2 = 1 - \frac{(n-1)}{(n-k-1)} (1-R^2)$$

$\bar{R}^2$ is very helpful in comparing the explanatory power of different sets of explanatory variables. The addition of explanatory variables will always increase the value of R^2 but it is not true in case of $\bar{R}^2$.

l) The normalized coefficients are given by

$$\hat{a}_i = \hat{A}_i \; \sigma x_i \; / \; \sigma y$$

m) The Durbin-Watson statistic 'd' is given by

$$d = \sum_{i=2}^{n} (e_i - e_{i-1})^2 \Big/ \sum_{i=1}^{n} e_i^2$$

and is used for testing the presence of auto-correlation.

n) The first order auto correlation coefficient can be approximately calculated using the formula

$$r \doteq 1 - d/2$$

o) The reduction in regression sum of squares by deleting the i^{th} independent variable x_i is given by

$$\hat{A}_i^2 \; / \; a_{ii}^{-1}$$

p) Regression analysis can be used to make prediction of y corresponding to some set of x values as shown below.

Let C be a vector given by $(1, x_{1,n+1}, \ldots\ldots, x_{k,n+1})$ for which prediction of the expected value of y i.e. y_{n+1} is required and it is given by

$$\hat{y}_{n+1} = C^1 \hat{A}$$

and the confidence interval for the conditional expectation, $E(y_{n+1} / C)$ is given by

$$C^1 \hat{A} \pm t_{\epsilon/2} \, S \, (C^1 \, (X^1 X)^{-1} C)^{1/2}$$

and the confidence intervals for the individual y values are given by

$$C^1 \hat{A} + t_{\epsilon/2} \, S \, (1 + C^1 (X^1 X)^{-1} \, C)^{1/2}$$

3. Programming System*:-

The system contains a main program which calls the 'REACSR' subroutine. The main program supplies only the object time dimensions to the subprogram 'REACSR' which contains a varied number of program control cards described in the next section. The subroutine 'REACSR' inturn calls 2 more subroutine which are named as 'MINCSR' and 'PRACSR'. The system is designed to give a complete regression analysis for any given data as described in the previous section.

The subroutine 'MINCSR' evaluates the inverse of the variance-covariance matrix or the raw moment matrix depending on the value of some parameter in the program control cards to be described in section 4.

The subroutine 'PRACSR' is used to obtain the predicted values for any set of values taken by the independent variables. It also provides the confidence limits for the predicted values. The revoking of this subroutine depends on the value of a certain parameter of a program control card to be described in the next section.

The program is quite versatile and one can fit any model to any data which can be brought under the purview of linear parametric models, as can be viewed from the control card information. Around 21 generally used transformations are provided to make some of the non linear (in variables) models as linear models. For example polynomial equations, cob-douglus production functions, exponential functions can be fitted using this package. This package has the facility to add, subtract, multiply and divide variables to form new variables which may go into the regression model as a variable.

4. Control Card Information:-

The program does a complete regression analysis for any given data and prints a detailed output which any user of the technique generally needs. All computations are done in double precision. The input data is read in single precision.

Input:

The input to the program consists of a minimum of 3 control cards and a maximum of 11 control cards. Each control card is either a single 80 column card or sometimes more than 1 depending on the type of control card and the number of parameters to be read (for details see different control cards).

Control Card No.1:

This control card specifies the object time format for reading the basic input data. This control card reads two 80 column cards. If the input format can be accommodated in a single 80 column card, the other card is a blank card.

Control Card No.2:

This card is very crucial and depending on this 1 to 7 additional control cards may be read. This control card reads 25 parameter values in 25I2 format.

Parameter 1:-

If new variables have to be generated for fitting the regression model using the given variables, the value of this parameter should be supplied as 1. If the value of this parameter is 1 then one has to supply control card(s) No.8 (for details see control card(s) No.8). If the value of this parameter is 0 control card(s) No.8 will not be read. For this parameter a value of 1 should be supplied when one is fitting polynomials.

Parameter 2:-

This parameter is used in conjuction with parameter 1. If the values taken by the variables used in regression analysis are positive then the value of this parameter should be 0.

It should also be zero when the power to which a variable raised is a fraction. If all or a few of the values are negative the value of this parameter must be specified as 1.

Parameter 3:-

This parameter specifies the input unit from which data is read. If 0 or 5 card reader.

Parameter 4:-

This parameter reads the number of independent variables in the regression equation.

Parameter 5:-

This parameter reads the number of dependent variables. Using the same independent variables and assuming the same model, a regression equation is fitted for each dependent variable. All these regression models (one for each dependent variable) constitute one set.

Parameter 6:-

This parameter specifies the sum of the number of terms in the regression equation excluding constant term, and the number of dependent variables. If the value of parameter 1 of this control card is 0, then the value of this parameter should be supplied a value equal to the sum of independent variables and dependent variables.

Parameter 7:-

If specified value of this parameter is 1, then control card No.4 will be read. If value is 0 there will not be control card No.4.

Parameter 8:-

If supplied value to this parameter is 1, then control card(s) No.6 will be read. If the value is 0, there will not be control card(s) No.6.

Parameter 9:-

If value given to this parameter is 1, then control card(s) No.7 will be read. If value is 0 control card(s) No.7 will not be present.

Parameter 10:-

If the value given to this parameter is 1, then control card (s) No.8 will be read. If the value read is 0 there will not be control card(s) No.8.

Parameter 11:-

This parameter specifies the output unit on which data has to be written. This parameter is used for fitting different models using the same data.

Parameter 12:-

This parameter should be given 1, if one wishes to test a linear combination of the parameters, against a constant

..15

value. When the parameter takes 1, then control card No.3 and control card(s) No.10 have to be supplied. If the value to this parameter is 0 then no testing is required. (For details see these control cards)

Parameter 13:-

This parameter is needed if the value of parameter 12 is 1. This parameter reads the constant value to be specified by the user against which testing has to be done.

Parameter 14:-

If the value given to this parameter is 1, input data will be printed. If the value taken by this parameter is 0, then data will not be printed, if 2 data used in the model only printed.

Parameter 15:-

If the value taken by this parameter is different from 0 then control card No.3 will be read. If it is non zero, original data, observed values of the dependent variables, estimated values of the dependent variables, the residuals and the percentage deviations will be written on the unit specified by this parameter. If the value of this parameter is 0, then they will not be written and no need to supply control card No.3.

Parameter 16:-

In a regression model a constant can be used as a variable i.e. a variable takes constant value. If one of the variables is constant the regression equation should not contain a constant term again. However, this parameter reads the variable number which is a constant. If the value read by this parameter is 0, no variable is a constant.

Parameter 17:-

If the value read by this parameter is other than 0, then Inverse matrix of the regressors will be written on the output unit specified by this parameter in the format specified by control card No.3. If given value is not 0 then control card 3 has to be supplied.

Parameter 18:-

When any transformation is used for dependent variable and if value given to this parameter is 1, then observed value, Calculated value, their differences and percentage differences are given in transformed dependent variable(s). If the value supplied is 0 then by retransformation, given in original variables.

Parameter 19:-

If the supplied value to this parameter is non zero then control card 3 has to be supplied, and transformed data will be written on the unit specified by this parameter. If the given value is 0, then one need not supply control card 3 and transformed data will not be written.

Parameter 20:-

This parameter reads the number of points on the curve at which extrapolation or interpolation is needed. If 0 neither extrapolation nor interpolation will be done. If the value of this parameter is other than 0, then the values of the independent variables at these points should be provided at the end of the data. These values are read in 5E14.8 format. For each point a different set of cards has to be supplied depending on the number of independent variables.

Parameter 21:-

If the value given to this parameter is 1, then the control card information does not form part of the output. On the other hand if the given value to this parameter is 0, it forms part of output.

Parameter 22 and 23:-

These are dummy parameters and does not play any role in program execution, hence they can take any numeric value including blanks or zeros.

Parameter 24:-

This parameter is used when combination of variables is nded. This parameter reads the total number of variables. When this parameter value is other than zero control card(s) No.5 will be required.

Parameter 25:-

When the value of parameter No.24 is other than zero, this parameter has to be given the number of new variables to be generated. New variables are formed by combining the variables originally supplied.

Control card No.3:-

This control card is required if the value specified for any one of the parameter numbers 12, 15, 17 and 19 of control card No.2 is other than zero. Even if more than one of the above parameters is different from zero, then also there will be only one control card. This control card reads the object time format for reading additional input or writing additional output. (For the types of input and output refer to the parameter numbers of control card No.2).

Control card No.4:-

This control card is present if the value taken by parameter No.7 is 1. This control card reads 40 parameters in 40I2 format. Out of these 40 parameters 8 are important and they will be discussed below. The rest of the parameters are mostly used for punching different portions of the output. Any one interested to have them can contact the author.

Minimum output will be provided in the absence of this control card. The minimum output consists of (1) value of constant (2) regression coefficients, their variances, standard deviations, T-tests and normalized coefficients (3) regression sum of squares, error sum of squares, mean error sum of squares, standard error, R^2, $\bar{\bar{R}}^2$, F-value. If more specific output is required than this minimum output then this control card should be supplied along with its parameter values for those specific outputs.

Parameter No.1:-

This parameter takes any value between 0 and 3 (including them). Depending on the value taken constant may or may not present in the regression model. If the value tàken by the parameter is either 0 or 2, then constant term will be included in the regression equation. If the value specified is 3 the constant term is present in the regression equation as a variable which takes a value 1 for each data point. If constant term is present as a variable it will be tested against 0 and the T-value will be provided. If constant is present (parameter value 0 or 2) but not as a variable, then the program does not provide T-value. If the constant is required to be tested against 0, then one has to introduce a variable which takes the constant value 1. The value supplied to this parameter 1 implies no constant term in the equation.

Parameter 2:-

If the value taken by this parameter is 0, then moment matrix of Regressors, Inverse matrix of the moment matrix of regressors, Matrix of regressors on regresond will form part of the printout. If the given value to this parameter is 1 they will not form part of the printout.

Parameter 3:-

If the value taken by this parameter is 1, then control Card No.10 has to be supplied. When the parameter value is 1, the program calculates and prints confidence limits for the coefficients. If the value given is 0, then there will not be control card No.10 and confidence limits will not be calculated.

Parameter 4:-

If the value taken by this parameter is 1 then expected values of the dependent variable(s) will be calculated using the estimated parameters of the equation. When the given value is 1, observed values, calculated values their differences and percentage differences for each dependent variable will be printed provided the number of observations are less than or equal to 600. If the number of observations are greater than 600 and if one desires to get the above calculations done, then one has to specify a non zero value to the 19th parameter of control card No.2 (see 19th parameter of control card No.2), in addition to specifying this parameter as 1. If the specified value of this parameter is 0, the calculations will not be done.

Parameter 5:-

If the value taken by this parameter is 1, then Durbin-Watson Statistic and serial correlation will be calculated and printed. If the supplied value is 0 then the above calculation is not done.

Parameter 6:-

If the value given to this parameter is 1 then correlation matrix of all the variables will be printed. If the value supplied to this parameter is 2, then loss in error sum of squares will be computed by deleting one independent variable at a time and the output will be provided in addition to the correlation matrix. If the specified value of this parameter is 0 then neither of the above will be calculated.

Parameter 7:-

If the value taken by this parameter is 1, then means and standard deviations of all the variables will be printed. On the other hand if the given value is 0 these will not be printed.

Parameter 8:-

This parameter is used along with parameter 4 of this control card. This parameter is used for printing or punching purposes. Specification of a value of 1 to the parameter 4 of this control card does the calculations (see parameter 4 of this control card) while the parameter 8 decides whether to print or punch or not do either. If the specified value of this parameter is either 0 or 2, the observed, expected values of the dependent variable(s), their differences and percentage differences will be printed. If the taken value of this parameter is 1 they will be punched. If the supplied value to this parameter is 3 they will neither be printed nor punched.

Control card(s) No.5:-

This card(s) will be required whenever the values taken by 24th and 25th parameters of control card No.2 are different from zero. The number of cards is equal to the value taken by the 25th parameter of control card 2. Each card reads as the first parameter in 'I2' format, the number of variables used for forming a new variable. This is followed by the number of the old variable and its corresponding coefficient, in forming the combination to make the new variable, upto a maximum of 13 variables in the format I2, F4.1. The format of this card is thus (I2, 13(I2,F4.1)). These new variables are the input for subsequent control card information. The new variables are the one that will be treated as independent and dependent variables for supplying information with regard to parameters 4 and 5 of the 2nd control card.

Example:

Suppose from 5 variables 3 new variables have to be generated as follows:

$$Nx_1 = 5x_1 + 3x_2 - x_5$$

$$Nx_2 = x_3 + 2x_5$$

$$Nx_3 = x_1 + x_3 + 2.5x_5$$

where Nx_1, Nx_2, Nx_3 are new variables. Parameter 24 and 25 of card No.2 should be 5 and 3 respectively. There will be 3 control cards of type 5, which are as follows:

0301ƀ5.002ƀ3.005-1.0 – 1st card

0203ƀ1.005ƀ2.0 – 2nd card

0301ƀ1.003ƀ1.005ƀ2.5 – 3rd card

Punching of the above cards is from col.no. 1 and ƀ is blank.

Control card(s) No.6:-

This card (some times 2) is used mainly for rearranging the variables so that dependent variables are kept in the end, which is a prerequisite to the execution of this program, if they are already not punched in the end. This card can also be used for rearranging for any other purpose the user requires. This card can also be used for suppressing certain variables (which can also be achieved by control card No.1). This control card reads the parameters in 40I2 format. The first parameter specifies the number of variables read by control card No.1. The maximum number of variables read is 99, however. The other parameters are nothing but the variable numbers of the variables (equal to the sum of the values specified by parameters 4 and 5 of control card No.2). The numbers of the dependent variables should be punched in the end. If the total number of variables used in the regressions are greater than 39, 2 card(s) are necessary. These cards will be present if the value of 8th parameter of the control card No.2 is 1.

Control card(s) No. 7:-

These card(s) are present if the 9th parameter of the control card No.2 is 1. The number of parameters read is equal to sum of the values of parameter numbers 4 and 5 of control card No.2. The values of the parameters are read in 40I2 format. If the total number of variables (including dependent variables) is greater than 40 then one will have more than one card of this type. Since the program handles a maximum of 50 variables a maximum of 2 cards of this type are needed. These parameters read the transformation required for each one of the variables.

Transformations and their corresponding codes are given below. Let x be the original variable and A any constant.

Code No.	Original variable	Transformed variable
1	x	e^{x}
2	x	e^{-x}
3	x	SIN $(2\pi x)$
4	x	COS $(2\pi x)$
5	x	10^{x}
6	x	10^{-x}
7	x	$LOG_{e}x$
8	x	$LOG_{10}x$
9	x	$x^{0.5}$
10	x	$\frac{1}{x^{0.5}}$
11	x	$1/x$
12	x	$1/LOG_{e}x$
13	x	$1/LOG_{10}x$
14	x	$LOG_{e}(1/x)$
15	x	$LOG_{10}(1/x)$
16	x	$x + A$
17	x	$x - A$
18	x	$x * A$
19	x	A/x
20	x	x^{A}
21	x	x/A

..25

Once transformation is used for any variable, the original variables will not exist and all the calculations will be carried out in terms of transformed variables.

Control card(s) No.8:-

These card(s) will be present if the value of 10th parameter of control card No.2 is 1. These card(s) are used for completing the transformations against code numbers 16 to 21 specified by control card(s) No.7. When transformation codes 16 to 21 are required for any one of the variables one has to specify 10th parameter value of control card 2 as 1, otherwise the transformations are not complete. These card(s) are used to read the constant value A for each variable (whether transformed by these codes (16 to 21), or not) in 5E14.8 format. Number of constants (parameters) read by these control cards is equal to the sum of the values of 4th and 5th parameters of control card No.2. Since the maximum number of variables that the program can handle is 50, one can have a maximum of 10 cards of this type. If no variable is to be transformed by these codes (16 to 21) then these cards will not be needed and hence value of 10th parameter of control card No.2 should be specified as 0.

Control card(s) No.9:-

These control card(s) will be required if the value of the parameter 1 of control card 2 is 1. These card(s) are generally used when the number of terms in a regresson model is greater or less than the number of independent variables. (The terms used here is the number of terms in single regresson equation).

They will generally be present when one is forming new variables, by taking products, quotients, powers of the given (or old) variables. For example when one is required to fit a polynomial these card(s) have to be supplied. These card(s) describe how each term of the regression equation is obtained from the original variables. The original variables here correspond to transformed variables by control card(s) No.7 but not to the original variables read. Each term (or new variable) is the product of a maximum of 3 given (or old) variables. Each term occupies 18 columns and 4 terms are described in one card. Each term reads 6 parameters, in 3 (I2, F4.1) format. The 1st, 3rd and 5th parameters read the required given (or old) variable numbers, while the 2nd, 4th and 6th parameters read the powers to which these variables have to be raised and multiplied to form the new variables or term. Since the program can handle upto a maximum of 50 terms one will have a maximum of 13 cards of this nature. If the 1st parameter value of control card No.2 is 0 these cards will not be present. The following example makes things clear. Suppose there are 3 independent variables x_1, x_2, x_3 and one dependent variable y. Let us suppose one wants to fit.

$$y = A_0 + A_1x_1 + A_2x_2^2 + A_3x_1x_2^3 + A_4x_1x_2x_3$$

Now there are 5 terms in the regression equation (excluding constant but including y).

So one requires 2 card(s) of this type.

Term	Coded	Starting column No. in card	Card No.
x_1	01ƀ1.0	1	1
x_2^2	02ƀ2.0	19	1
$x_1 x_2^3$	01ƀ1.002ƀ3.0	37	1
$x_1 x_2 x_3$	01ƀ1.002ƀ1.003ƀ1.0	55	1
y	04ƀ1.0	1	2

The first 4 terms are punched on the 1st card while the 5th in the 2nd card.

Control card(s) No.10:-

These control card(s) are required if value of the 12th parameter of control card No.2 is 1. These card(s) are necessary when one would like to test a linear combination of the parameters (coefficients) of the regression model against a given constant value. These card(s) read the weights of the linear combination to be tested in a format supplied by the user using control card No.3 (see control card No.3). The number of card(s) depend on the linear combination of the parameters (coefficients) used as well as the format (specified by control card No.3) used in reading the weights of the linear combination. If value of 12th parameter is 0 these card(s) are not required.

Control card No.11:-

This control card will be present if the 3rd parameter of the 4th control card is 1. This control card reads the T-value (generally 5% or 1%) for calculating the 95% or 99% confidence limits for the coefficients. The value is read in E14.8. If the 3rd parameter of 4th control card is 0, this control card will not be present. However for reading 4th control card 7th parameter of control card No.2 should be specified as 1.

Control card No.12:-

This control card reads 6 parameters in I5, I2, I1, I5, I3, A64 format.

Parameter 1:-

This parameter reads the number of observations. The analysis can be carried out using a maximum of 99, 999 data points.

Parameter 2:-

This parameter specifies where to branch, after processing one data set, whenever more than one data set, is processed.

Specified value	Start reading input from
0 or 1	first control card
2	second control card
3	third control card
4	fourth control card
5	fifth control card
6	sixth control card
7	seventh control card
8	eighth control card
9	ninth control card

..29

Specified value	Start reading input from
10	tenth control card
11	eleventh control card
12	twelth control card
13	direct data

Parameter 3:-

If the specified value of this parameter is 1, then the input unit will be rewound.

Parameter 4:-

This parameter is used for by pasing as many records as specified by this parameter. A maximum of 99, 999 records can be by passed.

Parameter 5:-

This parameter specifies the number of sets of data to be processed. A maximum of 999 sets can be processed at a time. If unspecified it assumes a value of 10. So without specifying one can process 10 sets.

Parameter 6:-

This parameter is used for title. A maximum of 64 columns are provided for title. The punching of the title starts from 17th column. One can punch the title any where between 17th to 80th column.

For the first data set control card No.1, 2 and 12 are compulsory, and depending on the parameter values of control card No.2 other control cards may be present. For subsequent sets in the same run, number of control cards depend on the parameter values of control card No.2 and 12.

5 Sample example:

Fit an exponential growth curve of the form $y = AB^x$ to the following data

x	y
3.0	29.0
4.0	42.0
5.0	55.0
9.0	200.00
15.0	300.0

Further suppose the data is supplied through cards punched in 2F10.4 format. The number of given data points is 5. It is required to print averages, standard deviations, correlation matrix, moment matrix and its inverse in the transformed variables as given below. Further, observed, expected value of the dependent variable, their difference and percentage difference in original variable are also required. Also Durbin-Watson statistic 'd' and an estimate of the auto correlation, are to be printed.

..31

Since the curve to be fitted is non-linear in variables it has to be transformed into a linear model by taking logarithms to the base 'e' on both sides of the equation.

$$Log_e y = Log_e A + x Log_e B$$

which can be written as

$$z = C + Dx$$

which is a linear function of z on x, where z is transformed variable and C and D are new coefficients to be estimated. Given the above information, one will proceed for preparing input to the program which is as follows:

Input:-

Control card No.1:

1.............................. column number
(2F10.4)

followed by a

BLANK CARD

SECOND CONTROL CARD:-

000000000111111111 column number

123456789012345678 column number

bbbbbb01 01 0201 bb01 values

FOURTH CONTROL CARD:-

00000000011111 column number

12345678901234 column number

bbbbbb01 01 01 01 values

SEVENTH CONTROL CARD:-

1234......	column number
~~bb~~07	values

TWELFTH CONTROL CARD:-

1234567......	column number
~~bbbb~~5~~bb~~	value

After these control cards, the data is provided according to the format specified by first control card, namely (2F10.4) with each card containing one observation(in the order of x and corresponding y) per card.

Output:-

The output is given below in capital letters. The parameters estimated are $C = Log_e A$ and $D = Log_e B$, but not A and B and to get A and B retransform them by using $A = e^C$ and $B = e^D$. The results are provided in scientific notation.

OUT PUT OF SET 1

OUT PUT OF CODE

NUMBER OF INDEPENDENT VARIABLES		1
NUMBER OF DEPENDENT VARIABLES		1
NUMBER OF TERMS		2
NUMBER OF OBSERVATIONS		5
DEGREES OF FREEDOM	1	3

..33

MEAN AND STANDARD DEVIATION OF VARIABLES

	MEAN	STD
1	0.72000000D 01	0.44000000D 01
2	0.44228792D 01	0.91252975D 00

CORRELATION MATRIX

1 0.95175761E 00

MOMENT MATRIX OF REGRESSORS

1 0.96800000D 02

VALUE OF DETERMENANT IS -.96800000D 02

INVERSE OF THE MATRIX IS

1 0.10330579D-01

MOMENT MATRIX OF REGRESSAND BY REGRESSORS

1 0.19107158D 02

D.VAR	CONSTANT
1	0.30016857D 01

D.VAR	COEF.	VAR.	STD.	T TEST
1	0.19738799D 00	0.13499629D-02	0.36741841D-01	0.53722945D 01

N.COEF.

0.95175765D 00

D.VAR	R.S.SQ	E.S.SQ	M.E.S.SQ
1	0.37715235D 01	0.39202923D 00	0.13067641D 00

STD.ERROR	R-SQUARE	CO.R-SQUARE	F-RATIO
0.36149192D 00	0.90584262D 00	0.87445682D 00	0.28861548D 02

OBSERVED	CALCULATED	DIFFERENCE	% DIFFERENCE	
0.28999984D 02	0.36373833D 02	-.73738496D 01	-.25427080D 02	1
0.41999974D 02	0.44311208D 02	-.23112341D 01	-.55029400D 01	1
0.54999979D 02	0.53980649D 02	0.10193297D 01	0.18533200D 01	1
0.19999992D 03	0.11888749D 03	0.81112425D 02	0.40556210D 02	1
0.29999988D 03	0.38858252D 03	-.88582642D 02	-.29527540D 02	1

D.VAR	DURBIN WATSON D STATISTIC	SERIAL CORRELATION
1	0.24331036D 01	-.21655181D 00

..34

6. How to execute the program ?

```
Job card
// EXEC RACSR
Control Card 1
Control card 2
Control Card 4
Control Card 7
Control Card 12
Data
/*
/*
/&
```

7. Acknowledgement:-

I am thankful to Shri S.R.Thakore, Dr.(Mrs.) S Rao and Dr.Dinesh Patel for their encouragement during this development. Thanks are due to the scientists from various organisations who used this package extensively and thereby helped in generalising and thoroughly testing the system. Thanks are also due to Mr.V.Ranganathan for typing the manuscript and Miss Meena Shah for stenciling.

8. Bibliography:-

a. Statistical Methods
Snedecor G.W., Cochran W.G., 1967, Oxford & I B H

b. Econometric Methods
J.Johnston, 1960 ; McGraw Hill Book Co., New York

c. Statistics in Psychology & Education
Garrett H.E., Woodworth R.S., 1969
Vakils, Feffer and Simons Pvt.Ltd., Bombay

d. Errors of observation and their treatment
J.Topping, 1972, Chapman and Hall