Matrix Example

Consider the following small data set:

X	Y
1	3
2	5
3	10
4	10

It can be shown that

$$\overline{X} = 2.5 \qquad \overline{Y} = 7.0$$
$$\sum (X_i - \overline{X})^2 = 5 \qquad \sum (Y_i - \overline{Y})^2 = 38$$
$$\sum (X_i - \overline{X})(Y_i - \overline{Y}) = 13$$

Using these quantities, we compute

$$\begin{aligned} b_1 &= \frac{13}{5} = 2.6 \\ b_0 &= 7.0 - 2.6(2.5) = 0.5 \end{aligned}$$

so the fitted and residual values are

X	Y	$\widehat{Y}$	e_i
1	3	3.1	-0.1
2	5	5.7	-0.7
3	10	8.3	1.7
4	10	10.9	-0.9

and the estimated variance,

$$\frac{-0.1^2 + -0.7^2 + 1.7^2 + -0.9^2}{3 - 1} = 2.1$$

This means the estimated variance of each parameter is

$$\begin{aligned} s^2(b_1) &= 2.1/5 = 0.42 \\ s^2(b_0) &= 2.1(1/4 + 2.5^2/5) = 3.15 \end{aligned}$$

With SAS, we can obtain some of the matrix manipulations using various proc reg options. Below is the SAS file that does this. The options are

covb	:	gives the covariance matrix for the estimated parameters
xpx	:	gives the X'X, X'Y and Y'Y matrices
i	:	gives the inverse of X'X, the parameters, and SSE

```
**** SAS CODE ****;

data example;
 input x y;
 cards;
      1      3
      2      5
      3     10
      4     10
;

proc reg;
 model y=x / covb xpx i;
run;
```

Below is the output

The REG Procedure

Model Crossproducts X'X X'Y Y'Y

Variable	Intercept	x	y
Intercept	4	10	28
x	10	30	83
y	28	83	234

This output above says

$$X'X = \begin{vmatrix} 4 & 10 \\ 10 & 30 \end{vmatrix} \quad X'Y = \begin{vmatrix} 28 \\ 83 \end{vmatrix} \quad Y'Y = 234$$

X'X Inverse, Parameter Estimates, and SSE

Variable	Intercept	x	y
Intercept	1.5	-0.5	0.5
x	-0.5	0.2	2.6
y	0.5	2.6	4.2

The output above says

$$\text{inv}(X'X) = \begin{vmatrix} 1.5 & -0.5 \\ -0.5 & 0.2 \end{vmatrix} \quad b0 = 0.5 \quad SSE = 4.2 \quad b1 = 2.6$$

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	33.80000	33.80000	16.10	0.0569
Error	2	4.20000	2.10000		
Corrected Total	3	38.00000			

Root MSE	1.44914	R-Square	0.8895
Dependent Mean	7.00000	Adj R-Sq	0.8342
Coeff Var	20.70197		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.50000	1.77482	0.28	0.8046
x	1	2.60000	0.64807	4.01	0.0569

Covariance of Estimates

Variable	Intercept	x
Intercept	3.15	-1.05
x	-1.05	0.42

The covariance of estimates shows that the parameters are negatively correlated. The pearson correlation coefficient is

$$r = \frac{-1.05}{\sqrt{3.15(0.42)}} = -0.913$$

You can also use SAS to set up the matrices and do the matrix manipulations. These are shown below.

```
*In SAS, matrix manipulations require PROC IML.  The functions are
      t(A) : transpose the matrix A
    inv(A) : compute the inverse of matrix A
       *   : multiply the two matrices together
 A comma separates rows and a space separates columns in a matrix;

proc iml;

X ={1 1, 1 2, 1 3, 1 4};
Y ={3,5,10,10};

XP   = t(X);
XPX  = XP*X;
IXPX = inv(XPX);
B    = IXPX*XP*Y;
E    = Y-X*B;
MSE  = t(E)*E / 2;
COV  = MSE*IXPX;
```

```
print X XP;
print XPX IXPX;
print B MSE;
print COV;

run;
```

Below is the output

```
                    X                       XP
                    1         1             1         1         1         1
                    1         2             1         2         3         4
                    1         3
                    1         4

                          XPX                     IXPX
                            4         10            1.5      -0.5
                           10         30           -0.5       0.2


                                       B           MSE
                                     0.5           2.1
                                     2.6


                                          COV

                                     3.15     -1.05
                                    -1.05      0.42
```