# **REGRESIÓN LINEAL**

En el contexto de una regresión lineal, un modelo se considera MELI (Modelo de Estimación Lineal Insesgado) cuando cumple con los supuestos de los estimadores de mínimos cuadrados ordinarios (MCO o OLS, por sus siglas en inglés), también conocidos como los supuestos de Gauss-Markov. Los principales son:

**1. Linealidad en los parámetros:** El modelo debe ser lineal en los parámetros. Esto significa que la relación entre la variable dependiente (Y) y las variables independientes (X) es una combinación lineal de los coeficientes, aunque las variables en sí pueden ser no lineales.

**2. Exogeneidad** (Esperanza condicional cero): Los errores (residuos) del modelo deben tener una esperanza condicional de cero dado cualquier valor de las variables independientes. Es decir, \( E(u | X) = 0 \). Esto implica que no debe haber correlación entre los errores y las variables explicativas.

Debe cumplirse:

**3. No colinealidad perfecta:** No debe existir multicolinealidad perfecta entre las variables explicativas, lo que significa que las variables independientes no deben estar perfectamente correlacionadas entre sí.

**4. Homoscedasticidad:** La varianza de los errores debe ser constante para todos los valores de las variables independientes. Es decir, los errores no deben mostrar patrones de aumento o disminución de la varianza a lo largo de las predicciones del modelo.

**5. No autocorrelación:** Los errores no deben estar correlacionados entre sí. Esto es importante especialmente en modelos de series temporales.

**6. Normalidad de los errores (en algunos casos):** Aunque no es un requisito para obtener estimadores insesgados, se asume que los errores tienen una distribución normal si se desea realizar pruebas de hipótesis precisas y obtener intervalos de confianza fiables.

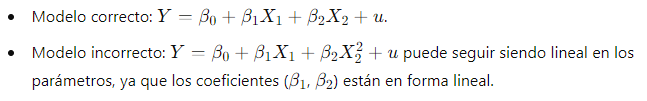
Cuando estos supuestos se cumplen, los coeficientes estimados por MCO serán insesgados, eficientes y consistentes, y el modelo puede ser considerado MELI.

Entiendo que quieras una explicación clara de cada uno de los supuestos de un modelo de regresión lineal para que sea MELI (Modelo de Estimación Lineal Insesgado), con ejemplos prácticos de lo que sucede si estos supuestos no se cumplen. Vamos paso por paso.

**1. Linealidad en los Parámetros**

**Supuesto:** La relación entre la variable dependiente (Y) y las variables independientes (X) debe ser lineal en los parámetros (coeficientes). Esto no implica que las variables independientes tengan que ser lineales, pero sí los parámetros.

**- Ejemplo:**



**- Qué pasa si no se cumple:** Si la relación no es lineal en los parámetros (por ejemplo, si incluimos parámetros en forma de logaritmos o productos no lineales), los estimadores obtenidos por mínimos cuadrados ya no serán MELI. El modelo no capturará bien la relación entre las variables y los resultados serán sesgados.

**2. Exogeneidad (Esperanza Condicional Cero)**

Supuesto: Los errores *‘u’* no deben estar correlacionados con las variables independientes *‘X’*. Es decir,



Los errores deben ser "ruido puro" no relacionado con los predictores.

**- Ejemplo:**

- Supongamos que estamos modelando los salarios (Y) en función de la educación (X). Si existe una variable omitida (como la experiencia laboral) que afecta tanto a la educación como al salario, entonces los errores (que contienen el efecto de la experiencia) estarán correlacionados con la educación.

**- Qué pasa si no se cumple:** Si los errores están correlacionados con alguna variable independiente, el modelo estará sesgado y mal especificado. Por ejemplo, el efecto de la educación en el salario podría estar sobre o subestimado si no se tiene en cuenta la experiencia.

**3. No Colinealidad Perfecta (No Multicolinealidad)**

**Supuesto:** No debe existir una relación lineal perfecta entre las variables explicativas. Las variables independientes deben contener información única.

**- Ejemplo:**

- Imagina que estás intentando modelar el consumo de energía en función de la temperatura y el uso de aire acondicionado. Si el uso de aire acondicionado está completamente determinado por la temperatura, entonces hay colinealidad perfecta.

**- Qué pasa si no se cumple:** Cuando existe colinealidad perfecta, los coeficientes no se pueden estimar de forma única porque las variables son redundantes. El software estadístico puede dar error o producir coeficientes sin sentido.

**4. Homoscedasticidad (Varianza Constante de los Errores)**

**Supuesto:** La varianza de los errores debe ser constante para todos los valores de *‘X’*. Esto significa que los errores no deben aumentar o disminuir a medida que cambian los valores de las variables independientes.

**- Ejemplo:**

- Supongamos que estás modelando los gastos de publicidad en función de las ventas de una empresa. Si a medida que las ventas crecen, la varianza en los gastos de publicidad también aumenta, esto es heteroscedasticidad.

**- Qué pasa si no se cumple:** Si los errores no tienen varianza constante (es decir, hay heteroscedasticidad), las estimaciones de los coeficientes serán insesgadas, pero las pruebas de hipótesis y los intervalos de confianza serán incorrectos. Los errores estándar estarán mal calculados y podrías cometer errores al interpretar los resultados.

**5. No Autocorrelación de los Errores**

**Supuesto:** Los errores no deben estar correlacionados entre sí. Esto es particularmente importante en modelos de series temporales, donde los errores en un momento pueden estar correlacionados con los errores en otro momento.

**- Ejemplo:**

- Estás modelando el precio de las acciones en función de varios factores económicos. Si los errores en un día están correlacionados con los errores del día anterior (por ejemplo, debido a shocks del mercado), tienes autocorrelación.

**- Qué pasa si no se cumple:** La autocorrelación de los errores puede llevar a estimaciones incorrectas de los coeficientes. Aunque los coeficientes sigan siendo insesgados, sus errores estándar estarán mal calculados, lo que afectará la precisión de las pruebas de significancia.

**6. Normalidad de los Errores**

**Supuesto:** Los errores deben seguir una distribución normal, aunque este supuesto no es estrictamente necesario para que los estimadores sean MELI. Es relevante para realizar inferencias precisas (pruebas de hipótesis, intervalos de confianza).

**- Ejemplo:**

- Estás modelando el rendimiento de estudiantes basado en horas de estudio. Si algunos estudiantes presentan valores atípicos (outliers) muy altos o muy bajos, los errores pueden no seguir una distribución normal.

**- Qué pasa si no se cumple:** Si los errores no son normales, las pruebas estadísticas pueden ser inexactas, especialmente si el tamaño de la muestra es pequeño. Las pruebas de significancia y los intervalos de confianza podrían estar mal calculados, llevando a conclusiones erróneas sobre la importancia de las variables.

Si un modelo no cumple con estos supuestos, sus resultados pueden ser sesgados (es decir, los coeficientes no representarán correctamente la verdadera relación entre las variables), ineficientes (pueden no ser los mejores estimadores posibles en términos de varianza), y las inferencias (pruebas de hipótesis y predicciones) no serán confiables.