Stat 6021: Homework Set 3 Solutions

2. (a) The table is displayed below.

x_i	70	75	80	80	85	90
y_i	75	82	80	86	90	91
$\hat{y}_i$	76	80	84	84	88	92
e_i	-1	2	-4	2	2	-1

(b)

	DF	SS	MS	F-stat	p-value
Regression	1	160	160	21.3333	0.0099
Residual	4	30	7.5	***	***
Total	5	190	***	***	***

$SS_{res} = \sum_i e_i^2 = (-1)^2 + 2^2 + (-4)^2 + 2^2 + 2^2 + (-1)^2 = 30.$

$$\begin{aligned} SS_T &= \sum_i (y_i - \bar{y})^2 \\ &= (75-84)^2 + (82-84)^2 + (80-84)^2 + (86-84)^2 + (90-84)^2 + (91-84)^2 \\ &= 190. \end{aligned}$$

(c) $\hat{\sigma^2} = \frac{SS_{res}}{n-2} = \frac{30}{4} = 7.5.$

(d) $R^2 = \frac{SS_{res}}{SS_T} = \frac{160}{190} = 0.842.$

(e) $H_0 : \beta_1 = 0, H_a : \beta_1 \neq 0.$ Since the p-value is less than 0.05, we reject the null hypothesis. Our data supports the claim that there is a linear relationship between scores on the second quiz and scores on the first quiz.

Critical value approach: Critical value is 7.71 (using qf(0.95,1,4) in R). Since the F-stat is greater than the critical value, we reject the null hypothesis. Our data supports the claim that there is a linear relationship between scores on the second quiz and scores on the first quiz.

3. We want to find $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize

$$Q = SS_{res} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2.$$

Taking partial derivatives of Q with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$, we obtain

$$\frac{\partial Q}{\partial \hat{\beta}_0} = \sum_i 2\left(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)\right)(-1) \tag{1}$$

and

$$\frac{\partial Q}{\partial \hat{\beta}_1} = \sum_i 2\left(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)\right)(-x_i). \tag{2}$$

To find the stationary points, set (1) and (2) equal 0. Thus we obtain

$$\sum_i y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_i x_i \tag{3}$$

and

$$\sum_i x_i y_i = \hat{\beta}_0 \sum_i x_i + \hat{\beta}_1 \sum_i x_i^2. \tag{4}$$

From (3),

$$\begin{aligned} \hat{\beta}_0 &= \frac{\sum_i y_i}{n} - \hat{\beta}_1 \frac{\sum_i x_i}{n} \\ &= \bar{y} - \hat{\beta}_1 \bar{x}. \end{aligned} \tag{5}$$

Sub (5) into (4),

$$\begin{aligned} \sum_i x_i y_i &= (\bar{y} - \hat{\beta}_1 \bar{x}) \sum_i x_i + \hat{\beta}_1 \sum_i x_i^2 \\ \Longrightarrow \hat{\beta}_1 \left(\sum_i x_i^2 - \bar{x} \sum_i x_i \right) &= \sum_i x_i y_i - \bar{y} \sum_i x_i \\ \Longrightarrow \hat{\beta}_1 \left(\sum_i x_i^2 - n\bar{x}^2 \right) &= \sum_i x_i y_i - n\bar{x}\bar{y} \\ \Longrightarrow \hat{\beta}_1 &= \frac{\sum_i x_i y_i - n\bar{x}\bar{y}}{\sum_i x_i^2 - n\bar{x}^2} \\ &= \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}. \end{aligned} \tag{6}$$

Equations (5) and (6) are the least squares estimators.