Matemática IV.
Trabajo práctico II

Integrantes:

- Aguado Thomas, 14245/6
- Fermín Moreno, 16276/2

Un investigador realiza mediciones de los pesos y de los perımetros de los craneos (MDC) en dos razas distintas de una misma especie de mamíferos. Los resultados obtenidos est´an dados a continuación:

Raza A:

MDC (en cm)	21	22	22	25	26	26	27	28	29	29	30	31
Peso (en kg)	4.8	5.9	5.0	7.0	6.5	6.1	8.1	8.8	9.4	9.0	11.4	10.4

Raza B:

MDC (en cm)	20	20	22	23	23	24	25	27	28	28
Peso (en kg)	4.2	4.9	5.7	5.8	5.5	6.2	5.7	7.6	6.9	6.9

1. Ajustar un modelo de regresión lineal considerando a "Medida del cráneo" como variable independiente y "Peso" como variable dependiente para ambas razas, y expresar los valores estimados de β0 y β1. ¿Qué suposiciones son necesarias para que esta modelización sea válida?

Raza 1 datos: x̄ = 26,3 ∑xᵢ = 316 ∑xᵢ^2 = 8442 ȳ = 7,7 ∑yᵢ = 92,4 ∑yᵢ^2 = 5866,168

La suposiciones necesarias para que estas modelizaciones sean válidas es que ambas tengan distribución normal.

Si asumimos que Xi yi son independientes y que las i tienen todas la misma distribución con E i 0 , entonces E Yi Xi 0 1Xi Si además suponemos que 2 i ~ N 0, entonces se puede probar que los estimadores de máxima verosimilitud para los parámetros 0 y 1 son

β1 = Sxy/Sxx

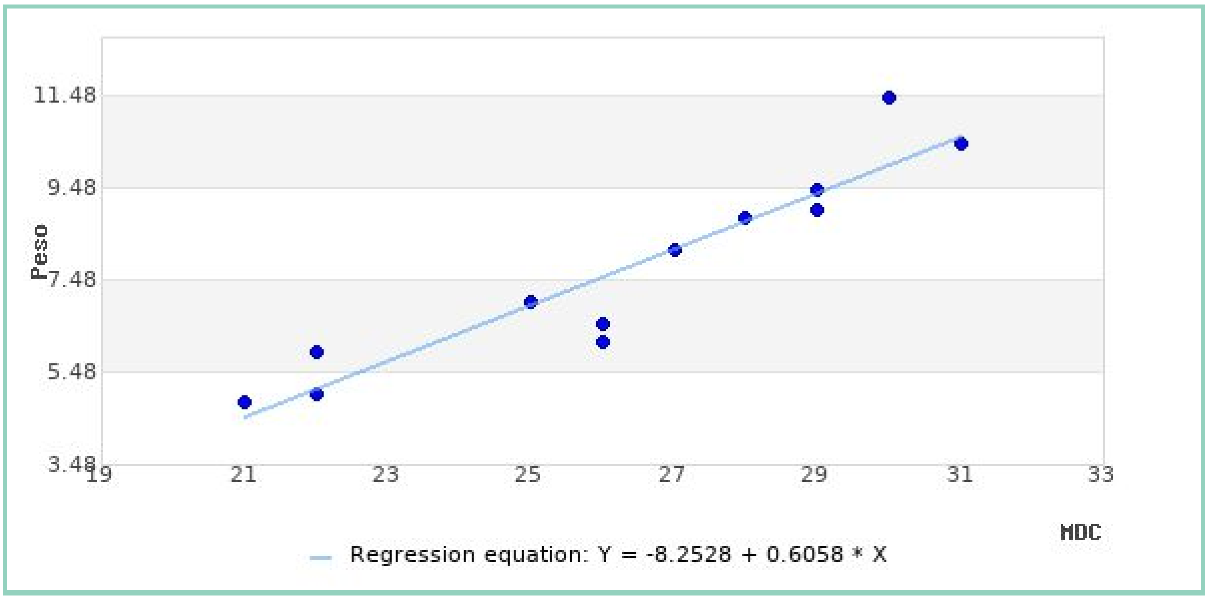
Sxx = ∑(xi - x̄)^2 = 8442-(316^2)/12= 120.66666666667

Sxy = ∑XiYi - 1/n (∑Xi)(∑Xi) = 2506,3 - (316* 92,4) /12 = 73,1

(2506,3 se obtiene de la suma de la multiplicación de xi,yi)

β1 = Sxy/Sxx
β1 = 73.1/120.66666666667
β1 = 0.6058

β0 = ȳ - β1*x̄
β0 =7,7 - 0.6058* 26.3 =
β0 = -8.2528



Raza 2 datos: $\bar{x} = 24$ $\sum x_i = 240$ $\sum x_i^2 = 5840$ $\bar{y} = 5{,}94$ $\sum y_i = 59{,}4$ $\sum y_i^2 = 361{,}94$

Sxx = ∑(xi - x̄)^2 = (20-24)^2 +(20-24)^2 +(22-24)^2 +(23-24)^2 +(23-24)^2 +(24-24)^2 +(25-24)^2 +(27-24)^2 +(28-24)^2 +(28-24)^2 =

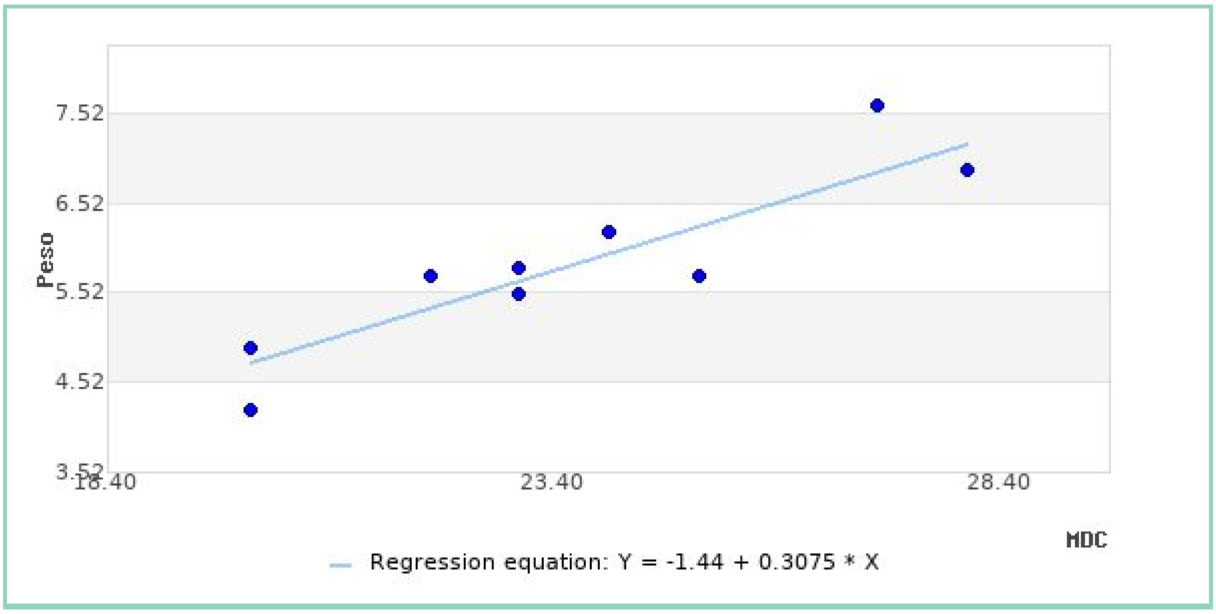
Sxx = 16 + 16 + 4 + 1 + 1 + 0 + 1 + 9 + 16 + 16
Sxx = 80

Sxy = ∑yi(xi - x̄) = 4.2*(20-24) +4.9*(20-24) +5.7*(22-24) +5.8*(23-24) +5.5*(23-24) +6.2*(24-24)+5.7*(25-24) +7.6*(27-24) +6.9*(28-24) +6.9*(28-24) =

Sxy = -16.8 - 19.6 - 11.4 - 5.8 - 5.5 + 0 + 5.7 + 22.8+ 27.6 + 27.6
Sxy = 24.6

β1 = 24.6/80
β1 = 0.3075

β0 = ȳ - β1*x̄
β0 =5.94 - 0.3075*24=
β0 = -1.44



2. Hallar el valor del coeficiente de determinación R2 de los modelos ajustados en el inciso anterior. ¿Que representa este valor?

R2 = 1 - SSR/Syy

Raza 1:

SSR = ∑(yi - Ŷi)^2

= (4.8-4.49)^2 +(5.9-5.09)^2 +(5.0-5.09)^2 +(7.0-6.91)^2 +(6.5-7.51)^2 +(6.1-7.51)^2 +(8.1-8.12)^2 +(8.8-8.72)^2 +(9.4-9.3)^2 +(9.0-9.3)^2 +(11.4-9.93)^2 +(10.4-10.54)^2 =

SSR = 0.0961 + 0.6561 + 0.0081 + 0.0081 + 1.0201 + 1.9881 + 0.0004 + 0.0064 + 0.01 + 0.09 + 2.1609 + 0.0196

SSR= 6.0639

Syy =∑(yi - ȳ)^2 = (4.8-7.7)^2 +(5.9-7.7)^2 +(5.0-7.7)^2 +(7.0-7.7)^2 +(6.5-7.7)^2 +(6.1-7.7)^2 +(8.1-7.7)^2 + (8.8-7.7)^2 +(9.4-7.7)^2 +(9.0-7.7)^2 +(11.4-7.7)^2 +(10.4-7.7)^2 =

Syy = 8.41 + 3.24 + 7.29 + 0.49 + 1.44 + 2.56 + 0.16 + 1.21 + 2.89 + 1.69 + 13.69 + 7.29

Syy = 50.36

R2 = 1 - SSR/Syy =

R2 = 1 - 6.0639/50.36

R2 = 1 - 0.1204

R2 = 0.8795

Raza 2

SSR = ∑(yi - Ŷi)^2 = (4.2-4.66)^2 + (4.9-4.66)^2 + (5.7-5.3)^2 + (5.8-5.62)^2 + (5.5-5.62)^2 + (6.2-5.94)^2 +(5.7-6.26)^2 + (7.6-6.9)^2 + (6.9-7.22)^2 +(6.9-7.22)^2

SSR = 0.2116 + 0.0576 + 0.16 + 0.0324 + 0.0144 + 0.0676 + 0.3136 + 0.49 + 0.1024 + 0.01024 =

SSR = 1.552

Syy = ∑(yi - ȳ)^2 = (4.2-5,94)^2 +(4.9-5,94)^2 +(5.7-5,94)^2 +(5.8-5,94)^2 +(5.5-5,94)^2 +(6.2-5,94)^2 +(5.7-5,94)^2 +(7.6-5,94)^2 +(6.9-5,94)^2 +(6.9-5,94)^2 =

Syy = 3,0276 + 1,0816 + 0,0576 + 0,0196 + 0,1936 + 0,0676 + 0,0576 + 2,7556 + 0,9216 + 0,9216

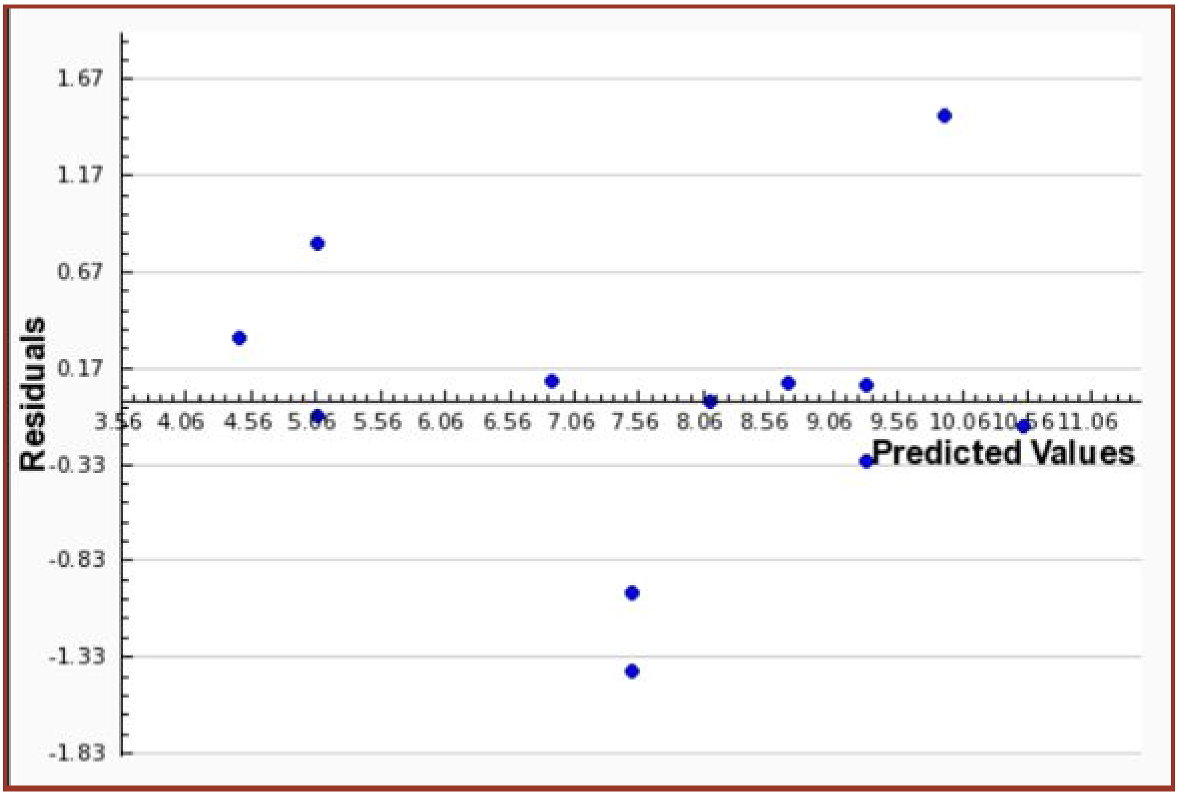
Syy = 9,104

R2 = 1 - SSR/Syy =
R2 = 1 - 1,552/9,104
R2 = 1 - 0,1704
R2 = 0,8295

3. Realizar un gráfico de los residuos de los modelos ajustados previamente. ¿Se detecta algún patrón? 0

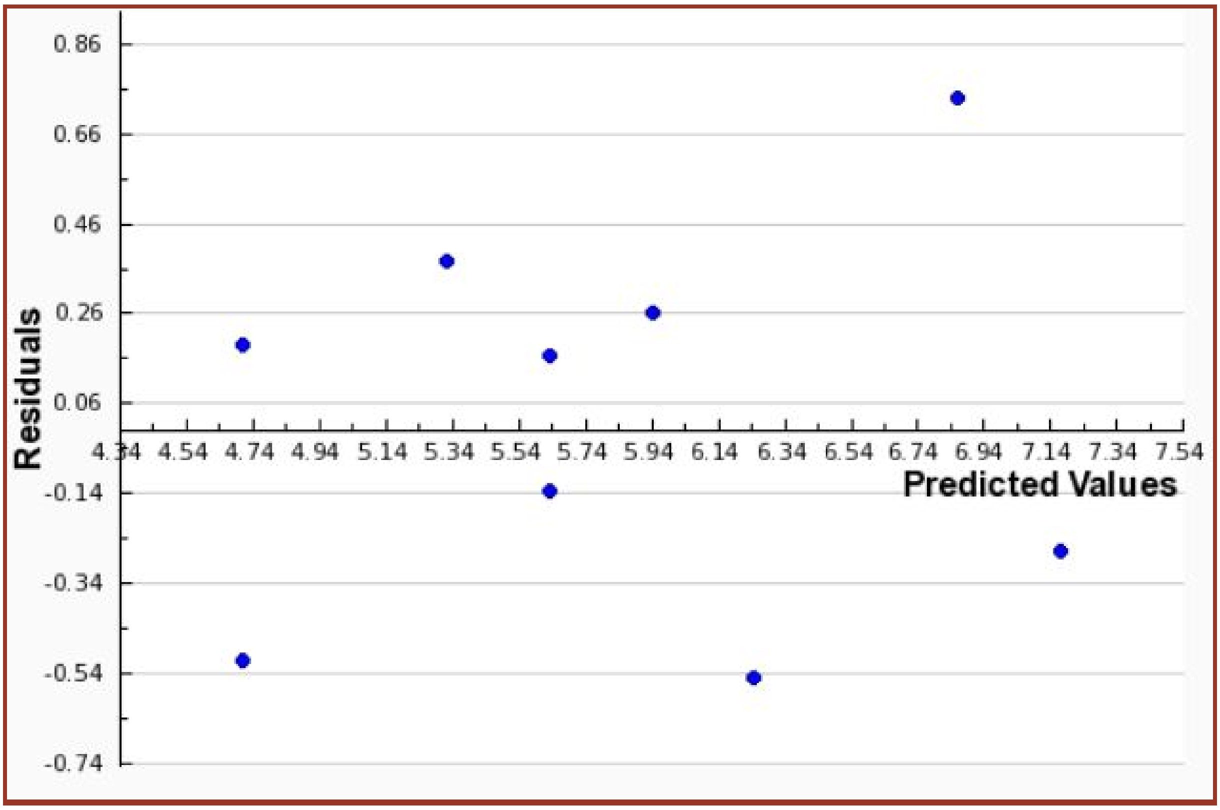
Raza 1

X	Y	Ŷ	Residuos (Y−Ŷ)
21	4.8	-8.2528+0.6058 x 21= 4.469	4.8 - 4.469 = 0.331
22	5.9	-8.2528+0.6058 x 22= 5.075	5.9 - 5.075 = 0.825
22	5.0	-8.2528+0.6058 x 22= 5.075	5.0 -5.075 = -0.075
25	7.0	-8.2528+0.6058 x 25= 6.892	7.0 - 6.892 = 0.108
26	6.5	-8.2528+0.6058 x 26= 7.498	6.5 - 7.498 = -0.998
26	6.1	-8.2528+0.6058 x 26= 7.498	6.1 - 7.498 = -1.398
27	8.1	-8.2528+0.6058 x 27= 8.104	8.1 - 8.104 = -0.00399
28	8.8	-8.2528+0.6058 x 28= 8.71	8.8 - 8.71 = 0.09
29	9.4	-8.2528+0.6058 x 29= 9.315	9.4 – 9.315 = 0.085
29	9.0	-8.2528+0.6058 x 29= 9.315	9.0 - 9.315 = -0.315
30	11.4	-8.2528+0.6058 x 30= 9.921	11.4 - 9.921 = 1.479
31	10.4	-8.2528+0.6058 x 31= 10.527	10.4 - 10.527 = -0.127



Raza 2

X	Y	Ŷ	Residuos (Y-Ŷ)
20	4.2	-1.44+0.3075 x 20= 4.71	4.2 - 4.71 = -0.51
20	4.9	-1.44+0.3075 x 20= 4.71	4.9 – 4.71 = 0.19
22	5.7	-1.44+0.3075 x 22= 5.325	5.7 -5.325 = 0.375
23	5.8	-1.44+0.3075 x 23= 5.633	5.8 – 5.633 = 0.167
23	5.5	-1.44+0.3075 x 23= 5.633	5.5 – 5.633 = -0.133
24	6.2	-1.44+0.3075 x 24= 5.94	6.2 – 5.94 = 0.26
25	5.7	-1.44+0.3075 x 25= 6.248	5.7 – 6.248 = -0.548
27	7.6	-1.44+0.3075 x 27= 6.863	7.6 – 6.863 = 0.737
28	6.9	-1.44+0.3075 x 28= 7.17	6.9 – 7.17 = -0.27
28	6.9	-1.44+0.3075 x 28= 7.17	6.9 – 7.17 = -0.27



Como ambos gráficos nos dan valores entre -2 y 2 y hay una distribución pareja no se detecto ningún patrón.

4. Hallar los intervalos de confianza de 95 % y 99 % de los parámetros de regresión β1 en ambas razas.

[β1 - tα/2, n-2 √(σ^2/Sxx);β1 + tα/2, n-2 √(σ^2/Sxx)]

95% = 1-95 = 0,05 = α
99% = 1-99 = 0,01 = α

Raza 1
con 95%

[β1 - tα/2, n-2 √(σ^2/Sxx);β1 + tα/2, n-2 √(σ^2/Sxx)]
t0,025,10 = 2.228
σ^2 = (SSR/n-2)
σ^2 = 6.0639/10 = 0,60639
[0.6058 - 2.228* √(0,60639/120.66666666667); 0.6058 + 2.228* √(0,60639/120.66666666667)]

[0.6058 - 2.228* 0,0708; 0.6058 + 2.228* 0,0708]
[0,4480576; 0,7635424]

con 99%

$t_{0,005,10} = 3.169$

[0.6058 -3.169 * 0,0708; 0.6058 + 3.169* 0,0708]

[0,3814348; 0,8301652]

Raza 2

con 95%

$[\beta_1 - t_{\alpha/2,\ n-2} \sqrt{(\sigma^2/S_{xx})}; \beta_1 + t_{\alpha/2,\ n-2} \sqrt{(\sigma^2/S_{xx})}]$

$\beta_1 = 0.3075$

$t_{0,025,8} = 2.306$

$\sigma^2 = (SSR/n-2)$

$\sigma^2 = 1.552/8 = 0,194$

[0.3075- 2.306 * √(0,194/80);0.3075 + 2.306 √(0,194/80)]

[0.3075- 2.306* 0,0492;0.3075 + 2.306 * 0,0492]

[0,1940448;0,4209552]

con 99%

$t_{0,005,8} = 3.355$

[0.3075- 3.355* √(0,194/80);0.3075 + 3.355*√(0,194/80)]

[0.3075- 3.355* 0,0492;0.3075 + 3.355 * 0,0492]

[0,142434;0,472566]

5. ¿Los datos evidencian el hecho de que el tamaño del cráneo influye en el peso en alguna de las dos razas? Justifique utilizando el p-valor. ¿Se podría concluir esto en el inciso anterior? (Considerar α = 0.05).

Para empezar necesitamos plantear las hipótesis:

Raza 1

$H_0: \beta_1 = 0$ contra $H_1: \beta_1 \neq 0$

Armamos el estadístico T

$T = |\beta_1/\sqrt{(\sigma^2/S_{xx})}|$

con $\alpha = 0,05$

$\beta_1 = 0.6058$

$\sigma^2 = 0,60639$

$S_{xx} = 120.66666666667$

$\sqrt{(\sigma^2/S_{xx})} = 0,0708$

$T = |0.6058/0,0708|$

$T = 8,5564$

por lo tanto

p-valor = $(|T| > 8,5564)$

por tabla sabemos que V= 10,0,0005 = 4.587

entonces sabemos que el p-valor es más chico que 0,05, en conclusión se rechaza H_0, por lo tanto el tamaño del cráneo no influye en esta raza

Raza 2

$H_0: \beta_1 = 0$ contra $H_1: \beta_1 \neq 0$

Armamos el estadístico T

$T = |\beta_1/\sqrt{(\sigma^2/S_{xx})}|$

con $\alpha = 0,05$

$\beta_1 = 0.3075$

$S_{xx} = 80$

$\sigma^2 = 0,194$

$\sqrt{(\sigma^2/S_{xx})} = 0,0492$

$T = |0.3075/0,0492|$

$T = 6,25$

por lo tanto

p-valor = $(|T| > 6,25)$

por tabla sabemos que V= 10,0,0005 = 4.587

entonces sabemos que el p-valor es más chico que 0,05, en conclusión se rechaza H_0, por lo tanto el tamaño del cráneo no influye en esta raza