Institute of Economic Development and Research
SCHOOL OF ECONOMICS
University of the Philippines

Discussion Paper No. 68-33 December 2, 1968

REGRES: A LINEAR REGRESSION PROGRAM

GUIDE FOR USERS

by

Mahar K. Mangahas, 1944-

NOTE: IEDR Discussion Papers are preliminary versions circulated privately to elicit critical comment. References in publications to Discussion Papers should be cleared with the author.

REGRES: A LINEAR REGRESSION PROGRAM
GUIDE FOR USERS

REGRES is a program now available at the U.P. Computer Center for doing linear regressions. It consists of REGRE, a sample main program accompanying the Center's Scientific Subroutine Package, and some modifications to provide for computations of interest to economists. This guide aims simply to provide a minimum amount of information necessary in order to use REGRES.

Figure 1 is a reproduction indicating all the cards necessary to run two regressions from data consisting of twelve observations on five variables. The only cards which change from problem to problem are those which are labelled; they are the user's control cards. The order in which they appear, except for subset selection cards, may not be altered.

1. The job card. There are to be no blank spaces other than the two indicated. The job name, which is RICE 0006 in the example, has a maximum of 8 spaces. The user's project number goes into the space occupied by 0801003. A maximum of 20 spaces are allowed for the name of the user.

2. The data format card (the eleventh card in the deck). This gives the format in FORTRAN of each joint observation on the variables as

they appear on the forthcoming data deck. Not all the variables punched in the data deck need to be provided for in the format statement if they are not all included in the regression. In the example there are six columns of numbers in the data deck but only the first five are formatted. The maximum number of formatted variables is 40.

3. The problem parameter card.

Col. 1-6	Description or number of problem
Col. 7-11	Number of observations (at least one greater than the number of formatted variables, limit 99,999)
Col. 12-13	Number of formatted variables
Col. 14-15	Number of selections, or equations to be estimated (no limit)

The latter three parameters need to be right-justified, i.e., aligned with the last column on the right of the respective fields provided for them. In the example, the problem is called simply "1"; there are 12 observations, 5 formatted variables, and 2 equations to be estimated.

4. The data deck. The data should be punched exactly in accordance with the format statement. They are scanned by the computer from left to right and identified in numerical order. In the example, for the first observation we have

$$X_1 = 8107, \quad X_2 = 42, \quad X_3 = 78, \quad X_4 = 2125 \quad \text{and} \quad X_5 = 1.$$

The sixth column, consisting entirely of ones, is not used in this problem. Its usefulness for some tests of hypotheses is touched upon later.

5. Subset selection cards. One card per selection of regression, with no limit on number of regressions.

Col. 2	0	if the table of residuals is not desired;
	1	if the table of residuals is desired.
Col. 4	0	if residuals are to be plotted;
	1	if actual and computed (from the regression equation) values of the dependent variable are to be plotted. This column is effective only if a 1 is punched in Col. 2.
Col. 5-6		Dependent variable.
Col. 7-8		Number of independent variables.
Col. 9-10		First independent variable.
Col. 11-12		Second independent variable.
Col. 13-14		Third independent variable, etc.

In the example, two regressions are desired. Residuals are called for by each. The first selection requests a plot of residuals, while the second requests a plot of actual and computed values of the dependent variable. In both selections the dependent variable is X_1; there are three independent variable, namely X_2, X_4 and X_5. X_3 is also available but not used.

A copy of the output produced by the sample problem is attached. The computations of the two selections are of course identical and only the plots differ. The headings are probably self-explanatory. In the column "CORRELATION X VS Y", X refers to the variable at left and Y refers to the dependent variable.

"R-SQUARED" is the coefficient of determination and "RBAR-SQUARED" is the same coefficient as "corrected for degrees of freedom", usually denoted by $\bar{R}^2$.

Unfortunately REGRES does not give the standard error of the intercept. Neither can the intercept be easily suppressed (the regression "forced through the origin") if the user so desires. An entirely different program would probably have to be written in order that these features be added. REGRES uses the correlation (CORRE) and multiple regression (MULTR) subroutines available in storage, by which the data are immediately converted into deviations from means so as to decrease by one the rank of the matrix to be inverted. For tests of hypotheses requiring the entire estimated covariance matrix of the regression coefficients, with the constant considered as one of the regressors, it is possible to use other subroutines available at the Center to separately compute the matrix usually denoted $(X'X)^{-1}$, where X is the matrix of regressors including a column of ones for the sake of the intercept. This accounts for the column of ones in the sample data deck.

The plots may be turned on their side so that the horizontal line of dots becomes the y-axis and the vertical line the x-axis. In the plot of residuals, the x-axis refers to a zero (approximately) residual. The residuals are denoted by E. In the second plot, with Y denoting an actual value and * a computed

value, the x-axis corresponds to Y = 0. Whenever the plotted points for Y and * coincide, only * is printed. Examination of the plots reveals a few rounding errors; for example, in case no. 1 (observation no. 1) a residual of -61.2 produces a Y and * which coincide, while in case no. 8 a residual of 59.5 produces a Y and * which do not. Such plotting errors seem unavoidably more frequent if the x-axis is to refer to Y = 0 rather than say to $Y = \bar{Y}$ (and running down the middle of the page). REGRES can be accordingly modified if so desired.