

Apuntes Unidad 2

Coeficiente de correlación



Curso: Probabilidad y estadística descriptiva

Unidad 2 : Media muestral, dispersión y correlación

Tema: Relación entre variables cuantitativas

Contenido: Coeficiente de correlación

COEFICIENTE DE CORRELACIÓN

La covarianza es una medida de la relación lineal entre dos variables. Tal como estudiamos en la lección anterior, un valor positivo indica una relación lineal positiva o creciente y un valor negativo indica una relación lineal negativa o decreciente.

La covarianza puede ser difícil de interpretar ya que depende de las unidades de medición. Para eliminar el efecto de las escalas de medición introduciremos una nueva medida llamada coeficiente de correlación lineal, el cual se obtiene normalizando la covarianza:

$$r_{xy} = \frac{s_{xy}}{s_x \cdot s_y}$$

El coeficiente de correlación es un valor tal que:

$$-1 \leq r_{xy} \leq 1$$

A continuación, veremos cómo interpretar los posibles valores del coeficiente de correlación lineal.

INTERPRETACIÓN DEL COEFICIENTE DE CORRELACIÓN

Un coeficiente de correlación lineal $r_{xy} > 0$ nos indica que hay una tendencia lineal positiva entre las variables. El grado de correlación lineal es mayor si el coeficiente r_{xy} es cercano a 1.

Por otro lado, un coeficiente de correlación lineal $r_{xy} < 0$ nos indica que hay una tendencia lineal negativa entre las variables. En este caso, el grado de correlación lineal es mayor si el coeficiente r_{xy} es cercano a -1.

Curso: Probabilidad y estadística descriptiva
Unidad 2 : Media muestral, dispersión y correlación
Tema: Relación entre variables cuantitativas
Contenido: Coeficiente de correlación

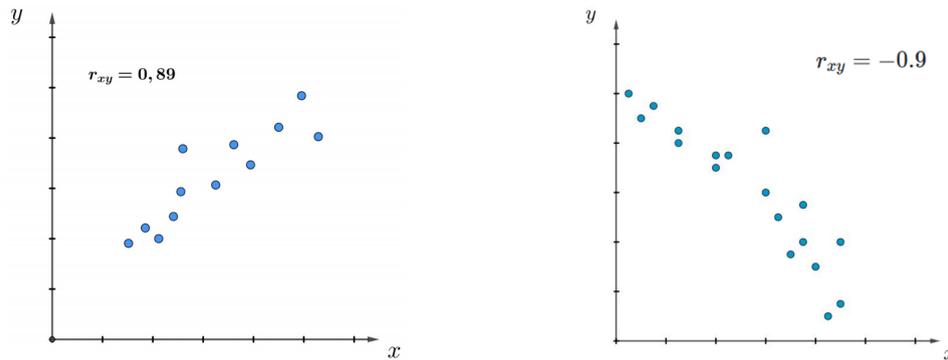


Figura 1: A la izquierda una distribución cuyo coeficiente de correlación es positivo y a la derecha una donde es negativo.

Por último, si el coeficiente de correlación es cercano a 0 ($r_{xy} \approx 0$), podemos presumir que la relación entre las variables x e y es inexistente o bien que tiene una tendencia que no es lineal.

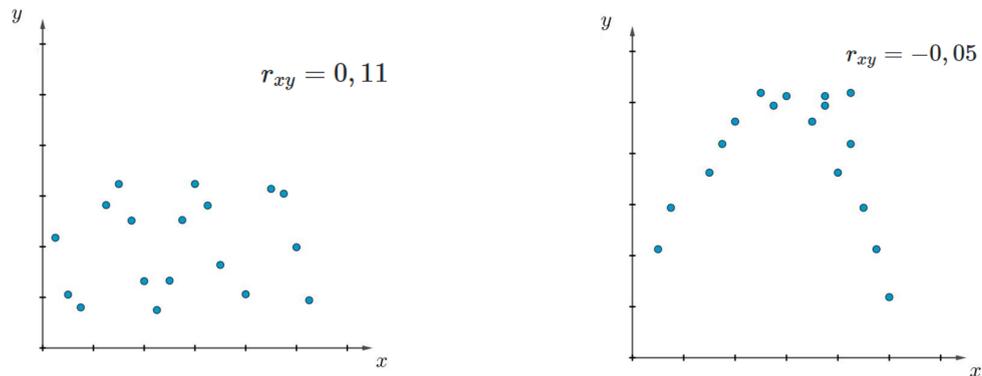


Figura 2: Distribuciones cuyo coeficiente de correlación es cercano a cero.

En algunas ocasiones, calcular manualmente el valor del coeficiente de correlación lineal entre dos variables puede ser complejo, por lo que es importante saber cómo hacerlo utilizando hojas de cálculo.

Curso: Probabilidad y estadística descriptiva

Unidad 2 : Media muestral, dispersión y correlación

Tema: Relación entre variables cuantitativas

Contenido: Coeficiente de correlación

CALCULANDO EL COEFICIENTE DE CORRELACIÓN EN MICROSOFT EXCEL Y GOOGLE SHEETS

Tanto en Microsoft Excel como en Google Sheets existe una función que permite calcular rápidamente el coeficiente de correlación entre dos variables cuantitativas, siendo idéntica en ambos programas computacionales. Para calcularlo debemos utilizar lo siguiente:

$$=COEF.DE.CORREL(A:B ; C:D)$$

Donde A:B delimitan la ubicación de los datos de la primera variable, donde A es la primera celda que los contiene y B la última. Análogamente, C:D delimitan la ubicación de los datos de la segunda variable.

SÍNTESIS

- El valor del coeficiente de correlación lineal no depende de las unidades de medida de las variables.
- Para calcular el coeficiente de correlación lineal entre dos variables x e y , necesitamos calcular la desviación estándar para cada variable y su respectiva covarianza.
- El coeficiente de correlación es un valor que se encuentra entre -1 y 1 .
 - Cuando este valor crece de 0 a 1 , crece la tendencia lineal positiva.
 - Cuando este valor decrece de 0 a -1 , crece la tendencia lineal negativa.
 - Si las variables no están asociadas, o se relacionan de forma no lineal, el valor del coeficiente de correlación será cercano a 0 .
- Se puede calcular el coeficiente de correlación con un solo comando y que es posible filtrar una base de datos en una hoja de cálculo.
- Que las relaciones observadas a partir de una muestra de datos no necesariamente representarán el comportamiento de las variables en la población.

Curso: Probabilidad y estadística descriptiva

Unidad 2 : Media muestral, dispersión y correlación

Tema: Relación entre variables cuantitativas

Contenido: Coeficiente de correlación

- El coeficiente de correlación lineal nos permite comparar el grado de asociación lineal entre distintas variables dentro de un mismo contexto. Si el valor absoluto del coeficiente de correlación se acerca a 1, mayor será la asociación lineal de la relación entre las variables.