MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.430J/15.064J Summer 1997 AGH
Problem Set 3 Solutions — VIII-1a,b,c,d, 2a,b,e, 9a, 11 07/17/97

8-1

The data points are: $(1,3)$, $(2,5)$ and $(3,10)$
$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 2$
$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = 6$

(a) $\widehat{\beta} = \frac{\sum_{i=1}^{n}(x_i-\overline{x})y_i}{\sum_{i=1}^{n}(x_i-\overline{x})^2} = \frac{(1-2)\times 3+(2-2)\times 5+(3-2)\times 10}{(1-2)^2+(2-2)^2+(3-2)^2} = \frac{7}{2}$

$\widehat{\alpha} = \overline{y} - \widehat{\beta}\overline{x} = 6 - \frac{7}{2}\times 2 = -1$

(b) $r_i = y_i - \left(\widehat{\alpha} + \widehat{\beta}x_i\right)$

$r_1 = 3 - \left(-1 + \frac{7}{2}\times 1\right) = 0.5$

$r_2 = 5 - \left(-1 + \frac{7}{2}\times 2\right) = -1.0$

$r_3 = 10 - \left(-1 + \frac{7}{2}\times 3\right) = 0.5$

(c) $V^2(e) = \frac{\sum_{i=1}^{n} r_i^2}{(n-2)} = \frac{0.5^2+(-1)^2+0.5^2}{1} = 1.5$

(d) The mean is zero because $\widehat{\beta}$ is an unbiased estimator of β. The variance on the other hand is,

$V^2(W) = \frac{\sum_{i=1}^{n} r_i^2}{(n-2)\sum_{i=1}^{n}(x_i-\overline{x})^2} = \frac{0.5^2+(-1)^2+0.5^2}{1\times\left((1-2)^2+(2-2)^2+(3-2)^2\right)} = 0.75$

8-2

The data points are: $(40,27)$, $(60,23)$, $(80,22)$ and $(120,16)$
$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 75$
$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = 22$

(a) $\widehat{\beta} = \frac{\sum_{i=1}^{n}(x_i-\overline{x})y_i}{\sum_{i=1}^{n}(x_i-\overline{x})^2} = \frac{(40-75)\times 27+(60-75)\times 23+(80-75)\times 22+(120-75)\times 16}{(40-75)^2+(60-75)^2+(80-75)^2+(120-75)^2} = -\frac{23}{175} \approx -0.131$

$\widehat{\alpha} = \overline{y} - \widehat{\beta}\overline{x} = 22 - \left(-\frac{23}{175}\right)\times 75 = \frac{223}{7} \approx 31.9$

(b) $r_1 = 27 - \left(\frac{223}{7} - \frac{23}{175}\times 40\right) = \frac{2}{5} = 0.4$

$r_2 = 23 - \left(\frac{223}{7} - \frac{23}{175}\times 60\right) = -\frac{34}{35} = -0.97143$

$r_3 = 22 - \left(\frac{223}{7} - \frac{23}{175}\times 80\right) = \frac{23}{35} = 0.65714$

$r_4 = 16 - \left(\frac{223}{7} - \frac{23}{175}\times 120\right) = -\frac{3}{35} = -0.085714$

$\sum_{i=1}^{n} r_i^2 = \left(\frac{2}{5}\right)^2 + \left(-\frac{34}{35}\right)^2 + \left(\frac{23}{35}\right)^2 + \left(-\frac{3}{35}\right)^2 = \frac{54}{35} = 1.5429$

$\sum_{i=1}^{n} (y_i - \overline{y})^2 = (27-22)^2 + (23-22)^2 + (22-22)^2 + (16-22)^2 = 62$

$R^2 = 1 - \frac{\sum_{i=1}^{n} r_i^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} = 1 - \frac{\frac{54}{35}}{62} = \frac{1058}{1085} \approx 0.975$

(e) The parameter estimate $\widehat{\beta} \approx -0.131$ implies that if we raise the price one cent, we expect to get 0.131 thousand fewer calls.

8-9

We are given:

$$\begin{array}{rcrcrcl} & & & & & n & = & 20 \\ y & = & 116.2 & + & 1.24x & R^2 & = & 0.688 \\ & & (23.37) & & (2.72) & V^2(\epsilon) & = & 15.35 \end{array}$$

(a) We want to find the probability that the fuel consumption actually drops as the temperature goes up. To put it another way, we want $P(\beta \leq 0)$. We are given that $t_\beta = 2.72$. So, $P(\beta \leq 0) = 0.5 - n(2.72) = 0.0033$.

8-11

R^2 is the proportional reduction in the SSR (sum of squared residuals) when the "best" horizontal line is replaced by the "best" unrestricted line.

(a) (i) Given: $R^2 = 0$ and $V^2(\epsilon) = 0$.
$R^2 = 0$ implies that the opportunity to tilt the line is of no value whatsoever from a least-squares perspective. $V^2(\epsilon) = 0$ implies the errors in the "best" unrestricted line is also zero. This means that the 3 points lie on a straight line. Hence, any 3 points on a horizontal line will satisfy the 2 conditions. For example,

$$(1,0), \quad (2,0), \quad (3,0)$$

(ii) Given: $R^2 = 0$ and $V^2(\epsilon) > 0$.
Again, tilting the line is of no value. $V^2(\epsilon) > 0$ implies that the 3 points are not on a straight line. Thus any 3 points forming an equilateral triangle with a horizontal base will satisfy the 2 conditions. For example,

$$(0,0), \quad (1,1), \quad (2,0)$$

(iii) Given: $R^2 = 1$ and $\hat{\beta} = 1$.
$R^2 = 1$ means that the tilted line gives a perfect fit. Furthermore, since $\hat{\beta} = 1$, the 3 points lie on a line of slope 1. For example,

$$(0,0), \quad (1,1), \quad (2,2)$$

(b) In (iv), we are given: $R^2 = 1$ and $V^2(\epsilon) > 0$.
$R^2 = 1$ means that the "best" tilted line gives a perfect fit. $V^2(\epsilon) > 0$ means that the "best" tilted line does have some errors. But if you have some errors, you do not have perfect fit. Therefore no data can be found to satisfy (iv).

Árni