# LÍNEAS BASES Y EVALUACIÓN DE LOS IMPACTOS SOCIOECONÓMICOS DEL CAMBIO CLIMÁTICO EN AMÉRICA LATINA
# DDSAH-CEPAL

**COMPONENTE SOCIOECONÓMICO**
**PROGRAMA EUROCLIMA**

NACIONES UNIDAS

CEPAL

# Análisis econométrico con datos de sección cruzada

NACIONES UNIDAS

CEPAL

# TABLA DE CONTENIDO

1. Introducción: el modelo de regresión y supuestos

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i$$

Parámetros a estimar

Variable dependiente

Variables independientes

Error

**Dirección del análisis**

1. Estimar el impacto ( “+” o “–”) de cada VI sobre la VD
2. Cuantificar cada impacto

# 1. Introducción: el modelo de regresión y supuestos

Econometría: “parte de los métodos cuantitativos que emplea el economista”. Es la unión de la teoría económica, la estadística y la matemática que busca establecer relaciones entre variables económicas para predecir el impacto de una o mas variables sobre la variable denominada “respuesta”. (Gujarati, Wooldridge, Judge)

1. Planteamiento de la hipótesis → Las variaciones del clima afectan la productividad agrícola
2. Especificación del modelo matematico de la teoria → $Produccion = f(Tierra, insumos, agua, temperatura, precipitacion, otros)$
3. Especificacion del modelo econometrico de la teoria → $Produccion_i = \beta_o + \beta_1 T_i + \beta_2 I_i + \beta_3 Clima_i + U_i$
4. Estimación de los parámetros del modelo econométrico (Datos) → $Produccion_i = \hat{\beta}_o + \hat{\beta}_1 T_i + \hat{\beta}_2 I_i + \hat{\beta}_3 Clima_i$
5. Pruebas de hipótesis → $H_o: \beta_3 = 0 \;\; Vs. H_a: \beta_3 \neq 0$
6. Validación del modelo → Se cumplen los supuestos del MRL
7. Uso del modelo → Medición del impacto

$Y_i = \beta_o + \beta_1 X_i + u_i \; con \; i = 1,2,\dots,n \; individuos$

| | |
|---|---|
| $Y_1$ | $X_1$ |
| $Y_2$ | $X_2$ |
| . | . |
| $Y_i$ | $X_i$ |
| . | . |
| Yn | Xn |

$Y_t = \beta_o + \beta_1 X_t + u_t \; con \; t = 1,2,\dots,t$

| | |
|---|---|
| $Y_{1980}$ | $X_{1980}$ |
| $Y_{1981}$ | $X_{1981}$ |
| . | . |
| $Y_t$ | $X_t$ |
| : | . |
| $Y_{2009}$ | $X_{2009}$ |
| $Y_{2010}$ | $X_{2010}$ |

$Y_{it} = \beta_o + \beta_1 X_{it} + u_{it} \; con \; i = 1,2,\dots,n \; individuos; t = 1,2,\dots t$

| Individuo | Tiempo | Y | X |
|---|---|---|---|
| i=1 | t=1 | $Y_{11}$ | $X_{11}$ |
| | T=2 | $Y_{12}$ | $X_{12}$ |
| | . | . | . |
| | . | . | . |
| | . | . | . |
| | T=t | $Y_{1t}$ | $Y_{1t}$ |
| i=2 | t=1 | $Y_{21}$ | $X_{21}$ |
| | T=2 | $Y_{22}$ | $X_{22}$ |
| | . | . | . |
| | . | . | . |
| | . | . | . |
| | T=t | $Y_{2t}$ | $Y_{2t}$ |
| i=i | t=1 | $Y_{i1}$ | $X_{i1}$ |
| | T=2 | $Y_{i2}$ | $X_{i2}$ |
| | . | . | . |
| | . | . | . |
| | . | . | . |
| | T=t | $Y_{it}$ | $Y_{it}$ |
| i=n | t=1 | $Y_{n1}$ | $X_{n1}$ |
| | T=2 | $Y_{n2}$ | $X_{n2}$ |
| | . | . | . |
| | . | . | . |
| | . | . | . |
| | T=t | $Y_{nt}$ | $Y_{nt}$ |

# 1. Introducción: el modelo de regresión y supuestos

Supuesto 1: **Modelo de Regresión Lineal**

El modelo de regresión es lineal en los parámetros,

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

ó

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i$$

Para el modelo de regresión lineal múltiple.

Supuesto 2: **Los valores de X son fijos en muestreo repetido.**

Los valores que toma el regresor $X_j$ $(\forall j = 1,2,3,\ldots,k)$ son considerados fijos en muestreo repetido. Más técnicamente, se supone no estocástica.

Supuesto 3: **El valor medio de la perturbación $u_i$ es igual a cero.**

Dado el valor de $X$, la media, o el valor esperado del término aleatorio de perturbación $u_i$ es cero. Técnicamente, el valor de la media condicional de $u_i$ es cero. Simbólicamente, se tiene

$$E[u_i / X_i] = 0$$

ó

$$\mathrm{E}[\mathrm{u}_{\mathrm{i}} / \mathrm{X}_{2\mathrm{i}}, \mathrm{X}_{3\mathrm{i}}, \ldots, \mathrm{X}_{\mathrm{ki}}] = 0$$

# 1. Introducción: el modelo de regresión y supuestos

Supuesto 4: **Homoscedasticidad o igual varianza de $u_i$**

Dado el valor de $X$, la varianza de $u_i$ es la misma para todas las observaciones.

Esto es, las varianzas condicionales de $u_i$ son idénticas. Simbólicamente, se tiene que

$$\operatorname{var}[u_i / X_i] = E[u_i - E[u_i]/ X_i]^2$$
$$\operatorname{var}[u_i / X_i] = E[u_i^2 / X_i]$$
$$\operatorname{var}[u_i / X_i] = \sigma^2$$

Supuesto 5: **No auto correlación entre las perturbaciones.**

Dados dos valores cualquiera de $X$, $X_i$ y $X_j$ $(i \neq j)$, la correlación entre dos $u_i$ y $u_j$ cualquiera $(i \neq j)$ es cero. Simbólicamente,

$$\operatorname{cov}(u_i, u_j / X_i, X_j) = E[u_i - E[u_i]/ X_i][u_j - E[u_j]/ X_j]$$
$$\operatorname{cov}(u_i, u_j / X_i, X_j) = E[u_i / X_i][u_j / X_j]$$
$$\operatorname{cov}(u_i, u_j / X_i, X_j) = 0$$

Supuesto 6: **La covarianza entre $u_i$ y $X_i$ es cero, o $E[u_i, X_i] = 0$.**

Formalmente,

$$\operatorname{cov}[u_i, X_i] = E[u_i - E[u_i]][X_i - E[X_i]]$$
$$\operatorname{cov}[u_i, X_i] = E[u_i (X_i - E[X_i])] \qquad E[u_i] = 0$$
$$\operatorname{cov}[u_i, X_i] = E[u_i X_i] - E[X_i] E[u_i] \qquad E[X_i]\_no\_estocastica$$
$$\operatorname{cov}[u_i, X_i] = E[u_i X_i], \qquad E[u_i] = 0$$
$$\operatorname{cov}[u_i, X_i] = 0$$

# 1. Introducción: el modelo de regresión y supuestos

Supuesto7: **El número de observaciones n debe ser mayor que el número de parámetros por estimar.**

Supuesto 8: **Variabilidad en los valores de $X$ .**

No todos los valores de $X$ en una muestra dada deben ser iguales. Técnicamente, $Var[X]$ debe ser un número positivo finito.

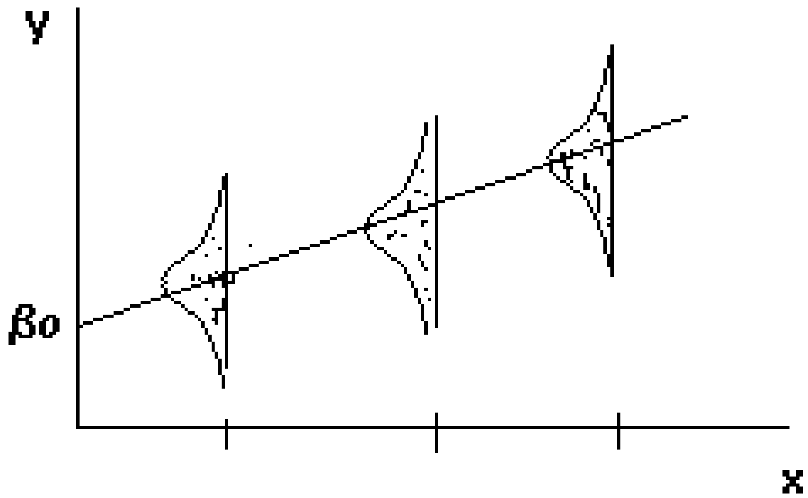
Supuesto 9: **El modelo de regresión esta correctamente especificado.**

Alternativamente, no hay un sesgo de especificación o error en el modelo utilizado en el análisis empírico.

La omisión de variables importantes del modelo, o la escogencia de una forma funcional equivocada, o la consideración de supuestos estocásticos equivocados sobre las variables del modelo, harán muy cuestionable la validez de la interpretación de la regresión estimada.

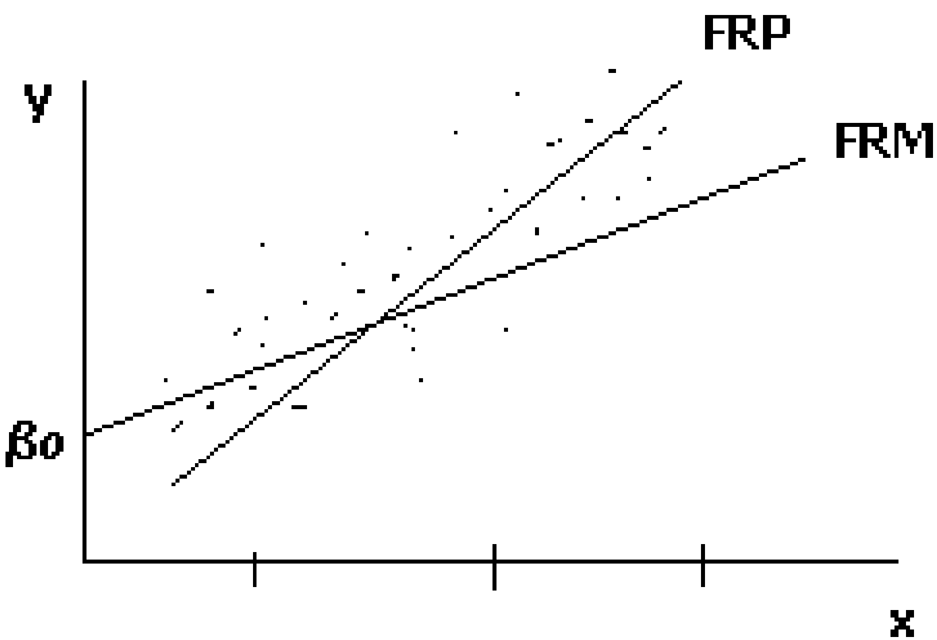
Supuesto 10: **No hay multicolinealidad perfecta**

2. Modelo de regresión lineal simple




La función de regresión poblacional denota la media poblacional de la distribución de Y dado un Xi, que esta relacionado con una forma funcional f(xi)

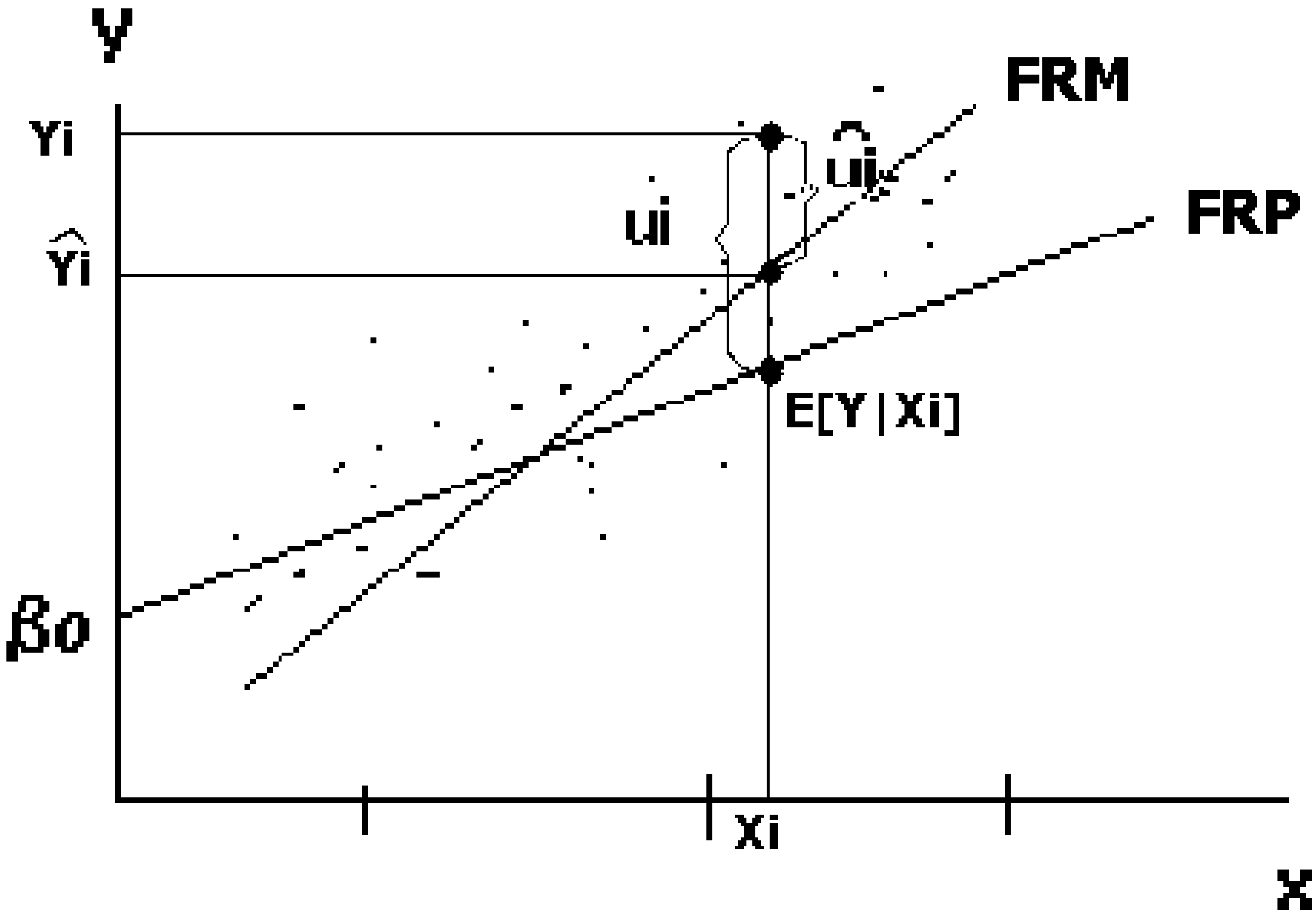
## 2. Modelo de regresión lineal simple




$$Y_i = \beta_o + \beta_1 X_i + U_i$$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

## 2. Modelo de regresión lineal simple




- Minimizar la suma de cuadrados de los errores estimados
- Son los estimadores que minimizan la SCE

# Estimadores de MCO

- Min {SCE}

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

Naciones Unidas

CEPAL

3. Modelo de regresión lineal múltiple

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i \qquad \text{i=1,2,...,n}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_N \end{bmatrix} = \begin{bmatrix} 1 & X_{21} & 0 & . & . & . & X_{k1} \\ 1 & X_{22} & 0 & . & . & . & X_{k2} \\ . & . & . & . & & & . \\ . & . & . & & . & & . \\ . & . & . & & & . & . \\ 1 & X_{2n} & 0 & . & . & . & X_{kn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_k \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ . \\ . \\ . \\ u_n \end{bmatrix}$$

Matriz de VI

Vector de Error

$$\boxed{Y_{(nx1)}} = \boxed{X_{(nx(k+1))}} B_{((k+1)x1)} + U_{(nx1)}$$

Vector de Parámetros

Vector de VD

## 4. Estimación: por Mínimos cuadrados ordinarios (MCO)

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i$$

i=1,2,…,n

$$\vec{Y} = \vec{X}\vec{B} + \vec{U}$$

$$\boxed{\hat{\mathrm{B}} = (X^t X)^{-1} X^t y}$$ → Estimador por MCO de B

$$\boxed{Var[\hat{\mathrm{B}}] = (X^t X)^{-1} \sigma^2}$$ → Estimador por de la VAR (B)

$$\begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_k \end{bmatrix} \rightarrow \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ . \\ . \\ . \\ \hat{\beta}_k \end{bmatrix} \rightarrow \begin{bmatrix} \mathrm{var}(\hat{\beta}_1) \\ \mathrm{var}(\hat{\beta}_2) \\ . \\ . \\ . \\ \mathrm{var}(\hat{\beta}_k) \end{bmatrix}$$

5. Inferencia: realmente es modelo es bueno?

# Pruebas de inferencia individual

$$Y_i = \beta_1 + \beta_2 \boxed{X_{2i}} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i$$

$$\downarrow$$

$$\boxed{\hat{\beta}_2}$$

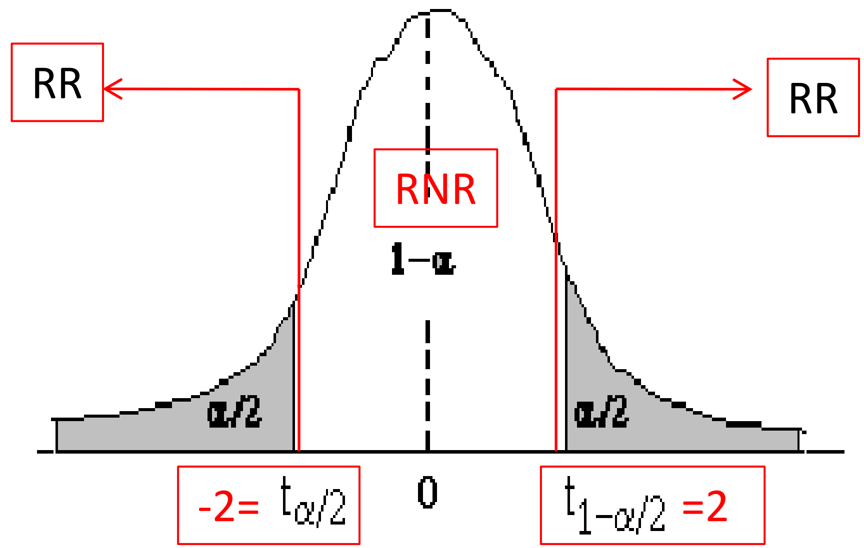
$$\downarrow$$

$$\boxed{\mathrm{var}(\hat{\beta}_2)}$$

$$\downarrow$$

$$\boxed{t = \frac{\hat{\beta}_2}{\sqrt{var(\hat{\beta}_2)}} \sim t_{n-(k+1)}}$$

Realmente la variable X2 afecta a la VD? $\boxed{H_o: \beta_2 = 0 \;\; Vs. H_a: \beta_2 \neq 0}$




T(0.95; n>30)= 2

5. Inferencia: realmente es modelo es bueno?

# Pruebas de inferencia Global

$$Y_i = \beta_1 + \boxed{\beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki}} + u_i$$

Al mismo tiempo, las variables X2, X3,…, Xk, explican a la VD ?

| *Fuente de Variación* | *Suma de cuadrados* | *Grados de libertad* | *Medias cuadráticas* | *F* |
|---|---|---|---|---|
| **Regresión** | SCReg | glReg=K+1 | $CMReg = \frac{SCReg}{glReg}$ | $F = \frac{CMReg}{CMErr} \sim F_{glReg,glErr}$ |
| **Error** | SCErr | glErr=N-(K+1) | $CMErr = \frac{SCErr}{glErr}$ | |
| **Total** | STC | N-1 | $Ho: \beta_1 = \beta_2 = \cdots = \beta_k = 0$ <br> $Ho: \beta_1 \neq \beta_2 \neq \cdots \neq \beta_k = 0$ | |

Naciones Unidas

CEPAL

# Pruebas de inferencia Global

$$\boxed{Y_i} = \beta_1 + \boxed{\beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki}} + \boxed{u_i}$$

TOTAL | REGRESION | ERROR

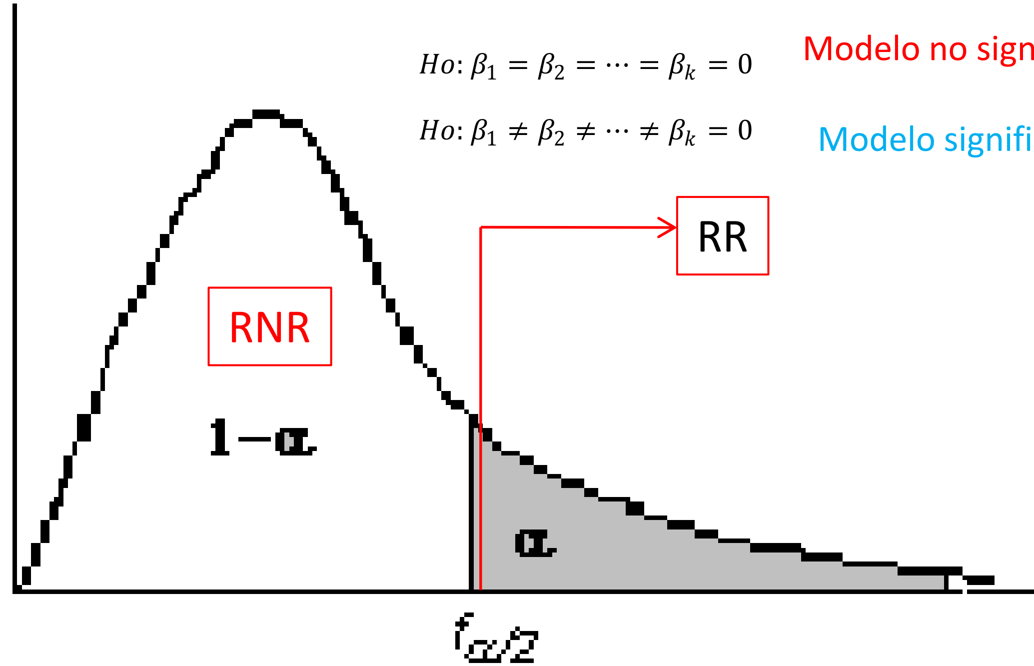
$$\boxed{\text{STC}} = \boxed{\text{SCReg}} + \boxed{\text{STErr}}$$

5. Inferencia: realmente es modelo es bueno?

# Pruebas de inferencia Global

$Ho: \beta_1 = \beta_2 = \cdots = \beta_k = 0$ Modelo no significativo globalmente

$Ho: \beta_1 \neq \beta_2 \neq \cdots \neq \beta_k = 0$ Modelo significativo globalmente




| | STC | SCReg | SCErr | |
|---|---|---|---|---|
| MODELO 1 | 100 | 90 | 10 | MODELO GLOBALMENTE SIGNIFICATIVO |
| MODELO 2 | 100 | 20 | 80 | MODELO NO SIGNIFICATIVO |

NACIONES UNIDAS
CEPAL

## 6. Formas funcionales de los modelos econométricos

Pueden ser consideradas distintas formas funcionales en que se relacionan la VD y las VI

$$\boxed{Y_i} = \beta_1 + \boxed{\beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki}} + u_i$$

Lineal

Lineal

$$\frac{\partial Y}{\partial X_k} = \frac{\partial}{\partial X_k}(\beta_o + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + U_i) = \boxed{\beta_k}$$

Efecto marginal "ante un cambio absoluto de la VI "Xk", la VD responde en promedio en $\beta_k$ manteniendo las demás variables constantes"

$$\eta = elasticidad = \frac{\%Y}{\%X_k} = \frac{\frac{\Delta Y}{Y}}{\frac{\Delta X_k}{X}} = \frac{\Delta Y}{\Delta X_k} * \frac{X_k}{Y} = \frac{\partial Y}{\partial X_k} * \frac{X_k}{Y} = \boxed{\beta_k * \frac{X_k}{Y}}$$

Elasticidad "ante un cambio porcentual de la VI "Xk", la VD responde en promedio en $\beta_k * \frac{X_k}{Y}$

## 6. Formas funcionales de los modelos econométricos

| MODELO | ECUACION | INTERPRETACION DE LOS COEFICIENTES (Bk) | ELASTICIDAD (η) |
|---|---|---|---|
| **Lineal -Lineal** | $Y_i = \beta_o + \beta_1 X_i + U_i$ | Efecto marginal | $\beta_k * \frac{X_k}{Y}$ |
| **Logaritmico - Lineal** | $LN\ Y_i = \beta_o + \beta_1 X_i + U_i$ | Semielasticidad | $\beta_k * X_k$ |
| **Logaritmico - Logaritmico** | $LN\ Y_i = \beta_o + \beta_1 LN\ X_i + U_i$ | Elasticidad | $\beta_k$ |
| **Lineal - Logaritmico** | $Y_i = \beta_o + \beta_1 LN\ X_i + U_i$ | Semielasticidad | $\frac{\beta}{Y}$ |

$$Y = a_0 + a_1 * X_1 + a_2 * X_2 + a_3 * X_1^{\ 2} + a_4 * X_2^{\ 2} + a_5 * X_1 X_2 + \varepsilon$$

$$Y = a_0 + a_1 * \sqrt{X_1} + a_2 * \sqrt{X_2} + a_3 * X_1 + a_4 * X_2 + a_5 * \sqrt{X_1 X_2} + \varepsilon$$

**Formas funcionales polinómicas**

$$Y = A * X_1^{\ \alpha} * X_2^{\ \beta} e^{u}$$

**Modelo Coob - Douglas**

NACIONES UNIDAS

CEPAL

7. Regresión con Variable Independiente Dicotómica

- ✓ Las variables explicadas no solo dependen de VI continuas…. También dependen de variables dicótomas, categóricas, de casilla, discretas, indicadoras, binarias…

- ✓ Cuatro formas de medir en estadísticas:

  - ✓ VARIABLES NOMINALES: Genero (dos categorías), Raza (mas de dos categorías)
  - ✓ VARIABLES ORDINALES: estrato (1, 2, 3, 4, 5, y 6) , Tamaño del predio (Grande, Mediano, Pequeño), Tamaño de la industria, ….
  - ✓ VARIABLES CONTINUA
  - ✓ VARIABLE DE RAZON

1 individuo

Mujer=1

Hombre=0

Pertenece a un programa=1

No Pertenece a un programa=1

1 empresa

Grande=1

Pyme=1

Sector=1

Sector=2

Sector=3

## 7. Regresión con Variable Independiente Dicotómica

El tratamiento de este tipo de variables será el siguiente:

1. Se define la variable dummy Di, como:
   - Di=1 si la unidad de análisis posee la característica
   - Di=0 si la unidad de análisis no posee la característica

2. Siempre se tomara el "1" como referencia:

3. Si la variable dummy presenta mas de dos categorías, se crearan (k-1) dummy , donde K es el numero de categorías.

   1. Tamaño de las empresas (grande (G), Mediana (M) y Pyme (P). Aquí K=3, entonces se crean 2 dummy
   - D1=1 si la empresa es grande
   - D1=0 de lo contrario

   - D2=1 si la empresa es Mediana
   - D2=0 de lo contrario

Se deja a la categoría ( P) como referencia

## 7. Regresión con Variable Independiente Dicotómica

Se pueden presentar varios casos:

- Una variable dependiente continua y una variable independiente dicótoma de dos categorías
- Una variable dependiente continua y dos variables independientes: una continua y una dicótoma de dos categorías
- Una variable dependiente continua y dos variables independientes: una continua y una dicótoma de dos categorías
- Una variable dependiente continua y dos variables independientes: una continua y una dicótoma de dos categorías mas una interacción

# 7. Regresión con Variable Independiente Dicotómica

| DESCRIPCION | MODELO |
|---|---|
| Una variable dependiente continua y una variable independiente dicótoma de dos categorías | $Y_i = \beta_o + \beta_1 D_i + U_i$<br>$D_i = 1\ si\ posee\ la\ caracteristica$<br>$D_i = 0\ si\ no\ posee\ la\ caracteristica$ |
| Una variable dependiente continua y dos variables independientes: una continua y una dicótoma de dos categorías | $Y_i = \beta_o + \beta_1 X_i + \beta_2 D_i + U_i$<br>$D_i = 1\ si\ posee\ la\ caracteristica$<br>$D_i = 0\ si\ no\ posee\ la\ caracteristica$ |
| Una variable dependiente continua y dos variables independientes: una continua y una dicótoma de dos categorías más la interacción | $Y_i = \beta_o + \beta_1 X_i + \beta_2 D_i + \beta_{3*} * X_i * D_i + U_i$<br>$D_i = 1\ si\ posee\ la\ caracteristica$<br>$D_i = 0\ si\ no\ posee\ la\ caracteristica$ |
| Una variable dependiente continua y tres variables independientes: una continua y dos dicótomas de dos categorías | $Y_i = \beta_o + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 X_i + U_i$<br>$D_{1i} = 1\ si\ posee\ la\ caracteristica$<br>$D_{1i} = 0\ si\ no\ posee\ la\ caracteristica$<br>$D_{2i} = 1\ si\ posee\ la\ caracteristica$<br>$D_{2i} = 0\ si\ no\ posee\ la\ caracteristica$ |
| Una variable dependiente continua y tres variables independientes: una continua y dos dicótomas de dos categorías con interacción entre las dummy | $Y_i = \beta_o + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{1i} * D_{2i} + \beta_4 X_i + U_i$<br>$D_{1i} = 1\ si\ posee\ la\ caracteristica$<br>$D_{1i} = 0\ si\ no\ posee\ la\ caracteristica$<br>$D_{2i} = 1\ si\ posee\ la\ caracteristica$<br>$D_{2i} = 0\ si\ no\ posee\ la\ caracteristica$ |
| Una variable dependiente continua y dos variables independientes: una continúa y una dicótoma de más de dos categorías con k=4 | $Y_i = \beta_o + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{3i} + \beta_4 X_i + U_i$<br>$D_{1i} = 1\ si\ posee\ la\ caracteristica$<br>$D_{1i} = 0\ si\ no\ posee\ la\ caracteristica$<br>$D_{2i} = 1\ si\ posee\ la\ caracteristica$<br>$D_{2i} = 0\ si\ no\ posee\ la\ caracteristica$<br>$D_{3i} = 1\ si\ posee\ la\ caracteristica$<br>$D_{3i} = 0\ si\ no\ posee\ la\ caracteristica$ |

7. Regresión con Variable Independiente Dicotómica

Una variable dependiente continua y una variable independiente dicótoma de dos categorías

$$Y_i = \beta_o + \boxed{\beta_1} D_i + U_i$$

Mide la diferencia promedio en la variable respuesta Y de tener y no tener la característica

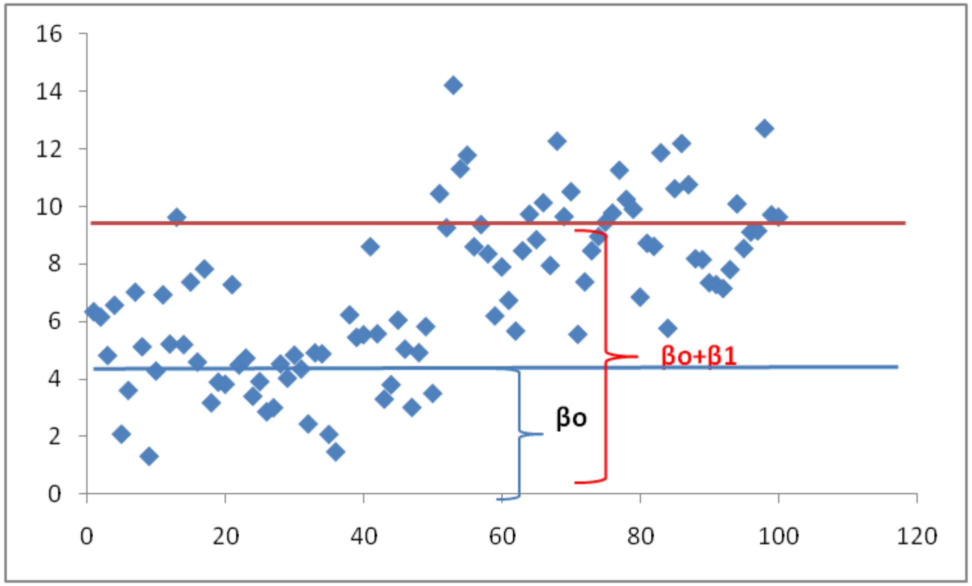
$$D_i = 1 \; si \; posee \; la \; caracteristica$$

$$D_i = 0 \; si \; no \; posee \; la \; caracteristica$$

$$E[Y_i / D_i = 1] = \beta_o + \beta_1$$

$$E[Y_i / D_i = 0] = \beta_o$$

$$E[Y_i / D_i = 1] - E[Y_i / D_i = 0] = \beta_1$$



## 7. Regresión con Variable Independiente Dicotómica

Una variable dependiente continua y dos variables independientes: una continua y una dicótoma de dos categorías

$$Y_i = \beta_o + \beta_1 X_i + \boxed{\beta_2} D_i + U_i$$

Mide la diferencia promedio en la variable respuesta Y de tener y no tener la característica

$$D_i = 1 \; si \; posee \; la \; caracteristica$$

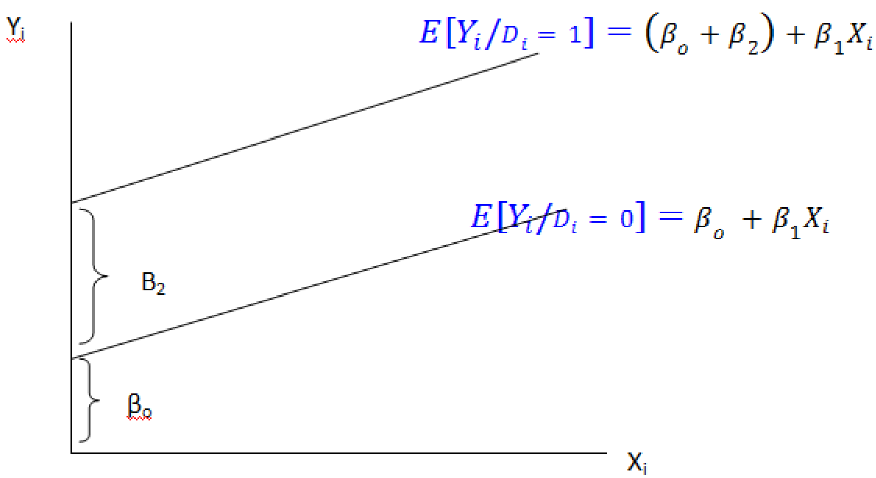
$$D_i = 0 \; si \; no \; posee \; la \; caracteristica$$

$$E[Y_i / D_i = 1] = (\beta_o + \beta_2) + \beta_1 X_i$$

$$E[Y_i / D_i = 0] = \beta_o + \beta_1 X_i$$

$$E[Y_i / D_i = 1] - E[Y_i / D_i = 0] = \beta_2$$

$$H_o: \beta_2 = 0 \;\; Vs. H_a: \beta_2 \neq 0$$



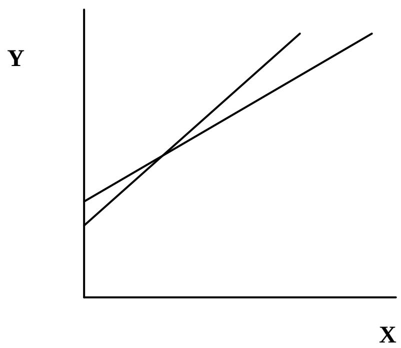
## 7. Regresión con Variable Independiente Dicotómica

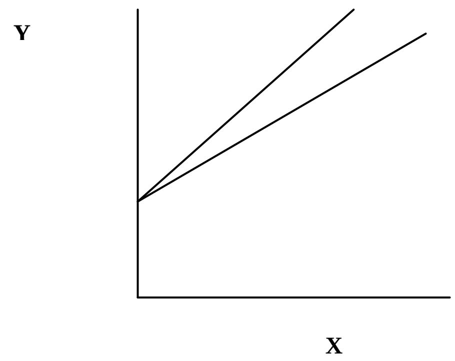
Una variable dependiente continua y dos variables independientes: una continua y una dicótoma de dos categorías más la interacción

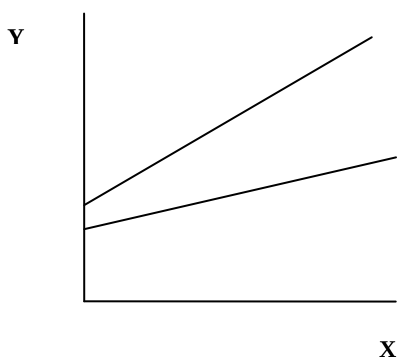
$$Y_i = \beta_o + \beta_1 X_i + \beta_2 D_i + \beta_{3*} * X_i * D_i + U_i$$

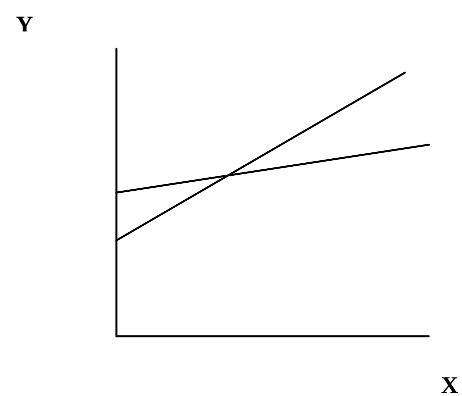
$$D_i = 1 \; si \; posee \; la \; caracteristica$$

$$D_i = 0 \; si \; no \; posee \; la \; caracteristica$$



8. Violación de los supuestos del modelo de regresión

- ✓MULTICOLINEALIDAD
- ✓HETEROCEDASTICIDAD
- ✓AUTOCORRELACION
- ✓NORMALIDAD DE LOS ERRORES

# MULTICOLINEALIDAD

Relaciones lineales entre variables independientes

$$Y_i = \beta_1 + \overbrace{\beta_2 \underbrace{X_{2i} + \beta_3 X_{3i}}_{Corr(X_2, X_3)} + \ldots + \beta_k X_{ki}}^{Corr(X_2, X_k)} + u_i$$

$$Corr(X_3, X_k)$$

La multicolinealidad puede deberse a los siguientes factores:

- El método de recolección de información empleado.
- Restricciones sobre el modelo o en la población que es objeto de muestreo.
- Especificación del modelo.
- Un modelo sobredeterminado

# MULTICOLINEALIDAD

## Consecuencias prácticas de la multicolinealidad

✓Aun cuando los estimadores de MCO son MELI, estos presentan varianzas y covarianzas grandes, que hacen difícil la estimación precisa.
✓Factor inflador de varianza
✓los intervalos de confianza tienden a ser mucho más amplios, conduciendo a una aceptación más fácil de la hipótesis nula .
✓El estadístico de los coeficientes tiende a ser no significativo.
✓Aun cuando el estadístico de uno o más coeficientes sea estadísticamente no significativo, el , la media global de bondad de ajuste, puede ser muy grande
✓Los estimadores MCO y sus errores estándar pueden ser sensibles a pequeños cambios en la información.

## Como detectarla?

✓Un elevado pero pocas razones significativas (y un valor  significativo)
✓Altas correlaciones entre parejas de Regresores
✓Regresiones Auxiliares:
✓Valores Propios e Índice de Condición
✓Factores de Tolerancia y de Inflación de Varianza

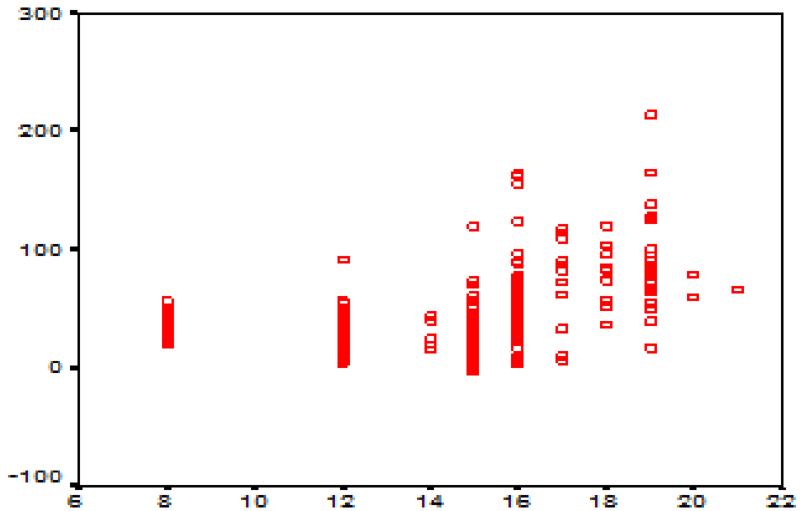
✓**Medidas remediales**

✓**Información a priori**
✓**Combinación de información de corte transversal y de series de tiempo**
✓**Eliminación de una(s) variable(s) y el sesgo de especificación**:
✓**Transformación de variables**:
✓**Datos nuevos o adicionales**
✓**Reducción de la colinealidad en las regresiones polinomiales**

# HETEROCEDASTICIDAD

La heteroscedasticidad se presenta cuando las varianzas de $u_i$ no son las mismas a lo largo de las observaciones. Simbólicamente,

$$E[u_i] = \sigma_i^2$$




✓A medida que aumentan las variables independientes las varianzas se hacen mayores.
✓Presencia de factores atípicos.
✓La inclusión o exclusión de una observación de este tipo
✓Incorrecta especificación del modelo

HETEROCEDASTICIDAD

Consecuencias de utilizar MCO en presencia de heteroscedasticidad

✓Los intervalos de confianza basados en los estimadores de MCO son muy amplios.
✓Como resultado, es probable que las pruebas  y  den resultados imprecisos
✓La característica más sobresaliente de estos resultados es que los MCO, con o sin corrección por heteroscedasticidad, sobreestiman consistentemente el verdadero error estándar obtenido mediante el procedimiento correcto.

Como detectarla?: Prueba de White.

Como corregirla:

**Cuando es conocida: Método de Mínimos Cuadrados Ponderados**

**Cuando es desconocida: Método de Mínimos Cuadrados Generalizados**

$$P\Omega P^{t} = I_{nxn}$$
$$PY = PXB + Pu$$
$$Y^{*} = X^{*}\mathrm{B}^{*} + u^{*}$$

NACIONES UNIDAS

CEPAL

# Autocorrelacion

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i$$

Dados dos valores cualquiera de $X$ , $X_i$ y $X_j$ $(i \neq j)$, la correlación entre dos $u_i$ y $u_j$ cualquiera $(i \neq j)$ es cero. Simbólicamente,

$$\operatorname{cov}(u_i, u_j / X_i, X_j) = E[u_i - E[u_i]/ X_i][u_j - E[u_j]/ X_j]$$

$$\operatorname{cov}(u_i, u_j / X_i, X_j) = E[u_i / X_i][u_j / X_j]$$

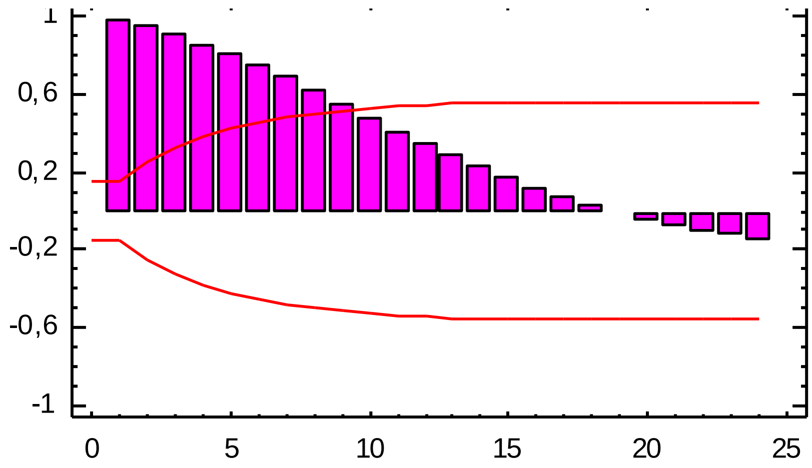
$$\operatorname{cov}(u_i, u_j / X_i, X_j) = 0$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_N \end{bmatrix} = \begin{bmatrix} 1 & X_{21} & 0 & . & . & . & X_{k1} \\ 1 & X_{22} & 0 & . & . & . & X_{k2} \\ . & . & . & . & & & . \\ . & . & . & & . & & . \\ . & . & . & & & . & . \\ 1 & X_{2n} & 0 & . & . & . & X_{kn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_k \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ . \\ . \\ . \\ u_n \end{bmatrix}$$

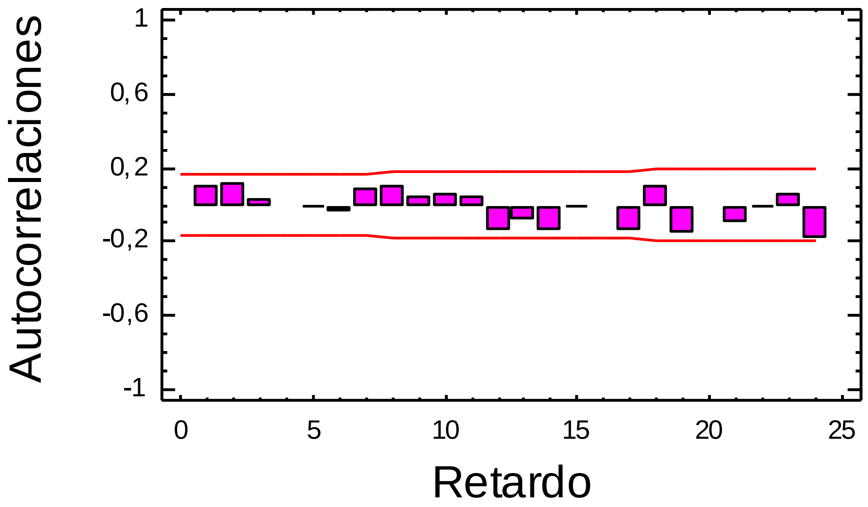
$Corr(u_1, u_2)$

$Corr(u_2, u_n)$

Errores autocorrelacionados




Errores  no autocorrelacionados



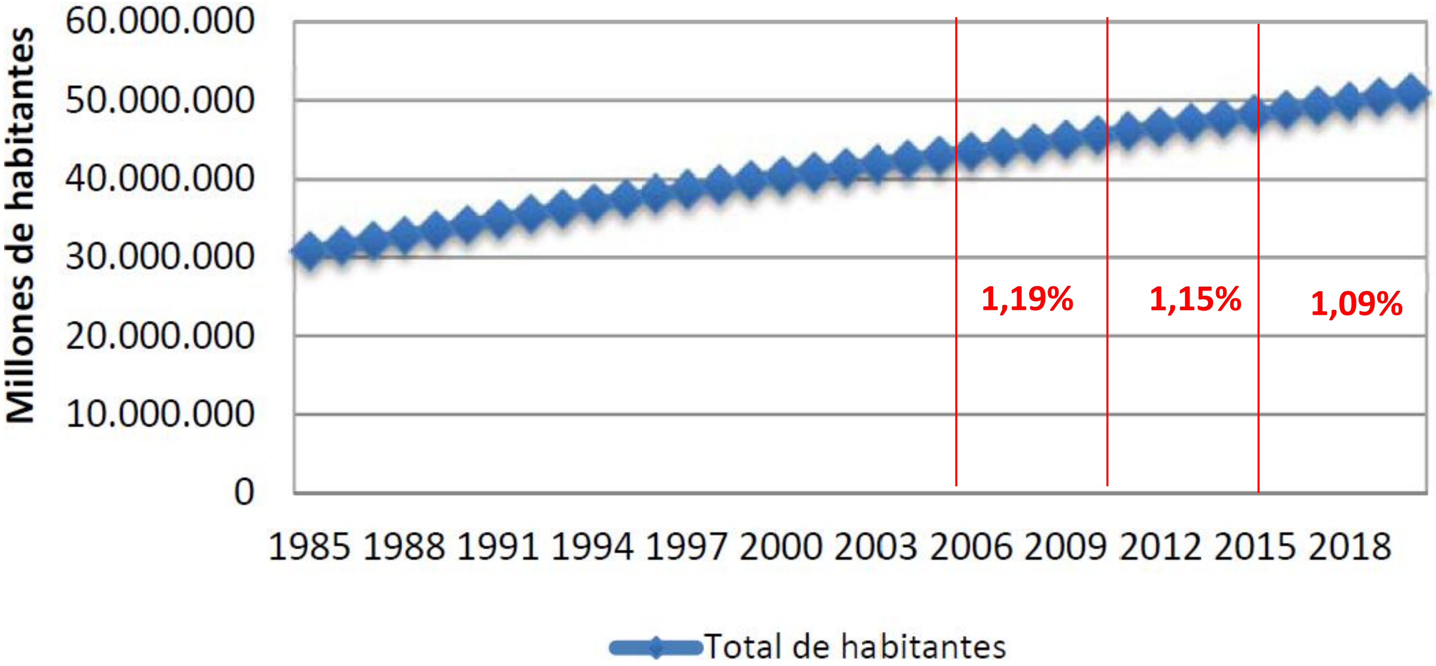
# Ejemplos……

NACIONES UNIDAS

CEPAL

# Proyecciones de población DANE. 1985 – 2020

- Población inicial: 42.888.592 a 2005
- Tasa de crecimiento 2000-2005: 1,25%
- Tasa de crecimiento proyectada 2015-2020: 1,09%
- Población final: 50.249.912 a 2020



Proyecciones de población DANE. 1985 – 2020

$$N(t) = k_1 + \frac{k_2}{(1 + \exp(\propto_o + \propto_1\ t))}$$

Dónde:

$N(t)$: Población estimada en el tiempo t

$k_1$: Límite inferior de la función logística

$k_2$: Límite superior de la función logística (Valor donde se estabiliza la población en el largo plazo)

$\propto_o$ y $\propto_1$: parámetros a estimar para la función logística con base en modelos econométricos y la información censal anterior.

Para la obtención de los pronósticos de población se emplearon valores para los parámetros $k_1$= 350.000 y $k_2$= 83.000.000. - Flórez C. (2000). Las transformaciones demográficas en Colombia durante el sigloXX.

$$N(t) - k_1 = \frac{k_2}{(1 + \exp(\propto_o + \propto_1 t))}$$

$$(1 + \exp(\propto_o + \propto_1 t)) = \frac{k_2}{N(t) - k_1}$$

$$\exp(\propto_o + \propto_1 t) = \left[\frac{k_2}{N(t) - k_1} - 1\right]$$

$$\propto_o + \propto_1 t = ln\left[\frac{k_2}{N(t) - k_1} - 1\right]$$

Reorganizando:

$$\underbrace{\boxed{ln\left[\frac{k_2}{N(t) - k_1} - 1\right]}}_{\text{Variable dependiente}} = \underbrace{\boxed{\propto_o + \propto_1 t}}_{\text{Variable independiente}}$$

| Censo | t | Población | U* | Ln (U)=Z |
|---|---|---|---|---|
| 1938 | 0 | 9.066.218 | 8,52247867 | 2,14270722 |
| 1961 | 13 | 12.379.910 | 5,89946974 | 1,77486247 |
| 1964 | 26 | 18.337.973 | 3,61419416 | 1,28486892 |
| 1973 | 35 | 23.881.851 | 2,52713435 | 0,92708599 |
| 1985 | 47 | 31.593.587 | 1,65654517 | 0,50473421 |
| 1993 | 55 | 37.422.791 | 1,23883872 | 0,21417443 |
| 2005 | 67 | 41.468.384 | 1,01856182 | 0,01839165 |

Variable independiente

Variable dependiente

**Tabla 1. Resultados de la estimación por MCO para los parámetros $\propto_o$ y $\propto_1$**

| | *Parámetros* | *Coeficientes* | *Error típico* | *Estadístico t* | *Probabilidad* |
|---|---|---|---|---|---|
| Intercepción | $\propto_o$ | 2,146394439*** | 0,057073084 | 37,60782294 | 2,50397E-07 |
| t | $\propto_1$ | -0,033571754*** | 0,001390491 | -24,14381077 | 2,27157E-06 |

***Significativo al 1%. R2= 0.99. Prob (F)= 0.000. n=7 observaciones.
Variable dependiente **Ln (U)=Z**

De esta manera, la función logística queda completamente especificada con los siguientes parámetros:

$$N(t) = 350.000 + \frac{83.000.000}{(1 + \exp(2,146394439 - 0,033571754t))}$$

Naciones Unidas

CEPAL

NACIONES UNIDAS

CEPAL

# Gracias por su atención

Harold Coronado Arango

Naciones Unidas

CEPAL