MAT-042: Taller 4

Felipe Osorio

http://fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Ejercicio 1.a)

Para el conjunto de datos $\boldsymbol{x} = \{x_1, x_2, \dots, x_{10}\}$. Tenemos,

$$\sum_{i=1}^{10} x_i = 140, \qquad \sum_{i=1}^{10} x_i^2 = 5230.$$

De este modo, $\overline{x} = 140/10 = 14$. Mientras que,

$$\begin{aligned} s^2 &= \frac{1}{10-1}\sum_{i=1}^{10}(x_i - \overline{x})^2 = \frac{1}{9}\Big(\sum_{i=1}^{n} x_i^2 - 10 \cdot \overline{x}^2\Big) = \frac{1}{9}\big(5230 - 10 \cdot 14^2\big) \\ &= \frac{1}{9}(5230 - 1960) = \frac{3270}{9} = \frac{1090}{3} = 363.3333. \end{aligned}$$

Además,

$$\text{CV} = \frac{s}{\overline{x}} = \frac{\sqrt{1090/3}}{14} = \frac{19.0613}{14} = 1.3615.$$

Ejercicio 1.b)

Tenemos que los valores ordenados, $x_{(1)} < x_{(2)} < \cdots < x_{(10)}$ son dados por:

$$2,\ 3,\ 5,\ 7,\ 8, 10, 11, 12, 15, 67.$$

Como $n = 10$, sigue que

$$\text{me} = \frac{8+10}{2} = 9.$$

Para calcular Q_1 y Q_3 considere los nuevos conjuntos de datos ordenados

$$\boldsymbol{D}_1 = \{2, 3, 5, 7, 8\}, \quad \text{y} \quad \boldsymbol{D}_2 = \{10, 11, 12, 15, 67\}.$$

De ahí que $Q_1 = 5$ y $Q_3 = 12$, y $IQR = Q_3 - Q_1 = 12 - 5 = 7$. Esto permite obtener

$$b_{\mathsf{G}} = \frac{(Q_3 - \text{me}) - (\text{me} - Q_1)}{IQR} = \frac{(12-9)-(9-5)}{7} = \frac{3-4}{7} = -\frac{1}{7} = -0.1429.$$

EX UMBRA IN SOLEM

Ejercicio 1.c)

Sabemos que

$$\overline{y} = -1.3\,\overline{x} + 7 = -1.3 \cdot 14 + 7 = -11.2$$
$$\mathsf{var}(\boldsymbol{y}) = (-1.3)^2\,\mathsf{var}(\boldsymbol{x}) = 1.69 \cdot 363.3333 = 614.0333.$$

De este modo,

$$\mathrm{CV}_y = \frac{\sqrt{\mathsf{var}(\boldsymbol{y})}}{|\overline{y}|} = \frac{\sqrt{614.0333}}{11.2} = \frac{24.7797}{11.2} = 2.2125.$$

Ejercicio 1.d)

Note que

$$y_i = ax_i + b, \qquad i = 1, 2, \dots, n,$$

con $a = -1.3$ y $b = 7$. Esto nos permite escribir

$$\begin{aligned}\mathsf{cov}(\boldsymbol{x}, \boldsymbol{y}) &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(ax_i + b - a\overline{x} - b) \\ &= a \cdot \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = a\,\mathsf{var}(\boldsymbol{x}).\end{aligned}$$

Como $\mathsf{var}(\boldsymbol{y}) = a^2\,\mathsf{var}(\boldsymbol{x})$, sigue que

$$\mathsf{corr}(\boldsymbol{x}, \boldsymbol{y}) = \frac{\mathsf{cov}(\boldsymbol{x}, \boldsymbol{y})}{\sqrt{\mathsf{var}(\boldsymbol{x})\,\mathsf{var}(\boldsymbol{y})}} = \frac{a\,\mathsf{var}(\boldsymbol{x})}{\sqrt{a^2\,\mathsf{var}^2(\boldsymbol{x})}},$$

y como en nuestro caso particular $a < 0$, sigue que

$$\mathsf{corr}(\boldsymbol{x}, \boldsymbol{y}) = \frac{a}{\sqrt{a^2}} = \frac{a}{|a|} = -1.$$

Ejercicio 2)

Desarrollando el cuadrado de binomio y sumando, obtenemos

$$\sum_{i=1}^{k} n_i(x_i - \overline{x})^2 = \sum_{i=1}^{k} n_i(x_i^2 - 2x_i\overline{x} + \overline{x}^2) = \sum_{i=1}^{k} n_i x_i^2 - 2\overline{x}\sum_{i=1}^{k} n_i x_i + \overline{x}^2 \sum_{i=1}^{k} n_i$$
$$= \sum_{i=1}^{k} n_i x_i^2 - 2n\overline{x}^2 + n\overline{x}^2 = \sum_{i=1}^{k} n_i x_i^2 - n\overline{x}^2,$$

lo que verifica el resultado.

Ejercicio 3)

Tenemos $n = 6$, y

$$\sum_{i=1}^{n} x_i = 21, \qquad \sum_{i=1}^{n} x_i^2 = 91, \qquad \sum_{i=1}^{n} x_i y_i = 280$$
$$\sum_{i=1}^{n} y_i = 65, \qquad \sum_{i=1}^{n} y_i^2 = 879.$$

Es decir,

$$\sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2 = 91 - 21^2/6 = 17.5000$$
$$\sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n} y_i^2 - n\overline{y}^2 = 879 - 65^2/6 = 174.8333$$
$$\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^{n} x_i y_i - n\overline{x}\,\overline{y} = 280 - 21 \cdot 65/6 = 52.5000.$$

Ejercicio 3)

De este modo,

$$\widehat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{52.5}{17.5} = 3.0,$$

y portanto

$$\widehat{\alpha} = \overline{y} - \widehat{\beta}\,\overline{x} = \frac{1}{6}(65 - 3 \cdot 21) = \frac{1}{3}.$$

Además, tenemos que los residuos, $e_i = y_i - \widehat{\alpha} - \widehat{\beta}x_i$, para $i = 1, \dots, n$, son dados por:

$$\begin{aligned}
\boldsymbol{e} &= \{5 - \tfrac{1}{3} - 3, 7 - \tfrac{1}{3} - 6, 7 - \tfrac{1}{3} - 9, 10 - \tfrac{1}{3} - 12, 16 - \tfrac{1}{3} - 15, 20 - \tfrac{1}{3} - 18\} \\
&= \{2 - \tfrac{1}{3}, 1 - \tfrac{1}{3}, -2 - \tfrac{1}{3}, -2 - \tfrac{1}{3}, 1 - \tfrac{1}{3}, 2 - \tfrac{1}{3}\} \\
&= \{(6-1)/3, (3-1)/3, (-6-1)/3, (-6-1)/3, (3-1)/3, (6-1)/3\} \\
&= \{5/3, 2/3, -7/3, -7/3, 2/3, 5/3\},
\end{aligned}$$

Ejercicio 3)

Tenemos que,

$$RSS = \sum_{i=1}^{n} e_i^2 = \frac{1}{3^2}\left(5^2 + 2^2 + (-7)^2 + (-7)^2 + 2^2 + 5^2\right) = \frac{156}{9} = 17.3333.$$

Es decir, $s^2 = RSS/(n-2) = 17.3333/4 = 4.3333$. Finalmente

$$R^2 = 1 - \frac{RSS}{\sum_{i=1}^{n}(y_i - \overline{y})^2} = 1 - \frac{17.3333}{174.3333} = 0.9009.$$

Ejercicio 4.a)

Tenemos que

$$\frac{\mathrm{d}}{\mathrm{d}\,\theta}S_1(\theta) = \sum_{i=1}^n \frac{\mathrm{d}}{\mathrm{d}\,\theta}(y_i - \theta x_i)^2 = 2\sum_{i=1}^n (y_i - \theta x_i)\frac{\mathrm{d}}{\mathrm{d}\,\theta}(y_i - \theta x_i)$$
$$= -2\sum_{i=1}^n (y_i - \theta x_i)x_i.$$

Resolviendo la condición $\mathrm{d}\,S_1(\theta)/\,\mathrm{d}\,\theta = 0$, sique que

$$\sum_{i=1}^n (y_i - \theta x_i)x_i = 0 \qquad \Longrightarrow \qquad \widehat{\theta} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}.$$

Notando que

$$\frac{\mathrm{d}^2}{\mathrm{d}\,\theta^2}S_1(\theta) = 2\sum_{i=1}^n x_i^2 > 0,$$

para cualquier $\theta \in \mathbb{R}$, sigue que $\widehat{\theta}$ es mínimo global.

Ejercicio 4.b)

Para los datos del Ejercicio 3, tenemos

$$\widehat{\theta} = \frac{280}{91} = 3.0769.$$

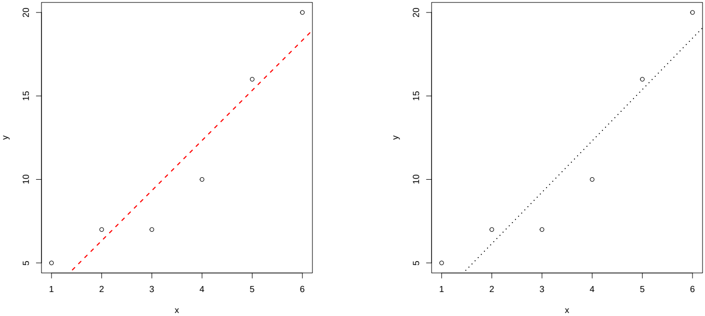
Calculando $e_i = y_i - \widehat{\theta} x_i$, para $i = 1, \ldots, 6$, resulta

$$\boldsymbol{e} = \{1.9231, 0.8462, -2.2308, -2.3077, 0.6154, 1.5385\}.$$

De este modo,

$$s_*^2 = \frac{1}{6-1} \sum_{i=1}^{6} e_i^2 = \frac{17.4620}{5} = 3.4924.$$

Ejercicio 3 y 4)[1]



(a) con intercepto

(b) sin intercepto

[1] Además, $R_0^2 = \{\text{corr}(\boldsymbol{y}, \widehat{\boldsymbol{y}})\}^2 = 0.9009.$

Ejercicio 3 y 4)

```
# ajuste del modelo 'con' intercepto
> fm <- lm(y ~ x, data = p3)
> fm

Call:
lm(formula = y ~ x, data = p3)

Coefficients:
(Intercept)            x
     0.3333       3.0000


# ajuste del modelo 'sin' intercepto
> f0 <- lm(y ~ -1 + x, data = p3)
> f0

Call:
lm(formula = y ~ -1 + x, data = p3)

Coefficients:
    x
3.077
```

Ejercicio 3 y 4)

```
# resumen de estimación del modelo 'con' intercepto
> summary(fm)

Call:
lm(formula = y ~ x, data = p3)

Residuals:
      1       2       3       4       5       6 
 1.6667  0.6667 -2.3333 -2.3333  0.6667  1.6667 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.3333     1.9379   0.172  0.87179
x             3.0000     0.4976   6.029  0.00382

Residual standard error: 2.082 on 4 degrees of freedom
Multiple R-squared:  0.9009,    Adjusted R-squared:  0.8761 
F-statistic: 36.35 on 1 and 4 DF,  p-value: 0.003815
```

Ejercicio 3 y 4)

```
# resumen de estimación del modelo 'sin' intercepto
> summary(f0)

Call:
lm(formula = y ~ -1 + x, data = p3)

Residuals:
      1       2       3       4       5       6
 1.9231  0.8462 -2.2308 -2.3077  0.6154  1.5385

Coefficients:
  Estimate Std. Error t value Pr(>|t|)
x   3.0769     0.1959   15.71  1.9e-05

Residual standard error: 1.869 on 5 degrees of freedom
Multiple R-squared:  0.9801,    Adjusted R-squared:  0.9762
F-statistic: 246.7 on 1 and 5 DF,  p-value: 1.902e-05
```