Name: Aniruddh Kulkarni

Roll No: I081

Subject: Statistical Methods

Tutorial: Tutorial 9

Lab Exercise 9

1. Calculate the least square estimates, the residuals, and the residual sum of squares of the data

X	30	20	60	80	40	50	60	30	70	60
Y	73	50	128	170	87	108	135	~~1~~69	148	132

Find the fitted equation
Find the fitted value of y_2 and its variance
Find the variance of the second residual $\{Var(e_2)\}$

$$Y = \begin{bmatrix} 73 \\ 50 \\ 128 \\ 170 \\ 87 \\ 108 \\ 135 \\ 69 \\ 148 \\ 132 \end{bmatrix} \qquad X = \begin{bmatrix} 1 & 30 \\ 1 & 20 \\ 1 & 60 \\ 1 & 80 \\ 1 & 40 \\ 1 & 50 \\ 1 & 60 \\ 1 & 30 \\ 1 & 70 \\ 1 & 60 \end{bmatrix}$$

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 30 & 20 & 60 & 80 & 40 & 50 & 60 & 30 & 70 & 60 \end{bmatrix}$$

$$Y' = \begin{bmatrix} 73 & 50 & 128 & 170 & 87 & 108 & 135 & 69 & 148 & 132 \end{bmatrix}$$

$$X'X = \begin{bmatrix} 10 & 500 \\ 500 & 28400 \end{bmatrix}$$

$$\left[X^{-1}X\right]^{-1} = \frac{1}{34000}\begin{bmatrix} 28400 & -500 \\ -500 & 10 \end{bmatrix}$$

$$Y'Y = 134660$$

$$X'Y = \begin{bmatrix} 1100 \\ 61800 \end{bmatrix}$$

$$\hat{\beta} = (X^{-1}X)^{-1}X^{-1}Y$$

$$= \frac{1}{34000}\begin{bmatrix} 28400 & -500 \\ -500 & 10 \end{bmatrix}\begin{bmatrix} 1100 \\ 61800 \end{bmatrix}$$

$$= \frac{1}{34000}\begin{bmatrix} 340000 \\ 680000 \end{bmatrix} = \begin{bmatrix} 10 \\ 2 \end{bmatrix}$$

$$\hat{\beta} = \begin{bmatrix} 10 \\ 2 \end{bmatrix}$$

$$\beta_0 = 10$$

$$\beta_1 = 2$$

$$\hat{y} = 10 + 2x$$

$$\hat{Y} = X\hat{\beta} = \begin{bmatrix} 70 \\ 150 \\ 130 \\ 170 \\ 90 \\ 110 \\ 130 \\ 70 \\ 150 \\ 130 \end{bmatrix}$$

$$e = Y - \hat{Y}$$

$$\hat{e} = [3 \quad 0 \quad 2 \quad 0 \quad 3 \quad -2 \quad -5 \quad -1 \quad -2 \quad 2]$$

Residual sum of squares = 60

$$SSE = Y'Y - \hat{\beta}X'Y$$
$$= 134660 - (102) \begin{bmatrix} 1100 \\ 61800 \end{bmatrix}$$
$$= 134660 - 134600$$
$$= 60$$

$$\sigma^2 = \frac{60}{8} = 7.5$$

Fitted value for y_2 is is $\hat{Y}_2 = 10 + 2(20) = 50$

$$h_{22} = x_2'(X'X)^{-1}x_2$$

$$= [1 \quad 20] \times \frac{1}{34000}\begin{bmatrix} 28400 & -500 \\ -500 & 10 \end{bmatrix}\begin{bmatrix} 1 \\ 20 \end{bmatrix}$$

$$= \frac{1}{34000}[18400 \quad -300]\begin{bmatrix} 1 \\ 20 \end{bmatrix}$$

$$= 0.3647$$

$$\therefore Var(\hat{Y}_2) = \sigma^2 h_{22} = 7.5(0.3647)$$
$$= 2.73525$$

$$Var(e_2) = \sigma^2(1-h_{22})$$
$$= 7.5(1-0.3647)$$
$$= 4.764$$

Variance - covariance matrix of $\hat{\beta}$

$$cov(\hat{\beta}) = \sigma^2 (X'X)^{-1}$$

$$= \sigma^2 \times \frac{1}{34000}\begin{bmatrix} 28400 & -500 \\ -500 & 10 \end{bmatrix}$$

$$Var(\hat{\beta}_0) = 7.5 \times \frac{1}{34000} \times 28400 = 62647$$

$$Var(\hat{\beta}_1) = 7.5 \times \frac{1}{34000} \times 10 = 0.002205$$

2. Consider the following data on one predictor variable z_1 and two responses y_1 & y_2

z_1	-2	-1	0	1	2
y_1	5	3	4	2	1
y_2	-3	-1	-1	2	3

$$Z = \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \qquad Z' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \end{bmatrix}$$

$$Z'Z = \begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix}$$

$$(Z'Z)^{-1} = \frac{1}{50}\begin{bmatrix} 10 & 0 \\ 0 & 5 \end{bmatrix}$$

$$= \frac{1}{10}\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

For y_1

$$Z'y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \end{bmatrix}\begin{bmatrix} 5 \\ 3 \\ 4 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 15 \\ -9 \end{bmatrix}$$

$$\hat{\beta} = (Z^{-1}Z)^{-1} Z'y$$

$$= \frac{1}{10}\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 15 \\ -9 \end{bmatrix}$$

$$= \frac{1}{10}\begin{bmatrix} 30 \\ -9 \end{bmatrix}$$

$$= \begin{bmatrix} 3 \\ -0.9 \end{bmatrix}$$

For y_2

$$Z'y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & 1 & 0 & 1 & 2 \end{bmatrix}\begin{bmatrix} -3 \\ -1 \\ -1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 15 \end{bmatrix}$$

$$\hat{\beta}(2) = (Z'Z)^{-1} Z^{-1} y(2) = \frac{1}{10}\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 15 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 1.5 \end{bmatrix}$$

$$\hat{\beta} = [\hat{\beta}(1) \quad \hat{\beta}(2)] = \begin{bmatrix} 3 & 0 \\ 0.9 & 1.5 \end{bmatrix}$$

The fitted values

$$\hat{y}_1 = 0.9z \qquad \hat{y}_2 = 0 + 1.5z$$

$$\hat{y} = Z\hat{\beta} = \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0.9 & 1.5 \end{bmatrix}$$

$$= \frac{1}{10} \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 30 & 0 \\ -9 & 15 \end{bmatrix}$$

$$\hat{y} = Z\hat{\beta}_1 = \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ -0.9 & 1.5 \end{bmatrix}$$

$$\hat{y} = \begin{bmatrix} 4.8 & -3 \\ 3.9 & -1.5 \\ 3 & 0 \\ 2.1 & 1.5 \\ 1.2 & 3 \end{bmatrix}$$

$$\hat{e} = y - \hat{y}$$

$$= \begin{bmatrix} 5 & -3 \\ 3 & -1 \\ 4 & -1 \\ 2 & 2 \\ 1 & 3 \end{bmatrix} - \begin{bmatrix} 4.8 & -3 \\ 3.9 & -1.5 \\ 3 & 0 \\ 2.1 & 1.5 \\ 1.2 & 3 \end{bmatrix}$$

$$\hat{e} = \begin{bmatrix} 0.2 & 0 \\ -0.9 & 0.5 \\ 1 & -1 \\ -0.1 & 0.5 \\ -0.2 & 0 \end{bmatrix}$$

$$(\hat{e})' = \begin{bmatrix} 0.2 & -0.9 & 1 & -0.1 & -0.2 \\ 0 & 0.5 & -1 & 0.5 & 0 \end{bmatrix}$$

$$(\hat{e}')\hat{e} = \begin{bmatrix} 1.9 & -1.5 \\ -1.5 & 1.5 \end{bmatrix}$$

$$(\hat{y}')\hat{y} = \begin{bmatrix} 53.1 & -13.5 \\ -13.5 & 22.5 \end{bmatrix}$$

$$y'y = \begin{bmatrix} 55 & -15 \\ -15 & 24 \end{bmatrix}$$

$$\therefore y'y = \hat{y}'\hat{y} + \hat{e}'\hat{e}$$

$$\begin{bmatrix} 55 & -15 \\ -15 & 24 \end{bmatrix} = \begin{bmatrix} 53.1 & -13.5 \\ -13.5 & 22.5 \end{bmatrix} + \begin{bmatrix} 1.9 & -1.5 \\ -1.5 & 1.5 \end{bmatrix}$$

Hence proved

3. The data in the accompanying table relate grams plant dry weight y to percent soil organic matter x_1 and kilograms of supplemental soil nitrogen added per 1000 square meters X_2

Y	X_1	X_2
78.5	7	2.6
74.3	1	2.9
104.3	11	5.6

a) Define y, X, β and e for the model

b) Compute $\hat{\beta}$ and write the regression equation. Interpret each estimated regression coefficient

c) Compute $\hat{y}$ and e

d) Compute $Var(\hat{\beta})$, $Var(e)$ and $Var(\hat{y})$

$$X = \begin{bmatrix} 1 & 7 & 2.6 \\ 1 & 1 & 2.9 \\ 1 & 11 & 5.6 \end{bmatrix} \quad Y = \begin{bmatrix} 78.5 \\ 74.3 \\ 104.3 \end{bmatrix}$$

$$X' = \begin{bmatrix} 1 & 1 & 1 \\ 7 & 1 & 11 \\ 2.6 & 2.9 & 5.6 \end{bmatrix}$$

$$X'y = \begin{bmatrix} 257.1 \\ 1771.1 \\ 1003.65 \end{bmatrix} \quad y' = \begin{bmatrix} 78.5 & 74.3 & 104.3 \end{bmatrix}$$

$$(X'X) = \begin{bmatrix} 3 & 19 & 11.1 \\ 19 & 171 & 82.7 \\ 11.1 & 82.7 & 46.53 \end{bmatrix}$$

$$(x^{+1} x^{-1}) = \begin{bmatrix} 3.03 & 0.09 & -0.88 \\ 0.09 & 0.04 & -0.10 \\ -0.88 & -0.10 & 0.41 \end{bmatrix}$$

$$\hat{\beta} = (X'X)^{-1} X'y = \begin{bmatrix} 52.39 \\ 1.05 \\ 7.18 \end{bmatrix}$$

$$Y = 52.39 + 1.05x_1 + 7.18x_2$$

Both the regression coefficients β_1 and β_2 are positive as well as β_0. It shows positive correlation between each independent variable (X_1, X_2) and dependent variable y.

$$\hat{y} = X\hat{\beta}$$

$$= \begin{bmatrix} 1 & 7 & 2.6 \\ 1 & 1 & 2.9 \\ 1 & 11 & 2.6 \end{bmatrix} \begin{bmatrix} 52.39 \\ 1.05 \\ 7.18 \end{bmatrix}$$

$$= \begin{bmatrix} 78.408 \\ 74.262 \\ 104.148 \end{bmatrix}$$

$$\hat{e} = y - \hat{y} = \begin{bmatrix} 78.5 \\ 74.3 \\ 104.3 \end{bmatrix} - \begin{bmatrix} 78.408 \\ 74.262 \\ 104.148 \end{bmatrix}$$

$$\hat{e} = \begin{bmatrix} 0.092 \\ 0.038 \\ 0.152 \end{bmatrix}$$

$$SSE = y'y - (\hat{\beta})'X'y$$

$$= 22561 - \begin{bmatrix} 52.39 & 1.05 & 7.18 \end{bmatrix} \begin{bmatrix} 257.1 \\ 1771.1 \\ 1003.65 \end{bmatrix}$$

$$SSE = 22561 - 22533.331$$
$$= 25.669$$

$$\sigma^2 = \frac{SSE}{n-k-1}$$
$$= \frac{25.669}{3-2-1}$$

$$cov(\hat{\beta}) = \sigma^2 (X'X)^{-1} = \sigma^2 \begin{bmatrix} 3.03 & 0.09 & -0.88 \\ 0.09 & 0.04 & -0.10 \\ -0.88 & -0.10 & 0.41 \end{bmatrix}$$
= Not possible

$Var(\hat{\beta}_0) = \sigma^2 (3.03)$ = Not possible

$Var(\hat{\beta}_1) = \sigma^2 (0.04)$ = Not possible

$Var(\hat{\beta}_2) = \sigma^2 (0.41)$ = Not possible

$Var(e) = \sigma^2 (I - P)$ = Not possible

$$P = X(X'X)^{-1}X'$$

$$= \begin{bmatrix} 1 & 7 & 2.6 \\ 1 & 1 & 2.9 \\ 1 & 11 & 5.6 \end{bmatrix} \begin{bmatrix} 3.03 & 0.09 & -0.88 \\ 0.09 & 0.04 & -0.1 \\ -0.88 & -0.10 & 0.41 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 7 & 1 & 11 \\ 2.6 & 2.9 & 5.6 \end{bmatrix}$$

$$P = \begin{bmatrix} 2.162 & 0.6331 & -1.423 \\ 0.6466 & 0.3821 & -0.46 \\ -1.412 & -0.43 & 0.9353 \end{bmatrix}$$

$$I - P = \begin{bmatrix} -1.162 & -0.633 & 1.4237 \\ -0.646 & 0.6178 & 0.4608 \\ 1.425 & 1.4307 & -0.935 \end{bmatrix}$$

$$\text{Var}(\hat{y}) = Py$$

$$= \begin{bmatrix} 26.002 \\ 31.17 \\ -45.23 \end{bmatrix}$$