Estadística (2)

Jesús García de Jalón de la Fuente

IES Ramiro de Maeztu
Madrid

2020

Supongamos que sobre una población se miden dos magnitudes X e Y.

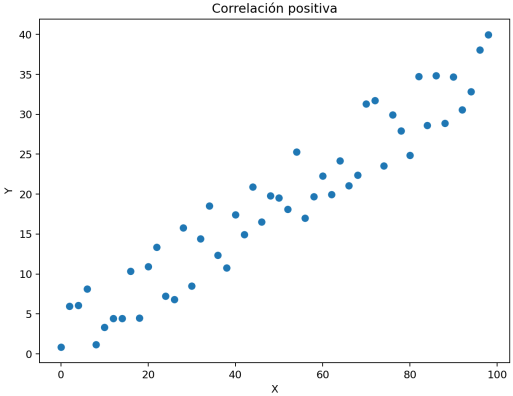
Para cada elemento de la población se obtienen un par de valores (x, y).

Considerando estos pares de números como coordenadas, podemos representar los puntos correspondientes en unos ejes de coordenadas.

Se obtiene un diagrama que se llama **nube de puntos**.

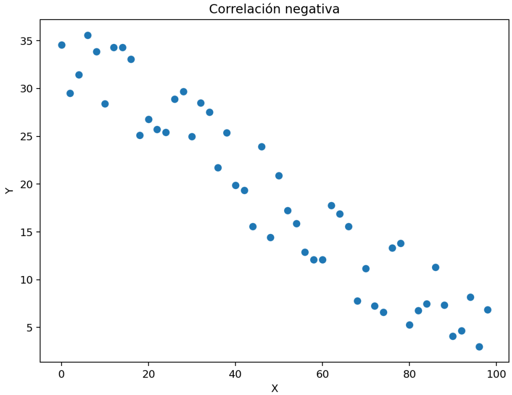
Veamos diferentes formas que puede tener la nube de puntos.

Correlación positiva



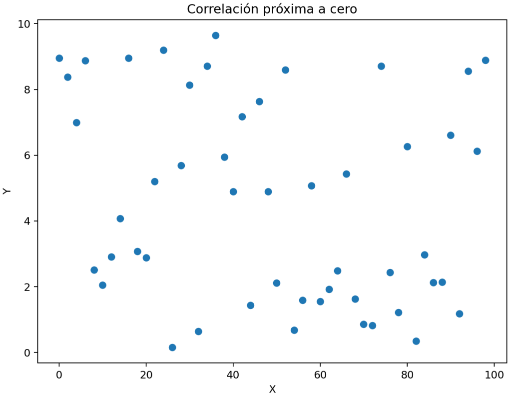
No podemos afirmar que cuando X aumenta necesariamente aumenta Y, pero sí que cuando X aumente es más probable que Y aumente. Entre X e Y existe una correlación positiva.

Correlación negativa



No podemos afirmar que cuando X aumenta necesariamente disminuye Y, pero sí que cuando X aumente es más probable que Y disminuya. Entre X e Y existe una correlación negativa.

Correlación cero



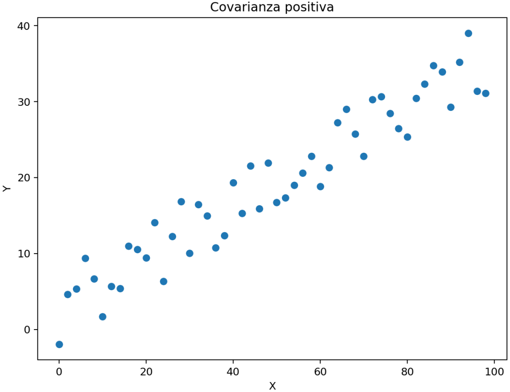
No existe correlación entre las variables.

Covarianza

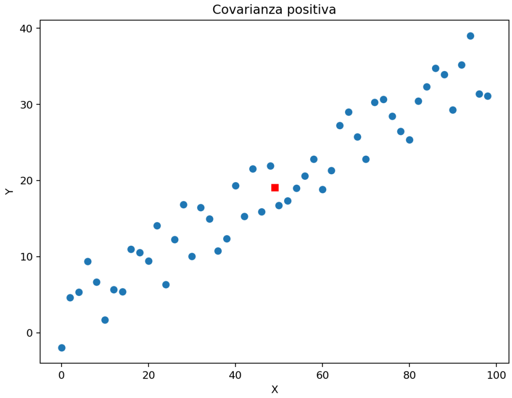
El concepto clave para estudiar la correlación entre las variables es el de **covarianza**:

$$\sigma_{xy} = \frac{1}{N} \sum (x_i - \bar{x})(y_i - \bar{y})$$

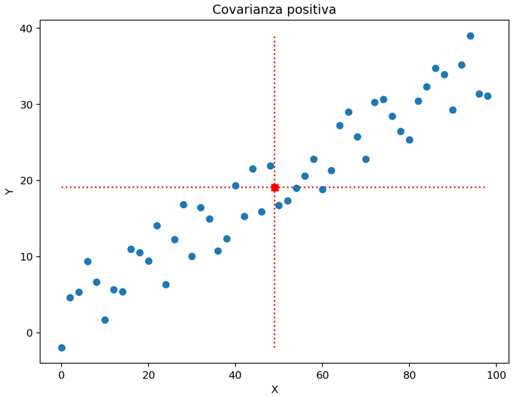
Si las dos variables están correlacionadas positivamente, la covarianza será positiva.



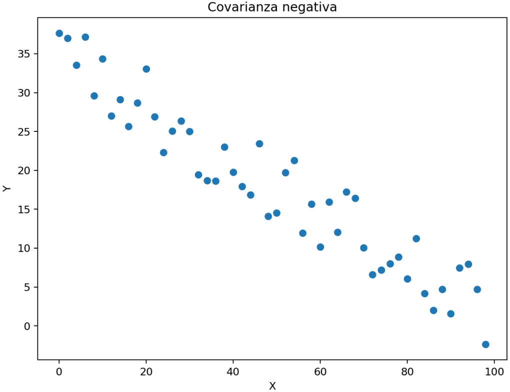
Si las dos variables están correlacionadas positivamente, la covarianza será positiva.



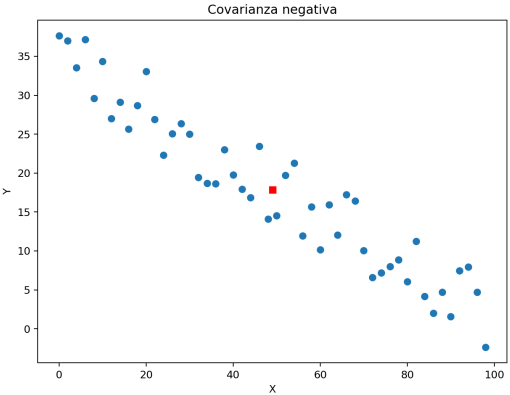
Si las dos variables están correlacionadas positivamente, la covarianza será positiva.



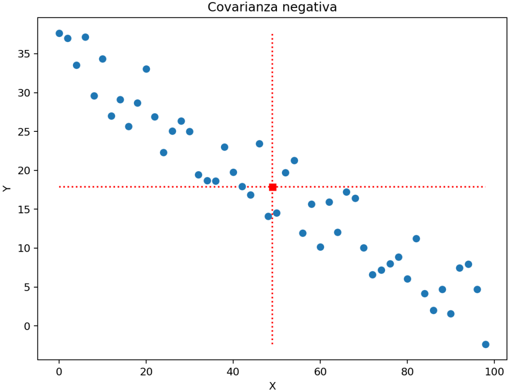
Si las dos variables están correlacionadas negativamente, la covarianza será negativa.



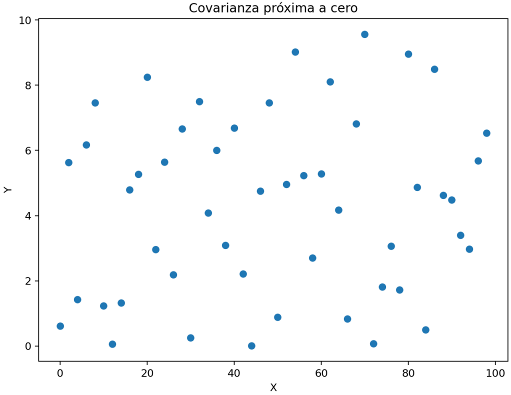
Si las dos variables están correlacionadas negativamente, la covarianza será negativa.



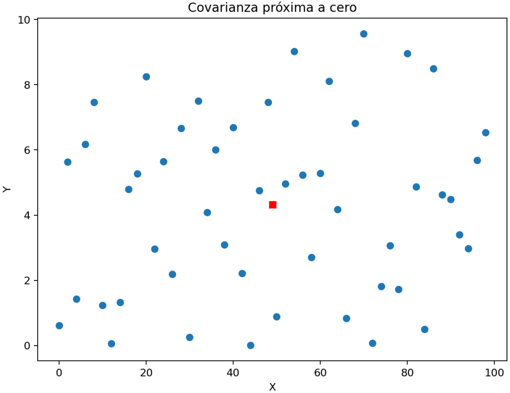
Si las dos variables están correlacionadas negativamente, la covarianza será negativa.



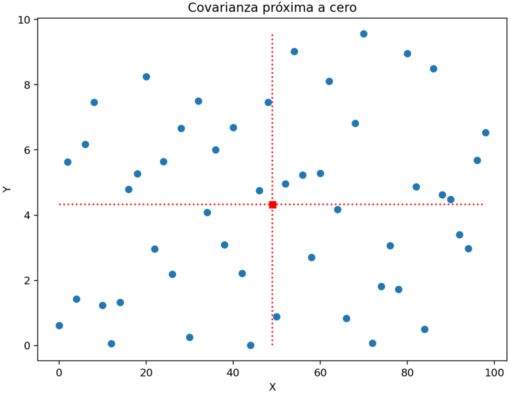
Si las variables no están correlacionadas, la covarianza resulta ser próxima a cero.



Si las variables no están correlacionadas, la covarianza resulta ser próxima a cero.



Si las variables no están correlacionadas, la covarianza resulta ser próxima a cero.



Coeficiente de correlación

La covarianza permite saber si la correlación es positiva o negativa pero no permite saber si es fuerte o débil pues su valor depende de la unidades utilizadas.

Por ello se utiliza el **coeficiente de correlación**:

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

que toma valores comprendidos entre -1 y $+1$.

La correlación es positiva y fuerte si el valor de r es próximo a $+1$; es negativa y fuerte si el valor de r es próximo a -1. Si r es práximo a cero, la correlación es débil.

Ejemplo

Los valores de las variables X e Y medidas sobre la misma población están dados en la siguiente tabla:

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

Calcular la covarianza y el coeficiente de correlación.

Calculamos en primer lugar las media de las variables:

$$\bar{x} = \frac{1}{10}(4+16+18+9+15+13+4+2+13+6) = \frac{100}{10} = 10$$
$$\bar{y} = \frac{1}{10}(3+18+16+7+17+14+6+4+11+7) = \frac{103}{10} = 10{,}3$$

La suma de los productos es:

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

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

La covarianza es la media de los productos menos el producto de las medias:

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

Para calcular el coeficiente de correlación debemos hallar las desviaciones típicas:

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

$$\sigma_y^2 = \frac{1}{10}\left(3^2 + 18^2 + 16^2 + 7^2 + 17^2 + 14^2 + 6^2 + 4^2 + 11^2 + 7^2\right) - 10{,}3^2 = 134{,}5 - 10{,}3^2 = 28{,}41$$

El coeficiente de correlación es:

$$r = \frac{27{,}5}{\sqrt{29{,}6} \cdot \sqrt{28{,}41}} \simeq 0{,}948$$

Agradecimiento

Gracias por vuestra atención