Estadística (3)

Jesús García de Jalón de la Fuente

IES Ramiro de Maeztu
Madrid

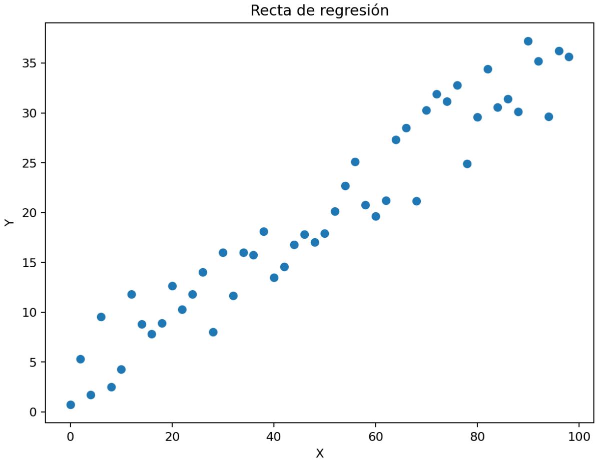
2020

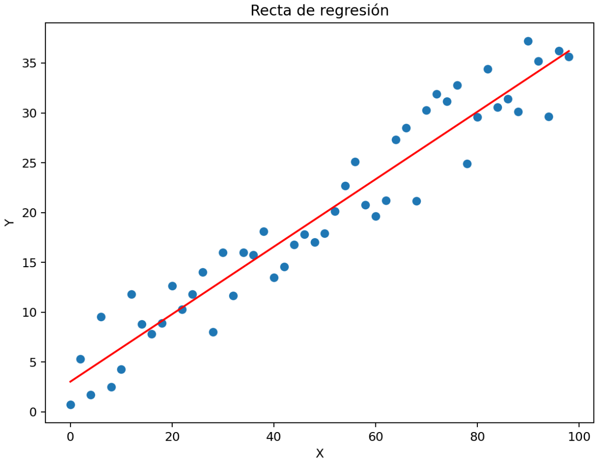
Recta de regresión

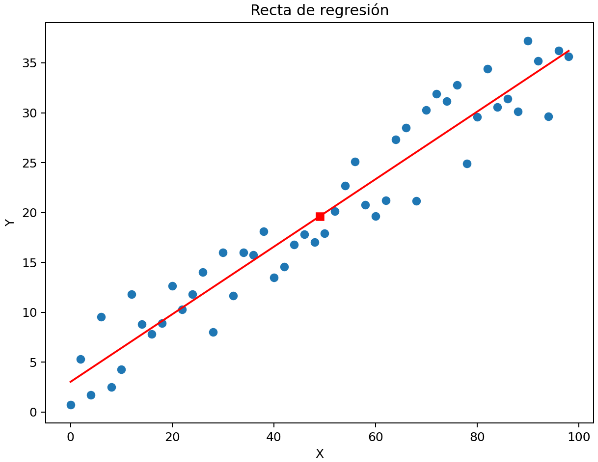
Supongamos ahora que dos variables X e Y están linealmente correlacionadas.

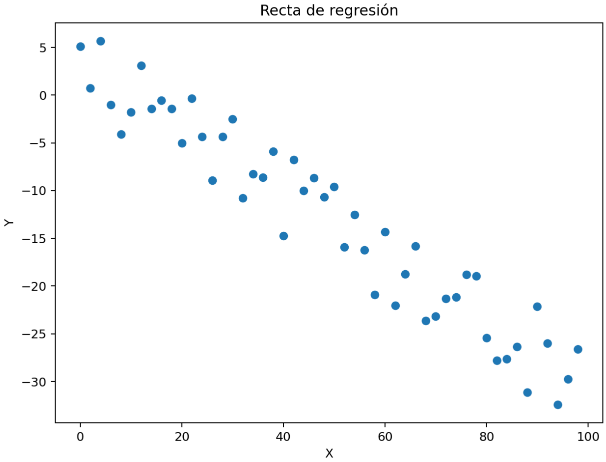
Puesto que la relación entre ambas variables tiene un carácter estadístico, si para un dato la primera variable toma un valor x no podemos conocer a partir de él el correspondiente valor de la otra variable, pero sí puede calcularse un valor probable o aproximado.

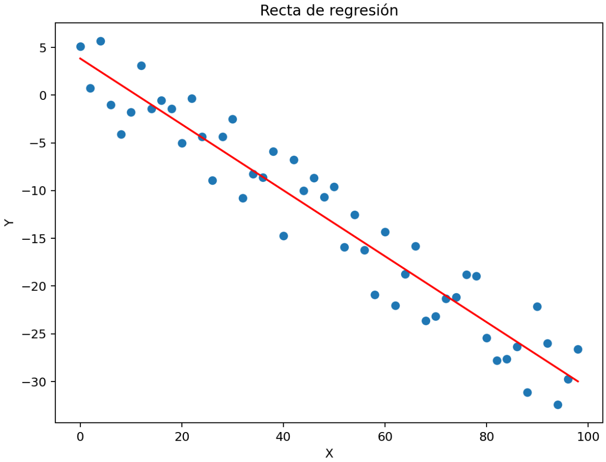
Se llama **recta de regresión** a la recta que mejor se ajusta a la nube de puntos.

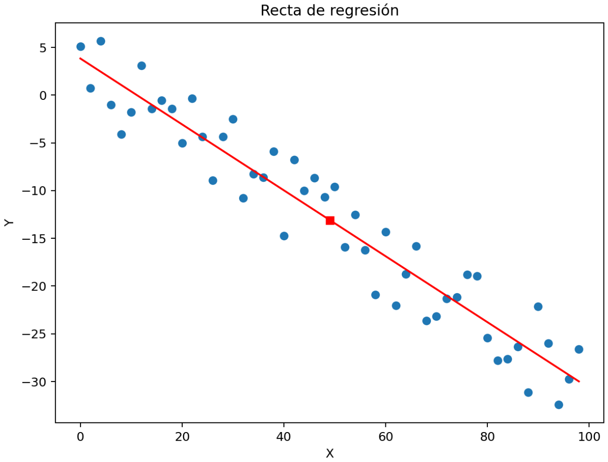

Recta de regresión
Y
X
0
5
10
15
20
25
30
35
0
20
40
60
80
100


Recta de regresión
35
30
25
20
15
10
5
0
Y
0
20
40
60
80
100
X


Recta de regresión
Y
X
0
5
10
15
20
25
30
35
0
20
40
60
80
100


Recta de regresión
Y
X
5
0
−5
−10
−15
−20
−25
−30
0
20
40
60
80
100


Recta de regresión
5
0
−5
−10
−15
−20
−25
−30
Y
0
20
40
60
80
100
X


Recta de regresión
Y
X
5
0
−5
−10
−15
−20
−25
−30
0
20
40
60
80
100

El criterio de ajuste de la recta a la nube es que la cantidad

$$S = \sum_{i=1}^{N} (y_i - mx_i - b)^2$$

sea mínima.

Es decir debe ser mínima la suma de los cuadrados de las diferencias entre los valores de la variable Y y las ordenadas de los puntos correspondientes de la recta.

Ecuación de la recta de regresión

Sea $y = mx + b$ la ecuación de esta recta.

La recta de regresión pasa por el punto $(\bar{x}, \bar{y})$ y tiene como pendiente el cociente de la covarianza y la varianza.

Su ecuación es: $$y - \bar{y} = \frac{\sigma_{xy}}{\sigma_x^2}(x - \bar{x})$$

Esta recta es la **recta de regresión de y sobre x**. Mediante esta ecuación, dado un valor para x podemos calcular el valor probable de y.

Si queremos calcular el valor probable de x conocido el de y basta intercambiar todas las x e y en la igualdad anterior. La **recta de regresión de x sobre y** es:

$$x - \bar{x} = \frac{\sigma_{xy}}{\sigma_y^2}(y - \bar{y})$$

Ejemplo

Con los siguientes datos

x	4	16	18	9	15	13	4	2	13	6
y	3	18	16	7	17	14	6	4	11	7

calcular la recta de regresión para estimar el valor de y correspondiente a $x = 3$. Aproximar al entero más cercano.

Debemos calcular la ecuación de la recta de regresión de y sobre x.

Calculamos en primer lugar las media de las variables:

$$\bar{x} = \frac{1}{10}(4 + 16 + 18 + 9 + 15 + 13 + 4 + 2 + 13 + 6) = \frac{100}{10} = 10$$

$$\bar{y} = \frac{1}{10}(3 + 18 + 16 + 7 + 17 + 14 + 6 + 4 + 11 + 7) = \frac{103}{10} = 10{,}3$$

x	4	16	18	9	15	13	4	2	13	6
y	3	18	16	7	17	14	6	4	11	7

La suma de los productos es:

$$\sum x_i y_i = 4\cdot 3 + 16\cdot 18 + 18\cdot 16 + 9\cdot 7 + 15\cdot 17 + 13\cdot 14 + 4\cdot 6 + 2\cdot 4 + 13\cdot 11 + 6\cdot 7 = 1305$$

Y la covarianza:

$$\sigma_{xy} = \frac{1305}{10} - 10\cdot 10{,}3 = 27{,}5$$

Calculamos ahora la varianza de x:

$$\sigma_x^2 = \frac{1}{10}\left(4^2 + 16^2 + 18^2 + 9^2 + 15^2 + 13^2 + 4^2 + 2^2 + 13^2 + 6^2\right) - 10^2 = 129{,}6 - 10^2 = 29{,}6$$

La pendiente de esta recta es:

$$m = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{27{,}5}{29{,}6} \simeq 0{,}93$$

Puesto que la recta pasa por $(\bar{x}, \bar{y})$, su ecuación es:

$$y - 10{,}3 = 0{,}93\,(x - 10)\ ; \qquad y = 0{,}93x + 1$$

Con la aproximación pedida resulta que para $x = 3$, $y = 4$.

♠♠♠♠

Agradecimiento

Gracias por vuestra atención