



# Regresión y correlación

ANÁLISIS DE REGRESIÓN

# Relación entre variables

Si los responsables de la toma de decisiones pueden determinar cómo lo conocido se relaciona con el evento futuro, pueden ayudar considerablemente al proceso de toma de decisiones. Los análisis de regresión y correlación nos mostrarán cómo determinar tanto la naturaleza como la fuerza de una relación entre dos variables.

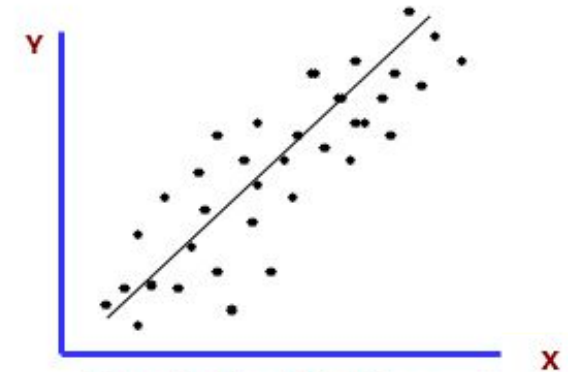
Los análisis de regresión y de correlación se basan en la relación, o asociación, entre dos (o más) variables. La variable (o variables) conocida(s) se llaman variable(s) independientes(s); la que tratamos de predecir es la variable dependiente. En el análisis de regresión desarrollaremos una fórmula matemática que relacione las variables antes mencionadas.

# TIPOS DE RELACIONES

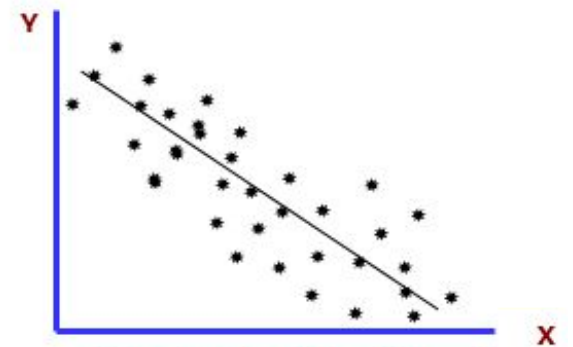
VARIABLES X Y Y

# Relación entre $x$ y $y$

**DIRECTA:** Al incrementar la variable independiente  $X$ , la variable dependiente  $Y$  también lo hace. Se dice que la pendiente de esta recta es positiva, porque  $Y$  crece si  $X$  crece.

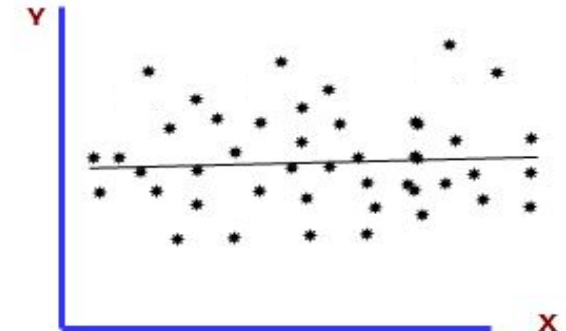


**INVERSA:** la variable dependiente  $Y$  disminuye al incrementar la variable independiente  $X$ . Se dice que la pendiente de esta recta es negativa porque  $Y$  decrece si  $X$  crece.

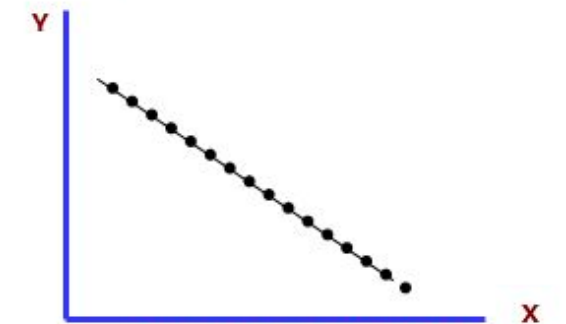


# Relación entre $x$ y $y$

**NULA:** la variable  $Y$  es independiente de la variable  $X$ , ambas son independientes. Se dice que las variables no están relacionadas.



**PERFECTA:** ya sea de manera directa o inversa, los valores de  $X$  vs  $Y$  se ajustan perfectamente a una línea recta.



# Relación de asociación, no causa-efecto

A menudo encontramos una relación causal entre variables, esto es, la variable independiente  $X$  “causa” cambios en la variable dependiente  $Y$ .

Por esta razón es importante considerar que las relaciones encontradas por la regresión son relaciones de asociación, pero no necesariamente de causa y efecto. A menos que tenga razones específicas para creer que los valores de la variable dependiente se originan por los valores de las variables independientes, no infiera causalidad en las relaciones encontradas por la regresión.

# Análisis de correlación

COEFICIENTE DE CORRELACIÓN

# Coeficiente de correlación

El análisis de correlación es la herramienta estadística que podemos usar para describir el grado en el que una variable esta linealmente relacionada con otra. Sin embargo, la correlación también se puede usar sola para medir el grado de asociación entre dos variables.

Una de las medidas que podemos usar para describir qué tan bien explica una variable a otra es el coeficiente de correlación, denotado por  $r$  y definido como:  $r = \sqrt{1 - \frac{\sum(Y - \hat{Y})^2}{\sum(Y - \bar{Y})^2}}$

donde  $\sum(Y - \hat{Y})^2$  es la variación de los valores de  $Y$  alrededor de la recta de regresión

$\sum(Y - \bar{Y})^2$  es la variación de los valores de  $Y$  alrededor de su propia media

cuando la pendiente de la ecuación de estimación es positiva,  $r$  es la raíz cuadrada positiva, pero si la pendiente es negativa,  $r$  es la raíz cuadrada negativa.



# Interpretación de $r$

El signo de  $r$  indica la dirección de la relación entre las dos variables  $X$  y  $Y$ :

- ▶ Si existe una relación inversa, entonces  $r$  caerá entre  $-1$  y  $0$
- ▶ si existe una relación directa, entonces  $r$  caerá entre  $0$  y  $1$

$r=A$  significa que el  $A*100\%$  de los datos están relacionados entre sí

# Nivel de asociación, según el valor de $r$

Relación negativa	Nivel de asociación	Relación positiva
$r = -1$	PERFECTA	$r = 1$
$-1 < r < -0.9$	MUY FUERTE	$0.9 < r < 1$
$-0.9 < r < -0.7$	FUERTE	$0.7 < r < 0.9$
$-0.7 < r < -0.4$	MODERADO	$0.4 < r < 0.7$
$-0.4 < r < -0.2$	DÉBIL	$0.2 < r < 0.4$
$-0.2 < r \leq 0$	DESPRECIABLE	$0 \leq r < 0.2$