

ANÁLISIS DE REGRESIÓN LINEAL

El análisis de regresión lineal es una técnica estadística que permite examinar la relación entre una variable dependiente y una o más variables independientes, con el propósito de explicar o predecir comportamientos. Su fundamento radica en representar esta relación mediante una línea recta que mejor se ajusta a los datos observados, lo que facilita estimar cómo los cambios en una variable explicativa afectan a la variable resultado.

En los negocios internacionales, este análisis se utiliza para comprender fenómenos como el efecto del tipo de cambio sobre las exportaciones, la influencia del costo logístico en los precios de venta o la relación entre la inversión extranjera y el crecimiento del comercio exterior, proporcionando así una herramienta clave para la toma de decisiones estratégicas basadas en evidencia.

La ecuación de regresión expresa una relación entre x (llamada variable explicativa, variable de pronóstico o variable independiente) y $\hat{y}$ (llamada variable de respuesta o variable dependiente). La ecuación típica de una línea recta $y = mx + b$ está expresada en la forma $\hat{y} = b_0 + b_1x$, donde b_0 es el intercepto y (o la intersección en y) y b_1 es la pendiente. La notación dada muestra que b_0 y b_1 son estadísticos muestrales utilizados para estimar los parámetros poblacionales β_0 y β_1 .

Requisitos:

1. La muestra de datos apareados (x, y) es una muestra aleatoria de datos cuantitativos.
2. El examen visual del diagrama de dispersión indica que los puntos se aproximan al patrón de una línea recta.
3. Se debe eliminar cualquier valor extremo si se sabe que es un error. Es importante tomar en cuenta los efectos de cualquier valor extremo que no sea un error conocido.

Nota: Los requisitos 2 y 3 representan una verificación simplificada de los siguientes requisitos formales del análisis de regresión:

- Para cada valor fijo de x , los valores correspondientes de y tienen una distribución en forma de campana.
- Para los distintos valores fijos de x , las distribuciones de los valores correspondientes de y tienen la misma varianza.
- Para los distintos valores fijos de x , las distribuciones de los valores correspondientes de y tienen medias que se ubican en la misma línea recta.
- Los valores de y son independientes.

Los resultados no se ven muy afectados si la distribución no se aleja demasiado de la normalidad y si las varianzas no son demasiado diferentes.

$$b_1 = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

Dado un conjunto de datos muestrales apareados, la ecuación de regresión

$$\hat{y} = b_0 + b_1x$$

describe algebraicamente la relación entre las dos variables. La gráfica de la ecuación de regresión se denomina recta de regresión (*o recta del mejor ajuste o recta de mínimos cuadrados*).

Notación para la ecuación de regresión

	Parámetro poblacional	Estadístico muestral
Intercepto y de la ecuación de regresión	β_0	b_0
Pendiente de la ecuación de regresión	β_1	b_1
Ecuación de la recta de regresión	$y = \beta_0 + \beta_1x$	$\hat{y} = b_0 + b_1x$

Cálculo de la pendiente b_1 y de b_0 (el intercepto y) en la ecuación de regresión $\hat{y} = b_0 + b_1x$

Pendiente:

$$b_1 = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

Intercepto y :

$$b_0 = \bar{y} - b_1\bar{x}$$

Uso de la ecuación de regresión para hacer predicciones

Se debe utilizar la ecuación de regresión para hacer predicciones **solo si existe una correlación lineal**. En ausencia de una correlación lineal, no debemos usar la ecuación de regresión para proyectar o pronosticar; en vez de ello, nuestro mejor estimado de la segunda variable es simplemente su media muestral.

Al pronosticar un valor de y con base en algún valor dado de x ...

1. Si no existe una correlación lineal, el mejor valor pronosticado de y es $\bar{y}$.
2. Si existe una correlación lineal, el mejor valor pronosticado de y , se calcula sustituyendo el valor de x en la ecuación de regresión.

Lineamientos para el uso de la ecuación de regresión

1. Si no existe una correlación lineal, no debe utilizarse la ecuación de regresión para hacer pronósticos.
2. Cuando se utilice la ecuación de regresión para hacer pronósticos, permanece en el ámbito de los datos muestrales disponibles. Si se calcula una ecuación de regresión que relaciona la estatura y el número de calzado de mujeres, es absurdo pronosticar el número de calzado de una mujer que mide más de 2m de estatura.

3. Una ecuación de regresión que está basada en datos antiguos no necesariamente es válida ahora. La ecuación de regresión que relaciona precios de automóviles usados con la antigüedad de los automóviles ya no es útil si está basada en datos de la década de 1990.
4. No se debe hacer pronósticos acerca de una población distinta de la población de donde se obtuvieron los datos muestrales. Si reunimos datos muestrales de hombres y desarrollamos una ecuación de regresión que relaciona la edad con el uso del control remoto del televisor, los resultados no necesariamente se aplican a las mujeres.
5. Falacia ecológica, esta propone que es un error inferir relaciones individuales a partir de datos agregados a nivel grupal, por ejemplo, supongamos que un estudio muestra que los países con mayores exportaciones presentan, en promedio, una alta productividad laboral. A partir de esta correlación, **podríamos concluir erróneamente** que todas las empresas exportadoras dentro de esos países son más productivas que las de los países con menores exportaciones.

Sin embargo, esta inferencia sería incorrecta, ya que la relación observada a nivel agregado (país) no necesariamente se cumple a nivel individual (empresa), es posible que solo unas pocas corporaciones multinacionales concentren la mayoría de las exportaciones, mientras que la mayoría de las pequeñas y medianas empresas tengan niveles de productividad bajos o incluso por debajo del promedio.

Referencia:

Triola, M. (2013). Estadística. Decimoprimera edición. Pearson Educación. México.