A PROGRAM FOR APPLYING THE PRINCIPLE OF PARSIMONY IN MULTIPLE REGRESSION

James B. Bartoo*, Danuta Hiz**, and Donald T. Laird**
The Pennsylvania State University, University Park, Pa.

1. Introduction.

A program is described for solving problems in multiple regression analysis on a medium scale computer, PENNSTAC. While the general principles involved are not new, several features of the routine to be discussed make it quite suitable for most of the applications which have been encountered.

PENNSTAC, which was built by the Department of Electrical Engineering at The Pennsylvania State University, is a magnetic drum computer with roughly the same capabilities as the IBM Type 650. It uses a binary-coded decimal number system, and the memory holds 2,500 eleven digit words. Numerical quantities are represented by ten digits with a sign; the position of the decimal point is the same for all numbers, but can be set by the operator to any suitable place for a given problem. The input-output medium is punched paper tape. Input speed is quite adequate, but use of output must be economized.

2. Statement of Problem.

Let y be a variable dependent on p-1 other variables, $x_1, x_2, \ldots, x_{p-1}$. The x's are not necessarily functionally independent of one another. For example, x_2 may be equal to x_1^2. The classical multiple regression problem assumes there exists a linear relationship between the variables of the form $y = \beta_0 + \sum_{i=1}^{p-1} \beta_i x_i$ and estimates of the β-coefficients are obtained on the basis of n sets of observations $(y_j, x_{1j}, x_{2j}, \ldots, x_{p-1,j})$ $j = 1, \ldots, n$, using the principle of least squares. Usually the goodness of this least squares fit is considered to be inversely proportional to the residual variance about the least squares function $y = b_0 + \sum_{i=1}^{p-1} b_i x_i$ where the b_i are estimates of the β_i.

In most estimation problems the economics of gathering data, etc., dictate that a minimal number of variables be used in the regression equation. In regression analysis this principle of parsimony has been applied in various ways. The escalator method is perhaps the best known of such methods. Here, independent variables are added one at a time to the regression equation according to some criterion and the effect of the additional variables on the residual variance noted. If this residual variance is significantly reduced the added variable is retained in the regression equation; otherwise it is discarded.

Another method of applying this principle of parsimony in regression is to start with all available variables in the regression equation and then to eliminate some according to some criterion. This might be called the descending escalator method as opposed to the ascending escalator method. The procedure outlined in this paper is of the descending escalator type.

3. The Elimination Procedure.

Let b_i be the least squares estimate of β_i, the true regression coefficient of x_i, and let $S_{b_i}^2$ be its sample variance. In the multiple regression program for PENNSTAC the procedure for eliminating variables is as follows:

---

*Dept. of Mathematics. Work supported by National Science Foundation grant G-3270.
**Computer Laboratory, Dept. of Electrical Engineering.

a. At any stage of the elimination procedure the variable x_i is a candidate for elimination if the quantity $b_i^2/S_{b_i}^2$ is the smallest in the entire set of b^2/S_b^2 at this stage. We note that $b_i^2/S_{b_i}^2$ is proportional to t^2 where t is Student's statistic under the hypothesis that the true regression coefficient, β_i, is zero. t^2 is used here simply as an ordering device.

b. If p is the total number of variables in the initial regression equation (including the dependent variables) we let RSS_p be the residual sum of squares when all p variables are present in the regression, RSS_{p-r} be the residual sum of squares after r variables have been eliminated, and ESS_r be the sum of squares due to the r variables eliminated from the regression equation. Then $RSS_p + ESS_r = RSS_{p-r}$ where the degrees of freedom are $(n-p) + (r) = (n-p+r)$. The candidate for elimination, x_i, if it is the r^{th} candidate to be considered, is eliminated if the F ratio $[ESS_r/r] / [RSS_p/(n-p)]$ is not significant at the $100\alpha\%$ level, for a prescribed α. In this event a new set of regression coefficients and the associated residual sum of squares is calculated and the whole procedure is repeated. x_i is not eliminated if this F ratio is significant. Here, the F ratio is not used in its statistical sense but only as a criterion for eliminating variables.

4. Statistical Interpretation.

If we assume that the residuals are normally distributed about the hyperplane, $y = \beta_0 + \sum_{i=1}^{p} \beta_i x_i$, with homogenous variances, then the last F ratio in our elimination procedure may be interpreted as a means for testing the composite hypothesis that $\beta_i = 0$ for i ranging over the subscripts of the variables eliminated. The sample variances and covariances of the sample regression coefficients may be printed out thus providing a means for obtaining interval estimates of the β_i. The multiple correlation coefficient may be easily calculated from the residual variance.

5. Description of the Computer Program.

In its usual form, the input data consists of n observations of the dependent variable y and as many as nineteen independent variables x_i. The data tape also contains a set of parameters which control the various optional features of the program, and, if the elimination procedure is to be used, a table of the critical F ratios. Provision can be made for (nonlinear) transformations of the raw data during read in.

The computer processes this data tape to obtain the matrix of correlation coefficients. The dependent variable, y, is assigned to the p^{th} column and row of this matrix. The elements of the correlation matrix are the coefficients of the least squares normal equations if the variables are considered to be in standard measure. Hence all following computations assume that this standard measure is used. This permits the computation to proceed in fixed-point arithmetic without difficulties except in pathological cases.

The p-1 by p-1 matrix, augmented by the p^{th} row and column, is now inverted by the Gauss-Jordan method. At each stage the largest remaining diagonal coefficient is selected as the pivotal element. (Since the matrix is positive definite, it is not necessary to consider off-diagonal elements in finding the largest.) Optionally, the inverse matrix can be printed out.

The elements produced in the p^{th} column are the regression coefficients, b_i, in standard measure. The p^{th} diagonal element is the residual sum of squares of the dependent variable, RSS_p, and is equal to one minus the square of the multiple correlation coefficient. The residual variance is RSS_p divided by the degrees of freedom, which at this stage is n-p. The product of this residual var-

iance with the i^{th} diagonal element, $c_{i,i}$, of the inverse gives the variance $S^2_{b_i}$ of the estimated regression coefficient b_i. These results are printed.

The computer then proceeds to eliminate a set of variables from the regression equation. The input parameters may have specified that certain of the x_i are to be retained notwithstanding their contribution. Of the others, the variable for which $b_i^2/S^2_{b_i}$, or actually $b_i^2/c_{i,i}$, is minimum is taken to be the candidate for the first elimination. The increase in residual sum of squares due to elimination of this variable, ESS_1, would $b_1^2/c_{1,1}$. The variable is eliminated if $[ESS_1/1]/[RSS_p/(n-p)]$ is less than the critical F ratio. The elimination is performed by pivoting about $c_{1,1}$. Then the p^{th} column gives the new regression coefficients, their variances may be obtained from the new diagonal elements, and the p^{th} diagonal element is the new residual sum of squares, RSS_{p-1}. These results are printed.

Another variable is selected as a candidate for elimination, and the process is continued until the critical F ratio would be exceeded by elimination of another variable. When r-1 variables have been eliminated and x_i has been selected as the next candidate, the quantity ESS_r required for computing the F ratio is obtained by adding $b_i^2/c_{i,i}$ to ESS_{r-1}.

When the process stops, the inverse matrix has already been partly inverted, and the inversion is then continued without printout and the final matrix compared with the original correlation matrix to check for possible loss of significant figures in the inversion process.

6. Conclusions.

The program described here has been successfully used by research groups in several fields of endeavour at The Pennsylvania State University. The several optional features available have all proven useful, and would seem to be adequate for a fairly large class of problems.

The use of fixed-point arithmetic occasionally causes overflow difficulties. The computer will not attempt to complete the problem, if, during the matrix inversion an element greater than or equal to one hundred is developed. It can perhaps be argued that when this situation arises, the problem is not well posed, and hence the solution would not be very useful even if it were obtained. Since largest available pivotal coefficients have been used, some information is now available as to which variables cause the matrix to be ill-conditioned.

The most novel feature of the program is presumably the elimination procedure, used to obtain a parsimonious regression equation. By means of this procedure, the program discards from the regression equation a certain set of independent variables which, with the data available, are not proven to contribute anything to the explanation of the variance of the dependent variable. No claim can be made that the set of variables eliminated is that which reduces the residual variance the least, since not all combinations have been considered. But the order of elimination seems to be a reasonable one, and the result is a useful regression equation of reduced complexity which is, in the statistical sense, as acceptable as the equation with all variables present.