Regresión y Correlación Lineal Simple Parte 2

Resolución Matricial del modelo de regresión

Dados los valores de las variables independiente $\boldsymbol{x_i}$ y dependiente $\boldsymbol{y_i}$.

$\boldsymbol{x_i}$	$\boldsymbol{y_i}$
x_1	y_1
x_2	y_2
x_3	y_3
$\vdots$	$\vdots$
x_n	y_n

$$B = (X'X)^{-1}(X'Y)$$

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

$$X' = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ x_1 & x_2 & x_3 & \dots & x_n \end{bmatrix}$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}$$

$$B = \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \end{bmatrix}$$

$$(X'X) = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ x_1 & x_2 & x_3 & \dots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ x_1 & x_2 & x_3 & \dots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix}$$

Sistema de Ecuaciones Normales de Gauss

$$n\ \hat{b}_0 + \hat{b}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$\hat{b}_0 \sum_{i=1}^{n} x_i + \hat{b}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i\, y_i$$

Dado que X’X = (X’X)’. Se trasponen las filas por las columnas

$$(X'X) = \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix}$$

$$(X'X)' = \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix}$$

Calculamos el determinante de X’X

$$Determinante de X'X = |X'X| = \begin{vmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{vmatrix} = n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2$$

Hallamos la inversa de X’X

$$(X'X)^{-1} = \frac{1}{|X'X|} adj(X'X)$$

Matriz adjunta

En primer lugar se tiene que aplicar la regla de signos de los adjuntos de la siguiente manera:

Si el número de la fila más el número de la columna da número par queda el signo del adjunto, si el número de la fila más el número de la columna da número impar cambia el signo del adjunto.

$$\begin{array}{ccc} & col\,1 & col\,2 \\ fila\,1 & + & (-1) \\ fila\,2 & (-1) & + \end{array}$$

Dada la matriz (X’X) para hallar la matriz adjunta se procede de la siguiente manera:

Se selecciona el elemento de la fila 1 columna 1, se tapa la fila y la columna y el elemento que queda visible es el adjunto de ese elemento adj(1,1).

$$(X'X) = \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix}$$

$$(X'X) = \begin{bmatrix} & \\ & \sum_{i=1}^{n} x_i^2 \end{bmatrix}$$

$$Adj(X'X) = \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & \\ & \end{bmatrix}$$

Se selecciona el elemento de la fila 1 columna 2 y se tapa la fila y la columna, lo que queda visible es el adjunto del elemento adj(1,2)

$$(X'X) = \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix}$$

$$(X'X) = \begin{bmatrix} & \\ \sum_{i=1}^{n} x_i & \end{bmatrix}$$

$$Adj(X'X) = \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ & \end{bmatrix}$$

Se selecciona el elemento de la fila 2 columna 1 y se tapa la fila y la columna, lo que queda visible es el adjunto del elemento adj(2,1)

$$(X'X) = \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix}$$

$$(X'X) = \begin{bmatrix} & \sum_{i=1}^{n} x_i \\ & \end{bmatrix}$$

$$Adj(X'X) = \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & \end{bmatrix}$$

Se selecciona el elemento de la fila 2 columna 2 y se tapa la fila y la columna, lo que queda visible es el adjunto del elemento adj(2,2)

$$(X'X) = \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix}$$

$$(\mathrm{X'X}) = \begin{bmatrix} \mathrm{n} & \\ & \end{bmatrix}$$

$$Adj(X'X) = \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & n \end{bmatrix}$$

$$(X'X)^{-1} = \frac{1}{|X'X|} adj(X'X)$$

$$(X'X)^{-1} = \frac{1}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & n \end{bmatrix}$$

Se multiplica la matriz adjunta (X'X) por el vector (X'Y) y luego se divide por el determinante:

$$B = (X'X)^{-1}(X'Y) = \frac{\begin{bmatrix} \sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & n \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix}}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \end{bmatrix}$$

De este modo se obtienen los estimadores del modelo de regresión lineal simple:

$$\hat{b}_0 = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i + \left(-\sum_{i=1}^{n} x_i\right) \sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$\hat{b}_1 = \frac{\left(-\sum_{i=1}^{n} x_i\right) \sum_{i=1}^{n} y_i + n\sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

Ecuación de la recta de regresión:

$$\hat{y}_i = \hat{b}_0 + \hat{b}_1 x_i$$

Bondad de Ajuste del modelo de Regresión Lineal Simple

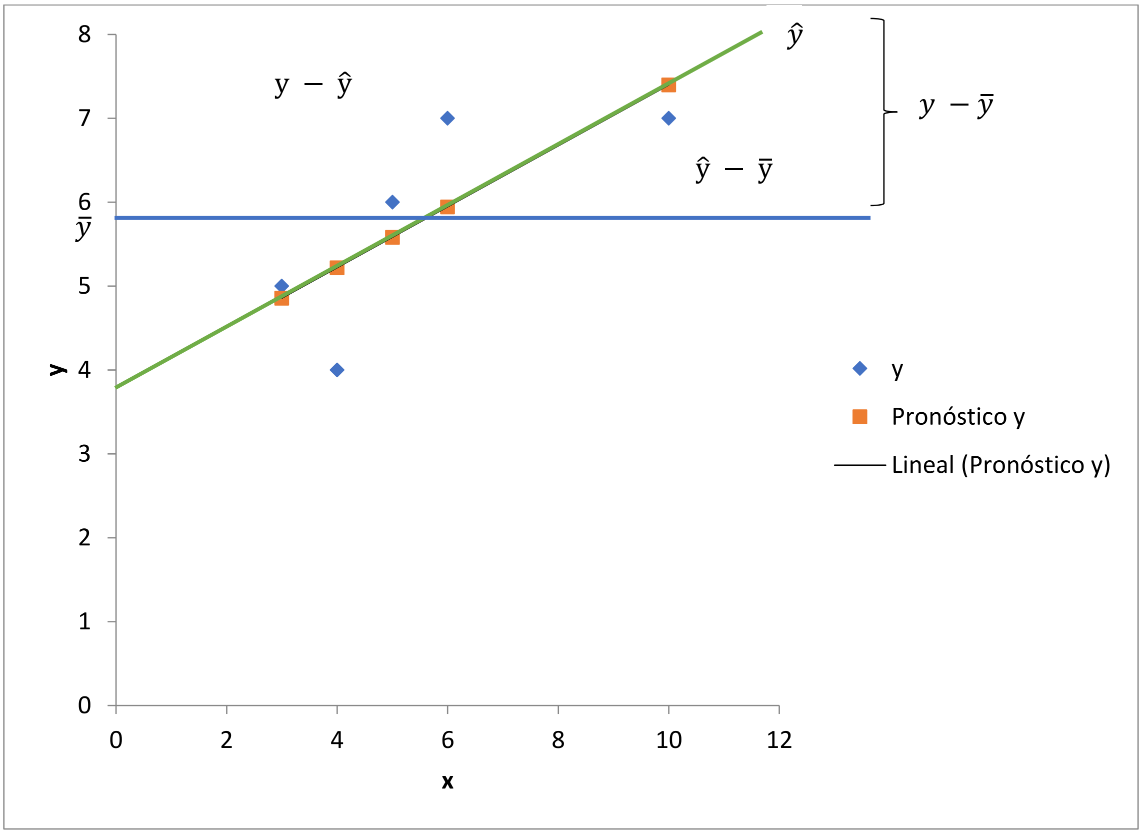
Medidas de la bondad del ajuste

Medida absoluta

Error estándar de la estimación

Medida relativa

Coeficiente de determinación



$y - \bar{y}$ = desviación total

$y - \hat{y}$ = desviación no explicada

$\hat{y} - \bar{y}$ = desviación explicada

Si se elevan al cuadrado y se suma se obtiene lo siguiente:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = Variación\ total = SCT = VT$$

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = Variación\ no\ explicada = SCE = VNE$$

$$\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = Varianción\ explicada = SCR = VE$$

Variación total = variación no explicada + variación explicada

Suma de cuadrados total = Suma de cuadrados del error + Suma de cuadrados de la regresión

Demostración:

Sumando y restando $\hat{y}_i$, luego aplicando el cuadrado del binomio queda:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \bar{y} + \hat{y}_i - \hat{y}_i)^2 = \sum_{i=1}^{n}\left((y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})\right)^2$$

$$= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + 2\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

$$= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

$$\boxed{\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}$$

$$SCT = SCE + SCR$$

$$VT = VNE + VE$$

$$\frac{SCT}{SCT} = \frac{SCE}{SCT} + \frac{SCR}{SCT}$$

$$1 = \frac{SCE}{SCT} + \frac{SCR}{SCT}$$

Medida absoluta de la bondad del ajuste. Error estándar de la estimación

$$S_e = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k-1}} = \sqrt{\frac{SCE}{n-k-1}}$$

Medida relativa de la bondad del ajuste. Coeficiente de determinación

$$R^2 = \frac{SCR}{SCT} = \frac{\text{Varianción explicada}}{Variación\ total}$$

$R^2 =$ es la proporción de la variación total explicada por la recta de regresión

$$R^2 100 = \_\_\_\%\ de\ la\ variación\ total\ explicado\ por\ la\ recta\ de\ regresión$$

¿Cuál es el rango de variabilidad de R^2?

$$\boxed{0 \leq R^2 \leq 1}$$

Otras medidas

Varianza de la estimación

$$S_e^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k-1} = \frac{SCE}{n-k-1}$$

Coeficiente de no determinación

$$1 - R^2 = 1 - \frac{SCR}{SCT}$$

Es la proporción de la variación total que no está explicada por la recta de regresión.

$(1 - R^2)100 =$ ___% es el porcentaje de la variación total que no está explicada por la recta de regresión.

Coeficiente de correlación lineal.

Mide el grado de asociación lineal entre las variables x e y.

$$R = \sqrt{\frac{SCR}{SCT}}$$

El signo es el mismo que la pendiente de la recta.

Otras formas de calcular R

$$R = \frac{Cov(x,y)}{D(x)D(y)} = \frac{\dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}}{\sqrt{\sum_{i=1}^{n}\dfrac{(x_i - \bar{x})^2}{n-1}}\sqrt{\sum_{i=1}^{n}\dfrac{(y_i - \bar{y})^2}{n-1}}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

$$R = \frac{Cov(x,y)}{D(x)D(y)} = \frac{n\sum_{i=1}^{n}x_i y_i - \sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{\sqrt{n\sum_{i=1}^{n}x_i^2 - \left(\sum_{i=1}^{n}x_i\right)^2}\sqrt{n\sum_{i=1}^{n}y_i^2 - \left(\sum_{i=1}^{n}y_i\right)^2}}$$

¿Qué valores puede tomar R?

$$\boxed{-1 \le R \le 1}$$

Interpretación de los valores de R.