

# Correlación

Una correlación existe entre dos variables cuando una de ellas está relacionada con la otra de alguna manera.

## Coeficiente de correlación lineal

*La inspección visual de los diagramas de dispersión es muy subjetiva, por lo que se necesitan medidas más precisas y objetivas. El coeficiente de correlación lineal  $r$  sirve para detectar patrones lineales.*

**Definición:** *el coeficiente de correlación lineal  $r$  mide la fuerza de la relación lineal entre los valores cuantitativos apareados  $x$  y  $y$  en una muestra. Su valor se calcula con la fórmula:*

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \sqrt{n(\sum y^2) - (\sum y)^2}}$$

*El coeficiente de correlación lineal también se conoce como coeficiente de correlación de Pearson, en honor de Karl Pearson (1857-1936), quien lo desarrolló originalmente.*

## Requisitos

*Dado cualquier conjunto de datos muestrales apareados, siempre se puede calcular el coeficiente de correlación lineal  $r$ , pero se deben satisfacer los siguientes requisitos cuando se prueban hipótesis o cuando se hacen inferencias acerca de  $r$ :*

# Correlación

1. La muestra de datos apareados  $(x, y)$  es una muestra aleatoria de datos cuantitativos. (Es importante que los datos muestrales no se hayan reunido por medio de algún método inapropiado, como una muestra de respuesta voluntaria).
2. El examen visual del diagrama de dispersión debe confirmar que los puntos se acercan al patrón de una línea recta.
3. Es necesario eliminar cualquier valor extremo si se sabe que se trata de un error. Los efectos de cualquier otro valor extremo deben tomarse en cuenta calculando  $r$  con y sin el valor extremo incluido.

## **Notación para el coeficiente de correlación lineal**

$n$ : representa el número de pares de datos presentes.

$\Sigma$ : denota la suma de los elementos indicados.

$\Sigma x$ : denota la suma de todos los valores de  $x$ .

$\Sigma x^2$ : indica que cada valor de  $x$  debe elevarse al cuadrado y después deben sumarse esos cuadrados.

$(\Sigma x)^2$ : indica que los valores de  $x$  deben sumarse y el total elevarse al cuadrado.

Es sumamente importante evitar confundirse entre  $\Sigma x^2$  y  $(\Sigma x)^2$ .

$\Sigma xy$ : indica que cada valor de  $x$  debe multiplicarse primero por su valor  $y$  correspondiente.

Después de obtener todos estos productos, se calcula su suma.

$r$ : representa el coeficiente de correlación lineal de una muestra.

# Correlación

$\rho$ : la letra griega rho se usa para representar el coeficiente de correlación lineal de una población

## **Propiedades del coeficiente de correlación lineal $r$**

1. El valor de  $r$  está siempre entre  $-1$  y  $+1$ , inclusive. Es decir,

$$-1 \leq r \leq +1$$

2. El valor de  $r$  no cambia si todos los valores de cualquiera de las variables se convierten a una escala diferente.

3. El valor de  $r$  no se ve afectado por la elección de  $x$  o  $y$ . Intercambie todos los valores de  $x$  y  $y$ , y el valor de  $r$  no sufrirá cambios.

4.  $r$  mide la fuerza de una relación lineal. No está diseñada para medir la fuerza de una relación que no sea lineal.

### **Referencia:**

Triola, M., (2013). Estadística. Decimoprimer edición. Pearson educación. México