A. MATRIX STRUCTURE AND NOTATION

1) A matrix is a rectangular arrangement of numbers. The matrix is usually denoted by a capital letter.

$$A = \begin{bmatrix} 1 & 3 \\ 7 & 9 \end{bmatrix} \qquad D = \begin{bmatrix} 4 & 2 & 4 \\ 1 & 6 & 0 \\ 3 & 0 & 5 \\ 2 & 3 & 0 \end{bmatrix}$$

2) The dimensions of a matrix are given by the number of rows and columns in the matrix (i.e. the dimensions are r by c). For the matrices above,

A is 2 by 2

D is 4 by 3

3) The individual elements of a matrix can be referred to by specifying the row and column in which it occurs. Lower case numbers are used to represent individual elements, and should match the upper case letter used to denote matrix. For example, individual elements from matrices A and D above can be referred to as,

$a_{11} = 1$

$a_{21} = 7$

$d_{22} = 6$

$d_{12} = 2$

B. TYPES OF MATRICES

1) Square matrix - the number of rows and columns are equal. Matrix A above is a square matrix (2 by 2), matrix D is not (4 by 3). A symmetric matrix is an important variation of the square matrix. In a symmetric matrix, the value in position "ij" equals the value in position "ji" (where $i \neq j$). For example, if $c_{31} = 5$ then c_{13} is also 5.

2) Scalar - a single number can be thought of as a 1 by 1 matrix and is called a scalar.

3) Vector - a single column or single row of numbers is called a vector. The dimensions of a row vector are (1 by c), where "c" is the number of columns, and the dimensions of a column vector (r by 1), where "r" is the number of rows.

4) Identity matrix - this special square matrix consists of all ones on the main diagonal, or principal diagonal, and zeros in all the off diagonal positions. The following are examples of identity matrices,

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad F = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The diagonal matrix is a generalization of the identity matrix. A diagonal matrix can have any value on the main diagonal, but also has zeros in the off diagonal positions.

C. MATRIX TRANSPOSE

The transpose of a matrix consists of a new matrix such that the rows of the original matrix become the columns of the transpose matrix. The transpose matrix is denoted with the same letter as the original matrix followed by a prime (e.g. the transpose of X is X).

$$D = \begin{bmatrix} 4 & 2 & 4 \\ 1 & 6 & 0 \\ 3 & 0 & 5 \\ 2 & 3 & 0 \end{bmatrix} \qquad D' = \begin{bmatrix} 4 & 1 & 3 & 2 \\ 2 & 6 & 0 & 3 \\ 4 & 0 & 5 & 0 \end{bmatrix}$$

D. MATRIX ADDITION AND SUBTRACTION

Matrices to be added or subtracted must be of the same dimensions. Each element of the first matrix, (a) is added (or subtracted) from the corresponding element of the second matrix, (b).

$$A = \begin{bmatrix} 1 & -2 \\ 3 & 4 \\ 9 & 0 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 4 \\ 1 & 4 \\ -4 & 4 \end{bmatrix} \quad A+B = \begin{bmatrix} 1+1 & -2+4 \\ 3+1 & 4+4 \\ 9-4 & 0+4 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 4 & 8 \\ 5 & 4 \end{bmatrix}$$

E. MATRIX MULTIPLICATION

Multiplication by a scalar - in this type of multiplication each element of the matrix is simply multiplied, element by element, by the scalar value.

$$A = \begin{bmatrix} 1 & -2 \\ 3 & 4 \\ 9 & 0 \end{bmatrix} \quad B = [7] \qquad A * B = 7 * \begin{bmatrix} 1 & -2 \\ 3 & 4 \\ 9 & 0 \end{bmatrix} = \begin{bmatrix} 7 & -14 \\ 21 & 28 \\ 63 & 0 \end{bmatrix}$$

Element by element multiplication - matrix multiplication is not usually done by matching each i,jth element of one matrix with the corresponding ijth element of the second matrix. This is called elementwise multiplication and it is not the normal mode of matrix multiplication and should not be used unless specifically requested.

The standard method of matrix multiplication requires that the number of columns in the first matrix equal the number of rows in the second matrix. If the first matrix is (r by c) and the second is (r by c), in order to multiply the matrices, c must equal r. The resulting matrix will have the dimensions (r by c).

Multiplication is accomplished by summing the cross products of each row of the first matrix and each column of the second matrix.

$$A = \begin{bmatrix} 1 & -2 \\ 3 & 4 \\ 9 & 0 \end{bmatrix} X = \begin{bmatrix} 1 & -2 \\ 3 & 4 \end{bmatrix}$$

Since A is 3 rows by 2 columns, and X is 2 by 2, then the columns of the first matrix equals the rows of the second matrix, and the matrices may be multiplied.

$$A*X = \begin{bmatrix} 1 & -2 \\ 3 & 4 \\ 9 & 0 \end{bmatrix} * \begin{bmatrix} 1 & -2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} (1*1)+(-2*3) & (1*-2)+(-2*4) \\ (3*1)+(4*3) & (3*-2)+(4*4) \\ (9*1)+(0*3) & (9*-2)+(0*4) \end{bmatrix} = \begin{bmatrix} -5 & -10 \\ 15 & 10 \\ 9 & -18 \end{bmatrix}$$

the new dimensions for the product of A * X are,

(3 x 2) ↓ must be equal ↓ x (2 x 3)

↑ new dimensions ↑

Note that though we can multiply A * X, we could not have done the multiplication the other way (i.e. X * A), since the dimensions would not have matched. That is, we could pre-multiply by A, but could not pre-multiply by X.

F. SIMPLE MATRIX INVERSION (2 by 2 matrix only)

Matrices are not "divided", but may be inverted. Instead of "dividing" A by B, one would multiply A by the inverse of B. The inverse of a (2 by 2) matrix is given by,

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad A^{-1} = \frac{1}{(a \times d) - (b \times c)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

The scalar value resulting from the calculation "(a×d) – (b×c)" is called the determinant. The matrix cannot be inverted unless the inverse of the determinant exists (is defined). It will not exist in a case such as the one below since (1÷0) is not defined.

$$A = \begin{bmatrix} 1 & 4 \\ 2 & 8 \end{bmatrix} \qquad \text{then } \frac{1}{\text{Determinant of A}} = \frac{1}{(1 \times 8) - (2 \times 4)} = \frac{1}{0}$$

This occurs in regression when two variables are linearly related.

An example of the inversion of a 2 * 2 matrix is given below.

$$B = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} \quad B^{-1} = \frac{1}{(2 \times 4) - (1 \times 3)} \begin{bmatrix} 4 & -3 \\ -1 & 2 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 4 & -3 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 0.8 & -0.6 \\ -0.2 & 0.4 \end{bmatrix}$$

Note that a matrix times its inverse (i.e. $B \times B^{-1}$) results in an identity matrix. By definition, the inverse of a matrix G is a matrix which when multiplied by G produces an identity matrix, or $G \times G^{-1} = I$.

G. SIMPLE LINEAR REGRESSION

Solving a simple linear regression with matrices requires the same values used for an algebraic solution from summation notation formulas. These are;

$$n, \qquad \sum_{i=1}^{n} X_i, \quad \sum_{i=1}^{n} Y_i, \quad \sum_{i=1}^{n} X_i^2, \quad \sum_{i=1}^{n} Y_i^2, \quad \sum_{i=1}^{n} X_i Y_i$$

where n is the size of the sample of data. To obtain these values in the matrix form we start with the matrix equivalent of the individual values of X and Y, the raw data matrices.

$$X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ 1 & X_4 \\ 1 & X_5 \\ 1 & X_6 \\ 1 & X_7 \end{bmatrix} \qquad Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \\ Y_7 \end{bmatrix}$$

The column of ones is necessary, and represents the intercept. Omitting this column would force the regression through the origin. The next step in the calculations is to obtain the X′X, X′Y and Y′Y matrices. These calculations provide the sums of squares and cross products.

$$X'X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ X_1 & X_2 & X_3 & X_4 & X_5 & X_6 & X_7 \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ 1 & X_4 \\ 1 & X_5 \\ 1 & X_6 \\ 1 & X_7 \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ X_1 & X_2 & X_3 & X_4 & X_5 & X_6 & X_7 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \\ Y_7 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_i Y_i \end{bmatrix}$$

$$Y'Y = \begin{bmatrix} Y_1 & Y_2 & Y_3 & Y_4 & Y_5 & Y_6 & Y_7 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \\ Y_7 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} Y_i^2 \end{bmatrix}$$

The regression coefficients, b_0 and b_1, are then given by, $B = (X'X)^{-1}X'Y$, where

$$(X'X)^{-1} = \frac{1}{\left(n\sum X_i^2\right) - \left(n\sum X_i\right)^2} \begin{bmatrix} \sum X_i^2 & -\sum X_i \\ -\sum X_i & n \end{bmatrix}$$

and since

$$\frac{1}{\text{Determinant } (X'X)^{-1}} = \frac{1}{\left(n\sum X_i^2\right) - \left(\sum X_i\right)^2} = \frac{1}{nS_{XX}}$$

where S_{xx} is the corrected sum of squares of X.

Then $$(X'X)^{-1} = \begin{bmatrix} \frac{\sum X_i^2}{nS_{XX}} & \frac{-\sum X_i}{nS_{XX}} \\ \frac{-\sum X_i}{nS_{XX}} & \frac{n}{nS_{XX}} \end{bmatrix}$$

and the regression coefficients can be obtained by,

$$(X'X)^{-1}X'Y = \begin{bmatrix} \frac{\sum X_i^2}{nS_{XX}} & \frac{-\sum X_i}{nS_{XX}} \\ \frac{-\sum X_i}{nS_{XX}} & \frac{n}{nS_{XX}} \end{bmatrix} \times \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix} = \begin{bmatrix} \frac{\sum X_i^2}{nS_{XX}} \sum Y_i & \frac{-\sum X_i}{nS_{XX}} \sum X_i Y_i \\ \frac{-\sum X_i}{nS_{XX}} \sum Y_i & \frac{n}{nS_{XX}} \sum X_i Y_i \end{bmatrix}$$

$$= \begin{bmatrix} \overline{Y} - b_1 \overline{X} \\ \frac{\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}}{\sum X_i^2 - \frac{\left(\sum X_i\right)^2}{n}} \end{bmatrix} = \begin{bmatrix} \overline{Y} - b_1 \overline{X} \\ \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2} \end{bmatrix} = \begin{bmatrix} \overline{Y} - b_1 \overline{X} \\ \frac{S_{XY}}{S_{XX}} \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

The remaining calculations usually needed to complete the compliment of calculations for the simple linear regression is the sum of squared deviations or error term. The matrix formula is

$$SSE = Y'Y - B'X'Y = \Sigma Y^2 - [b_0 \;\; b_1] \begin{bmatrix} \Sigma Y \\ \Sigma XY \end{bmatrix}$$

$$= \Sigma Y^2 - (b_0 * \Sigma Y + b_1 * \Sigma XY) = \quad UCSSTotal \; - \; UCSSReg$$

These calculations produce the same algebraic equations for b_0, b_1, and SSE that are given in most statistics texts. The advantage of using the matrix version of the formulas is that the matrix equations given above will work equally well for multiple regression with two or more independent variables.

The ANOVA table calculated with matrix formulas is

	Uncorrected		Corrected	
Source	d.f.	Sum of Squares	d.f.	Sum of Squares
Regression	2	B′X′Y	1	B′X′Y – CF
Error	n–2	Y′Y–B′X′Y	n–2	Y′Y – B′X′Y
Total	n	Y′Y	n–1	Y′Y – CF

where the correction factor is calculated as usual, $CF = \frac{(\Sigma Y)^2}{n} = n\overline{Y}^2$.

The value for R^2 is calculated as $\frac{SSRegression}{SSTotal} = \frac{B'X'Y - CF}{Y'Y - CF}$, and is often expressed as a percent. Note that this calculation employs corrected sums of squares for both SSRegression and SSTotal.

The Mean Squares (MS) for the SSRegression and SSError are calculated by dividing the SS (corrected sum of squares) by their d.f. (degrees of freedom). The test of hypotheses for $[H_0 := \beta_1]$ is then calculated as;

$$F = \frac{MSRegression}{MSTotal} = \frac{\left(B'X'Y - CF\right) / dfReg}{\left(Y'Y - CF\right) / dfError}$$

or

$$t = \frac{(b_1 - 0)}{S_{b_1}} = \sqrt{\text{F value}}$$

where S_{b_1} is obtained from the VARIANCE COVARIANCE matrix.

The VARIANCE COVARIANCE matrix is calculated as from the $(X'X)^{-1}$ matrix.

$$(X'X)^{-1} = \begin{bmatrix} c_{00} & c_{01} \\ c_{10} & c_{11} \end{bmatrix}$$

where the c_{ij} values are called Gaussian multipliers. The VARIANCE-COVARIANCE matrix is then calculated from this matrix by multiplying by the MSError.

$$MSE(X'X)^{-1} = \begin{bmatrix} MSEc_{00} & MSEc_{01} \\ MSEc_{10} & MSEc_{11} \end{bmatrix}$$

The individual values then provide the variances and covariances such that

$MSE*c_{00}$ = Variance of b_0 = $VAR(b_0)$

$MSE*c_{11}$ = Variance of b_1 = $VAR(b_1)$, so $S_{b_1} = \sqrt{MSE * c_{11}}$

$MSE*c_{01}$ = $MSE*c_{10}$ = Covariance of b_0 and $b_{0\backslash 1}$ = $COV(b_0,b_1)$

It is important to note that the variances and covariances calculated from the $(X'X)^{-1}$ are for the b_i (β_i estimates), not for the X_i values. Also, $COV(b_0,b_1) \neq COV(X_0,X_1)$.

```
1          ************************************************************************;
2          *** EXST7034 Example 1 using PC-SAS - Airline vial breakage          ***;
3          *** Problem from Neter, Kutner, Nachtsheim & Wasserman 1996, #1.21   ***;
4          ************************************************************************;
5          OPTIONS LS=88 PS=256 NOCENTER NODATE NONUMBER;
6          DATA ONE; INFILE CARDS MISSOVER;
7           TITLE1 `EXST7034 - Example 1 : Airline vial breakage - NKNW Example 1.21`;
8          *     LABEL X = `Breakage per 1000 vials`;
9          *     LABEL Y = `Number of airline transfers`;
10            INPUT X1 Y;
11            X0 = 1;
12         CARDS;
NOTE: The data set WORK.ONE has 10 observations and 3 variables.
NOTE: DATA statement used:
      real time           0.06 seconds
      cpu time            0.06 seconds
23         ;
24
25         PROC IML; ***RESET PRINT;
NOTE: IML Ready
25       !                          USE ONE;
26         READ ALL VAR{X0 X1} INTO X;
27         READ ALL VAR{Y} INTO Y;
28         CLOSE ONE;
29         ****************************************;
30         **** Calculation of the Full Model    ****;
31         ****************************************;
32         **** Intermediate calculations        ****;
33         ****************************************;
34             N = NROW(X);     P = NCOL(X);
35             YPY = Y`*Y;      XPX = X`*X;     XPY = X`*Y;
36         PRINT ,,"Intermediate statistics",
37               N P XPX YPY XPY;

EXST7034 - Example 1 : Airline vial breakage - NKNW Example 1.21

Intermediate statistics

        N          P       XPX                  YPY       XPY
       10          2        10        10       2194       142
                            10        20                  182

38         ****************************************;
39         **** Solution & Analysis of Variance  ****;
40         ****************************************;
41             XSumSq = VECDIAG(XPX);
42             XPXINV = INV(XPX);
43             B = XPXINV * XPY;          CF = (J(1,N,1)*Y)##2/N;
44             USSTOTAL = YPY;            USSREG = B`*XPY;
45             SSTOTAL = YPY - CF;        SSREG = B`*XPY - CF;
46             SSERROR = YPY - B`*XPY;    dfE = N-P;
47             MSE = SSERROR / (N-P);     RSquare = SSReg / SSTotal;
48         PRINT ,,"Reg coefficients and ANOVA table info",
49               B XPXINV XSumSq CF USSTOTAL USSREG,,
50               SSTOTAL SSREG SSERROR MSE DFE RSquare;

Reg coefficients and ANOVA table info

         B    XPXINV              XSUMSQ        CF  USSTOTAL    USSREG
      10.2       0.2      -0.1        10    2016.4      2194    2176.4
         4      -0.1       0.1        20

   SSTOTAL     SSREG   SSERROR       MSE       DFE     RSQUARE
     177.6       160      17.6       2.2         8   0.9009009
```

```
51          ********************************************;
52          **** Test of B and observation diagnostics ****;
53          ********************************************;
54              hatmatrix = X*XPXinv*X`;
55              VARCOV = MSE * XPXINV;
56              VarResid = (I(N) - HatMatrix) * MSE;
57              VarPred = HatMatrix * MSE;
58          PRINT ,,"Variance information",,
59                HatMatrix, VarCov, VarResid, VarPred;
```

Variance information

HATMATRIX

0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
0.1	0.2	0	0.2	-0.1	0.1	0.2	0.1	0	0.2
0.1	0	0.2	0	0.3	0.1	0	0.1	0.2	0
0.1	0.2	0	0.2	-0.1	0.1	0.2	0.1	0	0.2
0.1	-0.1	0.3	-0.1	0.5	0.1	-0.1	0.1	0.3	-0.1
0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
0.1	0.2	0	0.2	-0.1	0.1	0.2	0.1	0	0.2
0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
0.1	0	0.2	0	0.3	0.1	0	0.1	0.2	0
0.1	0.2	0	0.2	-0.1	0.1	0.2	0.1	0	0.2

VARCOV

0.44	-0.22
-0.22	0.22

VARRESID

1.98	-0.22	-0.22	-0.22	-0.22	-0.22	-0.22	-0.22	-0.22	-0.22
-0.22	1.76	0	-0.44	0.22	-0.22	-0.44	-0.22	0	-0.44
-0.22	0	1.76	0	-0.66	-0.22	0	-0.22	-0.44	0
-0.22	-0.44	0	1.76	0.22	-0.22	-0.44	-0.22	0	-0.44
-0.22	0.22	-0.66	0.22	1.1	-0.22	0.22	-0.22	-0.66	0.22
-0.22	-0.22	-0.22	-0.22	-0.22	1.98	-0.22	-0.22	-0.22	-0.22
-0.22	-0.44	0	-0.44	0.22	-0.22	1.76	-0.22	0	-0.44
-0.22	-0.22	-0.22	-0.22	-0.22	-0.22	-0.22	1.98	-0.22	-0.22
-0.22	0	-0.44	0	-0.66	-0.22	0	-0.22	1.76	0
-0.22	-0.44	0	-0.44	0.22	-0.22	-0.44	-0.22	0	1.76

VARPRED

0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22
0.22	0.44	0	0.44	-0.22	0.22	0.44	0.22	0	0.44
0.22	0	0.44	0	0.66	0.22	0	0.22	0.44	0
0.22	0.44	0	0.44	-0.22	0.22	0.44	0.22	0	0.44
0.22	-0.22	0.66	-0.22	1.1	0.22	-0.22	0.22	0.66	-0.22
0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22
0.22	0.44	0	0.44	-0.22	0.22	0.44	0.22	0	0.44
0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22
0.22	0	0.44	0	0.66	0.22	0	0.22	0.44	0
0.22	0.44	0	0.44	-0.22	0.22	0.44	0.22	0	0.44

```
60          ********************************************;
61              StdErrB = sqrt(vecdiag(VarCov));
62              t = B / StdErrB;
63              Probt = 1 - ProbF(t#t, 1, dfe);
64          PRINT ,,"Tests of regression coefficients",
65                StdErrB t Probt;
```

Tests of regression coefficients

STDERRB	T	PROBT
0.663325	15.377079	3.1783E-7
0.4690416	8.5280287	0.0000275

```
66         ********************************************;
67             hatvalues = vecdiag(hatmatrix);
68             YHat = X * B;
69             Resid = Y - YHat;
70             /*Must specify t value */ TVal = 2.306;
71             LowerCLM = YHAT - TVAL # SQRT(hatvalues*MSE);
72             UpperCLM = YHAT + TVAL # SQRT(hatvalues*MSE);
73             LowerCLI = YHAT - TVAL # SQRT(hatvalues*MSE+MSE);
74             UpperCLI = YHAT + TVAL # SQRT(hatvalues*MSE+MSE);
75         PRINT ,,"Observation information" ,
76               Y YHAT RESID HatValues LowerCLM UpperCLM LowerCLI UpperCLI;
77         quit;
NOTE: Exiting IML.
NOTE: The PROCEDURE IML printed page 1.
NOTE: PROCEDURE IML used:
      real time           0.07 seconds
      cpu time            0.07 seconds


NOTE: SAS Institute Inc., SAS Campus Drive, Cary, NC USA 27513-2414
NOTE: The SAS System used:
      real time           0.70 seconds
      cpu time            0.50 seconds
```

Observation information

Y	YHAT	RESID	HATVALUES	LOWERCLM	UPPERCLM	LOWERCLI	UPPERCLI
16	14.2	1.8	0.1	13.11839	15.28161	10.612706	17.787294
9	10.2	-1.2	0.2	8.6703726	11.729627	6.4531935	13.946807
17	18.2	-1.2	0.2	16.670373	19.729627	14.453193	21.946807
12	10.2	1.8	0.2	8.6703726	11.729627	6.4531935	13.946807
22	22.2	-0.2	0.5	19.781447	24.618553	18.010943	26.389057
13	14.2	-1.2	0.1	13.11839	15.28161	10.612706	17.787294
8	10.2	-2.2	0.2	8.6703726	11.729627	6.4531935	13.946807
15	14.2	0.8	0.1	13.11839	15.28161	10.612706	17.787294
19	18.2	0.8	0.2	16.670373	19.729627	14.453193	21.946807
11	10.2	0.8	0.2	8.6703726	11.729627	6.4531935	13.946807

Application of matrix procedures to multiple regression first requires calculation of the X′X, X′Y and Y′Y matrices, where for dependent variable Y and independent variables X_1 and X_2. For a 2 factor multiple regression, these matrices are;

$$X'X = \begin{pmatrix} n & \sum_{i=1}^{n} X_{1i} & \sum_{i=1}^{n} X_{2i} & \sum_{i=1}^{n} X_{3i} \\ \sum_{i=1}^{n} X_{1i} & \sum_{i=1}^{n} X_{1i}^2 & \sum_{i=1}^{n} X_{1i}X_{2i} & \sum_{i=1}^{n} X_{1i}X_{3i} \\ \sum_{i=1}^{n} X_{2i} & \sum_{i=1}^{n} X_{1i}X_{2i} & \sum_{i=1}^{n} X_{2i}^2 & \sum_{i=1}^{n} X_{2i}X_{3i} \\ \sum_{i=1}^{n} X_{3i} & \sum_{i=1}^{n} X_{1i}X_{3i} & \sum_{i=1}^{n} X_{2i}X_{3i} & \sum_{i=1}^{n} X_{3i}^2 \end{pmatrix} \quad X'Y = \begin{pmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_{1i}Y_i \\ \sum_{i=1}^{n} X_{2i}Y_i \\ \sum_{i=1}^{n} X_{3i}Y_i \end{pmatrix} \quad Y'Y = \left(\sum_{i=1}^{n} Y_i^2 \right)$$

As with the simple linear regression, these sums, sums of squares and cross products are required by any method of fitting multiple regression. Once these values are obtained, application of formulas for an algebraic solution is relatively easy for a two-factor model. However, matrix procedures are more easily expanded to more than two independent variables than are summation notation formulas.

The inversion technique we will use is called the sweepout technique, and it requires the application of "row operations". Row operations consist of **(1) multiplying any row by a scalar value, and (2) adding or subtracting any row from any other row**. These are the only operations required to complete the sweepout technique after the matrices have been obtained and augmented.

Obtaining a maximum of information from the technique requires reducing the X′X matrix one column at a time to an identity matrix. However, values of the regression coefficients, error sum of squares and inverse matrix will be correct even of the row operations are not applied in a column by column reduction.

By "sweeping" out each column of the X′X matrix one by one to obtain an identity matrix, the sequentially adjusted sums of squares error can also be obtained. This requires augmenting the X′X matrix with the X′Y matrix and an identity matrix prior to applying the row operations. The complete augmented matrix is given below. The matrix has separate sections that are recognizable as matrices seen earlier. This type of sectioned matrix is called a partitioned matrix.

$$\begin{bmatrix} X'X & X'Y & I \\ X'Y & Y'Y & 0 \end{bmatrix} \xrightarrow{\text{row operations}} \begin{bmatrix} I & B & (X'X)^{-1} \\ 0 & SSE & -B' \end{bmatrix}$$

Sections of the matrix may be left off if less information is required. For example, if only the regression coefficients are needed, then the sweepout technique need be applied only to the matrix,

$$\begin{bmatrix} X'X & X'Y \end{bmatrix} \xrightarrow{\text{row operations}} \begin{bmatrix} I & B \end{bmatrix},$$

and if only the inverse is required, the only matrix needed is

$$\begin{bmatrix} X'X & I \end{bmatrix} \xrightarrow{\text{row operations}} \begin{bmatrix} I & (X'X)^{-1} \end{bmatrix}.$$

The regression coefficients and sum of squares error can be obtained by sweeping out the matrix,

$$\begin{bmatrix} X'X & X'Y \\ X'Y & Y'Y \end{bmatrix} \xrightarrow{\text{row operations}} \begin{bmatrix} I & B \\ 0 & SSE \end{bmatrix}.$$

If the above matrix is swept out column by column, then it will also provide the sequentially adjusted sums of squares. Only the use of the complete augmented matrix provides the inverted X′X matrix necessary to obtain the variance - covariance matrix, confidence limits and other types of sums of squares.

The technique will be illustrated with an example using data from Snedecor and Cochran (1981; ex. 17.2.1). The example will employ the complete augmented matrix. The original data matrices are;

$$X'X = \begin{bmatrix} 17 & 188.2 & 700 \\ 188.2 & 3602.78 & 8585.1 \\ 700 & 8585.1 & 31712 \end{bmatrix} \quad X'Y = \begin{bmatrix} 1295 \\ 16203.8 \\ 54081 \end{bmatrix} \quad Y'Y = [103075]$$

The augmented matrix to be swept is then,

$$\left(\begin{array}{cccc|ccc} 17 & 188.2 & 700 & 1295 & 1 & 0 & 0 \\ 188.2 & 3602.78 & 8585.1 & 16203.8 & 0 & 1 & 0 \\ 700 & 8585.1 & 31712 & 54081 & 0 & 0 & 1 \\ 1295 & 16203.8 & 54081 & 103075 & 0 & 0 & 0 \end{array}\right)$$

The first step in the sweepout technique is to multiply through the first row by the inverse of 17. This will result in a value of 1 in the first row - first column. A multiple of this new first row is then subtracted from each of the other rows (2, 3 and 4). The multiplier should be such that value(i,1)– [value(1,1)*multiplier] = 0 for $i \neq 1$.

The multiplier which accomplishes this is simply the value(i,1) since the new value(1,1) is unity (1). Therefore, every value(i,j) will be processed in the same way. The calculations would be,

for row 2: value(2,j) – (value(1,j) * 118.2)

for row 3: value(3,j) – (value(1,j) * 700)

for row 4: value(4,j) – (value(1,j) * 1295)

After applying these transformations we obtain the following matrix,

COLUMN 1 SWEEP

$$\left(\begin{array}{cccc|ccc} 1 & 11.0706 & 41.1765 & 76.1765 & 0.05882 & 0 & 0 \\ 0 & 1519.30 & 835.688 & 1867.39 & -11.0706 & 1 & 0 \\ 0 & 835.688 & 2888.47 & 757.471 & -41.1765 & 0 & 1 \\ 0 & 1867.39 & 757.471 & 4426.47 & -76.1765 & 0 & 0 \end{array}\right)$$

At this point the effect of X (the intercept) has been removed from the model. The value replacing Y′Y is 4426.471. This is the corrected sum of squares of Y (i.e. Y was 103075, and has now been corrected for the mean, yielding 4426.47).

The sweepout now proceeds to the second column. A value of 1 is needed in the second column - second row to proceed with the development of the identity matrix. This is obtained by multiplying through the second row by the inverse of the value presently in that position (i.e. 1519.30). Then, the

appropriate multiple of the new row 2 is subtracted from each of the other rows. Note that the first column remains unchanged since the value subtracted is always a multiple of zero.

COLUMN 2 SWEEP

$$\left(\begin{array}{cc|c|c|ccc} 1 & 0 & 35.08709 & 62.5694 & 0.13949 & -0.00729 & 0 \\ 0 & 1 & 0.550050 & 1.22911 & -0.00729 & 0.00066 & 0 \\ 0 & 0 & 2428.800 & -269.686 & -35.0871 & -0.55005 & 1 \\ 0 & 0 & -269.686 & 2131.236 & -62.5694 & 1.22911 & 0 \end{array}\right)$$

The sweep then proceeds with the third column. Once again a value of 1 is required in row 3, column 3, and all rows other than row 3 will have a multiple of row 3 subtracted from them.

COLUMN 3 SWEEP

$$\left(\begin{array}{ccc|c|ccc} 1 & 0 & 0 & 66.4654 & 0.646369 & 0.000660 & -0.014446 \\ 0 & 1 & 0 & 1.29019 & 0.000660 & 0.000783 & 0.000226 \\ 0 & 0 & 1 & 0.11104 & -0.014446 & 0.000226 & 0.000412 \\ 0 & 0 & 0 & 2101.291 & -66.46541 & -1.290191 & -0.11104 \end{array}\right)$$

Once this swept out matrix has been obtained, most commonly desired calculations follow easily. Some of these results are discussed below.

There are also several checks which can be done on the calculations. As the matrix is swept out, the null matrix (matrix of zeroes in the original augmented matrix) is replaced by the negative values of the regression coefficients if the calculations have been done correctly. As a second check, the product of the original X′X matrix and its inverse should produce an identity matrix. (i.e. $X'X * (X'X)^{-1} = I$)

REGRESSION COEFFICIENTS

The regression coefficients are produced during the sweepout, replacing the X′Y matrix. The model for the analysis above is,

$$\hat{Y} = b_0 + b_1 X_{1i} + b_2 X_{2i}$$

$$\hat{Y} = 66.4654 + 1.2902 X_{1i} + 0.1110 X_{2i}$$

SEQUENTIALLY ADJUSTED SUMS OF SQUARES

As each column is swept out, the sums of squares are "adjusted" for the factor removed. The first sweep adjusts for the intercept (i.e. 1 = n) on the diagonal of X′X, so the reduction in the Y is the correction factor or the adjustment for the mean.

e.g. C.F. = 103075 – 4426.470 = 98648.530

The second sweep adjusts for the second term in the X matrix, usually X, and the reduction in the error term is that sum of squares attributable to X (given that X is already in the model).

e.g. SS(X|X) = 4426.470 – 2131.236 = 2295.234

The third sweep adjusts for X and the reduction in the sum of squares is attributable to X (given that X and X are already in the model).

e.g. SS(X|X X) = 2131.236 – 2101.291 = 29.945

Finally, the remaining sum of squares is the error sum of squares

SSE = 2101.291

Note that since the variables are adjusted sequentially, the sums of squares obtained are dependent on the order in which the variables are entered. That is, if we had entered X first and X second, the sums of squares attributable to these two variables would not be the same as the results obtained above. Only the correction factor would be the same (since it would have been entered first in both models).

Each adjustment of the sum of squares takes one degree of freedom. The residual sum of squares has (nk) degrees of freedom, where n is the number of observations, and k is the number of sweeps, or the number of columns in the X′X matrix. The mean square error is then,

$$MSE = \frac{SSE}{(n-k)} = \frac{2101.291}{(17 - 3)} = 150.092$$

PARTIAL SUMS OF SQUARES

Since the sequentially adjusted sums of squares are dependent on the order in which the variables are entered, another value of interest is the partial sum of squares or the uniquely attributable sum of squares. This is simply the sum of squares that would be accounted for by each variable if it had been entered into the model in last place. This value could be obtained by reversing the sweep operation, and observing the change in the sum of squares as each variable was swept back into the model.

The only change in sum of squares when a variable is swept back into the model is, bc,

So this calculation will give the partial SS due to variable X without actually doing all the calculations necessary to reverse the sweepout technique. The elements (c) are obtained from the $(X'X)^{-1}$ matrix and are called Gaussian multipliers.

The partial SS due to X above does not change since it was the variable in the last position. The partial SS due to X would be calculated as,

$$SS(X_1|X_0\ X_2) = \frac{(1.29019)^2}{(0.000783)} = 2125.913$$

VARIANCE COVARIANCE MATRIX

Another major result of the sweepout technique is the inverse of the X′X matrix. Multiplying this matrix by the mean square error (MSE) gives the variance - covariance matrix of the regression coefficients.

e.g. $VarCov = MSE * (X'X)^{-1} =$

$$= 150.092 \begin{pmatrix} 0.64637 & 0.00066 & -0.01445 \\ 0.00066 & 0.00078 & -0.00023 \\ -0.01445 & -0.00023 & 0.00041 \end{pmatrix} = \begin{pmatrix} 97.0149 & 0.0990 & -2.1683 \\ 0.0990 & 0.1175 & -0.0340 \\ -2.1683 & -0.0340 & 0.0618 \end{pmatrix}$$

so, $Var(b_0)=97.0149$, $Var(b_1)=0.1175$, $Var(b_2)=0.0618$, $Var(b_{12})=0.0340$, etc.

The variance - covariance matrix can also be used to obtain confidence intervals about estimates of $\hat{Y}$ for particular values of X and X. The most versatile approach is to use matrix algebra in these calculations. The equation is

$$S^2_{\hat{Y}} = MSE\ (L'(X'X)^{-1}L)$$

where L is a vector of values for X corresponding to $\hat{Y}$. It may also be a vector of hypothesized X values for which a variance is needed.

For example, if we wish to predict the response ($\hat{Y}$) and its variance when X = 4 and X = 24, first we would calculate the response,

$$\hat{Y}=66.4654+1.2902X_{1i}+0.1110X_{2i} = 66.4654+1.2902(4)+0.1110(24) = 68.9622$$

Using L = [1 4 24], (note that a 1 is included for the intercept) the variance of the estimate is then,

$$S^2_{\hat{Y}} = 150.092[1\ \ 4\ \ 24]\begin{pmatrix} 0.64637 & 0.00066 & -0.01445 \\ 0.00066 & 0.00078 & -0.00023 \\ -0.01445 & -0.00023 & 0.00041 \end{pmatrix}\begin{pmatrix} 1 \\ 4 \\ 24 \end{pmatrix} = 24.6782$$

and the standard error is $\sqrt{24.6782}$ = 4.9677.

The sweepout technique is not the only method of matrix inversion. However, its application to the augmented matrix described above is a relatively simple and versatile method of obtaining most of the results commonly desired from a multiple regression analysis.

REFERENCES

Goodnight, J. H. 1978. The Sweep Operator: Its importance in statistical computing. in Proc. Eleventh Annual Symposium on the INTERFACE. Gallant A. R. and T. M. Gerig (ed), Inst. Statistics, N. C. State University, Raleigh, N. C.

Three factor multiple regression from Snedecor and Cochran (1967), table 13.10.1, page 405.

Y = estimated plant available phosphorus in the soil (20 C)
X_1 = inorganic phosphorus
X_2 = organic phosphorus soluble in K_2CO_3 and hydrolized by hypobromite
X_3= organic phosphorus soluble in K_2CO_3 and NOT hydrolized by hypobromite

All least squares regression analyses start with the same three matrices.

$$X = \begin{bmatrix} 1 & 0.4 & 53 & 158 \\ 1 & 0.4 & 23 & 163 \\ 1 & 3.1 & 19 & 37 \\ 1 & 0.6 & 34 & 157 \\ 1 & 4.7 & 24 & 59 \\ 1 & 1.7 & 65 & 123 \\ 1 & 9.4 & 44 & 46 \\ 1 & 10.1 & 31 & 117 \\ 1 & 11.6 & 29 & 173 \\ 1 & 12.6 & 58 & 112 \\ 1 & 10.9 & 37 & 111 \\ 1 & 23.1 & 46 & 114 \\ 1 & 23.1 & 50 & 134 \\ 1 & 21.6 & 44 & 73 \\ 1 & 23.1 & 56 & 168 \\ 1 & 1.9 & 36 & 143 \\ 1 & 26.8 & 58 & 202 \\ 1 & 29.9 & 51 & 124 \end{bmatrix} \qquad Y = \begin{bmatrix} 64 \\ 60 \\ 71 \\ 61 \\ 54 \\ 77 \\ 81 \\ 93 \\ 93 \\ 51 \\ 76 \\ 96 \\ 77 \\ 93 \\ 95 \\ 54 \\ 168 \\ 99 \end{bmatrix}$$

$$X'X = \begin{bmatrix} 18 & 215 & 758 & 2214 \\ 215 & 4321.02 & 10139.5 & 27645 \\ 758 & 10139.5 & 35076 & 96598 \\ 2214 & 27645 & 96598 & 307894 \end{bmatrix} \qquad X'Y = \begin{bmatrix} 1463 \\ 20706.2 \\ 63825 \\ 187542 \end{bmatrix}$$

$$Y'Y = \begin{bmatrix} 131299 \end{bmatrix}$$

Create a fully augmented matrix of the form;

$$\begin{bmatrix} X'X & X'Y & I \\ (X'Y)' & Y'Y & 0 \end{bmatrix}$$

The resulting matrix contains;

X_0	X_1	X_2	X_3	$X'Y$	c_0	c_1	c_2	c_3
n	ΣX_1	ΣX_2	ΣX_3	ΣY	1	0	0	0
ΣX_1	ΣX_1^2	ΣX_1X_2	ΣX_1X_3	ΣX_1Y	0	1	0	0
ΣX_2	ΣX_1X_2	ΣX_2^2	ΣX_2X_3	ΣX_2Y	0	0	1	0
ΣX_3	ΣX_1X_3	ΣX_2X_3	ΣX_3^2	ΣX_3Y	0	0	0	1
ΣY	ΣX_1Y	ΣX_2Y	ΣX_3Y	ΣY^2	0	0	0	0

Numerically for this problem given previously the matrix is;

X_0	X_1	X_2	X_3	$X'Y$	c_0	c_1	c_2	c_3
18	215	758	2214	1463	1	0	0	0
215	4321.02	10139.5	27645	20706.2	0	1	0	0
758	10139.5	35076	96598	63825	0	0	1	0
2214	27645	96598	307894	187542	0	0	0	1
1463	20706.2	63825	187542	131299	0	0	0	0

The first step (divide row 1 by value$_{1,1}$) in the sweepout technique produces,

X_0	X_1	X_2	X_3	$X'Y$	c_0	c_1	c_2	c_3
1	11.944444	42.111111	123	81.277778	0.055556	0	0	0
215	4321.02	10139.5	27645	20706.2	0	1	0	0
758	10139.5	35076	96598	63825	0	0	1	0
2214	27645	96598	307894	187542	0	0	0	1
1463	20706.2	63825	187542	131299	0	0	0	0

And after sweeping out the first column (subtracting a multiple of row 1 from all other rows) we have;

X_0	X_1	X_2	X_3	$X'Y$	c_0	c_1	c_2	c_3
1	11.944444	42.111111	123	81.277778	0.055556	0	0	0
0	1752.964444	1085.611111	1200	3231.477778	-11.944444	1	0	0
0	1085.611111	3155.777778	3364	2216.444444	-42.111111	0	1	0
0	1200	3364	35572	7593	-123	0	0	1
0	3231.477778	2216.444444	7593	12389.61111	-81.277778	0	0	0

We start the second column sweep by dividing row 2 by value$_{2,2}$,

X_0	X_1	X_2	X_3	$X'Y$	c_0	c_1	c_2	c_3
1	11.944444	42.111111	123	81.277778	0.055556	0	0	0
0	1	0.6193	0.684555	1.843436	-0.006814	0.00057	0	0
0	1085.611111	3155.777778	3364	2216.444444	-42.111111	0	1	0
0	1200	3364	35572	7593	-123	0	0	1
0	3231.477778	2216.444444	7593	12389.61111	-81.277778	0	0	0

and finish sweeping the second column to obtain;

X_0	X_1	X_2	X_3	$X'Y$	c_0	c_1	c_2	c_3
1	0	34.713915	114.823375	59.258959	0.136943	-0.006814	0	0
0	1	0.6193	0.684555	1.843436	-0.006814	0.00057	0	0
0	0	2483.458674	2620.839842	215.189831	-34.713915	-0.6193	1	0
0	0	2620.839842	34750.53439	5380.87679	-114.823375	-0.684555	0	1
0	0	215.189831	5380.87679	6432.588616	-59.258959	-1.843436	0	0

The third column starts with,

X_0	X_1	X_2	X_3	$X'Y$	c_0	c_1	c_2	c_3
1	0	34.713915	114.823375	59.258959	0.136943	-0.006814	0	0
0	1	0.6193	0.684555	1.843436	-0.006814	0.00057	0	0
0	0	1	1.055318	0.086649	-0.013978	-0.000249	0.000403	0
0	0	2620.839842	34750.53439	5380.87679	-114.823375	-0.684555	0	1
0	0	215.189831	5380.87679	6432.588616	-59.258959	-1.843436	0	0

and after being swept out produces,

X_0	X_1	X_2	X_3	$X'Y$	c_0	c_1	c_2	c_3
1	0	0	78.189139	56.251024	0.622176	0.001843	-0.013978	0
0	1	0	0.030996	1.789774	0.001843	0.000725	-0.000249	0
0	0	1	1.055318	0.086649	-0.013978	-0.000249	0.000403	0
0	0	0	31984.71367	5153.782984	-78.189139	-0.030996	-1.055318	1
0	0	0	5153.782984	6413.942579	-56.251024	-1.789774	-0.086649	0

Finally the fourth column in the X´X matrix is started and swept out,

X_0	X_1	X_2	X_3	$X'Y$	c_0	c_1	c_2	c_3
1	0	0	78.189139	56.251024	0.622176	0.0018428	-0.0139781	0
0	1	0	0.030996	1.789774	0.001843	0.0007249	-0.0002494	0
0	0	1	1.055318	0.086649	-0.013978	-0.0002494	0.0004027	0
0	0	0	1	0.161133	-0.002445	-0.0000010	-0.0000330	0.000031
0	0	0	5153.782984	6413.942579	-56.251024	-1.7897741	-0.0866492	0

and the final result is;

X_0	X_1	X_2	X_3	$X'Y$	c_0	c_1	c_2	c_3
1	0	0	0	43.652198	0.813316	0.0019185	-0.0113982	-0.002445
0	1	0	0	1.78478	0.001919	0.0007249	-0.0002483	-0.000001
0	0	1	0	-0.083397	-0.011398	-0.0002483	0.0004375	-0.000033
0	0	0	1	0.161133	-0.002445	-0.0000010	-0.0000330	0.000031
0	0	0	0	5583.499658	-43.652198	-1.7847797	0.0833971	-0.161133

The resulting matrix is of the form;

I	B	$(X'X)^{-1}$
0	SSE	$-B'$

and contains the values

X_0	X_1	X_2	X_3	X´Y	c_0	c_1	c_2	c_3
1	0	0	0	b_0	c_{00}	c_{01}	c_{02}	c_{03}
0	1	0	0	b_1	c_{10}	c_{11}	c_{12}	c_{13}
0	0	1	0	b_2	c_{20}	c_{21}	c_{22}	c_{23}
0	0	0	1	b_3	c_{30}	c_{31}	c_{32}	c_{33}
0	0	0	0	SSE	$-b_0$	$-b_1$	$-b_2$	$-b_3$

The solution to the regression equation is then,

$$Y_i = 43.652 + 1.785X_{1i} - 0.083X_{2i} + 0.161X_{3i} + e$$

The sums of squares are given by the sequential reduction in the YY matrix

MATRIX	Y´Y VALUE	INTERPRETATION of the REPLACEMENT VALUE	DIFFERENCE from PREVIOUS VALUE	INTERPRETATION of the DIFFERENCE
Original	131299	ΣY^2 (uncorrected)		
Col 1 sweep	12389.6111	$\Sigma Y^2-(\Sigma Y)^2/n = SSY\|X_0$	118909.3840	$(Y)^2/n$ = C.F.
Col 2 sweep	6432.5886	$SSY\|X_0,X_1$	5957.0225	$SeqSSX_1$
Col 3 sweep	6413.9426	$SSY\|X_0,X_1,X_2$	18.6460	$SeqSSX_2$
Col 4 Sweep	5583.4997	$SSY\|X_0,X_1,X_2,X_3 = SSE$	830.4429	$SeqSSX_3$

Partial sums of squares, or fully adjusted sums of squares, are given by

PARTIAL SS = $\frac{b_k}{c_{kk}}$

Partial $SSX_1 = b_1^2/c_{11} = 1.7848^2 / 0.0007249 = 4394.1523$

Partial $SSX_2 = b_2^2/c_{22} = 0.08340^2 / 0.0004375 = 15.8979$

Partial $SSX_3 = b_3^2/c_{33} = 0.1611^2 / 0.00003127 = 830.4429$

Recall that I number the positions in the X´X matrix differently, from k = 0, 1, ... , p (where p is the number of parameters excluding the intercept) instead of starting at 1 as other matrices. This is done in order to be able to associate the matrix position with the regression coefficient subscript.

```
1          ************************************************************************;
2          *** EXST7034 Example 1 using PC-SAS - Airline vial breakage          ***;
3          *** Problem from Neter, Kutner, Nachtsheim & Wasserman 1996, #1.21   ***;
4          ************************************************************************;
5          OPTIONS LS=88 PS=256 NOCENTER NODATE NONUMBER;
6          DATA ONE; INFILE CARDS MISSOVER;
7             TITLE1 'EXST7015 Multiple Regression from Snedecor & Cochran (1967)';
8          *     LABEL Y ='Plant available phosphorus';
9          *     LABEL X1='Inorganic phosphorus';
10         *     LABEL X2='Hydrolized organic phosphorus';
11         *     LABEL X3='Nonhydrolized phosphorus';
12            INPUT X1 X2 X3 Y;
13            X0 = 1;
14         CARDS; RUN;
NOTE: The data set WORK.ONE has 18 observations and 5 variables.
NOTE: DATA statement used:
      real time           0.05 seconds
      cpu time            0.05 seconds
35         PROC IML; ***RESET PRINT;
NOTE: IML Ready
35       !                          USE ONE;
36         READ ALL VAR{X0 X1 X2 X3} INTO X;
37         READ ALL VAR{Y} INTO Y;
38         CLOSE ONE;
39         ****************************************;
40         **** Calculation of the Full Model    ****;
41         ****************************************;
42         **** Intermediate calculations        ****;
43         ****************************************;
44             N = NROW(X);   P = NCOL(X);
45             YPY = Y`*Y;    XPX = X`*X;   XPY = X`*Y;
46         PRINT ,,"Intermediate statistics",
47               N P XPX YPY XPY;

EXST7015 Multiple Regression from Snedecor & Cochran (1967)

Intermediate statistics
        N         P       XPX                                        YPY       XPY
       18         4        18       215       758      2214    131299      1463
                           215   4321.02   10139.5     27645             20706.2
                           758   10139.5     35076     96598               63825
                          2214     27645     96598    307894              187542

48         ****************************************;
49         **** Solution & Analysis of Variance  ****;
50         ****************************************;
51             XSumSq = VECDIAG(XPX);
52             XPXINV = INV(XPX);
53             B = XPXINV * XPY;        CF = (J(1,N,1)*Y)##2/N;
54             USSTOTAL = YPY;          USSREG = B`*XPY;
55             SSTOTAL = YPY - CF;      SSREG = B`*XPY - CF;
56             SSERROR = YPY - B`*XPY;  dfE = N-P;
57             MSE = SSERROR / (N-P);   RSquare = SSReg / SSTotal;
58         PRINT ,,"Reg coefficients and ANOVA table info",
59               B XPXINV XSumSq CF USSTOTAL USSREG,,
60               SSTOTAL SSREG SSERROR MSE DFE RSquare;


Reg coefficients and ANOVA table info
             B    XPXINV
          Col1      Col2      Col3      Col4      Col5
ROW1 43.652198 0.8133157 0.0019185 -0.011398 -0.002445
ROW2 1.7847797 0.0019185 0.0007249 -0.000248 -9.691E-7
ROW3 -0.083397 -0.011398 -0.000248 0.0004375 -0.000033
ROW4 0.1611327 -0.002445 -9.691E-7 -0.000033 0.0000313
```

```
        XSUMSQ        CF  USSTOTAL    USSREG
          Col6      Col7      Col8      Col9
ROW1        18 118909.39    131299  125715.5
ROW2   4321.02
ROW3     35076
ROW4    307894

  SSTOTAL     SSREG   SSERROR       MSE       DFE   RSQUARE
12389.611 6806.1115 5583.4997  398.8214        14 0.5493402

61         ********************************************;
62         **** Test of B and observation diagnostics ****;
63         ********************************************;
64             hatmatrix = X*XPXinv*X`;
65             VARCOV = MSE * XPXINV;
66             VarResid = (I(N) - HatMatrix) * MSE;
67             VarPred = HatMatrix * MSE;
68         PRINT ,,"Variance information",,
69               HatMatrix, VarCov, VarResid, VarPred;

Variance information
                                        HATMATRIX
          COL1      COL2      COL3      COL4      COL5      COL6      COL7      COL8      COL9
ROW1  0.2804123 0.0898651 -0.060047 0.1544019 -0.029023 0.3175679 0.0385927  -0.00039 0.0117839
ROW2  0.0898651 0.3037333  0.082666 0.2166979 0.0843716  -0.06289 -0.093709 0.1266609 0.2407672
ROW3  -0.060047  0.082666 0.3430106 0.0382104 0.2744963 -0.054615 0.1951249 0.1245554  0.026436
ROW4  0.1544019 0.2166979 0.0382104 0.1870174 0.0484735 0.0770947 -0.035366 0.0796078 0.1490809
ROW5  -0.029023 0.0843716 0.2744963 0.0484735 0.2226066   -0.0292  0.157257 0.1097935 0.0385842
ROW6  0.3175679  -0.06289 -0.054615 0.0770947   -0.0292  0.477301  0.170022 -0.055321 -0.141838
ROW7  0.0385927 -0.093709 0.1951249 -0.035366  0.157257  0.170022 0.2587848 0.0300573 -0.119465
ROW8   -0.00039 0.1266609 0.1245554 0.0796078 0.1097935 -0.055321 0.0300573 0.0985578 0.1192347
ROW9  0.0117839 0.2407672  0.026436 0.1490809 0.0385842 -0.141838 -0.119465 0.1192347 0.2500576
ROW10 0.1429617  -0.07592 -0.004399 0.0072644 0.0067074 0.2546866 0.1447647 -0.012308 -0.080453
ROW11 0.0251027 0.0714964 0.1051302 0.0554465 0.0935431 0.0140087 0.0670627  0.073386 0.0658236
ROW12 -0.049881 -0.028818 0.0465051 -0.033363 0.0448009  -0.03516 0.0683344 0.0535076 0.0421984
ROW13 -0.009053 -0.018226 -0.012155 -0.013942 0.0008233 0.0001648 0.0349846 0.0403041 0.0529216
ROW14 -0.076766 -0.087285 0.1361661   -0.0753 0.1111148 -0.010053 0.1648528 0.0523623 -0.029481
ROW15 0.0551725  0.005233 -0.107817 0.0205552 -0.070728 0.0501715 -0.024909  0.021222 0.0769911
ROW16  0.142468 0.1718181 0.0541392 0.1564136 0.0590982 0.0955999 0.0042628 0.0705234 0.1103796
ROW17 0.0523988 0.0423982 -0.185349 0.0393674  -0.12879 0.0014372 -0.109083 0.0242436 0.1412053
ROW18 -0.085567  -0.06886 -0.002058  -0.07166 0.0060701 -0.068979 0.0484307 0.0440025 0.0457739

                                        HATMATRIX
         COL10     COL11     COL12     COL13     COL14     COL15     COL16     COL17     COL18
ROW1  0.1429617 0.0251027 -0.049881 -0.009053 -0.076766 0.0551725  0.142468 0.0523988 -0.085567
ROW2   -0.07592 0.0714964 -0.028818 -0.018226 -0.087285  0.005233 0.1718181 0.0423982  -0.06886
ROW3  -0.004399 0.1051302 0.0465051 -0.012155 0.1361661 -0.107817 0.0541392 -0.185349 -0.002058
ROW4  0.0072644 0.0554465 -0.033363 -0.013942   -0.0753 0.0205552 0.1564136 0.0393674  -0.07166
ROW5  0.0067074 0.0935431 0.0448009 0.0008233 0.1111148 -0.070728 0.0590982  -0.12879 0.0060701
ROW6  0.2546866 0.0140087  -0.03516 0.0001648 -0.010053 0.0501715 0.0955999 0.0014372 -0.068979
ROW7  0.1447647 0.0670627 0.0683344 0.0349846 0.1648528 -0.024909 0.0042628 -0.109083 0.0484307
ROW8  -0.012308  0.073386 0.0535076 0.0403041 0.0523623  0.021222 0.0705234 0.0242436 0.0440025
ROW9  -0.080453 0.0658236 0.0421984 0.0529216 -0.029481 0.0769911 0.1103796 0.1412053 0.0457739
ROW10 0.1764697  0.033044 0.0525862 0.0638153 0.0790943 0.0771839  0.029232 0.0491451  0.056126
ROW11  0.033044  0.065554 0.0571015 0.0466681 0.0679039 0.0301959 0.0558435 0.0219846 0.0507046
ROW12 0.0525862 0.0571015 0.1358751 0.1243763 0.1439654  0.105446 -0.019752  0.115985 0.1762924
ROW13 0.0638153 0.0466681 0.1243763  0.127104 0.1115795 0.1314869 -0.006374 0.1572317 0.1682902
ROW14 0.0790943 0.0679039 0.1439654 0.1115795 0.2009727 0.0564609 -0.043702 0.0253717 0.1727436
ROW15 0.0771839 0.0301959  0.105446 0.1314869 0.0564609 0.1742998 0.0170395 0.2268269 0.1551691
ROW16  0.029232 0.0558435 -0.019752 -0.006374 -0.043702 0.0170395 0.1355044 0.0224105 -0.054905
ROW17 0.0491451 0.0219846  0.115985 0.1572317 0.0253717 0.2268269 0.0224105 0.3187639 0.1844514
ROW18  0.056126 0.0507046 0.1762924 0.1682902 0.1727436 0.1551691 -0.054905 0.1844514 0.2439751


                VARCOV
324.36769 0.7651496 -4.545863  -0.97495
0.7651496 0.2891165 -0.099046 -0.000386
-4.545863 -0.099046 0.1744779 -0.013159
 -0.97495 -0.000386 -0.013159 0.0124691
```

VARRESID

	COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8	COL9
ROW1	286.98696	-35.84012	23.948119	-61.57878	11.574855	-126.6529	-15.39159	0.1556033	-4.699675
ROW2	-35.84012	277.68605	-32.96898	-86.42376	-33.64919	25.081681	37.373055	-50.51509	-96.02312
ROW3	23.948119	-32.96898	262.02145	-15.23913	-109.475	21.781438	-77.82	-49.67534	-10.54326
ROW4	-61.57878	-86.42376	-15.23913	324.23485	-19.33228	-30.74702	14.104634	-31.7493	-59.45664
ROW5	11.574855	-33.64919	-109.475	-19.33228	310.04115	11.645401	-62.71748	-43.78801	-15.38821
ROW6	-126.6529	25.081681	21.781438	-30.74702	11.645401	208.46357	-67.80842	22.063003	56.567965
ROW7	-15.39159	37.373055	-77.82	14.104634	-62.71748	-67.80842	295.6125	-11.9875	47.645317
ROW8	0.1556033	-50.51509	-49.67534	-31.7493	-43.78801	22.063003	-11.9875	359.51446	-47.55336
ROW9	-4.699675	-96.02312	-10.54326	-59.45664	-15.38821	56.567965	47.645317	-47.55336	299.09308
ROW10	-57.01619	30.278631	1.7545655	-2.897204	-2.675035	-101.5745	-57.73525	4.9088065	32.086555
ROW11	-10.01151	-28.51429	-41.92816	-22.11326	-37.30698	-5.586959	-26.74603	-29.26792	-26.25185
ROW12	19.893595	11.493322	-18.54721	13.305931	-17.86754	14.022749	-27.25323	-21.33999	-16.82964
ROW13	3.6107134	7.2687205	4.8478018	5.5605225	-0.328363	-0.065707	-13.95262	-16.07413	-21.10627
ROW14	30.615914	34.811173	-54.30597	30.031299	-44.31496	4.0094046	-65.74683	-20.88321	11.757849
ROW15	-22.00399	-2.087014	42.999563	-8.197863	28.20801	-20.00946	9.9342765	-8.463777	-30.70568
ROW16	-56.81929	-68.52473	-21.59186	-62.38111	-23.56965	-38.12728	-1.700081	-28.12622	-44.02175
ROW17	-20.89776	-16.90933	73.921145	-15.70057	51.36413	-0.573176	43.504443	-9.668863	-56.31569
ROW18	34.126017	27.462993	0.8208314	28.579685	-2.420871	27.510172	-19.31521	-17.54915	-18.25561

	COL10	COL11	COL12	COL13	COL14	COL15	COL16	COL17	COL18
ROW1	-57.01619	-10.01151	19.893595	3.6107134	30.615914	-22.00399	-56.81929	-20.89776	34.126017
ROW2	30.278631	-28.51429	11.493322	7.2687205	34.811173	-2.087014	-68.52473	-16.90933	27.462993
ROW3	1.7545655	-41.92816	-18.54721	4.8478018	-54.30597	42.999563	-21.59186	73.921145	0.8208314
ROW4	-2.897204	-22.11326	13.305931	5.5605225	30.031299	-8.197863	-62.38111	-15.70057	28.579685
ROW5	-2.675035	-37.30698	-17.86754	-0.328363	-44.31496	28.20801	-23.56965	51.36413	-2.420871
ROW6	-101.5745	-5.586959	14.022749	-0.065707	4.0094046	-20.00946	-38.12728	-0.573176	27.510172
ROW7	-57.73525	-26.74603	-27.25323	-13.95262	-65.74683	9.9342765	-1.700081	43.504443	-19.31521
ROW8	4.9088065	-29.26792	-21.33999	-16.07413	-20.88321	-8.463777	-28.12622	-9.668863	-17.54915
ROW9	32.086555	-26.25185	-16.82964	-21.10627	11.757849	-30.70568	-44.02175	-56.31569	-18.25561
ROW10	328.44151	-13.17867	-20.97248	-25.45092	-31.54451	-30.78259	-11.65834	-19.60013	-22.38427
ROW11	-13.17867	372.67707	-22.77332	-18.61225	-27.08154	-12.04277	-22.27157	-8.767929	-20.22209
ROW12	-20.97248	-22.77332	344.6315	-49.60392	-57.41649	-42.05412	7.8773327	-46.2573	-70.30919
ROW13	-25.45092	-18.61225	-49.60392	348.12959	-44.50028	-52.4398	2.542002	-62.70737	-67.11773
ROW14	-31.54451	-27.08154	-57.41649	-44.50028	318.66919	-22.5178	17.429388	-10.11878	-68.89385
ROW15	-30.78259	-12.04277	-42.05412	-52.4398	-22.5178	329.30692	-6.795714	-90.46342	-61.88478
ROW16	-11.65834	-22.27157	7.8773327	2.542002	17.429388	-6.795714	344.77935	-8.937794	21.897309
ROW17	-19.60013	-8.767929	-46.2573	-62.70737	-10.11878	-90.46342	-8.937794	271.69155	-73.56318
ROW18	-22.38427	-20.22209	-70.30919	-67.11773	-68.89385	-61.88478	21.897309	-73.56318	301.51891

VARPRED

	COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8	COL9
ROW1	111.83444	35.84012	-23.94812	61.578779	-11.57485	126.65288	15.391587	-0.155603	4.6996747
ROW2	35.84012	121.13535	32.968984	86.423762	33.649188	-25.08168	-37.37306	50.515093	96.023123
ROW3	-23.94812	32.968984	136.79996	15.239133	109.47499	-21.78144	77.820004	49.675342	10.543256
ROW4	61.578779	86.423762	15.239133	74.586558	19.33228	30.747018	-14.10463	31.749302	59.456643
ROW5	-11.57485	33.649188	109.47499	19.33228	88.780258	-11.6454	62.717477	43.788012	15.388205
ROW6	126.65288	-25.08168	-21.78144	30.747018	-11.6454	190.35784	67.808419	-22.063	-56.56797
ROW7	15.391587	-37.37306	77.820004	-14.10463	62.717477	67.808419	103.20891	11.987497	-47.64532
ROW8	-0.155603	50.515093	49.675342	31.749302	43.788012	-22.063	11.987497	39.306941	47.553364
ROW9	4.6996747	96.023123	10.543256	59.456643	15.388205	-56.56797	-47.64532	47.553364	99.728329
ROW10	57.016188	-30.27863	-1.754566	2.8972043	2.6750354	101.57448	57.735252	-4.908806	-32.08656
ROW11	10.011508	28.514295	41.928156	22.11326	37.30698	5.5869588	26.746026	29.26792	26.251846
ROW12	-19.89359	-11.49332	18.547213	-13.30593	17.86754	-14.02275	27.253228	21.339992	16.829643
ROW13	-3.610713	-7.26872	-4.847802	-5.560523	0.3283627	0.0657071	13.952618	16.074133	21.106272
ROW14	-30.61591	-34.81117	54.305973	-30.0313	44.314959	-4.009405	65.746829	20.883211	-11.75785
ROW15	22.00399	2.0870137	-42.99956	8.1978632	-28.20801	20.009458	-9.934277	8.4637772	30.705683
ROW16	56.819294	68.524726	21.591861	62.381108	23.569646	38.127278	1.7000805	28.126224	44.021751
ROW17	20.897759	16.909326	-73.92114	15.700565	-51.36413	0.5731759	-43.50444	9.6688633	56.315688
ROW18	-34.12602	-27.46299	-0.820831	-28.57968	2.4208707	-27.51017	19.315205	17.549146	18.255611

	COL10	COL11	COL12	COL13	COL14	COL15	COL16	COL17	COL18
ROW1	57.016188	10.011508	-19.89359	-3.610713	-30.61591	22.00399	56.819294	20.897759	-34.12602
ROW2	-30.27863	28.514295	-11.49332	-7.26872	-34.81117	2.0870137	68.524726	16.909326	-27.46299
ROW3	-1.754566	41.928156	18.547213	-4.847802	54.305973	-42.99956	21.591861	-73.92114	-0.820831
ROW4	2.8972043	22.11326	-13.30593	-5.560523	-30.0313	8.1978632	62.381108	15.700565	-28.57968
ROW5	2.6750354	37.30698	17.86754	0.3283627	44.314959	-28.20801	23.569646	-51.36413	2.4208707
ROW6	101.57448	5.5869588	-14.02275	0.0657071	-4.009405	20.009458	38.127278	0.5731759	-27.51017
ROW7	57.735252	26.746026	27.253228	13.952618	65.746829	-9.934277	1.7000805	-43.50444	19.315205
ROW8	-4.908806	29.26792	21.339992	16.074133	20.883211	8.4637772	28.126224	9.6688633	17.549146
ROW9	-32.08656	26.251846	16.829643	21.106272	-11.75785	30.705683	44.021751	56.315688	18.255611
ROW10	70.379899	13.178666	20.972484	25.450916	31.544509	30.782589	11.658341	19.60013	22.384266
ROW11	13.178666	26.144335	22.773316	18.612247	27.081535	12.042766	22.271568	8.7679292	20.222093
ROW12	20.972484	22.773316	54.189903	49.60392	57.416487	42.054115	-7.877333	46.257299	70.309194
ROW13	25.450916	18.612247	49.60392	50.691809	44.500284	52.439803	-2.542002	62.707365	67.117728
ROW14	31.544509	27.081535	57.416487	44.500284	80.152218	22.517804	-17.42939	10.118777	68.893847
ROW15	30.782589	12.042766	42.054115	52.439803	22.517804	69.514481	6.7957143	90.46342	61.884775
ROW16	11.658341	22.271568	-7.877333	-2.542002	-17.42939	6.7957143	54.042052	8.9377938	-21.89731
ROW17	19.60013	8.7679292	46.257299	62.707365	10.118777	90.46342	8.9377938	127.12985	73.56318
ROW18	22.384266	20.222093	70.309194	67.117728	68.893847	61.884775	-21.89731	73.56318	97.302494

```
70         ********************************************;
71             StdErrB = sqrt(vecdiag(VarCov));
72             t = B / StdErrB;
73             Probt = 1 - ProbF(t#t, 1, dfe);
74           PRINT ,,"Tests of regression coefficients",
75                 StdErrB t Probt;

Tests of regression coefficients
  STDERRB         T     PROBT
18.010211 2.4237472 0.0294949
0.5376955  3.319313 0.0050639
0.4177056 -0.199655 0.8446212
0.1116652 1.4429978 0.1710214

76         ********************************************;
77             hatvalues = vecdiag(hatmatrix);
78             YHat = X * B;
79             Resid = Y - YHat;
80             /*Must specify t value */
80       !                                TVal = 2.306;
81             LowerCLM = YHAT - TVAL # SQRT(hatvalues*MSE);
82             UpperCLM = YHAT + TVAL # SQRT(hatvalues*MSE);
83             LowerCLI = YHAT - TVAL # SQRT(hatvalues*MSE+MSE);
84             UpperCLI = YHAT + TVAL # SQRT(hatvalues*MSE+MSE);
85           PRINT ,,"Observation information" ,
86                 Y YHAT RESID HatValues LowerCLM UpperCLM LowerCLI UpperCLI;
87           quit;

Observation information
        Y       YHAT      RESID  HATVALUES   LOWERCLM   UPPERCLM   LOWERCLI   UPPERCLI
       64  65.405031  -1.405031  0.2804123  41.018665  89.791397  13.294745  117.51532
       60  68.712606  -8.712606  0.3037333  43.332423  94.092789  16.129902  121.29531
       71   53.56238   17.43762  0.3430106  26.591055  80.533705  0.1934808  106.93128
       61  67.185398  -6.185398  0.1870174  47.269973  87.100824  17.011588  117.35921
       54  59.545962  -5.545962  0.2226066  37.818074  81.273849  8.6255517  110.46637
       77  61.084836  15.915164   0.477301  29.268901   92.90077  5.1112636  117.05841
       81   64.17176   16.82824  0.2587848  40.744695  87.598825  12.503448  115.84007
       93  77.945689  15.054311  0.0985578  63.488165  92.403213  29.677614  126.21376
       93  89.813083   3.186917  0.2500576  66.784428  112.84174  38.324191  141.30198
       51  79.350254  -28.35025  0.1764697  60.004591  98.695917  29.399861  129.30065
       76  77.906334  -1.906334   0.065554  66.115403  89.697265  30.368842  125.44383
       96  99.413471  -3.413471  0.1358751   82.43813  116.38881  50.332423  148.49452
       77  102.30254  -25.30254   0.127104  85.884236  118.72084  53.411354  151.19372
       93  90.296655  2.7033452  0.2009727  69.651551  110.94176   39.82877  140.76454
       95  107.28067  -12.28067  0.1742998  88.054311  126.50702  57.376359  157.18497
       54   67.08296  -13.08296  0.1355044  50.130793  84.035127  18.009922    116.156
      168  119.19607  48.803933  0.3187639  93.195485  145.19665  66.311123  172.08101
       99  112.74431  -13.74431  0.2439751  89.997463  135.49117  61.380842  164.10779
```