

Notación habitual en una distribución estadística bidimensional

Para la serie de definiciones y cálculos que debemos llevar a cabo con una distribución bidimensional es costumbre utilizar esta notación:

- * Las variables estadísticas reciben el nombre de X e Y.
- * La media de las variables estadísticas son \bar{x} y \bar{y} , respectivamente.
- * Las desviaciones típicas son σ_x y σ_y , respectivamente.

Centro de gravedad de una distribución estadística bidimensional

Es el punto de coordenadas (\bar{x}, \bar{y}) . No tiene por qué coincidir con ninguno de los datos de la distribución.

Covarianza de una distribución bidimensional

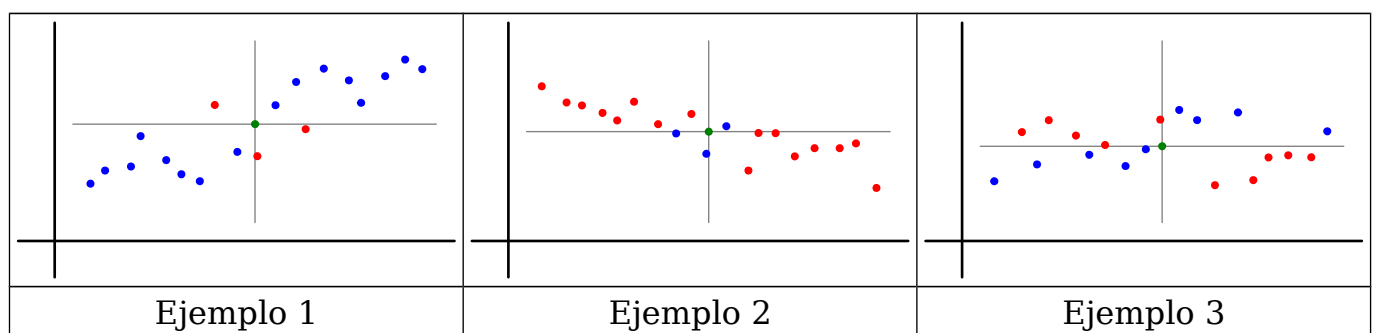
- * Se define como la media de los productos de las desviaciones de cada variable estadística considerando todos los datos.
- * Se designa σ_{xy} .
- * Llamando «n» al número de datos, se define así simbólicamente:

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

Observa la gran similitud de esta definición con la de varianza de una sola variable estadística: en la varianza se elevan al cuadrado las desviaciones (solo hay una para cada dato), en la covarianza se multiplican entre sí (hay dos para cada dato).

Visualización del valor de la covarianza

En las siguientes nubes de puntos, mostramos en verde el centro de gravedad de la distribución, en azul los puntos que tienen positivo el producto de las desviaciones y en rojo los que lo tienen negativo.



El centro de gravedad divide los datos de la distribución en cuatro zonas, señaladas con líneas grises en los ejemplos. Los puntos que están en la zona de arriba a la derecha y los que están abajo a la izquierda tienen el producto positivo (porque las dos desviaciones tienen el mismo signo), los de las otras dos zonas lo tienen negativo (porque las dos desviaciones tienen distinto signo).

- * Cuando hay más productos positivos que negativos (ejemplo 1), la covarianza resulta **positiva**, con valor absoluto mayor cuantos más puntos azules haya.
- * Cuando hay más productos negativos que positivos (ejemplo 2), la covarianza resulta **negativa**, con valor absoluto mayor cuantos más puntos rojos haya.
- * Y cuando el número de productos es similar (ejemplo 3), la covarianza puede tener cualquier signo, pero **valor absoluto pequeño**.