

# Taller de Reforzamiento

## Introducción a la Econometría

### Sesión 3

Chamorro Rodriguez, Gianfranco David  
E-mail: gianfranco.chamorrer@gmail.com

Enero 2023

---

El presente documento sirve de apoyo para el taller de reforzamiento de Econometría, durante el cual se trabajarán conceptos que serán necesarios para iniciar un curso introductorio de pregrado o reforzar conceptos olvidados. Se brindará una perspectiva amable junto a material que permitiera a los participantes abarcar temas mas complejos adelante.

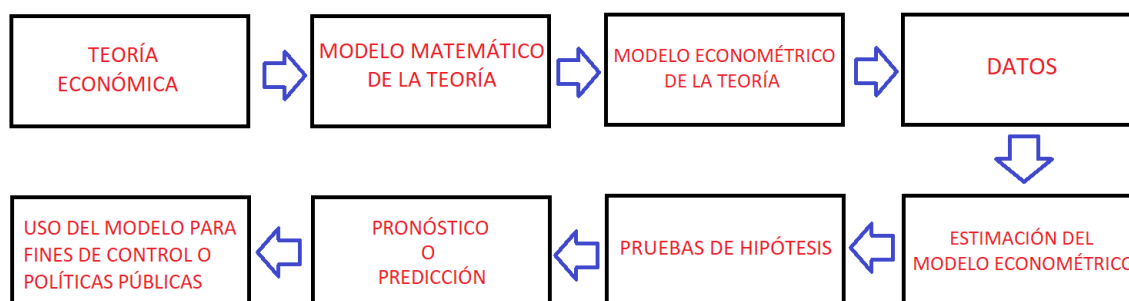
---

## 1. Introducción

El concepto de Econometría podría ser resumido en la aplicación de métodos matemáticos y estadísticos para el contraste de hipótesis planteadas por la teoría económica. Siguiendo esta definición, la econometría tiene como base las tres disciplinas antes mencionadas pero es necesario marcar una diferencia entre el análisis econométrico y estadístico, el primero analiza y soluciona los problemas causados por la violación de los supuestos convencionales de la teoría estadística.<sup>1</sup> La econometría está interesada en usar datos para entender mejor un fenómeno de interés. El objetivo es encontrar un modelo que sea capaz de explicar las principales características de los datos.

## 2. Metodología de la Econometría

Existen diferentes formas de aplicar las técnicas econométricas o de trabajar nuestra base de datos sin embargo la mayoría coincide en una primera etapa donde se seleccionan la(s) variable(s) que sera explicada(s), tomando como base la teoría económica y nuestra intuición. Luego de ello se realizará una correcta especificación y estimación que será sometida a un contraste de hipótesis que nos permita confiar en su nivel inferencial y así contribuir en la toma de decisiones o políticas a aplicar.



Metodología Según D. Gujarati <sup>2</sup>

---

<sup>1</sup>CASTRO, Juan Francisco. *Econometría Aplicada*(2014)

<sup>2</sup>GUJARATI, Damodar. *Econometría*(2010)

### 3. Tipo de Datos

#### 3.1. Corte Transversal

La información de corte transversal consiste en datos de una o más variables de distintos individuos (personas, distritos, empresas, países, etc.), recogidos en el mismo momento del tiempo. Así, se investigan las diferencias entre las características de los individuos. Así el análisis de una encuesta determinada para un periodo de tiempo fijo, significa un análisis de corte transversal. Por lo general, estos datos pertenecen a una misma unidad de tiempo o las diferencias del tiempo por lo general son ignorados.<sup>3</sup>

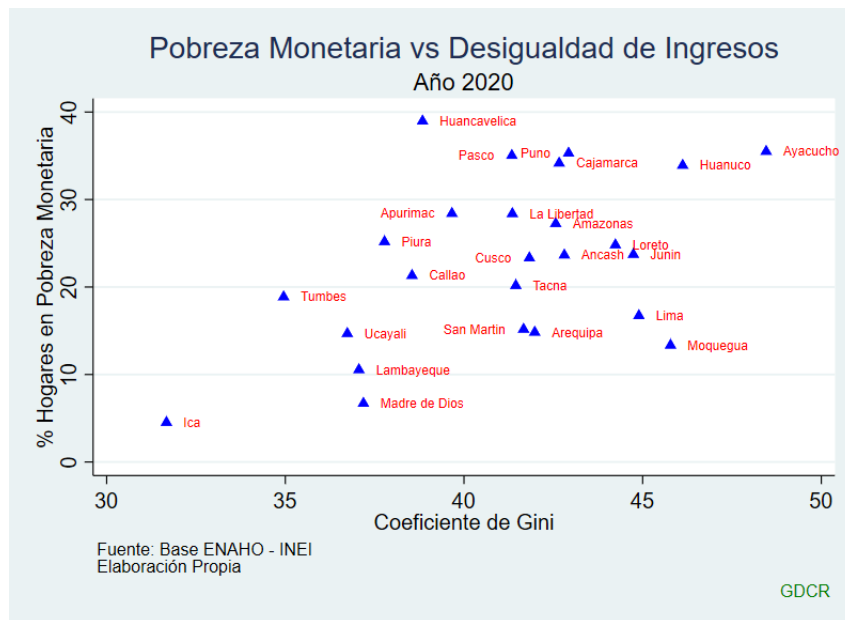


Gráfico 1 : Dispersión de Corte Transversal

#### 3.2. Series de Tiempo

Una Serie de Tiempo consiste en información de una o más variables de un individuo (personas, distritos, empresas, países, etc.) recogidos en distintas unidades de tiempo. La econometría de series de tiempo es un área de estudio cuyos resultados más interesantes, desafortunadamente, requieren del manejo de conceptos matemáticos y estadísticos que vas más allá de lo cubierto en el presente taller introductorio.

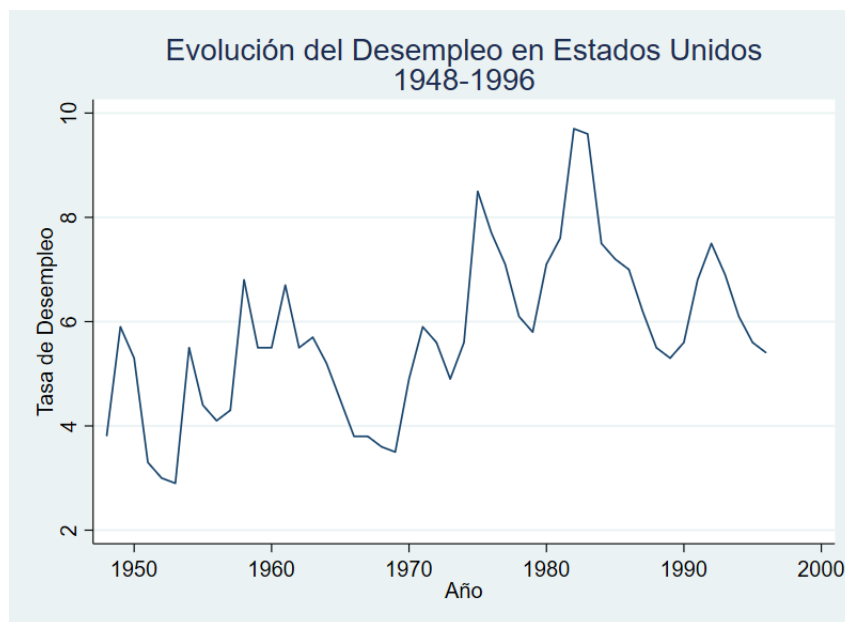


Gráfico 2 : Serie de Tiempo

<sup>3</sup>Las encuestas a hogares por lo general se realizan en distintos momentos pero son considerados dentro de una misma unidad temporal.

### 3.3. Datos de Panel

Combina fuente de variabilidad temporal y transversal. Es decir, las observaciones son distintos individuos, distritos, empresas o países a lo largo de tiempo. Por ello, permite estudiar tanto las diferencias en características de los individuos y como cambian tales en el tiempo.

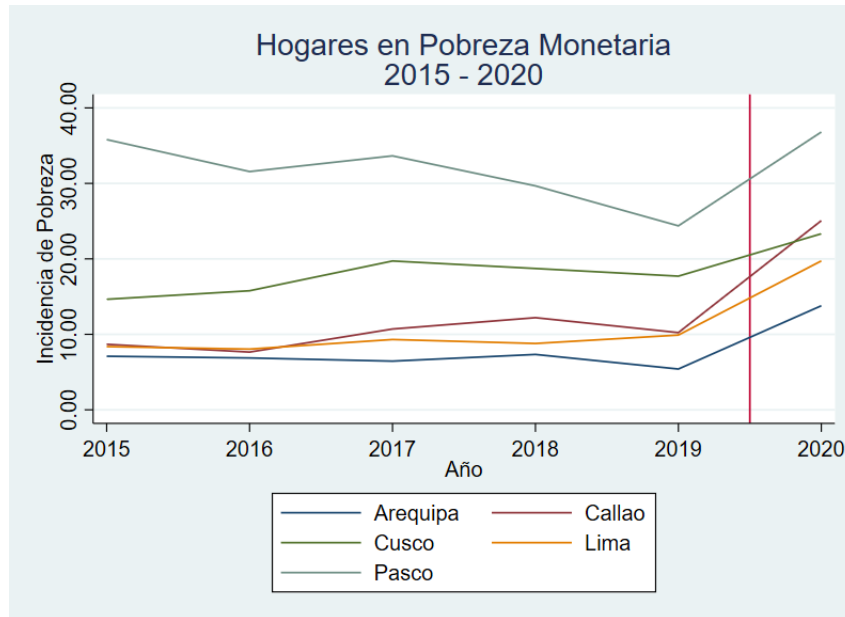


Gráfico 3 : Evolución de la Pobreza Monetaria por Departamento

A nivel de ecuaciones:

**Ecuación Lineal en Datos de Corte Transversal**

$$y_i = \alpha + \beta x_i + \mu_i$$

**Ecuación Lineal en Series de Tiempo**

$$y_t = \alpha + \beta x_t + \mu_t$$

**Ecuación Lineal en Datos de Panel**

$$y_{it} = \alpha + \beta x_{it} + \mu_{it}$$

Siendo :

i= Individuo de análisis  
t= Periodo temporal

## 4. Modelo de Regresión Lineal

La regresión lineal es una técnica de modelado estadístico que se emplea para describir una variable de respuesta continua como una función de una o varias variables predictoras. Puede ayudar a comprender y predecir el comportamiento de sistemas complejos o a analizar datos experimentales, financieros, sociales, entre otros.

Las técnicas de regresión lineal permiten crear un modelo lineal. Este modelo describe la relación entre una variable dependiente  $Y$  como una función de una o varias variables independientes  $X$ .

La ecuación general correspondiente a un modelo de regresión lineal es:

$$\underbrace{Y_i}_{\text{Dependiente}} = \underbrace{\beta_1 + X_1\beta_2 + \dots + X_{k-1}\beta_k}_{\text{Determinístico}} + \underbrace{e_i}_{\text{Estocástico}}$$

Donde  $\beta$  representa las estimaciones de parámetros lineales que se deben calcular y  $\epsilon$  representa los términos de error.

### 4.1. Enfoque Matricial

Al trabajar con amplias bases de datos, se trabaja las relaciones lineales con un enfoque matricial explicado a continuación:

$$Y_{n1} = X_{nk}\beta_{k1} + e_{n1}$$

$$\begin{bmatrix} Y_{11} \\ Y_{21} \\ \vdots \\ Y_{n1} \end{bmatrix} = \begin{bmatrix} 1 & X_{12} & \dots & X_{1k} \\ 1 & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n2} & \dots & X_{nk} \end{bmatrix} * \begin{bmatrix} B_{11} \\ \vdots \\ B_{k1} \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{21} \\ \vdots \\ e_{n1} \end{bmatrix}$$

### 4.2. Estimadores

Dentro de la estadística inferencia la estimación de los parámetros es el proceso de uso de los resultados derivados de las muestras para obtener conclusiones acerca de las características de una población.

Un estimador  $\bar{\theta}$  es un valor que puede calcularse a partir de los datos muestrales que proporcionan información sobre el valor del parámetro  $\theta$ .

#### 4.2.1. Propiedades

##### ■ INSESGADEZ:

Un estimador insesgado es aquel cuya esperanza matemática coincide con el valor del parámetro que se desea estimar. En caso de no coincidir se dice que el estimador tiene sesgo.

$$\begin{aligned} E(\bar{\theta}) &= \theta \\ E(\bar{\theta}) - \theta &= 0 \\ \text{Sesgo} &= 0 \end{aligned}$$

##### ■ EFICIENCIA:

Un estimador es eficiente u óptimo cuando posee varianza mínima o bien en términos relativos cuando presenta menor varianza que otro ( $\text{VAR}(\bar{\theta}_1) < \text{VAR}(\bar{\theta}_2)$ ).

Quedando claro que el hecho puede plantearse también en términos más coherentes de Error Cuadrático Medio (ECM). Cuanto menor es la eficiencia, menor es la confianza de que el estadístico obtenido en la muestra aproxime al parámetro poblacional.

■ CONSISTENCIA:

Un estimador consistente es aquel cuyo error de medida o sesgo se aproxima a cero cuando el tamaño de la muestra tiende a infinito. De la definición de estimador insesgado, podemos extraer la conclusión de que, en ocasiones, tenemos errores de estimación. Ahora bien, existen casos en los que cuando la muestra se va haciendo más grande el error disminuye.

### 4.3. Método de Mínimos Cuadrados Ordinarios

Uno de los puntos determinantes en la econometría se basa en el procesamiento estadístico y para ello el método de Mínimos Cuadrados Ordinarios MCO permite encontrar los Mejores Estimadores Lineales Insesgados.

Este método presenta muchas ventajas en cuanto a lo fácil de su uso y por lo adecuado del planteamiento estadístico matemático que permite adecuarse a los supuestos para los modelos econométricos. El término de MCO está vinculado con la regresión y la correlación, ambas determinan la existencia de relación entre dos o más variables.

De:

$$Y = X\hat{\beta} + e$$

Se obtiene:

$$e = Y - X\hat{\beta}$$

$$\text{Min. } \sum_{i=1}^n e_i^2 \Rightarrow e'e$$

$$e'e = (Y - X\beta)'(Y - X\beta)$$

$$e'e = (Y' - \beta'X')(Y - X\beta)$$

$$e'e = Y'Y - \underbrace{Y'X\beta - \beta'X'Y}_{Y'X\beta = \beta'X'Y} + \beta'X'X\beta$$

$$e'e = Y'Y - 2\beta'X'Y + \beta'X'X\beta$$

Si es mínimo cumple que :

$$\frac{de'e}{d\beta} = 0$$

$$\frac{d^2e'e}{d\beta\beta'} \geq 0$$

entonces:

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

$$-2X'Y + 2X'X\beta = 0$$

$$X'X\beta = X'Y$$

$$\hat{\beta}_{MCO} = (X'X)^{-1}X'Y$$

y:

$$\frac{d(-2(X'Y) + 2(X'X)\beta)}{d\beta'} \geq 0$$

$$2(X'X) \geq 0$$

#### 4.3.1. Supuestos del Modelo

- **Linealidad de Parámetros :**

El supuesto de linealidad permite que el efecto marginal solo dependa del parámetro.

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + e$$

$$\frac{dY}{dX_1} = \beta_1$$

- **Ausencia de Multicolinealidad - No Singularidad :**

La multicolinealidad es la relación de dependencia lineal fuerte entre más de dos variables explicativas en una regresión múltiple que incumple el supuesto de Gauss-Markov cuando es exacta. Resaltamos que la relación lineal (correlación) entre variables explicativas tiene que ser fuerte. Es muy habitual que las variables explicativas de la regresión estén correlacionadas. Entonces, se debe puntualizar que esta relación debe ser fuerte, pero nunca perfecta, para que sea considerada un caso de multicolinealidad. La relación lineal sería perfecta si el coeficiente de correlación fuese 1.

Cabe resaltar que si la matriz  $X'X$  es singular no se podría estimar el matriz de vectores  $\hat{\beta}$ .

- **El término de error tiene distribución normal :**

El término de error  $e$  sigue una distribución normal con media 0 y varianza ( $\sigma^2$ ). Este supuesto nos permitirá validar la inferencia estadística usando las pruebas correspondientes.

$$e_i \sim N(0, \sigma^2)$$

- **Exogeneidad Estricta - Ortogonalidad :**

Del anterior supuesto observamos que los valores de  $X$  no tiene relación con el error del modelo, es decir por LEY DE LAS EXPECTATIVAS INTEGRADAS implica que la esperanza no condicional es cero ,  $E(E(e_i/X)) = E(e) = 0$ .

El incumplimiento de este supuesto genera , endogeneidad, problemas de sesgo e inconsistencia en los estimadores, con solución en la Técnica de Variables Instrumentales.

- **Varianza Constante - Homocedasticidad :**

El concepto de Homocedasticidad corresponde a la igualdad de varianza , este supuesto nos indica que el segundo momento condicional es una constante.

$$E(e_i^2/X) = \sigma^2 > 0$$

$$VAR[e_i^2/X] = E[e_i^2/X] - [e_i/X]^2$$

$$E[e_i^2/X] = \sigma^2 > 0$$

Presentar diferentes varianzas (Heterocedasticidad) genera problemas de Eficiencia en los estimadores,

- **Ausencia de Autocorrelación :**

Existe independencia de los errores ,  $COV(u_i, u_j), COV(u_t, u_{t-1}) = 0$  , La violación de este supuesto también genera problemas de Eficiencia en los estimadores.

- **Valores de X :**

La matriz  $X$  son valores fijos (Predeterminados).

Hallando los  $\hat{\beta}_{MCO}$  en un modelo de Regresión lineal simple

Sea :

$$Y = \alpha + \beta_1 X + e$$

$$RSS = \min \sum_{i=1}^n e_i^2 = e'e$$

$$e = Y - \alpha - \beta_1 X$$

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y - \alpha - \beta_1 X)^2$$

---

$$\frac{dRSS}{d\alpha} = 2 \sum_{i=1}^n (Y - \alpha - \beta_1 X)(-1) = 0$$

$$\sum_{i=1}^n Y - n\alpha - \beta_1 \sum_{i=1}^n X = 0$$

$$\boxed{\hat{\alpha}_{mco} = \bar{Y} - \beta_1 \bar{X}}$$

---

$$\frac{dRSS}{d\beta_1} = 2 \sum_{i=1}^n (Y - \alpha - \beta_1 X)(-X) = 0$$

$$\sum_{i=1}^n YX - (\bar{Y} - \beta_1 \bar{X}) \sum_{i=1}^n X - \beta_1 \sum_{i=1}^n X^2 = 0$$

$$\sum YX - \bar{Y} \sum X + n\beta_1 \bar{X}^2 - \beta_1 \sum X^2 = 0$$

$$\beta_1 (\sum X^2 - n\bar{X}^2) = \sum YX - \bar{Y} \sum X$$

$$\frac{\sum YX - \bar{Y} \sum X}{\sum X^2 - n\bar{X}^2} = \frac{\sum YX - n\bar{Y}\bar{X}}{\sum X^2 - n\bar{X}^2}$$

$$\boxed{\hat{\beta}_{mco} = \frac{COV(XY)}{VAR(X)}}$$

Sea los vectores X,Y que contienen información económica y se desea obtener los estimadores por MCO :

Y	X
20	32
23	39
37	48
22	39
32	49
45	62
32	43
15	30
39	52
35	56
42	62
42	60
18	41
26	49
16	32

Donde :

$$\sum YX = 21991$$

$$n\bar{Y}\bar{X} = 20542,4$$

$$\sum X^2 = 33754$$

$$n\bar{X}^2 = 32109,06667$$

Entonces:

$$\hat{\alpha}_{mco} = -11,14444354$$

$$\hat{\beta}_{mco} = 0,880643592$$

Verificando :

Source	SS	df	MS	Number of obs	=	15
Model	1275.70031	1	1275.70031	F(1, 13)	=	86.42
Residual	191.899692	13	14.7615148	Prob > F	=	0.0000
Total	1467.6	14	104.828571	R-squared	=	0.8692
				Adj R-squared	=	0.8592
				Root MSE	=	3.8421

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x	.8806436	.0947308	9.30	0.000	.6759901 1.085297
_cons	-11.14444	4.493744	-2.48	0.028	-20.85259 -1.4363

Resultado 1 : Regresión Lineal <sup>4</sup>

---

<sup>4</sup>output de software STATA



Considerando la recta generada :

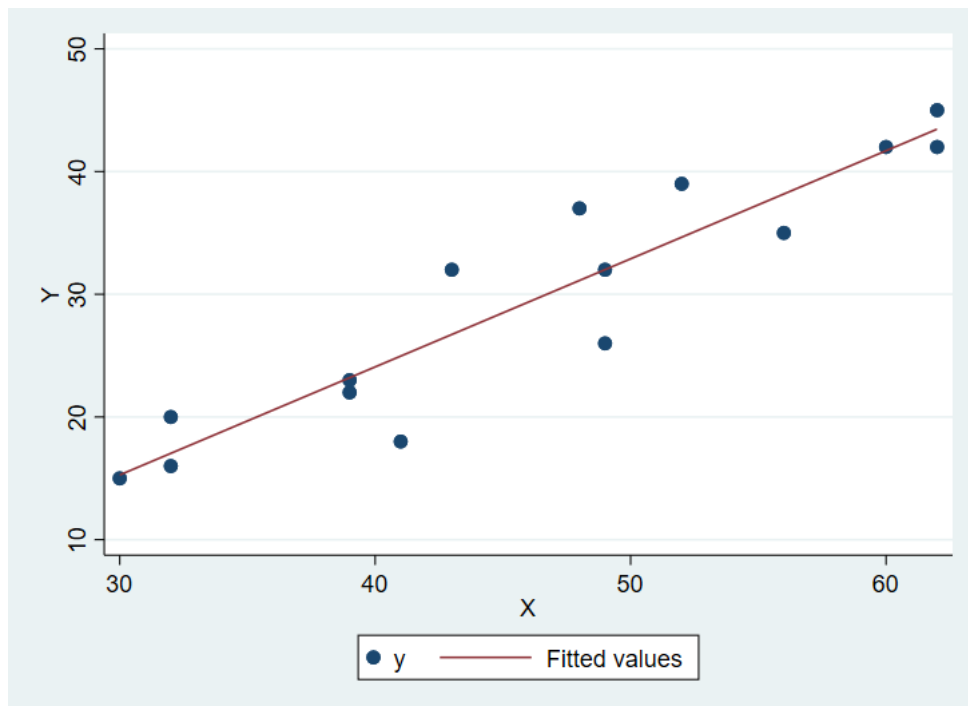


Gráfico 4 : Dispersión (Y vs X)

Una vez obtenido nuestros parámetros podemos estimar los valores de Y para así encontrar el término de error. Podemos observar un promedio cercano a 0 y una distribución que se asemeja a una distribución normal.

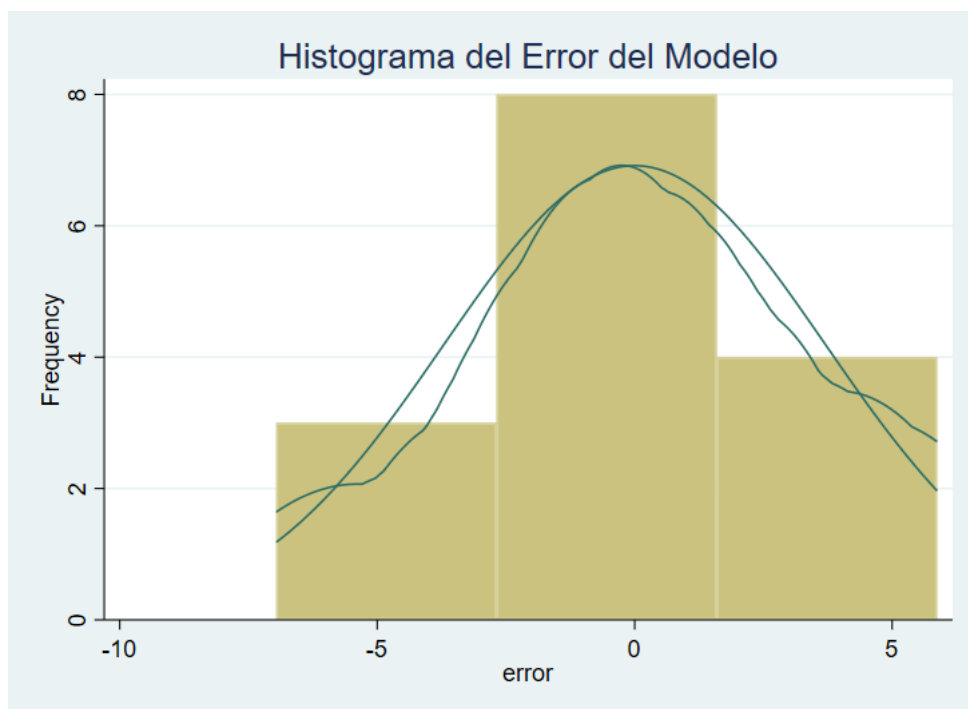


Gráfico 5 : Histograma del error

El presente taller finaliza en este punto, se considerará sus recomendaciones para los siguientes temas a tratar, por último se dejan softwares y bibliografía recomendada para una mejor aplicación y reforzamiento de los temas tratados. Es necesario mencionar que existe un análisis matemático mayor en los temas tratados que se recomienda revisar y según la necesidad de los participantes se podrán realizar.

## 5. Softwares Recomendados

- Eviews
- SPSS
- Stata
- R - RStudio
- Python
- Julia

## 6. Bibliografía Recomendada

- Juan Francisco Castro y Roddy Rivas-Llosa (2014) Econometría Aplicada (1<sup>a</sup> Edición)
- Damodar N. Gujarati Dawn C. Porter (2010) Econometría (5<sup>a</sup> Edición)
- William H. Greene (2012) Econometric Analysis (7<sup>a</sup> Edición)
- Jeffrey M. Wooldridge (2010) Introducción a la econometría (4<sup>a</sup> Edición)
- A. Colin Cameron A. Colin Cameron (2005) Microeconometrics Methods and Applications (1<sup>a</sup> Edición)
- James D. Hamilton (1994) Time Series Analysis (1<sup>a</sup> Edición)