MAT-032: Lista de ejercicios 1

Felipe Osorio

http://fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Ejercicio 1 a)

Breve descripción:

Las plantas sometidas al tratamiento Cross tienen una altura promedio de 161.5 unid., su distribución presenta asimetría negativa (hacia la izquierda) con curtosis positiva indicando que la distribución es aguda (leptocúrtica). Se puede apreciar además que existe una única observación atípica. Comparativamente con las plantas sometidas al tratamiento Self la distribución tiene un nivel promedio y variabilidad mayores.

EX UMBRA IN SOLEM

Ejercicio 1 b)

Note que

$$\overline{z} = \frac{1}{n}\sum_{i=1}^{n} z_i = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i) = \frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{n}\sum_{i=1}^{n} y_i = \overline{x} - \overline{y}$$
$$= 161.5 - 140.6 = 20.9.$$

En el enunciado se nos indica que:

$$\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = 0 \qquad \Rightarrow \qquad \mathsf{cov}(\boldsymbol{x}, \boldsymbol{y}) = 0$$

esto lleva a[1]

$$s_z^2 = \mathsf{var}(\boldsymbol{x}) + \mathsf{var}(\boldsymbol{y}) = 28.94^2 + 16.41^2 = 1106.812,$$

y

$$CV_z = \frac{s_z}{\overline{z}} = \frac{\sqrt{1106.812}}{20.9} = \frac{33.2688}{20.9} = 1.5918.$$

[1] Propiedad (b) en página 7 de las slides 4ta sesión de clases.

Ejercicio 1 c)

Tenemos la transformación

$$u_i = -1.76 y_i + 408.96, \qquad i = 1, \ldots, n.$$

Sabemos que[2]

$$\overline{u} = -1.76\overline{y} + 408.96 = 161.504,$$

y

$$s_u = |-1.76| s_y = 1.76 \cdot 16.41 = 28.882.$$

[2] ver Propiedad (d) en página 38 de las slides 2da sesión de clases.

Ejercicio 2

Puede ser notado que

$$(x_i - x_j)^2 = ((x_i - \overline{x}) - (x_j - \overline{x}))^2 = (x_i - \overline{x})^2 + (x_j - \overline{x})^2 - 2(x_i - \overline{x})(x_j - \overline{x}),$$

de este modo,

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2 &= \sum_{i=1}^{n}\sum_{j=1}^{n}\{(x_i - \overline{x})^2 + (x_j - \overline{x})^2 - 2(x_i - \overline{x})(x_j - \overline{x})\} \\ &= n\sum_{i=1}^{n}(x_i - \overline{x})^2 + n\sum_{j=1}^{n}(x_j - \overline{x})^2 - 2\sum_{i=1}^{n}(x_i - \overline{x})\sum_{j=1}^{n}(x_j - \overline{x}).\end{aligned}$$

Sabemos que $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$ (análogamente para $\sum_{j=1}^{n}(x_j - \overline{x}) = 0$), luego

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

Finalmente, tenemos

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{2n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2.$$

EX UMBRA IN SOLEM

Ejercicio 2

Puede ser notado que

$$(x_i - x_j)^2 = ((x_i - \overline{x}) - (x_j - \overline{x}))^2 = (x_i - \overline{x})^2 + (x_j - \overline{x})^2 - 2(x_i - \overline{x})(x_j - \overline{x}),$$

de este modo,

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2 &= \sum_{i=1}^{n}\sum_{j=1}^{n}\{(x_i - \overline{x})^2 + (x_j - \overline{x})^2 - 2(x_i - \overline{x})(x_j - \overline{x})\} \\ &= n\sum_{i=1}^{n}(x_i - \overline{x})^2 + n\sum_{j=1}^{n}(x_j - \overline{x})^2 - 2\sum_{i=1}^{n}(x_i - \overline{x})\sum_{j=1}^{n}(x_j - \overline{x}).\end{aligned}$$

Sabemos que $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$ (análogamente para $\sum_{j=1}^{n}(x_j - \overline{x}) = 0$), luego

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

Finalmente, tenemos

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{2n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2.$$

Ejercicio 3

Breve descripción:

Los números sorteados en la lotería de Maryland tienen un promedio de 518.96 y una desviación estándar de 291.7 unidades. Desde el boxplot y el coeficiente de asimetría

$$b_1 = \frac{1}{218}\sum_{i=1}^{218}\left(\frac{x_i - \overline{x}}{s}\right)^3 = \frac{1}{218}\sum_{i=1}^{218} z_i^3 = -\frac{20.07}{218} = -0.0921,$$

se aprecia que los datos son simétricos y que su curtosis es negativa

$$\begin{aligned} b_2 &= \left\{\frac{1}{218}\sum_{i=1}^{218}\left(\frac{x_i - \overline{x}}{s}\right)^4\right\} - 3 = \frac{1}{218}\sum_{i=1}^{218} z_i^4 - 3 = \frac{390.38}{218} - 3 \\ &= 1.7907 - 3 = -1.2093, \end{aligned}$$

indicando que la distribución es achatada (platicúrtica). Además, la información provista por el histograma indica que los datos siguen un comportamiento bastante uniforme.

Ejercicio 4

Desde el conjunto de datos $(x_1, y_1), \dots, (x_6, y_6)$, 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,

$$\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^n x_i y_i - n\,\overline{x}\,\overline{y} = -10.59,$$
$$\sum_{i=1}^n (x_i - \overline{x})^2 = \sum_{i=1}^n x_i^2 - n\,\overline{x}^2 = 130.$$

Luego, tenemos que la estimación de los coeficientes de regresión[3] es dada por:

$$\widehat{\beta} = \frac{\operatorname{cov}(\boldsymbol{x}, \boldsymbol{y})}{\operatorname{var}(\boldsymbol{x})} = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2} = -\frac{10.59}{130} = -0.0815,$$
$$\widehat{\alpha} = \overline{y} - \widehat{\beta}\,\overline{x} = \frac{0.41}{6} = 0.0683.$$

[3] ver página 23 de las slides 4ta sesión de clases.

Ejercicio 4

El vector de valores predichos $\widehat{\boldsymbol{y}} = (\widehat{y}_1, \widehat{y}_2, \dots, \widehat{y}_6)^\top$, con $\widehat{y}_i = \widehat{\alpha} + \widehat{\beta} x_i$, está dado por:

$$\widehat{\boldsymbol{y}} = (0.3942, 0.3127, 0.2313, 0.1498, 0.0683, -0.7463)^\top.$$

Note que

$$\overline{\widehat{y}} = \frac{1}{n}\sum_{i=1}^{n} \widehat{y}_i = \frac{1}{n}\sum_{i=1}^{n}(\widehat{\alpha} + \widehat{\beta} x_i) = \widehat{\alpha} + \widehat{\beta}\overline{x} = \overline{y} - \widehat{\beta}\overline{x} + \widehat{\beta}\overline{x} = \overline{y}$$

De este modo,

$$\sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n} y_i^2 - n\overline{y}^2 = 10.5009 - 6\Big(\frac{0.41}{6}\Big)^2 = 10.4729$$

$$\sum_{i=1}^{n}(\widehat{y}_i - \overline{y})^2 = \sum_{i=1}^{n} \widehat{y}_i^2 - n\overline{y}^2 = 0.8907 - 6\Big(\frac{0.41}{6}\Big)^2 = 0.8627,$$

y

$$\sum_{i=1}^{n}(\widehat{y}_i - \overline{y})(y_i - \overline{y}) = \sum_{i=1}^{n} y_i\widehat{y}_i - n\overline{y}^2 = 0.8907 - 6\Big(\frac{0.41}{6}\Big)^2 = 0.8627.$$

Ejercicio 4

Tenemos,

$$\begin{aligned} R^2 = \{\mathsf{cor}(\boldsymbol{y}, \widehat{\boldsymbol{y}})\}^2 &= \left\{ \frac{\mathsf{cov}(\boldsymbol{y}, \widehat{\boldsymbol{y}})}{\sqrt{\mathsf{var}(\boldsymbol{y})\,\mathsf{var}(\widehat{\boldsymbol{y}})}} \right\}^2 \\ &= \frac{\{\sum_{i=1}^n (\widehat{y}_i - \overline{y})(y_i - \overline{y})\}^2}{\sum_{i=1}^n (y_i - \overline{y})^2 \sum_{i=1}^n (\widehat{y}_i - \overline{y})^2} \\ &= \frac{0.8627^2}{10.4729 \cdot 0.8627} = \frac{0.8627}{10.4729} = 0.0824. \end{aligned}$$

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

Ejercicio 4

Observación:

Comparando el valor de R^2 con el modelo ajustado se aprecia que un ajuste pobre. Un mejor modelo puede ser obtenido considerando, por ejemplo, una regresión cuadrática.