

# Apuntes Unidad 2

Covarianza

---

Curso: Probabilidad y estadística descriptiva

Unidad 2 : Media muestral, dispersión y correlación

Tema: Relación entre variables cuantitativas

Contenido: Covarianza

## COVARIANZA

La covarianza de una muestra con  $n$  datos, donde  $\bar{x}$  es la media de la variable  $x$  e  $\bar{y}$  la media muestral de la variable  $y$ , se calcula utilizando la siguiente fórmula:

$$s_{xy} = \frac{(x_1 - \bar{x}) \cdot (y_1 - \bar{y}) + (x_2 - \bar{x}) \cdot (y_2 - \bar{y}) + \dots + (x_n - \bar{x}) \cdot (y_n - \bar{y})}{n - 1}$$

Por ejemplo, para los  $n = 5$  datos correspondientes al gráfico de dispersión de más arriba, cuyas medias son  $\bar{x} = 4,2$  e  $\bar{y} = 6,6$ , tenemos que:

Dato ( $x_i, y_i$ )	Diferencia respecto a la media en $x$ : ( $x_i - \bar{x}$ )	Diferencia respecto a la media en $y$ : ( $y_i - \bar{y}$ )	Producto de las diferencias: ( $x_i - \bar{x}$ ) · ( $y_i - \bar{y}$ )
(3,5)	$3 - 4,2 = -1,2$	$5 - 6,6 = -1,6$	$-1,2 \cdot -1,6 = 1,92$
(4,7)	$4 - 4,2 = -0,2$	$7 - 6,6 = 0,4$	$-0,2 \cdot 0,4 = -0,08$
(2,4)	$2 - 4,2 = -2,2$	$4 - 6,6 = -2,6$	$-2,2 \cdot -2,6 = 5,72$
(5,7)	$5 - 4,2 = 0,8$	$7 - 6,6 = 0,4$	$0,8 \cdot 0,4 = 0,32$
(7,10)	$7 - 4,2 = 2,8$	$10 - 6,6 = 3,4$	$2,8 \cdot 3,4 = 9,52$

Luego la covarianza entre las variables  $x$  e  $y$ , para este conjunto de datos, es igual a:

$$s_{xy} = \frac{1,92 - 0,08 + 5,72 + 0,32 + 9,52}{5 - 1} = \frac{17,4}{4} = 4,35$$

Podemos escribir la covarianza de la siguiente manera:

$$s_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n - 1}$$

Ahora que sabemos cómo calcular la covarianza, veamos cómo interpretar los posibles valores que esta tome.

### INTERPRETACIÓN DEL SIGNO DE LA COVARIANZA

Dependiendo de la distribución de puntos en el gráfico de dispersión, podemos describir de manera general las distintas situaciones que pueden presentarse.

Cuando la relación entre las variables sigue una tendencia lineal positiva, la mayoría de los puntos se encuentran en el primer grupo y por tanto la suma de los productos de las diferencias es positiva. Luego, la covarianza es positiva.

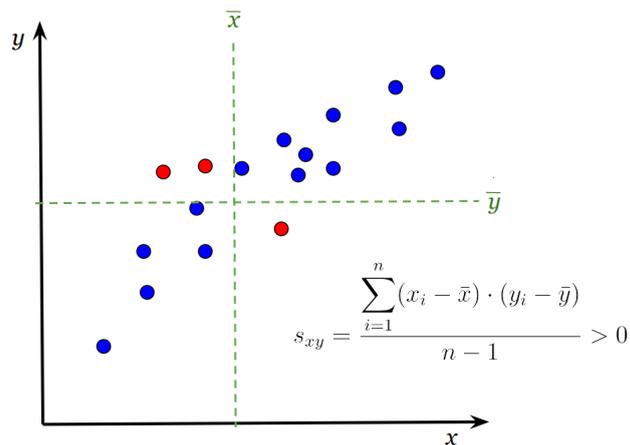


Figura 2: Ejemplo de distribución con covarianza positiva.

Por otro lado, si la relación entre las dos variables sigue una tendencia lineal negativa, la covarianza será negativa. Esto sucede porque hay muchos más puntos en el segundo grupo.

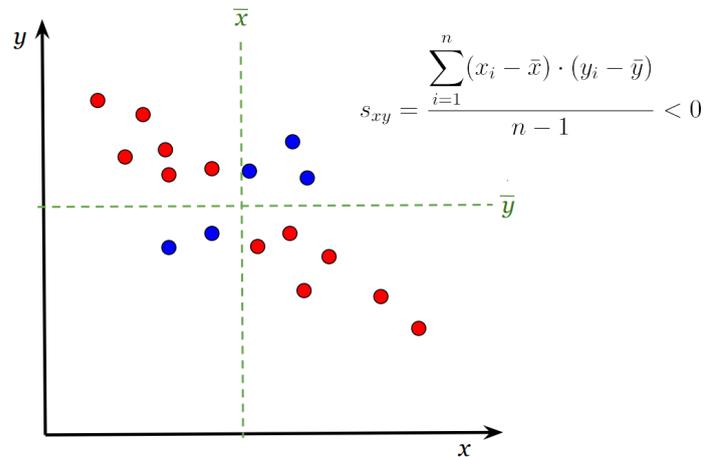


Figura 3: Ejemplo de distribución con covarianza negativa.

Por último, cuando los puntos están distribuidos de forma pareja, la suma de los productos negativos anula casi exactamente a la suma de productos positivos, por lo que la covarianza será aproximadamente cero. Esto ocurre, por ejemplo, en situaciones en las que no se observa ningún patrón claro entre las dos variables.

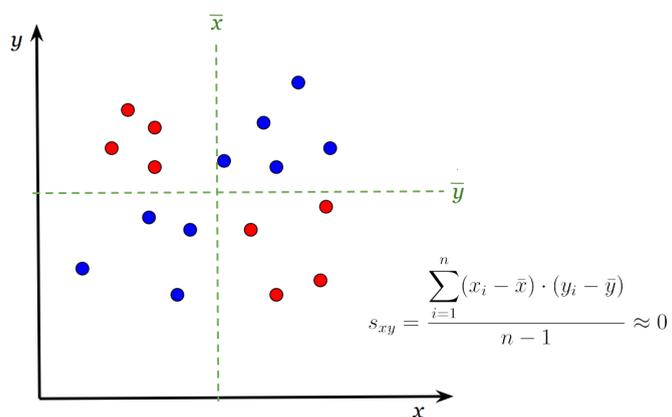


Figura 4: Ejemplo de distribución donde la covarianza será aproximadamente cero.

Tal y como hemos aprendido anteriormente, muchas veces puede resultar complejo calcular estas medidas manualmente. Por esta razón, en la siguiente tabla se muestran los comandos necesarios para calcularlos utilizando Microsoft Excel o Google Sheets, donde A

Curso: Probabilidad y estadística descriptiva

Unidad 2 : Media muestral, dispersión y correlación

Tema: Relación entre variables cuantitativas

Contenido: Covarianza

son las celdas que contienen los datos de la variable 1 y B son las celdas donde se ubican los datos de la variable 2.

	Microsoft Excel	Google Sheets
Comando	=COVARIANZA.M(A; B)	=COVARIANZA.S(A; B)

## SÍNTESIS

- La covarianza puede usarse para analizar si la tendencia lineal entre dos variables es positiva, negativa o no existe relación entre estas.
- Al realizar un gráfico de dispersión de dos variables cuantitativas, estas pueden relacionarse de la siguiente forma:
  - Si las dos variables siguen una tendencia lineal positiva, la covarianza entre los datos correspondientes a estas variables es positiva.
  - Si las dos variables siguen una tendencia lineal negativa, la covarianza entre los datos correspondientes a estas variables es negativa.
  - Si las dos variables no siguen alguna tendencia o no están relacionadas de forma lineal, la suma del producto de las diferencias será cercana a cero, por lo que la covarianza será parecida a ese valor.