

CURSO: ECONOMETRÍA Y ANÁLISIS DE POLÍTICAS FISCALES

INSTRUCTOR: HORACIO CATALÁN ALONSO



Análisis Estadístico con STATA



-
- Stata se considera un “paquete para análisis estadístico”
 - Trabaja con datos de sección cruzada, series de tiempo y datos panel
 - Se puede utilizar en sistemas operativos de Windows, Macintosh, Unix y Linux

<http://www.stata.com>

- 
- Stata/MP: The fastest version of Stata (for dual-core and multicore/multiprocessor computers)
 - Stata/SE: Stata for large datasets
 - Stata/IC: Stata for moderate-sized datasets
 - Small Stata: A version of Stata that handles small datasets (for educational purchases only)
 - Numerics by Stata: Stata for embedded and web applications

Historial de instrucciones

Stata/MP 13.0 - [Results]

File Edit Data Graphics Statistics User Window Help

Review

Command _rc

1 clear

(R)

STATA 13.0

Statistics/Data Analysis

MP - Parallel Edition

Copyright 1985-2013 StataCorp LP
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC <http://www.stata.com>
979-696-4600 stata@stata.com
979-696-4601 (fax)

Notes:
1. (/v# option or -set maxvar-) 5000 maximum variables
2. New update available; type -update all-

Command

Lista de variables

Variables

Variable Label

There are no items to show.

Properties

Variables

Name	
Label	
Type	
Format	
Value Label	
Notes	

Data

Filename	
Label	
Notes	
Variables	0
Observations	0
Size	0
Memory	64M
Sorted by	

Ventana de Resultados

Editor de comandos

Características de las variables

Comandos básicos

Syntax

[prefix :] command [varlist] [=exp] [if] [in] [weight]
[using filename] [, options]

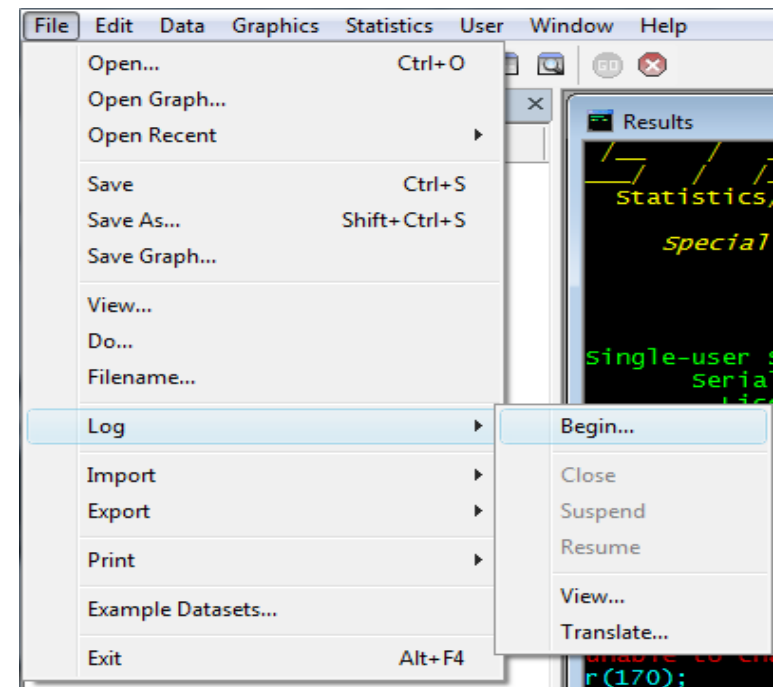
Cambio de directorio

```
cd c:\practica\stata_1\
```

Log files guarda los resultados de Stata la extensión es .log
se puede cambiar a txt

File/Log/Begin..

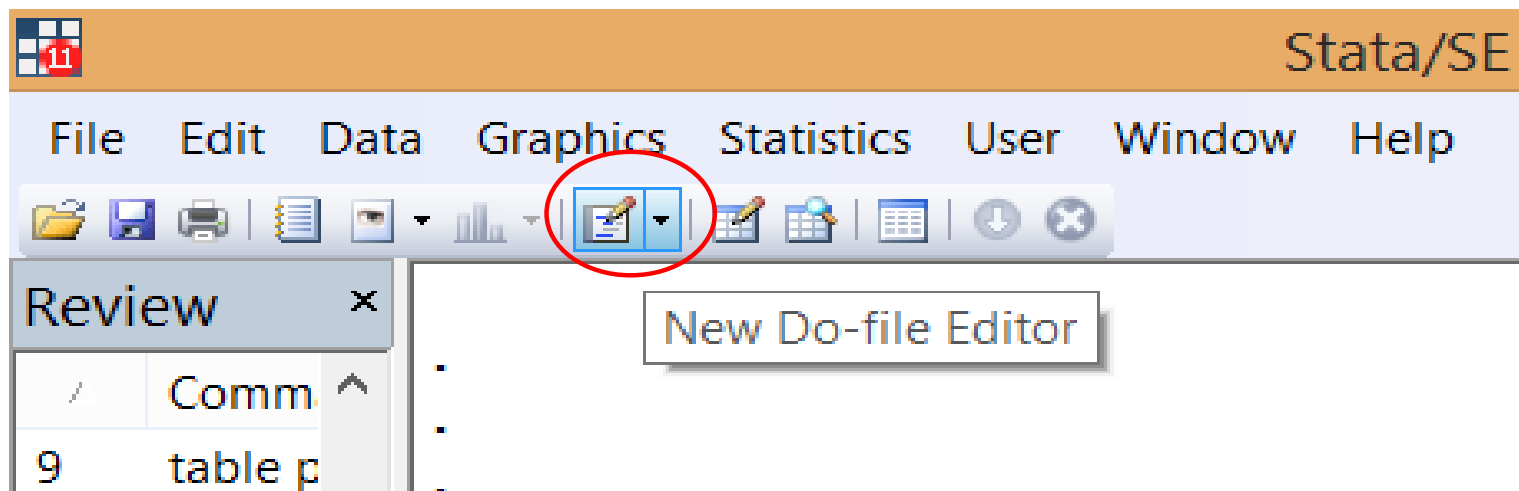
Se debe iniciar abriendo este archivo



Comandos básicos

Syntax

command [*nombre var*] [=operación] [if] [in] [, *opciones*]



- Crear un archivo con extensión “do”
- Permite editar y ejecutar las instrucciones


```
. tsset tiempo
```

```
time variable: tiempo, 1965 to 2014
```

```
delta: 1 unit
```

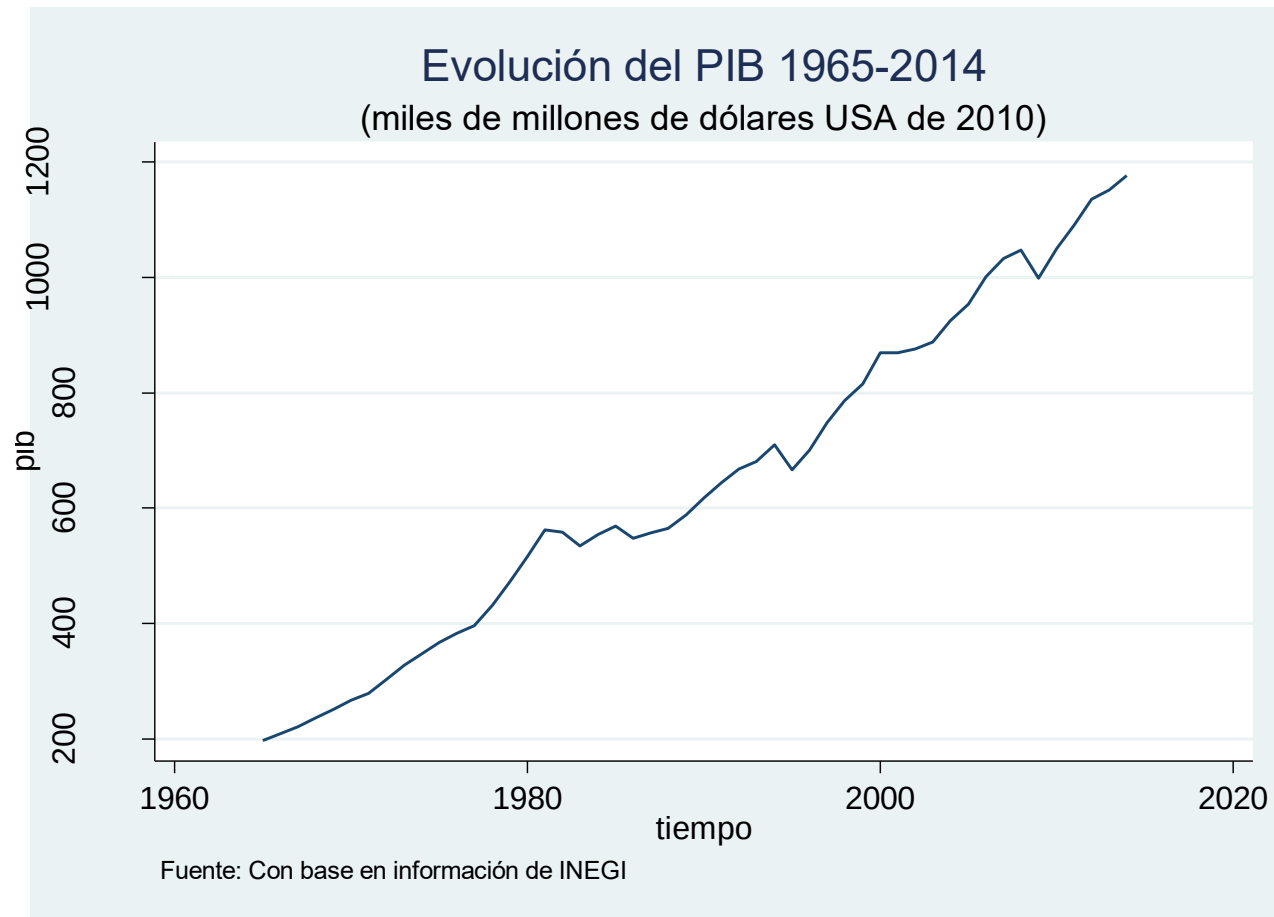
```
obs:      50  
vars:     10  
size:    1,900
```

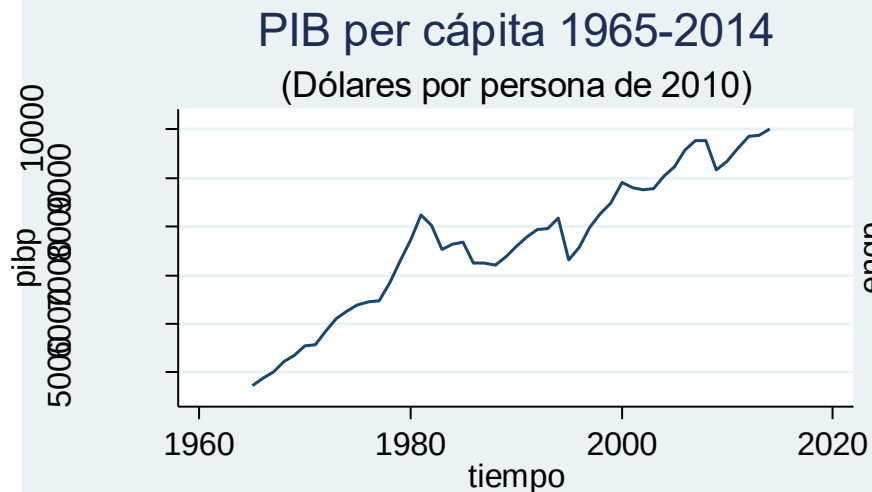
variable name	storage type	display format
tiempo	int	%8.0g
eng	float	%8.0g
pib	float	%8.0g
co2	float	%8.0g
p	float	%8.0g
pe	float	%8.0g
engp	float	%8.0g
pibp	float	%8.0g
co2p	float	%8.0g
ieng	float	%8.0g

```
Sorted by: tiempo
```

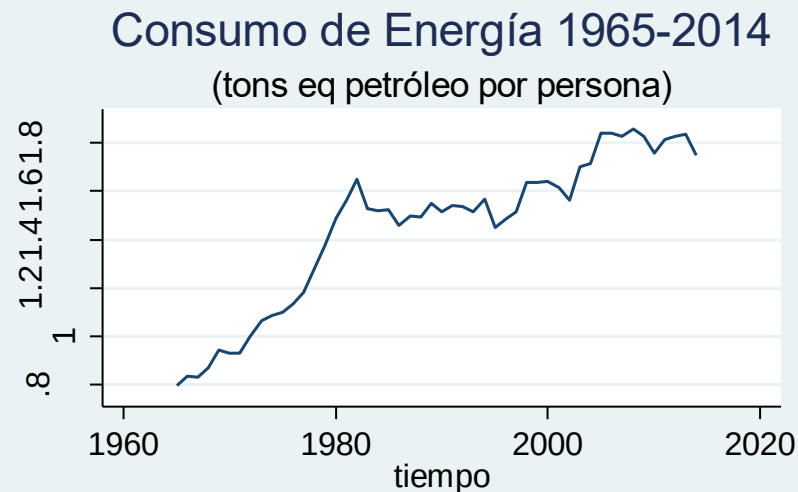
```
. summarize engp co2p pibp
```

Variable	Obs	Mean	Std. Dev.	Min	Max
engp	50	1.449195	.3167181	.7965844	1.856623
co2p	50	3.022972	.7354895	1.342707	3.935247
pibp	50	7669.737	1468.869	4729.16	9996.854

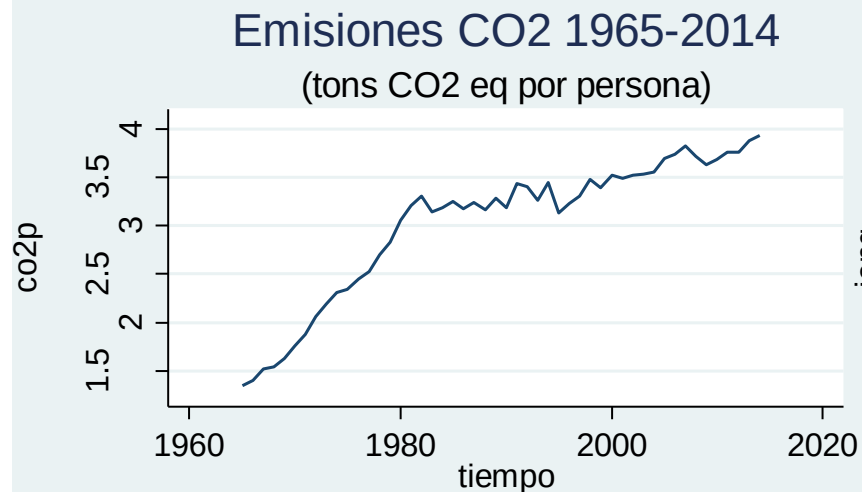




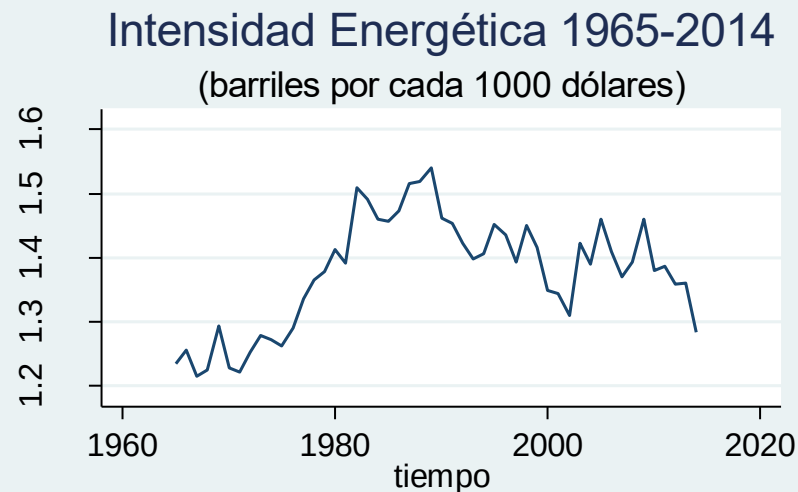
Fuente: Con base en información de INEGI



Fuente: Con base en información de SENER



Fuente: Con base en información de IEA



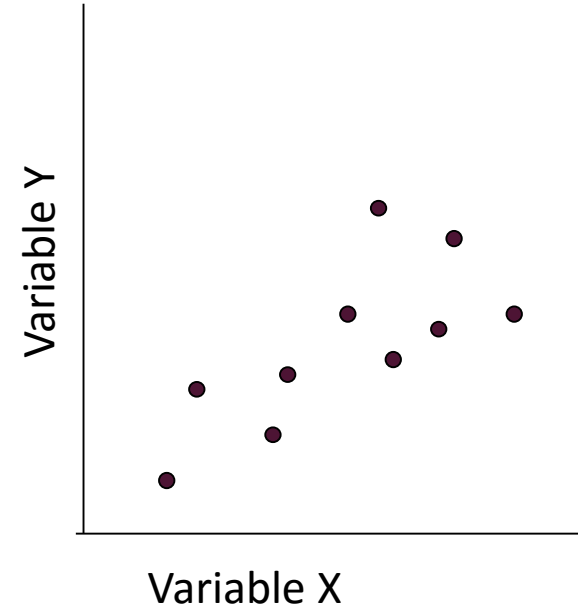
Fuente: Con base en información de IEA

GRÁFICOS DE DISPERSIÓN: Permite representar la evolución conjunta de ambas variables

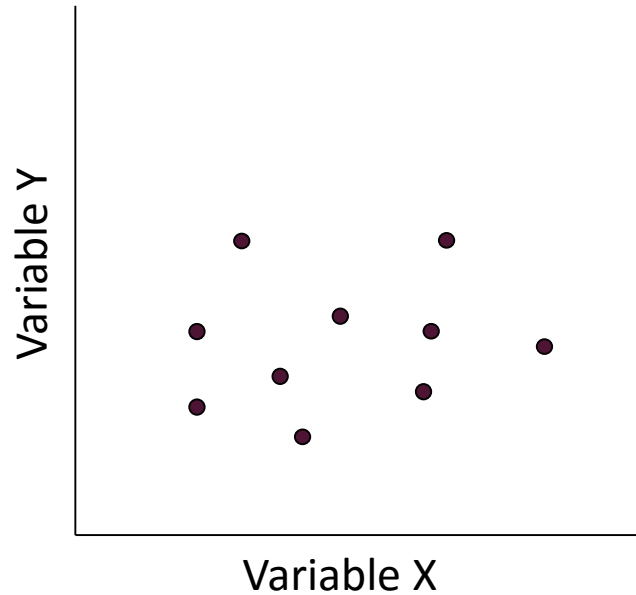
Dadas dos variables (Y, X) tomadas sobre el mismo elemento de la población, el diagrama de dispersión es simplemente un gráfico de dos dimensiones, donde en un eje (la abscisa) se sitúa una variable, y en el otro eje (la ordenada) se sitúa la otra variable

Si las variables están correlacionadas, el gráfico mostraría algún nivel de correlación (tendencia) entre las dos variables. Si no hay ninguna correlación, el gráfico presentaría una figura sin forma, una nube de puntos dispersos en el gráfico.

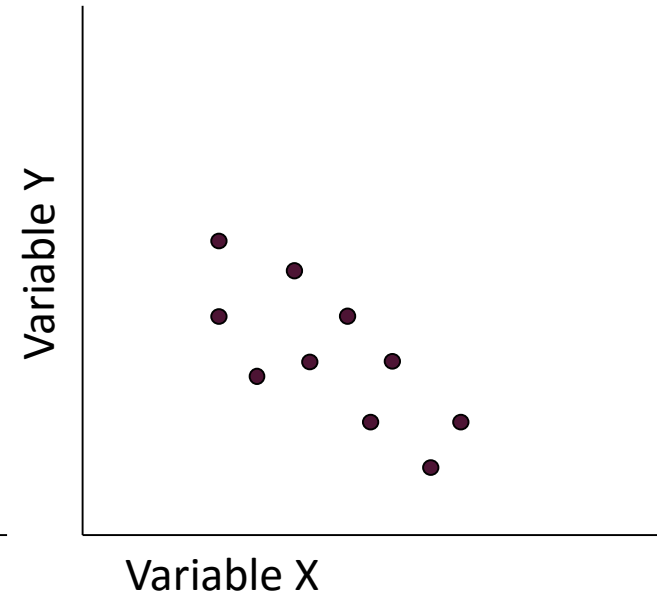
Representación gráfica de una relación



Relación lineal positiva

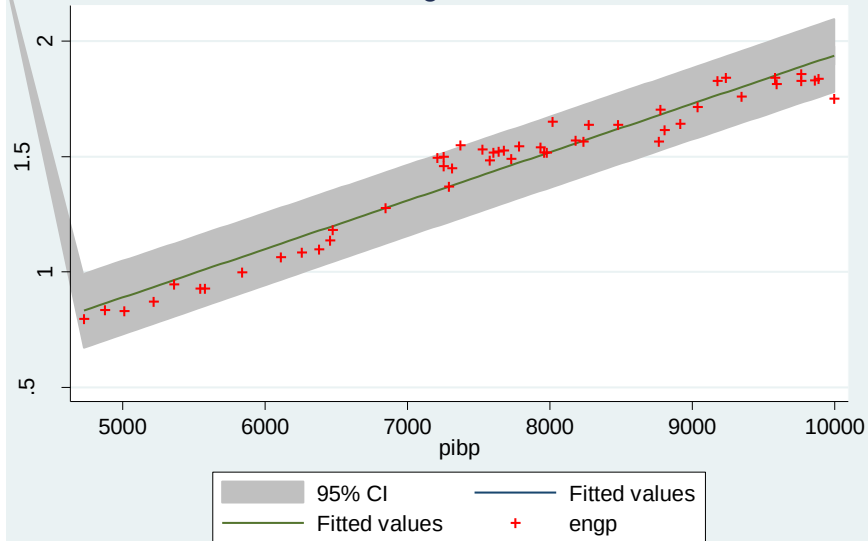


Sin relación

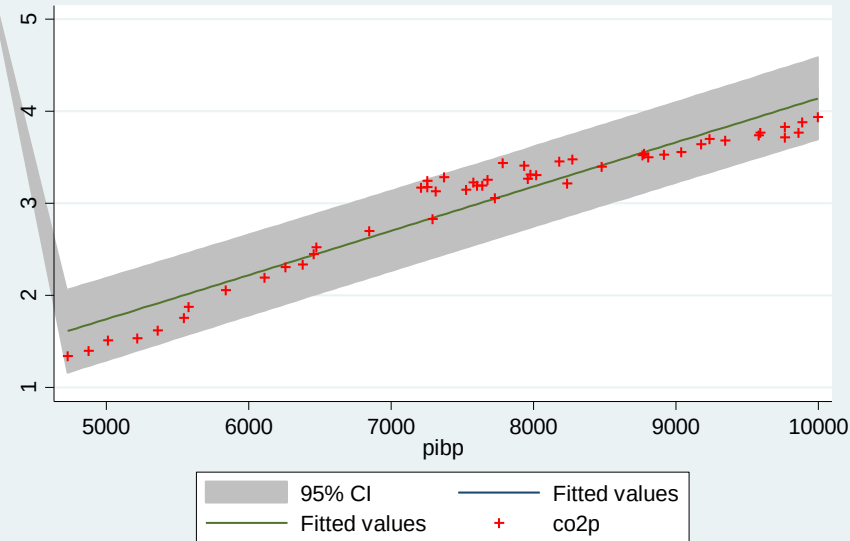


Relación lineal negativa

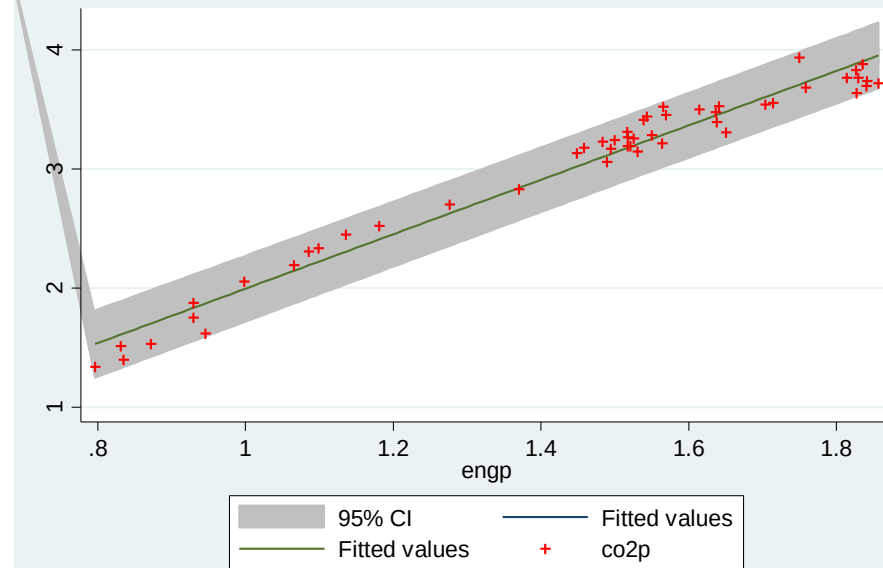
Energía vs PIB



Emisiones CO2 vs PIB

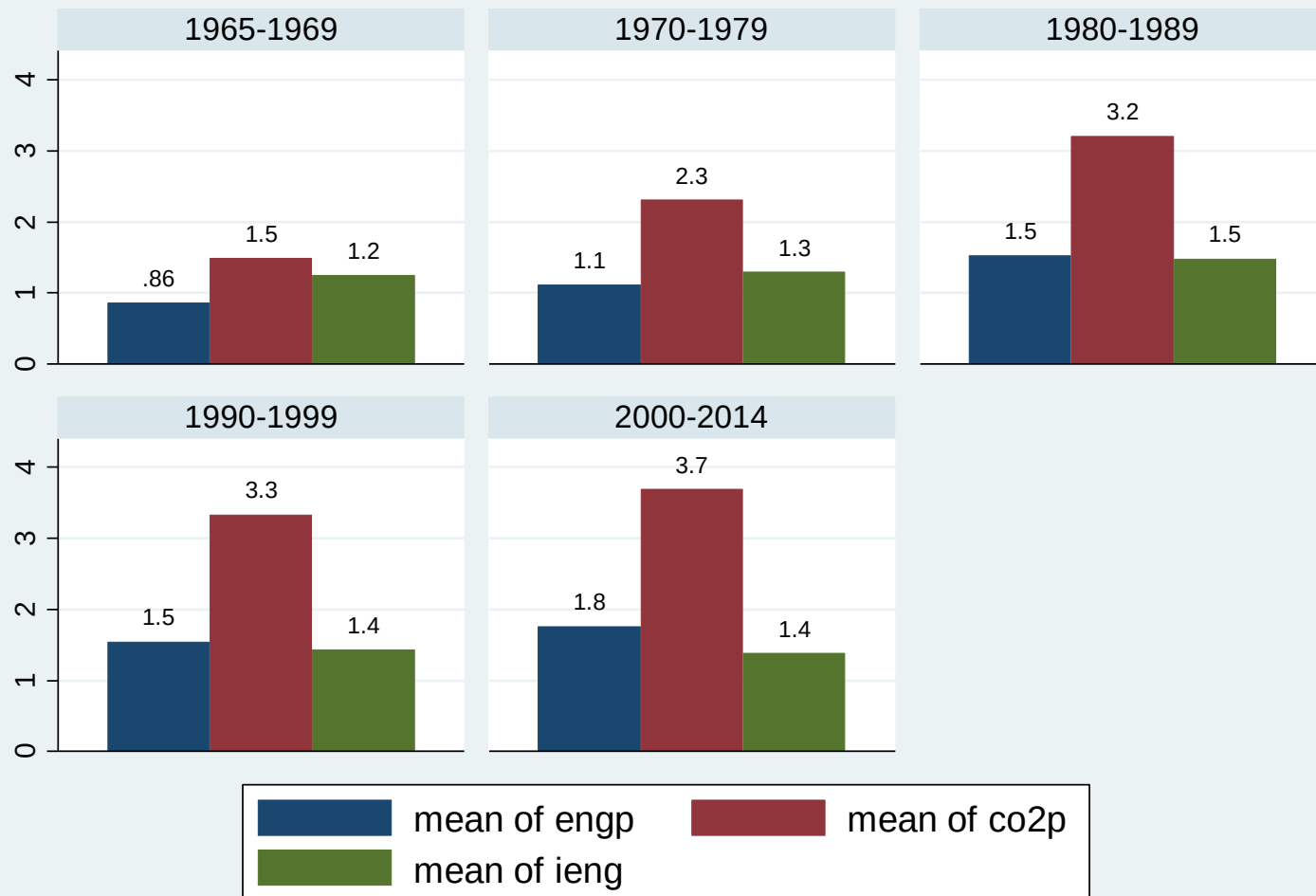


Emisiones CO2 vs ENG



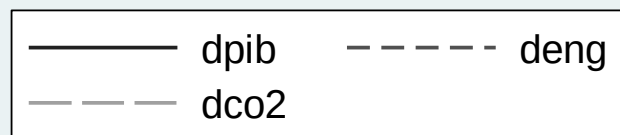
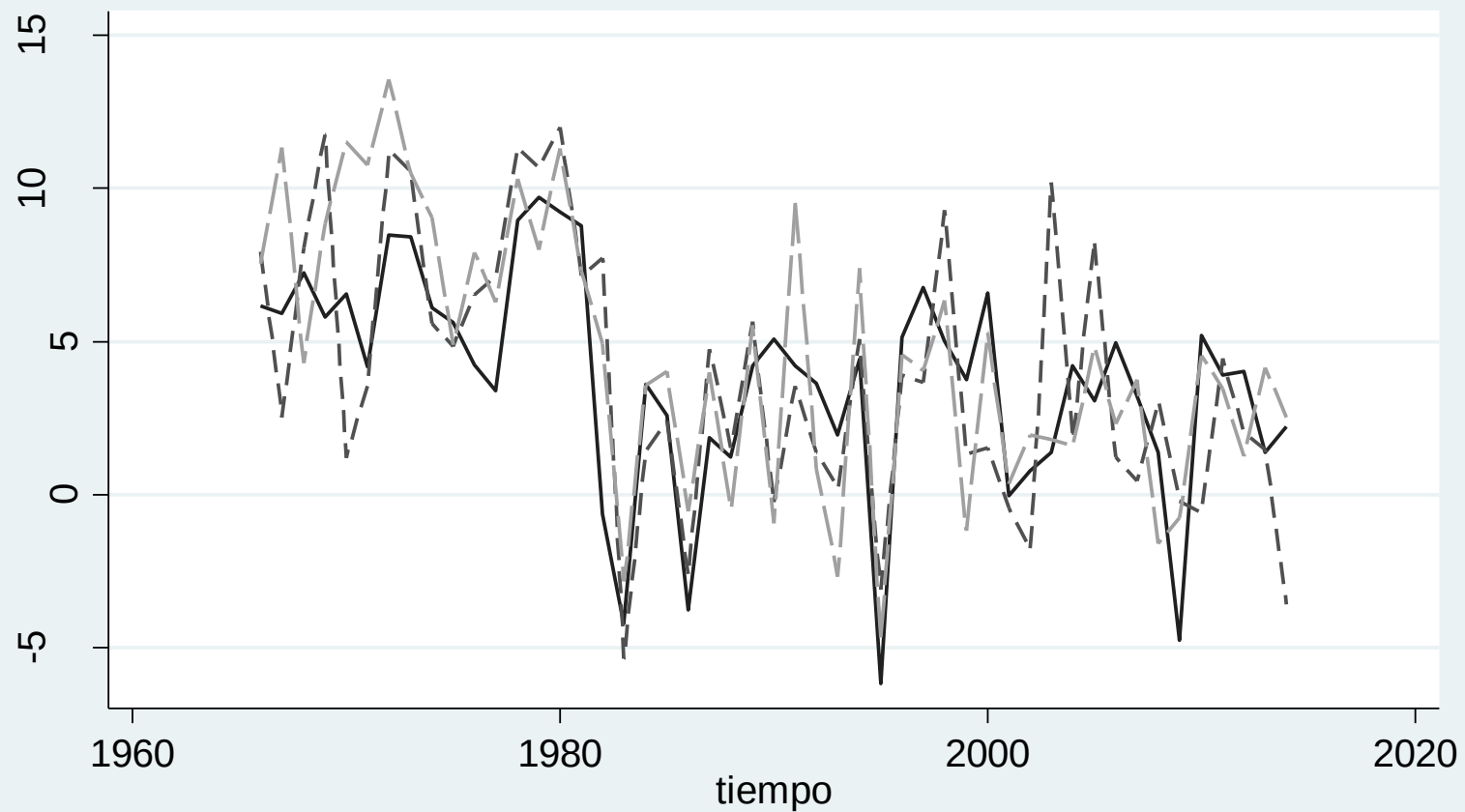
Summary statistics: mean
by categories of: periodo

periodo	pibp	engp	co2p	ieng
1965-1969	5038.483	.8558482	1.482975	1.244533
1970-1979	6277.621	1.106903	2.303677	1.28847
1980-1989	7591.397	1.528382	3.206242	1.477013
1990-1999	7908.306	1.540923	3.332397	1.428738
2000-2014	9368.08	1.761229	3.687371	1.378415
Total	7669.737	1.449195	3.022972	1.376822



Graphs by periodo

Tasas de crecimiento



periodo	dpib	engp	co2p	ieng
1965-1969	6.287934	.8558482	1.482975	1.244533
1970-1979	6.564035	1.106903	2.303677	1.28847
1980-1989	2.293078	1.528382	3.206242	1.477013
1990-1999	3.383059	1.540923	3.332397	1.428738
2000-2014	2.508442	1.761229	3.687371	1.378415
Total	3.779185	1.449195	3.022972	1.376822

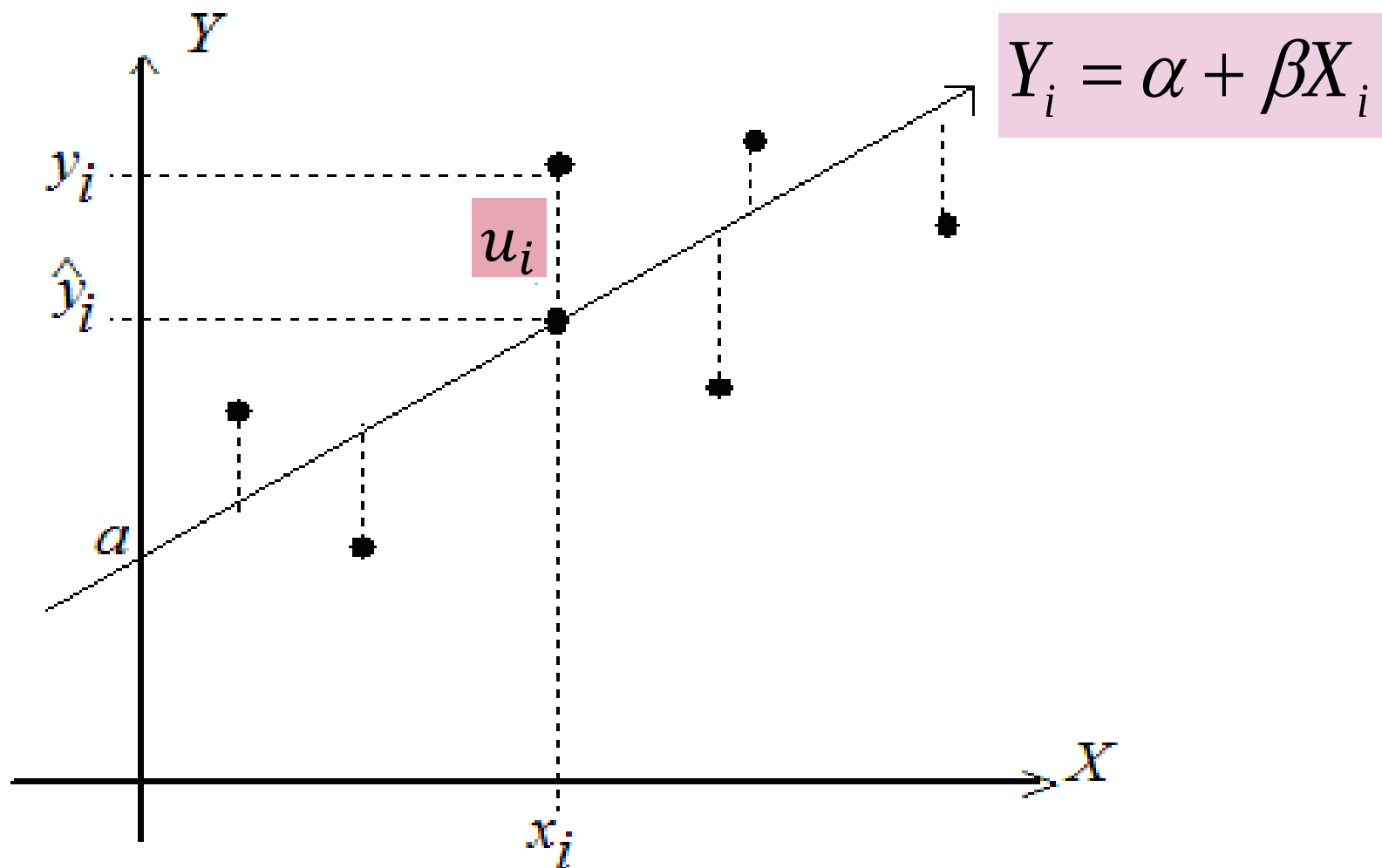
Modelo de Regresión Lineal

$$CO2p = \alpha + \beta PIBp$$

El modelo establece que un cambio en una unidad de del PIB per cápita produce u ocasiona un cambio en las emisiones per cápita de CO2, medido por el parámetro β

Este modelo podría ser adecuado si la línea recta pasará por todos los puntos del diagrama de dispersión

Debido a que los valores observados de la línea recta del modelo se define un erro



La diferencia entre el valor observado de la variable y la recta estimada se denomina error

```
. reg co2p pibp
```

Source	SS	df	MS
Model	24.3218194	1	24.3218194
Residual	2.18447801	48	.045509958
Total	26.5062974	49	.540944845

```
Number of obs =      50
F( 1,      48) =   534.43
Prob > F       =   0.0000
R-squared      =   0.9176
Adj R-squared  =   0.9159
Root MSE      =   .21333
```

co2p	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
pibp	.0004796	.0000207	23.12	0.000	.0004379	.0005214
_cons	-.655754	.1619649	-4.05	0.000	-.9814062	-.3301017

$$CO_2p = \alpha + \beta PIBp$$

$$CO_2p = -0.655 + 0.000479 PIBp$$

TABLA DE ANALISIS DE LA VARIANZA (ANOVA)

Source	SS	df	MS
Model	24.3218194	1	24.3218194
Residual	2.18447801	48	.045509958
Total	26.5062974	49	.540944845

$$RegSS = \sum_{i=1}^N (\hat{Y}_i - \bar{Y})^2$$

Regression Sum of Squares
Variabilidad explicada por el modelo

$$RSS = \sum_{i=1}^N (Y_i - \hat{Y}_i)^2$$

Residual Sum of Squares
Variabilidad **NO EXPLICADA**

$$TSS = \sum_{i=1}^N (\hat{Y}_i - \bar{Y})^2$$

Total Sum of Squares
Variabilidad Total

Fuente de Variación (Source)	SS	d.f.	MS
Model	RegSS	1	RegSS/1
Residual	RSS	n - 2	RSS/(n-2)
Total	TSS	n - 1	

MS (Mean Square)

$$s_e^2 = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{(n-2)} = \frac{RSS}{(n-2)}$$

$$s_e = \sqrt{s_e^2}$$

$$TSS = RegSS + RSS$$

Number of obs = 50

F(1, 48) = 534.43

Prob > F = 0.0000

R-squared = 0.9176

Adj R-squared = 0.9159

Root MSE = .21333

$$F = \frac{\text{RegMS}}{\text{RMS}} = \frac{\text{RegMS}}{\text{RSS}/(n-2)}$$

$$R^2 = 1 - \frac{\sum_{i=1}^N a_i^2}{\sum_{i=1}^N (Y_i - \bar{Y})^2}$$

$$R^2 = \frac{\text{TSS} - \text{RSS}}{\text{TSS}} = \frac{\text{RegSS}}{\text{TSS}}$$

$$\bar{R}^2 = 1 - (1 - R^2) \frac{N-1}{N-k}$$

$$s_e = \sqrt{s_e^2}$$

co2p	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
pibp	.0004796	.0000207	23.12	0.000	.0004379	.0005214
_cons	-.655754	.1619649	-4.05	0.000	-.9814062	-.3301017

co2p	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
pibp	β	$se(\beta)$	$t = \frac{\hat{\beta}}{se(\beta)}$	$P(t_{(N-K)} \geq t)$	$\beta \pm t_{(0.975, N-k)} se(\beta)$	
cons	α	$se(\alpha)$	$t = \frac{\hat{\alpha}}{se(\alpha)}$		$\alpha \pm t{(0.975, N-k)} se(\alpha)$	

$$\hat{\beta} = \frac{\sum_{i=1}^N CO2_i YP_i - \overline{CO2} \sum_{i=1}^N YP_i}{\left(\sum_{i=1}^N YP_i^2 - \overline{YP} \sum_{i=1}^N YP_i \right)}$$

$$\hat{\alpha} = \overline{CO2} - \hat{\beta} \overline{YP}$$


```
. reg co2p pibp pibp_2 engp
```

Source	SS	df	MS
Model	26.1230164	3	8.70767213
Residual	.383281013	46	.008332196
Total	26.5062974	49	.540944845

Number of obs = **50**
 F(3, 46) = **1045.06**
 Prob > F = **0.0000**
 R-squared = **0.9855**
 Adj R-squared = **0.9846**
 Root MSE = **.09128**

co2p	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
pibp	.0010882	.0001469	7.41	0.000	.0007926	.0013839
pibp_2	-5.64e-08	7.48e-09	-7.53	0.000	-7.14e-08	-4.13e-08
engp	1.113316	.2261746	4.92	0.000	.6580503	1.568582
_cons	-3.501951	.4303906	-8.14	0.000	-4.368282	-2.63562

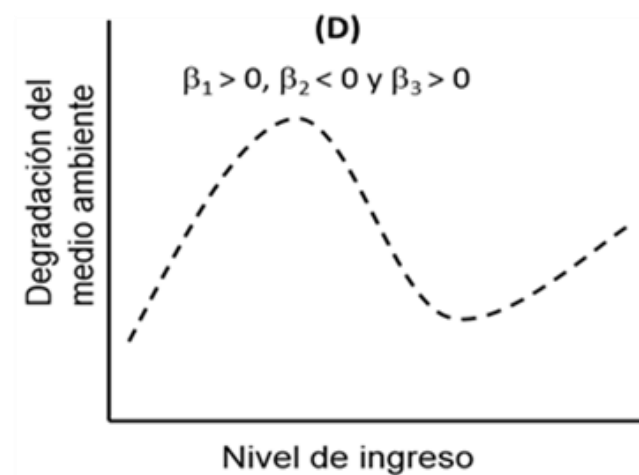
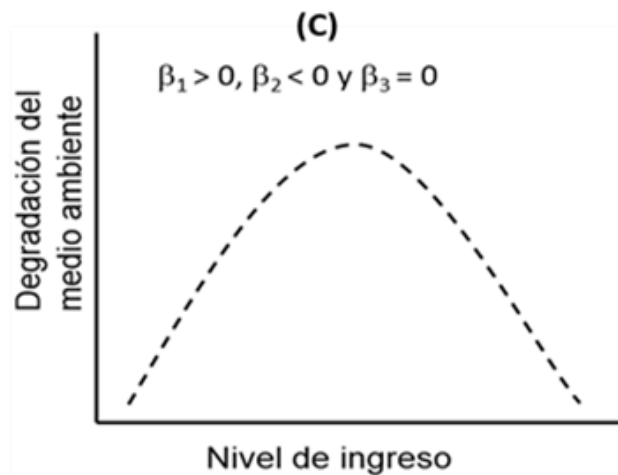
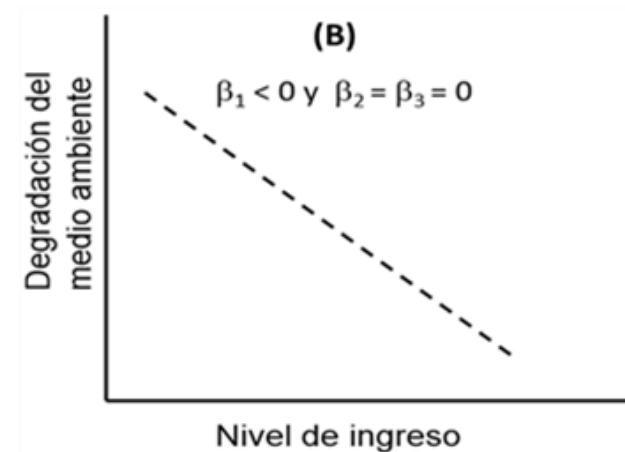
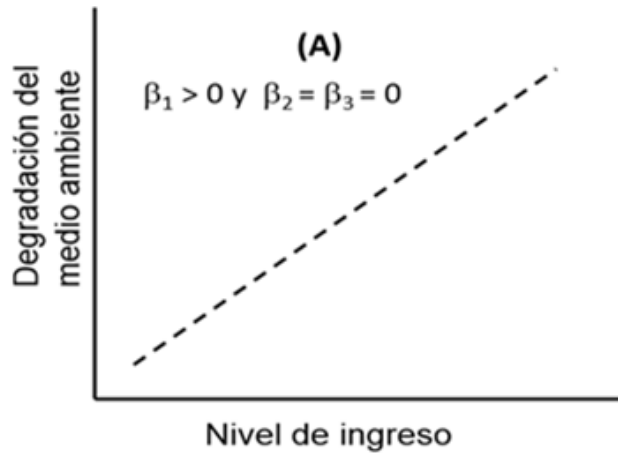
La curva ambiental de Kuznets, representada por una U invertida, es una de las hipótesis básicas para explicar la relación entre nivel de ingreso per cápita y el medio ambiente

Las causas económicas que originan este comportamiento son múltiples

Tres efectos principales: de escala, de composición y tecnológico



$$CO2_t = \beta_0 + \beta_1 YP_t + \beta_2 YP_t^2 + \beta_3 YP_t^3 + \sum_{j=1}^k \gamma_j X_{j,t} + u_t$$

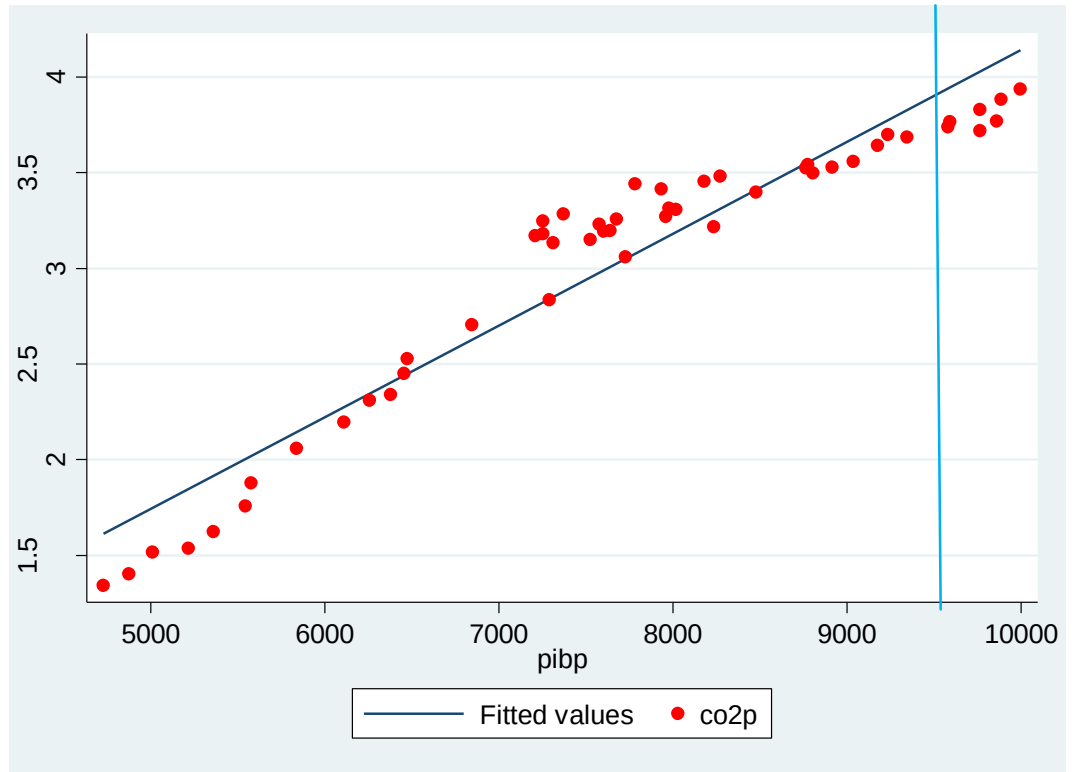
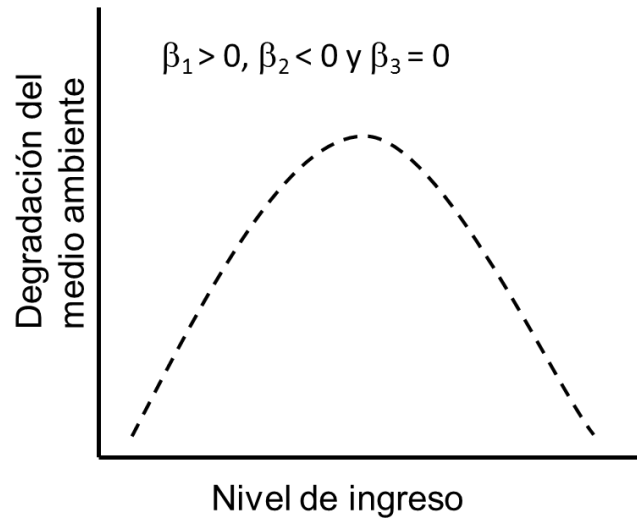


```
. reg co2p pibp pibp_2 pibp_3 engp
```

Source	SS	df	MS	Number of obs =	50
Model	26.1359516	4	6.5339879	F(4, 45) =	793.93
Residual	.370345793	45	.008229907	Prob > F =	0.0000
				R-squared =	0.9860
				Adj R-squared =	0.9848
Total	26.5062974	49	.540944845	Root MSE =	.09072

co2p	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
pibp	.0020272	.0007631	2.66	0.011	.0004903	.0035642
pibp_2	-1.91e-07	1.07e-07	-1.78	0.083	-4.07e-07	2.56e-08
pibp_3	6.10e-12	4.86e-12	1.25	0.216	-3.70e-12	1.59e-11
engp	1.20461	.2362831	5.10	0.000	.728711	1.680508
_cons	-5.694726	1.800603	-3.16	0.003	-9.321327	-2.068125

co2p	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
pibp	.0010882	.0001469	7.41	0.000	.0007926	.0013839
pibp_2	-5.64e-08	7.48e-09	-7.53	0.000	-7.14e-08	-4.13e-08
engp	1.113316	.2261746	4.92	0.000	.6580503	1.568582
_cons	-3.501951	.4303906	-8.14	0.000	-4.368282	-2.63562



CURSO: ECONOMETRÍA Y ANÁLISIS DE POLÍTICAS FISCALES

INSTRUCTOR: HORACIO CATALÁN ALONSO

