# Tarea #4

## EQUIPO #3

Sofía Alejandra Díaz Miranda 172360
David Isaac López Romero 173993
Sofía Oliva Ruiz 164595
Adriana Alavez Lujano 163480
Diego Carlos Krafft de Silva 173246

Se pretende explicar el ingreso a través de la edad y el género de la persona, así que:

a) **Ajuste un modelo** de regresión lineal simple para los hombres y otro para las mujeres. En cada caso, proporcione los resultados que considere más relevantes e interprételos.

b) Para cada una de las regresiones del inciso anterior, **realice un análisis** de residuos y, de ser posible, **obtenga un mejor modelo**.

Resultados para a) ∧ b):

Regresión Hombres:

SUMMARY OUTPUT

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.933190961 |
| R Square | 0.870845369 |
| Adjusted R Square | 0.79335259 |
| Standard Error | 20666.36556 |
| Observations | 9 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 3 | 14398898895 | 4799632965 | 11.23776156 | 0.011634241 |
| Residual | 5 | 2135493327 | 427098665.4 | | |
| Total | 8 | 16534392222 | | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
|---|---|---|---|---|---|---|---|---|
| Intercept | -1149950.833 | 434005.3922 | -2.649623378 | 0.045447014 | -2265597.211 | -34304.45564 | -2265597.211 | -34304.45564 |
| Persona | -393484.1667 | 132770.4371 | -2.963642926 | 0.03138655 | -734781.4405 | -52186.89281 | -734781.4405 | -52186.89281 |
| Edad | 80215 | 26199.40017 | 3.061711317 | 0.02804727 | 12867.29782 | 147562.7022 | 12867.29782 | 147562.7022 |
| Género | 0 | 0 | 65535 | #N¡NUM! | 0 | 0 | 0 | 0 |

Nuestro modelo se explica 87.08 %

Tenemos $\beta_{persona} = -393484.1667$  Error estandar = 20666.36556
$\beta_{edad} = 80215$
$\beta_{género} = 0$

En cuanto al inciso b), podemos ver específicamente la tabla de ANOVA, donde se ven los grados de libertad (3+5 = 8) y suma de cuadrados 16534392222.

Regresión Mujeres:

SUMMARY OUTPUT

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.944111619 |
| R Square | 0.891346748 |
| Adjusted R Square | 0.826154797 |
| Standard Error | 6761.191043 |
| Observations | 9 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 3 | 1875082478 | 625027492.8 | 13.67265029 | 0.007613081 |
| Residual | 5 | 228568521.6 | 45713704.32 | | |
| Total | 8 | 2103651000 | | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
|---|---|---|---|---|---|---|---|---|
| Intercept | 1388830.595 | 249260.7279 | 5.571798681 | 0.002564617 | 748085.4957 | 2029575.695 | 748085.4957 | 2029575.695 |
| Persona | -220476.3095 | 43437.06625 | -5.075764285 | 0.00384822 | -332134.843 | -108817.776 | -332134.843 | -108817.776 |
| Edad | 44165.35714 | 8571.374066 | 5.152657766 | 0.003606845 | 22131.93866 | 66198.77562 | 22131.93866 | 66198.77562 |
| Género | 0 | 0 | 65535 | #N¡NUM! | 0 | 0 | 0 | 0 |

Nuestro modelo se explica 89.13%

Tenemos $\beta_{persona} = -220476.3095$ Error estándar = 6761.191043
$\beta_{edad} = 44165.3514$
$\beta_{género} = 0$

En cuanto al inciso b), podemos ver específicamente la tabla de ANOVA, donde se ven los grados de libertad (3+5 = 8) y suma de cuadrados 2103651000.

b) El ajuste que realizamos fue agregar una nueva variable explicativa
Agregamos la edad al cuadrado y al hacerlo se reduce la tendencia cuadrática de los residuos

c) Considere ahora un modelo de regresión en donde aparezcan las 18 observaciones e **incluya la variable género**. ¿Cómo se interpreta el coeficiente de esta variable?

Tenemos que calcular $b = (X'X)^{-1} X'Y$

$$X = \begin{pmatrix} 1 & 21 & 441 & 1 \\ 1 & 27 & 729 & 1 \\ 1 & 32 & 1024 & 1 \\ 1 & 37 & 1369 & 1 \\ 1 & 42 & 1764 & 1 \\ 1 & 47 & 2209 & 1 \\ 1 & 52 & 2704 & 1 \\ 1 & 57 & 3249 & 1 \\ 1 & 62 & 3844 & 1 \\ 1 & 21 & 441 & 0 \\ 1 & 27 & 729 & 0 \\ 1 & 32 & 1024 & 0 \\ 1 & 37 & 1369 & 0 \\ 1 & 42 & 1764 & 0 \\ 1 & 47 & 2209 & 0 \\ 1 & 52 & 2704 & 0 \\ 1 & 57 & 3249 & 0 \\ 1 & 62 & 3844 & 0 \end{pmatrix} \quad Y = \begin{pmatrix} 141,080 \\ 205,230 \\ 224,710 \\ 260,170 \\ 273,140 \\ 280,930 \\ 272,880 \\ 267,060 \\ 259,510 \\ 111,540 \\ 145,170 \\ 162,150 \\ 161,990 \\ 157,770 \\ 159,060 \\ 162,170 \\ 157,550 \\ 152,400 \end{pmatrix} \quad b = \begin{pmatrix} -102805.525 \\ 11584.1381 \\ -119.550586 \\ 90545.5556 \end{pmatrix}$$

$$\therefore \hat{y} = \hat{Y} = -102805.52 + 11584.13 * edad - 119.55 * edad^2 + 90545 * Género$$

El coeficiente lo podemos interpretar como la diferencia que gana un hombre o mujer solo por el género

d) **Realice** un análisis de varianza con el modelo del inciso previo. ¿Se podría decir que el coeficiente asociado con la variable género es diferente de cero?

| Values | |
|---|---|
| ebarra | 6863.60482016762 |
| edad | num [1:18] 21 27 32 37 42 47 52 57 62 21 ... |
| eiAnterior | num [1:17] -32613 11703 14654 33586 30027 ... |
| eiNueva | num [1:17] 11703 14654 33586 30027 21289 ... |
| esum | 123544.886763017 |
| FdeR | 3.28738210463651 |
| gen | num [1:18] 1 1 1 1 1 1 1 1 1 0 ... |
| i | 17 |
| ingreso | num [1:18] 141080 205230 224710 260170 273140 ... |
| ingresoBarra | 197472.777777778 |
| n | 18L |
| per | int [1:18] 1 2 3 4 5 6 7 8 9 10 ... |

| Data | |
|---|---|
| b | num [1:2, 1] 3306 104273 |
| CME | num [1, 1] 2.46e+11 |
| CMR | num [1, 1] 1.39e+09 |
| e | num [1:18, 1] -32613 11703 14654 33586 30027 ... |
| Fcalculada | num [1, 1] 176 |
| m | num [1:2, 1:2] 34666 377 377 9 |
| m1 | num [1:2, 1:2] 0.000053 -0.002219 -0.002219 0.20408 |
| m2 | num [1:2, 1] 1.54e+08 2.18e+06 |
| p_value | num [1, 1] 6.41e-12 |
| R | num [1, 1] 0.0276 |
| SCE | num [1, 1] 7.37e+11 |
| SCR | num [1, 1] 2.09e+10 |
| SCT | num [1, 1] 7.57e+11 |
| SCTformula | num [1, 1] 7.57e+11 |
| x | num [1:2, 1:18] 21 1 27 1 32 1 37 1 42 1 ... |
| X | num [1:18, 1:2] 21 27 32 37 42 47 52 57 62 21 ... |
| Yg | num [1:18, 1] 173693 193527 210056 226584 243113 ... |

Tendríamos que restringir el modelo a solo género para verificarlo

e) **Pronostique** el ingreso para un hombre de 33 años y para una mujer de la misma edad.

Con el modelo $\hat{Y} = -102805.52 + 11584.13 * Edad - 119.55 * edad^2 * + 90845 * género$

Hombre 33 (Género = 1)

$\hat{Y} = -102805.52 + 11584.13 * 33 + -119.55 * 1089 + 90845$

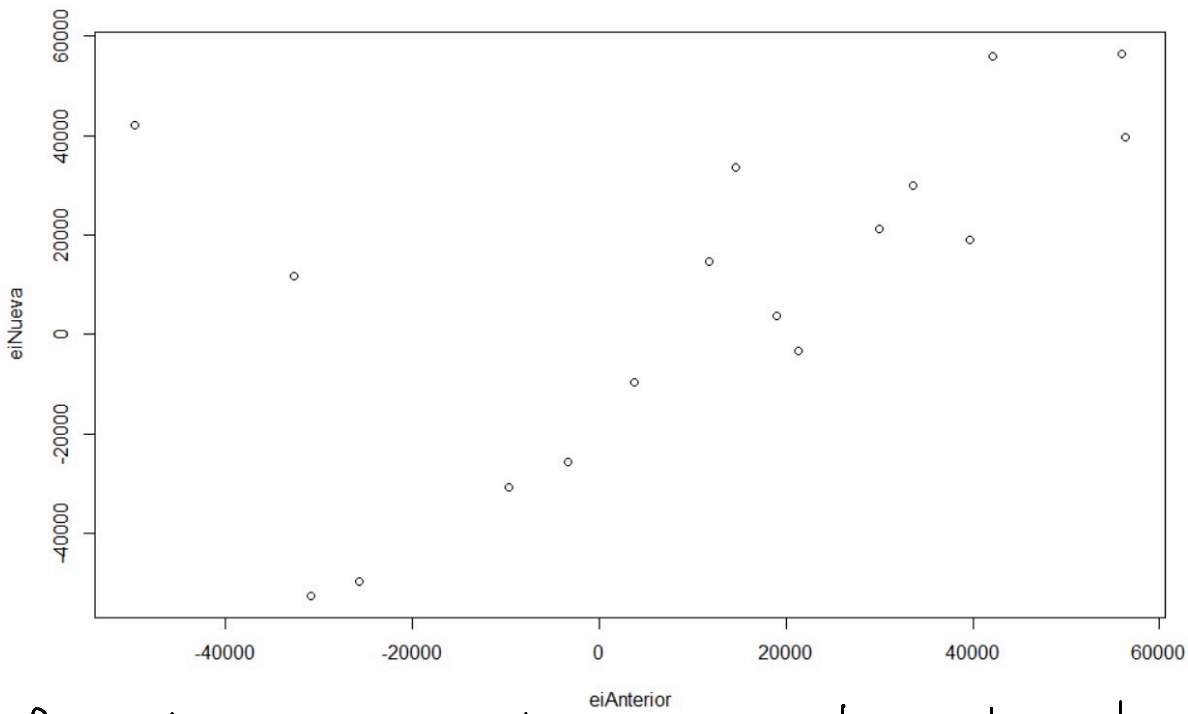
El ingreso estimado será 240,128.82

Mujer 33 (Género = 0)

$\hat{Y} = -102805.52 + 11584.13 * 33 + -119.55 * 1089 + 90845 * 0$

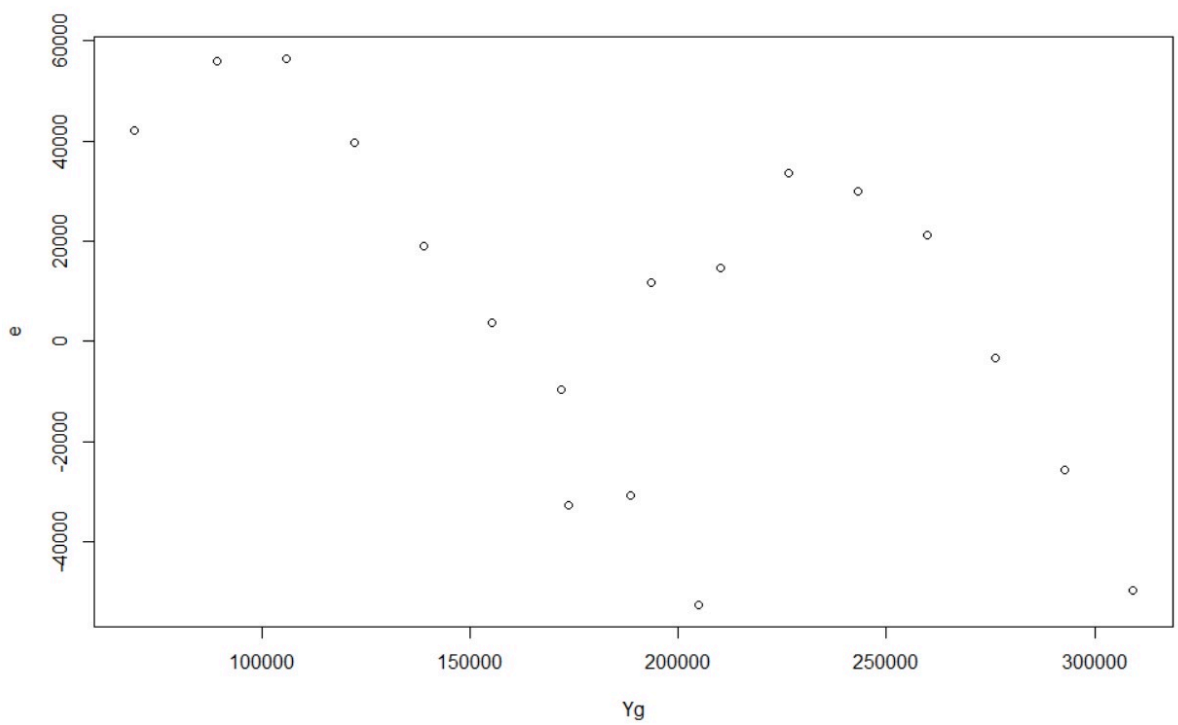
El ingreso estimado será 149283.82

f) **Verifique** todos los supuestos del modelo del inciso c).




Gráfica de $e_i$ con $e_{i+1}$, se ve que hay autocorrelación porque hay una ligera tendencia línea no capturada por el modelo

Son correlacionados en alguna medida porque hay una tendencia lineal que no capturamos




Gráfica de los valores estimados con los errores para verificar si hay varianza constante o no

Vemos que la varianza no es constante