

ITAM - Métodos Estadísticos para C.Pol y R.I.

Assignment 02 - Respuestas

1. • Sabemos que:

Los estimadores de mínimos cuadrados de α y β son

$$b = \frac{S_{xy}}{S_{xx}}$$

y

$$a = \bar{y} - b\bar{x}$$

Entonces, para encontrar la recta de predicción de mínimos cuadrados para los datos de la tabla:

Paso 1:

$$S_{xx} = \sum x_i^2 - \frac{\sum x_i^2}{n} = 23634 - \frac{460^2}{10} = 2474$$

$$S_{xy} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n} = 36854 - \frac{460 \cdot 760}{10} = 1894$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{760}{10} = 76$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{460}{10} = 46$$

Paso 2:

$$b = \frac{S_{xy}}{S_{xx}} = \frac{1894}{2474} = 0.76556$$

y

$$a = \bar{y} - b\bar{x} = 76 - (0.76556)(46) = 40.78424$$

Entonces, la recta de regresión de mínimos cuadrados es:

$$\hat{y} = a + bx$$

o bien:

$$\hat{y} = \beta_0 + \beta_1 x = 40.78424 + 0.76556x$$

- Sabemos que:

$$S_{xx} = S_{x^2} = \sum x_i^2 - \frac{\sum x_i^2}{n} = 274.88$$

$$S_{xy} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n} = 210.44$$

$$S_{yy} = S_{y^2} = \sum y_i^2 - \frac{\sum y_i^2}{n} = 228.44$$

Entonces:

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$$

Entonces:

$$r = \frac{210.44}{\sqrt{274.88 \cdot 228.44}} = 0.8397$$

- Para realizar la prueba de hipótesis necesitamos:

– Para β_1

$H_o : \beta_1 = 0$ contra: $H_a : \beta_1 \neq 0$

Bajo H_o el estadístico de prueba es: $T_{\beta_1} = \frac{(\beta_1 - 0)}{\sqrt{(\sigma_{\beta_1}^2)}} = 1.786$

El valor p : (en R)

$2 * (1 - pt(T_{\beta_1}, df = 5 - 2))$

$valorp = 0.1720621$

– Para β_0

$H_o : \beta_0 = 0$ contra: $H_a : \beta_0 \neq 0$

Bajo H_o el estadístico de prueba es: $T_{\beta_0} = \frac{(\beta_0 - 0)}{\sqrt{(\sigma_{\beta_0}^2)}} = 6.4164$

El valor p : (en R)

$2 * (1 - pt(T_{\beta_0}, df = 5 - 2))$

$valorp = 0.007671261$

- Se debe realizar una prueba de hipótesis para determinar si la correlación entre las variables es significativa :

$H_o : \rho = 0$ contra: $H_a : \rho \neq 0$

Bajo H_o el estadístico de prueba es:

$$\rho = \frac{S_{xy}}{(\sqrt{S_{x^2}} * \sqrt{S_{y^2}})}$$

Sustituyendo tenemos: $\rho = 0.8797$

Entonces, el valor p

$$T_{obs} = \frac{(\rho * \sqrt{(5-2)})}{\sqrt{(1-\rho^2)}}$$

con la formula en R

$$2 * (1 - pt(T_{obs}, df = 5 - 2)) = 0.0751$$

concluimos que "hay una relación lineal significativa entre las calificaciones esperadas y la puntuación final del examen"

- Para estimar el promedio de las calificaciones cuyo aprovechamiento es de 50, con un intervalo de confianza debemos:

La estimación puntual de $E(y | x_0 = 50)$ el promedio de calificación es:

$$\hat{y} = 40.78424 + 0.76556(50) = 79.06$$

2. • Sabemos que:

$$S_{xx} = \sum x_i^2 - \frac{\sum x_i^2}{n}$$

$$S_{xy} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$$

$$S_{yy} = \sum y_i^2 - \frac{\sum y_i^2}{n}$$

$$\text{Entonces: } S_{xx} = 60.4 \quad S_{xy} = 328$$

$$S_{yy} = 2610$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$$

Entonces:

$$r = \frac{328}{\sqrt{60.4 * 2610}} = 0.8261$$

El valor obtenido para r es cerca a 1, indica que hay una relación lineal positiva bastante fuerte entre estatura y peso.

- Para probar si es significativamente diferente de cero, tenemos:

$$H_o : \rho = 0 \text{ contra } H_a : \rho \neq 0$$

El valor del estadístico de prueba es

$$t = r \sqrt{\frac{n-2}{1-r^2}}$$

$$= 0.8261 * \sqrt{\frac{10-2}{1-(0.8261)^2}} = 4.15$$

La t observada o el valor de tablas: tiene una distribución, con $n = 10$, de 8 grados de libertad.

Dado que la t observada es mayor que

$$t_{0.005} = 3.355$$

y el *valor p* es menor a $2(0.005) = 0.01$, entonces rechazamos H_o . Concluimos que la correlación es significativamente diferente de 0

3. • Tenemos que:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Entonces:

$$S_{xx} = \sum x_i^2 - \frac{\sum x_i^2}{n}$$

$$S_{xy} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$$

$$S_{yy} = \sum y_i^2 - \frac{\sum y_i^2}{n}$$

$$\hat{\beta}_1 = \frac{7 - \frac{1}{5}(0)(5)}{10 - \frac{1}{5}(0)^2} = 0.7$$

Ahora:

$$\bar{y} = \frac{\sum(y_i)}{n} = 0.7$$

$$\bar{x} = \frac{\sum(x_i)}{n} = 0$$

$$\hat{\beta}_0 = \frac{5}{5} - (0.7)(0) = 1$$

Entonces, la recta de regresión de mínimos cuadrados es:

$$\hat{y} = \bar{y} + \hat{\beta}_0 x$$

$$\hat{y} = 1 + 0.7x$$

- Ahora, para determinar el intervalo de confianza al 90% para E(y) cuando x= 1 tenemos que:

$$E(y) = \beta_0 + \beta_1 x$$

Para estimar dicho valor fijo, usaremos el estimador insesgado $E(\hat{Y}) = \hat{\beta}_0 + \hat{\beta}_1 x^*$ Entonces:

$$E(\hat{Y}) = 1 + 0.7x^* \text{ En este caso, } x^* = 1 \text{ y como } n = 5, \bar{x} = 0 \text{ y } S_{xx} = 10$$

Entonces, el error estándar es

$$\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}} = \frac{1}{5} + \frac{(1-0)^2}{10} = 0.3$$

Ahora bien, el valor de la $t_{0.05}$ de tablas es con n-2= 3 grados de libertad es 2.353

El intervalo está dado por:

$$\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\alpha/2} * S * \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}}$$

sustituyendo:

$$((1 + (0.7)(1)) \pm (2.353)(0.606)\sqrt{0.3})$$

= 1.7 ± 0.781 El intervalo está dado por:

$$(0.919, 2.481)$$

Es decir: Con un nivel de confianza de 90% cuando la variable independiente tome el valor de 1, la variable dependiente

- Para demostrar que la pendiente difiere de cero dicho ejercicio implica una prueba de hipótesis

$$H_o : \beta_1 = 0 \text{ contra } H_a : \beta_1 \neq 0$$

El valor del estadístico de prueba es

$$t = \frac{\hat{\beta}_1}{s\sqrt{c_{11}}}$$

Sustituyendo tenemos:

$$c_{11} = \frac{1}{S_{xx}} = 1/10 = 0.1$$

$$t = \frac{0.7-0}{0.606\sqrt{0.1}} = 3.65$$

Ahora, si tomamos con un $\alpha = 0.05$, el valor de t de tablas, $t_{\alpha/2} = t_{0.025}$ con 3 grados de libertad es 3.182

Bajo la hipótesis alternativa la región de rechazo está dado por:

$$-3.182 \leq t \leq 3.182$$

Ahora bien, dado que es una prueba de dos colas para calcular el valor p:

$$valorp = 2P(t > 3.65)$$

$$= 0.01 < P(t > 3.65) < 0.025$$

Entonces:

$$= 0.02 < valorp < 0.05$$

Dado que el p value es menor a 0.1, rechazamos la hipótesis nula

- Para calcular el intervalo de confianza para β_1 tenemos:

El valor de tablas para $t_{0.025}$ con 3 grados de libertad es 3.182. Entonces el intervalo de confianza está dado por

$$\hat{\beta}_1 \pm t_{0.025} * s * \sqrt{c_{11}}$$

siendo $c_{11} = 1/S_{xx}$

Sustituyendo, obtenemos

$$0.7 \pm (3.182)(0.606)\sqrt{0.1}$$

O bien,

$$0.7 \pm 0.610$$

El parámetro de β_1 variará entre 0.9 y 1.31 unidades

- Para determinar el pronóstico puntual cuando $x = 2$ con un nivel de confianza de 0.90. Tenemos:

$$\hat{\beta}_0 = 1$$

$$\hat{\beta}_1 = 0.7$$

De modo que el valor pronóstico para Y cuando $x = 2$ es

$$\hat{\beta}_0 + \hat{\beta}_1 x^* = 1 + (0.7)(2) = 2.4$$

Además, el intervalo de predicción está dado por:

$$\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\alpha/2} S \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}}$$

Sustituyendo, tenemos:

$$2.4 \pm (2.353)(0.606)\sqrt{1 + 0.6}$$

$$2.4 \pm 1.804$$

4. • Para ajustar el modelo de regresión de mínimos cuadrados:

$$S_{xx} = \sum x_i^2 - \frac{\sum x_i^2}{n}$$

$$S_{xy} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$$

$$S_{yy} = \sum y_i^2 - \frac{\sum y_i^2}{n}$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Sustituyendo, obtenemos:

$$S_{xy} = 8.709 - \frac{1}{6}(8.74)(6.148) = -0.247$$

$$S_{xx} = 12.965 - \frac{1}{6}(8.74)^2 = 0.234$$

$$S_{yy} = 6.569 - \frac{1}{6}(6.148)^2 = 0.269$$

$$\hat{\beta}_1 = \frac{-0.247}{0.234} = -1.056$$

$$\hat{\beta}_0 = \frac{6.148}{6} - (-1.056)\left(\frac{8.74}{6}\right) = 2.563$$

Entonces, la recta de minimos cuadrados está dada por

$$\hat{\beta}_0 + \hat{\beta}_1 x = -1.056 + 2.563x$$

- Prueba de hipótesis para β_1
- Para realizar la prueba de hipótesis necesitamos:

– Para β_1

$$H_o : \beta_1 = 0 \text{ contra: } H_a : \beta_1 \neq 0$$

Bajo Ho el estadístico de prueba es:

$$T_{\beta_1} = \frac{(\beta_1 - 0)}{\sqrt{(\sigma_{\beta_1}^2)}} = -9.8263$$

El valor p : (en R)

$$2 * (1 - pt(T_{\beta_1}, df = 5 - 2))$$

se obtiene :

$$valorp = 1.9977$$

- Para calcular intervalo de confianza para β_1 al 90%

Necesitamos:

$$\beta_1 \epsilon (\hat{\beta}_1 \pm t_{\alpha/2, n-2} * \sqrt{Var(\hat{\beta}_1)})$$

Sustituyendo:

$$t_{-10/2, 6-2} = 0.07510391 \text{ cálculo de T en R}$$

$$pt(0.05, df = 4, lower.tail = false)$$

El intervalo está dado por:

$$\beta_1 \epsilon (-1.0655 \pm 0.07510391 * \sqrt{0.1084})$$

Bibliografía Wackerly. (2008). Estadística Matemática con Aplicaciones (7.a ed.). Cengage Learning.