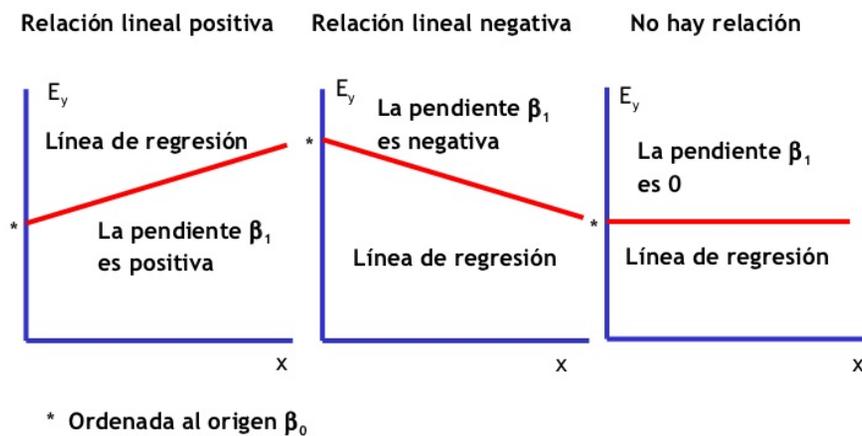


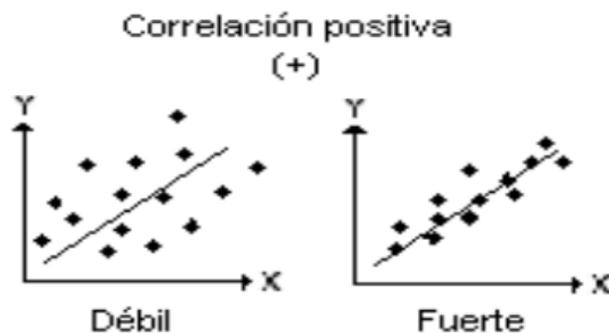
Correlación Simple

A menudo el análisis de correlación simple se utiliza junto con el análisis de regresión lineal simple para medir la eficacia con que la línea de regresión explica la variación de la variable dependiente, Y .

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



Ejemplo de correlación lineal positiva



Correlación Simple

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 Simple

- 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.

Σ : denota la suma de los elementos indicados.

Σx : denota la suma de todos los valores de x .

Σx^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.

Correlación Simple

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

$\sum xy$: 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.

ρ : 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.

Correlación Simple

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.

En general, $r > 0$ indica una relación positiva y $r < 0$ indica una relación negativa, mientras que $r = 0$ indica que no hay relación (o que las variables son independientes y no están relacionadas). Aquí, $r = 1$ describe una correlación positiva perfecta y $r = -1$ describe una correlación negativa perfecta.

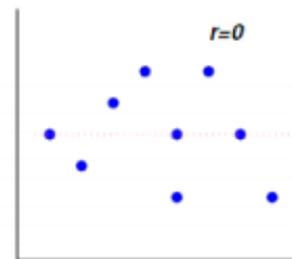
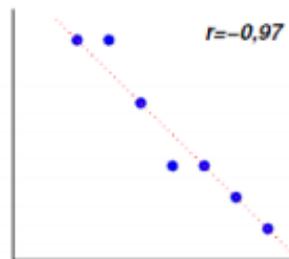
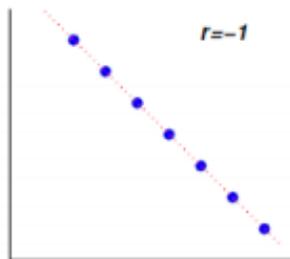
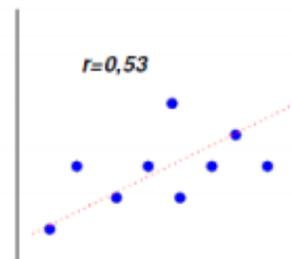
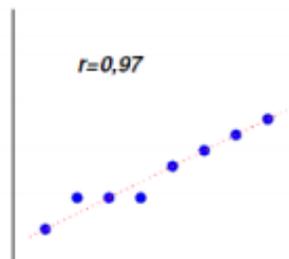
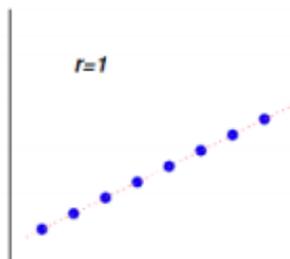
Cuanto más cerca estén los coeficientes de $+1$ y -1 , mayor será la fuerza de la relación entre las variables.

Como norma, las siguientes directrices sobre la fuerza de la relación son útiles (aunque muchos expertos podrían disentir con la elección de los límites).

Correlación Simple

Rango		Relacion Lineal
$\pm 0,96$	$\pm 1,0$	Perfecta
$\pm 0,85$	$\pm 0,95$	Fuerte
$\pm 0,70$	$\pm 0,84$	Significativa
$\pm 0,50$	$\pm 0,69$	Moderada
$\pm 0,20$	$\pm 0,49$	Débil
$\pm 0,10$	$\pm 0,19$	Muy Débil
$\pm 0,09$	$\pm 0,0$	Nula

Veamos los siguientes diagramas de dispersión:



Correlación Simple

En la figura vemos que $r = \pm 1$ es lo mismo que decir que las observaciones de ambas variables están perfectamente alineadas. La relación lineal es tanto más perfecta cuanto r está cercano a ± 1 .

En la correlación no se distingue la variable dependiente de la independiente, la correlación de X con respecto a Y es la misma que la correlación de Y con respecto a X .

Aunque la interpretación de la magnitud del coeficiente de correlación depende del contexto particular de aplicación, en términos generales se considera que una correlación es débil por debajo de 0.50 en **valor absoluto**, que existe una asociación moderada entre 0.50 y 0.60, y fuerte por encima de 0.85.

REFERENCIAS:

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

http://asesorias.cuautitlan2.unam.mx/Laboratoriovirtualdeestadistica/CARPETA%203%20INFERENCIA_ESTADISTICA/DOC_%20INFERENCIA/TEMA%204/09%20REGRESION%20Y%20CORRELACION%20LINEAL%20SIMPLE.pdf

<http://www.ics-aragon.com/cursos/salud-publica/2014/pdf/M2T04.pdf>

<https://platzi.com/tutoriales/1269-probabilidad-estadistica/2308-coeficiente-de-correlacion-que-es-y-para-que-sirve/>