

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.

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

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

$$Y'Y = [131299]$$

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	$\Sigma X_1 X_2$	$\Sigma X_1 X_3$	$\Sigma X_1 Y$	0	1	0	0
ΣX_2	$\Sigma X_1 X_2$	ΣX_2^2	$\Sigma X_2 X_3$	$\Sigma X_2 Y$	0	0	1	0
ΣX_3	$\Sigma X_1 X_3$	$\Sigma X_2 X_3$	ΣX_3^2	$\Sigma X_3 Y$	0	0	0	1
ΣY	$\Sigma X_1 Y$	$\Sigma X_2 Y$	$\Sigma X_3 Y$	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;

$$\left[\begin{array}{c|c|c} \mathbf{I} & \mathbf{B} & (\mathbf{X}'\mathbf{X})^{-1} \\ \hline \mathbf{0} & \mathbf{SSE} & -\mathbf{B}' \end{array} \right]$$

and contains the values

$$\left[\begin{array}{cccc|cccc} \mathbf{X}_0 & \mathbf{X}_1 & \mathbf{X}_2 & \mathbf{X}_3 & \mathbf{X}'\mathbf{Y} & \mathbf{c}_0 & \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3 \\ \hline 1 & 0 & 0 & 0 & \mathbf{b}_0 & \mathbf{c}_{00} & \mathbf{c}_{01} & \mathbf{c}_{02} & \mathbf{c}_{03} \\ 0 & 1 & 0 & 0 & \mathbf{b}_1 & \mathbf{c}_{10} & \mathbf{c}_{11} & \mathbf{c}_{12} & \mathbf{c}_{13} \\ 0 & 0 & 1 & 0 & \mathbf{b}_2 & \mathbf{c}_{20} & \mathbf{c}_{21} & \mathbf{c}_{22} & \mathbf{c}_{23} \\ 0 & 0 & 0 & 1 & \mathbf{b}_3 & \mathbf{c}_{30} & \mathbf{c}_{31} & \mathbf{c}_{32} & \mathbf{c}_{33} \\ \hline 0 & 0 & 0 & 0 & \mathbf{SSE} & -\mathbf{b}_0 & -\mathbf{b}_1 & -\mathbf{b}_2 & -\mathbf{b}_3 \end{array} \right]$$

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	$\Sigma Y^2 - (\Sigma Y)^2/n = \text{SSY} X_0$	118909.3840	$(Y)^2/n = \text{C.F.}$
Col 2 sweep	6432.5886	$\text{SSY} X_0, X_1$	5957.0225	SeqSSX ₁
Col 3 sweep	6413.9426	$\text{SSY} X_0, X_1, X_2$	18.6460	SeqSSX ₂
Col 4 Sweep	5583.4997	$\text{SSY} X_0, X_1, X_2, X_3 = \text{SSE}$	830.4429	SeqSSX ₃

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

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

$$\text{Partial SSX}_1 = b_1^2/c_{11} = 1.7848^2 / 0.0007249 = 4394.1523$$

$$\text{Partial SSX}_2 = b_2^2/c_{22} = 0.08340^2 / 0.0004375 = 15.8979$$

$$\text{Partial SSX}_3 = b_3^2/c_{33} = 0.1611^2 / 0.00003127 = 830.4429$$

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.