Econometría

Supuestos básico (código de R de Edward Rubin)

Ana María Díaz

27 July 2023

Bienvenidos

MCO: Supuestos y propiedades

Los supuestos del modelo clásico de regresión lineal están resumidos en la siguiente tabla:

Supuesto	Implicación
A1. Lineal	$y = X\beta + \epsilon$
A2. Exogeneidad Estricta	$E[\epsilon_i \mid X] = 0$
A3. Colinealidad Imperfecta	X es una matriz nxK con rango K
A4. Perturbaciones Esféricas	$Var[\epsilon_i \mid X] = \sigma^2$
	$Cov\left[\epsilon_i, \epsilon_j \mid X\right] = 0$
A5. Regresores no estocásticos	X es una matriz nxK no estocástica
A6. Normalidad	$\epsilon \mid X \sim N(0, \sigma^2 I)$
A.2, A.4-A.6	$\epsilon \mid X \sim i.i.d \quad N(0, \sigma^2 I)$

S1: Lineal en parámetros

El valor esperado de la distribución de y está relacionada con el valor de X_i de una manera lineal:

$$E[Y|X_i = x] = f(X_i) = X_i\beta$$

Por lo tanto el proceso generador de datos es igual a

$$Y_i = X\beta + \epsilon$$

S1: Lineal en parámetros

S1: Lineal en parámetros

S1: Lineal en parámetros

Modelos de Regresión

- **Lineal**: $y_i = \beta_1 + \beta_2 x_i + \epsilon_i$
- **Log-log**: $ln(y_i) = \beta_1 + \beta_2 ln(x_i) + \epsilon_i$
- **Log-lineal**: $ln(y_i) = \beta_1 + \beta_2 x_i + \epsilon_i$
- **Lineal-log**: $y_i = \beta_1 + \beta_2 ln(x_i) + \epsilon_i$
- **Recíproco**: $y_i = \beta_1 + \beta_2 \frac{1}{x_i} + \epsilon_i$
- **Cuadrático**: $y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2 + \epsilon_i$
- **Interactuado**: $y_i = \beta_1 + \beta_2 x_{i1} + \beta_3 x_{i2} + \beta_4 (x_{i1} \times x_{i2}) + \epsilon_i$

En general, un modelo de regresión es no lineal cuando ni es lineal en su formación original, ni se puede convertir en un modelo lineal mediante

S2: Exogeneidad Estricta

Para muchas de las aplicaciones de la economía el supuesto más importante es la **EXOGENEIDAD**

$$E[\epsilon \mid X] = 0$$

Pero qué quiere decir?

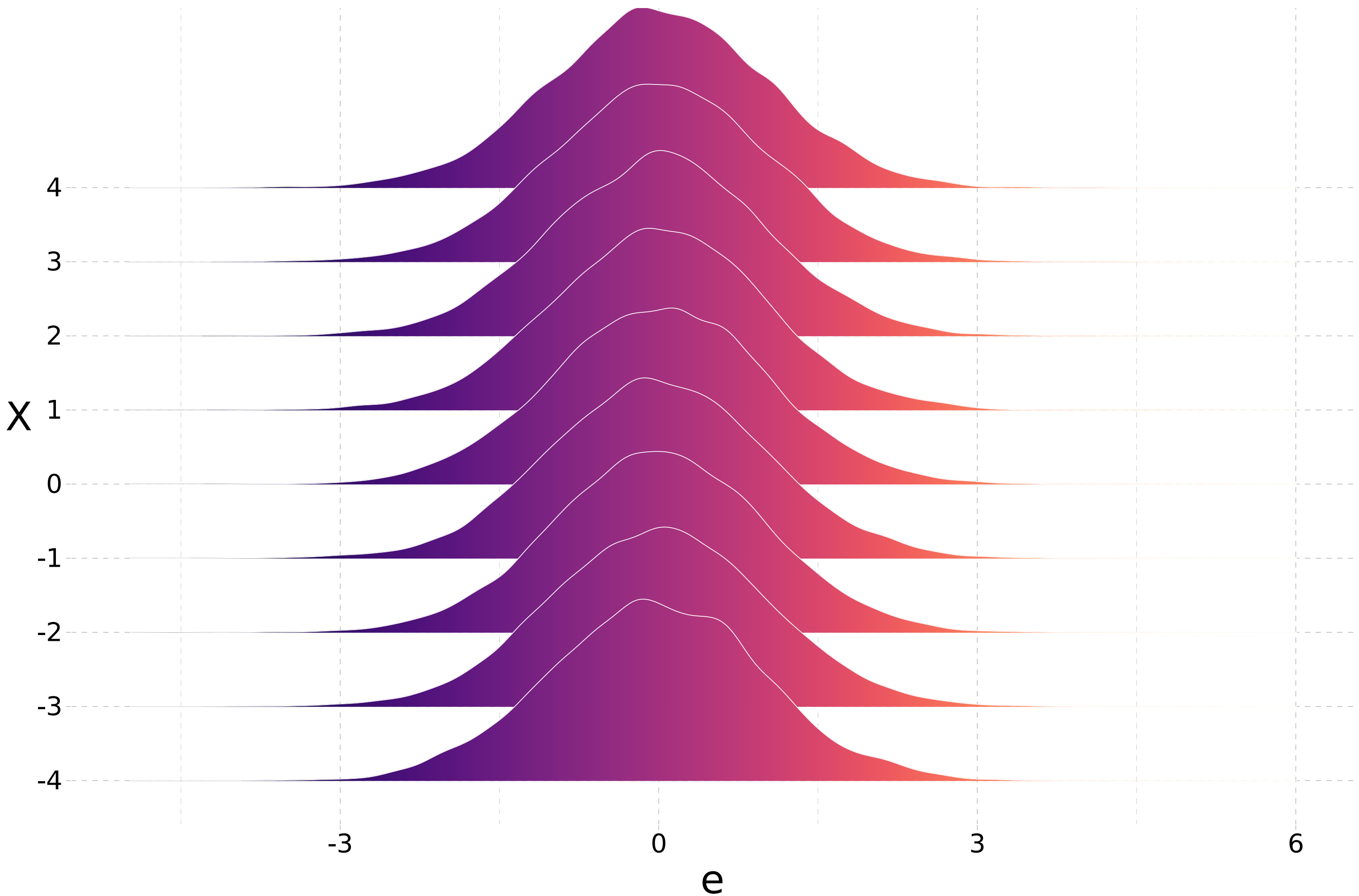
Una forma de pensar en esta definición es:

> Para *cualquier* valor de X, el valor esperado de los residuos debe ser igual a cero

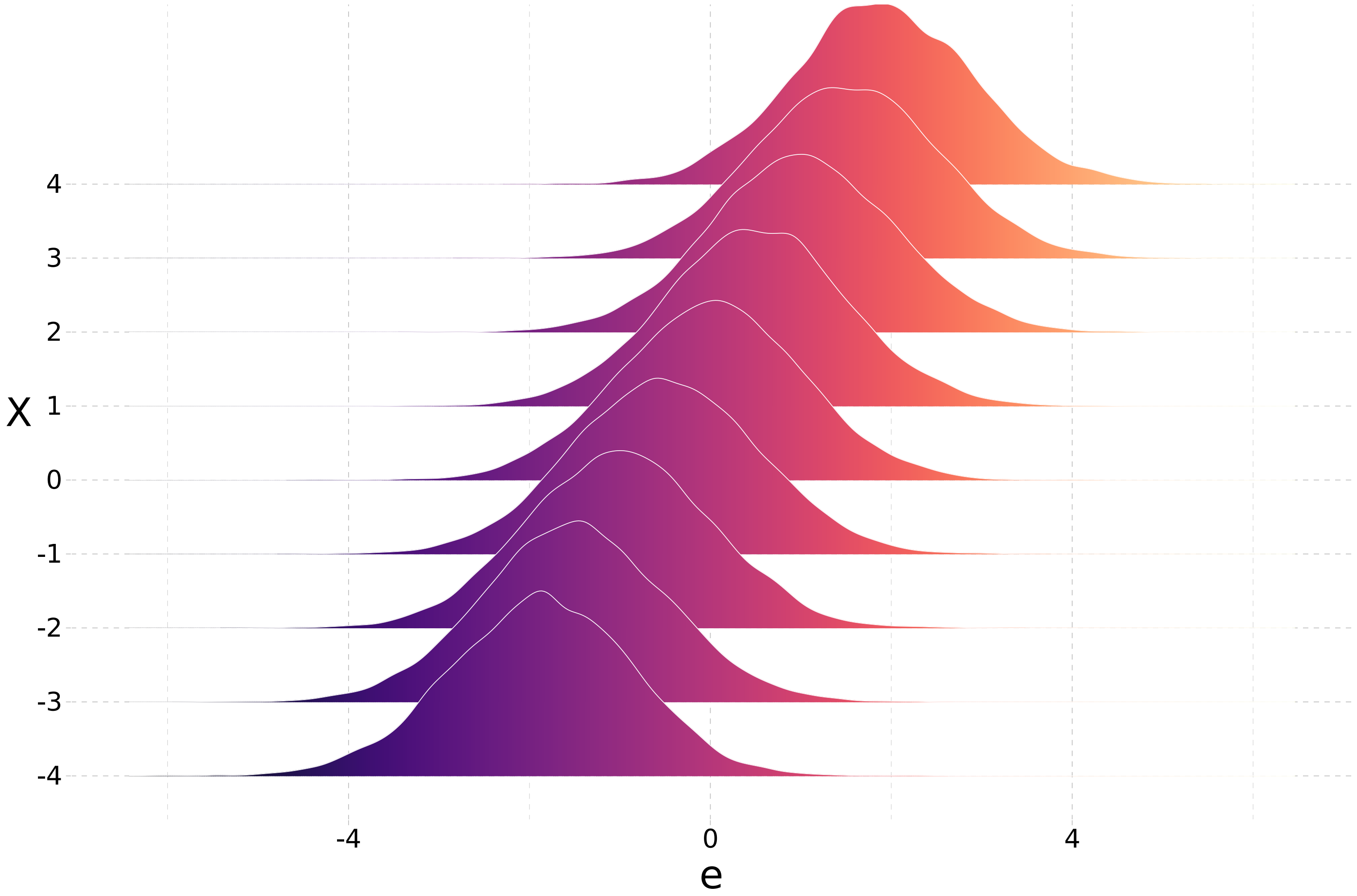
- *E.g.,* $E[u \mid X = 1] = 0$ *and* $E[u \mid X = 100] = 0$
- *E.g.,* $E[u \mid X_2 = \textbf{Mujer}] = 0$ *and* $E[u \mid X_2 = \textbf{Hombre}] = 0$
- Note: $E[u \mid X] = 0$ es más restrictivo que $E[u] = 0$

Graficamente...

Exogeneidad Estricta se cumple, $E[\epsilon \mid X] = 0$



Exogeneidad Estricta se Incumple, *i.e.*, $E[\epsilon \mid X] \neq 0$



S3: Colinealidad Imperfecta

$$X \text{ es una matriz } nxK \text{ con rango } K$$

Wooldridge (2003), este supuesto permite que las variables independientes estén correlacionadas, siempre y cuando no lo hagan de forma perfecta.

i. Si hay un vector x_k cuyos valores son iguales a cero, o cuando no existe variación en alguno de los regresores.

ii. Cuando dos variables independientes x_k son iguales.

iii. Si una variable es proporcional a otra.

iv. Cuando una variable es una combinación lineal de otras.

En estos casos específicos, el rango de la matriz X será menor que K, lo que implica que las columnas de X no son linealmente independientes.

S4: Perturbaciones Esféricas

Homocedasticidad

$$Var(\epsilon_i|X) = \sigma^2 \text{ para } i = 1, \ldots, n$$

Homocedasticidad significa que la dispersión alrededor de la recta de regresión es igual para los diversos valores de X.

No Autocorrelación

$$Cov(\epsilon_i, \epsilon_j|X) = 0 \text{ para } i \neq j$$

La no autocorrelación significa que los errores no se encuentran relacionados entre sí. La autocorrelación generalmente aparece en datos de series de tiempo aunque también puede presentarse en el caso de una muestra de corte transversal (e.g., correlación espacial).

S4: Perturbaciones Esféricas

$$
\begin{aligned}
Var(\epsilon|X) &= E[\epsilon\epsilon'|X] - E[\epsilon|X]E[\epsilon'|X] \\
&= E[\epsilon\epsilon'|X] - \underbrace{E[\epsilon|X]E[\epsilon'|X]}_{0} \text{ por supuesto A2.}
\end{aligned}
$$

$$
= E\left[\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \begin{bmatrix} \epsilon_1 & \epsilon_2 & \cdots & \epsilon_n \end{bmatrix} |X \right]
$$

S4: Perturbaciones Esféricas

$$
\begin{aligned}
Var(\epsilon|X) &= \begin{bmatrix} E[\epsilon_1\epsilon_1|X] & E[\epsilon_1\epsilon_2|X] & \cdots & E[\epsilon_1\epsilon_n|X] \\ E[\epsilon_2\epsilon_1|X] & E[\epsilon_2\epsilon_2|X] & \cdots & E[\epsilon_2\epsilon_n|X] \\ \vdots & \vdots & \ddots & \vdots \\ E[\epsilon_n\epsilon_1|X] & E[\epsilon_n\epsilon_1|X] & \cdots & E[\epsilon_n\epsilon_n|X] \end{bmatrix} \\
&= \begin{bmatrix} Var[\epsilon_1|X] & Cov[\epsilon_1\epsilon_2|X] & \cdots & Cov[\epsilon_1\epsilon_n|X] \\ Cov[\epsilon_2\epsilon_1|X] & Var[\epsilon_2|X] & \cdots & Cov[\epsilon_2\epsilon_n|X] \\ \vdots & \vdots & \ddots & \vdots \\ Cov[\epsilon_n\epsilon_1|X] & Cov[\epsilon_n\epsilon_1|X] & \cdots & Var[\epsilon_n|X] \end{bmatrix} \\
&= \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} \text{ por supuesto A4.}
\end{aligned}
$$

S4: Perturbaciones Esféricas

$$Var[\epsilon|X] = E[\epsilon\epsilon'|X] = \sigma^2 I$$

S5. Regresores no estocásticos

La implicación de este supuesto es que no hay necesidad de distinguir entre la distribución condicional del error, $f(\epsilon_i|x_1, \cdots, x_n)$, y la distribución no condicional, $f(\epsilon_i)$, por lo tanto los supuestos A.2 y A.4 pueden escribirse como:

- A.2. $E(\epsilon_i|X) = E(\epsilon_i) = 0$
- A.4. $E[\epsilon\epsilon'|X] = E[\epsilon\epsilon'] = \sigma^2 I$

S6. Normalidad

$$\epsilon | X \sim N(0, \sigma^2 I)$$

Como dice Greene (2003): el supuesto de normalidad generalmente es considerado como innecesario y posiblemente inapropiado para ser incluido al modelo de regresión. Excepto en aquellos casos en los que se asume explícitamente alguna distribución alternativa".