RAPIER - A FORTRAN IV PROGRAM FOR MULTIPLE LINEAR REGRESSION ANALYSIS PROVIDING INTERNALLY EVALUATED REMODELING

by Steven M. Sidik and Bert Henry

Lewis Research Center
Cleveland, Ohio

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • FEBRUARY 1970
RAPIER is a very flexible, easy to use, sophisticated multiple linear regression program which computes the variance-covariance matrix of the independent variables, regression coefficients, t-statistics for individual tests, and analysis of variance tables. The major value of the program is its comprehensiveness and options, such as a choice of three strategies for the variance estimate, an analysis of more than one set of response variables for the same independent variables, a backward rejection based on the first response variable, the use of weighted regression, computation of predicted values for any combination of independent variables, and a chi-square test for normality.
CONTENTS

<table>
<thead>
<tr>
<th>Summary</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Symbols</td>
<td>2</td>
</tr>
<tr>
<td>Estimation of Basic Linear Model</td>
<td>5</td>
</tr>
<tr>
<td>Basic Linear Model</td>
<td>5</td>
</tr>
<tr>
<td>Estimating $b$</td>
<td>8</td>
</tr>
<tr>
<td>Correlation Matrix</td>
<td>10</td>
</tr>
<tr>
<td>Estimating $\sigma^2$</td>
<td>11</td>
</tr>
<tr>
<td>Hypothesis Testing</td>
<td>12</td>
</tr>
<tr>
<td>Test NE - Normality of $e$</td>
<td>12</td>
</tr>
<tr>
<td>Analysis of Variance Table</td>
<td>15</td>
</tr>
<tr>
<td>Choice of Estimator for $\sigma^2$</td>
<td>16</td>
</tr>
<tr>
<td>Test OR - Overall Regression</td>
<td>17</td>
</tr>
<tr>
<td>Test SF - Sequential F-Test</td>
<td>17</td>
</tr>
<tr>
<td>Test TT - t-Tests</td>
<td>18</td>
</tr>
<tr>
<td>Predicting Values from Estimated Regression Equation</td>
<td>19</td>
</tr>
<tr>
<td>User's Guide to Input</td>
<td>20</td>
</tr>
<tr>
<td>Sample Regression Problem</td>
<td>20</td>
</tr>
<tr>
<td>Types of Input Cards</td>
<td>22</td>
</tr>
<tr>
<td>Sample Output</td>
<td>33</td>
</tr>
<tr>
<td>Appendixes</td>
<td></td>
</tr>
<tr>
<td>A - Program Documentation and Listings</td>
<td>46</td>
</tr>
<tr>
<td>B - Borrowed Routines</td>
<td>85</td>
</tr>
<tr>
<td>References</td>
<td>87</td>
</tr>
</tbody>
</table>
RAPIER - A FORTRAN IV PROGRAM FOR MULTIPLE LINEAR REGRESSION ANALYSIS PROVIDING INTERNALLY EVALUATED REMODELING

by Steven M. Sidik and Bert Henry

Lewis Research Center

SUMMARY

RAPIER is a digital computer program which can be used with ease to perform extensive regression analyses or a simple least-squares curve fit, and it includes a backward term rejection option. The program is written in FORTRAN IV, version 13, for the IBM 7094/7044 DCS. The major value of the program is its comprehensiveness of calculations and options.

RAPIER computes the variance-covariance matrix of the independent variables, regression coefficients, t-statistics for individual tests, and analysis of variance tables for overall testing of regression. There is a provision for a choice of three strategies for the variance estimate to be used in computing t-statistics.

Also, more than one set of response or dependent variables can be analyzed for the same set of independent variables.

A backward rejection option method based on the first dependent variable may be used. In this case, a critical significance level is supplied as input. The least significant independent variable is deleted and the regression recomputed. This process is repeated until all remaining variables have significantly nonzero coefficients.

The algorithm uses the triangular form of symmetric matrices throughout. It also allows for the use of weighted regression, computation of predicted values at any combination of independent variables, a table of residuals, and a chi-square test for the normality of the distribution of residuals.

INTRODUCTION

RAPIER is an almost entirely new multiple regression computer program. It is the result of 5 years of development in meeting the needs of several statistical investigators posing a variety of problems. The problems included analysis of nuclear reactor components, determining predictive models from corrosion and fracture data of both metals and alloys, investigating the behavior of processing variables in the manufacture of
solar cells, optimizing fuel-cell experiment procedures, and predicting personnel performance from academic histories.

The nucleus of the program is based on a program written by Kunin (ref. 1). However, in its present expanded form, it allows the user to choose from a number of sets of options which include options of input, of methods for calculation, and of output, thereby providing great flexibility.

With the aid of a few control cards, the program can be used readily for a wide range of applications which can vary from a simple least-squares curve-fitting problem to a complete regression analysis. It can provide the variance-covariance matrix of independent variables, regression coefficients, the variance-covariance matrix of the regression coefficients, individual t-statistics with their significance levels, analysis of variance tables for significance of regression, special usage of replicated data to estimate the error due to lack of fit, any one of three pooling procedures which may be used to estimate the error variance, tests for normality of distribution of the residuals, weighted regression, and the use of more than one dependent variable.

The mathematical analysis of the computations and their reliability is aided further by the option of obtaining an eigenvector decomposition of both the variance-covariance matrix and the correlation matrix of the independent variables.

The program also provides an option to perform a backward rejection regression at any given level of significance.

Despite its sophistication, RAPIER is relatively easy to use, but it presupposes that the user has at least a basic knowledge and/or experience in the application of statistical techniques.

To provide a framework for the discussion of the calculations and statistical options available in RAPIER, a brief description of multiple linear regression is presented, with no attempt to make the discussion thorough or rigorous. Notable presentations of applied regression analysis are those by Draper and Smith (ref. 2) and Graybill (ref. 3). Reference 4 by Kendall and Stuart is a useful guide to both applications and theoretical justifications. Rao (ref. 5) presents a more mathematically sophisticated treatment of the subject of linear statistical models.

After discussion of the calculations and options available, the card input necessary is described in detail and illustrated by an example which uses almost all of the options.

SYMBOLS

A matrix
A' transpose of A
A^(-1) inverse of A

2
B  matrix
b  vector (column)
b_i  true regression coefficient
\( \hat{b}_i \)  estimated regression coefficient
b_0  constant term
b_1, \ldots, b_J  unknown parameters
C  correlation matrix
C_{ij}  elements of C
D  indicator variable, equal to 0 if no \( b_0 \) coefficient is estimated and equal to 1 if \( b_0 \) is estimated
E(x)  expected value of x (i.e., average of x over all possible values of x)
e  vector of observation errors
F_{a,b}  statistic distributed as variance ratio with a and b degrees of freedom
f  expected number of observations in each cell of a partitioned range of studentized residuals
f_j(z_1, \ldots, z_K)  term of regression equation
H_0  statistical hypothesis to be tested
H_1  alternate hypothesis to be accepted if \( H_0 \) is judged to be false
J  number of coefficients estimated, excluding \( b_0 \)
K  number of independent variables observed
k  number of segments or cells in range of possible studentized residuals
LOF  lack of fit
M  total number of independent and dependent variables
MS(source)  mean square due to source, where source is REG, RES, etc.
m  moment about origin
N  number of observations
\( N(\mu,\sigma^2) \)  normal distribution with mean \( \mu \) and variance \( \sigma^2 \)
n_i  number of studentized residuals in \( i^{th} \) cell
\( p_i \)  
\( R \)  
\( \text{REG} \)  
\( \text{REP} \)  
\( \text{RES} \)  
\( r \)  
\( S \)  
\( S_c \)  
\( s_j \)  
\( \text{SSQ(source)} \)  
\( \text{TOT} \)  
\( t_n \)  
\( V(x) \)  
\( W, X \)  
\( x \)  
\( x(J) \)  
\( \bar{x}_{.j} \)  
\( y \)  
\( Z_i \)  
\( z_1, \ldots, z_K \)  
\( \mu_x \)  
\( \hat{\mu} \)  
\( \sigma_x^2 \)  
\( \hat{\sigma}^2 \)  
\( \chi^2 \)  
\( \sim \)  
\( ^\top \)  
\( 4 \)
ESTIMATION OF BASIC LINEAR MODEL

Basic Linear Model

In multiple linear regression, a dependent or response variable Y (such as temperature or pressure) measured on an object or experiment is assumed to be correlated with a function of one or more other variables \( z_1, \ldots, z_K \) measured on the same object or experiment. This function includes a number of unknown parameters \( b_1, \ldots, b_J \) and can be represented as

\[
y = h(b_1, \ldots, b_J, z_1, \ldots, z_K) + e
\]  \hspace{1cm} (1)

The only restriction imposed on this function is that it be linear in the parameters

\[
y = \sum_{j=1}^{J} b_j f_j(z_1, \ldots, z_K)
\]  \hspace{1cm} (2)

where \( f_j(z_1, \ldots, z_K) \) is a TERM of the regression equation. (A TERM is a quantity which may be a variable or a function of a variable, e.g., \( T \) is a TERM and \( Z \), after it is defined as \( Z = \log T \), is also a TERM.)

Suppose that there are \( N \) observations of the dependent variable. Let the subscript \( i \) indicate that the values are associated with the \( i^{th} \) observation; in particular, the value of the response variable \( y_i \) would depend on the observed values of the variables \( z_{i1}, \ldots, z_{iK} \). Also, let the subscript \( j \) denote the \( j^{th} \) term in the regression model so that \( x_{ij} = f_j(z_{i1}, \ldots, z_{iK}) \) describes the transformations of the \( (z_{i1}, \ldots, z_{iK}) \) to produce the value of \( x_{ij} \) for the \( j^{th} \) term at the \( i^{th} \) observation.

The regression model can now be rewritten as

\[
y_i = b_1 x_{i1} + b_2 x_{i2} + \ldots + b_J x_{iJ} + e_i \hspace{1cm} i = 1, \ldots, N
\]  \hspace{1cm} (3)

where \( e_i \) denotes the difference between the observed value and the expected value of \( y_i \). For the \( N \) observations, it is convenient to write this regression model in matrix notation as \( y = Xb + e \) where
More often than not, the analyst feels the model

$$y_i = b_0 + b_1 x_{i1} + \ldots + b_J x_{iJ} + e_i \quad i = 1, \ldots, N$$

is more appropriate. Let $a_0 = b_0 + b_1 \bar{x}_1 + \ldots + b_J \bar{x}_J$. Then, as a result of adding
this equation to, and subtracting it from, equation (5) and rearranging terms

\[ y_i = (b_0 + b_1 x_{i1} + \ldots + b_J x_{iJ}) \]

\[ + b_1(x_{i1} - \bar{x}_{i1}) + \ldots + b_j(x_{iJ} - \bar{x}_{iJ}) + e_i \quad i = 1, \ldots, N \] (6)

If then, a dummy variable \( x_0 \) is introduced such that for all values of \( i \), \( x_{i0} = 1.0 \), equation (6) may be written as

\[ y_i = a_0 x_{i0} + b_1(x_{i1} - \bar{x}_{i1}) + \ldots + b_j(x_{iJ} - \bar{x}_{iJ}) + e_i \quad i = 1, \ldots, N \] (6a)

Equation (6a) now resembles equation (3) and may be written in matrix notation, similar to equation (4), as \( y = Xb + e \) where now

\[
\begin{align*}
\begin{pmatrix}
y_1 \\
y_2 \\
\vdots \\
y_N \\
\end{pmatrix}
= 
\begin{pmatrix}
1.0 & x_{11} - \bar{x}_{11} & \ldots & x_{1J} - \bar{x}_{1J} \\
1.0 & x_{21} - \bar{x}_{11} & \ldots & x_{2J} - \bar{x}_{1J} \\
\vdots & \vdots & \ddots & \vdots \\
1.0 & x_{N1} - \bar{x}_{11} & \ldots & x_{NJ} - \bar{x}_{1J} \\
\end{pmatrix}
\begin{pmatrix}
a_0 \\
b_1 \\
\vdots \\
b_J \\
\end{pmatrix}
+ 
\begin{pmatrix}
e_1 \\
e_2 \\
\vdots \\
e_N \\
\end{pmatrix}
\end{align*}
\] (7)
Estimating $\hat{b}$

Equations (4) and (7) are similar in form and for $N > J$ are an overdetermined set of linear equations. There will be some vector $\hat{b}$ which is a "best" vector to use. If the vector $e$ is composed of random variables $e_i$ such that $E(e_i) = 0$, $\text{V}(e_i) = \sigma^2 < +\infty$, and the $e_i$ are uncorrelated, then as is well known, the method of least squares gives the linear minimum variance estimators $\hat{b}$ for $b$. And $\hat{b}$ is given by

$$\hat{b} = (X'X)^{-1} X'y$$

(8)

The matrix $X'X$ divided by $N - 1$ is called the variance-covariance matrix of the independent variables. The variance-covariance matrix of $\hat{b}$ is given by

$$\text{V}(\hat{b}) = \sigma^2 (X'X)^{-1}$$

(9)

It is important to note that when the form of equation (7) is used, $X'X$ is

$$X'X = \begin{pmatrix}
N & 0 & \ldots & 0 \\
0 & \sum_{i=1}^{N} (x_{i1} - \bar{x}_1)^2 & \ldots & \sum_{i=1}^{N} (x_{i1} - \bar{x}_1)(x_{iJ} - \bar{x}_J) \\
\vdots & \vdots & \ddots & \vdots \\
0 & \sum_{i=1}^{N} (x_{i1} - \bar{x}_1)(x_{iJ} - \bar{x}_J) & \ldots & \sum_{i=1}^{N} (x_{iJ} - \bar{x}_J)^2
\end{pmatrix}$$

(10)

This is seen to be symmetric and of the form

$$X'X = \begin{pmatrix}
A & 0 \\
0 & B
\end{pmatrix}$$
Hence,

$$(X'X)^{-1} = \begin{pmatrix} A^{-1} & 0 \\ 0 & B^{-1} \end{pmatrix}$$

RAPIER uses this relation to advantage by storing only the upper triangular part of $B$ and computing only the coefficients $b_1, \ldots, b_J$ by matrix manipulations. Then $b_0$ is given by the simple equation

$$b_0 = \bar{y} - \hat{b}_1 \bar{x}_1 - \hat{b}_2 \bar{x}_2 - \ldots - \hat{b}_J \bar{x}_J$$

where $\bar{y} = \sum y_i / N = \hat{a}_0$. It can also be shown that

$$V(\hat{b}_0) = V(\bar{y}) + V(\hat{b}' \bar{x}) = \left[ \frac{1}{N} + \bar{x}' (X'X)^{-1} \bar{x} \right] \sigma^2$$

$$\text{COV}(\hat{b}_0, \hat{b}) = -(X'X)^{-1} \bar{x} \sigma^2$$

When there is no $b_0$ term in the regression model,

$$X'X = \begin{bmatrix} \sum_{i=1}^{N} x_{i1}^2 & \sum_{i=1}^{N} x_{i1} x_{i2} & \ldots & \sum_{i=1}^{N} x_{i1} x_{iJ} \\ \sum_{i=1}^{N} x_{i1} x_{i2} & \sum_{i=1}^{N} x_{i2}^2 & \ldots & \sum_{i=1}^{N} x_{i2} x_{iJ} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{N} x_{i1} x_{iJ} & \sum_{i=1}^{N} x_{i2} x_{iJ} & \ldots & \sum_{i=1}^{N} x_{iJ}^2 \end{bmatrix}$$

(12)

Comparing this to equation (10) shows this form of $X'X$ to be similar to the lower right submatrix in equation (10). This similarity is used to simplify notation by assuming that $X'X$ represents either the form of equation (12) or the lower right portion of equation (10) and considering the calculation of $b_0$ as a special case. Thus, further reference to $b$ implies
Another matrix of interest both computationally and statistically is the correlation matrix $C$. The elements of $C$, which are denoted $C_{ij}$, are the sample correlation coefficients between the terms $X_i$ and $X_j$. These are

$$C_{ij} = \sum_{l=1}^{N} \frac{(x_{li} - \bar{x}_i)(x_{lj} - \bar{x}_j)}{\sqrt{\sum_{l=1}^{N} (x_{li} - \bar{x}_i)^2} \sqrt{\sum_{l=1}^{N} (x_{lj} - \bar{x}_j)^2}}$$  \hspace{1cm} (13)$$

and all these numbers are between 1.0 and -1.0.

The calculation of $C$ can be expressed in matrix notation conveniently by defining a diagonal matrix $S = \text{diag}(s_1, s_2, \ldots, s_J)$ with elements

$$s_j = \frac{1.0}{\sqrt{(X'X)_{jj}}} \hspace{1cm} j = 1, \ldots, J$$  \hspace{1cm} (14)$$

Then

$$C = S(X'X)S$$  \hspace{1cm} (15)$$

and

$$(X'X)^{-1} = S^{-1}(X'X)^{-1}S^{-1}S = SC^{-1}S$$  \hspace{1cm} (16)$$

The algorithm of RAPIER performs the following operations: (1) constructs the $X'X$ matrix, (2) computes $C$, (3) inverts $C$, (4) computes $(X'X)^{-1}$ from $C^{-1}$ by equation (16), and (5) computes the $b$ estimates. Because $C$ is a normalized matrix, the
inversion of \( C \) is likely to be more accurate than direct inversion of \( X'X \). Examination of the structure of \( X'X \) and/or \( C \) is of assistance in evaluating the possible numerical problems.

It may also be that the independent variables are random variables. Then \( X'X \) divided by \( N - 1 \) represents the variance-covariance matrix and \( C \) the sample correlation matrix. If the independent variables are considered to be from a multivariate distribution, it is useful in some cases to consider the eigenvalues and eigenvectors of \( X'X \) and/or \( C \).

For these reasons, RAPIER includes options to compute and print these quantities. As a partial check on the accuracy of the inversion process, it is also possible to have \( C \cdot C^{-1} \) computed and printed. This should be the identity matrix.

### Estimating \( \sigma^2 \)

For any regression model \( y = Xb + e \), there are possibly two methods of estimating \( \sigma^2 \). First, if the assumed regression model is, in reality, the true model, it is well known that an unbiased estimator is given by

\[
\sigma^2_{RES(J)} = \frac{y'y - b'X'y}{N - J - D}
\]

\[
= \frac{SSQ(RES)}{N - J - D}
\]

\[
= MS(RES(J))
\]

(17)

Second, where there are replicated data points, another estimator of \( \sigma^2 \), depending only on \( V(e_i) = \sigma^2 \) for all \( i \) and not on the correctness of the assumed model, is the pooled mean squares computed from the replicated data points.

Assume the observations are grouped into replicate sets in sequence. Let \( R \) be the number of sets of replicates and \( r_i \) be the number of replicates in the \( i \)th replicate set. Let

\[
SSQ(i) = \sum_{n=r^*+1}^{r^*+r_1} (y_n - \bar{y}_i)^2
\]

(18)
where

\[ r^* = \sum_{j=1}^{i-1} r_j \]

It is assumed \( y_n \) is from the \( i^{th} \) replicate set and \( \bar{y}_i \) is calculated only from those \( y_n \) in the \( i^{th} \) replicate set. Then define the pooled sum of squares due to replication as

\[ SSQ(REP) = \sum_{i=1}^{R} SSQ(i) \]

and the pooled degrees of freedom as

\[ NPDEG = \sum_{i=1}^{R} (r_i - 1) \].

The second estimator of \( \sigma^2 \) becomes

\[ \sigma^2_{REP} = \frac{SSQ(REP)}{NPDEG} = MS(REP) \] \hspace{1cm} (19)

It can be shown (ref. 2) that the sums of squares due to residuals can be partitioned into a component due to replication and a component due to lack of fit; that is,

\[ SSQ(RES) = SSQ(LOF) + SSQ(REP) \] \hspace{1cm} (20)

This partitioning is used later to determine the estimate of \( \sigma^2 \) to use in tests of hypotheses.

**HYPOTHESIS TESTING**

**Test NE - Normality of e**

As stated before, the only assumption necessary for \( \hat{\beta} \) to be a linear minimum variance estimator is that \( E(e_i) = 0 \), \( V(e_i) = \sigma^2 < +\infty \), and \( e_i \) be uncorrelated. If it can further be assumed that \( e_i \sim N(0, \sigma^2) \), a number of standard tests become available. RAPIER computes a chi-square statistic and the sample skewness and kurtosis for testing this hypothesis.

Under the hypothesis \( e_i \sim N(0, \sigma^2) \), the studentized residuals defined by

\[ Z_i = \frac{y_i - \hat{y}_i}{\hat{\sigma}} = \frac{e_i}{\hat{\sigma}} \]
will be distributed as Student's $t$ with the degrees of freedom associated with the estimate $\hat{\sigma}$. If the degree of freedom is 30 or more, the $t$ distribution is very close to the normal.

The range of possible studentized residuals is $(-\infty, +\infty)$ and may be divided into $k$ segments or cells each with probability $p_i$, so that each segment will have $Np_i$ as the expected number of observations falling into it. Let $n_i$ denote the number of studentized residuals in the $i$th cell. Then a chi-square goodness-of-fit statistic may be calculated as

$$
\chi^2_{(k-1)} = \sum_{i=1}^{k} \frac{(n_i - Np_i)^2}{Np_i}
$$

RAPIER computes this statistic by using an even number of cells greater than or equal to four and less than or equal to 20, such that the expected numbers of observations per cell is five or more. The bounding values for the $i$th cell are $Z_{i-1}$, $Z_i$ where $F(Z_i) = (i \cdot k)/N$ and $F(Z)$ is the cumulative normal distribution function. Then each cell has the same expected number of observations, say $f = N/k$. Then

$$
\chi^2_{(k-1)} = \sum_{i=1}^{k} \frac{(n_i - f)^2}{f} = \frac{k}{N} \sum_{i=1}^{k} n_i^2 - N
$$

This statistic is not computed for less than 30 degrees of freedom for the estimate $\hat{\sigma}^2$.

Two other statistics which may be used to test the normality of an empirical distribution are skewness and kurtosis. Define the moments about the origin as

$$
m_2 = \frac{1}{N} \sum Z_i^2
$$

$$
m_3 = \frac{1}{N} \sum Z_i^3
$$

$$
m_4 = \frac{1}{N} \sum Z_i^4
$$
where \( Z_i \) is the \( i^{th} \) studentized deviate. Then skewness is \( \text{RELSKW} = m_3^2 / m_2^3 \), which should be nearly zero. Kurtosis is \( \text{RELKUR} = m_4^2 / m_2^2 \), which should be nearly 3. Probability points for these are tabulated in reference 6.

If these statistics indicate nonnormality, there are three possible courses of action. First, perhaps a transformation of the response variable or the independent variables can be found which will bring the distribution of residuals closer to normal. RAPIER makes this task quite easy. Second, a different candidate model might be used (see ref. 2, "Analysis of Residuals"). As the last choice, it is possible to do nothing and simply rely on the robustness of the tests involved. See reference 4 for definition and discussion of robustness.

Also note that the individual observations may be weighted to perform a weighted regression analysis. RAPIER permits the use of weights (ref. 2). In this case, the \( X'X \) and \( X'y \) matrices take the form

\[
X'X = \begin{pmatrix}
\sum_{i=1}^{N} \left( (x_{i1} - \bar{x}_1)^2 w_i \right) & \ldots & \sum_{i=1}^{N} \left( w_i (k_{i1} - \bar{x}_1)(k_{ij} - \bar{x}_j) \right) \\
\vdots & \ddots & \vdots \\
\sum_{i=1}^{N} \left( (x_{ij} - \bar{x}_j)(x_{i1} - \bar{x}_1)w_i \right) & \ldots & \sum_{i=1}^{N} \left( (x_{ij} - \bar{x}_j)^2 w_i \right)
\end{pmatrix}
\]

\[
X'y = \begin{pmatrix}
\sum_{i=1}^{N} x_{i1} y_i w_i \\
\vdots \\
\sum_{i=1}^{N} x_{ij} y_i w_i
\end{pmatrix}
\]
Analysis of Variance Table

For most hypothesis testing of the regression model, it is convenient to summarize the available information in an Analysis of Variance (ANOVA) table, as follows:

<table>
<thead>
<tr>
<th>Source</th>
<th>Sums of squares</th>
<th>Degrees of freedom</th>
<th>Mean squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression (REG)</td>
<td>SSQ(REG) = ( \hat{b}'X'y - S_c )</td>
<td>( J )</td>
<td>MS(REG) = SSQ(REG)/J</td>
</tr>
<tr>
<td>Residual (RES)</td>
<td>SSQ(RES) = ( y'y - \hat{b}'X'y )</td>
<td>( N - J - D )</td>
<td>MS(RES) = SSQ(RES)/(N - J - D)</td>
</tr>
<tr>
<td>Total</td>
<td>SSQ(TOT) = ( y'y - S_c )</td>
<td>( N - D )</td>
<td></td>
</tr>
</tbody>
</table>

\( a \) if no \( b_0 \) coefficient is estimated.

\( S_c = \begin{cases} 0 & \text{if no } b_0 \text{ coefficient is estimated.} \\ \frac{N\gamma^2}{2} & \text{if a } b_0 \text{ coefficient is estimated.} \end{cases} \)

\( b \) if no \( b_0 \) coefficient is estimated.

\( D = \begin{cases} 0 & \text{if no } b_0 \text{ coefficient is estimated.} \\ 1 & \text{if } b_0 \text{ is estimated.} \end{cases} \)

If there are replicated data points, another ANOVA table can be constructed to show the separation of the residual sums of squares into components from lack of fit and replication, as in the following table:

<table>
<thead>
<tr>
<th>Source</th>
<th>Sums of squares</th>
<th>Degrees of freedom</th>
<th>Mean squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lack of fit (LOF)</td>
<td>SSQ(LOF) = SSQ(RES) - SSQ(REP)</td>
<td>( N - J - D - NPDEG )</td>
<td>MS(LOF) = SSQ(LOF)/(N - J - D - NPDEG)</td>
</tr>
<tr>
<td>Replication (REP)</td>
<td>SSQ(REP)</td>
<td>NPDEG</td>
<td>MS(REP) = SSQ(REP)/NPDEG</td>
</tr>
<tr>
<td>Residual (RES)</td>
<td>( y'y - \hat{b}'X'y )</td>
<td>( N - J - D )</td>
<td></td>
</tr>
</tbody>
</table>

NPDEG
Choice of Estimator for \( \sigma^2 \)

As mentioned previously, there are two possible methods of estimating \( \sigma^2 \) depending on whether there are replicated data points. This is true for any given model equation. When the backward rejection option of RAPIER is used, there is no longer one hypothetical model but a series of models. Thus, there is the choice of estimator for \( \sigma^2 \) to be made after each rejection of a term in the previous model.

As an example, consider the model

\[
y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + e
\]  

(22)

with replicated data points. The first step is to estimate \( b_0, b_1, b_2, \) and \( b_3 \). There will then be the estimators \( \hat{\sigma}^2_{\text{RES}(J)} \) and \( \hat{\sigma}^2_{\text{REP}} \). If the model in equation (22) has not left out any significant terms, both estimators are valid.

The ratio \( F = \frac{\text{MS(LOF)}}{\text{MS(REP)}} \) can be used to test the hypothesis that there is no lack of fit, where \( F \sim F_{a,b} \) with \( a = N - J - D - \text{NPDEG} \) and \( b = \text{NPDEG} \) degrees of freedom. If the test accepts the hypothesis of no lack of fit, \( \text{MS(RES)} \) is a pooled estimate of \( \sigma^2 \) with more degrees of freedom and will usually make tests using \( \hat{\sigma}^2_{\text{RES}(J)} \) more sensitive than those using \( \hat{\sigma}^2_{\text{REP}} \). But there is the possibility that the hypothesis was accepted as a result of random fluctuation when there really is some lack of fit; that is, there is the possibility that \( \hat{\sigma}^2_{\text{RES}(J)} \) is a biased estimator. If lack of fit is not concluded to be significant, the decision to pool or not is usually made on the basis of the number of degrees of freedom for replication. If this is "large" (no definition of large is given herein), \( \hat{\sigma}^2_{\text{REP}} \) is used. If "small," the pooled estimate \( \hat{\sigma}^2_{\text{RES}(J)} \) is used.

In testing equation (22), should it be decided that \( b_3 \) is not significantly different from zero (see section Test TT - \( t \)-Tests), the coefficients of the following model would be estimated:

\[
y = b_0 + b_1 x_1 + b_2 x_2 + e
\]

From this model there is an estimate \( \hat{\sigma}^2_{\text{RES}(J-1)} \). This estimate could also be biased since \( b_3 \) may be small but nonzero and the decision of \( b_3 = 0 \) may have been due to random fluctuation.

At the first step, the lack of fit can be considered a random sample of an infinite possibility of biases. But the biases due to pooling mean squares after rejecting terms can be considered to be systematic biasing and hence less desirable.
RAPIER provides three strategies of pooling estimates for use in the decision procedure:

1. Never pool. This is appropriate only when there are replicated data points. The estimator used in all t-tests is $\hat{\sigma}_{\text{REP}}^2$.

2. Always pool initial residual. This will always pool the lack of fit and replication from the first model only. Additional mean squares due to rejected terms will be ignored.

3. Always pool. This strategy will always use $\hat{\sigma}_{\text{RES}(J-i)}^2$ for the model with $i$ rejected terms.

A rule for pooling lack of fit and replication mean squares is discussed by Draper and Smith (ref. 2). Related work as applied to factorial designs is presented by Holms (ref. 7) and Bozivich, Bancroft, and Hartley (ref. 8).

Test OR - Overall Regression

One of the first tests usually applied to a regression model is the test of the overall significance of the model. In the notation of hypothesis testing this is stated $H_0: b = 0$; $H_1: b \neq 0$ where

$$b = \begin{pmatrix} b_1 \\ \vdots \\ \vdots \\ b_J \end{pmatrix}$$

The statistic for this test is $F = \text{MS(REG)}/\hat{\sigma}^2$. Then $F \sim F_{a, b}$ with $a = J - D$, and $b$ equals the degrees of freedom associated with $\hat{\sigma}^2$.

Another useful statistic for judging the significance of overall regression is $R^2 = \text{SSQ(REG)}/\text{SSQ(TOT)}$. The sampling distribution of $R$ does not lend itself to very simple tests except in the case of $H_0: R = 0$. The main value of $R^2$ is that it must be a number in the range 0 to 1 and 100 $R^2$ is a measure of the percentage of variation in the $y$ values that is accounted for by the regression model.

Test SF - Sequential F-Test

There is often reason to consider a partitioned form of $b' = \{w_1', w_2'\}$ for testing the hypotheses.
Partition the matrix $X$ corresponding to the partitioning of $b$ and denote it as $X = (W_1, W_2)$ where $W_1$ is $N \times p$ and $W_2$ is $N \times (J - p)$. Then the test statistic is $F = \frac{(SSQ(\text{REG})(p)/p) \hat{\gamma}_{\text{RES}(J)}}{\hat{\gamma}_{\text{RES}(J)}}$, where $SSQ(\text{REG})(p) = \hat{\gamma}_1 W_1' y$ and $\hat{\gamma}_1 = (W_1' W_1)^{-1} W_1' y$. Then $F \sim F_{a, b}$ with $a = p$ and $b = N - p - D$. Sometimes this test is performed with $p = 1, p = 2, \ldots, p = J$. This is then referred to (ref. 2) as the sequential F-test. RAPIER computes regressions for $p = 1, \ldots, p = J$ upon request.

Test II - t-Tests

In many cases, the regression model contains terms whose estimated coefficients are "small." This may be an indication that the term does not have a real effect on the dependent variable and that the coefficient is nonzero due to random sampling variation. If this is true, it is desirable to delete the term from the regression model. A test statistic for deciding this is

$$ t = \frac{\hat{b}_i}{\hat{\sigma}^2 (X'X)_{ii}^{-1}} $$

where $(X'X)_{ii}^{-1}$ denotes the $i^{th}$ diagonal element of the $(X'X)^{-1}$ matrix. The statistic $t \sim t_{N-J-D}$. An equivalent test statistic is

$$ F = t^2 = \frac{\hat{b}_i^2}{\hat{\sigma}^2 (X'X)_{ii}^{-1}} $$

where $F \sim F_{1, N-J-D}$. This is often referred to (ref. 2) as the partial F-test. The quantity $\frac{\hat{b}_i}{\sqrt{(X'X)_{ii}^{-1}}}$ is called the sum of squares due to $b_i$, if $x_i$ were last to enter the equation. RAPIER computes and prints the t-statistics and the probability associated with the interval $(-t, t)$.

This particular test is the basis for the rejection option of RAPIER. The analyst has chosen which $\hat{\sigma}^2$ estimator to use by the choice of strategy. Then the analyst may choose a significance level which all coefficients must meet. For example, suppose a
significance level of 0.900 is chosen. The t-statistic is then computed for each coefficient, and the coefficient with minimum \(|t|\) is identified. If \(\min|t| > t_{N-J-D, 0.950}\), all terms are concluded to be significant at the 0.900 (or 90.0 percent) level of significance. If \(\min|t| < t_{N-J-D, 0.950}\), the term corresponding to the minimum \(|t|\) is dropped from the hypothetical model, and the regression is recomputed. This process is repeated until all remaining coefficients are significant at the specified level of probability.

PREDICTING VALUES FROM ESTIMATED REGRESSION EQUATION

Regression equations are often used to predict an estimated response at some condition of the independent variables. Useful estimates of parameters to know are the variance of the regression equation and the variance of a single further observation at the desired combination of the independent variables.

Let \(x' = (x_1, \ldots, x_J)\) denote the vector of independent variables at which a prediction is desired. Let \(x^* = x - \overline{x}\). Let \(\hat{\sigma}^2_{\mu \cdot x}\) denote the estimated variance of the regression equation at \(x\). Let \(\hat{\sigma}^2_{y \cdot x}\) denote the estimated variance of a single further observation at \(x\). Then,

\[
\hat{\sigma}^2_{\mu \cdot x} = \hat{\sigma}^2 \left[ \frac{D}{N} + x^* (X'X)^{-1} x^* \right]
\]

(25)

\[
\hat{\sigma}^2_{y \cdot x} = \hat{\sigma}^2 \left[ 1.0 + \frac{D}{N} + x^* (X'X)^{-1} x^* \right]
\]

(26)

where, as before, \(D = 1\) if a \(b_0\) coefficient is estimated and \(D = 0\) if a \(b_0\) coefficient is not estimated. The quantity \(s = \hat{\sigma}_{\text{RES}(J)}\) is called the standard error of estimate and often is used as a simple approximation to \(\hat{\sigma}_{y \cdot x}\). This approximation is close if \(N\) is very large and \(x = \overline{x}\), in which case,

\[
\hat{\sigma}_{y \cdot x} = s \left( 1.0 + \frac{D}{N} \right) \approx s
\]

When \(x \neq \overline{x}\), this may be a poor approximation. RAPIER accepts input vectors \(x\) and computes \(\hat{y} = \hat{b}_0 + \hat{b}_1 x_1 + \ldots + \hat{b}_J x_J\), as well as \(\hat{\sigma}^2_{\mu \cdot x}, \hat{\sigma}^2_{\mu \cdot x'}, \hat{\sigma}^2_{y \cdot x}, \hat{\sigma}^2_{y \cdot x'}, \hat{\sigma}_{y \cdot x'}, \hat{\sigma}_{y \cdot x'}\), and the standard error of estimate.
USER'S GUIDE TO INPUT

Sample Regression Problem

Let

\[ x_1 \] temperature

\[ x_2 \] time

\[ x_3 \] pressure

\[ y_1 \] output, lb

\[ y_2 \] cost of operation

The data are coded into standardized units, as is often done in experimental design analysis. The \( y_1 \) variable is assumed to be of primary interest in this problem.

Table I contains the \( x \) and \( y \) data. Table II contains a summary of the type of in-

<table>
<thead>
<tr>
<th>Group</th>
<th>( x_1 )</th>
<th>( x_2 )</th>
<th>( x_3 )</th>
<th>( y_1 )</th>
<th>( y_2 )</th>
<th>Group</th>
<th>( x_1 )</th>
<th>( x_2 )</th>
<th>( x_3 )</th>
<th>( y_1 )</th>
<th>( y_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>9.17</td>
<td>38.9</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9.61</td>
<td>49.1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>12.76</td>
<td>42.1</td>
<td>10</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>10.01</td>
<td>49.3</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>12.97</td>
<td>44.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10.12</td>
<td>50.1</td>
</tr>
<tr>
<td></td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>8.96</td>
<td>58.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.95</td>
<td>51.8</td>
</tr>
<tr>
<td>3</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>9.11</td>
<td>58.3</td>
<td>9</td>
<td>-2</td>
<td>0</td>
<td>0</td>
<td>11.78</td>
<td>50.1</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>8.96</td>
<td>58.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22.90</td>
<td>58.1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>17.03</td>
<td>63.2</td>
<td>10</td>
<td>0</td>
<td>-2</td>
<td>0</td>
<td>7.99</td>
<td>28.6</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>9.05</td>
<td>38.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22.90</td>
<td>58.1</td>
</tr>
<tr>
<td></td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>8.86</td>
<td>39.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11.70</td>
<td>70.2</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>12.60</td>
<td>44.1</td>
<td>12</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>12.11</td>
<td>71.0</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>13.21</td>
<td>43.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11.70</td>
<td>70.2</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>17.20</td>
<td>62.1</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>-2</td>
<td>10.11</td>
<td>48.7</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>17.04</td>
<td>62.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10.02</td>
<td>56.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type of input</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Identification</td>
</tr>
<tr>
<td>2</td>
<td>Definition of problem size</td>
</tr>
<tr>
<td>3</td>
<td>Definition of problem logic</td>
</tr>
<tr>
<td>4</td>
<td>Terms, transformations, and constants</td>
</tr>
<tr>
<td>5</td>
<td>Control rejection option</td>
</tr>
<tr>
<td>6</td>
<td>Provide replication information</td>
</tr>
<tr>
<td>7</td>
<td>Data input unit and format</td>
</tr>
<tr>
<td>8</td>
<td>Data</td>
</tr>
<tr>
<td>9</td>
<td>Prediction information</td>
</tr>
</tbody>
</table>
### Figure 1 - Sample Input Form

<table>
<thead>
<tr>
<th>SAMPLE INPUT (Concluded)</th>
<th>FORTRAN STATEMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 26 0.0</td>
<td></td>
</tr>
<tr>
<td>27 1 -1</td>
<td>9.17 -59.5</td>
</tr>
<tr>
<td>28 1 -1</td>
<td>12.76 45.1</td>
</tr>
<tr>
<td>29 1 -1</td>
<td>12.97 44.0</td>
</tr>
<tr>
<td>30 1 1</td>
<td>9.11 58.3</td>
</tr>
<tr>
<td>31 1 1</td>
<td>-8.86 58.7</td>
</tr>
<tr>
<td>32 0 0</td>
<td>10.11 48.7</td>
</tr>
<tr>
<td>33 0 0</td>
<td>10.01 49.9</td>
</tr>
<tr>
<td>34 0 0</td>
<td>10.02 50.8</td>
</tr>
</tbody>
</table>

**Notes:**
- **Type Line:** Indicates the type of the line in the FORTRAN statement.
- **Sample Input (Concluded):** Additional sample input data.
- **FORTRAN Statement:** Code examples for the input data.
put cards and their basic functions. Figure 1 shows a sample set of data for a complete regression as it would be written on a FORTRAN coding sheet.

Use the model equations

\[ y_i = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{23} x_2 x_3 + b_{11} x_1^2 + b_{22} x_2^2 + b_{33} x_3^2 + e \quad i = 1, 2 \] (27)

Test whether the interaction terms as a group are significant, given that the linear terms are in the model. Predict the response at the point \((x_1, x_2, x_3) = (-1, +1, +1)\).

**Types of Input Cards**

Nine types of data card may be used to define a regression analysis for RAPIER. A summary of the types and their functions is presented in table II.

**Type 1.** - In type 1 input, as many as 100 cards with Hollerith information may be read to identify and describe the problem. At least one card is read. The first two columns of the first card used to specify the additional number of identification cards to be read, and columns 3 to 80 are used for Hollerith information. Each following card uses columns 1 to 78. (See lines 1 to 16 in fig. 1.)

**Type 2.** - In type 2 input, one card with three four-column fields followed by a five-column field specifies

1. Number of independent variables to read
2. Number of dependent variables to read
3. Number of terms in the model equation (not counting \(b_0\))
4. Number of observations

(See line 17 in fig. 1.)

**Type 3.** - In type 3 input, one card with 10 one-column fields specifies

1. The \(b_0\) term in the model equation (T or F)
2. Computation of t-statistics and their confidence levels (T or F)
3. Weighting factor either of 1.0 (T) or supplied with each data point (F)
4. Computation of residuals and chi-square test (T or F)
5. Computation of eigenvalues and eigenvectors of correlation matrix (T or F)
6. Computation of eigenvalues and eigenvectors of \(X'X\) (T or F)
7. Computation of product of correlation matrix and its inverse (T or F)
8. Use of bordering inversion technique for computation of sequential regression (T or F); see Test SF - Sequential F-Test

22
(9) Use of an economy version of output which does not print the matrices $X'X$, $(X'X)^{-1}$, $x'y$, $C$, or $C^{-1}$ (T or F) (If item 7 of this set is set T, then $C \cdot C^{-1}$ is printed.)

(10) The pooling strategy:

1. Never pool. Always use replication error. (If there is no replication, the program sets this to 3.)
2. Pool initial residual.
3. Pool all residuals.

(See line 18 in fig. 1.)

Type 4, type 4A, type 4B, and type 4C. - Type 4 card has two four-column fields specifying

1. Number of transformations
2. Number of constants

(See line 19 of fig. 1.)

If the number of transformations is zero, and therefore the number of constants is zero, the type 4A, type 4B, and type 4C cards are not expected by the program. In this case the program assumes the independent and the dependent variables are arranged on the input cards as

$$x_1, x_2, \ldots, x_J, y_1, \ldots, y_{\text{NODEP}}$$

where NODEP is the number of dependent variables.

If a weighting factor other than 1.0 is to be used (i.e., if item 3 of the type 3 card contains an F), the value of the weighting factor for each observation must appear as the last item in the list, so that in this case the data for one observation is entered on the cards as

$$x_1, x_2, \ldots, x_J, y_1, \ldots, y_{\text{NODEP}}, \text{WT}$$

For each observation, RAPIER reads a total of $M$ numbers, where $M$ is the sum of the number of independent and dependent variables. These numbers are stored consecutively in an array called VAR, beginning with location 01 and ending with location $M$. If the weighting factor is not identically 1.0, then $M + 1$ numbers are read, but the last number, being the weighting factor, is treated and stored separately. The data in VAR are used with appropriate weighting factors to cumulatively create $X'X$ and $X'y$ as shown in equations (21).

When a more complex model is desired, information must be supplied instructing the program as to (1) where to find the values for the TERMS of the equation, (2) how to
create the TERMS from the variables and constants, and (3) what the values are for the constant terms. This can be achieved easily by use of the type 4A(TERMS), type 4B(TRANSFORMATIONS), and type 4C(CONSTANTS) control cards (see p. 29). These three types and their functions can best be described by considering the sample model given by equation (27) as an example which illustrates their application.

An array called CON has a twofold purpose. First, if the number of constants designated in the second field of the type 4 card is nonzero, that number of constants will be read from the type 4C card and stored consecutively in this array beginning with location 01. If the number of constants is zero, the type 4C card is not expected by the program. Second, all the intermediate and final results of transformations are also stored in the CON array as the program obeys the instructions of the type 4B cards. The type 4A card must identify the relative location in the CON array where the value for each TERM is to be found for constructing the X'X and X'y matrices.

The VAR and CON arrays for this example are illustrated in figure 2. Five numbers are read for each observation: \( x_1, x_2, x_3, y_1, \) and \( y_2 \). These numbers automatically enter the VAR array beginning with location 01. Using transformation codes packed in fields of eight columns each on the type 4B cards, the program stores the result of each transformation into the appropriate relative location in the CON array as designated by

![Figure 2. Map of VAR and CON arrays. Data transferred into any location of CON array beyond location 60 are immediately duplicated in same relative location in VAR array.](image)
the last two digits of the field. Each transformation code is made up of four subfields of two card columns each, with the following interpretation:

<table>
<thead>
<tr>
<th>Subfield</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Operand VI</td>
</tr>
<tr>
<td>2</td>
<td>Operation OP</td>
</tr>
<tr>
<td>3</td>
<td>Operator CI</td>
</tr>
<tr>
<td>4</td>
<td>Result CS</td>
</tr>
</tbody>
</table>

Thus, subfield 1 always references the VAR array and subfields 3 and 4 refer to the CON array. The result of every transformation is a term which is stored in the designated location of the CON array, with the added feature that if the term is stored in relative location 61 or beyond, it is also stored in the VAR array. This is illustrated by the arrows in figure 2. This feature allows successive transformations to be performed more easily. The OP (operation codes) are tabulated in table III.

<table>
<thead>
<tr>
<th>Operation code (OP)</th>
<th>Resulting operation</th>
<th>Operation code (OP)</th>
<th>Resulting operation</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>No operation</td>
<td>01</td>
<td>VAR + CONST</td>
</tr>
<tr>
<td>01</td>
<td>VAR + CONST</td>
<td>02</td>
<td>VAR*CONST</td>
</tr>
<tr>
<td>02</td>
<td>VAR*CONST</td>
<td>03</td>
<td>CONST/VAR</td>
</tr>
<tr>
<td>03</td>
<td>CONST/VAR</td>
<td>04</td>
<td>EXP(VAR)</td>
</tr>
<tr>
<td>04</td>
<td>EXP(VAR)</td>
<td>05</td>
<td>VAR**CONST</td>
</tr>
<tr>
<td>05</td>
<td>VAR**CONST</td>
<td>06</td>
<td>A LOG(VAR)</td>
</tr>
<tr>
<td>06</td>
<td>A LOG(VAR)</td>
<td>07</td>
<td>A LOG10(VAR)</td>
</tr>
<tr>
<td>07</td>
<td>A LOG10(VAR)</td>
<td>08</td>
<td>SIN(VAR)</td>
</tr>
<tr>
<td>08</td>
<td>SIN(VAR)</td>
<td>09</td>
<td>COS(VAR)</td>
</tr>
<tr>
<td>09</td>
<td>COS(VAR)</td>
<td>10</td>
<td>SIN(π<em>CONST</em>VAR)</td>
</tr>
<tr>
<td>10</td>
<td>SIN(π<em>CONST</em>VAR)</td>
<td>11</td>
<td>COS(π<em>CONST</em>VAR)</td>
</tr>
<tr>
<td>11</td>
<td>COS(π<em>CONST</em>VAR)</td>
<td>12</td>
<td>1.0/VAR</td>
</tr>
<tr>
<td>12</td>
<td>1.0/VAR</td>
<td>13</td>
<td>EXP(CONST/VAR)</td>
</tr>
<tr>
<td>13</td>
<td>EXP(CONST/VAR)</td>
<td>14</td>
<td>EXP(CONST/VAR**2)</td>
</tr>
<tr>
<td>14</td>
<td>EXP(CONST/VAR**2)</td>
<td>15</td>
<td>SQRT(VAR)</td>
</tr>
<tr>
<td>15</td>
<td>SQRT(VAR)</td>
<td>16</td>
<td>1.0/SQRT(VAR)</td>
</tr>
<tr>
<td>16</td>
<td>1.0/SQRT(VAR)</td>
<td>17</td>
<td>CONST**VAR</td>
</tr>
<tr>
<td>17</td>
<td>CONST**VAR</td>
<td>18</td>
<td>10.0**VAR</td>
</tr>
<tr>
<td>18</td>
<td>10.0**VAR</td>
<td>19</td>
<td>SINH(VAR)</td>
</tr>
<tr>
<td>19</td>
<td>SINH(VAR)</td>
<td>20</td>
<td>COSH(VAR)</td>
</tr>
<tr>
<td>20</td>
<td>COSH(VAR)</td>
<td>21</td>
<td>(1.0-COS(VAR))/2.0</td>
</tr>
<tr>
<td>21</td>
<td>(1.0-COS(VAR))/2.0</td>
<td>22</td>
<td>ATAN(VAR)</td>
</tr>
<tr>
<td>22</td>
<td>ATAN(VAR)</td>
<td>23</td>
<td>ATAN2(VAR/CONST)</td>
</tr>
<tr>
<td>23</td>
<td>ATAN2(VAR/CONST)</td>
<td>24</td>
<td>VAR**2</td>
</tr>
<tr>
<td>24</td>
<td>VAR**2</td>
<td>25</td>
<td>VAR**3</td>
</tr>
<tr>
<td>25</td>
<td>VAR**3</td>
<td>26</td>
<td>ARCSIN(SQRT(VAR))</td>
</tr>
<tr>
<td>26</td>
<td>ARCSIN(SQRT(VAR))</td>
<td>27</td>
<td>2.0<em>π</em>VAR</td>
</tr>
<tr>
<td>27</td>
<td>2.0<em>π</em>VAR</td>
<td>28</td>
<td>1.0/(2.0<em>π</em>VAR)</td>
</tr>
<tr>
<td>28</td>
<td>1.0/(2.0<em>π</em>VAR)</td>
<td>29</td>
<td>ERF(VAR)</td>
</tr>
<tr>
<td>29</td>
<td>ERF(VAR)</td>
<td>30</td>
<td>GAMMA(VAR)</td>
</tr>
<tr>
<td>30</td>
<td>GAMMA(VAR)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All function names and operations are consistent with FORTRAN IV mathematical subroutines.
Table IV shows the sequence of transformations used to construct the terms of the example in equation (27). (See lines 21 and 22 of fig. 1.)

The arrays VAR and CON are shown in figure 3 both before and after one set of transformations performed on an observation. Note that CON now contains $x_1$, $x_2$, $x_3$, $x_1x_2$, $x_1x_3$, $x_2x_3$, $x_1^2$, $x_2^2$, $x_3^2$, $y_1$, $y_2$ along with unused locations. It may be that not all of these quantities are needed to express the model equation. The type 4A(TERMS) card must be used to supply the locations of CON which contain the terms of the model equation. (Note that this allows the user to somewhat arbitrarily assign terms to locations in CON.) The terms identifying the independent variables must be first, and the terms identifying the dependent variables last, just as in the assumed convention when no transformations are performed. Thus, in this case the terms needed are, according to equation (27),

$$x_1 \ x_2 \ x_3 \ x_1x_2 \ x_1x_3 \ x_2x_3 \ x_1^2 \ x_2^2 \ x_3^2 \ y_1 \ y_2$$

(See line 20 of fig. 1.)

After the set of transformations has been performed on an observation, the contents of the relative locations of the CON array specified on the terms card are transferred

<table>
<thead>
<tr>
<th>Transformation number</th>
<th>VI</th>
<th>OP</th>
<th>CI</th>
<th>CS</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>01</td>
<td>00</td>
<td>00</td>
<td>11</td>
<td>$x_1 \rightarrow$ CON(11)</td>
</tr>
<tr>
<td>2</td>
<td>01</td>
<td>00</td>
<td>00</td>
<td>61</td>
<td>$x_1 \rightarrow$ VAR(61), CON(61)</td>
</tr>
<tr>
<td>3</td>
<td>02</td>
<td>00</td>
<td>00</td>
<td>12</td>
<td>$x_2 \rightarrow$ CON(12)</td>
</tr>
<tr>
<td>4</td>
<td>02</td>
<td>00</td>
<td>00</td>
<td>62</td>
<td>$x_2 \rightarrow$ VAR(62), CON(62)</td>
</tr>
<tr>
<td>5</td>
<td>03</td>
<td>00</td>
<td>00</td>
<td>13</td>
<td>$x_3 \rightarrow$ CON(13)</td>
</tr>
<tr>
<td>6</td>
<td>03</td>
<td>00</td>
<td>00</td>
<td>63</td>
<td>$x_3 \rightarrow$ VAR(63), CON(63)</td>
</tr>
<tr>
<td>7</td>
<td>61</td>
<td>02</td>
<td>61</td>
<td>17</td>
<td>$x_1^2 \rightarrow$ CON(17)</td>
</tr>
<tr>
<td>8</td>
<td>62</td>
<td>02</td>
<td>62</td>
<td>18</td>
<td>$x_2^2 \rightarrow$ CON(18)</td>
</tr>
<tr>
<td>9</td>
<td>63</td>
<td>02</td>
<td>63</td>
<td>19</td>
<td>$x_3^2 \rightarrow$ CON(19)</td>
</tr>
<tr>
<td>10</td>
<td>61</td>
<td>02</td>
<td>62</td>
<td>14</td>
<td>$x_1x_2 \rightarrow$ CON(14)</td>
</tr>
<tr>
<td>11</td>
<td>61</td>
<td>02</td>
<td>63</td>
<td>15</td>
<td>$x_1x_3 \rightarrow$ CON(15)</td>
</tr>
<tr>
<td>12</td>
<td>62</td>
<td>02</td>
<td>63</td>
<td>16</td>
<td>$x_2x_3 \rightarrow$ CON(16)</td>
</tr>
<tr>
<td>13</td>
<td>04</td>
<td>00</td>
<td>00</td>
<td>20</td>
<td>$y_1 \rightarrow$ CON(20)</td>
</tr>
<tr>
<td>14</td>
<td>05</td>
<td>00</td>
<td>00</td>
<td>21</td>
<td>$y_2 \rightarrow$ CON(21)</td>
</tr>
</tbody>
</table>
Figure 3. - Arrays VAR and CON before and after transformations and terms selection, for the first example.
back to VAR in consecutive locations beginning with location 01. Thus, for this example, after selection of proper terms, the VAR array contains

\[ x_1, x_2, x_3, x_1^2 x_3, x_1 x_3^2, x_2^2, x_3^2, y_1, y_2 \]

in consecutive locations as required by equation (27) and the convention on independent and dependent variables. The \( b_0 \) term is accounted for by setting the proper item of type 3 input to \( T \).

There are three important facts to note concerning types 4A, 4B, and 4C. First, the transformation with \( OP = 00 \) is an identity transformation which simply transfers data from VAR to CON. If transformations are desired at all, a minimum requirement is that all variables at least be moved to CON so that the selection of terms will have a number to move back to VAR. Second, constants used in the transformations are stored in CON in locations beginning with 01 in sequence. Suppose there are NC constants initially supplied. Then if a transformation has any of the locations 01 through NC referenced in subfield 4, a term will replace the constant. Third, the use of TRANSFORMATIONS and TERMS overrides the convention that independent variables must precede dependent variables on the input data cards. The convention holds true for the terms card data in this case.

The sequence of input and formats is as follows:

1. Type 4A(TERMS): One or more cards, as necessary, with two-column fields denoting the relative locations of the CON array containing the final terms to be used in regression model (Up to 60 independent and nine dependent terms may be supplied. See line 20 of fig. 1.)

2. Type 4B(TRANSFORMATIONS): As many cards as necessary, with 10 transformations per card, each transformation being composed of four two-column subfields (A maximum of 100 transformations may be performed. See lines 21 and 22 of fig. 1.)

3. Type 4C(CONSTANTS): As many cards as necessary, containing the required number of constants in \( 5E15.7 \) format. As many as 60 constants may be supplied.

A second example is shown to illustrate the flexibility afforded by the TERMS, TRANSFORMATIONS, and CONSTANTS. Consider a model with two independent and two dependent variables with four terms given by

\[ b_0 + b_1 x_1 + b_2 x_2 + b_3 x_1 x_2 + b_4 x_1^3 = y_1 = \frac{k}{y_2} \]
where \( k = 1.0 \). Then a sequence of transformations which could be used is

<table>
<thead>
<tr>
<th>VI</th>
<th>OP</th>
<th>CI</th>
<th>CS</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>00</td>
<td>00</td>
<td>61</td>
<td>( x_1 = \text{CON}(61), \text{VAR}(61) )</td>
</tr>
<tr>
<td>01</td>
<td>02</td>
<td>61</td>
<td>62</td>
<td>( x_2 = \text{CON}(62), \text{VAR}(62) )</td>
</tr>
<tr>
<td>61</td>
<td>02</td>
<td>62</td>
<td>63</td>
<td>( x_1 = \text{CON}(63), \text{VAR}(63) )</td>
</tr>
<tr>
<td>01</td>
<td>00</td>
<td>00</td>
<td>14</td>
<td>( x_1 = \text{CON}(14) )</td>
</tr>
<tr>
<td>02</td>
<td>00</td>
<td>00</td>
<td>02</td>
<td>( x_2 = \text{CON}(02) )</td>
</tr>
<tr>
<td>02</td>
<td>02</td>
<td>61</td>
<td>12</td>
<td>( x_1x_2 = \text{CON}(12) )</td>
</tr>
<tr>
<td>03</td>
<td>00</td>
<td>00</td>
<td>11</td>
<td>( y_1 = \text{CON}(11) )</td>
</tr>
<tr>
<td>04</td>
<td>12</td>
<td>01</td>
<td>09</td>
<td>( k = \text{CON}(09) )</td>
</tr>
</tbody>
</table>

The TERMS information should then be

\[
x_1 \quad x_2 \quad x_1x_2 \quad x_1^3 \quad y_1 \quad \frac{k}{y_2}
\]

The total process is illustrated by figure 4 and the type 4 input given in figure 5.

**Type 5.** - In type 5 input, one card with one one-column field and one three-column field specifies

1. Use of \( t \)-statistics to reject insignificant terms (T or F)
2. Probability level that \( t \)-statistic must meet to be considered significant (This is written without a decimal point; e.g., 95-percent significance level is supplied as 950, 99.9 percent as 999, etc. See line 23 of fig. 1.)

**Type 6.** - In type 6 input, one card with one one-column field specifies that the data contain replicated points (T or F). If there are replicated data points, as many cards as are needed are read, containing 20 four-column fields specifying

1. The number of replicate sets
2. The number of replicates in each replicate set

Note that it is not safe for the program to assume that all data points with the same levels of the independent variables are true replicates. For this reason, the user must arrange replicate sets. RAPIER does check that all independent terms are the same within a replicate set. If not, the program stops. A nonreplicated data point is considered to be a group of size 1. Note that the data in table I are grouped to clearly indicate the repli-
Figure 4. - Arrays VAR and CON before and after transformations and terms selection, for the second example.
cated data points. There are 14 such groups. Thus, the first field of the second type 6 card contains a 14, and the remaining fields contain the count of the replicates in each group. (See lines 24 and 25 of fig. 1.)

Type 7. - In type 7 input, one card with one two-column field specifies the input unit number the data will be on. The remainder of the card is used to supply the format in which the data will be supplied. Note that if a weighting factor other than 1.0 is to be used, it will be read with each data point, and the format must allow for this. The current example uses a weighting factor of 1.0.

The format is (5F6.0) since there are three independent variables \((x_1, x_2, x_3)\) and two dependent variables \((y_1, y_2)\). If a weighting factor other than 1.0 is used, it must appear with every data point, and the format could, for example, be (5F6.0, F10.3). (See line 26 of fig. 1.)

Type 8. - Type 8 input consists of the input variables. Each observation consisting of the given \(x\)'s and \(y\)'s is read by execution of one READ statement. Thus, there will be at least one card for each observation. If the transformation option is not used, the program expects the first variables read to be the independent variables and the last ones to be the dependent variables; that is, the data must be arranged as

\[ x_1, x_2, \ldots, x_J, y_1, y_2, \ldots, y_{\text{NODEP}} \]
Otherwise, if transformations are used, appropriate use of the terms card information allows for more flexibility of input. (See lines 27 to 50 of fig. 1.)

Type 9. - In type 9 input, one card with one column is used to indicate if predictions are desired (T or F). (See line 51 of fig. 1.) If this is false, a new case is started. If it is true, the following cards are read: One card with one four-column field specifies the number of predictions desired. This is followed by cards with the values of the independent variables at which predictions are desired. Only the final regression model is used, but the number of independent and dependent variables originally supplied on the type 2 data cards are read. All transformations in type 4 input are performed. Then the proper terms are chosen by the program to correspond to the final model. The dependent variables are not needed and, hence, may be left off the data card unless one of the transformations of the dependent variables might lead to an impossible operation (e.g., \( \log(y_2) \)). (See lines 52 and 53 of fig. 1.)
SAMPLE RAPIER PROBLEM

MODEL EQUATION: Y = 1.2

Y = 80 + B1*X1 + B2*X2 + B3*X3
+ B12*X1*X2 + B13*X1*X3 + B23*X2*X3
+ B11*X1**2 + B22*X2**2 + B33*X3**2

X1 = TEMP
X2 = TIME
X3 = PRESS

Y = POUNDS OUTPUT
Y2 = COST OF OPERATION

THE DATA IS FROM AN INCOMPLETE FACTORIAL DESIGN WITH ONE
REPICATION (FICTITIOUS DATA)

There is a 80 to estimate

The transformations are

1 0 0 1 1 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 1.0333 1.000000 1.000000 1.000000 1.000000

(5f, 6d)

SAMPLE RAPIER PROBLEM

OBSERVED VARIABLES: WEIGHT = 1.000000

-1.000000 -1.000000 -1.000000

9.170000 38.503000

9.170000 38.503000

** REPLICATE SET 1

OBSERVED VARIABLES: WEIGHT = 1.000000

1.000000 -1.000000 -1.000000

12.760000 43.103000

12.760000 43.103000

** REPLICATE SET 2

OBSERVED VARIABLES: WEIGHT = 1.000000

1.000000 -1.000000 -1.000000

12.970000 44.003000

12.970000 44.003000

** REPLICATE SET 3

OBSERVED VARIABLES: WEIGHT = 1.000000

1.000000 -1.000000 -1.000000

17.030000 63.203000

17.030000 63.203000

** REPLICATE SET 4

OBSERVED VARIABLES: WEIGHT = 1.000000

1.000000 -1.000000 -1.000000

17.030000 63.203000

17.030000 63.203000

** REPLICATE SET 5

OBSERVED VARIABLES: WEIGHT = 1.000000

1.000000 -1.000000 -1.000000

17.030000 63.203000

17.030000 63.203000

** REPLICATE SET 6

OBSERVED VARIABLES: WEIGHT = 1.000000

1.000000 -1.000000 -1.000000

17.030000 63.203000

17.030000 63.203000

OBSERVED VARIABLES. WEIGHT = 1.000000

OBSERVATION = .

-1.000000 -1.000000 -1.000000

9.170000 38.503000

9.170000 38.503000

-1.000000 -1.000000 -1.000000

12.760000 43.103000

12.760000 43.103000

-1.000000 -1.000000 -1.000000

12.970000 44.003000

12.970000 44.003000

-1.000000 -1.000000 -1.000000

17.030000 63.203000

17.030000 63.203000

-1.000000 -1.000000 -1.000000

17.030000 63.203000

17.030000 63.203000

-1.000000 -1.000000 -1.000000

17.030000 63.203000

17.030000 63.203000

-1.000000 -1.000000 -1.000000

17.030000 63.203000

17.030000 63.203000

-1.000000 -1.000000 -1.000000

17.030000 63.203000

17.030000 63.203000

-1.000000 -1.000000 -1.000000

17.030000 63.203000

17.030000 63.203000

-1.000000 -1.000000 -1.000000

17.030000 63.203000

17.030000 63.203000

-1.000000 -1.000000 -1.000000

17.030000 63.203000

17.030000 63.203000
** REPLICATE SET 4 *******************************************************

OBSERVED VARIABLES, WEIGHT = 1.000000 OBSERVATION = 7
-1.000000 1.000000 1.000000
TERMS OF THE EQUATION, OBSERVATION = 7
-1.000000 -1.000000 1.000000 1.000000 1.000000 1.000000

OBSERVED VARIABLES, WEIGHT = 1.000000 OBSERVATION = 8
-1.000000 1.000000 1.000000
TERMS OF THE EQUATION, OBSERVATION = 8
-1.000000 -1.000000 1.000000

** REPLICATE SET 5 *******************************************************

OBSERVED VARIABLES, WEIGHT = 1.000000 OBSERVATION = 9
-1.000000 1.000000 1.000000
TERMS OF THE EQUATION, OBSERVATION = 9
-1.000000 1.000000 1.000000

OBSERVED VARIABLES, WEIGHT = 1.000000 OBSERVATION = 10
-1.000000 1.000000 1.000000
TERMS OF THE EQUATION, OBSERVATION = 10
-1.000000 1.000000

** REPLICATE SET 6 *******************************************************

OBSERVED VARIABLES, WEIGHT = 1.000000 OBSERVATION = 11
-1.000000 1.000000 1.000000
TERMS OF THE EQUATION, OBSERVATION = 11
1.000000 1.000000 1.000000

OBSERVED VARIABLES, WEIGHT = 1.000000 OBSERVATION = 12
-1.000000 1.000000 1.000000
TERMS OF THE EQUATION, OBSERVATION = 12
1.000000 1.000000

** REPLICATE SET 7 *******************************************************

OBSERVED VARIABLES, WEIGHT = 1.000000 OBSERVATION = 13
-1.000000 1.000000 1.000000
TERMS OF THE EQUATION, OBSERVATION = 13
1.000000 1.000000

OBSERVED VARIABLES, WEIGHT = 1.000000 OBSERVATION = 14
-1.000000 1.000000 1.000000
TERMS OF THE EQUATION, OBSERVATION = 14
1.000000 1.000000

OBSERVED VARIABLES, WEIGHT = 1.000000 OBSERVATION = 15
-1.000000 1.000000 1.000000
TERMS OF THE EQUATION, OBSERVATION = 15
1.000000 1.000000

** REPLICATE SET 8 *******************************************************

OBSERVED VARIABLES, WEIGHT = 1.000000 OBSERVATION = 16
-1.000000 1.000000 1.000000
TERMS OF THE EQUATION, OBSERVATION = 16
1.000000 1.000000

OBSERVED VARIABLES, WEIGHT = 1.000000 OBSERVATION = 17
-1.000000 1.000000 1.000000
TERMS OF THE EQUATION, OBSERVATION = 17
1.000000 1.000000

** REPLICATE SET 9 *******************************************************

OBSERVED VARIABLES, WEIGHT = 1.000000 OBSERVATION = 18
-1.000000 1.000000 1.000000
TERMS OF THE EQUATION, OBSERVATION = 18
1.000000 1.000000

34
OBSERVED VARIABLES, WEIGHT = 1.00000
2.000000 0 0 22.90330 58.10300
TEAMS OF THE EQUATION, OBSERVATION = 19
0 0 0 0 4.000000 0 3
22.900000 58.100000
** REPLICATE SET 10 **************************************************************************
DEF. VAR. 1 SSQ = 0.4303461 SUM = 46.730000 MEAN = 23.365000
DEF. VAR. 2 SSQ = 0.125000 SUM = 115.700000 MEAN = 57.850000
******************************************************************************
OBSERVED VARIABLES, WEIGHT = 1.00000
2.000000 0 0 7.99330 28.60000
TEAMS OF THE EQUATION, OBSERVATION = 20
0 0 0 0 0 4.000000 0
7.990000 28.600000
** REPLICATE SET 11 **************************************************************************
******************************************************************************
OBSERVED VARIABLES, WEIGHT = 1.00000
2.000000 0 0 7.99330 28.60000
TEAMS OF THE EQUATION, OBSERVATION = 21
0 0 0 0 0 4.000000 0
7.990000 28.600000
** REPLICATE SET 12 **************************************************************************
DEF. VAR. 1 SSQ = 0.4303461 SUM = 46.730000 MEAN = 23.365000
DEF. VAR. 2 SSQ = 0.125000 SUM = 115.700000 MEAN = 57.850000
******************************************************************************
OBSERVED VARIABLES, WEIGHT = 1.00000
2.000000 0 0 7.99330 28.60000
TEAMS OF THE EQUATION, OBSERVATION = 22
0 0 0 0 0 4.000000 0
7.990000 28.600000
** REPLICATE SET 13 **************************************************************************
DEF. VAR. 1 SSQ = 0.4303461 SUM = 46.730000 MEAN = 23.365000
DEF. VAR. 2 SSQ = 0.125000 SUM = 115.700000 MEAN = 57.850000
******************************************************************************
OBSERVED VARIABLES, WEIGHT = 1.00000
2.000000 0 0 7.99330 28.60000
TEAMS OF THE EQUATION, OBSERVATION = 23
0 0 0 0 0 4.000000 0
7.990000 28.600000
** REPLICATE SET 14 **************************************************************************
******************************************************************************
SUMS OF INDEP AND DEP VARIABLES
4.000000 0 -2.000000 24.000000 24.000000
24.000000 308.10000 1282.2000
X TRANSPOSE X MATRIX
ROW 1 24.00000
ROW 2 24.00000
ROW 3 2.000000 -2.000000 24.000000
ROW 4 -2.000000 2.000000 4.000000
ROW 5 4.000000 2.000000 -2.000000
ROW 6 -2.000000 2.000000 4.000000
ROW 7 4.000000 -2.000000 0
ROW 8 6.000000 0 2.000000 -2.000000
ROW 9 2.000000 -2.000000 8.000000
ROW 10 4.000000 0 2.000000 -2.000000 12.000000
ROW 11 12.000000 12.000000 12.000000
ROW 12 12.000000 12.000000 12.000000
X TRANSPOSE Y MATRIX
ROW 1 127.5600 264.5000
ROW 2 23.3600 238.5900
ROW 3 -12.2400 -110.3000
ROW 4 8.740000 12.900000
ROW 5 26.620000 139.700000
ROW 6 -9.680000 -95.900000
ROW 7 382.0000 1260.1000
ROW 8 275.1600 1276.1000
ROW 9 275.1600 1276.1000
ROW 10 275.1600 1276.1000
MEANS OF INDEP AND DEP VARIABLES
8.160000 -8.800000E-01 0
0.960000 12.324000 51.188000
3.8003000E-01 -0.8000000E-01 0.950333
0.950333

<table>
<thead>
<tr>
<th>X TRANSPOSE X MATRIX WHERE X IS DEVIATION FROM MEAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROW</td>
</tr>
<tr>
<td>ROW</td>
</tr>
<tr>
<td>ROW</td>
</tr>
<tr>
<td>ROW</td>
</tr>
<tr>
<td>ROW</td>
</tr>
<tr>
<td>ROW</td>
</tr>
<tr>
<td>ROW</td>
</tr>
<tr>
<td>ROW</td>
</tr>
<tr>
<td>ROW</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>X TRANSPOSE Y MATRIX WHERE X AND Y ARE DEVIATIONS FROM MEAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROW</td>
</tr>
<tr>
<td>ROW</td>
</tr>
<tr>
<td>ROW</td>
</tr>
<tr>
<td>ROW</td>
</tr>
<tr>
<td>ROW</td>
</tr>
<tr>
<td>ROW</td>
</tr>
<tr>
<td>ROW</td>
</tr>
<tr>
<td>ROW</td>
</tr>
<tr>
<td>ROW</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CORRELATION COEFFICIENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROW</td>
</tr>
<tr>
<td>ROW</td>
</tr>
<tr>
<td>ROW</td>
</tr>
<tr>
<td>ROW</td>
</tr>
<tr>
<td>ROW</td>
</tr>
<tr>
<td>ROW</td>
</tr>
<tr>
<td>ROW</td>
</tr>
<tr>
<td>ROW</td>
</tr>
<tr>
<td>ROW</td>
</tr>
</tbody>
</table>

SAMPLE RAPID PROBLEM
EACH COLUMN CONTAINS THE COEFFICIENTS FOR ONE DEPENDENT TERM

<table>
<thead>
<tr>
<th>SAMPLE RAPID PROBLEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>REGRESSION COEFFICIENTS (B1,...,Bk)</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
</tbody>
</table>

ANOVA OF REGRESSION ON DEPENDENT VARIABLE

<table>
<thead>
<tr>
<th>ANOVA OF REGRESSION ON DEPENDENT VARIABLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>-------------------------------------------</td>
</tr>
<tr>
<td>SOURCE</td>
</tr>
<tr>
<td>REGRESSION</td>
</tr>
<tr>
<td>RESIDUAL</td>
</tr>
<tr>
<td>TOTAL</td>
</tr>
</tbody>
</table>

R SQUARED = SS(REGR) / SS(TOTAL) = 0.997453
R = 0.997453

ANOVA OF LACK OF FIT

<table>
<thead>
<tr>
<th>ANOVA OF LACK OF FIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>SOURCE</td>
</tr>
<tr>
<td>LACK OF FIT</td>
</tr>
<tr>
<td>REPLICATION</td>
</tr>
<tr>
<td>RESIDUAL</td>
</tr>
<tr>
<td>F = MS(REGR) / MS(REP) = 2.524</td>
</tr>
</tbody>
</table>

36
### ANOVA of Regression on Dependent Variable Z

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>253,849,480</td>
<td>9</td>
<td>28,215,930</td>
</tr>
<tr>
<td>Residual</td>
<td>10,003,580</td>
<td>15</td>
<td>0.693,43872</td>
</tr>
<tr>
<td>Total</td>
<td>263,853,060</td>
<td>24</td>
<td></td>
</tr>
</tbody>
</table>

### ANOVA of Lack of Fit

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lack of fit</td>
<td>3,528,704</td>
<td>5</td>
<td>0.705,7430</td>
</tr>
<tr>
<td>Residual</td>
<td>10,401,580</td>
<td>15</td>
<td>0.693,43872</td>
</tr>
<tr>
<td>F = MS Lack of fit / MS Residual = 1.021</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Standard Deviation of Regression Coefficients

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.128439</td>
<td>0.397453</td>
</tr>
<tr>
<td>0.6160818e-01</td>
<td>0.190876</td>
</tr>
<tr>
<td>0.602399e-01</td>
<td>0.186412</td>
</tr>
<tr>
<td>0.848938e-01</td>
<td>0.253559</td>
</tr>
<tr>
<td>0.859870e-01</td>
<td>0.26498</td>
</tr>
<tr>
<td>0.862156e-01</td>
<td>0.26608</td>
</tr>
<tr>
<td>0.522540e-01</td>
<td>0.161947</td>
</tr>
<tr>
<td>0.522100e-02</td>
<td>0.161225</td>
</tr>
<tr>
<td>0.523440e-01</td>
<td>0.161978</td>
</tr>
</tbody>
</table>

### IX Transpose X Matrix

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.525339e-01</td>
<td>0.353463</td>
</tr>
<tr>
<td>0.505117e-01</td>
<td>0.546779e-01</td>
</tr>
<tr>
<td>-0.116754e-01</td>
<td>0.115899e-01</td>
</tr>
<tr>
<td>-0.191222e-01</td>
<td>0.107343</td>
</tr>
<tr>
<td>-0.24718e-01</td>
<td>0.13364e-01</td>
</tr>
<tr>
<td>-0.890203e-02</td>
<td>0.41931e-02</td>
</tr>
<tr>
<td>-0.222912e-01</td>
<td>0.19344e-01</td>
</tr>
<tr>
<td>-0.107955e-03</td>
<td>0.15050e-01</td>
</tr>
<tr>
<td>0.859870e-01</td>
<td>0.26498</td>
</tr>
<tr>
<td>0.862156e-01</td>
<td>0.26608</td>
</tr>
<tr>
<td>0.522540e-01</td>
<td>0.161947</td>
</tr>
<tr>
<td>0.522100e-02</td>
<td>0.161225</td>
</tr>
<tr>
<td>0.523440e-01</td>
<td>0.161978</td>
</tr>
</tbody>
</table>

### Sample RAPIM Problem

Calculated t statistics can be used to test the net regression coefficients with t = (β - β0) / SE(β), where β is the coefficient, β0 is the hypothesized value, and SE(β) is the standard error of the coefficient. The t-statistic can be compared to the critical t-value from the t-distribution. The desired value of probability is 95.0 percent. The term in (b) is being deleted.
CORRELATION COEFFICIENTS

<table>
<thead>
<tr>
<th>ROW</th>
<th>CORRELATION COEFFICIENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.000000</td>
</tr>
<tr>
<td>2</td>
<td>-0.383120E-01 0.836125E-01 1.000200</td>
</tr>
<tr>
<td>3</td>
<td>0.119455 0.117851 0.136592 1.003000</td>
</tr>
<tr>
<td>4</td>
<td>-0.192644E-01 0.237289 0.128566 0.16779 1.000000</td>
</tr>
<tr>
<td>5</td>
<td>0.209427E-01 0.671519E-01 0.464812E-01 0.382472E-02 0.382475E-02 0.298731 1.000000</td>
</tr>
<tr>
<td>6</td>
<td>-0.262038E-01 0.671519E-01 0.234675 0.382427E-02 0.382427E-02 0.298731 1.000000</td>
</tr>
</tbody>
</table>

SAMPLE RAPID PROBLEM

EACH COLUMN CONTAINS THE COEFFICIENTS FOR ONE DEPENDENT TERM

REGRESSION COEFFICIENTS

<table>
<thead>
<tr>
<th>REGRESSION Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.989932 2.940208 6.205272E-01 1.905733 0.293722E-01</td>
</tr>
</tbody>
</table>

ANOVA OF REGRESSION ON DEPENDENT VARIABLE

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>SUMS OF SQUARES</th>
<th>DEGREES OF FREEDOM</th>
<th>MEAN SQUARES</th>
</tr>
</thead>
<tbody>
<tr>
<td>REGRESSION</td>
<td>425.388105</td>
<td>6</td>
<td>53.1735206</td>
</tr>
<tr>
<td>RESIDUAL</td>
<td>1.08843231</td>
<td>16</td>
<td>0.06027023</td>
</tr>
<tr>
<td>TOTAL</td>
<td>426.476507</td>
<td>24</td>
<td></td>
</tr>
</tbody>
</table>

R SQUARED = SS(REG) / SS(TOT) = 0.997448
R = 0.997448

ANOVA OF LACK OF FIT

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>SUMS OF SQUARES</th>
<th>DEGREES OF FREEDOM</th>
<th>MEAN SQUARES</th>
</tr>
</thead>
<tbody>
<tr>
<td>LACK OF FIT</td>
<td>0.17165756</td>
<td>6</td>
<td>0.28609539</td>
</tr>
<tr>
<td>REPLICATION</td>
<td>0.91077925</td>
<td>16</td>
<td>0.06027023</td>
</tr>
<tr>
<td>RESIDUAL</td>
<td>1.08843231</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>426.476507</td>
<td>24</td>
<td>425.388105</td>
</tr>
</tbody>
</table>

Sums of squares due to each variable if it were last to enter regression:

<table>
<thead>
<tr>
<th></th>
<th>SUMS OF SQUARES</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>146.5792</td>
</tr>
<tr>
<td>2</td>
<td>20.16795</td>
</tr>
<tr>
<td>3</td>
<td>0.327725E-02</td>
</tr>
<tr>
<td>4</td>
<td>10.46794</td>
</tr>
<tr>
<td>5</td>
<td>0.655261E-01</td>
</tr>
<tr>
<td>6</td>
<td>0.886623E-01</td>
</tr>
<tr>
<td>7</td>
<td>0.886623E-01</td>
</tr>
<tr>
<td>8</td>
<td>0.886623E-01</td>
</tr>
<tr>
<td>9</td>
<td>0.886623E-01</td>
</tr>
</tbody>
</table>

STANDARD DEVIATION OF REGRESSION COEFFICIENTS (DERIVED FROM DIAGONAL ELEMENTS OF (X TRANSPOSE X) INVERSE MATRIX)

<table>
<thead>
<tr>
<th></th>
<th>STANDARD DEVIATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.090130E-01</td>
</tr>
<tr>
<td>1</td>
<td>0.016574E-01</td>
</tr>
<tr>
<td>2</td>
<td>0.927375E-01</td>
</tr>
<tr>
<td>3</td>
<td>0.628695E-01</td>
</tr>
<tr>
<td>4</td>
<td>0.089097E-01</td>
</tr>
<tr>
<td>5</td>
<td>0.854686E-01</td>
</tr>
<tr>
<td>6</td>
<td>0.865131E-01</td>
</tr>
<tr>
<td>7</td>
<td>0.474089E-01</td>
</tr>
<tr>
<td>8</td>
<td>0.474089E-01</td>
</tr>
<tr>
<td>9</td>
<td>0.474089E-01</td>
</tr>
</tbody>
</table>

(A X TRANSPOSE X) INVERSE MATRIX

| A (X TRANSPOSE X) INVERSE MATRIX |
|-----------------|-----------------|
| 1 | 0.925463E-01 |
| 2 | -0.300312E-02 0.485234E-01 |
| 3 | -0.116609E-01 0.117411E-01 0.345828E-01 |
| 4 | 0.106634E-01 0.117411E-01 0.284234E-01 3.138159 |
| 5 | 0.886423E-02 0.216811E-01 0.191941E-01 0.315285E-01 0.100920 |
| 6 | -0.247074E-01 0.717373E-01 0.193574E-01 0.312835E-01 0.133154E-01 0.102493 |
| 7 | -0.420823E-02 0.462259E-02 0.272532E-02 0.442653E-02 0.462653E-02 0.496919E-02 |
| 8 | -0.219391E-02 0.560052E-02 0.903552E-02 0.381839E-02 0.315285E-02 0.407641E-02 0.470715E-02 3.17731E-01 |
SAMPLE RAPIER PROBLEM

CALCULATED T STATISTICS
THE T STATISTICS CAN BE USED TO TEST THE NET REGRESSION COEFFICIENTS BI1.

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.767303</td>
<td>16.430365</td>
<td>0.212736</td>
<td>12.01326</td>
<td>0.951245</td>
<td>0.238269</td>
</tr>
<tr>
<td></td>
<td>6.68147</td>
<td>0.61247</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

UNDER NULL HYPOTHESIS THE INTERVAL [-T, T] WHERE T IS GIVEN ABOVE, HAS APPROX PROBABILITY LISTED BELOW.
MINUS SIGN INDICATES PROB EXCEEDS .999.

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.999</td>
<td>-0.999</td>
<td>0.165</td>
<td>-0.999</td>
<td>0.644</td>
<td>0.185</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

THE DESIRED VALUE OF PROBABILITY IS 95.0 PERCENT
THE TERM XI (31) IS BEING DELETED

CORRELATION COEFFICIENTS

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ROW 1</td>
<td>ROW 2</td>
<td>ROW 3</td>
<td>ROW 4</td>
<td>ROW 5</td>
<td>ROW 6</td>
</tr>
<tr>
<td></td>
<td>1.000000</td>
<td>-0.119455</td>
<td>-0.192414E-01</td>
<td>0.259760</td>
<td>0.256642</td>
<td>0.626203E-01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ROW 7</td>
<td>ROW 8</td>
<td>ROW 9</td>
<td>ROW 10</td>
<td>ROW 11</td>
<td>ROW 12</td>
</tr>
<tr>
<td></td>
<td>-0.626203E-01</td>
<td>-0.301868E-02</td>
<td>0.1000000</td>
<td>0.209642</td>
<td>0.259760</td>
<td>0.318004E-01</td>
</tr>
</tbody>
</table>

SAMPLE RAPIER PROBLEM

EACH COLUMN CONTAINS THE COEFFICIENTS FOR ONE DEPENDENT TERM

CONSTANT TERM (BO)  
9.498235

REGRESSION COEFFICIENTS (B1,...,BK)

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.937156</td>
<td>0.991988</td>
<td>1.061119</td>
<td>0.766166E-01</td>
<td>0.25268EE-01</td>
<td>1.906408</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.25268EE-01</td>
<td>0.1000000</td>
<td>0.209642</td>
<td>0.259760</td>
<td>0.318004E-01</td>
<td>0.382427E-02</td>
</tr>
</tbody>
</table>

ANOVA OF REGRESSION ON DEPENDENT VARIABLE 1

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SOURCE</td>
<td>SUMS OF SQUARES</td>
<td>DEGREES OF FREEDOM</td>
<td>MEAN SQUARES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>REGRESSION</td>
<td>425.304892</td>
<td>7</td>
<td>60.769269</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RESIDUAL</td>
<td>1.09170532</td>
<td>17</td>
<td>0.0617960E-01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>426.476597</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

STANDARD ERROR OF ESTIMATE 0.259413

USING POOLING STRATEGY 1 THE ERROR MEAN SQUARE = 0.259413 WITH DEGREES OF FREEDOM = 15

ANOVA OF LACK OF FIT

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SOURCE</td>
<td>SUMS OF SQUARES</td>
<td>DEGREES OF FREEDOM</td>
<td>MEAN SQUARES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LACK OF FIT</td>
<td>0.17679367</td>
<td>7</td>
<td>0.2499032E-01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APPLICATION</td>
<td>0.91677475</td>
<td>10</td>
<td>0.91677475E-01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RESIDUAL</td>
<td>1.09170532</td>
<td>17</td>
<td>0.0617960E-01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F = MS(LDF/MS(RES)) = 0.273</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SUNS OF SQUARES DUE TO EACH VARIABLE IF IT WERE LAST TO ENTER REGRESSION

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>172.4090</td>
<td>2</td>
<td>21.2825</td>
<td>4</td>
<td>11.3265</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.623348E-01</td>
<td>6</td>
<td>0.667608E-02</td>
<td>7</td>
<td>0.337156E-01</td>
</tr>
</tbody>
</table>

39
STANDARD DEVIATION OF REGRESSION COEFFICIENTS (DERIVED FROM DIAGONAL ELEMENTS OF (X TRANSPOSE X) INVERSE MATRIX)

| 0  | 0.903708E-01 |
| 1  | 0.601993E-01 |
| 2  | 0.578642E-01 |
| 3  | 0.826629E-01 |
| 4  | 0.825796E-01 |
| 5  | 0.832182E-01 |
| 6  | 0.477436E-01 |
| 7  | 0.466049E-01 |

(X TRANSPOSE X) INVERSE MATRIX

| ROW 1 | 0.500441E-01 |
| ROW 2 | 0.836083E-02 |
| ROW 3 | -0.293328E-01 |
| ROW 4 | 0.477268E-02 |
| ROW 5 | 0.205804E-01 |
| ROW 6 | 0.956328E-01 |

SAMPLE RAPID PROBLEM

CALCULATED T STATISTICS

THE T STATISTICS CAN BE USED TO TEST THE NET REGRESSION COEFFICIENTS IN TT.

| 1  | -0.999 |
| 2  | -0.999 |
| 3  | -0.632 |
| 4  | -0.323 |
| 5  | -0.999 |
| 6  | 0.495 |

THE DESIRED VALUE OF PROBABILITY IS 95.0 PERCENT
THE TERM XI 61 IS BEING DELETED

CORRELATION COEFFICIENTS

| ROW 1 | 1.000000 |
| ROW 2 | -0.119455 |
| ROW 3 | -0.192414E-01 |
| ROW 4 | 0.209642 |
| ROW 5 | 0.626203E-01 |
| ROW 6 | 0.626203E-01 |

SAMPLE RAPID PROBLEM

EACH COLUMN CONTAINS THE COEFFICIENTS FOR ONE DEPENDENT TERM
CONSTANT TERM 8(0)
9.987362

REGRESSION COEFFICIENTS (B1.....B6)

| 1  | 2.942794 |
| 2  | 0.991139 |
| 3  | 0.606665 |
| 4  | 0.783369E-01 |
| 5  | 0.318936E-01 |

ANOVA OF REGRESSION ON DEPENDENT VARIABLE

<table>
<thead>
<tr>
<th>Source</th>
<th>Sums of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>425.378620</td>
<td>5</td>
<td>85.075640</td>
</tr>
<tr>
<td>Residual</td>
<td>1.08937523</td>
<td>18</td>
<td>0.06029751</td>
</tr>
<tr>
<td>Total</td>
<td>426.476977</td>
<td>24</td>
<td></td>
</tr>
</tbody>
</table>

R Squared = SS(REG/SS(Total) = 0.997425
R = 0.997111

STANDARD ERROR OF ESTIMATE = 0.297024

USING MISSING STRATEGY 2 THE ERROR MEAN SQUARE = 0.7241516E-01 WITH DEGREES OF FREEDOM = 15
F(MS(REG)/MS(Err)) = 979.03 COMPARING TO F (6, 15)
### ANOVA OF LACK OF FIT

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>SUMS OF SQUARES</th>
<th>DEGREES OF FREEDOM</th>
<th>MEAN SQUARES</th>
</tr>
</thead>
<tbody>
<tr>
<td>LACK OF FIT</td>
<td>0.18160248</td>
<td>8</td>
<td>0.22703113E-01</td>
</tr>
<tr>
<td>REPLICATION</td>
<td>0.9167475</td>
<td>13</td>
<td>0.9167475E-01</td>
</tr>
<tr>
<td>RESIDUAL</td>
<td>1.0983773</td>
<td>13</td>
<td>0.61020957E-01</td>
</tr>
</tbody>
</table>

### Sums of Squares Due to Each Variable If It Were Last to Enter Regression

1. 189.8499
2. 21.29583
3. 12.67660
4. 0.654731E-01
5. 1.15.8460
6. 0.334046E-01

### Standard Deviation of Regression Coefficients (Derived from Diagonal Elements of \( X^T X \) Inverse Matrix)

<table>
<thead>
<tr>
<th>INDEX</th>
<th>STANDARD DEVIATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.90160E-01</td>
</tr>
<tr>
<td>1</td>
<td>0.57796E-01</td>
</tr>
<tr>
<td>2</td>
<td>0.80619E-01</td>
</tr>
<tr>
<td>3</td>
<td>0.82835E-01</td>
</tr>
<tr>
<td>4</td>
<td>0.47617E-01</td>
</tr>
<tr>
<td>5</td>
<td>0.46602E-01</td>
</tr>
</tbody>
</table>

### \( X^T X \) Inverse Matrix

<table>
<thead>
<tr>
<th>ROW</th>
<th>1</th>
<th>0.950152E-01</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROW</td>
<td>2</td>
<td>0.192791E-02</td>
</tr>
<tr>
<td>ROW</td>
<td>3</td>
<td>0.48359E-02</td>
</tr>
<tr>
<td>ROW</td>
<td>4</td>
<td>0.337152E-02</td>
</tr>
<tr>
<td>ROW</td>
<td>5</td>
<td>0.376179E-02</td>
</tr>
<tr>
<td>ROW</td>
<td>6</td>
<td>0.199262E-03</td>
</tr>
</tbody>
</table>

### Sample Raper Problem

**Calculated t Statistics**

The t statistics can be used to test the net regression coefficients \( b_i \).

<table>
<thead>
<tr>
<th>INDEX</th>
<th>T-STATISTIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.000000</td>
</tr>
<tr>
<td>2</td>
<td>-0.999</td>
</tr>
<tr>
<td>3</td>
<td>0.999</td>
</tr>
<tr>
<td>4</td>
<td>0.643</td>
</tr>
<tr>
<td>5</td>
<td>-0.999</td>
</tr>
<tr>
<td>6</td>
<td>0.999</td>
</tr>
</tbody>
</table>

The desired value of probability is 0.05 percent. The term \( x_1 \) is being deleted.

### Correlation Coefficients

<table>
<thead>
<tr>
<th>ROW</th>
<th>1</th>
<th>1.000000</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROW</td>
<td>2</td>
<td>-0</td>
</tr>
<tr>
<td>ROW</td>
<td>3</td>
<td>-0.19455</td>
</tr>
<tr>
<td>ROW</td>
<td>4</td>
<td>-0.209842</td>
</tr>
<tr>
<td>ROW</td>
<td>5</td>
<td>0.209842</td>
</tr>
</tbody>
</table>

### Sample Raper Problem

Each column contains the coefficients for one dependent term.

**Constant Term**

1. \( 10.02689 \)

### Regression Coefficients (\( b_1 \) to \( b_4 \))

<table>
<thead>
<tr>
<th>INDEX</th>
<th>COEFFICIENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.942962</td>
</tr>
<tr>
<td>2</td>
<td>0.987325</td>
</tr>
<tr>
<td>3</td>
<td>1.087618</td>
</tr>
<tr>
<td>4</td>
<td>0.80070E-01</td>
</tr>
<tr>
<td>5</td>
<td>1.895660</td>
</tr>
</tbody>
</table>
### ANOVA OF REGRESSION ON DEPENDENT VARIABLE

<table>
<thead>
<tr>
<th>Source</th>
<th>Sums of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>425.344822</td>
<td>5</td>
<td>85.068964</td>
</tr>
<tr>
<td>Residual</td>
<td>1.13177490</td>
<td>19</td>
<td>0.05957103E-01</td>
</tr>
<tr>
<td>Total</td>
<td>426.476597</td>
<td>24</td>
<td></td>
</tr>
</tbody>
</table>

R Squared = SS(Reg) / SS(Total) = 0.997345  \( R = 0.99872 \)

Using pooling strategy, the error mean square = 0.7241516E-01 with degrees of freedom = 15

\[ F = \frac{MS(Reg)/MS(Res)}{17.74} \]

**ANOVA OF LACK OF FIT**

<table>
<thead>
<tr>
<th>Source</th>
<th>Sums of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replication</td>
<td>0.21500015</td>
<td>9</td>
<td>2.7488905E-01</td>
</tr>
<tr>
<td>Residual</td>
<td>0.5917103E-01</td>
<td>10</td>
<td>0.05957103E-01</td>
</tr>
</tbody>
</table>

R Lack = SS(Res) / SS(Total) = 0.997345

**Sample Raper Problem**

The t statistics can be used to test the net regression coefficients below.

<table>
<thead>
<tr>
<th>t</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>17.16396</td>
<td>13.24464</td>
<td>0.912368</td>
<td>-7.28688</td>
</tr>
</tbody>
</table>

Under null hypothesis, the interval (-t, t) where t is given above, has approx probability listed below.

Minus sign indicates prob exceeds .999.

<table>
<thead>
<tr>
<th>t</th>
<th>-0.999</th>
<th>-0.999</th>
<th>-0.999</th>
<th>-0.999</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
</tr>
</tbody>
</table>

The desired value of probability is 95.0 percent, the term x1 51 being deleted.

### Correlation Coefficients

<table>
<thead>
<tr>
<th>Row</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.000000</td>
<td>-0.000000</td>
<td>-0.119455</td>
<td>-0.209642</td>
</tr>
<tr>
<td>2</td>
<td>0.000000</td>
<td>1.000000</td>
<td>0.517851</td>
<td>0.671619E-01</td>
</tr>
<tr>
<td>3</td>
<td>0.119455</td>
<td>0.517851</td>
<td>1.000000</td>
<td>1.000000</td>
</tr>
<tr>
<td>4</td>
<td>0.209642</td>
<td>0.671619E-01</td>
<td>1.000000</td>
<td></td>
</tr>
</tbody>
</table>

**Sample Raper Problem**

Each column contains the coefficients for one dependent term.

Constant term (b0) = 10.03236

Regression coefficients (b1,...,b8) = 1.04006 1.002149 1.051323 1.097115
ANOVA OF REGRESSION ON DEPENDENT VARIABLE

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>SUMS OF SQUARES</th>
<th>DEGREES OF FREEDOM</th>
<th>MEAN SQUARES</th>
</tr>
</thead>
<tbody>
<tr>
<td>REGRESSION</td>
<td>425.72363</td>
<td>4</td>
<td>176.319091</td>
</tr>
<tr>
<td>RESIDUAL</td>
<td>1.20023346</td>
<td>23</td>
<td>0.06011672E-01</td>
</tr>
<tr>
<td>TOTAL</td>
<td>427.923972</td>
<td>27</td>
<td>1.57946298E-01</td>
</tr>
</tbody>
</table>

R SQUARE = SSREG/SSTOT = 0.997105  R = 0.99852

USING POOLING STRATEGY 2 THE ERROR MEAN SQUARE = 3.7241516E-01 WITH DEGREES OF FREEDOM = 15

F = MSREG/MSRES = 468.199 COMPARTE TO F 4.151

ANOVA OF LACK OF FIT

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>SUMS OF SQUARES</th>
<th>DEGREES OF FREEDOM</th>
<th>MEAN SQUARES</th>
</tr>
</thead>
<tbody>
<tr>
<td>LACK OF FIT</td>
<td>0.28345871</td>
<td>13</td>
<td>0.28345871E-01</td>
</tr>
<tr>
<td>REPLICAITION</td>
<td>0.1677475</td>
<td>13</td>
<td>0.1677475E-01</td>
</tr>
<tr>
<td>RESIDUAL</td>
<td>1.20023346</td>
<td>23</td>
<td>0.66011672E-01</td>
</tr>
</tbody>
</table>

SQUARES DUE TO EACH VARIABLE IF IT WERE LAST TO ENTER REGRESSION

1 190.0073
2 23.63953
4 17.67437
7 126.8280

STANDARD DEVIATION OF REGRESSION COEFFICIENTS (DERIVED FROM DIAGONAL ELEMENTS OF IX TRANSPOSE XI INVERSE MATRIX)

0 0.663378E-01
1 0.573971E-01
2 0.554661E-01
4 0.600117E-01
7 0.454038E-01

IX TRANSPOSE XI INVERSE MATRIX

| ROW | 1   | 0.460737E-01 |
| ROW | 2   | -0.123740E-02 |
| ROW | 3   | 0.179983E-02 |
| ROW | 4   | -0.273692E-02 |

SAMPLE TAPLR PROBLEM

CALCULATE T STATISTICS

THE T STATISTICS CAN BE USED TO TEST THE NET REGRESSION COEFFICIENTS H11. 91.22363
18.50778
41.78336

UNDER NULL HYPOTHESIS THE INTERVAL [-T,T] WHERE T IS GIVEN ABOVE HAS APPROX PROBABILITY LISTED BELOW.

MINUS SIGN INDICATES P4H EXCEEDS .999.

1 -0.999
2 -0.999
4 -0.999
7 -0.999

THE DECIMAL VALUE OF PROBABILITY IS 95.0 PERCENT

SAMPLE TAPLR PROBLEM

FOR EACH DEPENDENT TERM AND OBSERVATION IS PRINTED

OBSERVED RESPONSE (Y OBSERVED)
CALCULATED RESPONSE (Y CALC)
RESIDUAL (Y OBSERVED - Y CALC + Y DIFF)
STANDARDIZED RESIDUAL (Z)

| Y OBSERVED | 9.1700 |
| Y CALC    | 9.0386 |
| Y DIFF    | 0.1314 |
| STUDENTIZED | 0.4984 |

| Y OBSERVED | 12.740 |
| Y CALC    | 12.816 |
| Y DIFF    | -0.2086 |
| STUDENTIZED | 0.0208 |

| Y OBSERVED | 12.970 |
| Y CALC    | 12.816 |
| Y DIFF    | 0.1539 |
| STUDENTIZED | 0.5720 |

| Y OBSERVED | 9.1100 |
| Y CALC    | 9.0602 |
| Y DIFF    | 0.1498 |
| STUDENTIZED | 0.0610 |

| Y OBSERVED | 8.9400 |
| Y CALC    | 8.9902 |
| Y DIFF    | 0.0496 |
| STUDENTIZED | 0.0002 |

43
| Y OBSERVED | 13.030 |
| Y CALC    | 16.923 |
| Y DIF     | 0.1070 |
| STUDENTIZED | 0.3975 |
| Y OBSERVED | 9.0500 |
| Y CALC    | 9.0386 |
| Y DIF     | 0.0114E-01 |
| STUDENTIZED | 0.4252E-01 |
| Y OBSERVED | 9.8600 |
| Y CALC    | 9.2866 |
| Y DIF     | -0.1786 |
| STUDENTIZED | -0.6635 |
| Y OBSERVED | 12.6000 |
| Y CALC    | 12.816 |
| Y DIF     | -0.2161 |
| STUDENTIZED | -0.8030 |
| Y OBSERVED | 13.210 |
| Y CALC    | 12.816 |
| Y DIF     | 0.3939 |
| STUDENTIZED | 1.4638 |
| Y OBSERVED | 17.200 |
| Y CALC    | 16.923 |
| Y DIF     | 0.2770 |
| STUDENTIZED | 1.0292 |
| Y OBSERVED | 17.040 |
| Y CALC    | 16.923 |
| Y DIF     | 0.1170 |
| STUDENTIZED | 0.347 |
| Y OBSERVED | 9.6100 |
| Y CALC    | 10.032 |
| Y DIF     | -0.4224 |
| STUDENTIZED | -1.5695 |
| Y OBSERVED | 10.010 |
| Y CALC    | 10.032 |
| Y DIF     | -0.0236E-01 |
| STUDENTIZED | -0.8308E-01 |
| Y OBSERVED | 10.120 |
| Y CALC    | 10.032 |
| Y DIF     | 0.0764E-01 |
| STUDENTIZED | 0.3257 |
| Y OBSERVED | 9.9500 |
| Y CALC    | 10.032 |
| Y DIF     | -0.6236E-01 |
| STUDENTIZED | -0.3060 |
| Y OBSERVED | 11.780 |
| Y CALC    | 11.741 |
| Y DIF     | 0.3936E-01 |
| STUDENTIZED | 0.1663 |
| Y OBSERVED | 23.830 |
| Y CALC    | 23.501 |
| Y DIF     | 0.3290 |
| STUDENTIZED | 1.2226 |
| Y OBSERVED | 22.900 |
| Y CALC    | 23.501 |
| Y DIF     | -0.6010 |
| STUDENTIZED | -2.2333 |
| Y OBSERVED | 7.9900 |
| Y CALC    | 9.0281 |
| Y DIF     | -0.3080E-01 |
| STUDENTIZED | -0.1414 |
| Y OBSERVED | 12.110 |
| Y CALC    | 12.037 |
| Y DIF     | 0.7335E-01 |
| STUDENTIZED | 0.2726 |
| Y OBSERVED | 11.700 |
| Y CALC    | 12.037 |
| Y DIF     | -0.3367 |
| STUDENTIZED | -1.2510 |
| Y OBSERVED | 10.110 |
| Y CALC    | 10.032 |
| Y DIF     | 0.7764E-01 |
| STUDENTIZED | 0.2089 |
| Y OBSERVED | 10.010 |
| Y CALC    | 10.032 |
| Y DIF     | -0.2336E-01 |
| STUDENTIZED | -0.8308E-01 |
| Y OBSERVED | 10.020 |
| Y CALC    | 10.032 |
| Y DIF     | -0.1336E-01 |
| STUDENTIZED | -0.4592E-01 |
SAMPLE RAPID PROBLEM

SKENEWS (SHOULD BE NEAR ZERO)
0.1928
KURTOSIS (SHOULD BE NEAR THREE)
1.6949

CHI-SQUARE IS NOT COMPUTED FOR LESS THAN 30 DEGREES OF FREEDOM FOR ERROR.

FOR EACH SET OF INDEP VARIABLES THERE IS COMPUTED...

PREDICTED RESPONSE
VARIANCE OF REGRESSION LINE
STANDARD DEVIATION OF REGRESSION
VARIANCE OF PREDICTED VALUE
STANDARD DEVIATION OF PREDICTED VALUE

INPUT DATA FOR THIS PREDICTED RESPONSE
-1.000000 1.000000 1.000000 -0 -0

INDEPENDENT TERMS ACCORDING TO FINAL REGRESSION MODEL
-1.000000 1.000000 -1.000000 1.000000

PREDICTED RESPONSE FOR ABOVE INDEP VARIABLES
8.940211
0.192847E-01
0.138869
0.916998E-01
0.302820

Lewis Research Center,
National Aeronautics and Space Administration,
Cleveland, Ohio, October 16, 1969,
129-04.
APPENDIX A

PROGRAM DOCUMENTATION AND LISTINGS

The contents of this appendix include a flow chart of the program, a listing of the routines used in RAPIER and their major functions, the call structure of the program, a dictionary of the program, and the listing.

General Mathematical and Logical Flow of Program

The flow of operation in RAPIER is illustrated in figure 6.

START
Given data
Perform transformations

Yes
Calculate replication error

Yes
Calculate Y'X
Y'X

Do sequential regressions?

Yes
j = 0

j = 1

j = j + 1

No

Calculate C

No

No

START

START

START

START

START

START

START

START
<table>
<thead>
<tr>
<th>FORTRAN name</th>
<th>Function of routine</th>
</tr>
</thead>
<tbody>
<tr>
<td>BORD</td>
<td>Inverts symmetric matrix $A$ of order $n$ by adding bordering column to already inverted matrix of order $n - 1$</td>
</tr>
<tr>
<td>CHISQ</td>
<td>Computes residuals at observed points and chi-square statistic to test goodness of fit</td>
</tr>
<tr>
<td>HIST</td>
<td>Prints histogram of residuals</td>
</tr>
<tr>
<td>INVXTX</td>
<td>Inverts symmetric matrix by Gauss elimination</td>
</tr>
<tr>
<td>LOC</td>
<td>When given row and column indices of symmetric matrix element, it computes location this element would have if only upper triangular part were stored as vector.</td>
</tr>
<tr>
<td>MATINV</td>
<td>Controls inversion process; computes regression coefficients; computes eigenvalues and eigenvectors of $X'X$ if requested</td>
</tr>
<tr>
<td>MFX</td>
<td>Prints and truncates $X'X$ and computes $C$</td>
</tr>
<tr>
<td>PREDCT</td>
<td>Computes predicted values, variances, and standard deviations of regression line and further observations at specified points</td>
</tr>
<tr>
<td>RAPIER</td>
<td>Executes overall problem control; computes replication error; controls deletion of variables when given results of t-tests; controls most input and output</td>
</tr>
<tr>
<td>RECT</td>
<td>Writes rectangular matrix</td>
</tr>
<tr>
<td>RSTATS</td>
<td>Computes regression statistics and writes regression and lack-of-fit analysis of variance tables</td>
</tr>
<tr>
<td>SUMUPS</td>
<td>Constructs $X'X$ and $X'y$ matrices one observation at a time, in double precision</td>
</tr>
<tr>
<td>TRAN</td>
<td>Performs transformations</td>
</tr>
<tr>
<td>TRIANG</td>
<td>Writes lower triangular part of symmetric matrix</td>
</tr>
<tr>
<td>TTEST</td>
<td>Computes $t$-statistics and their significance levels; determines which variable should be deleted</td>
</tr>
<tr>
<td>EIGEN</td>
<td>Computes eigenvalues and eigenvectors of input symmetric matrix</td>
</tr>
</tbody>
</table>
Call Structure of Program

The call structure of the program is illustrated in figure 7.

![Call structure of RAPIER](image)

Figure 7. - Call structure of RAPIER.

Dictionary of Program

<table>
<thead>
<tr>
<th>FORTRAN name</th>
<th>Mathematical symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACTDEV</td>
<td>$e_i$</td>
<td>Error in observation $i$; difference between observed and predicted response</td>
</tr>
<tr>
<td>ZEAN</td>
<td>$E(X), \mu$</td>
<td>Expected or average value (of $X$)</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>Regression coefficients other than the constant</td>
</tr>
<tr>
<td>BO</td>
<td></td>
<td>Constant regression coefficient</td>
</tr>
<tr>
<td>CHI</td>
<td>$\chi^2$</td>
<td>Chi-squared statistic</td>
</tr>
<tr>
<td>CON</td>
<td></td>
<td>Constants used in transformations, and results of transformations</td>
</tr>
<tr>
<td>CORR</td>
<td>$C$</td>
<td>Correlation matrix</td>
</tr>
<tr>
<td>FORTRAN name</td>
<td>Mathematical symbol</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>DELETE</td>
<td></td>
<td>Logical variable set to TRUE when deletion of terms is desired</td>
</tr>
<tr>
<td>ERRMS</td>
<td>$\hat{\sigma}^2$</td>
<td>Estimate of $\sigma^2$ used in hypothesis tests</td>
</tr>
<tr>
<td>EIG</td>
<td></td>
<td>Overflow area for storing part of modal matrix</td>
</tr>
<tr>
<td>FMT</td>
<td></td>
<td>Variable input format</td>
</tr>
<tr>
<td>FMTTRI</td>
<td></td>
<td>Format for printing matrix</td>
</tr>
<tr>
<td>IDENT</td>
<td></td>
<td>First identification printed at top of each page</td>
</tr>
<tr>
<td>IDOUT</td>
<td></td>
<td>Original sequence number of each term relating reduced models to original model</td>
</tr>
<tr>
<td>IFCHI</td>
<td></td>
<td>Logical variable set to TRUE if chi-square option is desired</td>
</tr>
<tr>
<td>IFSSR</td>
<td></td>
<td>Logical variable set to TRUE if sequential regressions are desired</td>
</tr>
<tr>
<td>IFTT</td>
<td></td>
<td>Logical variable set to TRUE if t-statistics are desired</td>
</tr>
<tr>
<td>IFWT</td>
<td></td>
<td>Logical variable set to TRUE if all weights of observations are 1.0</td>
</tr>
<tr>
<td>INPUT</td>
<td></td>
<td>Input logical tape unit number for data</td>
</tr>
<tr>
<td>INPUT5</td>
<td></td>
<td>Set equal to 5 to denote input device is card reader</td>
</tr>
<tr>
<td>INTER</td>
<td></td>
<td>Tape unit where input data is stored for chi-square or prediction routines</td>
</tr>
<tr>
<td>IOUT</td>
<td></td>
<td>Sequence number of term among those remaining which is to be deleted</td>
</tr>
<tr>
<td>JCOL</td>
<td></td>
<td>Total number of independent and dependent terms in regression model</td>
</tr>
<tr>
<td>KONNO</td>
<td></td>
<td>Number of constants originally supplied for transformations</td>
</tr>
<tr>
<td>LENGTH</td>
<td></td>
<td>Number of locations in correlation matrix storage area currently needed</td>
</tr>
<tr>
<td>LIST</td>
<td></td>
<td>Set equal to 6 to denote output device is printer</td>
</tr>
<tr>
<td>FORTRAN name</td>
<td>Mathematical symbol</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>NARAY</td>
<td></td>
<td>Number of replications per replicate set</td>
</tr>
<tr>
<td>NCON</td>
<td></td>
<td>Array containing addresses in CON array for use in transformations</td>
</tr>
<tr>
<td>NERROR</td>
<td></td>
<td>Degrees of freedom for error mean square estimate</td>
</tr>
<tr>
<td>NODEP</td>
<td></td>
<td>Number of dependent variables</td>
</tr>
<tr>
<td>NOOB</td>
<td>N</td>
<td>Number of observations</td>
</tr>
<tr>
<td>BZERO</td>
<td></td>
<td>Logical variable set to TRUE if constant $b_0$ coefficient should be in regression model</td>
</tr>
<tr>
<td>NOTERM</td>
<td>J</td>
<td>Number of terms in current regression model</td>
</tr>
<tr>
<td>NOVAR</td>
<td>K</td>
<td>Number of independent variables to be read</td>
</tr>
<tr>
<td>NPDEG</td>
<td>NPDEG</td>
<td>Pooled degrees of freedom for replication error</td>
</tr>
<tr>
<td>NRES</td>
<td>$N - J - D$</td>
<td>Degrees of freedom for estimation of residual variance</td>
</tr>
<tr>
<td>NTERM</td>
<td></td>
<td>Array containing locations of terms in CON array that should be in regression model</td>
</tr>
<tr>
<td>NTRAN</td>
<td></td>
<td>Array containing transformation codes for use in performing transformations</td>
</tr>
<tr>
<td>NTRANS</td>
<td></td>
<td>Number of transformations to perform</td>
</tr>
<tr>
<td>NWHERE</td>
<td></td>
<td>Location in $X$ array of first dependent variable; used in prediction routine to adjust for deleted terms</td>
</tr>
<tr>
<td>NXCOD</td>
<td></td>
<td>Array containing addresses of variables (or terms with address &gt; 60) for use in transformations</td>
</tr>
<tr>
<td>P</td>
<td></td>
<td>Probability that the interval $(-t, t)$ must have before a term is considered to be significant</td>
</tr>
<tr>
<td>POOLED</td>
<td>SSQ(REP)</td>
<td>Array containing pooled sums of squares from replications for each dependent term</td>
</tr>
<tr>
<td>PREDCT</td>
<td></td>
<td>Logical variable set to TRUE if prediction option is desired</td>
</tr>
<tr>
<td>RELSKW</td>
<td>RELSKW</td>
<td>Skewness of distribution of residuals</td>
</tr>
<tr>
<td>FORTRAN name</td>
<td>Mathematical symbol</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>RELKUR</td>
<td>RELKUR</td>
<td>Kurtosis of distribution of residuals</td>
</tr>
<tr>
<td>REPS</td>
<td></td>
<td>Logical variable set to TRUE if there are replicate sets in the data</td>
</tr>
<tr>
<td>REPVAR</td>
<td></td>
<td>Array containing replication variance of each dependent term</td>
</tr>
<tr>
<td>RESMS</td>
<td></td>
<td>Array containing residual mean square or variance of each dependent term</td>
</tr>
<tr>
<td>RWT</td>
<td></td>
<td>Reciprocal of total weight</td>
</tr>
<tr>
<td>SUMX</td>
<td>$\sum x, \sum y$</td>
<td>Array containing sums of independent and dependent terms</td>
</tr>
<tr>
<td>SUMX2</td>
<td>$\sum x^2, \sum y^2$</td>
<td>Array containing sum of squared independent and dependent terms</td>
</tr>
<tr>
<td>SUMXX</td>
<td>$X'X$</td>
<td>Sums of squares and crossproducts matrix, and variance-covariance matrix of independent terms</td>
</tr>
<tr>
<td>SUMXXI</td>
<td>$(X'X)^{-1}$</td>
<td>Variance-covariance matrix of estimated regression coefficients</td>
</tr>
<tr>
<td>SUMXY</td>
<td>$X'y$</td>
<td>Array containing sums of crossproducts of independent terms with dependent terms</td>
</tr>
<tr>
<td>TOTWT</td>
<td>$\sum w_i$</td>
<td>Sum of weight of observations</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td>Before transformations are performed, this contains the variables as read in. After the transformations are performed, appropriate data from the CON array are placed here according to information on the terms cards.</td>
</tr>
<tr>
<td>NLOF</td>
<td>$N - J - NPDEG - D$</td>
<td>Degrees of freedom for estimating variance due to lack of fit</td>
</tr>
<tr>
<td>NREG</td>
<td>$J$</td>
<td>Degrees of freedom for determining variance due to regression</td>
</tr>
<tr>
<td>NTOT</td>
<td>$N-D$</td>
<td>Total degrees of freedom</td>
</tr>
<tr>
<td>RNLOF</td>
<td></td>
<td>Reciprocal of degrees of freedom for lack of fit</td>
</tr>
<tr>
<td>RNREG</td>
<td></td>
<td>Reciprocal of degrees of freedom for regression</td>
</tr>
</tbody>
</table>
FORTRAN Mathematical symbol | Description
--- | ---
RNRES | Reciprocal of degrees of freedom for residual
STORY1 | Logical variable set to TRUE if product $C \cdot C^{-1}$ is to be computed and printed
STORYC | Logical variable set to TRUE if eigenvectors and eigenvalues of $C$ are to be computed and printed
STORYX | Logical variable set TRUE if eigenvectors and eigenvalues of $X'X$ are to be computed and printed
SATRTD | Logical variable indicating that there are no degrees of freedom for residual if TRUE
ECONMY | Logical variable indicating suppress printout of $x'x$, $x'x$ deviations, and $C$ if TRUE
XCHK | Array used in checking if all values of independent terms are the same within a replicate set

**Program Listing**

```
SIBFTC BLCK

BLCK DATA
BYPASS, BZERO, DELETE, FIRST, IFCHI, IFSSL,
X IFIT, IFMT, INPUT, INPUTS, INTER,
X ISTRAT, JCOL, KONNO, LENGTH, LIST,
X NERRC, NODEP, NODEP, NROW, NOTERM,
X NOVAR, NPEEG, NRES, NTRANS, NWHERE,
X P, PREDICT, REPS, RWT,
X STORY, STORYC, STORYX, TOTWT, WEIGHT,
X ERRFX, ECONMY
LOGICAL ECONMY
DOUBLE PRECISION RWT, TOTWT, WEIGHT
COPPCX, /FORMS/: FMTRI(1), FMTRI(14)
DATA (INTER/3, INPUTS/5, LIST/6)
DATA (FMTRI(1), I=1, 4) /6H(MW RO, 6HW 15, 2, 6HX, (8G1, 6H9.6)) /
ENC
```
$IBFTC RAPIER

C THIS IS RAPIER, MAIN PROGRAM FOR REGRESSION ANALYSIS PROVIDING
C INTERNAL EVALUATION OF RESULTS.
C**************************************************************************
C
C COMCEN/BIG/SUMX1(1830), SUMXX1(1830), EIG1(1830), SUMXY(60,9),
C E(60,9), CONR1(1830)
C COMCEN/MED/ B049, CON(99), ERRMS(9),
C ICENT(13), IDOUT(60), NCON(200)
C NTERM(60), NTRAN(100), NXCOD(100), POOLED(9),
C REPVAR(9), RESMS(9), SUMX(138),
C XSUMX2(691), X(s9), ZEOAN(69), SUMY2(18)
C COMCEN/FMT/ FMT(13), FMTTRI(14)
C COMCEN/SMALL/ BYPASS, BZERO, DELETE, FIRST, IFGHI, IFSSR,
C X IFFT, IFWT, INPUT, INPUTS, INTER,
C X ISTRAT, JCOL, KONNO, LENGTH, LIST,
C X NERROR, NODEP, NOOB, NOTERM,
C X NNEW, NPEEG, NRES, NTRANS, NWHERE,
C X P, PREOCT, REPS, RWT,
C X STORY, STORYC, STORYX, TOTWT, WEIGHT,
C X ERRFXD, ECONMY
C LOGICAL ECONMY
C DOUBLE PRECISION RWT, TOTWT, WEIGHT
C LOGICAL BYPASS, BZERO, DELETE, IFCHI,
C X XFRSR, XFFT, IFWT, REPS, PREDCT,
C X STORY, STORYC, STORYX, FIRST, ERRFXD
C LOGICAL XSAVE
C DIMENSION XCHK(60)
C COMCEN/CNTRS/ I, IBC, IC, ICOL,
C X INEW, INOCH, IODD, IOUT, IR,
C X IRC, IREP, IS4, ITC, J,
C X K, KBAR
C**************************************************************************
C
C EQUVALENCE (NARAY, CBRR)\$ (SB0, SSQ, REPVAR), IZOUT, SUMXX
C DIMENSION NARAY1(1830), S(9), SSQ(9)
C DIMENSION ZOUT(1)
C**************************************************************************
C
C ZERQUT ALL DATA ARRAYS EACH NEW DATA SET
C 100 IZOUT=9470
C DO ICI J=1, IZOUT
C 101 SUMXX(J) = 0.0
C**************************************************************************
C
C REAB IDENTIFICATION CARD AND OPTIONS CARD
C
C REAB(INPUT5, 110) I, IDENT
C WRITE(LIST, 111) IDENT
C FIRST=, TRUE.
C ERRFXD=.FALSE.
C 113 (F(I) 120, 120, 115
C 115 REAB(INPUT5, 300) FMT
C WRITE(LIST, 301) FMT
C (=I-1
C 120 REAB(INPUT5, 1282) NOWAR, NODEP, NOTERM, NOOB
C WRITE(LIST, 1283) NOWAR, NODEP, NOTERM, NOOB
C REAB(INPUT5, 117) BZERO, IFFT, IFWT, IFCHI, STORYC, STORYX, STORYI, IFSSR
C , ECONMY, ISTRAT
C WRITE(LIST, 118) BZERO, IFFT, IFWT, IFCHI, STORYC, STORYX, STORYI, IFSSR
C , ECONMY, ISTRAT
C
C
C*********************************************************
C THESE ARE INITIALIZATIONS MADE BEFORE EACH SET OF DATA
ICOL DETERMINES THE NUMBER OF VARIABLES READ PER OBSERVATION
JCOL IS THE NUMBER OF TERMS IN THE TOTAL REGRESSION EQUATION
LENGTH IS THE NUMBER OF STORES NEEDED FOR THE MATRICES
   LENGTH= NOTERM*(NOTERM+1)/2
   ICOL=NOVAR + NODEP
   JCOL = NOTERM +NODEP
   NWHERE= NOTERM
REMEMBER
DO 140 J=1,60
   IDCT(J) = J
140 NTERM(J)=J
DO 145 J=1,160
   NXCC(J)=J
   NTRAN(J)=0
145 NCCN(2*J)=J
C*********************************************************
C*********************************************************
C IF(BZER) WRITE(LIST,190)
C IF(NOT.BZER) WRITE(LIST,170)
C*********************************************************
C READ(INPUT5,282) NTRANS,KONNO
C (F(NTRANS,EQ;0) GO TO 255
C 220 READ (INPUT5,230) (TERM(K),K=1,JCOL)
C WRITE(LIST,235) (TERM(K),K=1,JCOL)
C READ (INPUT5,230) (NXCOD(I),NTRAN(I),NCON(2*I-1),NCON(2*I),I=1,NTRANS
C WRITE(LIST,240) (NXCOD(I),NTRAN(I),NCON(2*I-1),NCON(2*I),I=1,
C X NTRANS)
C (F(KONNO) 255,255,256
C 250 READ (INPUT5,260) (CON(I),I=1,KONNO)
C WRITE(LIST,262) (CON(I),I=1,KONNO)
C*********************************************************
C 255 READ(INPUT5,257) DELETE,P
C IF(DELETE) [FTT=.TRUE.]
C*********************************************************
C IF THERE ARE REPLICATED POINTS READ IN THE NUMBER OF POINTS AND
C THE NUMBER OF REPLICATIONS. SINGLE DATA POINTS ARE DATA POINTS
C REPLICATED ONCE. OBSERVED DATA MUST BE ARRANGED IN THE ORDER
C IMPLIES HERE.
C READ(INPUT5,257) REP3
C KSAVE=.FALSE.
C 265 IF(.NOT.REP3) GO TO 290
C READ(INPUT5,282) IREP,(NARAY(I),I=1,IREP)
C NREP=0
C IREP=1
C C=NARAY(I)
C KSAVE=.TRUE.
C DO 315 I=1,NODEP
C POCLED(I)=C.O
C S(I)=0.0
C 315 SSC(I)=0.0
C*********************************************************
C READ VARIABLE FORMAT FOR DATA
C 290 READ(INPUT5,110) INPUT,FMT
C WRITE(LIST,111) FMT
C 310 TOTWT=0.OCC
C WEIGHT=1.OCC
C WRITE(LIST,301) IDENT
C
C**********************************************************************************************
C READ IN INPUT VARIABLES
    DO 490 J=1, NOOB
    330 IF (.NOT. ACT (FMT)) GO TO 350
    340 READ (INPUT, FMT) (X(I), I=1, ICOL)
        GO TO 360
    350 READ (INPUT, FMT) (X(I), I=1, ICOL), WEIGHT
    360 CONTINUE
        IF (ECONMY) WRITE (LIST, 381) J (X(I), I=1, ICOL)
    381 FORMAT (1H14, 15X, 9G14.6)
        IF (ECONMY) GO TO 390
        WRITE (LIST, 370) WEIGHT, J
    390 CALL TRAN
    420 DO 430 K=1, JCOL
        IF (TERM(K)) CCNTINUE
    430 CONTINUE
    450 CONTINUE
        IF (ECONMY) GO TO 4609
        WRITE (LIST, 480) J
    4609 CONTINUE
        IF (IFCH(I)) WRITE (INTER) (X(I), I=1, ICOL), WEIGHT
        IF (.NOT. XSAVE) GO TO 4610
    4610 XCH(K) = X(K)
    4611 CONTINUE
C**********************************************************************************************
C COMPUTE THE ERROR VARIANCE FROM REPLICATED DATA
    IF (.NOT. REPS) GO TO 480
        GCTC = 1
    480 IF (NARAY (REPS) .GT. 1) IGOTO=2
        IF (J .GE. IC) WRITE (6, 462) REPS
        DO 475 I=1, NODEP
            IF (I-1) 46290, 464
        4629 DO 463 K=1, NOTERM
            IF (X(K) .NE. XCH(K)) GO TO 2001
        463 CONTINUE
        464 CONTINUE
            KBAR = NOTERM+1
            S(I) = S(I) + X(KBAR)
            SSC(I) = SSC(I) + X(KBAR) ** 2
            IF (J-IC) 475, 465
        465 GO TO (468, 466, IGOTO)
        466 ZEAN(I) = S(I) / FLOAT(NARAY(REPS))
            SSC(I) = SSC(I) - ZEAN(I) * S(I)
            POOLED(I) = POOLED(I) + SSC(I)
            WRITE (6, 467) I, SSC(I), S(I), ZEAN(I)
        468 IF (I.EQ. NODEP) GO TO 469
            REP = REP + 1
            (C = IC + NARAY(REPS) - 1
            IF (REP.EQ. REP+1)
                WRITE (LIST, 4671)
        469 S(I) = 0.0
            SSC(I) = 0.0
            XSAVE = .TRUE.
        475 CONTINUE
C*******CALCULATE SUMS, SUMS OF SQUARES AND SUMS OF CROSS PRODUCTS.**************
CALCULATE SUMS, SUMS OF SQUARES AND SUMS OF CROSS PRODUCTS.
CALL SUMUP
CONTINUE

C CONTINUE IS THE END OF THE LOOP FOR READING DATA CARDS
DO 493 I=1,NODEP
REPVARI(I)=POOLED(I)/FLOAT(NPDEG)
CONTINUE
CONTINUE

C ALL DATA HAS BEEN READ IN AND THE XTRANSPOSEX AND XTRANSPOSEY
C MATRIX HAS BEEN CALCULATED.
CALL MXF
REPRINT INTER
GO TO 640

C THIS CODING DELETES THE DATA FROM THE CORRELATION MATRIX
C CORRESPONDING TO THE INDEPENDENT TERM DELETED
6500 CONTINUE
IR=ICT-1
IC=NTERM - IOUT
IF (IC.EQ.0) GO TO 6700
NEW=INCH+IR/2
NEW = INCH
NEW = INCH + IOUT
RC=C
BC=G
TC=C
DO 6600 I=IOLD,LENGTH
NEW = NEW+1
LC=IOLD + 1
IF(ITC.GT.0) GO TO 6540
IRC=IRC + 1
IF (IRC.GT.IR) GO TO 6530
CORR(NEW) = CORR(IOCD)
GO TO 6600
6540 IRC=IRC
CORR(NEW) = CORR(IOCD)
6600 CONTINUE

C INVERT THE CORR COEF MATRIX AND COMPUTE REGRESSION COEFS
C AND SUMS OF SQUARES DUE TO REGRESSION IN THE MATRIX INVERSION
6400 CONTINUE
CALL MATINV
FIRST=.FALSE.

C
IF A VARIABLE HAS BEEN DELETED ADJUST COUNTERS AND RECOMPUTE THE
REGRESSION. IF NO VARIABLE HAS BEEN DELETED CONTROL WILL PASS
FROM THE TTEST ROUTINE TO THE CHI-SQUARE OPTION.

**CONTINUE**

**GO TO 1020**

**WRITE (LIST,3011)** IDENT

**CALL TTEST**($1020)

**IF(NCEP=1) 985,990,985**

**WRITE (LIST,986)** NODEP

**NCEP=1**

**J=JCEL-1**

**DD 995 K=OUT.J**

**NTERM(K)=NTERM(K+1)**

**ZEAN(K) = ZEAN(K+1)**

**SUMX(K) = SUMX(K+1)**

**SUMX2(K) = SUMX2(K+1)**

**4DOUT(K) = 4DOUT(K+1)**

**SUMXY(K+1) = SUMXY(K+1,1)**

**CONTINUE**

**IF(NTERM.EQ.1) GO TO 1020**

**GO TO 6500**

**GO TO 1000**

**WRITE (LIST,1005)**

**NCEP=C**

**GO TO 1035**

**GO TO 1025**

**GO TO 1030**

**WRITE (LIST,331)** IDENT

**CALL CHERC**

**GO TO 1035**

**READ INPUTS,117** PRECCT

**IF (.NOT. PRECCT) GO TO 1CC**

**CALL FRECCT**

**GO TO 1CC**

**GO TO 2001**

**WRITE (LIST,1206)**

**STCF**

**GO TO 800**

**FCPFT(11)**

**FCPFT(1H2)**

**FCPFT (12,13A6)**

**FCPFT (1H1,13A6,A2)**

**FCPFT(9L1,11)**

**FCPFT(1H 9L11)**

**FCPFT(3H THERE IS NO BO TERM IN THE MODEL)**

**FCPFT(26H THERE IS A BO TO ESTIMATE)**

**FCPFT(4012)**

**FCPFT(1H NTERM(K)= / (1H 3014))**

**FCPFT(25H THE TRANSFORMATIONS ARE / (1H 5(414,5X)))**

**FCPFT(11L, F3 = 3)**

**FCPFT(5E15.7)**

**FCPFT(19H THE CONSTANTS ARE / (1H 8G15(7)))**

**FCPFT(2014)**

**FCPFT(1H 2014)**
300 FORMAT(13A6,1A2)
301 FORMAT (1H 13A6,A2)
370 FORMAT(1HO,29H OBSERVED VARIABLES, WEIGHT = G14.6,6X,15H OBSERVATION
1 = ,15)
380 FORMAT(1H 961A6)
460 FORMAT(1H ,37H TERMS OF THE EQUATION, OBSERVATION = ,15)
462 FORMAT(18H K** REPLICATE SET 15,3X,100(1H*))
467 FORMAT(1H 125(1H*))
467 FORMAT(1H DEP. VAR I6,8H SSQ=G14.7,8H SUM=G14.7,8H N
X=AN= G14.78)
500 FORMAT(1H 60H TOTAL OF THE WEIGHTS ,F8.0)
510 FORMAT(1H 60H TOTAL WEIGHT TOO HIGH, EXCEEDS NUMBER OF OBSERVATION
1 IS BY ONE PERCENT. )
520 FORMAT(1H 60H TOTAL WEIGHT TOO LOW, LESS THAN 95 PERCENT OF NUMBER
1 OF OBSERVATIONS. )
540 FORMAT(1H 8G14.7)
560 FORMAT(21H X TRANSPOSE X MATRIX )
670 FORMAT(25H2 CORRELATION COEFFICIENTS )
700 FORMAT(32H (X TRANSPOSE X) INVERSE MATRIX )
986 FORMAT(39H THE NUMBER OF DEPENDENT VARIABLES WAS 13,83H IT IS BE
XING SET TO ONE AND THE REJECTION OPTiON EXERCISED ON DEPENDENT VAR
XABLE 1 )
1005 FORMAT(39H THERE IS NO EVIDENCE OF A REGRESSION. /)
1282 FORMAT(314,15)
1283 FORMAT(1H 314,15)
1306 FORMAT(40H REPLICATE SETS ARE NOT GROUPED PROPERLY )
ENC

$IBFTC TRANSX

SUBLTLINE TRANS

C******************************************************************************C
CCMCN/BIG/SUMXX(I830),SUMXX(I1830),EIG(I830),SUMXY(60,9),
X (60,91,CRRR(1830)
CCMCN/SMALL/ BYPASS,BZERO,DELETE, FIRST, IFCHI, IFSSR,
X (99), BYPASS,BZERO,DELETE, FIRST, IFCHI, IFSSR,
X IFIT, IFMT, INPUT, INPUT5, INTER,
X ISTRAT, ICOL, KONNO, LENGTH, LIST,
X NERRO, NODEP, NCOB, NOTERM,
X NOWAR, NPDGE, NRES, NTRANS, NWHERE,
X P, PREDCT, REPS, RWT,
X STORYI, STORYC, STORYX, TOTWT, WEIGHT,
X ERRFXC, ECONMY
LOGICAL ECONMY
X IFSSR, IFIT, IFMT, REPS, PREDCT,
X STORYC, STORYX, STORYI, FIRST ,ERRFXD

58
**DOUBLE PRECISION** RWT,TOTWT,WEIGHT

**COMMON/CNTRS/** I, IBC, IC, ICOL, X INEW, INOCH, IOLD, IOUT, IR, X IRC, IREP, IS, ITC, J, X K, KBAR

C

C********************************************************************

C THIS SUBROUTINE PERFORMS TRANSFORMATIONS IF THIS OPTION IS

C REQUESTED.

C

C K TRANSFORMATION SET NUMBER.

C NCON(2*K-1) CONSTANT NUMBER TO USE.

C NCON(2*K) DERIVED CONSTANT.

C NTRAN(K) NUMBER OF TRANSFORMATION REQUESTED.

C NXCOD(K) VARIABLE NUMBER

C

80 DO 500 K=1,NTRANS

I=NCON(2*K-1)

IF(I).EQ.100,100,110

100 CONS=0.0

GO TO 120

110 CONS=CON(I)

120 I=NXCOD(K)

Y=X(I)

MTRAN = NTRAN(K)

IF(MTRAN.LE.0) MTRAN=31

140 GO TO 150,160,170,180,190,200,210,220,230,240,250,260,270,280,290,

A300,310,320,330,340,350,360,370,380,390,400,410,420,430,440,450),

B MTRAN

150 CONS=Y+CONS

GO TO 460

160 CONS=Y*CONS

GO TO 460

170 CONS=CONS/Y

GO TO 460

180 CONS=EXP(Y)

GO TO 460

190 CONS=Y**CONS

GO TO 460

200 CCNS=ALOG(Y)

GO TO 460

210 CCNS=ALOG10(Y)

GO TO 460

220 CCNS=SIN(Y)

GO TO 460

230 CCNS=COS(Y)

GO TO 460

240 CCNS=SIN(3.14159265*(CONS*Y))

GO TO 460

250 CCNS=COS(3.14159265*(CONS*Y))

GO TO 460

260 CONS=1.0/Y

GO TO 460

270 CONS=EXP(CONS/Y)

GO TO 460

280 CONS=EXP(CONS/(Y*Y))

GO TO 460

290 CONS=SQRT(Y)

GO TO 460

300 CONS=1.0/SQRT(Y)

GO TO 460

310 CONS=CONS**Y

GO TO 460

320 CONS=10.0**Y
GO TC 460
330 CONS=SINH(Y)
   GO TO 460
340 CONS=COSH(Y)
   GO TC 460
350 CONS=(1.0-COS(Y))/2.0
   GO TO 460
360 CONS=ATAN(Y)
   GO TC 460
370 CONS=ATAN2(Y/CONS)
   GO TC 460
380 CONS=Y*Y
   GO TC 460
390 CONS=Y*Y*Y
   GO TO 460
400 CONS=ARSIN(SQRT(Y))
   GO TC 460
410 CONS=2.0*3.14159265*Y
   GO TO 460
420 CONS=1.0/(2.0*3.14159265*Y)
   GO TC 460
430 CONS=ERF(Y)
   GO TO 460
440 CONS=GAMMA(Y)
   GO TO 460
450 CONS=Y
460 I=NCON(2*K)
   IF(I)470,470,480
470 CON(K)=CONS
   GO TC 500
480 CON(I)=CONS
   IF(I)500,500,490
490 X(I)=CONS
500 CONTINUE
   RETURN
END
C SUBROUTINE SUMUPS
C
C PURPOSE
C 1) CALCULATE (X TRANSPOSE X) AND (X TRANSPOSE Y) MATRICES ONE
C OBSERVATION AT A TIME.
C 2) COMPUTE TOTAL OF THE WEIGHTS
C ** BOTH CALCULATIONS ARE IN DOUBLE PRECISION
C
C SUBROUTINES NEEDED
C LOC
C
C*****************************************************************************
C*****************************************************************************
C*****************************************************************************
SUBROUTINE SUMUP
COMMON/BIG/SUMXX(1830),EIG(1830),SUMXY(60,9),CORR(1830)
DOUBLE PRECISION SUMXX, SUMXY
COMMON/MED/, R0(9), CON(99), ERRMS(9),
X IDEN(13), IDOUT(60), NCON(200),
X NTERM(60), NTRAN(100), NXCOD(100), POOLED(9),
X XREPVAR(9), RESMS(9), SUMX(69),
X SUMX2(69), X(99), ZEAN(69), SUMY2(9)
DOUBLE PRECISION SUMX, SUMY2
COMMON/SMALL/ BYPASS, BZERO, DELETE, FIRST, IFCHI, IFSSR,
X IFFT, IFWT, INPUT, INPUTS, INTER,
X ISTRAT, JCOL, KONNO, LENGTH, LIST,
X NERROR, NODEP, NOOB, NOTERM,
X NOVAR, NPDEG, NRES, NTRANS, NWHERE,
X P, PREDCT, REPS, RWT,
X STORYI, STORYC, STORYX, TOTWT, WEIGHT,
X ERFXQ, ECONMY
LOGICAL ECONMY
LOGICAL BYPASS, BZERO, DELETE, IFCHI,
X IFSSR, IFFT, IFWT, REPS, PREDCT,
X STORYC, STORYX, STORYI, FIRST, ERFXQ
DOUBLE PRECISION RWT, TOTWT, WEIGHT
COMMON/CNTRS/ I, IBC, IC, ICOL,
X INEW, INOCH, IOLD, IDOUT, IR,
X IRC, IREP, IS, ITC, J,
X K, KBAR
C*****************************************************************************
C*****************************************************************************
C*****************************************************************************
DO 110 I=1,JCOL
SUMX(I)=SUMX(I)+X(I)*WEIGHT
110 CONTINUE
DO 100 K=1,NOTERM
DO 90 J=1,NODEP
KBAR=J+NOTERM
SUMXY(K,J)=SUMXY(K,J) + X(K)*X(KBAR)*WEIGHT
90 CONTINUE
DO 50 I=1,K
C CALL LOC(K,I,IR)
SUMXX(IR) = SUMXX(IR) + X(I)*X(K)*WEIGHT
50 CONTINUE
DO 100 J=1,NODEP
KBAR=NOTERM + J
15 SUMY2(J)=SUMY2(J) + X(KBAR)*WEIGHT
100 CONTINUE
TOTWT=TOTWT+WEIGHT
RETURN
END
SUBROUTINE MFIX
C THIS ROUTINE USES THE SUMXX MATRIX COMPUTE IN DOUBLE PRECISION
C AND THE SUMX ARRAY COMPUTED IN DOUBLE PRECISION TO COMPUTE THE
C SUMXX DEVIATIONS FORM OF X TRANSPOSE X IN DOUBLE PRECISION AND
C THE RESULT IS TRUNCATED TO SINGLE.
C THE MATRICES ARE ALSO PRINTED
C
C*****************************************************************************
COMMON/BIG/SUMXX(1830),EIG(1830),SUMXY(60,9),CORR(1830)
DOUBLE PRECISION SUMXX,SUMXY
COMMON/MED/ 80(9), CON(99), ERRMS(9),
X ICENT(13),IODUT(60),NCON(200),
X NTERM(60),NTRAN(100),NXCOD(100), POOLED(9),
XREPVAR(9), RESMS(9), SUMX(69),
XSUMX(69), X(30), ZEAN(69), SUMY(9),
DOUBLE PRECISION SUMX,SUMX2, SUMY2
COMMON /FRMTS/ FMT(13),FMTTRI(14)
COMMON /SMALL/ BYPASS,BZERO,DELETE, FIRST, IFCH1, IFSSR,
X IFTT, IFWT, INPUT, INPUT5, INTER,
X ISTRAT, JCOL, KCONNO, LENGTH, LIST,
X NERROR, NNODE, NOOB, NOTERM,
X NOVAR, NPDEG, NRES, NTRANS, NWHERE,
X P, PREDCT, REPS, RWT,
X STORYI, STORYC, STORYX, TOTWT, WEIGHT,
X ERRFXD, ECONMY
LOGICAL ECONMY
DOUBLE PRECISION RWT,TOTWT,WEIGHT
LOGICAL BYPASS, BZERO, DELETE, IFCHI,
X XIFSSR, IFTT, IFWT, REPS, PREDCT,
XSTORYC, STORYXY, STORY1, FIRST, ERRFXD
COMMON/CNTRS/ I, IBC, IC, ICOL,
X INEW, INOCH, IOLD, IOUT, IR,
X IRC, IREP, IS, ITCLJ,
X K, KBAR
EQUIVALENCE (EIG,SONGLX),(EIG170),RITEXY),(SONGLXX,SUMXX),(SONGLXY,
X SUMXY),(SONGLX2,SUMX2),(SONGLY2,SUMY2)
DIMENSION SONGLX(69),SONGLXY(60,9),SONGLXX(1830),RITEXY(60,9)
X ,SONGLX2(69),SONGLY2(9)
C*****************************************************************************
C DO 10 J=1,JCOL
10 SONGLX(J)= SONGL(SUMX(J))
IF(ECONMY) GO TO 500
WRITE(LIST,530)
WRITE (LIST,540) (SONGLX(I),I=1,JCOL)
WRITE(LIST,560)
DO 20 I=1,LENGTH
20 CORR(I)= SONGL(SUMXX(I))
CALL TRIANG(CORR,NOTERM,FMTTRI)
DO 30 I=1,NOTERM
DO 30 J=1,NODEP
30 RITEXY(I,J)= SONGL(SUMXY(I,J))
WRITE(LIST,565)
CALL RECT(NOTERM,NODEP,60,9,RITEXY,FMTTRI)
C*****************************************************************************
C COMPUTE AND PRINT MEANS, COMPUTE AND PRINT THE(X TRANSPOSE X)
C MATRIX IN TERMS OF DEVIATIONS FROM MEAN. THE DEVIATIONS FORM
C OF (X T X) IS THE VARIANCE-COVARIANCE MATRIX OF THE
C INDEPENDENT VARIABLES.
500 CONTINUE
   RWT=1.000/TOTWT
   DO 570 I=1,JCOL
570  ZEAN(I)=SUMX(I)*RWT
      WRITE(LIST,580)
      WRITE(LIST,540) (ZEAN(I),I=1,JCOL)
      IR = 0
      DO 600 J=1,NOTERM
      IR=IR + J
      IF(.NOT.BZERO) GO TO 601
      SUMXZ(J)=SUMXX(IR)-SUMX(J)**2*RWT
      GO TO 600
601  SUMXZ(J)=SUMXX(IR)
600  CONTINUE
602  CONTINUE
   IR=1
   DO 620 J=1,NOTERM
      DO 618 K=1,NODEP
      IF(BZERO) GO TO 617
      SNGLXY(J,K)=SNGL(SUMXY(J,K))
      GO TO 618
617  KBAR=NOTERM+K
      SNGLXY(J,K)=SNGL(SUMXY(J,K)-SUMX(J)*SUMX(KBAR)*RWT)
618  CONTINUE
619  IR=IR+1
   DO 640 J=1,NOTERM
      DO 650 J=1,NODEP
      IF(BZERO) GO TO 645
      SNGLY2(J)=SNGL(SUMY2(J))
      GO TO 650
645  K=NOTERM+J
      SNGLY2(J)=SNGL(SUMY2(J)-SUMX(K)**2*RWT)
650  CONTINUE
C******************************************************************************
IF(ECONMY) GO TO 622
   IF(.NOT.BZERO) GO TO 621
      WRITE(LIST,625)
      CALL TRIANG(SNGLXX,NOTERM,8,FMTTRI)
      WRITE(LIST,630)
      CALL RECT(NOTERM,NODEP,60,9,SNGLXY,FMTTRI)
621  WRITE(LIST,670)
      CALL TRIANG(CORR,NOTERM,8,FMTTRI)
622  CONTINUE
   DO 640 J=1,NOTERM
540  SNGLX2(J)=SNGL(SUMX2(J))
   DO 650 J=1,NODEP
      IF(BZERO) GO TO 645
      SNGLY2(J)=SNGL(SUMY2(J))
      GO TO 650
645  K=NOTERM+J
      SNGLY2(J)=SNGL(SUMY2(J)-SUMX(K)**2*RWT)
650  CONTINUE
C******************************************************************************
RETURN
C
530 FORMAT(1HO,32H SUMS OF INDEP AND DEP VARIABLES )
540 FORMAT(1H 8G15.7)
560 FORMAT(2IH2X TRANPOSE X MATRIX   )
565 FORMAT(2IH2X TRANPOSE Y MATRIX   )
580 FORMAT(33H MEANS OF INDEP AND DEP VARIABLES )
625 FORMAT(53H2X TRANPOSE X MATRIX WHERE X IS DEVIATION FROM MEAN )
630 FORMAT(60H2X TRANPOSE Y MATRIX WHERE X AND Y ARE DEVIATIONS FROM XMEAN )
670 FORMAT(25H2CORRELATION COEFFICIENTS )
END
SUBROUTINE MATINV

C PURPOSE
C  1) COMPUTE EIGENVALUES AND EIGENVECTORS OF (X-TRANSPOSE X)
C     AND/OR CORRELATION MATRIX IF REQUESTED.
C     STORYC = .TRUE. IF FOR CORRELATION
C     STORYX = .TRUE. IF FOR XTX
C  2) INVERT CORRELATION COEFFICIENT MATRIX BY EITHER BORDERING
C     OR GAUSS ELIMINATION
C  3) COMPUTE PRODUCT OF CORRELATION AND INVERTED CORRELATION
C     MATRIX IF REQUESTED
C     STORYI = .TRUE. IF THE PRODUCT IS TO BE PRINTED
C  4) COMPUTE (X TRANSPOSE X) INVERSE FROM INVERTED CORRELATION
C     MATRIX
C  5) COMPUTE REGRESSION COEFFICIENTS
C  6) COMPUTE OTHER REGRESSION STATISTICS

C SUBROUTINES NEEDED
C BORD
C LOC
C EIGEN
C INVXX
C RECT
C RSTATS
C TRIANG

C REMARKS
C THE EIGENVALUES ARE COMPUTED AS AN AID IN DETERMINING THE
C CONDITION OF THE SYSTEM OF EQUATIONS FOR THE REGRESSION
C COEFFICIENTS. EXAMINATION OF THEM AND THEIR ASSOCIATED
C EIGENVECTORS MAY SHOW THAT CERTAIN SETS OF INDEPENDENT
C VARIABLES ARE HIGHLY CORRELATED AND NOT EASILY LIABLE TO
C INDEPENDENT STUDY.

C SUBROUTINE MATINV

COMMON/BIG/SUMXX(1830), SUMXX(1830), EIG(1830), SUMXY(60, 9),
 X B(60, 9), CORR(1830)
 COMMON/MED/ B0(9), CON(99), ERRMS(9),
 X IDENT(13), IDOUT(60), NCON(200),
 X NTERM(60), TRAN(100), NXCOD(100), POOLED(9),
 XREPVAR(9), X REEM(138), RESMS(9), SUMX(138),
 XSUMX(69), X(99), ZEAN(69), SUMY(18),
 COMMON/FRMTR/FMT(13), FMTTRI(14)
 COMMON/SMALL/ BYPASS, BZERO, DELETE, FIRST, IFCHI, IFSSR,
DIMENSION A(1), C(1), XTX(3), CMAT(3), HOL(3)

EQUIVALENCE (SUMXX, A), (SUMXX, C)

DATA (XTX(1), I=1,3) /6HX TRANS, 6HPOSE, 6HX /
DATA (CMAT(I), I=1,3) / 6HCORREL, 6HATION, 1H /

C
C******************************************************************************
C ORDER= NOTERM
IF(NCTERM-1) 10,10.12
10 SUMXX(I)= 1.0/SUMXX(I)
GO TO 350
C
C TRANSFER CORR TO A FOR INVERSION
C AT THIS POINT THE INFORMATION IN A MAY BE USED FOR ANY DESIRED
C CALCULATIONS --- EIGENVALUES,RANK,ETC.
C JUST PUT CORR INTO A BEFORE PROCEEDING TO REMAINDER OF ROUTINE
C
C
12 BYPASS= .FALSE.
IF(.NCT. STORYX) GO TO 30
15 DO 14 I=1,3
14 HOL(I)=XTX(I)
16 CALL EIGEN(A, SUMXX, IORDER, 0)
WRITE(LIST,17) (HOL(I), I=1,3)
J=0
DO 18 I=1, IORDER
J=J+1
18 A(I)=A(J)
WRITE(LIST,19) (A(I), I=1, IORDER)
WRITE(LIST,20)
CALL RECT(IORDER, IORDER, IORDER, SUMXX, FMTRI)
30 DO 35 I=1, LENGTH
35 A(I)=CORR(I)
IF(BYPASS) GO TO 49
IF(.NCT. STORYC) GO TO 49
36 DO 38 I=1, 3
38 HOL(I)= CMAT(I)
BYPASS= .TRUE.
GO TO 16
C
C******************************************************************************
C NO SUBMODELS TO ANALYZE SO INVERT A DIRECTLY BY GAUSS
C
49 IF(IFSSR) GO TO 50
CALL INVXTX(A,NOTERM,D,1.0)
GO TO 60
C
C**************************************************************************************
C SUBMODELS HAVE BEEN REQUESTED SO WE USE BORDERING
50 IORDER=0
55 IORDER=IORDER +1
CALL BORD(IORDER,A)
60 IF(.NOT.STORY1) GO TO 200
C
C**************************************************************************************
C WRITE INVERSE OF CORRELATION COEFFICIENT MATRIX
WRITE(LIST,65)
CALL TRIANG(A,IORDER,R,FMTTRI)
C COMPUTE A TIMES A INVERSE AND WRITE IT
WRITE(LIST,70)
ITC= 0
DO 150 IC=1,IORDER
DO 150 IR=1,IC
ITC= ITC+1
C(ITC)= 0.0
DO 130 IR=1,IORDER
CALL LOC(IR,I,IRC)
CALL LOC(I,IC,IBC)
C(ITC)= C(ITC) + A(IRC)*CORR(IBC)
130 CONTINUE
150 CONTINUE
C
CALL TRIANG(C,IORDER,R,FMTTRI)
C
200 CONTINUE
C
C**************************************************************************************
C COMPUTE SUMXXI FROM CORR INVERSE. SUMXXI TIMES THE ERROR MEAN SQUARE IS THE
C VARIANCE-COVARIANCE MATRIX OF THE REGRESSION
C COEFFICIENTS.
IR=0
DO 340 I=1,IORDER
DO 340 J=1,I
IR= IR+1
SUMXXI(IR) = A(IR)/SQRT(SUMX2(I)*SUMX2(J))
340 CONTINUE
C
C**************************************************************************************
C COMPUTE COEFFICIENTS AND PRINT THEM
C
350 DO 370 J=1,NODEP
DO 370 K=1,IORDER
B(K,J)=0.0
DO 370 L=1,IORDER
CALL LOC(L,K,IR)
B(K,J) = B(K,J) + SUMXXI(IR)*SUMXY(L,J)
370 CONTINUE
C
WRITE(LIST,380) IDENT
WRITE(LIST,382)
IF(NOT.BZERO) GO TO 400
DO 390 J=1,NODEP
SUM=0.0
KBAR= NOTERM + J
DO 385 K=1,IORDER
SUM = SUM + B(K,J)*ZEAN(K)
385 CONTINUE
B0(J)= ZEAN(KBAR) -SUM
390 CONTINUE
WRITE(LIST,395)
WRITE(LIST,397) (B0(K),K=1,NODEP)
400 WRITE(LIST,410)
DO 430 J=1,IORDER
WRITE(LIST,432) IDOUT(J),(B(J,K),K=1,NODEP)
430 CONTINUE
C*************************************************************************
C COMPUTE REGRESSION STATISTICS IN RSTATS
C
CALL RSTATS(IORDER)
C*************************************************************************
C IF IORDER IS LESS THAN NOTERM WE HAVE USED THE BORDERING OPTION
C AND MUST GO BACK TO FINISH
C
IF(IORDER-NOTERM) 55,500
500 STORYC=.FALSE.
STORYX=.FALSE.
RETURN
17 FORMAT(34H2THE FOLLOWING ARE EIGENVALUES OF 2A6,A1, 7H MATRIX)
19 FORMAT(1H 8G16.7)
20 FORMAT(132H2THIS IS THE MODAL MATRIX OR MATRIX OF EIGENVECTORS. EIGENVECTORS ARE WRITTEN IN COLUMNS LEFT TO RIGHT IN SAME ORDER AS EIGENVALUES )
65 FORMAT(14H2CORR INVERSE )
70 FORMAT(118H2AS A PARTIAL CHECK ON INVERSION ACCURACY THE (CORR)*(C XORR INVERSE) MATRIX follows. It should be the identity matrix. )
380 FORMAT(1HL,13A6,1A2)
382 FORMAT( 61H EACH COLUMN CONTAINS THE COEFFICIENTS FOR ONE DEPENDENT XT TERM )
395 FORMAT(20H CONSTANT TERM (BO) )
397 FORMAT(4X,9G14.6)
410 FORMAT(36H REGRESSION COEFFICIENTS (B1,...,BK) )
432 FORMAT(1H 13,9G14.6)
END
SUBROUTINE RSTATS( ORDER)

COMMON /BIG/ SUMXX(1830), SUMXY(1830), EIG(1830), SUMXY(60, 9),
X B(60, 9), CORR(1830)
COMMON /MED/ B0(9), CON(99), ERRMS(9),
X IDENT(13), IDOUT(60), NCON(200),
X NTERM(60), NTRAN(100), NCOD(100), POOLED(9),
X REPVAR(9), RESMS(9), SUMX(138),
X SUMX2(69), X(99), ZEAN(69), SUMY2(10),
COMMON /FRMTS/ FMT(13), FMTTRI(14)
COMMON /SMALL/ BYPASS, BZERO, DELETE, FIRST, IFCHI, IFSSR,
X IFFT, IFWT, INPUT, INPUTS, INTER,
X ISTAT, JCOL, LENNO, LENGTH,
X NERRR, NODEP, NOOB, NOTERM,
X NOVAR, NPDIG, NRES, NTRANS, NWHERE,
X P, PREDCT, REPS, RWT,
X STORYI, STORY X, STORY Y, TOTWT,
X ERRFXD, ECONMY
LOGICAL ECONMY
LOGICAL SATRD
LOGICAL BYPASS, BZERO, DELETE, IFCHI,
X IFSSR, IFFT, IFWT, REPS, PREDCT,
X STORYC, STORY X, STORY Y, TOTWT, WEIGHT
DOUBLE PRECISION RWT, TOTWT, WEIGHT
COMMON / CNTRS/ I, IBC, IC, ICOL,
X INEW, IINCH, IOLD, IOUT, IR,
X IRC, IREP, IS, ITC, J,
X K, KKRAR

C*****************************************************************************
C PROGRAMMING NOTE*************
C SOME OF THESE EQUIVALENCES ARE USED TO COMMUNICATE WITH
C SUBROUTINE PREDCT. BE CAREFUL ABOUT CHANGES INVOLVING THE
C ARRAY EIG
C DIMENSION SSQREG(9), SSQRES(9), REGMS(9),
X XLOF(9), XLOFMS(9), FRATIO(9), RSQD(9), R9,
X SQLST(9), DEVR(60, 9)
EQUIVALENCE (EIG(1), SSQREG), (EIG(10), SSQRES),
X (EIG(19), REGMS), (EIG(37), XLOF),
X (EIG(46), XLOFMS), (EIG(55), FRATIO),
X (EIG(64), RSQR), (EIG(73), R),
X (EIG(82), SSQSLT), (EIG(91), DEVB).

C*****************************************************************************
C COMPUTE DEGREES OF FREEDOM AND RECIPROCALS
C*****************************************************************************
NREG = IORDER
NTOT = IFIX(TOTWT) - 1
IF (.NOT. AZERO) NTOT = NTOT + 1
NRES = NTOT - NREG
NLOF = NRES - NPD
RNREG = 1.0/FLOAT(NREG)
IF (NRES.EQ.0) GO TO 980
RNRES = 1.0/FLOAT(NRES)
SATRTD = .FALSE.
IF (NLOF.EQ.0) GO TO 90
RNLOF = 1.0/FLOAT(NLOF)
GO TO 100
90 SATRTD = .TRUE.
100 CONTINUE
NXTERM = IORDER
RNOOB = RWT

C*****************************************************************************
C COMPUTE RESIDUAL SUM OF SQUARES, RESIDUAL VARIANCE, VARIANCE
C FROM REPLICATIONS IF APPLICABLE, AND THE F-RATIO OF MEAN SQUARE
C LACK-OF-FIT AND MEAN SQUARE RESIDUALS.
C*****************************************************************************
DO 210 J = 1, NODEP
SSQREG(J) = 0.0
DO 210 I = 1, NXTERM
SSQREG(J) = SSQREG(J) + B(I, J)*SUMXY(I, J)
210 CONTINUE
SSQRES(J) = SUMY2(J) - SSQREG(J)
REGMS(J) = SSQREG(J) * RNREG
RESMS(J) = SSQRES(J) * RNRES
RSQR(J) = SSQREG(J) / SUMY2(J)
R(J) = SQRT(RSQD(J))
IF (.NOT. REPS) OR SATRTD) GO TO 210
XLOF(J) = SSQRES(J) - POOLED(J)
XLOFMS(J) = XLOF(J) * RNLOF
FRATIO(J) = XLOFMS(J) / REPVAR(J)
210 CONTINUE

C*****************************************************************************
C DETERMINE WHICH ESTIMATE OF SIGMA SQUARED SHOULD BE USED IN
C HYPOTHESIS TESTS. PUT THE PROPER ONE IN ERRMS AND SET ERRFXD
C TO TRUE IF THE PRESENT VALUE IS TO BE USED FOR ALL FOLLOWING
C TESTS AND T-STATISTICS.
C*****************************************************************************
IOUT = ISTAT
IF (ERRFXD) GO TO 250
IF (ISTAT .NE. 3) GO TO 214
211 DO 213 J = 1, NODEP
213 ERRMS(J) = RESMS(J)
NERRO = NRES
980 CONTINUE
IOUT=3
GO TO 250
214 IF(ISTRAT.NE.1) GO TO 218
   IF(.NOT.REPS) GO TO 211
   DO 215 J=1,NODEP
215 ERRMS(J)= REPVAR(J)
   NERROR= NPDEC
   ERRFXD= .TRUE.
   IOUT=1
   GO TO 250
218 IF(FIRST.AND.(ORDER.EQ.NOTERM)) GO TO 220
   GO TO 211
220 ERRFXD= .TRUE.
   DO 222 J=1,NODEP
222 ERRMS(J)= RESMS(J)
   NERRCR= NRES
   IOUT=2
C
C************************************************************************************************************
C WRITE ANOVA TABLES
250 IS=2
   IF(REPS) IS=4
   DO 500 J=1,NODEP
      IF(ERRMS(J).EQ.0.0) ERRMS(J)=1.0E-30
      WRITE(LIST,1001) IS,J
      WRITE(LIST,1002)
      WRITE(LIST,1003) SSQREG(J), NREG, REGMS(J)
      WRITE(LIST,1004) SSQRES(J), NRES, RESMS(J)
      WRITE(LIST,1005)
      WRITE(LIST,1006) SUMY2(J), NTOT
      WRITE(LIST,1007)
      WRITE(LIST,1500) RSQD(J), R(J)
      STD=SQR(REGMS(J))
      WRITE(LIST,1600) STD
      WRITE(LIST,1700) IOUT,ERRMS(J), NERROR
      F=REGMS(J)/ERRMS(J)
      WRITE(LIST,1750) F, NREG, NERROR
      IF(.NOT.REPS).OR.SATRTD) GO TO 500
      WRITE(LIST,2001)
      WRITE(LIST,1002)
      WRITE(LIST,2005) XLOF(J), NLOF, XLOFSM(J)
      WRITE(LIST,2006) POOLED(J), NPDEC, REPVAR(J)
      WRITE(LIST,1004) SSQRES(J), NRES, RESMS(J)
      WRITE(LIST,1005)
      WRITE(LIST,2008) FRATIO(J)
      WRITE(LIST,1007)
      500 CONTINUE
C
C**************************************************************************************************************
C COMPUTE CONTRIBUTION OF EACH INDEPENDENT VARIABLE TO REG SUM
C OF SQUARES AS IF IT WERE LAST TO ENTER
C WRITE(LIST,370)
   IR= 0
   DO 8635 K=1,NXTERM
      IR= IR+K
   DO 8632 J=1,NODEP
8632 SSQLST(J)= B(K,J)**2/SUMXXI(IR)
WRITE(LIST,380) IDOUT(K),(SSQLST(J),J=1,NODEP)
8635 CONTINUE
C
C******************************************************************************
C COMPUTE STANDARD DEVIATION OF REGRESSION COEFFICIENTS
C******************************************************************************
WRITE(LIST,375)
IF(.NOT.BZERO) GO TO 959
DO 910 J=1,NXTERM
R(J)=0.0
DO 910 I=1,NXTERM
CALL LOC(I,J,IR)
R(J)=R(J)+ZEAN(I)*SUMXXI(IR)
910 CONTINUE
XXT=0.0
DO 920 J=1,NXTERM
XXT=XXT+ZEAN(J)*R(J)
920 CONTINUE
IR=0
DO 970 J=1,NXTERM
IR= IR+J
DO 960 K=1,NODEP
DEVB(J,K)=SQRT(ERRMS(K)*(RNOOB+XXT))
960 CONTINUE
WRITE(LIST,380) IDOUT(J),(DEVB(J,K),KR=1,NODEP)
970 CONTINUE
C
C******************************************************************************
C******************************************************************************
C FORMAT
1001 FORMAT(11,42H ANOVA OF REGRESSION ON DEPENDENT VARIABLE I5) 195
1002 FORMAT(1H 79(1H*)/79H /1H 79(1H*)) 196
1003 FORMAT(17H REGRESSION G20.8, 5X,I10,5X,G20.8) 198
1004 FORMAT(17H RESIDUAL G20.8, 5X,I10,5X,G20.8) 199
1005 FORMAT(1H 79(1H*)) 200
1006 FORMAT(17H TOTAL G20.8, 5X,I10) 201
1007 FORMAT(1H 79(1H*)) 202
2001 FORMAT(1X/1X/22H ANOVA OF LACK OF FIT ) 203
2005 FORMAT(17H LACK OF FIT G20.8, 5X,I10,5X,G20.8) 204
2006 FORMAT(17H REPLICAATION G20.8, 5X,I10,5X,G20.8) 205
2008 FORMAT(28H F = MS(LDF)/MS(REPS) = F10.3 ) 206
370 FORMAT(74H SUYS OF SQUARES DUE TO EACH VARIABLE IF IT WERE LAST T
207 XD ENTER REGRESSION ) 208
375 FORMAT(115H2 STANDARD DEVIATION OF REGRESSION COEFFICIENTS (DERIVE
209 XD FROM DIAGONAL ELEMENTS OF (X TRANSPOSE X)INVERSE MATRIX )) 210
380 FORMAT(1H I3,9G14.6) 211
1500 FORMAT(4OH R SQUARED = SSQ(REG) / SSQ(TOT) = F8.6, 212
X 5X, 4HR = F7.6) 213
1600 FORMAT(34H STANDARD ERROR OF ESTIMATE G14.6) 214
1700 FORMAT(24H USING POOLING STRATEGY I2,25H THE ERROR MEAN SQUARE = 215
X G14.7, 26H WITH DEGREES OF FREEDOM = I6) 216
1750 FORMAT(5X,19HF=MS(REG)/MS(ERR)= F6.2,5X,13HCOMPARE TO F(I2,1H,I3,1 217
XH)) 218
RETURN
980 WRITE(LIST,981)
981 FORMAT(41H ZERO RESIDUAL DEGREES OF FREEDOM. STOP. ) 219
STOP 220
END 221
71
SUBROUTINE TTEST

PURPOSE
COMPUTE THE T-STATISTICS FOR EACH REGRESSION TERM AND ITS TWO TAILED SIGNIFICANCE LEVEL. THEN DETERMINE THE TERM WITH THE LEAST SIGNIFICANCE AND RETURN THIS INFORMATION TO RAPIER.

SUBROUTINE TTEST(*)

COMMON/BIG/SUMX(1830),SUMXXI(1830),EIG(1830),SUMXY(60,9),
X B(60,9),CORR(1830)
COMMON/MED/ BO(9), CON(99), ERRMS(9),
X IDENT(13),IDOUT(60),NCON(200),
X NTERM(60),NTRAN(100),NXCOD(100),POOLED(9),
XREPVAR(9), XRESMS(9), SUMX(138),
XSUMX2(69), X(99), ZEAN(69), SUMY(181)
COMMON/SMALL/ BYPASS,BZERO,DELETE, FIRST, IFCHI, IFSSR,
X IFIT, IFWT, INPUT, INPUTS, INTER,
X ISTRAT, JCOL, KONNO, LENGTH, LIST,
X VERROR, NNODEP, NOOB, NOTERM,
X NOVAR, NPDEG, NRES, NTRANS, NWHERE,
X P, PREDCT, REPS, RWT,
X STORY, STORYC, STORYX, TOTWT, WEIGHT,
X ERRFXD, ECONMY,
LOGICAL ECONMY
LOGICAL BYPASS, BZERO, DELETE, IFCHI,
XIFSSR, IFIT, IFWT, REPS, PREDCT,
XSTORY, STORYX, STORYI, FIRST, ERRFXD
DOUBLE PRECISION WEIGHT,RWT
COMMON/CNTRS/ I, IBC, IC, ICOL,
X INEW, INOCH, IOLD, IOOUT, IR,
X IRC, IREP, IS, ITC, J,
X K, KBAR

LOGICAL MAKENU,NOZERO
DIMENSION T(35,13),PLEVEL(13)
DIMENSION DEVB(60,9),PROB(60,9),TT(60,9)
EQUIVALENCE (EIG(91),DEVB,TT), (EIG(650),PROB)
EQUIVALENCE (P,PWANT)

DATA (PLEVEL(JJ),JJ=1,13)/0.10,0.20,0.30,0.40,0.50,0.60,0.70,
10.80,0.90,0.95,0.98,0.99,0.999 /
DATA (T(1,JJ),JJ=1,13)/0.158,0.325,0.510,0.727,1.000,1.376,
<table>
<thead>
<tr>
<th>DATA</th>
<th>DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.963</td>
<td>3.0786</td>
</tr>
<tr>
<td>(T1(2,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.386</td>
<td>1.886</td>
</tr>
<tr>
<td>(T1(3,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.250</td>
<td>1.638</td>
</tr>
<tr>
<td>(T1(4,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.190</td>
<td>1.533</td>
</tr>
<tr>
<td>(T1(5,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.156</td>
<td>1.476</td>
</tr>
<tr>
<td>(T1(6,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.134</td>
<td>1.440</td>
</tr>
<tr>
<td>(T1(7,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.119</td>
<td>1.415</td>
</tr>
<tr>
<td>(T1(8,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.108</td>
<td>1.397</td>
</tr>
<tr>
<td>(T1(9,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.100</td>
<td>1.383</td>
</tr>
<tr>
<td>(T1(10,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.093</td>
<td>1.372</td>
</tr>
<tr>
<td>(T1(11,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.088</td>
<td>1.363</td>
</tr>
<tr>
<td>(T1(12,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.083</td>
<td>1.356</td>
</tr>
<tr>
<td>(T1(13,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.079</td>
<td>1.350</td>
</tr>
<tr>
<td>(T1(14,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.076</td>
<td>1.345</td>
</tr>
<tr>
<td>(T1(15,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.074</td>
<td>1.341</td>
</tr>
<tr>
<td>(T1(16,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.071</td>
<td>1.377</td>
</tr>
<tr>
<td>(T1(17,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.069</td>
<td>1.333</td>
</tr>
<tr>
<td>(T1(18,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.067</td>
<td>1.330</td>
</tr>
<tr>
<td>(T1(19,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.066</td>
<td>1.328</td>
</tr>
<tr>
<td>(T1(20,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.064</td>
<td>1.325</td>
</tr>
<tr>
<td>(T1(21,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.063</td>
<td>1.323</td>
</tr>
<tr>
<td>(T1(22,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.061</td>
<td>1.312</td>
</tr>
<tr>
<td>(T1(23,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.060</td>
<td>1.319</td>
</tr>
<tr>
<td>(T1(24,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.059</td>
<td>1.318</td>
</tr>
<tr>
<td>(T1(25,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.058</td>
<td>1.316</td>
</tr>
<tr>
<td>(T1(26,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.058</td>
<td>1.315</td>
</tr>
<tr>
<td>(T1(27,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.057</td>
<td>1.314</td>
</tr>
<tr>
<td>(T1(28,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.056</td>
<td>1.313</td>
</tr>
<tr>
<td>(T1(29,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
<tr>
<td>1.056</td>
<td>1.312</td>
</tr>
<tr>
<td>(T1(30,J1),J1,J1=1,13)</td>
<td>/</td>
</tr>
</tbody>
</table>
The table below provides the t-statistic values for different degrees of freedom and probability levels:

<table>
<thead>
<tr>
<th>JJ</th>
<th>Probability Level</th>
<th>JJ</th>
<th>Probability Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.10</td>
<td>8</td>
<td>0.80</td>
</tr>
<tr>
<td>2</td>
<td>0.20</td>
<td>9</td>
<td>0.90</td>
</tr>
<tr>
<td>3</td>
<td>0.30</td>
<td>10</td>
<td>0.95</td>
</tr>
<tr>
<td>4</td>
<td>0.40</td>
<td>11</td>
<td>0.98</td>
</tr>
<tr>
<td>5</td>
<td>0.50</td>
<td>12</td>
<td>0.99</td>
</tr>
<tr>
<td>6</td>
<td>0.60</td>
<td>13</td>
<td>0.999</td>
</tr>
<tr>
<td>7</td>
<td>0.70</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The t-statistic at the tabulated degrees of freedom and at the tabulated probability levels (JJ) is the t-statistic at the tabulated degrees of freedom, except for JJ = 31 is for 40 degrees, JJ = 32 is for 60, JJ = 33 is for 120, JJ = 34 is for infinity.

The t-statistic can be used to test the net regression coefficients B(I).

Calculating t-statistics:

1. Calculate the t-statistics for each degree of freedom.
2. Write the calculated t-statistics.
3. Search the table of tabulated degrees of freedom.
4. If the degree of freedom is not found, use the next closest value.
5. Continue searching until the degree of freedom is found.
6. Output the t-statistic.

The t-statistic can be used to test the net regression coefficients B(I).
MAKENU=.TRUE.
320 II=31
GO TO 400
330 IF(NCEG=60)340,350,360
340 FINV=1.0/60.0
FM1INV=1.0/40.0
MAKENU=.TRUE.
350 II=32
GO TO 400
360 IF(NCEG=120)370,380,390
370 FINV=1.0/120.0
FM1INV=1.0/60.0
MAKENU=.TRUE.
380 II=33
GO TO 400
390 II=34
FINV=0.0
FM1INV=1.0/120.0
MAKENU=.TRUE.

C
C
400 WRITE(LIST,410)
410 FORMAT(104H UNDER NULL HYPOTHESIS THE INTERVAL (-T,T) WHERE T IS G
XIVEN ABOVE, HAS APPROX PROBABILITY LISTED BELOW. /42H MINUS S
XIGN INDICATES PROB EXCEEDS .999. )
IF(.NOT.MAKENU) GO TO 430
FNDEG=NDEG
DO 420 JJ=1,13
T(35,JJ)=T(II,JJ)+((1.0/FNDEG - FINV)/(FM1INV-FINV))*T(II-1,JJ)
1-T(II,JJ)
420 CONTINUE
II=35
430 DO 560 J=1,NOTERM
DO 540 K=1,NDEP
DO 440 JJ=1,13
IF(T(II,JJ)-TT(J,J)K)440,450,460
440 CONTINUE
PROB(J,K)=-0.999
GO TO 540
450 PROB(J,K)=PLEVEL(JJ)
GO TO 540
460 IF(JJ.LE.9) GO TO 470
JJ1=JJ-2
JJ2=JJ-1
JJ3=JJ
GO TO 490
470 IF(JJ.LE.4)GO TO 480
JJ1=JJ-1
JJ2=JJ
JJ3=JJ+1
GO TO 490
480 JJ1=JJ
JJ2=JJ+1
JJ3=JJ+2
C
C PERFORM A THREE-POINT LAGRANGE INTERPOLATION
C 490 X=ALOG(TT(J,K))
X1=ALOG(T(I1,J1))
X2=ALOG(T(I2,J2))
X3=ALOG(T(I3,J3))
IF(TT(J,K).LE.1.0) GO TO 500
Y1=ALOG(1.0-PLEVEL(JJ1))
Y2=ALOG(1.0-PLEVEL(JJ2))
Y3=ALOG(1.0-PLEVEL(JJ3))
GO TO 510
500 Y1=ALOG(PLEVEL(JJ1))
Y2=ALOG(PLEVEL(JJ2))
Y3=ALOG(PLEVEL(JJ3))
GO TO 510
510 PROB(J,K)=( (X-X2)*(X-X3)*Y1)/((X1-X2)*(X1-X3)) + ((X-X1)*(X-X3)*Y2)/((X2-X1)*(X2-X3)) + ((X-X1)*(X-X2)*Y3)/((X3-X1)*(X3-X2))
IF(TT(J,K).LE.1.0) 520, 520, 530
520 PROB(J,K)=EXP(PROB(J,K))
GO TO 540
530 PROB(J,K)=1.0-EXP(PROB(J,K))
540 CONTINUE
C******************************************************************
C WRITE THE PROBABILITIES (1.0-ALPHA)
C WRITE(LIST,550) IDOUT(J),(PROB(J,K),K=1,NODEP)
550 FORMAT(13X,9(8X,F6.3))
560 CONTINUE
C******************************************************************
C LIST THE DESIRED VALUE OF PROBABILITY (PWANT)
570 PERCEN=PWANT*100.0
WRITE(LIST,580) PERCEN
580 FORMAT(1HO,36HTHE DESIRED VALUE OF PROBABILITY IS ,F5.1, 8H PERCENT)
C DELETE THE TERM WITH THE LOWEST COMPUTED PROBABILITY IF THAT
C PROBABILITY IS LESS THAN THAT DESIRED (PWANT)
C IF(.NOT.DEDELETE) GO TO 660
IOUT=0
590 AMIN=PWANT
GO TO 620, J=1, NOTERM
600 AMIN=ABS(PROB(J,1))-AMIN
IOUT=J
610 AMIN=ABS(PROB(J,1))
CONTINUE
620 AMIN=ABS(PROB(J,1))
GO TO 670
630 WRITE(LIST,650) IDOUT(IOUT)
640 FORMAT(1HO,11HTHE TERM X(I2,18H) IS BEING DELETED )
GO TO 670
C ALL VARIABLES REMAINING HAVE BEEN CONCLUDED SIGNIFICANT
660 RETURN
670 RETURN
END
LIBFTC PRECIX

SUBROUTINE PREDIC

PURPOSE

1) READ INPUT LEVELS OF INDEPENDENT VARIABLES AND COMPUTE

A PREDICTED RESPONSE FROM THE ESTIMATED REGRESSION EQUATION.

2) COMPUTE VARIANCE AND STANDARD DEVIATION OF THE PREDICTED

MEAN VALUE AND A SINGLE FURTHER OBSERVATION.

SUBROUTINES NEEDED

TRANS

LOC

REMARKS

VALUES FOR DEPENDENT VARIABLES ARE NOT NECESSARY FOR THE

PREDICTING OF VALUES. HOWEVER, A DUMMY VALUE MAY NEED TO

BE SUPPLIED IF A ZERO (BLANK) INPUT VALUE WILL CAUSE AN

IMPOSSIBLE OPERATION TO BE ATTEMPTED DURING THE

TRANSFORMATIONS.

SUBROUTINE PREDIC

COMMON/BIG/SUMXX(1830), SUMXY(60,9),

x 1(60,9), CORR(1830),

COMMON/MED/ CON(99), EIGMS(9),

X IDENT(13), DOUT(60), NCON(200),

X NTERM(60), NTRANS(100), NXCOD(100), POOLED(9),

XREPVAR(9), RESMS(9), SUMX(138),

XSUMX2(69), X(99), ZEAN(69),

COMMON /FRMTS/ FMT(13), FMTTRI(14)

COMMON/SMALL/ BYPASS, BZERO, DELETE, FIRST, IFCHI, IFSSR,

X IFIT, IFWT, INPUT, INPUTS, INTER,

X ISTRAT, JCOL, KONNO, LENGTH, LIST,

X NERROR, NODEP, NOOB, NOTERM,

X NOVAR, NDEG, NRES, NTRANS, NWHERE,

X P, PREDCT, REPS, RWT,

X STORY, STORYC, STORYX, TOTWT, WEIGHT,

X ERRFXD, ECONMY

LOGICAL ECONMY

LOGICAL BYPASS, BZERO, DELETE, IFCHI,

X IFSSR, IFIT, IFWT, REPS, PREDCT,

XSTORYC, STORY, STORYX, TOTWT, WEIGHT,

X ERRFXD, ECONMY

DOUBLE PRECISION WEIGHT, RWT, TOTWT

COMMON/CNTRS/ I, IBC, IC, ICOL,

X INEW, INOCH, IOLD, IOUT, IR,

X IRC, IREP, IS, ITC, J,

X K, KBAR

DIMENSION YCALC(9), V(60), VARM(9), SEEM(9),

X VARP(9), SEEP(9)

EQUIVALENCE (YCALC(1), SUMXX(1)), (V(1), SUMXX(10)),

X (VAR(1), SUMXX(1)), (SEEM(1), SUMXX(80)), (VARP(1), SUMXX(89))

X (SEEP(1), SUMXX(98))

EQUIVALENCE (NDEG, RWT)

IF(NTERM.EQ.0) RETURN

WRITE(6,3)

READ(5,5) NPRED

DO 500 KK=1,NPRED

}
105 READ(5,FMT) (X(I),I=1,ICOL)
   WRITE(6,110)(X(I),I=1,ICOL)
125 CALL TRANS
   DO 130 K=1,JCOL
      I=NTERM(K)
      X(K) = CON(I)
   WRITE(6,135) (X(I),I=NOTERM)
130 CONTINUE
   CALL TRANS
   DO 140 K=1,NODEP
      YCALC(K) = BO(K)
      IF(.NOT.BZERO) YCALC(K)=0.0
   DO 150 J=1,NOTERM
      YCALC(K) = YCALC(K) + B(J,K)*X(J)
   150 CONTINUE
140 CONTINUE
   WRITE(6,310) (YCALC(K),K=1,NODEP)
   WRITE(6,320)(VARM(K),K=1,NODEP)
   WRITE(6,320)(SEEK(K),K=1,NODEP)
   WRITE(6,320)(SEEP(K),K=1,NODEP)
   RETURN
   3 FORMAT(54H1FOR EACH SET OF INDEP VARIABLES THERE IS COMPUTED... / 111
      X 20H PREDICTED RESPONSE / 112
      X 29H VARIANCE OF REGRESSION LINE / 113
      X 34H STANDARD DEVIATION OF REGRESSION / 114
      X 29H VARIANCE OF PREDICTED VALUE / 115
      X 39H STANDARD DEVIATION OF PREDICTED VALUE )
   5 FORMAT(14)
   110 FORMAT(39HINPUT DATA FOR THIS PREDICTED RESPONSE /I1H 9G14.6)
   135 FORMAT(56H INDEPENDENT TERMS ACCORDING TO FINAL REGRESSION MODEL
      X /I1H 9G14.6)
   310 FORMAT( 55HPREDICTED RESPONSE FOR ABOVE INDEP VARIABLES
      X /I1H 9G14.6)
   320 FORMAT(IH 9G14.6)
   END
SUBROUTINE CHISQ

1) COMPUTE PREDICTED VALUE OF DEPENDENT VARIABLES AND RESIDUALS
   AT INPUT DATA POINTS.
2) COMPUTE STANDARDIZED RESIDUALS.
3) COMPUTE SKEWNESS AND KURTOSIS OF SAMPLE DISTRIBUTION OF
   RESIDUALS. ALSO USE THE SAMPLE DISTRIBUTION TO COMPUTE
   THE CHI-SQUARE STATISTIC.
4) PRINT HISTOGRAMS OF THE DISTRIBUTION OF RESIDUALS.

SUBROUTINES NEEDED
HIST

COMMON/BIG/SUMXX(1830), SUMXI(1830), EIG(1830), SUMXY(60,9),
X SUM(60,9), CORR(1830)
COMMON/MED/R0(9), CON(99), ERRMS(9),
X IDENT(13), IDOUT(60), NCON(200),
X NTERM(60), NTRAN(100), NXCOD(100), POOLED(9),
X REPVAR(9), RESMS(9), SUMX(138),
X SUMX2(69), X(99), ZEAN(69), SUMY(18),
COMMON.SMALL/BYPASS, BZERO, DELETE, FIRST, IFCHI, IFSSR,
X IFTT, IFWT, INPUT, INPUTS, INTER,
X ISTRAT, JCOL, KONNO, LENGTH, LIST,
X NERROR, NODEP, NOOB, NOTERM,
X NOVAR, NPDEG, NRES, NTRANS, NWHERE,
X P, PREDCT, REPS, RWT,
X STORYC, STORYX, TOTWT, WEIGHT,
X ERRFXD, ECONMY
LOGICAL ECONMY
DOUBLE PRECISION RWT, TOTWT, WEIGHT
LOGICAL BYPASS, BZERO, DELETE, IFCHI,
X IFSSR, IFTT, IFWT, REPS, PREDCT,
X STORYC, STORYX, STORYI, FIRST, ERRFXD
COMMON/CNTRS/ I, IBC, IC, ICOL,
X INEW, INOCH, IOLD, IOUT, IR,
X IRC, IREP, IS, ITC, J,
X K, KBAR

INTEGER CELLS, PLUS1
DIMENSION BOUND(45), CELLD(21), CHI(9),
X OBS(20,9), RCT(212), RELKUR(9), RELSKW(9),
X STDERR(9), VAR(9), YCALC(9), YDIFR(9),
X 2(9)
EQUIVALENCE (CELLBD, SUMXX(11)), (CHI, SUMXX(22)),
(OBS, SUMXX(31)), (RCT, SUMXX(221)),
(REL KUR, SUMXX(434)), (REL SKW, SUMXX(443)),
(ST DERR, SUMXX(452)), (VAR, RESMS),
(YCALC, SUMXX(461)), (YDIFR, SUMXX(470)),
(X, (Z, SUMXX(479)))

C
J COL=NOTERM + NODEP
NUVAR=NOTERM + 1
BYPASS=.FALSE.
KOUNT= 0

C******************************************************************************
DETERMINE IF SAMPLE SIZE IS LARGE ENOUGH TO PERMIT CHI-SQUARE
C CALCULATION. IF SO, DETERMINE NUMBER OF CELLS AND CELL BOUNDARIES
IF (ERROR-30) 110,120,120

110 BYPASS=.TRUE.
GO TO 125

120 CELLS=NOOB/5
CELLS=MNO(CELLS, 20)
I= MOD(CELLS, 2)
IF (I.NE.0) CELLS=CELLS + 1
CELLS= FLOAT(CELLS)
PLUS1= CELLS + 1
MINUS1= CELLS -1
NDEGCH = CELLS-3
IR= CELLS/2-1
IC=IR+(IR-1)/2
IS=IR+2
DO 122 J=lrIR
IC=IC+l
IBC=IS-J
IRC=IS+J
CELLBD(IBC)= -BOUND( IC)
CELLBD(IRC)= -BOUND( IC)
CONTINUE

122 CONTINUE
CELLBD(1)= -1.E+37
CELLBD(PLUS1)= -1.E+37
CELLBD(IS)= 0.0
DO 124 K=1, NODEP
CHI(K)= 0.0
DO 124 I=1, CELLS
OBS(I,K)= 0.0

80
**Initialize skewness and kurtosis arrays. Compute standard error of estimate.**

```
125 DO 130 K=1,NODEP
     RELKUR(K)=0.0
     RELSKW(K)=0.0
     STDERR(K)= SQRT(ERRMS(K))
130 CONTINUE
     WRITE(LIST,135)
135 FORMAT(51H FOR EACH DEPENDENT TERM AND OBSERVATION IS PRINTED)
     X /31H OBSERVED RESPONSE (Y OBSERVED)
     X /29H CALCULATED RESPONSE (Y CALC)
     X /28H RESIDUAL (YORS-YCALC=YDIF)
     X /28H STANDARDIZED RESIDUAL (Z)
```

**Read data and initialize variables.**

```
140 CONTINUE
     DO 142 I=1,NOTERM
        K= IDOUT(I)
        X(I)= X(K)
142 CONTINUE
     KBAR=NWHERE
     DO 143 I=1,NODEP
        IC= NOTERM* I
        KBAR=KBAR+1
        X(IC)= X(KBAR)
143 CONTINUE
```

**Calculate predicted values.**

```
150 CONTINUE
     ACTDEV= X(KBAR)- YCALC(K)
     YDIFR(K)= ACTDEV
     Z(K)=ACTDEV/STDERR(K)
     ACTDE3=ACTDEV**3
     RELSKW(K)= RELSKW(K)+ACTDE3
     RELKUR(K)= RELKUR(K)+ACTDE3*ACTDEV
160 CONTINUE
     WRITE(LIST,180) X(K),K=NUVAR,JCOL
     WRITE(LIST,190) YCALC(K),K=1,NODEP
     WRITE(LIST,200) YDIFR(K),K=1,NOTERM
     WRITE(LIST,210) Z(K),K=1,NODEP
180 FORMAT(16H KY OBSERVED,9G13.4)
190 FORMAT(12H Y CALC,9G13.4)
200 FORMAT(12H Y DIF,9G13.4)
210 FORMAT(12H STUDENTIZED,9G13.4)
     IF (BYPASS) GO TO 410
```

81
C******************************************************************************
DO 250 K=1,NODEP
   DO 230 I=1,PLUS1
     IF(ZIK)-CELLBD(I))220,220,230
220 OBS(I-1,K)=OBS(I-1,K)+ 1.0
     GO TO 250
230 CONTINUE
250 CONTINUE
C
410 KOUNT = KOUNT +1
     IF(KOUNT.LT.10) GO TO 430
     WRITE(LIST,270) IDENT
270 FORMAT(1H113A69A2)
     KOUNT=0
430 CONTINUE
C
C******************************************************************************
C PRINT SKEWNESS AND KURTOSIS
   DO 440 K=1,NODEP
      RELSKW(K)=RELSKW(K)**2/(FLOAT(NOOB)**2*ERRMS(K)**3)
      RELKUR(K)=RELKUR(K)/(FLOAT(NOOB)*ERRMS(K)**2)
440 CONTINUE
   WRITE(LIST,450) IDENT,(RELSKW(K),K=1,NODEP)
450 FORMAT(1H135A69A2)  // 186
      X 12X,9F12.4)
      WRITE(LIST,460) (RELKUR(K),K=1,NODEP)
460 FORMAT(10X,31H KURTOSIS (SHOULD BE NEAR THREE) //12X,9F12.4)
      IF(NOT .BYPASS) GO TO 480
      WRITE(LIST,470)
470 FORMAT(74HKCHI-SQUARE IS NOT COMPUTED FOR LESS THAN 30 DEGREES
      X FREEDOM FOR ERROR. )
      RETURN
C
C******************************************************************************
C COMPUTE CHI-SQUARED AND PRINT HISTOGRAMS OF RESIDUALS
   DO 580 K=1,NODEP
      DO 570 I=1,CELLS
         CHI(K)= CHI(K) +OBS(I,K)**2
570 CONTINUE
      CHI(K)=FCELLS*CHI(K)/FLOATO-FLOAT(NOOB)
580 CONTINUE
   WRITE(LIST,590) NDEGCH,(CHI(K),K=1,NODEP)
590 FORMAT(55HKTHE CHI-SQUARED VALUES ARE LISTED BELOW. COMPARE WITH
      X 110,20H DEGREES OF FREEDOM / 1H 9614.6)
   RELFRQ = TOTWT/FCELLS
   DO 650 K=1,NODEP
      WRITE(LIST,620) K,K,RELFRQ
620 FORMAT(1H113A69A2)  // 205
      X 15X,9G14.6)
      DO 640 I=1,CELLS
         RCT(I)=OBS(I,K)
640 CONTINUE
   CALL HIST(K,RCT,CELLS)
650 CONTINUE
   RETURN
END
$IBFTC_RECTXX$

SUBROUTINE RECT(IROW, JCOL, IMAX, JMAX, A, FMT)
DIMENSION A(IMAX, JMAX), FMT(14), XOUT(8)
DATA J8/8/
LOGICAL OUT
OUT = .FALSE.
JTIMES = 0
JCOL = JCOL
5 JNXT = JCOL - J8
IF(JNXT) 10, 20, 30
10 JP = JCOL
GO TO 40
20 JP = J8
GO TO 40
30 JCOL = JNXT
JP = J8
GO TO 50
40 OUT = .TRUE.
50 DO 100 I = IROW
DO 60 J = JP
JJ = JTIMES + J
60 XOUT(J) = A(I, JJ)
WRITE (6, FMT) (XOUT(K), K = 1, JP)
100 CONTINUE
IF(OUT) RETURN
WRITE(6, 110)
110 FORMAT(1H /
JTIMES = JTIMES + JP
GO TO 5
END

$IBFTC_LOCXXX$

SUBROUTINE LOC(I, J, IR)
IX = I
JX = J
20 IF(IX - JX) 22, 24, 24
22 IRX = IX + (JX*JX - JX)/2
GO TO 36
24 IRX = JX + (IX*IX - IX)/2
36 IR = IRX
RETURN
END
SUBROUTINE RORD

PURPOSE
TO COMPLETE THE INVERSION OF A SYMMETRIC POSITIVE DEFINITE
MATRIX A OF ORDER N GIVEN THAT THE UPPER LEFT SUB-
MATRIX OF ORDER N-1 HAS ALREADY BEEN INVERTED.

SUBROUTINES NEEDED
LOC

REMARKS
ONLY THE UPPER TRIANGULAR PART OF A IS STORED AS A
VECTOR IN THE ORDER A(1,1), A(1,2), A(2,2), A(1,3), ... ETC
SUBROUTINE RORD(IORDER,A)

DIMENSION BETA(60), A(I)

ALPHA = 0.0
NM1 = IORDER - 1
100 A(I) = 1.0 / A(I)
GO TO 600
200 M = NM1 * (NM1 + 1) / 2
LEN = M + IORDER
DO 400 I = 1, NM1
BETA(I) = 0.0
MI = M + I
DO 350 J = 1, NM1
CALL LOC(I, J, II)
MJ = M + J
BETA(I) = BETA(I) - A(II) * A(MJ)
350 CONTINUE
ALPHA = ALPHA + A(MI) * BETA(I)
400 CONTINUE
ALPHA = ALPHA + A(LEN)
RALPHA = 1.0 / ALPHA
A(LEN) = RALPHA
DO 500 I = 1, NM1
DO 500 J = 1, I
CALL LOC(I, J, II)
A(II) = A(II) + BETA(I) * BETA(J) * RALPHA
500 CONTINUE
DO 550 J = 1, NM1
MJ = M + J
A(MJ) = BETA(J) * RALPHA
550 CONTINUE

RETURN
END
APPENDIX B

BORROWED ROUTINES

Some of the routines used in the program were taken from the literature. Both
INVXTX and TRIANG are by Webb and Galley (ref. 9), and EIGEN and HIST are from the
IBM programmer's manual (ref. 10).

Listing of INVXTX and TRIANG are given here, as follows:

```fortran
SUBROUTINE INVXTX(A, NN, D, FACT)

C ASSUMES THE MATRIX A IS SYMMETRIC AND POSITIVE DEFINITE, AND ONLY
C THE UPPER TRIANGLE IS STORED AS A ONE-DIMENSIONAL ARRAY IN THE
C ORDER A(1,1), A(1,2), A(2,2), A(1,3), A(2,3), ..., A(N,N).
C NN IS THE ORDER N OF THE INPUT MATRIX A.
C D IS ON EXIT THE DETERMINANT OF A, DIVIDED BY FACT*NN.
C
DIMENSION A(1)
D = 1.000
N = NN
ITR1 = 0
FACTOR = FACT
DO 145 K=1,N

C ITR1 = ITR1+K-1
KP1 = K+1
KM1 = K-1
KJ = ITR1+K

C CONTINUED PRODUCT OF PIVOTS
D = D*(KJ)/FACTOR
PV = 1.000/A(KJ)

C ITR2 = 0
IF (K-1) 150,80,50

C REDUCE TOP PART OF TRIANGLE, LEFT OF PIVOTAL COLUMN
50 DO 60 J=1,KM1

ITR2 = ITR2+J-1
KJ = ITR1+J
F = A(KJ)*PV
DO 60 I=1,J
IJ = ITR2+I
IK = ITR1 + I
60 A(IJ) = A(IJ) + A(IK)*F

C IF (K-N) 70,120,150
C
C REDUCE REST OF TRIANGLE, RIGHT OF PIVOTAL COLUMN
70 ITR2 = ITR1
80 DO 110 J=1,KP1,N

ITR3 = ITR1
ITR2 = ITR2+J-1
KJ = ITR2+K
F = A(KJ)*PV
DO 110 I=1,J
IJ = ITR2+I
IK = ITR1 + I
110 A(IJ) = A(IJ) - A(IK)*F
GO TO 100

C DIVIDE PIVOTAL ROW-COLUMN BY PIVOT, INCLUDING APPROPRIATE SIGNS
120 ITR2 = ITR1
```

85
DO 140 I=1,N
   IF (I-K) 125,130,135
125  IK = ITR1+I
   A(IK) = -A(IK)*PV
   GO TO 140
C (REPLACE PIVOT BY RECIPROCAL)
130  A(IK) = PV
   GO TO 140
135  ITR2 = ITR2+I-1
   K1 = ITR2+K
   A(K1) = A(K1)*PV
140  CONTINUE
C 145 CONTINUE
C 150 RETURN
END

SIBFTC TRIANX

SUBROUTINE TRIANG1(NN,NKOL,FORMAT)
DIMENSION FORMAT(1)
DIMENSION A(1)
1 FORMAT (IH1)
3 FORMAT (IH /IH /IH )
n = NN
NCOL = NKOL
KLUMPS = N/NCOL

C KEEPTR = 0
   K1 = 1
   K2 = NCOL - 1
   K3 = NCOL
1F (KLUMPS .EQ. 0) GO TO 120
C DO 90 KLUMP=1,KLUMPS
   ITR1 = KEEPTR
   I = -1
   ILO = (KLUMP-1)*NCOL + ITR1 + 1
   DO 30 K=K1,K2
      I = I + 1
      ITR1 = ITR1 + K - 1
      ILO = ILO + K - 1
   30  IHI = ILO + 1
50  WRITE (6,FORMAT) K, (A(J), J=ILO,IHI)
   KEEPTR = ITR1 + K2
   DO 60 K=K3,N
      ITR1 = ITR1 + K - 1
      ILO = ILO + K - 1
   60  IHI = ILO + NCOL - 1
   60  WRITE (6,FORMAT) K, (A(J), J=ILO,IHI)
   K1 = K1 + NCOL
   K2 = K2 + NCOL
   K3 = K3 + NCOL
90  WRITE(6,3)
C 120 ITR1 = KEEPTR
   IF (K1 .GT. N) GO TO 180
      I = -1
      ILO = KLUMPS*NCOL + ITR1 + 1
      DO 150 K=K1,N
         I = I + 1
         ITR1 = ITR1 + K - 1
         ILO = ILO + K - 1
      150  IHI = ILO + 1
   150  WRITE (6,FORMAT) K, (A(J), J=ILO,IHI)
C 180 RETURN
END
REFERENCES


"The aeronautical and space activities of the United States shall be conducted so as to contribute . . . to the expansion of human knowledge of phenomena in the atmosphere and space. The Administration shall provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof."

—NATIONAL AERONAUTICS AND SPACE ACT OF 1958

NASA SCIENTIFIC AND TECHNICAL PUBLICATIONS

TECHNICAL REPORTS: Scientific and technical information considered important, complete, and a lasting contribution to existing knowledge.

TECHNICAL NOTES: Information less broad in scope but nevertheless of importance as a contribution to existing knowledge.

TECHNICAL MEMORANDUMS: Information receiving limited distribution because of preliminary data, security classification, or other reasons.

CONTRACTOR REPORTS: Scientific and technical information generated under a NASA contract or grant and considered an important contribution to existing knowledge.

TECHNICAL TRANSLATIONS: Information published in a foreign language considered to merit NASA distribution in English.

SPECIAL PUBLICATIONS: Information derived from or of value to NASA activities. Publications include conference proceedings, monographs, data compilations, handbooks, sourcebooks, and special bibliographies.

TECHNOLOGY UTILIZATION PUBLICATIONS: Information on technology used by NASA that may be of particular interest in commercial and other non-aerospace applications. Publications include Tech Briefs, Technology Utilization Reports and Notes, and Technology Surveys.

Details on the availability of these publications may be obtained from:

SCIENTIFIC AND TECHNICAL INFORMATION DIVISION

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Washington, D.C. 20546