

Regresión por mínimos cuadrados

Regresión Lineal

Prof. Lic. Mariel Ugarte

Análisis Numérico - Com 302 - 303

FRRO - UTN - 2023

REGRESIÓN POR MÍNIMOS CUADRADOS

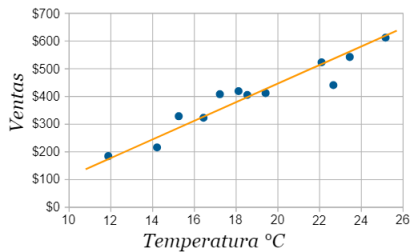
REGRESIÓN LINEAL

Dados un conjunto de pares ordenados (puntos del plano) y una familia de funciones, se intenta encontrar la función continua, dentro de dicha familia, que mejor se aproxime a los datos (es decir, que mejor **se ajuste** a la forma o a la tendencia general de los datos, sin coincidir necesariamente en todos los puntos), de acuerdo con algún criterio.

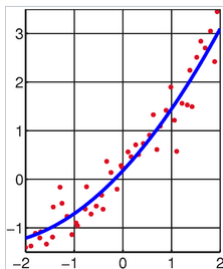
Si dicho criterio es el del *mínimo error cuadrático*, hablaremos de **regresión por mínimos cuadrados**.

Si la familia de funciones son las *funciones lineales*, se llamará **regresión lineal (por mínimos cuadrados)**.

Regresión lineal



Regresión cuadrática



REGRESIÓN LINEAL

Sean los puntos del plano

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

Buscamos la recta

$$y = a_0 + a_1x$$

que mejor “ajuste” dichos puntos.

¿¿según qué criterio??

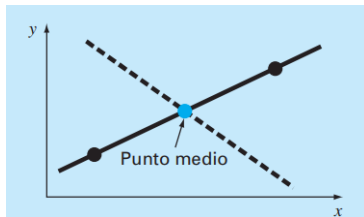
POSIBLES CRITERIOS

Opciones:

- *Primera opción:* Minimizar la suma de los errores residuales de todos los datos disponibles, es decir, minimizar

$$\sum_{i=1}^n e_i = \sum_{i=1}^n (y_i - a_0 - a_1 x_i)$$

Problema:



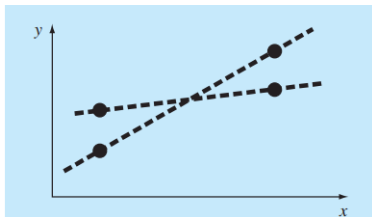
Con dos puntos, cualquier recta que pase por el punto medio de éstos, hace mínima e igual a 0 la suma.

POSIBLES CRITERIOS

- *Segunda opción:* Minimizar la suma de los valores absolutos de las discrepancias, es decir, minimizar

$$\sum_{i=1}^n |e_i| = \sum_{i=1}^n |y_i - a_0 - a_1 x_i|$$

Problema:



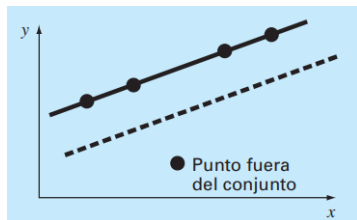
Con cuatro puntos, cualquier línea recta que esté dentro de las líneas punteadas minimizará el valor absoluto de la suma.

POSIBLES CRITERIOS

- *Tercera opción:* (Criterio *minimax*) Minimizar la máxima distancia a que un punto se encuentra de la línea, es decir, minimizar

$$\max \{|e_i|, i = 1 \dots n\} = \max \{|y_i - a_0 - a_1 x_i|, i = 1 \dots n\}$$

Problema:



Se da excesiva influencia a puntos fuera del conjunto; es decir, a un solo punto con un gran error.

POSIBLES CRITERIOS

- *Cuarta opción:* Minimizar la suma de los cuadrados de los residuos, es decir, minimizar

$$\sum_{i=1}^n (e_i)^2 = \sum_{i=1}^n (y_i - a_0 - a_1 x_i)^2$$

¿¿ a_0, a_1 ??

AJUSTE LINEAL POR MÍNIMOS CUADRADOS

Sea

$$S(a_0, a_1) = \sum_{i=1}^n (y_i - a_0 - a_1 x_i)^2$$

Para hallar a_0 y a_1 que minimicen $S(a_0, a_1)$, igualamos a 0 sus derivadas parciales:

$$\frac{\partial S}{\partial a_0} = -2 \sum_{i=1}^n (y_i - a_0 - a_1 x_i) = 0$$

$$\frac{\partial S}{\partial a_1} = -2 \sum_{i=1}^n (y_i - a_0 - a_1 x_i) x_i = 0$$

Luego:

$$\sum_{i=1}^n y_i - \sum_{i=1}^n a_0 - \sum_{i=1}^n a_1 x_i = 0 \quad \Longrightarrow \quad n a_0 + \left(\sum_{i=1}^n x_i \right) a_1 = \sum_{i=1}^n y_i$$

$$\sum_{i=1}^n y_i x_i - \sum_{i=1}^n a_0 x_i - \sum_{i=1}^n a_1 x_i^2 = 0 \quad \Longrightarrow \quad \left(\sum_{i=1}^n x_i \right) a_0 + \left(\sum_{i=1}^n x_i^2 \right) a_1 = \sum_{i=1}^n x_i y_i$$

AJUSTE LINEAL POR MÍNIMOS CUADRADOS

$$na_0 + \left(\sum_{i=1}^n x_i \right) a_1 = \sum_{i=1}^n y_i$$

$$\left(\sum_{i=1}^n x_i \right) a_0 + \left(\sum_{i=1}^n x_i^2 \right) a_1 = \sum_{i=1}^n x_i y_i$$

Entonces

$$a_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

$$a_0 = \frac{\sum y_i}{n} - a_1 \frac{\sum x_i}{n} = \bar{y} - a_1 \bar{x}$$

EJEMPLO

Ajuste a una línea recta los valores x y y de la tabla

x_i	y_i
1	0,5
2	2,5
3	2,0
4	4,0
5	3,5
6	6,0
7	5,5

$$n = 7 \quad \sum x_i y_i = 119,5 \quad \sum x_i^2 = 140$$

$$\sum x_i = 28 \quad \bar{x} = \frac{28}{7} = 4$$

$$\sum y_i = 24 \quad \bar{y} = \frac{24}{7} = 3,428571$$

EJEMPLO

Luego

$$a_1 = \frac{7 \cdot 119,5 - 28 \cdot 24}{7 \cdot 140 - 28^2} = 0,8392857$$

$$a_0 = 3,428571 - 0,8392857 \cdot 4 = 0,07142857$$

Por lo tanto el ajuste lineal por mínimos cuadrados es

$$y = 0,07142857 + 0,8392857x$$

x_i	y_i	$(y_i - \bar{y})^2$	$(y_i - a_0 - a_1 x_i)^2$
1	0.5	8.5765	0.1687
2	2.5	0.8622	0.5625
3	2.0	2.0408	0.3473
4	4.0	0.3265	0.3265
5	3.5	0.0051	0.5896
6	6.0	6.6122	0.7972
7	<u>5.5</u>	<u>4.2908</u>	<u>0.1993</u>
Σ	24.0	22.7143	2.9911

EJEMPLO

