Mundo Social **01**, Apuntes (2025)
Disponible en https://mundo-social.com/

Modelo de Regresión Lineal Clásico

Jiampier Asis Villanueva Mas
Undergraduate Economist
Universidad Nacional Mayor de San Marcos
villanuevamasj@gmail.com

Benjamin Cano Aviles
Undergraduate Economist
Universidad Nacional Mayor de San Marcos
benjamin.cano@unmsm.edu.pe

Disponible en: https://mundosocial-peru.blogspot.com

Este documento es un apunte introductorio del curso de Econometría I, cuyo objetivo es ayudar a comprender cómo se construye y aplica el Modelo de Regresión Lineal Clásico (MRLC). Busca ofrecer una base teórica sencilla que permita entender los principales conceptos de la econometría y su aplicación práctica en el análisis de datos.

Índice:

1. Antecedentes históricos

Antes de construir el modelo de regresión lineal, haremos un repaso de su origen y entender la historia detrás de ella.

Todo se remonta en la segunda mitad del siglo XVIII, cuando astrónomos y otros científicos enfrentaban problemas para resolver mediciones geodésicas, pues disponian de más observaciones que incógnitas. Cada una de estas observaciones (por ejemplo, mediciones de la órbita del planeta) estaban sujeta a error. Por lo tanto, no se sabía qué método aplicar para asegurar que se obtuviera la órbita más exacta a partir de conjunto de datos dado. Esta situación llevó al desarrollo de métodos para ajustar datos con errores.

Boscovich (1711-1787) propuso un método para estas mediciones, que Laplace denominó "método de situación", donde minimiza la suma del valor absoluto de los errores, ofreciendo una alternativa menos sensible a errores extremos. Posteriormente, en 1805, Legendre en Nouvelles méthodes pour la détermination des orbites des comètes formalizo el método de los mínimos cuadrados que minimiza la suma de los errores al cuadrado, lo que lo hace más sensible a errores grandes, pero permite un tratamiento algebraico más sistemático.

En 1809, Gauss publicó Theoria Motus, donde indica haber usado el método desde 1795 y lo aplica al cálculo de la órbita del asteroide Ceres. Introduce además la distribución de errores que hoy se conoce como la curva de Gauss o Normal como justificación probabilística del método (Este fue el primer vínculo con la estadística).

Laplace aparece con una relación más sofisticada, indicando que cualquiera fuese la distribución de los errores de las mediciones individuales, sus promedios tenderían a la distribución Normal. Con ello, mostró que las estimaciones con el método de mínimos cuadrados tendían a la misma distribución. En 1812 publica su tratado Theorie analytique des probabilities donde sintetizó todo lo desarrollado hasta ese entonces.

En el siglo XIX, las matemáticas y la estadística comenzaron a aplicarse de forma más sistemática al estudio de la sociedad. Adolphe Quetelet introdujo la idea del "hombre promedio", vinculando las distribuciones normales de la astronomía con las ciencias sociales, y proponiendo que las leyes estadísticas podían aplicarse también a fenómenos humanos.

Los conceptos de **regresión** y **correlación** aparecieron en el estudio de Francis Galton con guisantes (1886). Dividió un lote de semillas en siete grupos de acuerdo con el tamaño y observó que los descendientes (las nuevas semillas) no eran del tamaño promedio del su grupo original. Encontró que, aunque la variabilidad (varianza) se mantenía igual, las medias de los grupos tendían a acercarse al promedio de toda la población. Asi concluyó que las medias "regresaban" hacia a la media poblacional en las generaciones posteriores ("Regresión hacia la media").

Galton tambien introdujo el concepto de **correlación**, al ubicar los puntos en un diagrama de dispersión (en este caso, las variables era el tamaño de semillas originales y descendientes) y notó no estaban distribuidos al azar, sino que tendían a agruparse en torno a una línea inclinada ("Linea de regresión"). Descubrió que si las dos variables están relacionadas, los puntos se alinean más estrechamente en torno a esa línea. A partir de ese análisis, propuso un coeficiente de correlación para medir el grado de asociación entre variables.

Karl Pearson, discípulo de Galton, formalizó el concepto matemático de correlación, donde el resultado siempre queda acotado entre -1 y +1, y extendió la técnica de regresión a un marco estadístico más general.

A inicios del siglo XX, Ronald A. Fisher (1922) incorporó la teoría de la inferencia estadística, permitiendo que los modelos de regresión pudieran ser evaluados con base en pruebas de hipótesis y niveles de significancia. Esto hizo posible no solo describir relaciones entre variables, sino también realizar inferencias válidas sobre poblaciones.

En el campo de la economía, la regresión fue adoptada y desarrollada por pioneros como Ragnar Frisch, Jan Tinbergen y Trygve Haavelmo, quienes establecieron las bases de la econometría. De hecho, Haavelmo (1944) fue quien introdujo rigurosamente la probabilidad en el análisis económico, integrando la teoría estadística con los modelos econométricos.

2. Nociones previas

2.1. Álgebra Lineal

2.2.1. Matriz:

Es un arreglo rectangular numérico.
La cual se representa mediante su dimensión dada por n (filas) y m (columnas)
Dada una matriz A de orden $m \times n$ es un conjunto de elementos a_{ij} con $i = 1, 2, \ldots, m$ y $j = 1, 2, \ldots, n$

Notación general de una matriz:

$$A = [a_{ij}] = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

Ejemplo: Matriz 2×3

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} = \begin{pmatrix} 3 & -1 & 5 \\ 0 & 4 & 2 \end{pmatrix}$$

Donde:

- **Fila 1:** $\begin{pmatrix} 3 & -1 & 5 \end{pmatrix}$
- **Fila 2:** $\begin{pmatrix} 0 & 4 & 2 \end{pmatrix}$
- **Columna 1:** $\begin{pmatrix} 3 \\ 0 \end{pmatrix}$
- **Columna 2:** $\begin{pmatrix} -1 \\ 4 \end{pmatrix}$
- **Columna 3:** $\begin{pmatrix} 5 \\ 2 \end{pmatrix}$

2.2.2. Matriz transpuesta (A' o A^T):

Se define como el intercambio de filas por columnas dentro de una matriz, es decir, el elemento de la fila i y columna j de A se convierte en el elemento en la fila j y columna i de A^T

$$(A^T)_{ij} = a_{ji}$$

Ejemplo:

$$M_{2x3} = \begin{pmatrix} 2 & 7 & 4 \\ 3 & 5 & 9 \end{pmatrix} \rightarrow \begin{pmatrix} 2 & 3 \\ 7 & 5 \\ 4 & 9 \end{pmatrix} = M_{3x2}^T$$

Propiedades:

a) $(A')' = A$
b) $(\alpha A)' = \alpha A'$
c) $(A + B)' = A' + B'$
d) $(AB)' = B'A'$
e) $(ABC)' = (A(BC))' = (BC)'A' = C'B'A'$

OJO: **Matriz simétrica**: $A' = A$.

2.1.3. Traza de una matriz:

Se define como

$$\text{Tr}(A) = a_{11} + a_{22} + \cdots + a_{mm} = \sum_{i=1}^{m} a_{ii}$$

Propiedades:

a) $\text{Tr}(A) = \text{Tr}(A')$
b) $\text{Tr}(A + B) = \text{Tr}(A) + \text{Tr}(B)$
c) $\text{Tr}(AB) = \text{Tr}(BA)$

2.1.4. Matriz identidad ($\mathbf{I}_n$):

Es la matriz cuadrada de dimensión n con unos en la diagonal principal y ceros en las demás posiciones.

$$\mathbf{I}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Cumple que $\mathbf{AI} = \mathbf{IA} = \mathbf{A}$ para cualquier matriz $\mathbf{A}$ de dimensión compatible.

2.1.5. Matriz inversa:

($\mathbf{A}^{-1}$): es la matriz que satisface

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

Además:

$$A^{-1} = \frac{1}{\det(A)} \text{Adj}(A)$$

donde $\text{Adj}(A) = C^T$ es la traspuesta de la matriz de cofactores.

Ejemplo:

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{A}^{-1} = \begin{pmatrix} 0.5 & -0.5 \\ 0 & 1 \end{pmatrix}$$

En el MRLC, la condición de existencia de $(\mathbf{X}'\mathbf{X})^{-1}$ es fundamental para poder calcular $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.

Propiedades:

a) $(A^{-1})^{-1} = A$
b) $(AB)^{-1} = B^{-1}A^{-1}$
c) $(A')^{-1} = (A^{-1})'$

2.1.6. Determinante de una matriz:

El determinante de una matriz cuadrada es un número asociado que refleja información clave sobre su estructura. Se define como:

$$|A| = \det(A) = \sum_{j=1}^{m} a_{ij}(-1)^{i+j} M_{ij} = \sum_{j=1}^{m} a_{ij} C_{ij}$$

Donde:
M_{ij} es la submatriz de orden $(m-1) \times (m-1)$.
C_{ij} el **cofactor** del elemento a_{ij}.

Si $\det(\mathbf{A}) = 0$, la matriz es singular y no tiene inversa.
Si $\det(\mathbf{A}) \neq 0$, la matriz es no singular y sí tiene inversa.

Propiedades:

a) $|AB| = |A||B|$
b) $|A'| = |A|$
c) $|\alpha A| = \alpha^m |A|$
d) $|I| = 1$
e) $|A^{-1}| = |A|^{-1}$

Ejemplo:

$$\det \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = (1)(4) - (2)(3) = -2 \neq 0$$

En el MRLC, que $\det(\mathbf{X}'\mathbf{X}) \neq 0$ significa que los regresores no son combinación lineal exacta.

2.1.7. Rango de una matriz:

El rango de una matriz es el número máximo de columnas (o filas) linealmente independientes.
Si el rango de $\mathbf{X}$ es igual al número de columnas, entonces $\mathbf{X}$ tiene rango completo.
Si no, existe multicolinealidad perfecta y el modelo no puede estimarse.

Ejemplo:

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$$

Aquí la segunda columna es el doble de la primera $\rightarrow$ rango $= 1$ (no completo).
En el MRLC, esto implica que no podemos distinguir el efecto de cada variable porque son perfectamente dependientes.

2.1.8. Espacio columna:

El espacio columna de una matriz $\mathbf{X}$ es el conjunto de todas las combinaciones lineales de sus columnas.
Intuitivamente, es el "espacio" que abarcan las variables explicativas.
En el MRLC, los valores ajustados $\hat{\mathbf{y}}$ son la proyección ortogonal de $\mathbf{y}$ sobre el espacio columna de $\mathbf{X}$.
Por ejemplo: Si

$$\mathbf{X} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$$

El espacio columna está generado por las columnas: $\{(1,0,1)', (0,1,1)'\}$. Cualquier combinación de esas columnas está dentro del espacio columna.
En regresión, esto explica por qué $\hat{\mathbf{y}}$ está dentro del subespacio generado por los regresores.

2.1.9. Relación con la proyección en regresión

En relación a lo anterior, podemos definir la matriz de proyección:

$$\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

se cumple que:

$$\hat{\mathbf{y}} = \mathbf{P}\mathbf{y} \quad \text{(proyección de } \mathbf{y} \text{ en el espacio columna de } \mathbf{X})$$

Los residuos son $\hat{\mathbf{u}} = (\mathbf{I} - \mathbf{P})\mathbf{y}$, ortogonales a los regresores.
Esto es clave porque muestra que la regresión lineal es, en esencia, un problema de proyecciones ortogonales en álgebra lineal.

2.2. Estadística

El Modelo de Regresión Lineal Clásica (MRLC) se fundamentan en conceptos estadísticos esenciales.

2.2.1. Variables Aleatorias y Distribuciones

Una **variable aleatoria (VA)** es una función que asigna valores numéricos a los resultados de un experimento aleatorio.

- **VA Discreta**: Toma valores en un conjunto finito o infinito numerable. Ejemplos: Bernoulli, Binomial, Poisson.
- **VA Continua**: Toma valores en un intervalo real. Ejemplos: Uniforme, Exponencial, Normal.

La **función de densidad de probabilidad (fdp)** $f(x)$ para una VA continua satisface:

$$\int_{-\infty}^{\infty} f(x)\,dx = 1, \quad P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

La **función de distribución acumulativa (FDA)** se define como:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\,du$$

y verifica:

$$\frac{d}{dx}F(x) = f(x)$$

2.2.2. Esperanza Matemática y Varianza

El **valor esperado** o **esperanza** de una VA X es un promedio ponderado de sus valores posibles.

- Para VA discreta con función de masa $p(x)$:

$$E(X) = \sum_i x_i p(x_i)$$

- Para VA continua con densidad $f(x)$:

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$$

Propiedad clave:

$$E(aX + b) = aE(X) + b$$

La **varianza** mide la dispersión de X alrededor de su media:

$$\text{Var}(X) = E\left[(X - E(X))^2\right] = E(X^2) - [E(X)]^2$$

Propiedad:

$$\text{Var}(aX) = a^2\text{Var}(X)$$

La **desviación estándar** es $\sigma = \sqrt{\text{Var}(X)}$.

2.2.3. Covarianza y Correlación

La **covarianza** entre dos VA X e Y mide su asociación lineal:

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - E(X)E(Y)$$

Si X e Y son independientes, $\text{Cov}(X, Y) = 0$.

El **coeficiente de correlación** estandariza la covarianza:

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}, \quad -1 \leq \rho_{XY} \leq 1$$

2.2.4. Esperanza Condicional

La **esperanza condicional** $E(Y|X)$ es el valor esperado de Y dado X. Se cumple:

- **Ley de esperanzas iteradas**:
$$E(Y) = E[E(Y|X)]$$
- **Descomposición de la varianza**:
$$\text{Var}(Y) = \text{Var}[E(Y|X)] + E[\text{Var}(Y|X)]$$

2.2.5. Muestreo y Estimadores

Un **estimador** es una regla para calcular un parámetro poblacional a partir de una muestra. La **media muestral**:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

es un estimador de $\mu = E(X)$.
Propiedades deseables:

- **Insesgamiento**: $E(\hat{\theta}) = \theta$
- **Consistencia**: $\hat{\theta} \to \theta$ en probabilidad cuando $n \to \infty$
- **Eficiencia**: Mínima varianza entre estimadores insesgados

El **error cuadrático medio (MSE)** mide la calidad global:

$$\text{MSE}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = \text{Var}(\hat{\theta}) + [\text{Sesgo}(\hat{\theta})]^2$$

Nota: ¿Cómo surge el Error Cuadrático Medio?

Cuando evaluamos un estimador, buscamos que nos dé valores cercanos al parámetro. Por ello, usamos la esperanza porque nos interesa el valor promedio del estimador sobre todas las posibles muestras.
Sin embargo, que un estimador sea **insesgado**

$$E(\hat{\theta}) = \theta$$

no basta, porque puede tener una **varianza grande**, de modo que los resultados en distintas muestras pueden estar muy lejos del parámetro verdadero.
Por otro lado, un estimador puede tener una **varianza pequeña**, pero presentar la **presencia de sesgo**.
Entonces surge el **Error Cuadrático Medio (ECM)** como una medida de integrar ambos aspectos, teniendo como definición:

$$ECM(\hat{\theta}) = E\big[(\hat{\theta} - \theta)^2\big]$$

ya que evalúa la desviación del estimador respecto al parámetro real.
Por lo tanto, al descomponer obtenemos:

$$ECM(\hat{\theta}) = \text{Var}(\hat{\theta}) + \big(E(\hat{\theta}) - \theta\big)^2,$$

lo cual combina **precisión** (varianza pequeña) y **exactitud** (ausencia de sesgo).

2.2.6. Teorema del Límite Central (TLC)

Si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria de una v.a. X i.i.d.
(Independientes Identicamente Distribuidas) tales que:

$$\mathbb{E}(X_i) < \infty, \quad \forall i, \quad \mathbb{E}(X_i) = \mu, \quad \mathbb{V}(X_i) = \sigma^2,$$

es decir, tienen segundo momento finito para que existan $\mathbb{E}(X_i)$ y $\mathbb{V}(X_i)$.

Entonces:

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1)$$

y de manera equivalente:

$$\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{(Normal no estandarizada)}.$$

Obs: ¿n grande?
De manera práctica, se suele considerar que $n \geq 30$.

OJO: Esto justifica la normalidad asintótica de los estimadores MCO en regresión.

Nota: ¿Porqué usamos el TLC?

Cuando no podemos observar toda la población, trabajamos con **muestras aleatorias**. A partir de ellas, calculamos *estadísticos* (como $\overline{X}$) para estimar los parámetros (*estimadores*).

Al tomar muchas muestras de tamaño n y calcular $\overline{X}$ de cada una, obtenemos la **distribución de la media muestral**.

El **TLC** establece que dicha distribución se aproxima a una normal estándar, independientemente de la forma de la distribución poblacional, siempre que ésta tenga media y varianza finita.

Conforme el tamaño de la muestra n **crece**, la forma de la distribución muestral de $\overline{X}$ se va pareciendo cada vez más a una normal.

Entonces, aunque solo tengamos una muestra, gracias al **TLC** podemos ubicar su media muestral dentro de la distribución de la media muestral *teórica* y construir **intervalos de confianza** para realizar **pruebas de hipótesis**.

2.2.7. Distribuciones Continuas Conjuntas

Para dos VA continuas X e Y, la **función de densidad conjunta** $f_{X,Y}(x,y)$ satisface:

$$P((X,Y) \in A) = \iint_A f_{X,Y}(x,y)\,dx\,dy$$

Las **densidades marginales** son:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$$

X e Y son independientes si:

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

2.2.8. Función Generatriz de Momentos (FGM)

La FGM de una VA X se define como:

$$M_X(t) = E(e^{tX})$$

¿Qué es momento? Mide cómo se distribuye la masa de probabilidad alrededor de un punto (normalmente el 0 o la **media**).

Expansión de la exponencial con series de Taylor:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

Entonces:

$$e^{tX} = 1 + tX + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \dots$$

$$M_X(t) = E\left[1 + tX + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \dots\right]$$

$$M_X(t) = 1 + tE[X] + \frac{t^2}{2!}E[X^2] + \frac{t^3}{3!}E[X^3] + \dots$$

- 1er momento: $E[X]$
- 2do momento: $E[X^2]$
- 3er momento: $E[X^3]$

Derivamos para extraer los momentos respecto a t:

$$M'_X(0) = E(X) = \mu$$

$$M''_X(0) = E(X^2)$$

$$M'''_X(0) = E(X^3)$$

En general:

$$M_X^{(k)}(0) = E[X^k]$$

t: Número que sirve como parámetro en la función generadora, permite codificar todos los momentos de X dentro de una sola función.

OJO: Sólo me muestra la dispersión de los valores respecto a 0, qué tan "grandes" son los valores de X, pero no toma en cuenta dónde está la media.

Momentos Centrales (respecto a la media)

$$\mu_k = E\left[(X - \mu)^k\right]$$

$$\mu_1 = E[X - \mu] = E[X] - \mu = \mu - \mu = 0$$

$$\mu_2 = E\left[(X - \mu)^2\right] \quad \rightarrow \text{Me muestra la dispersión de los valores respecto a la media.}$$

$$\mu_3 = E\left[(X - \mu)^3\right]$$

Útil para demostrar propiedades de distribuciones y teoremas límite.

2.2.9. Pruebas de Hipótesis: t, χ^2, F

Estas pruebas derivan de la distribución normal:

- **Prueba** t: Para contrastar medias.
$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$
- **Prueba** χ^2: Para varianzas y bondad de ajuste.
$$\chi^2 = \sum_{i=1}^{k} Z_i^2 \sim \chi_k^2, \quad Z_i \sim N(0,1)$$
- **Prueba** F: Para comparar varianzas.
$$F = \frac{U/d_1}{V/d_2} \sim F_{d_1,d_2}, \quad U \sim \chi_{d_1}^2, V \sim \chi_{d_2}^2$$

2.3. Terminología de una regresión

- **Variable dependiente:** Variable que se intenta predecir o explicar.
- **Variable independiente:** Variable(s) que se utiliza(n) para predecir o explicar la variable dependiente.
- **Coeficientes:** Parámetros que indican la relación entre las variables independientes y la dependiente.
- **Intercepto:** Valor de la variable dependiente cuando todas las independientes son cero.
- **Error estándar:** Medida de la precisión de la estimación de los coeficientes.
- **Valor p:** Probabilidad de obtener resultados al menos tan extremos como los observados, asumiendo que la hipótesis nula es cierta.
- **R-cuadrado:** Proporción de la varianza de la variable dependiente que es explicada por el modelo.
- **Residuales:** Diferencias entre los valores observados y los valores predichos por el modelo.

2.3.1. Regresión Simple (Población)

$$y = \beta_0 + \beta_1 x + u$$

donde y es la variable dependiente/explicada/endógena, x la independiente/explicativa/exógena, y u resume variables no observadas que afectan a y. "Independiente" aquí es terminología: no significa independencia estadística entre v.a. (cuidado con la confusión).

Interpretación *ceteris paribus*: si los demás factores (en u) no cambian,

$$\frac{\partial E[y \mid x]}{\partial x} = \beta_1.$$

Esta es la lectura "marginal" que perseguimos empíricamente.

Objetivo: estudiar relación causa-efecto (cuando podemos defender exogeneidad) o, al menos, medir asociación estructurada entre X e Y.

3. Fundamentos del Modelo de Regresión Lineal Clásico

3.1. Definición

El MRLC es un modelo estadístico que busca representar la relación entre una variable dependiente/endógena y y un conjunto de variables explicativas/exógenas x_j. La forma general es:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + u_i$$

donde:

- y_i: variable dependiente en la observación i
- x_{ji}: valor de la variable independiente j en la observación i
- β_j: parámetros poblacionales que miden el efecto de cada x_j sobre y
- u_i: término de error, que captura los factores no observados

En el caso de un solo regresor:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

3.2. Función de Regresión Poblacional (FRP)

En la práctica, es muy dificil obtener los datos de una población (implica un mayor costo). Por ello, se trabaja a nivel muestral reconociendo que esta proviene de una población de la cual no tenemos información completa. Se asume que una variable **Y** (característica de la población), puede ser explicada por variables explicativas **X**. Esto se representa teóricamente mediante la **función de regresión poblacional**.

Tambien llamado función de esperanza condicional. Nos indica que el valor esperado de y_i dado x_i se relaciona funcionalmente x_i. En otras palabras, dice cómo la media o respuesta promedio de y varía con x.

$$E(y_i \mid x_i) = \beta_0 + \beta_1 x_i$$

En la realidad, hay factores que no incluimos en el modelo, variables omitidas, azar o variabilidad natural. Para recoger todos estos efectos **no observados**, se introduce el **término de perturbación** o **error**.

Esto implica que β_1 mide el cambio promedio en y asociado a una variación de una unidad en x, manteniendo constantes las demás variables. La regresión, entonces, busca estimar esta relación sistemática, diferenciándola de la parte aleatoria (u_i).

3.3. Función de regresión muestral (FRM)

Se utiliza para estimar la **FRP** en base a información muestral. Se supone la existencia de una línea de regresión poblacional, pero, debido a **fluctuaciones muestrales**, las estimaciones obtenidas constituyen, en el mejor de los casos, sólo una aproximación de la verdadera recta poblacional.

La línea de regresión muestral se denota con:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

En esta función, también se integra una parte aleatoria, denominado **residuo muestral**, que representa la diferencia entre lo observado y lo estimado por la recta de la muestra ($\hat{u}_i$).

Por lo tanto, el objetivo principal del análisis de regresión es estimar la FRP ($Y_i = \beta_0 + \beta_1 x_i + u_i$) a partir de la FRM ($\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$).

3.4. Forma matricial

La representación matricial en el modelo de regresión lineal clásico nos facilita los calculos y la comprensión de las estimaciones, propiedades y pruebas de hipótesis. Sabemos que:

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$$

donde y_i representa la variable dependiente, X_{ki} son las variables explicativas, β_k los coeficientes a estimar y u_i el término de perturbación. Ahora, expandiendo este modelo para n observaciones, se obtiene:

$$\begin{aligned}
y_1 &= \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} + \cdots + \beta_k X_{k1} + u_1 \\
y_2 &= \beta_0 + \beta_1 X_{12} + \beta_2 X_{22} + \cdots + \beta_k X_{k2} + u_2 \\
\vdots \quad & \qquad \vdots \qquad \vdots \qquad\qquad \vdots \quad \vdots \\
y_n &= \beta_0 + \beta_1 X_{1n} + \beta_2 X_{2n} + \cdots + \beta_k X_{kn} + u_n
\end{aligned}$$

Podemos expresar este sistema de ecuaciones en forma matricial como:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \beta_0 + \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1n} \end{bmatrix} \beta_1 + \cdots + \begin{bmatrix} x_{k1} \\ x_{k2} \\ \vdots \\ x_{kn} \end{bmatrix} \beta_k + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

Notamos que los coeficientes $\beta_0, \beta_1, \ldots, \beta_k$ son los mismos en todas las ecuaciones, lo que varía entre observaciones son los valores de las variables explicativas X_{ji} y los términos de error u_i. Por tanto, podemos agrupar los términos de manera que cada observación se exprese como un producto entre un vector de regresores y un vector de parámetros.

Para el coeficiente β_0, esta se encuentra multiplicado por 1 en todas las observaciones, por lo que incorporamos una columna de unos para representar el intercepto dentro de la matriz de regresores. De esta forma, quedaría:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

Entonces, para n observaciones y k regresores, el modelo se expresa en notación compacta:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

donde:

- $\mathbf{y}$ es un vector $n \times 1$ de la variable dependiente
- $\mathbf{X}$ es la matriz $n \times k$ de variables explicativas (con una columna de unos para el intercepto)
- $\boldsymbol{\beta}$ es el vector de parámetros de dimensión $k \times 1$
- $\mathbf{u}$ es el vector $n \times 1$ de perturbaciones aleatorias

Esta notación nos va a permitir usar herramientas algebraicas para derivar los estimadores y propiedades del modelo.

3.5. Estimación por Mínimos Cuadrados Ordinarios (MCO)

El principio fundamental para estimar los coeficientes $\beta_0, \beta_1, \ldots, \beta_k$ es el criterio de mínimos cuadrados, que busca minimizar la discrepancia entre los valores observados y los valores ajustados.

$$S(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} e_i^2 = \mathbf{e}'\mathbf{e} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

Nota: ¿Porqué decimos u′u?

Se utiliza $\mathbf{e}'\mathbf{e}$ porque es una forma matricial de representarlo que nos va a servir para simplificar el proceso.

Sabemos que $\mathbf{e}$ es un vector columna $n \times 1$. Al transponerlo ($\mathbf{e}'$) obtenemos un vector fila de $1 \times n$, y el producto $\mathbf{e}'\mathbf{e}$ resulta en un escalar 1×1, que corresponde a la suma de los cuadrados de los elementos de $\mathbf{e}$:

$$\mathbf{e}'\mathbf{e} = [\mathbf{e}']_{1\times n}\,[\mathbf{e}]_{n\times 1} = [escalar]_{1\times 1}$$

Si en cambio multiplicáramos $\mathbf{e}\mathbf{e}'$, obtendríamos una matriz de dimensión $n \times n$, lo que no tiene sentido para el criterio de minimización de MCO.

Entonces tenemos que resolver:

$$\min_{\hat{\boldsymbol{\beta}}} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

Expandiendo la operación:

$$\min_{\hat{\boldsymbol{\beta}}} [(\mathbf{y}' - \hat{\boldsymbol{\beta}}'\mathbf{X}')(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})]$$

$$\min_{\hat{\boldsymbol{\beta}}} [\ \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}\ ]$$

Observación:

Al analizar los componentes:

$$\mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}} = [\mathbf{y}']_{1\times n}\,[\mathbf{X}]_{n\times k}\,[\hat{\boldsymbol{\beta}}]_{k\times 1} = [.]_{1\times k}[\hat{\boldsymbol{\beta}}]_{k\times 1} = [escalar]_{1\times 1}$$

$$\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = [\hat{\boldsymbol{\beta}}']_{1\times k}\,[\mathbf{X}']_{k\times n}\,[\mathbf{y}]_{n\times 1} = [.]_{1\times n}[\mathbf{y}]_{n\times 1} = [escalar]_{1\times 1}$$

Tenemos que ambos son escalares, entonces podemos decir que:

$$\mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}} = (\mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}})' = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$$

Simplificando la expresión:

$$S(\hat{\boldsymbol{\beta}}) = Y'Y - 2Y'X\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}'X'X\hat{\boldsymbol{\beta}}$$

Para encontrar el valor de $\hat{\boldsymbol{\beta}}$ que minimiza $S(\hat{\boldsymbol{\beta}})$, derivamos con respecto a $\hat{\boldsymbol{\beta}}$ y igualamos a cero.
Las derivadas de cada término son:

$$\frac{\partial(Y'Y)}{\partial\hat{\boldsymbol{\beta}}} = 0, \quad \frac{\partial(-2Y'X\hat{\boldsymbol{\beta}})}{\partial\hat{\boldsymbol{\beta}}} = -2X'Y, \quad \frac{\partial(\hat{\boldsymbol{\beta}}'X'X\hat{\boldsymbol{\beta}})}{\partial\hat{\boldsymbol{\beta}}} = 2X'X\hat{\boldsymbol{\beta}}$$

Por tanto:

$$\frac{dS(\hat{\boldsymbol{\beta}})}{d\hat{\boldsymbol{\beta}}} = -2X'Y + 2X'X\hat{\boldsymbol{\beta}} = 0$$

Despejando $\hat{\boldsymbol{\beta}}$:

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$$

$$\therefore\ \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

4. Supuestos del MRLC

Los supuestos del Modelo de Regresión Lineal Clásico (MRLC) establecen las condiciones bajo las cuales los estimadores de Mínimos Cuadrados Ordinarios (MCO) poseen propiedades deseables como insesgamiento, eficiencia y validez de la inferencia estadística. El cumplimiento de estos supuestos garantiza que el Teorema de Gauss-Markov se aplique, haciendo que MCO sea el Mejor Estimador Lineal Insesgado (MELI).

4.1. Linealidad en Parámetros

Establece que el MRLC debe ser lineal en los parámetros, es decir, que los coeficientes β_j aparecen de forma lineal. Se permiten transformaciones de variables (como $\ln y$, X^2, o términos de interacción) siempre que el modelo mantenga linealidad en los parámetros.

Veamos algunos ejemplos:

1. $Y = \beta_0 + \beta_1 X + \beta_2 X + u \longrightarrow$ **Sí cumple**
2. $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + u \longrightarrow$ **Sí cumple**
3. $Y = \beta_0 (\beta_1 X_1 + \beta_2 X_2) + u \longrightarrow$ **No cumple**
4. $Y = \beta_0 + ln(\beta_1) X + \beta_2 X + u \longrightarrow$ **No cumple**
5. Modelo Coob-Douglas

   $Y = AK^{\beta_1} L^{\beta_2} e^u$

   A primera vista, el modelo parece *no lineal*, ya que los parámetros β_1 y β_2 se encuentran como exponentes de las variables K (capital) y L (trabajo). Sin embargo, al aplicar logaritmos naturales a ambos lados de la ecuación, se obtiene:

   $\ln(Y) = \ln(A) + \beta_1 \ln(K) + \beta_2 \ln(L) + u$

   Si definimos $\beta_0 = \ln(A)$, el modelo transformado adopta la forma:

   $\ln(Y) = \beta_0 + \beta_1 \ln(K) + \beta_2 \ln(L) + u \longrightarrow$ **Sí cumple**

Nota:

No existe una demostración matemática adicional más allá del enunciado, ya que se trata de una condición sobre la forma funcional del modelo. En esencia, la elección de la forma funcional es una cuestión empírica y depende del comportamiento de los datos y del sustento teórico.

Por ejemplo, un economista podría sostener que el consumo mantiene una relación lineal con el ingreso, lo que justifica la adopción de una función lineal como punto de partida o hipótesis de trabajo. Bajo esta premisa, se asume que el modelo es lineal en los parámetros.

Además, esta especificación no solo facilita el tratamiento matemático, sino que también permite una interpretación más sencilla y directa de los coeficientes estimados.

4.2. Regresores no Estocásticos

Establece que las variables explicativas $X_1, X_2, \ldots, X_k$ se consideran fijas y conocidas para todos los valores de la muestra. En otras palabras, los regresores son **determinísticos**, es decir, no presentan variabilidad aleatoria en muestras repetidas.

Este supuesto es fundamental para garantizar que el estimador de Mínimos Cuadrados Ordinarios (MCO) sea **insesgado y consistente**. Si las variables explicativas fueran estocásticas, es decir, dependieran de factores aleatorios que también influyen en la variable dependiente Y—, entonces los errores u podrían estar correlacionados con los regresores X. En tal caso, los estimadores $\hat{\beta}$ resultarían **sesgados e inconsistentes**.

Nota:

Asumir $\mathbf{X}$ fijo simplifica las demostraciones, aunque no es estrictamente necesario. Si $\mathbf{X}$ es aleatorio pero se cumple $E[\mathbf{u} \mid \mathbf{X}] = 0$, entonces $E[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$ por la ley de esperanzas iteradas.

La suposición de "$\mathbf{X}$ fijo" es un artificio matemático que facilita el cálculo de varianzas y derivaciones.

4.3. Exogeneidad / Independencia

Significa que, condicionalmente en los regresores, el término de error tiene media cero. No existe correlación sistemática entre X y los factores no observados contenidos en u. Esta es la condición central que garantiza el insesgamiento de los estimadores MCO.

$$E[u_i \mid X_i] = 0 \quad \text{para todo } i$$

Consecuencias directas

a) $E[u_i] = E(E[u_i \mid X_i]) = E(0) = 0 \longrightarrow E[u_i] = 0$ (Error tiene media 0)

b) $E[X_i u_i] = E(E[X_i u_i \mid X_i]) = E(X_i E[u_i \mid X_i]) = E(0) = 0 \longrightarrow E[X_i u_i] = 0$ (Ortogonalidad)

c) $\text{Cov}(X_i, u_i) = 0$

Ley de Expectativas Iteradas

También conocida como **Ley de la Esperanza Total**, establece que:

$$E[Y] = E\big(E[Y \mid X]\big)$$

Donde $E[Y \mid X]$ es la esperanza condicional de Y dado X, es decir, la expectativa de Y cuando conocemos el valor de X.

La ley dice que la expectativa total de Y es igual a la expectativa de la expectativa condicional de Y, dado X. En términos simples, equivale a decir que el **promedio total** de Y es igual al **promedio de los promedios** de Y dentro de cada grupo definido por X.

Por ejemplo, si el ingreso promedio (Y) depende del nivel educativo (X). Podemos dividir la población en tres grupos según X: Personas con educación básica, con educación media y con educación universitaria.

Calculamos el promedio del ingreso dentro de cada grupo:

$E[Y \mid X = \text{básica}] = 1000 \qquad E[Y \mid X = \text{media}] = 2000] \qquad E[Y \mid X = \text{universitaria}] = 4000$

y las proporciones son 0.4, 0.4 y 0.2 respectivamente, entonces:

$$E[Y] = E\big(E[Y \mid X]\big) = (0.4)(1000) + (0.4)(2000) + (0.2)(4000) = 2000$$

Por tanto, el ingreso de toda la población es $E[Y] = 2000$

4.4. Perturbaciones Esféricas o Residuos Esféricos

A) Homocedasticidad

Establece que la varianza del término de error u es constante para todas las observaciones del modelo. Matemáticamente se expresa como:

$$\text{Var}(u_i \mid X_i) = \sigma^2 \quad \text{para todo } i$$

Esto significa que los errores u_i deben presentar la misma dispersión, sin depender del valor de las variables explicativas X. En otras palabras, el grado de variabilidad del error debe ser el mismo en todos los niveles de X.

La homocedasticidad garantiza que los estimadores de MCO sean **eficientes** en el sentido de que poseen la menor varianza posible entre todos los estimadores lineales insesgados (Teorema de Gauss–Markov). Cuando el supuesto se cumple, los intervalos de confianza y las pruebas estadísticas se interpretan correctamente.

B) No autocorrelación

Establece que los errores del modelo no deben estar correlacionados entre sí. Formalmente:

$$\text{Cov}(u_i, u_j \mid \mathbf{X}) = 0 \quad \text{para } i \neq j$$

Los errores de distintas observaciones deben ser **independientes**. Si existe correlación entre ellos, se dice que hay **autocorrelación**, lo cual suele aparecer en series de tiempo o en modelos con estructura secuencial.

Juntos implican que la matriz de varianza-covarianza condicional de $\mathbf{u}$ es $\sigma^2\mathbf{I}_n$.

Nota: Demostración de la varianza de $\hat{\beta}$

Partimos de:

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$$

Entonces (condicional en $\mathbf{X}$):

$$\text{Var}(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\text{Var}(\mathbf{u} \mid \mathbf{X})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

Bajo el supuesto esférico $\text{Var}(\mathbf{u} \mid \mathbf{X}) = \sigma^2\mathbf{I}_n$:

$$\text{Var}(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\sigma^2\mathbf{I}_n)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

Nota sobre heterocedasticidad o autocorrelación

Si $\text{Var}(\mathbf{u} \mid \mathbf{X}) = \boldsymbol{\Omega}$ no es diagonal, entonces:

$$\text{Var}(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

y la fórmula $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ no es válida. En estos casos se utilizan errores estándar robustos.

Nota: Obtención de la Matriz de Varianzas y Covarianzas

$$\text{MNC} = E[uu'] = E\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} = E\begin{bmatrix} u_1^2 & u_1u_2 & \cdots & u_1u_n \\ u_2u_1 & u_2^2 & \cdots & u_2u_n \\ \vdots & \vdots & \ddots & \vdots \\ u_nu_1 & u_nu_2 & \cdots & u_n^2 \end{bmatrix}$$

$$= \begin{bmatrix} E(u_1^2) & E(u_1u_2) & \cdots & E(u_1u_n) \\ E(u_2u_1) & E(u_2^2) & \cdots & E(u_2u_n) \\ \vdots & \vdots & \ddots & \vdots \\ E(u_nu_1) & E(u_nu_2) & \cdots & E(u_n^2) \end{bmatrix}$$

Covarianza

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$$

$$E[(u_i - E(u_i))(u_j - E(u_j))] = E(u_iu_j) \Rightarrow \text{Cov}(u_i, u_j) = E(u_iu_j) = 0$$

Ausencia de Autocorrelación

$$E[uu'] = \begin{bmatrix} E(u_1^2) & 0 & \cdots & 0 \\ 0 & E(u_2^2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & E(u_n^2) \end{bmatrix}$$

Varianza

$$\text{Var}(X) = E[(X - E(X))^2]$$

$$\sigma_u^2 = \text{Var}(u_i) = E(u_i^2)$$

Homoscedasticidad

$$E[uu'] = \begin{bmatrix} \sigma_u^2 & 0 & \cdots & 0 \\ 0 & \sigma_u^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_u^2 \end{bmatrix} = \sigma_u^2\mathbf{I}$$

4.5. Condición de Rango Completo

Para que los estimadores de Mínimos Cuadrados Ordinarios (MCO) existan y sean únicos, la matriz $\mathbf{X'X}$ debe tener **rango completo**. Esto implica que sus columnas son linealmente independientes. Entonces:

$$\mathbf{X'X} \quad \text{debe tener rango completo} \quad \text{y} \quad \mathbf{n > k}$$

donde n es el número de observaciones y k el número de parámetros a estimar.

Ejemplos:

$$A_1 = \begin{pmatrix} 5 & 3 \\ 2 & 7 \end{pmatrix}, \quad \text{rango}(A_1) = 2 \Rightarrow \text{Rango completo}$$

$$A_2 = \begin{pmatrix} 2 & 7 \\ 4 & 14 \end{pmatrix}, \quad \text{rango}(A_2) = 1 \Rightarrow \text{Rango incompleto (matriz singular)}$$

Sabemos que $\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'Y}$

Si es rango incompleto $\Rightarrow \det(\mathbf{X'X}) = 0 \Rightarrow (\mathbf{X'X})^{-1}$ no existe y los estimadores no pueden calcularse.

4.6. Variabilidad de X

Establece que:

$$Var(X_i) > 0 \quad \forall\ X_i \text{ distinto de la constante.}$$

Implica que la variable X_i debe presentar cierta variabilidad dentro de la muestra. Si todos los valores de X_i fueran iguales, la varianza sería cero, lo cual imposibilitaría la estimación de los parámetros del modelo de regresión.

Por ejemplo:

$$X_i = \{4, 4, 4, 4, 4\} \Rightarrow Var(X_i) = 0$$

Por tanto, **no todos los valores de X_i deben ser idénticos**. Al menos uno de los valores debe diferir de los demás para que el modelo pueda identificar correctamente la relación entre la variable dependiente y la explicativa.

Ejemplo: Supongamos que queremos analizar el efecto del acceso a la educación superior sobre el salario, mediante el modelo:

$$w = \beta_0 + \beta_1 X + u$$

donde:

- w: Salario del individuo
- X: Variable dicotómica que indica acceso a la educación superior,
  - $X = 1$: Si la persona tuvo acceso a la educación superior,
  - $X = 0$: Si la persona no tuvo acceso a la educación superior.

Si todos los individuos tuvieran el mismo valor de X (por ejemplo, $X = \{1, 1, 1, 1\}$), entonces no existiría variabilidad, y no podríamos medir ni estimar el efecto de la educación superior sobre el salario. En otras palabras, la brecha salarial no podría calcularse.

4.7. Modelo Correctamente Especificado

Para que un modelo de regresión esté correctamente especificado, deben cumplirse los siguientes criterios:

(i) Forma funcional correcta

La forma funcional del modelo debe corresponder al verdadero proceso generador de datos (PGD). Si en el PGD la relación entre las variables es lineal, la estimación debe realizarse bajo una forma lineal. De igual forma, si la relación es cuadrática, cúbica o logarítmica, el modelo debe ajustarse en esa forma. Una especificación funcional incorrecta puede llevar a estimaciones sesgadas o inconsistentes.

(ii) No omisión de variable relevante

No se debe excluir ninguna variable que forme parte del proceso generador de datos. Si se omite una variable relevante que está correlacionada con las variables incluidas, los estimadores obtenidos serán sesgados, afectando la validez del modelo.

(iii) No inclusión de variable irrelevante

Tampoco se deben incluir variables que no formen parte del proceso generador de datos. La inclusión de variables irrelevantes no sesga los coeficientes, pero reduce la eficiencia de los estimadores y aumenta la varianza, afectando la precisión del modelo.

4.8. Normalidad de los errores

$$u_i \sim \text{i.i.d. } N(0, \sigma^2)$$

La normalidad permite distribuciones exactas de los estimadores y estadísticas de prueba (t, F) en muestras finitas. Si los errores son normales y se cumplen los supuestos anteriores, entonces $\hat{\boldsymbol{\beta}}$ sigue una distribución normal exacta.

Nota: Demostración de la distribución de $\hat{\beta}$

Si $\mathbf{u} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ y $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$, entonces:

$$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$$

De aquí se obtienen las estadísticas:

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_j)}} \sim t_{n-k}$$

Si la normalidad no se cumple, por el Teorema del Límite Central se puede usar inferencia aproximada basada en normalidad cuando n es grande.

5. Propiedades de los estimadores

Insesgadez

Si el valor esperado del estimador es igual al valor del parámetro verdadero.

$$E[\hat{\beta}] = \beta$$

Nota: Demostración de la insesgadez

Supuestos utilizados:
Linealidad en parámetros, Exogeneidad ($E[\mathbf{u} \mid \mathbf{X}] = 0$) y regresores no estocásticos.

El estimador MCO es:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Bajo el modelo $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$$

Tomando esperanza condicional en $\mathbf{X}$:

$$E[\hat{\boldsymbol{\beta}} \mid \mathbf{X}] = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{u} \mid \mathbf{X}]$$

Por exogeneidad, $E[\mathbf{u} \mid \mathbf{X}] = 0$, entonces:

$$\boxed{E[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}}$$

El estimador MCO es **insesgado**.

Consistencia

Cuando el número de observaciones n tiende a infinito y se mantienen los supuestos del modelo, el estimador converge en probabilidad al valor verdadero del parámetro:

$$\hat{\beta} \xrightarrow{p} \beta$$

En otras palabras, conforme aumenta el tamaño de la muestra, la estimación se vuelve más precisa.

Nota: Demostración de la consistencia

Partimos de la expresión del estimador MCO:

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'Y = \boldsymbol{\beta} + (X'X)^{-1}X'u.$$

Reescribiendo en términos por observación (dividiendo por n):

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \Big(\frac{X'X}{n}\Big)^{-1}\Big(\frac{X'u}{n}\Big).$$

Por la Ley de los Grandes Números:

$$\frac{\mathbf{X'X}}{n} \xrightarrow{p} E[\mathbf{X_i X_i'}], \quad \frac{\mathbf{X'u}}{n} \xrightarrow{p} E[\mathbf{X_i u_i}] = 0$$

Por tanto:

$$\boxed{\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}}$$

El estimador MCO es **consistente**.

Eficiencia

Se refiere a la capacidad del estimador de tener la menor varianza entre todos los estimadores insesgados.

$$Var(\hat{\beta}_1) < Var(\hat{\beta}_2) \quad \forall \text{ estimadores no sesgados}$$

6. Teorema de Gauss-Markov: MCO es BLUE

Bajo los supuestos de linealidad, $E[\mathbf{u} \mid \mathbf{X}] = 0$, y $\text{Var}(\mathbf{u} \mid \mathbf{X}) = \sigma^2\mathbf{I}$, el estimador MCO $\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'y}$ es el Mejor Estimador Lineal Insesgado (BLUE): es lineal en $\mathbf{y}$, es insesgado, y tiene mínima varianza entre todos los estimadores lineales insesgados.

Nota: Demostración matricial esquemática

Sea $\tilde{\boldsymbol{\beta}} = \mathbf{Cy}$ cualquier estimador lineal (matriz $\mathbf{C}$ de tamaño $k \times n$). El insesgamiento implica:

$$E[\tilde{\boldsymbol{\beta}} \mid \mathbf{X}] = \mathbf{C}E[\mathbf{y} \mid \mathbf{X}] = \mathbf{CX}\boldsymbol{\beta} = \boldsymbol{\beta} \quad \text{para todo } \boldsymbol{\beta}$$

lo que exige:

$$\mathbf{CX} = \mathbf{I}_k \tag{1}$$

Escribimos $\mathbf{C}$ como:

$$\mathbf{C} = (\mathbf{X'X})^{-1}\mathbf{X'} + \mathbf{D}$$

donde $\mathbf{DX} = \mathbf{0}$ (para satisfacer (1)). Entonces la varianza de $\tilde{\boldsymbol{\beta}}$ es:

$$\text{Var}(\tilde{\boldsymbol{\beta}} \mid \mathbf{X}) = \mathbf{C}\,\text{Var}(\mathbf{y} \mid \mathbf{X})\mathbf{C'} = \mathbf{C}(\sigma^2\mathbf{I})\mathbf{C'} = \sigma^2\mathbf{CC'}$$

Sustituyendo la expresión de $\mathbf{C}$:

$$\mathbf{CC'} = [(\mathbf{X'X})^{-1}\mathbf{X'} + \mathbf{D}][\mathbf{X}(\mathbf{X'X})^{-1} + \mathbf{D'}] = (\mathbf{X'X})^{-1} + \mathbf{DD'}$$

pues $\mathbf{X'D'} = (\mathbf{DX})' = \mathbf{0}$.
Así:

$$\text{Var}(\tilde{\boldsymbol{\beta}} \mid \mathbf{X}) = \sigma^2[(\mathbf{X'X})^{-1} + \mathbf{DD'}] \succeq \sigma^2(\mathbf{X'X})^{-1} = \text{Var}(\hat{\boldsymbol{\beta}} \mid \mathbf{X})$$

pues $\mathbf{DD'}$ es semidefinida positiva. Por tanto, $\hat{\boldsymbol{\beta}}$ tiene varianza menor o igual que cualquier otro estimador lineal insesgado.

7. Violaciones de Supuestos

El incumplimiento de los supuestos del Modelo de Regresión Lineal Clásico tiene consecuencias importantes para las propiedades de los estimadores MCO y la validez de la inferencia estadística. A continuación se detallan las principales violaciones, sus efectos y soluciones.

7.1. Endogeneidad

Ocurre cuando una o más de las **variables explicativas** $(X_1, X_2, ..., X_k)$ están **correlacionadas** con el **término de error** (u) en el modelo de regresión. Esto viola el supuesto de *exogeneidad*, que establece que las variables explicativas deben ser independientes del error.

Causas de la heterocedasticidad

- **Variables omitidas**: Falta una variable relevante que afecta tanto a Y como a X.
  Ejemplo: Estimar el efecto de la educación sobre el salario sin incluir la habilidad del individuo.
- **Simultaneidad**: Cuando Y y X se determinan de manera conjunta o simultánea.
  Ejemplo: Precio y cantidad en un mercado de equilibrio.
- **Errores de medición**: Si una variable X está medida con error, el componente de error entra correlacionado con u.

Planteamiento y efecto matricial

Consideremos el verdadero modelo completo:

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}, \quad E[\boldsymbol{\varepsilon} \mid \mathbf{X}_1, \mathbf{X}_2] = \mathbf{0}$$

Si estimamos omitiendo $\mathbf{X}_2$, el error efectivo es $\mathbf{u} = \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$. Entonces:

$$\hat{\boldsymbol{\beta}}_1^{\text{omit}} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y} = \boldsymbol{\beta}_1 + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\boldsymbol{\beta}_2 + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\boldsymbol{\varepsilon}$$

Tomando esperanza condicional:

$$E[\hat{\boldsymbol{\beta}}_1^{\text{omit}} \mid \mathbf{X}_1, \mathbf{X}_2] = \boldsymbol{\beta}_1 + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\boldsymbol{\beta}_2$$

El sesgo por variable omitida (OVB) es:

$$\text{Sesgo}(\hat{\boldsymbol{\beta}}_1^{\text{omit}}) = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\boldsymbol{\beta}_2$$

Consecuencias en las propiedades

Es **sesgado**, **inconsistente** e **ineficiente**. A diferencia de otras violaciones (como heterocedasticidad o autocorrelación), que solo afectan la *eficiencia* o la *validez de los errores estándar*, la endogeneidad afecta la *validez misma del modelo*. Por ello, se la conoce como la **violación más crítica** dentro del MRLC.

Perspectiva FWL (Frisch-Waugh-Lovell)

La estimación de $\boldsymbol{\beta}_1$ en la regresión conjunta equivale a regresar $\mathbf{M}_{\mathbf{X}_2}\mathbf{y}$ sobre $\mathbf{M}_{\mathbf{X}_2}\mathbf{X}_1$. Al omitir $\mathbf{X}_2$, no se realiza esta residualización, lo que introduce sesgo.

Soluciones

- **Variables instrumentales (IV/2SLS)**
  Para un instrumento $\mathbf{Z}$ válido $(n \times \ell)$:
  - **Relevancia**: plim $\frac{1}{n}\mathbf{Z}'\mathbf{X}$ tiene rango k
  - **Exogeneidad**: plim $\frac{1}{n}\mathbf{Z}'\mathbf{u} = \mathbf{0}$

El estimador 2SLS es:

$$\hat{\boldsymbol{\beta}}_{\text{2SLS}} = (\mathbf{X}'\mathbf{P}_Z\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_Z\mathbf{y}, \quad \text{donde } \mathbf{P}_Z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$$

Bajo exogeneidad del instrumento, $\operatorname{plim} \hat{\boldsymbol{\beta}}_{\text{2SLS}} = \boldsymbol{\beta}$.

- **Instrumentos débiles**: Regla empírica: estadístico $F > 10$ en la primera etapa. Para instrumentos débiles, usar tests robustos (Anderson-Rubin) o estimadores LIML/Kleibergen

7.2. Heterocedasticidad

En un modelo de regresión clásico, el supuesto de *homocedasticidad* indica que los errores tienen una **varianza constante**. Sin embargo, cuando esta varianza cambia dependiendo de los valores de las variables explicativas, se dice que existe heterocedasticidad.

Causas de la heterocedasticidad

- **Variables omitidas**: Si el modelo no incluye una o más variables que son relevantes para la variable dependiente, la dispersión de los errores podría depender de las variables no incluidas, causando heterocedasticidad.
- **Datos con escalas muy diferentes**: Si las variables explicativas tienen escalas muy distintas, como ingreso (en miles) y edad (en años), los errores pueden tener una varianza que depende de estas escalas
- **Outliers**: Los valores extremos en los datos pueden hacer que los errores tengan una mayor dispersión en ciertos rangos de las variables explicativas.
- **Cambios estructurales en los datos**:Factores externos como crisis económicas, reformas regulatorias o avances tecnológicos pueden cambiar la relación entre las variables.

Consecuencias en las propiedades

Es insesgado, consistente e **ineficiente**. Los errores estándar de los coeficientes son incorrectos, lo que afectará la significancia de las pruebas t y los intervalos de confianza.

Efecto matricial

Si $\operatorname{Var}(\mathbf{u} \mid \mathbf{X}) = \boldsymbol{\Omega}$ (diagonal con σ_i^2), entonces:

$$\operatorname{Var}(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

La fórmula clásica $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ no es válida.

Formas de detección

1. **Visualmente**: Gráfico de residuos vs valores ajustados.
2. **Pruebas formales**:
   - **Breusch-Pagan**: Prueba para heterocedasticidad bajo la suposición de normalidad.
   - **White**: Test más general que no asume que los errores sigan una distribución normal.

Soluciones

- **Errores robustos**: Para obtener estimaciones válidas de los errores estándar.

  Errores estándar robustos de White

$$\widehat{\operatorname{Var}}_{\text{White}}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\left(\sum_{i=1}^{n}\hat{u}_i^2\mathbf{x}_i\mathbf{x}_i'\right)(\mathbf{X}'\mathbf{X})^{-1}$$

- **Transformaciones logarítmicas**: Las transformaciones logarítmicas o de Box-Cox pueden ayudar a estabilizar la varianza.
- **Mínimos Cuadrados Generalizados (GLS/FGLS)**: Ajusta el modelo para corregir la heterocedasticidad.

  Si $\mathbf{\Omega}$ es conocida:

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{y}, \quad \text{Var}(\hat{\boldsymbol{\beta}}_{\text{GLS}}) = (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}$$

  Si $\mathbf{\Omega}$ es desconocida, se estima $\hat{\mathbf{\Omega}}$ y se usa FGLS.

7.3. Autocorrelación

Ocurre cuando los errores de diferentes observaciones están correlacionados entre sí. Es decir, el error en una observación depende de los errores anteriores. Este fenómeno es común en datos de series temporales.

Si $\mathbf{u}$ tiene estructura serial, $\mathbf{\Omega}$ no es diagonal:

$$\text{Var}(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

Causas de la Autocorrelación

- **Variables omitidas**: Factores no observados que afectan tanto a la variable dependiente como a las explicativas, lo que genera dependencia temporal en los errores.
- **Mala especificación**: Un modelo que no captura adecuadamente la estructura de los datos puede inducir autocorrelación.
- **Inercia natural**:En series de tiempo, fenómenos como tendencias o ciclos inherentes pueden causar que los errores de un periodo estén correlacionados con los errores del periodo anterior.

Consecuencias en las propiedades

Es insesgado, consistente e **ineficiente**. Los errores estándar se calculan incorrectamente, lo que invalida las pruebas de hipótesis.

Formas de detección

1. **Durbin-Watson**: Prueba común para detectar autocorrelación en modelos de regresión.
2. **Breusch-Godfrey**:Prueba que generaliza la Durbin-Watson para detectar autocorrelación de orden superior.
3. **ACF/PACF (Funciones de autocorrelación y autocorrelación parcial)**: Muestran la correlación entre los errores de diferentes periodos.

Soluciones

- **Prais-Winsten para AR(1)**

  Para errores AR(1): $u_t = \rho u_{t-1} + \varepsilon_t$, $|\rho| < 1$, $\varepsilon_t \sim \text{i.i.d.}(0, \sigma^2)$.

  Transformación para $t \geq 2$:

$$y_t - \rho y_{t-1} = (\mathbf{x}_t - \rho \mathbf{x}_{t-1})'\boldsymbol{\beta} + \varepsilon_t$$

  En la práctica, se estima $\hat{\rho}$ y se aplica Prais-Winsten o Cochrane-Orcutt.

- **Newey-West (HAC)**

  Para formas desconocidas de $\mathbf{\Omega}$:

$$\widehat{\text{Var}}_{\text{NW}}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\hat{\mathbf{S}}(\mathbf{X}'\mathbf{X})^{-1}$$

  donde $\hat{\mathbf{S}}$ incluye covarianzas hasta el lag L con pesos Bartlett.

7.4. Multicolinealidad

Ocurre cuando las variables explicativas en un modelo de regresión están altamente correlacionadas entre sí. Esto dificulta la separación de los efectos individuales de cada variable sobre la variable dependiente.

En este caso, la matriz $X'X$ se vuelve casi singular debido a la alta correlación entre las variables. Esto hace que la matriz sea difícil de invertir y afecta la estimación precisa de los coeficientes β.
Efecto Matricial
Si $\mathbf{X'X}$ está mal condicionado, $(\mathbf{X'X})^{-1}$ es grande $\rightarrow$ varianzas de $\hat{\boldsymbol{\beta}}$ son grandes.
Para la componente j:

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma^2}{(1-R_j^2)\sum_i (x_{ij}-\bar{x}_j)^2}, \quad \text{VIF}_j = \frac{1}{1-R_j^2}$$

Causas de la Multicolinealidad

- **Inclusión de variables redundantes**
- **Variables naturalmente correlacionadas**: Como variables económicas que tienden a moverse juntas (por ejemplo, PIB y consumo).
- **Muestras pequeñas**: Muestras con poca variabilidad en las variables explicativas.

Consecuencias en las propiedades

- **Multicolinealidad Perfecta**: No se puede obtener los estimadores
- **Multicolinealidad imperfecta**: El estimador es insesgado, consistente e **ineficiente**. Puede haber confusión sobre el efecto de una variable debido a la colinealidad. Las variables correlacionadas pierden poder explicativo y parecen no tener efecto.

Formas de detección

1. **Valores cercanos a ± 1**: Correlación entre pares de variables que indica posible multicolinealidad.
2. **VIF (Factor de Inflación de Varianza)**: Un valor de VIF superior a 5 o 10 sugiere problemas graves de multicolinealidad.

Soluciones

- **Eliminar o combinar variables redundantes**
- **Regresión por componentes principales (PCR)**
- **Regresión Ridge**

$$\hat{\boldsymbol{\beta}}_{\text{ridge}} = (\mathbf{X'X} + \lambda \mathbf{I})^{-1}\mathbf{X'y}$$

$$\text{Var}(\hat{\boldsymbol{\beta}}_{\text{ridge}}) = \sigma^2(\mathbf{X'X} + \lambda \mathbf{I})^{-1}\mathbf{X'X}(\mathbf{X'X} + \lambda \mathbf{I})^{-1}$$

7.5. Error de medición en regresores

Caso clásico: Atenuación

Verdadero: $y = \beta x + u$. Observamos $\tilde{x} = x + v$ con $E[v] = 0$, v independiente de x, u.

$$\text{plim}\,\hat{\beta} = \beta \cdot \frac{\text{Var}(x)}{\text{Var}(x) + \text{Var}(v)}$$

El sesgo es hacia cero (attenuation bias).

Soluciones

- Variables instrumentales
- Datos de validación/medidas repetidas
- SIMEX (simulation extrapolation)

7.6. Inestabilidad de parámetros

Ocurre cuando los coeficientes β en el modelo de regresión no son constantes a lo largo del tiempo o entre subgrupos de datos. Es decir, los parámetros β cambian dependiendo del periodo o del subgrupo de datos que estemos analizando.

Causas de la Inestabilidad de Parámetros

- **Cambios estructurales en la economía**: Como crisis económicas o cambios en la política económica que afectan la relación entre las variables.
- **Crisis financieras o políticas**: Eventualmente modifican las relaciones económicas subyacentes.
- **Reformas regulatorias o cambios tecnológicos**: Disruptivos en sectores clave que alteran las relaciones entre variables.

Consecuencias en las propiedades

- Los coeficientes pueden no reflejar la realidad de los datos en el momento actual.
- Los pronósticos basados en un modelo con parámetros inestables no serán precisos.
- Las conclusiones sobre las relaciones económicas pueden estar equivocadas debido a la inestabilidad.

Formas de detección

1. **Prueba de Chow**: Se utiliza para comprobar si hay diferencias significativas en los coeficientes de diferentes subgrupos o periodos.
2. **Residuos recursivos**: Análisis de residuos que ayuda a detectar cambios en los coeficientes a lo largo del tiempo.
3. **Prueba CUSUM**: Detecta cambios estructurales en los parámetros del modelo.
4. **Prueba de Quandt-Andrews**: Identifica puntos de ruptura en las series temporales.

Soluciones

- **Segmentación**: Dividir los datos en subgrupos o períodos para estimar modelos específicos.
- **Variables Dummy**: Utilizar variables indicadoras para modelar cambios estructurales.
- **Modelos Tiempo-Variables**: Incluir variables de tiempo que puedan capturar los cambios en las relaciones.
- **Ventanas Rodantes**: Estimar los parámetros en ventanas móviles para ajustar continuamente el modelo

7.7. No-normalidad de los errores

Ocurre cuando los **errores** (u) del modelo de regresión no siguen una distribución **normal**.

Causas de la No Normalidad

- **Outliers**: Valores extremos en los datos.
- **Especificación incorrecta del modelo**: Omisión de variables importantes o uso de transformaciones incorrectas.
- **Errores de medición**: Imposibilidad de medir con precisión las variables.
- **Asimetrías inherentes en los datos económicos**

Consecuencias en las propiedades

No afecta el insesgamiento pero afecta la exactitud de pruebas t y F en muestras pequeñas. (**Ineficiente**)
Formas de detección

1. **Prueba de Shapiro-Wilk**: Prueba de normalidad para muestras pequeñas.
2. **Prueba de Jarque-Bera**: Prueba para evaluar asimetría y curtosis en los errores.
3. **Prueba de Kolmogorov-Smirnov**: Prueba de ajuste de distribuciones.

Soluciones

- **Inferencia asintótica (TLC)**
- **Transformaciones Box-Cox, logarítmicas, raíz cuadrada**: Pueden ayudar a hacer los errores más normales.

7.8. Correlación por clusters

Cuando los errores están correlacionados dentro de grupos pero no entre grupos.

$$\widehat{\text{Var}}_{\text{cluster}}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1} \left(\sum_{g=1}^{G} \mathbf{X}_g' \hat{\mathbf{u}}_g \hat{\mathbf{u}}_g' \mathbf{X}_g \right) (\mathbf{X}'\mathbf{X})^{-1}$$

Requiere un número grande de clusters (G) para consistencia.

8. Aplicaciones del MRLC

8.1. Retornos a la educación (Ecuación de Mincer)

Modelo teórico

La ecuación de Mincer en forma log-lineal:

$$\ln w_i = \beta_0 + \beta_1 \text{educ}_i + \beta_2 \text{exper}_i + u_i$$

donde w_i es el salario, educ_i son los años de educación, exper_i es la experiencia potencial, y u_i captura factores no observados (habilidad, motivación, etc.).

Problema de endogeneidad

El problema principal es la **variable omitida** (habilidad a_i). El modelo verdadero es:

$$\ln w_i = \beta_0 + \beta_1 \text{educ}_i + \beta_2 \text{exper}_i + \gamma a_i + \varepsilon_i, \quad E[\varepsilon_i \mid X, a] = 0$$

Si estimamos omitiendo a_i, el término de error efectivo es $u_i = \gamma a_i + \varepsilon_i$ y:

$$\text{Cov}(\text{educ}_i, u_i) = \gamma \, \text{Cov}(\text{educ}_i, a_i) \neq 0$$

Sesgo por variable omitida

Forma escalar:

$$\operatorname{plim} \hat{\beta}_1^{OLS} = \beta_1 + \gamma \cdot \frac{\operatorname{Cov}(\text{educ}, a)}{\operatorname{Var}(\text{educ})}$$

Forma matricial:

$$E[\hat{\boldsymbol{\beta}}^{\text{omit}} \mid X, A] = \boldsymbol{\beta} + (X'X)^{-1}X'A\gamma$$

Error de medición

Si observamos $\widetilde{\text{educ}} = \text{educ} + v$ con $E[v] = 0$:

$$\operatorname{plim} \hat{\beta}_1^{OLS} = \beta_1 \cdot \frac{\operatorname{Var}(\text{educ})}{\operatorname{Var}(\text{educ}) + \operatorname{Var}(v)}$$

Este es el **sesgo de atenuación**.

Solución: Variables instrumentales (IV/2SLS)

Para un instrumento $\mathbf{Z}$ válido:

- **Relevancia**: $\operatorname{plim} \frac{1}{n}\mathbf{Z}'\mathbf{X}$ tiene rango completo
- **Exogeneidad**: $\operatorname{plim} \frac{1}{n}\mathbf{Z}'\mathbf{u} = 0$

El estimador 2SLS:

$$\hat{\boldsymbol{\beta}}_{\text{2SLS}} = (\mathbf{X}'\mathbf{P}_Z\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_Z\mathbf{y}, \quad \mathbf{P}_Z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$$

Bajo exogeneidad del instrumento, $\operatorname{plim} \hat{\boldsymbol{\beta}}_{\text{2SLS}} = \boldsymbol{\beta}$.

Varianza asintótica

Bajo homocedasticidad:

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{\text{2SLS}} - \boldsymbol{\beta}) \xrightarrow{d} N(0, \sigma^2(\operatorname{plim} \frac{1}{n}\mathbf{X}'\mathbf{P}_Z\mathbf{X})^{-1})$$

Aspectos prácticos

- **Instrumentos débiles**: Estadístico $F > 10$ en primera etapa
- **Interpretación local**: IV identifica efectos locales (LATE)
- **Ejemplos de instrumentos**: Política educativa, distancia a universidades, fecha de nacimiento

8.2. Función de consumo en series temporales

Modelo teórico

$$C_t = \alpha + \beta Y_t + u_t, \quad t = 1, \ldots, T$$

donde C_t es consumo, Y_t es ingreso, y u_t puede mostrar autocorrelación.

Problema de autocorrelación

Supongamos u_t sigue AR(1):

$$u_t = \rho u_{t-1} + \varepsilon_t, \quad |\rho| < 1, \quad \varepsilon_t \sim \text{i.i.d.}(0, \sigma^2)$$

La matriz de varianza-covarianza de $\mathbf{u}$ es $\boldsymbol{\Omega}$ con elementos:

$$\Omega_{ij} = \frac{\sigma^2 \rho^{|i-j|}}{1 - \rho^2}$$

La varianza correcta de OLS es:

$$\operatorname{Var}(\hat{\boldsymbol{\beta}} \mid X) = (X'X)^{-1}X'\boldsymbol{\Omega}X(X'X)^{-1}$$

Solución 1: GLS analítico (Prais-Winsten/Cochrane-Orcutt)

El estimador GLS:

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = (X'\boldsymbol{\Omega}^{-1}X)^{-1}X'\boldsymbol{\Omega}^{-1}y$$

Para AR(1), la transformación es:

$$y_t^* = y_t - \rho y_{t-1}, \quad x_t^* = x_t - \rho x_{t-1} \quad \text{para } t \geq 2$$

Estimación de ρ:

$$\hat{\rho} = \frac{\sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=2}^{T} \hat{u}_{t-1}^2}$$

Solución 2: Newey-West (HAC)

Estimador de varianza robusta:

$$\widehat{\text{Var}}_{\text{NW}}(\hat{\boldsymbol{\beta}}) = (X'X)^{-1}\hat{S}(X'X)^{-1}$$

donde $\hat{S}$ incluye términos de covarianza hasta el lag L con pesos Bartlett:

$$w_h = 1 - \frac{h}{L+1}$$

Propiedades de los estimadores

- **GLS**: BLUE si $\boldsymbol{\Omega}$ es conocida
- **FGLS**: Asintóticamente eficiente si $\boldsymbol{\Omega}$ se estima consistentemente
- **Newey-West**: Inferencia válida sin especificar la estructura de $\boldsymbol{\Omega}$

Demostración esquemática

$$\hat{\boldsymbol{\beta}}^{OLS} = \boldsymbol{\beta} + (X'X)^{-1}X'\mathbf{u}$$
$$\text{Var}(\hat{\boldsymbol{\beta}}^{OLS} \mid X) = (X'X)^{-1}X'\boldsymbol{\Omega}X(X'X)^{-1}$$

La transformación GLS equivale a OLS sobre datos transformados:

$$\mathbf{y}^* = \mathbf{C}\mathbf{y}, \quad \mathbf{X}^* = \mathbf{C}\mathbf{X}, \quad \text{donde } \mathbf{C}'\mathbf{C} = \boldsymbol{\Omega}^{-1}$$

9. Aplicación en Sotfwares (Stata - Python)

9.1. Aplicación en Stata

A continuación vamos a mostrar, paso a paso, cómo estimar y validar un Modelo de Regresión Lineal Clásico (MRLC) utilizando el software `Stata`. A través de este ejemplo se explica la lógica detrás de cada comando y su utilidad dentro del proceso econométrico, desde la carga de los datos hasta la verificación de supuestos y robustez del modelo.

En el siguiente enlace podrá descargar al archivo DO-FILE que contiene los códigos para efectuar el modelo:
Molde STATA.

Ejemplo: Determinantes del gasto en salud de los hogares peruanos

Se analiza una base de datos hipotética llamada `salud_hogares.dta`, que contiene información sobre hogares peruanos. El propósito del ejercicio es estudiar los **determinantes del gasto en salud** de los hogares, considerando variables como ingreso, edad, educación, entorno urbano, tenencia de seguro de salud y género del jefe del hogar.

El modelo base se plantea de la siguiente forma:

$$GastoSalud_i = \beta_0 + \beta_1 Ingreso_i + \beta_2 Edad_i + \beta_3 Educ_i + \beta_4 Urbano_i + \beta_5 Seguro_i + \beta_6 MujerJefe_i + u_i$$

A. Carga y exploración de los datos

```
use "salud_hogares.dta", clear
summarize gasto_salud ingreso edad educ urbano seguro mujer_jefe
```

El comando `use` carga la base de datos `salud_hogares.dta` y elimina cualquier base previa con `clear`.

El comando `summarize` muestra estadísticas descriptivas (media, desviación estándar, mínimo y máximo) de las variables principales. Esta etapa permite analizar los datos antes de la regresión.

B. Análisis gráfico

```
scatter gasto_salud ingreso || lfit gasto_salud ingreso
```

Se grafica la relación entre el gasto en salud (eje Y) y el ingreso (eje X), añadiendo una línea ajustada (`lfit`) para visualizar la tendencia lineal. Este paso permite comprobar si el supuesto de linealidad parece razonable.

```
scatter gasto_salud ingreso, ytitle("Gasto en salud (S/.)")
xtitle("Ingreso del hogar (S/.)") ///
legend(label(1 "Datos observados") label(2 "Línea ajustada"))
```

Los comandos `ytitle()` y `xtitle()` permiten añadir títulos a los ejes, mientras que `legend()` personaliza las etiquetas de la leyenda. La barra diagonal triple (`///`) sirve para continuar la línea de código en la siguiente fila. Este bloque busca mejorar la presentación visual de los resultados.

C. Correlaciones y covarianzas

```
correlate gasto_salud ingreso edad educ urbano seguro mujer_jefe
correlate gasto_salud ingreso edad educ urbano seguro mujer_jefe, covariance
```

`correlate` genera la matriz de correlaciones entre las variables, útil para identificar colinealidad o relaciones preliminares. La opción `covariance` muestra la matriz de covarianzas, conservando las unidades originales de las variables. Ambas sirven como diagnóstico previo al modelo.

D. Estimación del MRLC

```
regress gasto_salud ingreso
estimate store m1

regress gasto_salud ingreso edad educ
estimate store m2

regress gasto_salud ingreso edad educ urbano seguro mujer_jefe
estimate store m3
```

El comando `regress` estima un modelo lineal mediante Mínimos Cuadrados Ordinarios (MCO). Cada regresión incluye progresivamente más variables explicativas, almacenando los resultados con `estimate store`. Esto permite comparar modelos ($m1, m2, m3$) según la inclusión de controles adicionales.

E. Exportación de resultados

```
ssc install outreg2
outreg2 [m1 m2 m3] using gasto_salud_resultados.xls, replace
```

`outreg2` exporta los resultados de las regresiones a un archivo Excel, ideal para informes o papers. La opción `replace` sobrescribe el archivo si ya existe. Este paso facilita la presentación profesional de los resultados.

F. Prueba de hipótesis

```
regress gasto_salud ingreso edad educ urbano seguro mujer_jefe
test ingreso
test edad educ mujer_jefe
```

Las pruebas `test` permiten contrastar hipótesis sobre los coeficientes del modelo:

- **Individual:** `test ingreso` evalúa $H_0 : \beta_{ingreso} = 0$
- **Conjunta:** `test edad educ mujer_jefe` evalúa $H_0 : \beta_{edad} = \beta_{educ} = \beta_{mujer\_jefe} = 0$

Esto permite determinar si las variables tienen efecto estadísticamente significativo sobre el gasto en salud.

G. Verificación de los supuestos clásicos

```
predict uhat, resid
kdensity uhat
histogram uhat, normal kdensity
sktest uhat, noadj

hettest ingreso edad educ urbano seguro mujer_jefe, iid
regress gasto_salud ingreso edad educ urbano seguro mujer_jefe, robust
```

`predict` genera los residuos del modelo ($\hat{u}_i$) para su análisis. `kdensity` y `histogram` permiten observar gráficamente la distribución de los errores. `sktest` contrasta formalmente la normalidad (asimetría y curtosis). `hettest` evalúa la presencia de heterocedasticidad. Finalmente, `robust` corrige los errores estándar en caso de heterocedasticidad.

H. Estabilidad de parámetros: Test de Chow

```
reg gasto_salud ingreso edad educ urbano mujer_jefe if seguro==1
scalar ssrum=e(rss)
scalar dfum=e(df_r)

reg gasto_salud ingreso edad educ urbano mujer_jefe if seguro==0
scalar ssruf=e(rss)
scalar dfuf=e(df_r)
```

Se divide la muestra entre hogares con y sin seguro (`if seguro==1` o `if seguro==0`), estimando por separado ambos modelos. Luego se recuperan las sumas de residuos cuadrados (`rss`) y grados de libertad (`df_r`) para aplicar el Test de Chow. Este procedimiento verifica si los coeficientes difieren significativamente entre grupos.

I. Interacciones y modelo final

```
gen i1=seguro*ingreso
gen i2=seguro*edad
gen i3=seguro*educ
gen i4=seguro*urbano

reg gasto_salud ingreso edad educ urbano seguro mujer_jefe i1 i2 i3 i4, robust
test i1 i2 i3 i4
```

Las variables de interacción (`i1-i4`) permiten analizar si los efectos de las variables explicativas cambian dependiendo de la tenencia de seguro. La prueba conjunta `test i1 i2 i3 i4` evalúa si existen diferencias estructurales significativas.

El modelo final incluye los términos relevantes y se estima con errores robustos.
Entonces, en la especificación final estimamos el siguiente modelo con interacción entre `ingreso` y la variable `seguro`:

$$\text{gasto\_salud}_i = \beta_0 + \beta_1\,\text{ingreso}_i + \beta_2\,\text{edad}_i + \beta_3\,\text{educ}_i + \beta_4\,\text{urbano}_i + \beta_5\,\text{seguro}_i + \beta_6\,\text{mujer\_jefe}_i + \beta_7(\text{ingreso}_i \times \text{seguro}_i) + u_i.$$

En `Stata` se estima con:

```
reg gasto_salud ingreso edad educ urbano seguro mujer_jefe i1, robust
```

donde `i1 = seguro*ingreso` y la opción `robust` produce errores estándar consistentes frente a heterocedasticidad.

Interpretación del término de interacción: El efecto del ingreso sobre el gasto en salud depende de la tenencia de seguro.

Esto significa que la relación entre ingreso y gasto no es uniforme para todos los hogares: la pendiente del ingreso cambia según si el hogar cuenta o no con seguro. Por tanto, el impacto marginal del ingreso en el gasto en salud será distinto entre hogares asegurados y no asegurados.

9.2. Aplicación en Python

En este caso se presenta un planteamiento y caso arbitrario sobre el MRLC para ejecutar en Python; para ello, previamente es necesario instalar en el terminal con el presente comando las siguientes librerías para el programa:

Instalación de librerías requeridas

```
py -m pip install numpy pandas matplotlib scikit-learn seaborn
```

Instalación individual alternativa

```
py -m pip install numpy
py -m pip install pandas
py -m pip install matplotlib
py -m pip install scikit-learn
py -m pip install seaborn
```

(En caso de que no se puedan instalar todas simultáneamente, puede instalarse una a una)
Consideraciones importantes:

- **Obs:** Los paquetes Seaborn y Scikit-learn pueden presentar dificultades de instalación en Jupyter Notebook
- El código anterior se incluye en el notebook para garantizar la ejecución correcta
- En caso de errores, ejecutar individualmente cada instalación

Acceso al archivo:

En el siguiente enlace podrá descargar el archivo que contiene los códigos para efectuar el modelo: MRLC.

Instrucciones de uso:

1. Crear una carpeta con nombre breve (ej: `MRLC`)
2. Descargar el archivo `.ipynb` en dicha carpeta
3. Abrir Jupyter Notebook o entorno similar
4. Navegar a la carpeta y ejecutar el archivo
5. Ejecutar las celdas secuencialmente para observar cada paso

Entornos recomendados:

- Jupyter Notebook (nativo)
- JupyterLab
- Google Colab
- VS Code con extensión Jupyter

10. Conclusión

En conclusión, el Modelo de Regresión Lineal Clásico constituye uno de los pilares fundamentales de la econometría, al ofrecer un marco analítico didáctico y sólido para comprender la relación entre variables económicas y sociales. A través de su formulación algebraica y estadística, permite estimar parámetros con propiedades deseables como insesgadez, consistencia y eficiencia, siempre que se cumplan los supuestos establecidos. La revisión detallada de dichos supuestos —linealidad, exogeneidad, homocedasticidad, no autocorrelación, normalidad y rango completo— demuestra que cada uno cumple una función clave para garantizar la validez de los resultados y la solidez de las inferencias.

El análisis de las posibles violaciones, junto con las estrategias de corrección, como el uso de errores robustos, variables instrumentales o pruebas de estabilidad estructural, refuerza la importancia de evaluar críticamente las condiciones del modelo antes de interpretar los coeficientes. Desde una perspectiva práctica, la implementación en `Stata` y `Python` ilustra de manera clara cómo los conceptos teóricos se trasladan al trabajo empírico, permitiendo al investigador diagnosticar, estimar y validar sus modelos de forma sistemática.

En conjunto, el MRLC no solo constituye una herramienta estadística, sino también un enfoque metodológico que fomenta el razonamiento lógico y la interpretación económica rigurosa. Su aplicación adecuada permite obtener resultados confiables, identificar patrones relevantes en los datos y sustentar decisiones basadas en evidencia cuantitativa, consolidando su papel como base del análisis econométrico moderno.

11. Bibliografía

- **Córdova, M.** (2003). *Estadística descriptiva e inferencial: aplicaciones* (5a ed.). Lima: Moshera.
- **García, L.** (2023). *Econometría 1.*Fondo Editorial PUCP. Pontificia Universidad Católica del Perú (PUCP).
- **Greene, W. H.** (2018). *Econometric Analysis* (8th ed.). Pearson.
- **Johnston, J., & DiNardo, J.** (1997). *Econometric Methods* (4th ed.). McGraw–Hill.
- **Moya Calderón, R.** (2005). *Estadística descriptiva: conceptos y aplicaciones.* Lima, Perú: San Marcos de Aníbal Jesús Paredes Galván.
- **Stigler, S. M.** (1986). *The History of Statistics: The Measurement of Uncertainty before 1900.* Harvard University Press.
- **Universidade de Vigo.** (s. f.). *Modelo de regresión lineal clásico (MRLC).* Apuntes de curso.
- **Wooldridge, J. M.** (2010). *Introducción a la econometría: Un enfoque moderno.* Cengage Learning.

	Dependent variable: lnsalario			
	Modelo 1	Modelo 2	Modelo 3	Modelo 4
	(1)	(2)	(3)	(4)
Intercept	0.584***	0.074	0.335***	0.325***
	(0.097)	(0.106)	(0.101)	(0.121)
casado			0.081*	0.081*
			(0.042)	(0.042)
educ	0.083***	0.085***	0.076***	0.077***
	(0.008)	(0.008)	(0.007)	(0.009)
educ:femenino				-0.002
				(0.013)
exper		0.042***	0.036***	0.036***
		(0.005)	(0.005)	(0.005)
exper2		-0.001***	-0.001***	-0.001***
		(0.000)	(0.000)	(0.000)
femenino			-0.332***	-0.307*
			(0.036)	(0.169)
urban		0.158***	0.177***	0.177***
		(0.044)	(0.041)	(0.041)
Observations	526	526	526	526
R^2	0.186	0.317	0.423	0.423
Adjusted R^2	0.184	0.312	0.417	0.415
Residual Std. Error	0.480 (df=524)	0.441 (df=521)	0.406 (df=519)	0.406 (df=518)
F Statistic	119.582*** (df=1; 524)	60.440*** (df=4; 521)	63.475*** (df=6; 519)	54.308*** (df=7; 51

Note: *p<0.1; **p<0.05; ***p<0.
R-cuadrado ajustado se muestra aba