

■ Interpretación sobre la normalidad condicional en OLS

Tu razonamiento es correcto: si los residuos globales del modelo (los errores observados para todos los X_i) se comportan aproximadamente como una distribución normal centrada en cero, entonces es muy probable que la distribución condicional de los errores $\varepsilon|X_i$ sea también normal para cada valor de X_i .

----- ■ Fundamento estadístico

En regresión lineal, el modelo asume que: $\varepsilon_i | X_i \sim N(0, \sigma^2)$

y que los errores son independientes e idénticamente distribuidos (i.i.d.). Esto significa que todos los errores provienen de la misma distribución, independientemente del valor de X_i .

Por lo tanto, si: - la varianza es constante (homoscedasticidad), - los errores son independientes, - y los residuos globales tienen forma de distribución normal,

entonces la normalidad condicional ($\varepsilon|X_i$) se cumple implícitamente para todos los valores de X .

----- ■ Analogía intuitiva

Imaginá que los errores del modelo son como bolitas extraídas de una misma urna (la distribución de errores). Si las bolitas (residuos) que observás forman una distribución normal cuando las mirás todas juntas, entonces la “urna” original —que representa la distribución de errores para cada X — también debe ser normal.

----- ■■ Matiz importante

Este razonamiento es válido solo si la varianza de los errores no depende de X (es decir, si hay homoscedasticidad). Si la varianza cambia con X , cada $\varepsilon|X_i$ puede tener su propia distribución normal con diferente σ^2 , lo cual rompe la suposición de “idénticamente distribuidos”.

----- ■ En resumen

Condición | Consecuencia ----- | ----- Errores
independientes ■ | No hay relación entre los residuos Homoscedasticidad ■ | Todos los errores tienen la misma varianza Residuos globales ~ Normal ■ | Implica normalidad condicional $\varepsilon|X_i$

----- ■ Conclusión final

Si los residuos globales del modelo se comportan como una distribución normal centrada en cero, entonces es muy probable que exista normalidad en la distribución de errores condicional a cada X_i .