MAT-042: Taller 2

Felipe Osorio

http://fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Ejercicio 1.a)

Desde el conjunto de datos $(x_1, y_1), (x_2, y_2), \dots, (x_6, y_6)$, podemos obtener:

$$\sum_{i=1}^{n} x_i = 0, \quad \sum_{i=1}^{n} x_i^2 = 130, \quad \sum_{i=1}^{n} y_i = 0.41, \quad \sum_{i=1}^{n} y_i^2 = 10.5009, \quad \sum_{i=1}^{n} x_i y_i = -10.59,$$

con $n = 6$. De este modo,

$$\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^{n} x_i y_i - n\,\overline{x}\,\overline{y} = -10.59,$$

$$\sum_{i=1}^{n} (x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - n\,\overline{x}^2 = 130.$$

Ejercicio 1.a)

Luego, tenemos que la estimación de los coeficientes de regresión es dada por:

$$\widehat{\beta} = \frac{\operatorname{cov}(\boldsymbol{x}, \boldsymbol{y})}{\operatorname{var}(\boldsymbol{x})} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = -\frac{10.59}{130} = -0.08146,$$

$$\widehat{\alpha} = \overline{y} - \widehat{\beta}\,\overline{x} = 0.06833.$$

El vector de valores predichos está dado por:

$$\widehat{\boldsymbol{y}} = (0.3942, 0.31272, 0.2313, 0.1498, 0.0683, -0.7463)^\top,$$

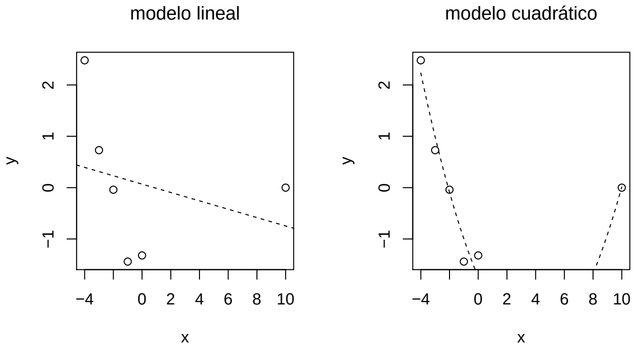
luego

$$R^2 = \{\operatorname{cor}(\boldsymbol{y}, \widehat{\boldsymbol{y}})\}^2 = \frac{\{\operatorname{cov}(\boldsymbol{y}, \widehat{\boldsymbol{y}})\}^2}{\operatorname{var}(\boldsymbol{y})\operatorname{var}(\widehat{\boldsymbol{y}})} = 0.08237.$$

De este modo, un 8.24% de la variabilidad de la respuesta es explicada por el modelo.

Ejercicio 1.a)

Comparando el valor de R^2 con el gráfico de los datos y el modelo ajustado se aprecia que el ajuste es pobre. Un ajuste bastante mejor puede ser obtenido considerando, por ejemplo, un modelo cuadrático.



Ejercicio 2.a)

Sabemos que $\widehat{\alpha} = \overline{y} - \widehat{\beta}\,\overline{x}$, de este modo

$$e_i = y_i - \widehat{\alpha} - \widehat{\beta}\,x_i = y_i - \overline{y} - \widehat{\beta}(x_i - \overline{x}).$$

En efecto,

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} \big(y_i - \overline{y} - \widehat{\beta}(x_i - \overline{x})\big) = \sum_{i=1}^{n}(y_i - \overline{y}) - \widehat{\beta}\sum_{i=1}^{n}(x_i - \overline{x}) = 0.$$

Ejercicio 2.a)

Sabemos que $\widehat{\alpha} = \overline{y} - \widehat{\beta}\,\overline{x}$, de este modo

$$e_i = y_i - \widehat{\alpha} - \widehat{\beta}\,x_i = y_i - \overline{y} - \widehat{\beta}(x_i - \overline{x}).$$

En efecto,

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} \big(y_i - \overline{y} - \widehat{\beta}(x_i - \overline{x})\big) = \sum_{i=1}^{n}(y_i - \overline{y}) - \widehat{\beta}\sum_{i=1}^{n}(x_i - \overline{x}) = 0.$$

Ejercicio 2.b)

Por otro lado,

$$
\begin{aligned}
\sum_{i=1}^{n} e_i\,\widehat{y}_i &= \sum_{i=1}^{n} (y_i - \overline{y} - \widehat{\beta}(x_i - \overline{x}))(\overline{y} + \widehat{\beta}(x_i - \overline{x})) \\
&= \overline{y} \sum_{i=1}^{n} (y_i - \overline{y} - \widehat{\beta}(x_i - \overline{x})) + \widehat{\beta} \sum_{i=1}^{n} (y_i - \overline{y} - \widehat{\beta}(x_i - \overline{x}))(x_i - \overline{x}),
\end{aligned}
$$

el primer término es cero pues, $\sum_{i=1}^{n} e_i = 0$. Es decir, tenemos que:

$$
\sum_{i=1}^{n} e_i\,\widehat{y}_i = \widehat{\beta}\Big[\sum_{i=1}^{n} (y_i - \overline{y})(x_i - \overline{x}) - \widehat{\beta} \sum_{i=1}^{n} (x_i - \overline{x})^2\Big].
$$

Notando que $\widehat{\beta} = \operatorname{cov}(\boldsymbol{x}, \boldsymbol{y}) / \operatorname{var}(\boldsymbol{x})$, se verifica el resultado.

Ejercicio 2.c)

Para notar la parte (c), considere que

$$\begin{aligned}(y_i - \overline{y})^2 &= \{(y_i - \widehat{y}_i) + (\widehat{y}_i - \overline{y})\}^2 \\ &= (y_i - \widehat{y}_i)^2 + 2(y_i - \widehat{y}_i)(\widehat{y}_i - \overline{y}) + (\widehat{y}_i - \overline{y})^2.\end{aligned}$$

Sumando, tenemos

$$\sum_{i=1}^n (y_i - \overline{y})^2 = \sum_{i=1}^n (y_i - \widehat{y}_i)^2 + \sum_{i=1}^n (\widehat{y}_i - \overline{y})^2 + 2\sum_{i=1}^n (y_i - \widehat{y}_i)(\widehat{y}_i - \overline{y}).$$

Recuerde que $e_i = y_i - \widehat{y}_i$, para $i = 1, \dots, n$. De este modo, el último término es

$$\sum_{i=1}^n e_i(\widehat{y}_i - \overline{y}) = \sum_{i=1}^n e_i\widehat{y}_i - \overline{y}\sum_{i=1}^n e_i = 0,$$

por partes (a) y (b), y la verificación es completa.