TP2 – Regresión Lineal

Agustina Sol Rojas y Antonio Felix Glorioso Ceretti

Ejercicio 1.

Suponga que (x1. y1), … , (xn, yn) son pares observados generados por los siguientes modelos y deduzca los estimadores de mínimos cuadrados de β1 y β0 .

a) $Y = \beta 1x + \varepsilon$

Se utilizará el método de los mínimos cuadrados:

$$f(\beta 1) = \sum_{i=1}^{n} \varepsilon^2 = \sum_{i=1}^{n} (yi - \beta 1xi)^2$$

$$\frac{\partial f(\beta 1)}{\partial \beta 1} = \sum_{i=1}^{n} 2 * (yi - \beta 1xi) * (-xi) = 0$$

$$= 2 \sum_{i=1}^{n} (-yixi + \beta 1xi^2) = 0$$

$$= \sum_{i=1}^{n} -yixi + \sum_{i=1}^{n} (\beta 1xi^2) = 0$$

$$= -\sum_{i=1}^{n} yixi + \beta 1 \sum_{i=1}^{n} xi^2 = 0$$

$$= \beta 1 \sum_{i=1}^{n} xi^2 = \sum_{i=1}^{n} yixi$$

$$\widehat{\beta 1} = \frac{\sum_{i=1}^{n} yixi}{\sum_{i=1}^{n} xi^2}$$

La recta de regresión estimada es $\hat{y}i = \widehat{\beta 1}x$

b) $Y = \beta 1(ax + c) + \beta 0 + \varepsilon$

$$Y = \beta 1X + \beta 0 + \varepsilon$$

$$\hat{y} = \widehat{\beta 1}x + \widehat{\beta 0}$$

$$y = \beta 1(ax + c) + \beta 0 + \varepsilon$$

$$y = \underbrace{a\beta 1}_{}x + \underbrace{\beta 1c + \beta 0}_{} + \varepsilon$$

$\beta 1'$ $\beta 0'$

$y = \beta 1' x + \beta 0' + \varepsilon$

$$\widehat{\beta 1} = \frac{Sxy}{Syy}$$

$$\widehat{\beta 0} = \hat{y} - \widehat{\beta 1}\bar{x}$$

$\widehat{\beta 1}' = a\widehat{\beta 1}$

$\widehat{\beta 0}' = \widehat{\beta 1}c + \widehat{\beta 0}$

La recta de regresión estimada es $\hat{y}' = \widehat{\beta 1}' x + \widehat{\beta 0}'$

Ejercicio 2.

Una cadena de supermercados financia un estudio sobre los gastos mensuales en alimentos, de familias de 4 miembros. La investigación se limitó a familias con ingresos netos entre $688.000 y $820.000, con lo cual se obtuvo la siguiente recta de estimación $\hat{y} = 0{,}85x - 18.000$.

y = gastos ; x = Ingresos

a) Estime los gastos en alimentos en un mes, para una familia de 4 miembros con un ingreso de $700.000

$\hat{y} = 0{,}85 * 700000 - 18000 = 577000$

b) Uno de los directivos de la compañía se preocupa por el hecho de que la ecuación aparentemente indica que para una familia que tiene un ingreso de $12.000 no gastaría nada en alimentos ¿Cuál sería su respuesta?

No se pueden estimar los gastos porque el ingreso está fuera del rango.

Ejercicio 3.

La empresa META quiere pronosticar el precio de sus acciones en función de los días en el periodo del 03/09/23 al 30/08/24, pero durante las fechas del 02/02/24 al 24/04/24 implementaron una serie de actualizaciones en sus distintas plataformas que dispararon el precio de sus acciones y querían saber en qué porcentaje afectaron dichas actualizaciones al ajuste y a la linealidad.

Utilizando los datos proporcionados en el archivo “META” haga los cálculos necesarios y responda.

Sugerencia: Realice dos análisis diferentes y para una de ellas desestimar los datos del periodo de actualización.

<u>Sin actualización:</u>

Cálculos auxiliares:

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i) * (\sum_{i=1}^{n} y_i)}{n} = 14896028.792634 - \frac{31878 * 106855.010034}{252}$$

$$= 14896028.792634 - 13517158.7693 = 1378870.02333$$

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n} = 5366130 - \frac{1016206884}{252} = 5366130 - 4032567.0$$

$$= 1333563.0$$

$$S_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n} = 47061434.07557989 - \frac{11417993169.366241}{252}$$

$$= 47061434.0755798 - 45309496.70383429 = 1751937.3717456013$$

$$\widehat{\beta_1} = \frac{S_{xy}}{S_{xx}} = \frac{1378870.02333}{1333563.0} = 1.0339744154066948$$

$$\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{x} = 424.0278175952381 - 1.0339744154066948 * 126.5$$

$$= 424.0278175952381 - 130.79776354894688$$

$$= 293.23005404629123$$

$$SS_R = Syy - \widehat{\beta_1} S_{xy} = 1751937.3717456013 - 1.0339744154066948 * 1378870.02333$$

$$= 1751937.3717456013 - 1425716.3262975523 = 326221.045448049$$

Ajuste y linealidad:

$$R^2 = 1 - \frac{SS_R}{Syy} = 1 - \frac{326221.045448049}{1751937.3717456013} = 1 - 0.18620588310357677$$
$$= 0.8137941168964232 * 100 = 81.37941\%$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{1378870.02333}{\sqrt{1333563.0 * 1751937.3717456013}} = \frac{1378870.02333}{\sqrt{2336318857277.179}}$$
$$= \frac{1378870.02333}{1528502.1613583604} = 0.9021053801504696$$

<u>Con actualización:</u>

Cálculos auxiliares:

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i) * (\sum_{i=1}^{n} y_i)}{n} = 1325517.5464050155$$
$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n} = 1312812.9538461538$$
$$S_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n} = 1397233.2607275099$$

$$\widehat{\beta_1} = \frac{S_{xy}}{S_{xx}} = 1.009677382083747$$
$$\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{x} = 278.98103545163457$$

$$SS_R = Syy - \widehat{\beta_1} S_{xy} = 58888.17456722213$$

Ajuste y linealidad:

$$R^2 = 1 - \frac{SS_R}{Syy} = 0.9578537269170359 * 100 = 95.78537\%$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = 0.9787000188602408$$

Dichas actualizaciones afectaron al ajuste un 14.40596% y a la linealidad un 7.65946387%

Ejercicio 4.

Los siguientes datos corresponden a los tiempos relativos en segundos que tardaron en ejecutarse seis programas elegidos al azar en el entorno Windows y en DOS:

a) Realizar el grafico de dispersión de los puntos

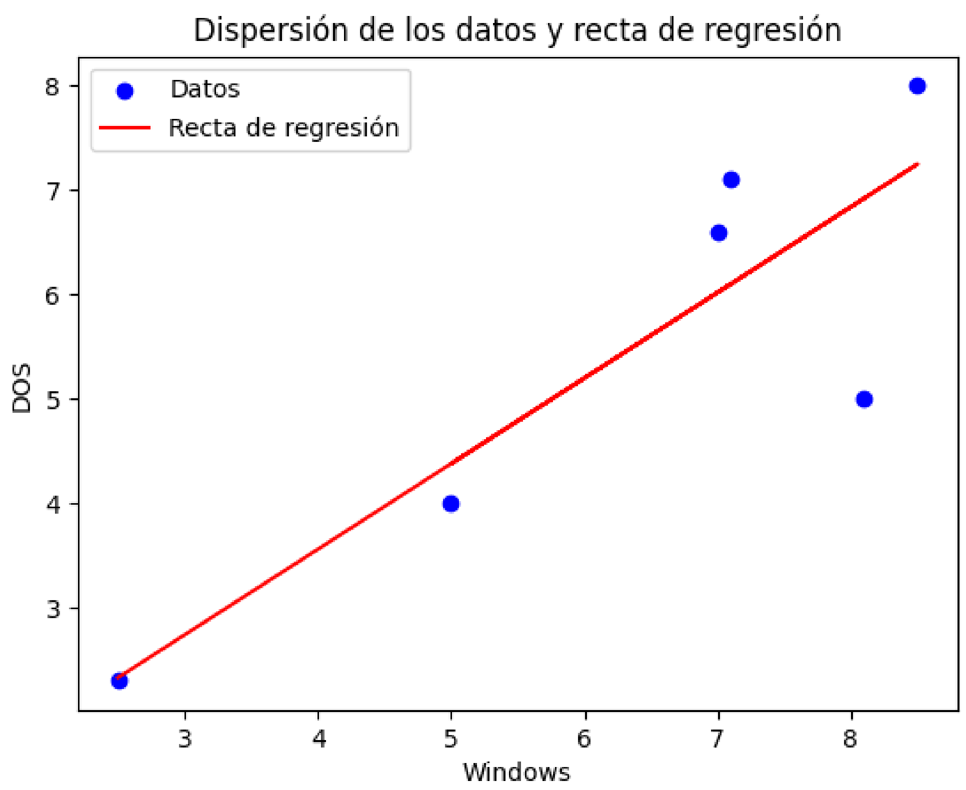
$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i) * (\sum_{i=1}^{n} y_i)}{n} = 20.759999999999962$$

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n} = 25.313333333333276$$

$$S_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n} = 22.75999999999999$$

$$\widehat{\beta_1} = \frac{S_{xy}}{S_{xx}} = \frac{20.759999999999962}{25.313333333333276} = 0.8201211482749543$$

$$\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{x} = 5.5 - 0.8201211482749543 * 6.3666 = 5.5 - 5.22138330261$$
$$= 0.2785620226494574$$



b) Si un programa tarda 6 segundos en ejecutarse en Windows, ¿Cuánto tardara en ejecutarse en DOS?

c) Se estima que los tiempos de Windows mejoraran reduciéndose en un 10% en los próximos años, estime la recta de regresión considerando esta mejora. Suponga que los tiempos DOS no se modifican.