Matemática 4 - Entrega 2

Integrantes:

Milla, Andrés - 14934/6
Mandarino, Leonel - 15379/5
Almandos, Julián - 15295/1

Código fuente

El código fuente se encuentra en este link.
Para ejecutarlo entrar a este link.

Ejercicio 1

Para ajustar un modelo de regresión lineal, primero debemos calcular $\bar{x}$ e $\bar{y}$, de manera que se puedan calcular Sxx y Sxy y posteriormente los estimadores de β_0 y β_1 y la recta estimada, a través del **método de mínimos cuadrados**.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$S_{xy} = \sum_{i=1}^{n} y_i(x_i - \bar{x})$$

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

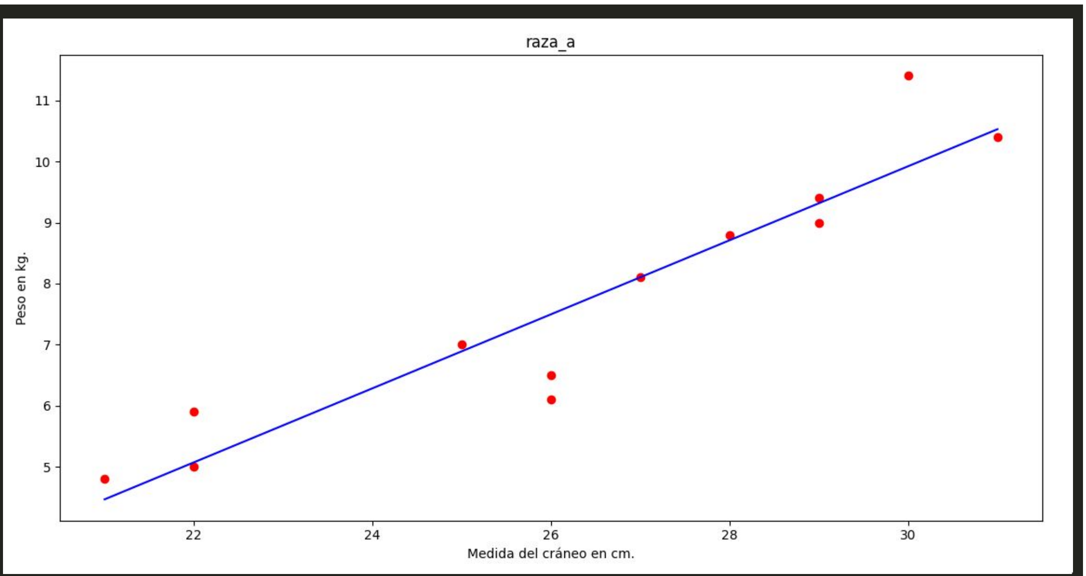
$$\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1\bar{x}$$

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x$$

Para calcular los estimadores, realizamos un código en el lenguaje de programación **Python** y obtuvimos los siguientes resultados:

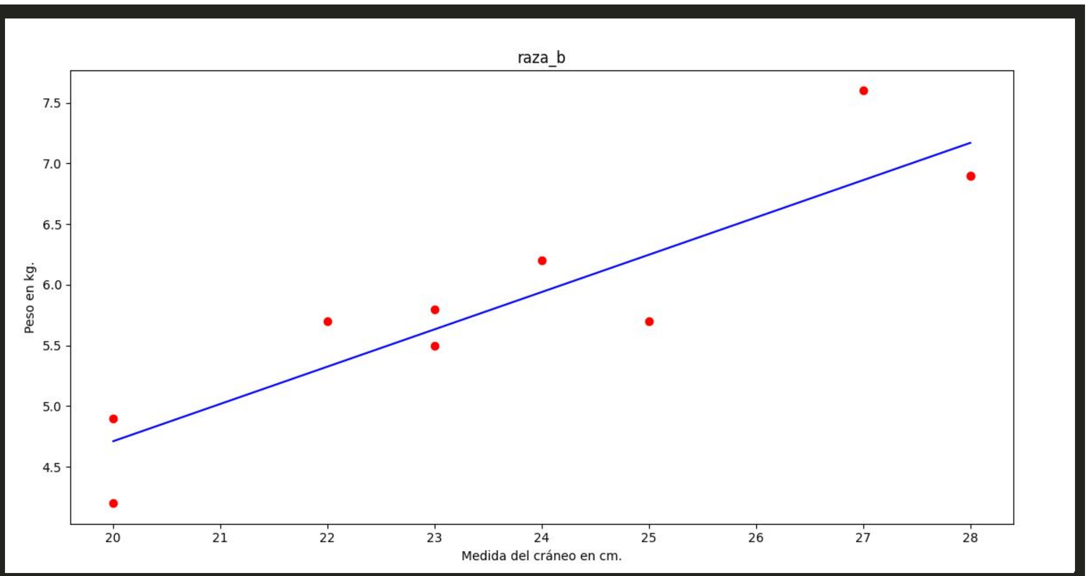
Raza A

```
----------
x̄ = 26.33
ȳ = 7.70
Sxx = 120.67
Sxy = 73.10
----------
ESTIMADORES
----------
β1 = 0.61
β0 = -8.25
----------
MODELO FINAL
----------
y = -8.25 + 0.61x
```



Raza B

```
----------
x̄ = 24.00
ȳ = 5.94
Sxx = 80.00
Sxy = 24.60
----------
ESTIMADORES
----------
β1 = 0.31
β0 = -1.44
----------
MODELO FINAL
----------
y = -1.44 + 0.31x
```



Para que esta modelización sea válida debemos suponer que:
(1) La frecuencia de los errores tienen distribución normal
(2) Los errores son independientes
(3) Los errores tienen varianza constante (homocedasticidad)
(4) La relación entre los parámetros es lineal

Ejercicio 2

El valor del coeficiente de determinación R^2 se calcula de la siguiente manera:

$$R^2 = 1 - \frac{SS_R}{S_{YY}}$$

Por lo tanto, deberemos calcular SS_R y S_{YY} respectivamente:

$$SS_R = \sum_{i=1}^{n} (y_i - \widehat{y_i})^2$$

$$S_{YY} = \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

Este valor representa la proporción de la variación de la respuesta Y que es explicada por el modelo, es decir, qué tan bien el modelo representa las variaciones de la variable Y. Esta variable toma valores entre 0 y 1. Además, podemos calcular el **índice de ajuste** $\sqrt{R^2}$ (también llamado **coeficiente de correlación muestral)** que se utiliza como indicador de qué tan bien el modelo de regresión ajusta los datos.
Hay que tener en cuenta que si el coeficiente de determinación toma un valor alto (cercano a 1), esto no necesariamente significa que el modelo de regresión sea correcto. Puede que el modelo represente bien la variación sobre los datos dados por la muestra, pero no nos sirva como estimador de valores a futuro.

Para calcular el coeficiente de determinación R^2 para ambos modelos utilizamos el código de **Python** manualmente, con las librerías *numpy* y *statsmodel*. Nos dió la siguiente salida:

Raza A

```
Manualmente:
R2: 0.879
=================
Con statsmodel:
R-squared: 0.879
=================
Con numpy:
R2: 0.879
```

Raza B

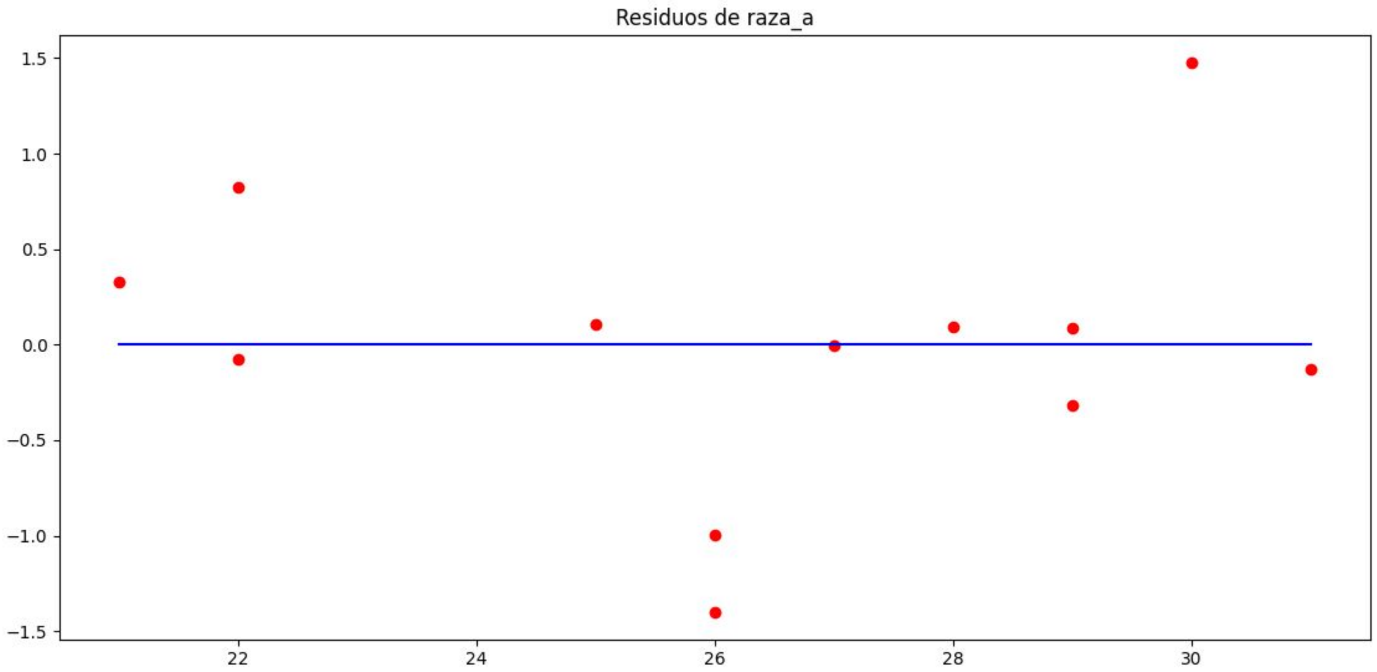
```
Manualmente:
R2: 0.831
=================
Con statsmodel:
R-squared: 0.831
=================
Con numpy:
R2: 0.831
```

Esto significa que el modelo de regresión lineal simple explica el 87,9% de la variabilidad total de las observaciones en el caso de la Raza A y el 83,1% en la Raza B

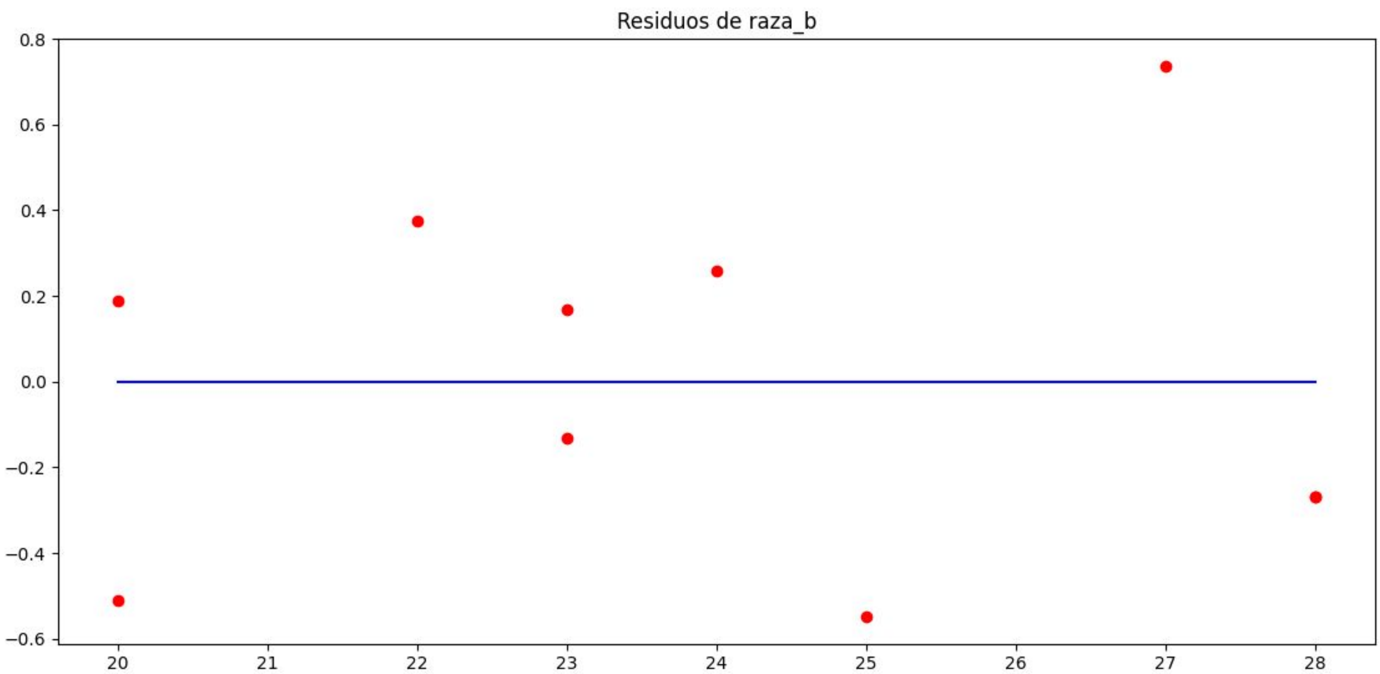
Ejercicio 3

En particular, no detectamos ningún patrón distinguible respecto a los residuos de los modelos ajustados previamente. Por lo tanto los ejemplos se adecuan bien al modelo de regresión lineal.

Raza A



Raza B



Ejercicio 4

Ambos ejercicios los resolvimos con código hecho en Python, por lo que dejamos expresados los resultados al final de cada paso.

Raza A

Para construir un intervalo de confianza para la pendiente de nivel 1 - ∝ = 0.95 buscamos:

$$\frac{\widehat{\beta}_1 - \beta_1}{s_r/\sqrt{S_{xx}}} \sim T_{(n-2)}$$

Lo que nos da como resultado $t_{0,025,\ 10}$ = 2.228

Luego calculamos el intervalo de confianza:

$$IC_{1-\alpha}(\beta_1) = \left(\widehat{\beta}_1 - t_{\alpha/2,n-2} \times \frac{s_r}{\sqrt{S_{xx}}}\ , \widehat{\beta}_1 + t_{\alpha/2,n-2} \times \frac{s_r}{\sqrt{S_{xx}}}\right)$$

$IC_{1-\alpha}\ \beta_1$ = $IC_{0.95}$ (0.61) = (0.45, 0.76)

En el caso de que el intervalo de confianza de la pendiente sea de nivel 1 - ∝ = 0.99 aplicamos el mismo procedimiento.

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

Y luego el intervalo de confianza
$IC_{1-\alpha}\ \beta_1$ = $IC_{0.95}$ (0.61) = (0.38, 0.83)

Raza B

Para construir un intervalo de confianza para la pendiente de nivel 1 - ∝ = 0.95 buscamos:

$$\frac{\widehat{\beta}_1 - \beta_1}{s_r/\sqrt{S_{xx}}} \sim T_{(n-2)}$$

Lo que nos da como resultado $t_{0,025,\ 10}$ = 2.306

Luego calculamos el intervalo de confianza:

$$IC_{1-\alpha}(\beta_1) = \left(\widehat{\beta}_1 - t_{\alpha/2,n-2} \times \frac{s_r}{\sqrt{S_{xx}}}\ , \widehat{\beta}_1 + t_{\alpha/2,n-2} \times \frac{s_r}{\sqrt{S_{xx}}}\right)$$

$IC_{1-\alpha}\ \beta_1$ = $IC_{0.95}$(0.31) = (0.19, 0.42)

En el caso de que el intervalo de confianza de la pendiente sea de nivel 1 - ∝ = 0.99 aplicamos el mismo procedimiento.

Obtenemos
$t_{0,005,\ 10}$ = 3.355

Y luego el intervalo de confianza
$IC_{1-\alpha}\ \beta_1$ = $IC_{0.95}$(0.31) = (0.14, 0.47)

Este procedimiento lo hicimos en el código, con el método **conf_int** de la librería **statsmodels**. Obtuvimos las mismas salidas:

```
          Raza A:

 β1 IC 95%: [0.45, 0.76]
 β1 IC 99%: [0.38, 0.83]
-------------------------
          Raza B:

 β1 IC 95%: [0.19, 0.42]
 β1 IC 99%: [0.14, 0.47]
```

Ejercicio 5

Lo primero que debemos hacer es calcular el p-valor para la siguiente hipótesis nula:

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

El **p-valor** de la raza A da como resultado 0.00000663 < 0.05, por lo tanto hay evidencia suficiente para rechazar la hipótesis nula y concluir que efectivamente **el tamaño del cráneo influye en el peso de la raza A**.

El **p-valor** de la raza B da como resultado 0.00024053 < 0.05, por lo tanto hay evidencia suficiente para rechazar la hipótesis nula y concluir que efectivamente **el tamaño del cráneo influye en el peso de la raza B**.

Si, se podía concluir esto en el inciso anterior. Partiendo de la hipótesis nula antes declarada, podríamos rechazarla si se cumple la siguiente desigualdad:

$$|t| > t_{\alpha/2,\, n/2}$$

En nuestro caso:

$$|t| > t_{0.025,\, 10}$$

Calculamos |t|, siendo $\beta_1 = 0$, para cada una de las razas:

$$\frac{\widehat{\beta}_1 - \beta_1}{s_r / \sqrt{S_{xx}}} \sim T_{(n-2)}$$

Raza A

$$|8.54| > 2.228$$

Concluimos entonces que tenemos evidencia suficiente para rechazar la hipótesis nula y por lo tanto concluir que **el tamaño del cráneo influye en el peso de la raza A**.

Raza B

$$|6.27| > 2.306$$

Concluimos entonces que tenemos evidencia suficiente para rechazar la hipótesis nula y por lo tanto concluir que **el tamaño del cráneo influye en el peso de la raza A**.

También se puede decir que, por la relación de test de hipótesis con intervalos de confianza y los datos del ejercicio anterior, se puede afirmar con **α = 0.05** (y también con **α = 0.01**) que las variables son dependientes, ya que H_0 no pertenece a ningún intervalo calculado en ambas razas.