

(1) Consider a multiple regression based on thirty observations of Y on five regressor variables. The model is: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \varepsilon_i$. The sample size n=22. The regressor variables are characteristics of the beak or mouth of the squid.

Y: weight of squids that were eaten by sharks and tuna.

X1: rostral length in inches

X2: wing length in inches

X3: rostral to notch length

X4: notch to wing length

X5: width in inches.

The sequential sums of squares due to regression are as follows:

Sequential Sum of Squares Regression

$$R(\beta_1 | \beta_0) = 199.145$$

$$R(\beta_2 | \beta_1, \beta_0) = 0.126664$$

$$R(\beta_3 | \beta_2, \beta_1, \beta_0) = 4.119539$$

$$R(\beta_4 | \beta_3, \beta_2, \beta_1, \beta_0) = 0.263496$$

$$R(\beta_5 | \beta_0, \beta_1, \beta_2, \beta_3, \beta_4) = 4.352193$$

Partial Sum of Squares

$$R(\beta_5 | \beta_0, \beta_1, \beta_2, \beta_3, \beta_4) = 4.352193$$

$$R(\beta_4 | \beta_0, \beta_1, \beta_2, \beta_3, \beta_5) = 0.982690$$

$$R(\beta_3 | \beta_0, \beta_1, \beta_2, \beta_4, \beta_5) = 0.078273$$

$$R(\beta_2 | \beta_0, \beta_1, \beta_3, \beta_4, \beta_5) = 0.868761$$

$$R(\beta_1 | \beta_0, \beta_2, \beta_3, \beta_4, \beta_5) = 0.298731$$

a. If SS Total=215.925, give the MSE.

b. Using the appropriate SS, test at significance level 0.05 the following: $H_0 : \begin{bmatrix} \beta_1 \\ \beta_4 \end{bmatrix} = \underline{0}$. Comment on your result. Give the p-value.

c. Using the appropriate SS, test at significance level 0.05 the following: $H_0 : \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \underline{0}$. Comment

on your result. Give the p-value.

d. Is the variable “rostral length” important in the regression model in the presence of all of the other variables? Is it important as a single regressor? Discuss.

(2) In a study on students taking a statistics course, three variables are measured: Y (how many statistics books were used), X1 (attendance in the statistics course) and X2 (final grade in the statistics course). There were forty students in this study.

For the Full Model:

$$X'X = \begin{bmatrix} 40 & 564 & 2542 \\ 564 & 8666 & 37186 \\ 2542 & 37186 & 172428 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 80 \\ 1234 \\ 5543 \end{bmatrix} \text{ Note that } \sum_{i=1}^n Y_i = 80 \text{ and that } Y'Y = 240, \text{ MSE} = 2.5.$$

- (a) Give the least squares regression line. Comment on the meaning of the coefficients of the regression line, referring to Y (how many statistics books were used), X1 (attendance in the statistics course) and X2 (final grade in the statistics course).
- (b) Estimate the regression variance for the full model. Then give the estimated variances of $\hat{\beta}_1$ and of $\hat{\beta}_2$. Comment on the results.

[blank page]

(4) Consider a simple linear regression of potato yield (tons/acre) on the amount of fertilizer used (cwt/acre).

<u>Fertilizer</u>	<u>Yield</u>
0	8.4
0	8.0
2	12.0
2	12.8
5	44.2
7	100.6
<u>10</u>	<u>400.0</u>

- (a) Create a scatter plot of the data. Does the plot indicate a linear association between potato yield and amount of fertilizer used?