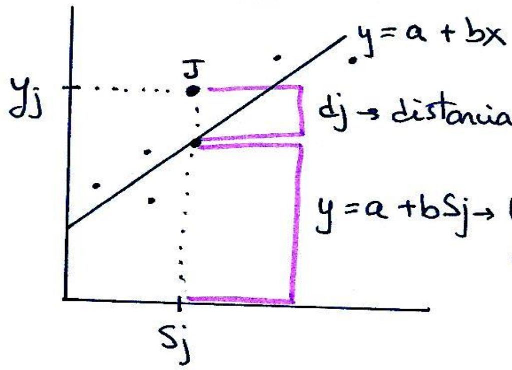
AJUSTE POR MÍNIMOS CUADRADOS

Recta de regresión



dj → distancia J a recta.

y = a + bSj → la coordenada y de la recta

$$dj = yj - (a + bSj)$$

Queremos hacer mínimo el error.

$$\boxed{f(a,b) = \sum_{j=1}^{n} dj^2 = \sum_{j=1}^{n} \left(yj - (a+bSj)\right)^2}$$

FUNCIÓN A MINIMIZAR (el error^2)

Nota: El objetivo es encontrar una recta de la forma $y = a + bx$ que minimice los errores respecto a los puntos dato que tengo.

→ "a" y "b" son las variables o incógnitas a despejar.
Como queremos minimizar hay que hallar la 1ª derivada.
Al tener dos variables empleamos las derivadas parciales que luego igualaremos a 0.

$$\frac{\partial f}{\partial a} = 2\sum_{j=1}^{n} \left(yj - (a+bSj)\right)(-1)$$

$$\frac{\partial f}{\partial b} = 2\sum_{j=1}^{n} \left(yj - (a+bSj)\right)(-Sj)$$

En las derivadas parciales hay que tener en cuenta que lo que no contiene la variable parcial se considera constante.

Ahora desarrollamos e igualamos a 0 (para minimizar).
(operamos)

Para ello tenemos en mente que la a = cte y por ello su Σ será n·a (a multiplicado n número de veces).

$$(1) \rightarrow \quad -\sum_{j=1}^{n} y_j + \sum_{j=1}^{n} a + \sum_{j=1}^{n} b \cdot s_j = 0$$

$$(2) \rightarrow \quad -\sum_{j=1}^{n} s_j y_j + a\sum_{j=1}^{n} s_j + b\sum_{j=1}^{n} s_j^2 = 0$$

Pasamos el término del $\sum y_j$ a la derecha para obtener unas ecuaciones con forma de recta.

$$\left.\begin{aligned} (1) &\rightarrow \quad n \cdot a + b\sum_{j=1}^{n} s_j = \sum_{j=1}^{n} y_j \\ (2) &\rightarrow \quad a\sum_{j=1}^{n} s_j + b\sum_{j=1}^{n} s_j^2 = \sum_{j=1}^{n} s_j y_j \end{aligned}\right\}$$

Planteo un sistema de ecuaciones de forma matricial con estas 2 expresiones para hallar "a" y "b".

Sistema

CLAVE

$$\boxed{\begin{pmatrix} n & S_x \\ S_x & S_{x2} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} S_y \\ S_{xy} \end{pmatrix}} \text{ donde} \rightarrow \boxed{\begin{aligned} S_x &= \sum_{j=1}^{n} s_j \\ S_{x2} &= \sum_{j=1}^{n} s_j^2 \\ S_y &= \sum_{j=1}^{n} y_j \\ S_{xy} &= \sum_{j=1}^{n} s_j y_j \end{aligned}}$$

CLAVE

Recordemos que s_j e y_j son las coordenadas "x" e "y" respectivamente de los vectores dato genéricos.

Resolvemos por Cramer $\longrightarrow$ p. ejemplo.

$$a = \frac{\begin{vmatrix} S_y & S_x \\ S_{xy} & S_{x2} \end{vmatrix}}{n S_{x2} - (S_x)^2} = \boxed{\frac{S_{x2} \cdot S_y - S_{xy} \cdot S_x}{n S_{x2} - (S_x)^2}} \rightarrow a$$

CLAVE

$$b = \frac{\begin{vmatrix} n & S_y \\ S_x & S_y \end{vmatrix}}{n S_{x2} - (S_x)^2} = \boxed{\frac{n S_{xy} - S_x S_y}{n S_{x2} - (S_x)^2}} \rightarrow b$$

La recta de regresión entonces es $y = a + bx$.

Si consideramos los valores en forma de Σ de Sx, Sx2, Sy...
obtenemos que:

$$a = \frac{\left(\sum_{j=1}^{n} s_j^2 \cdot \sum_{j=1}^{n} y_j\right) - \left(\sum_{j=1}^{n} s_j y_j \cdot \sum_{j=1}^{n} s_j\right)}{D} \qquad D = n \sum_{j=1}^{n} s_j^2 - \left(\sum_{j=1}^{n} s_j\right)^2$$

$$b = \frac{n \sum_{j=1}^{n} s_j y_j - \sum_{j=1}^{n} s_j \cdot \sum_{j=1}^{n} y_j}{D}.$$

Realmente hemos sustituido únicamente.

Acordarse del valor en Σ de Sx, Sx2, Sy... es muy fácil, pues todos ellos son $\sum_{j=1}^{n}$ y lo que va dentro del sumatorio tiene relación directa con la variable que le asignemos (Sx, Sy...).

$$x \rightarrow s_j \qquad x2 \rightarrow s_j^2$$
$$y \rightarrow y_j \qquad xy \rightarrow s_j y_j$$

→ Recordando que s_j e y_j son las coordenadas x, y del punto dato respectivamente.

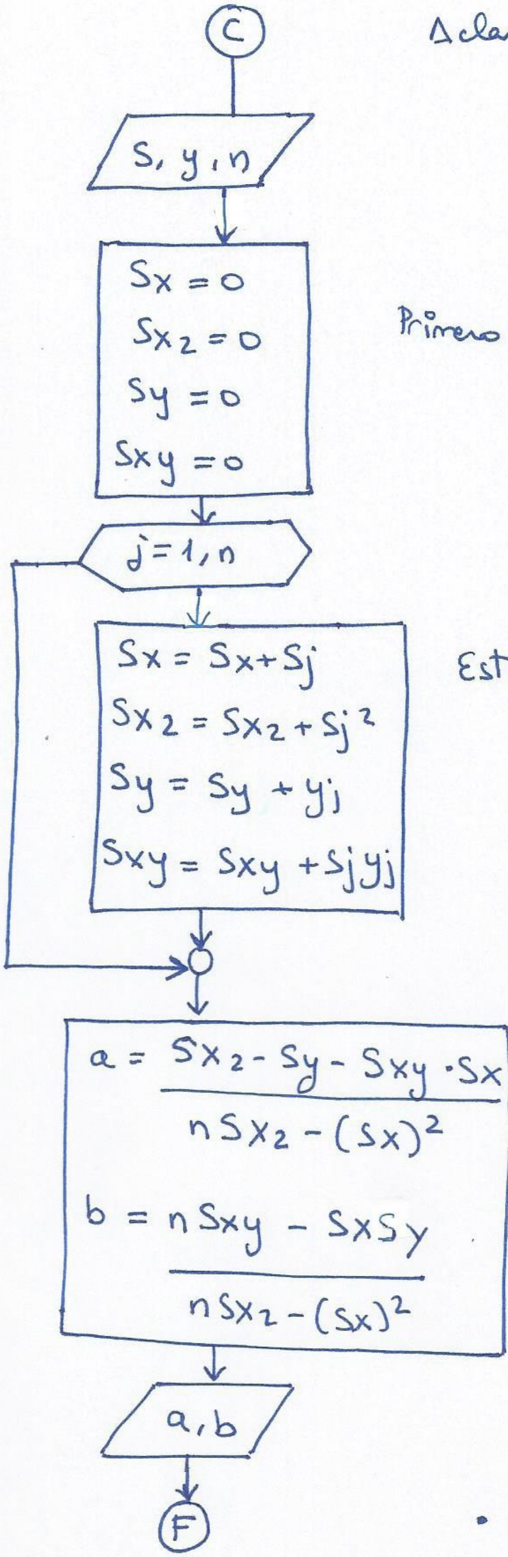
ALGORITMO (organigrama o Pseudocódigo)

Datos: vectores s, y, n

Resultado: a, b

Calcular: Sx, Sx2, Sy, Sxy.

Aclaración: Son todas letras "S", no hay ningún 5. número.



Primero inicializamos todos los sumatorios a 0.

Esto es simplemente la fórmula del sumatorio.

Con los datos obtenemos "a" y "b".

NOTA

- Una vez entendido el procedimiento podemos quedarnos con lo importante.

Con las fórmulas remarcadas y señaladas como "CLAVE" tendremos suficiente para hallar la recta de regresión.

Ahora aplicaremos la teoría y todas esas variables que pueden parecer un lío a un ejercicio con datos numéricos:

$\boxed{\text{EJERCICIO}}$ → Calcule la recta de regresión que minimice el error.

Pts dato

(0'1 , -1)
(0'8 , 0'95)
(1'2 , 1'8)
(1'2 , 1'9)
(1'7 , 2'1)
(2'5 , 3'6)

Las fórmulas a tener en cuenta son:

①

$Sx = \sum_{j=1}^{n} s_j$	$Sy = \sum_{j=1}^{n} y_j$
$Sx_2 = \sum_{j=1}^{n} s_j^2$	$Sxy = \sum_{j=1}^{n} s_j y_j$

②

$$a = \frac{(Sx_2 \cdot Sy) - (Sxy \cdot Sx)}{nSx_2 - (Sx)^2}$$

$$b = \frac{(nSxy) - (SxSy)}{nSx_2 - (Sx)^2}$$

Hallo ①

	s_j	s_j^2	y_j	$s_j y_j$
	0'1	0'01	-1	-0'1
	0'8	0'64	0'95	0'76
	1'2	1'44	1'8	2'16
	1'2	1'44	1'9	2'28
	1'7	2'29	2'1	3'57
	2'5	6'25	3'6	9'00
$\sum_{j=1}^{n}$	7'5	12'67	9'35	17'67
	Σs_j	Σs_j^2	Σy_j	$\Sigma s_j y_j$

Sustituyo en "a" y "b".

↓

Si uno no quiere aprenderse la fórmula de "a" y "b" es más sencillo deducirla del sistema en forma de matriz o sencillamente sustituir lo calculado en el sistema matricial. de la siguiente forma:

Se recomienda operar con la coma de los decimales arriba, pues se puede confundir con la coma que separa la coordenada de los puntos.

← Solo tenemos que sumar las columnas.

s_j → coordenadas x.

s_j^2 → coordenadas x al cuadrado.

y_j → coordenadas y.

$s_j y_j$ → coord. x por coord. y.

$$\begin{pmatrix} n & Sx \\ Sx & Sx2 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} Sy \\ Sxy \end{pmatrix} \rightarrow$$ se sustituyen $Sx, Sx2, Sy, Sxy$.

$\downarrow$

llego a este sistema.

$$\begin{cases} 6a + 7'5b = 9'35 \\ 7'5a + 12'67b = 17'67 \end{cases}$$

$$\boxed{r(x) = -0{,}7112038 + 1{,}815629x}$$

recta solución.

$$\boxed{\begin{array}{l} a = -0{,}7112038\ldots \\ b = 1{,}815629\ldots \end{array}}$$